Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015

Editors

Elena Babatsouli, David Ingram

Institute of Monolingual and Bilingual Speech
Chania, Greece

Publisher: Institute of Monolingual and Bilingual Speech
Chania 73100, Greece

Symposium dates: 7-10 September 2015
Publication date: December 2015

ISBN: 978-618-82351-0-6

URL: http://ismbs.eu/publications

CONTENTS

ACKNOWLEDGMENTS

The editors wish to express their gratitude to all the participants of the International Symposium of Monolingual and Bilingual Speech 2015 and its Proceedings: plenary speakers, invited speakers, contributed paper authors and speakers, members of the organizing committee, and members of the scientific committee. Special thanks to all for bringing valuable expertise to pave the way of research on monolingual and bilingual speech into the future. Aegean Air and the Institute of Monolingual and Bilingual Speech are also gratefully acknowledged for sponsoring the early career research awards.

FOREWORD

The Proceedings contain papers from a number of talks that were given at the inaugural International Symposium on Monolingual and Bilingual Speech which took place in Chania, Greece on 7-10 September 2015. This Symposium sprang from yearning for a specialized conference on speech that cuts across dividing boundaries between language subfields: first language, second language, bilingual, multilingual; child or adult; typical or impaired. The Symposium encouraged investigations that go to the heart of matters, widening existing horizons and perspectives, kindling a holistic viewpoint, fostering collaborations across the board and, ultimately, sparking innovative thought and approaches. Participant affiliations covered forty countries in Europe, North and South America, Africa, Asia, Australia, and New Zealand. Research included thirty eight languages among which Bengali, Comorian, Farsi, Maori, and Swahili that are not as common in the literature.

SCIENTIFIC COMMITTEE

Elena Babatsouli (Chania, Greece)
Anna Balas (Poznan, Poland)
Martin J. Ball (Linkoping, Sweden)
Hans Basbøll (Odense, Denmark)
B. May Bernhardt (Vancouver, BC, Canada)
Ocke-Schwen Bohn (Aarhus, Denmark)
Ferenc Bunta (Houston, TX, USA)
M. Grazia Busa (Padova, Italy)
Juli Cebrián (Barcelona, Spain)
Laura Colantoni (Toronto, ON, Canada)
Cynthia Core (Washington, DC, USA)
Elise de Bree (Amsterdam, The Netherlands)
Katarzyna Dziubalska-Kołaczyk (Poznan, Poland)
Leah Fabiano-Smith (Tucson, AZ, USA)
Cécile Fougeron (Paris, France)
Maria João Freitas (Lisbon, Portugal)
Helen Grech (Msida, Malta)
Tetsuo Harada (Tokyo, Japan)
David Ingram (Tempe, AZ, USA)
Margaret Kehoe (Geneva, Switzerland)
Ghada Khattab (Newcastle, UK)
Jong-mi Kim (Chuncheon, Korea)
Tanja Kocjancic Antolik (Paris, France)
Kristian Emil Kristoffersen (Oslo, Norway)
Malgorzata Kul (Poznan, Poland)
Shinsook Lee (Seoul, Korea)
Juana M. Liceras (Ottawa, ON, Canada)
Conxita Lleó (Hamburg, Germany)
Andrea MacLeod (Montreal, QC, Canada)
Sharynne McLeod (Bathurst, Australia)
Konstantinos Minas (Rhodes, Greece)
Peggy Mok (Hong Kong, China)
Joan C. Mora (Barcelona, Spain)
Eleni Morfidi (Ioannina, Greece)
Nicole Müller (Linkoping, Sweden)
Elena Nicoladis (Edmonton, AB, Canada)
F. Hulya Ozcan (Eksisehir, Turkey)
Lucrecia Rallo Fabra (Mallorca, Spain)
Michelle Pascoe (Cape Town, South Africa)
Elinor Payne (Oxford, UK)
Brechtje Post (Cambridge, UK)
Eric Raimy (Madison, WI, USA)
Yvan Rose (Newfoundland, Canada)
Eirini Sanoudaki (Bangor, UK)
Ellen Simon (Ghent, Belgium)
Anna Sosa (Flagstaff, AZ, USA)
Dimitrios Sotiropoulos (Chania, Greece)
Joseph P. Stemberger (Vancouver, BC, Canada)
Isao Ueda (Osaka, Japan)
Virve-Anneli Vihman (Manchester, UK)
Magdalena Wrembel (Poznan, Poland)
Mehmet Yavaş (Miami, FL, USA)

The relative perceptual weight of two Swedish prosodic contrasts

Åsa Abelin[1], Bosse Thorén[2]
asa.abelin@ling.gu.se, bth@du.se

[1]Department of Philosophy, University of Gothenburg
[2]School of Humanities and Media Studies, Dalarna University

Abstract. In addition to 9 vowel and 18 consonant phonemes, Swedish has three prosodic phonemic contrasts: word stress, quantity and tonal word accent. There are also examples of distinctive phrase or sentence stress, where a verb can be followed by either an unstressed preposition or a stressed particle. This study focuses on word level and more specifically on word stress and tonal word accent in disyllabic words. When making curriculums for second language learners, teachers are helped by knowing which phonetic or phonological features are more or less crucial for the intelligibility of speech and there are some structural and anecdotal evidence that word stress should play a more important role for intelligibility of Swedish, than the tonal word accent. The Swedish word stress is about prominence contrasts between syllables, mainly signaled by syllable duration, while the tonal word accent is signaled mainly by pitch contour. The word stress contrast, as in *armen* [´arːmən] 'the arm' - *armén* [ar´meːn] 'the army', the first word trochaic and the second iambic, is present in all regional varieties of Swedish, and realized with roughly the same acoustic cues, while the tonal word accent, as in *anden* [´anːdən] 'the duck' - *anden* [`anːdən] 'the spirit' is absent in some dialects (as well as in singing), and also signaled with a variety of tonal patterns depending on region. The present study aims at comparing the respective perceptual weight of the two mentioned contrasts. Two lexical decision tests were carried out where in total 34 native Swedish listeners should decide whether a stimulus was a real word or a non-word. Real words of all mentioned categories were mixed with nonsense words and words that were mispronounced with opposite stress pattern or opposite tonal word accent category. The results show that distorted word stress caused more non-word judgments and more loss, than distorted word accent. Our conclusion is that intelligibility of Swedish is more sensitive to distorted word stress pattern than to distorted tonal word accent pattern. This is in compliance with the structural arguments presented above, and also with our own intuition.

Keywords: second language pronunciation, intelligibility, word stress, tonal word accent

Introduction

In the field of second language teaching, there are four main skills that normally are considered; listening comprehension, reading comprehension, oral proficiency and writing proficiency. Oral proficiency can be further divided into pragmatics, like turn-taking, fluency and pronunciation. Pronunciation can be divided into segmental – including phonotactics – and prosodic features. Finally, prosodic features can be divided into dynamic, temporal and tonal variables. This study looks particularly at the perceptual weight of temporal vs tonal prosodic features in Swedish. The result could provide some guidelines as to what phonological features could be given higher or lower priority when Swedish is taught as a second language. This paper reports an expanded version of our experiment presented at Fonetik 2015 (Abelin & Thorén, 2015)

According to Munro and Derwing (1995) a foreign accent per se decreases intelligibility to some degree, but increased perceived degree of foreign accent does not seem to reduce intelligibility. We believe however, that specific details in a foreign accent may be more crucial to intelligibility than the perceived degree of global foreign accent. For English, some 'Lingua Franca Core' features were suggested by Jenkins (2002), and for Swedish Bannert (1980) suggested that some phonological features were more crucial to intelligibility than others. Thorén (2008) discussed differentiated priority among Swedish prosodic contrasts and their respective acoustic correlates.

Standard Swedish has three prosodic phonological contrasts: stress placement, quantity and a tonal word accent. There is some structural and anecdotal evidence that word stress should play a more important role in the perception and understanding of Swedish, than tonal word accent. Henceforth we will discuss only the two latter contrasts. Although both contrasts are phonemic, some dialects like standard Finland-Swedish lack the tonal word accent contrast but are still easily understood by speakers of other regional varieties. Also, in singing the tonal word accent is totally neutralized. The aim of the study is to find out which of two distortions causes the most difficulty in identifying some disyllabic words: 1) changing the word stress category from trochaic to iambic and vice versa, or 2) changing the tonal word accent category from accent II to accent I and vice versa.

Swedish word stress is about prominence contrasts between syllables, mainly signaled by syllable duration (Fant & Kruckenberg, 1994), although F0 gestures, voice source parameters and differences in vowel quality combine to signal syllable prominence (ibid.). Tonal word accent, however, is mainly signaled by changes in the F0 curve and the timing of those changes within the word. According to Bruce (1977, 2012) and Elert (1970), word stress in Swedish is variable, and words can have different meanings depending on where the main stress is placed, as found in *banan* [`bɑːnan] 'the path/course' and *banan* [ba´nɑːn] 'banana'. A great number of disyllabic trochaic-iambic minimal pairs can be created. A smaller number of trisyllabic minimal pairs, such as *Israel* [`iːsrael] 'the state of Israel' and *israel* [ɪsra´eːl] 'Israeli citizen', are also possible.

According to standard accounts Swedish has two word accent categories: accent I (acute), as in *tomten* [´tɔmːtən] 'the plot', and accent II (grave), as in *tomten* [`tɔmːtən] 'Santa Claus' (see Elert, 1970), even though only the grave accent can be considered a real word accent. It is the only one of these two that predicts that the main stressed syllable and the following syllable belong to the same word (in a disyllable word) i.e. having a cohesive function, and it is limited to the word, simple or compound. The word accent is connected with a primary stressed syllable. Pronounced in isolation, words usually carry sentence accent and accent II then tends to involve two F0 peaks.

The purpose was thus to investigate the relative perceptual weights of the two prosodic contrasts, and the weight of the categories of each contrast. The purpose of the first experiment was to test the recognition of words with trochaic stress mispronounced with iambic stress, and words with accent II mispronounced with accent I. The purpose of the second experiment was to test the recognition of words with iambic stress mispronounced with trochaic stress, and words with accent I mispronounced with accent II.

Method

Material and design

The material for the first experiment consisted of 10 trochaic (accent I) words, e.g., *bilen* [´biːlen] 'the car', 10 originally trochaic words pronounced with iambic stress, e.g., *vägen* *[vɛˈgɛn] 'the road', 10 iambic words, e.g., *kalas* [ka´lɑːs] 'the party', 10 accent II words, e.g., *gatan* [`gɑːtan] 'the street', 10 originally Accent II words pronounced with trochaic stress and accent I, e.g., *sagan* *[´sɑːgan] 'the fairy tale', and finally 26 disyllabic non-words with varying stress or tonal accents. Furthermore, the material for the second experiment consisted of 10 trochaic (accent I) words, e.g., *köket* [´ɕøːkət] 'the kitchen', 10 originally iambic words pronounced with trochaic stress, e.g., *kanel*, *[´kaneːl] 'cinnamon', 10 iambic words, e.g., *kalas* [ka´lɑːs] 'the party', 10 accent II words, e.g., *gatan* [`gɑːtan] 'the street', 10 originally accent I words pronounced with accent II, e.g., *djuret* *[`jʉːrət] 'the animal'. The same 26 disyllabic non-words as in the first experiment were used.

All trochaic words (with one exception) were nouns in the definite form. The words were recorded by a male phonetician with a neutral dialect. Recordings were made with a Røde NT3 condenser microphone to a laptop in a silent studio in the University of Umeå, Sweden, and editing was made with the Praat software (Boersma & Weenink, 2013).

There was some deliberation about how to treat vowel quality in the stressed and unstressed syllables, since these vary according to degree of stress. We decided to choose vowels which do not vary so much in unstressed vs. stressed position, e.g., /e/ rather than /a/, and keep the quality of the original word, e.g., not changing [e] to [ɛ] or [ə] in unstressed position. Each word was presented until it selfterminated, in all cases just below 1000 ms. Simultaneously the subjects had 1000 ms to react to each stimulus. The time allotted for reaction to the stimuli thus started when the word started. Between each word there was a 1000 ms pause.

For building and running the experiment, the PsyScope software was used (Cohen, MacWhinney, Flatt, & Provost, 1993).

Procedure

Two lexical decision tests were performed. In the first experiment there were 18 female L1 speakers of Swedish, approximately 20–25 years of age, who were presented with the above described 76 words of experiment 1, one by one in random order. In the second experiment, there were 16 female L1 speakers of Swedish, approximately 20–25 years of age, who were presented with the above described 76 words of experiment 2, one by one in random order. The subjects were instructed to press one key on a keyboard if the word was a real word and another key if the word was a non-word. The subjects were instructed to decide as quickly as possible, whether the word they heard was a real word or not. Reactions that were not registered within the 1000 ms period were categorized as loss. The subjects had no reported hearing impairment.

Results

Accuracy

Figure 1 shows the main results of experiments 1 and 2. It turned out that the task was quite difficult, and that the loss in the experiment was large.

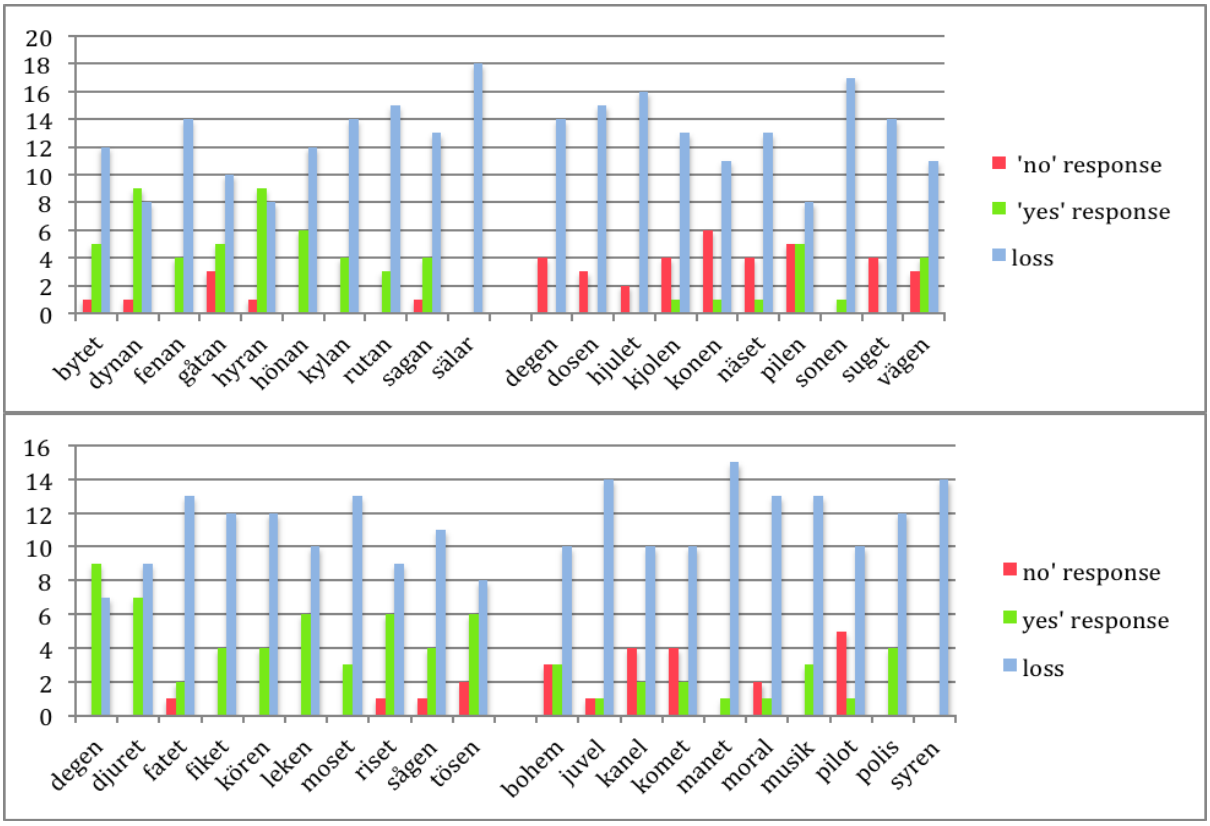


Figure 1. Main results of experiment 1 (above) and experiment 2 (below). The ten bars to the left show the effect of wrong tonal word accent, while the ten bars to the right show the effect of wrong stress placement.

It is evident from Figure 1 that wrong stress placement produced more rejections than wrong tonal word accent in both experiments.

Wrong tonal accent produced more acceptance than wrong stress placement in both experiments. An unpaired t-test showed a significant difference between the two groups ($p < .0001$). The difference in number of 'yes' responses between accent I mispronounced as accent II and accent II mispronounced as accent I is not significant. Neither is the difference between trochaic as iambic and iambic as trochaic significant.

Figure 2 shows a comparison between the wrongly pronounced words with the correctly pronounced words. The figure shows that the correctly pronounced words are, as expected, the most robust; they exhibit a smaller loss and they are more often assessed as real words. The words, which were most frequently judged as non-words were the words with wrong stress placement. The difference in number of 'yes' responses between correctly pronounced accent I words and accent I words pronounced with accent II was significant in an unpaired t-test ($p = .0233$). The difference in number of 'yes' responses between correctly pronounced accent II words and accent II words pronounced with accent I was not significant. When comparing the numbers for loss, accent II pronounced as accent I showed a larger loss than the reverse condition.

The difference in number of 'yes' responses between correctly pronounced trochaic words and trochaic words pronounced with iambic stress was significant ($p<.0001$). Likewise, the difference in number of 'yes' responses between correctly pronounced iambic words and iambic words pronounced with trochaic stress was significant ($p<.0001$).

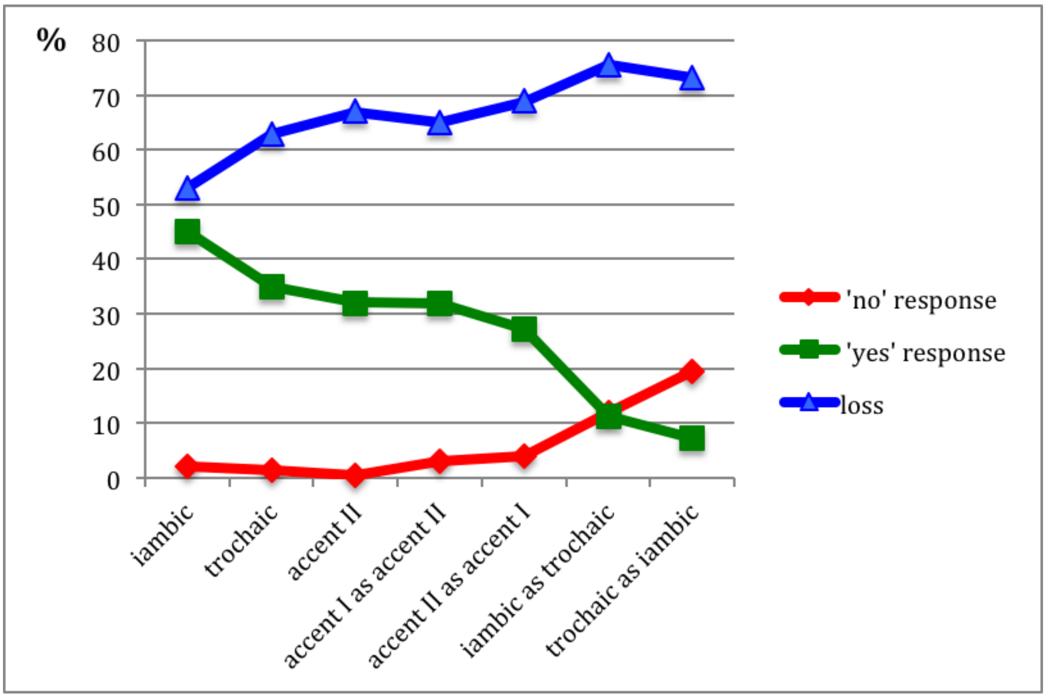


Figure 2. Results for Experiments 1 and 2, including the correctly pronounced words, shown to the left in the diagram.

There is interaction between loss, 'no' responses and 'yes' responses. There is a negative correlation between number of 'yes' responses and loss ($r^2 = .8473$). Furthermore, where there are more 'no' responses the loss is greater.

Reaction times

There was no large reaction time difference in mean total between the wrongly pronounced groups. However, to compare reaction times for the 'yes' responses is not possible since there were so few 'yes' responses for the words with wrong stress placement.

Durations of sound stimuli

The durations of the sound stimuli were measured and we found that the wrongly pronounced trochaic accent I words, pronounced as iambic, were slightly longer. However, this did not correlate with reaction times.

In general, reaction times were longer than the word durations, but not if deducting 200 ms for motor activation. There is a tendency that when durations are shorter, loss is smaller and the 'yes' responses are more numerous.

Further experiments

We are also performing an experiment with longer reaction times in order to see what happens to the variable loss. Preliminary results show that loss is diminished when longer reaction times are allowed. Reaction times for 'yes' responses are now possible to measure, and the length of reaction times partly reflects the same order as shown in diagram 2: trochaic accent I words have the shortest reaction times, thereafter iambic words and then trochaic accent II words. Furthermore, accent II mispronounced as accent I have longer reaction times than the correctly pronounced words, and the longest reaction times are the responses to iambic words pronounced as trochaic. Also here we can preliminarily conclude that wrong stress placement is more detrimental to identification of words, since wrong stress placement produces longer reactions times.

We plan to undertake further experiments for testing mispronunciations of the quantity contrast, such as the one in the minimal pair *vila* [`viːla] 'to rest' and *villa* [`vɪlːa] 'villa'. Together with the present results, the quantity data could help provide a more complete ranking among the Swedish prosodic contrasts with respect to their importance for communication and education.

Discussion

The results suggest a greater perceptual weight for stress pattern when compared with tonal word accent. Furthermore, the results can be discussed in relation to "left-to-right" models of speech perception and to where the actual recognition point is (cf. Marslen-Wilson, 1987). One question is whether an early absence of stress placement would be more detrimental for recognition than a late absence, i.e. would a stress-placement-changed trochaic word (which ought to have stress on the first syllable) be more difficult to process than a stress-placement-changed iambic word (which ought to have stress on the second syllable)? There is some evidence for this, although the difference was not significant: wrongly pronounced trochaic words were more difficult to identify than wrongly pronounced iambic words. Thus, an early absence of stress placement is more difficult to process.

The words of the present experiments were not checked for frequency or number of phonological neighbors. It could be the case that some of the iambic words (which often are loan words) have a lower frequency. On the other hand, correctly pronounced iambic words were words that had the least loss, the highest number of 'yes' responses and the lowest number of 'no' responses, which might indicate an effect of few phonological neighbors, as concerns "stress related neighbors". The reason that words were not balanced for frequency was that it was difficult to find suitable words. We made a check for possible correlations between rankings of frequencies and rankings of reaction times, and found no correlation between lower frequencies and longer reaction times. However, frequency is not a main issue since the results mainly concern correct interpretation or misinterpretation, not reaction time.

Another reflection is the following: What does it entail that the iambic (correct) words are not in the definite form? Morphology, such as different inflectional forms, can affect processing. Söderström (2012) studied perception of accent I and accent II in a mismatch condition where accent I words were followed by accent II inducing suffixes, and accent II words were followed by accent I inducing suffixes. He found that there is a stronger relation between suffixes and accent II compared with accent I, which could imply that accent II could indeed be very important to perception, identification and comprehension in certain contexts.

In relation to the studies of Söderström (2012) and Söderström, Roll, and Horne (2012), the question arises whether accent II might be more important to comprehension where there are other errors, e.g., in the speech of learners of Swedish as a second language, which might use the wrong suffixes on nouns or verbs. When adding further learner errors such as word order mistakes or wrong lexical choices, the picture becomes complicated.

There was an interaction between loss and 'yes' responses, where there was a negative correlation between number of 'yes' responses and loss. Furthermore, where there were more 'no' responses, the loss was greater. This could be due to the simple fact that 'no' responses generally have longer reaction times than 'yes' responses; thus, it could be that in some cases when a 'no' response is intended, the response time exceeds 1000 ms. But the result could also be due to an impossibility to interpret the wrongly pronounced word. This is further explored in an experiment with the possibility for longer reaction times.

We are well aware that our experiment does not show high ecological validity since it tested deliberately mispronounced words that were judged out of context. Follow-up studies will hopefully be made in more natural scenarios.

However, the present results suggest that learners of Swedish as a second language benefit more from proficiency in stress placement than in choice of word accent category or precise realization of word accent category.

This is also indicated by the fact that word accent categories are realized differently in different geographical regions, and that some varieties do not utilize the contrast at all.

Conclusion

We conclude that Swedish L1 listeners perceive and identify words with incorrect stress placement and incorrect tonal word accent with greater difficulty than words pronounced with correct stress and correct word accent. Thus, correctly pronounced words were easier to identify (they produced smaller loss, more 'yes' responses and less 'no' responses) than the wrongly pronounced words. There was a difference in the order of ease for identifying the correctly pronounced words: the easiest were the iambic words, intermediate were the trochaic accent I words, and the most difficult were accent II words.

Regarding incorrectly pronounced words, the result was that wrong stress placement produced larger loss, less 'yes' answers and more 'no' answers than wrong tonal word accent. When the 'yes' responses of mispronounced words were compared with correctly pronounced words, we saw a highly significant difference between correctly pronounced stress and mispronounced stress, for both types of stress change. There was a significant difference of amount of 'yes' answers between correct accent I and accent I as accent II, but not vice versa. This suggests that identification and comprehensibility of speech is more affected by wrong word stress placement than wrong word accent.

The study also shows that experimental methods combine well with phonetic, phonological and pedagogical issues. In further studies, we will test the perceptual weight of the third prosodic distinction of Swedish, quantity contrast, in relation to the two contrasts in the present study, and with respect to different positions in the word.

Pedagogical implication

Since the present experiment implies that the stress pattern of Swedish is more crucial for comprehensibility than tonal word accent, we suggest that second language learners of Swedish can benefit more from proficiency in perceiving and producing stress pattern. We can imagine a second language learner of Swedish going to school outside the Stockholm (capital) region. Her teacher may use a teaching material that describes the general Swedish stress patterns and also the Stockholm variety of the word accent contrast. In addition to this, the teacher may unintentionally introduce her own local accent, despite her effort to comply with the tonal patterns described in the material. Even if the teacher succeeds to mimic the Stockholm tonal patterns, the learner will probably receive diverse input of tonal word accents from society outside school and from the media, as well. This may confuse her, not allowing her to discern what ‘correct’ Swedish word accent patterns are. The results of the present study suggest that the learner in the hypothetical situation, who is very likely to represent actual learners, can reduce her confusion and acquire appropriate pronunciations successfully, when the focus of teaching and learning lies on stress placement rather than on tonal word accents.

References

Abelin, Å., & Thorén, B. (2015). What affects recognition most – wrong word stress or wrong word accent? In *Proceedings from Fonetik 2015*, Institutionen för lingvistik, Lunds universitet.

Bannert, R. (1980). Svårigheter med svenskt uttal: Inventering och prioritering. In *Praktisk lingvistik nr 5.* Institutionen för lingvistik, Lunds Universitet.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (http://www.praat.org).

Bruce, G. (1977). *Swedish word accents in sentence perspective* (Vol. 12). Lund University.

Bruce, G. (2012). *Allmän och svensk prosodi*, Studentlitteratur, Lund.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257-271.

Elert, C.-C. (1970). *Ljud och ord i svenskan,* Stockholm: Almqvist & Wiksell.

Fant, G., & Kruckenberg, A. (1994). Notes on stress and word accent in Swedish *STLQPSR*. 2-3/1994

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, *23*(1), 83-103.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71-102.

Munro, M. J., & Derwing, T. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73-97.

Söderström, P. (2012). *Processing Swedish word accents - evidence from response and reaction times*. MA thesis in General linguistics, Lund University.

Söderström, P., Roll, M., & Horne, M. (2012). Processing morphologically conditioned word accents. *Mental Lexicon*, *7*, 77-89.

Thorén, B. (2008). *The priority of temporal aspects in L2-Swedish prosody: Studies in perception and production.* PhD thesis, Stockholm University.

A perspective into noun-before-verb bias: Evidence from Turkish-Dutch speaking bilingual children

Feyza N. Altınkamış[1], F. Hülya Özcan[2], Steven Gillis[3],
feyza.altinkamis@ugent.be, fozcan@anadolu.edu.tr, steven.gillis@ua.ac.be

[1]Ghent University [2]Anadolu University [3]Antwerp University

Abstract: Nouns and verbs are considered as fundamental categories of lexical development from both linguistic and cognitive perspectives (Kauschke et al., 2007). From a linguistic point of view, nouns and verbs are the lexical units which categorically highlight language-general and language-specific characteristics. Cognitive representations of nouns and verbs are also significant to consider in terms of acquisition of early lexicon. The aim of this research is, therefore, to investigate the Turkish-Flemish bilingual children's early language period especially in terms of two syntactic categories; namely, nouns and verbs. Besides, the differences of typological characteristics between Turkish and Flemish are striking in terms of nouns and verbs. In addition to the linguistic and typological motivations, methodologically, this research aims to utilize a very fruitful data collection tool, Communicative Development Inventory (CDI), which has become an acceptable tool to use in bilingual language acquisition studies (David & Wei, 2003; Xuan & Dollaghan, 2012; De Houwer et al., 2006; Marchman & Martinez-Sussman, 2002). In line with this background, this study addresses the nature of Turkish-Flemish bilingual children's early lexicon with respect to the noun bias phenomenon by means of data collected from 19 Turkish-Danish bilingual children living in Flanders. The results of this study, of which data analysis is still in progress, were evaluated regarding the early trajectories of bilingual children's lexical acquisition with respect to noun-before-verb pattern. Findings of this study have shown that both language-general and language-specific characteristics operate on acquisition.

Keywords: lexicon acquisition, noun bias, noun dominance, noun-before-verb pattern

Introduction

Early lexical development is characterized by the acquisition of nouns and verbs. The acquisition of these two fundamental word categories is accepted cross linguistically. In spite of this mutual agreement on word categories, there have been different views on the acquisition order of these categories. One claim is that early lexical development is characterized by nouns and, therefore, nouns precede verbs in acquisition. That is, children's first words are nouns. This view brings up the cognitive aspects, like conceptual readiness, focusing on the availability of perceptual-cognitive information and claims that children rely on perceptual categories to produce language. Gentner (1981, 1982, 2006) claimed that children acquire nouns before verbs because nouns have perceptual-cognitive dominance. That is nouns, especially concrete nouns, are "entities that can be individuated on the basis of perceptual dominance", and can be inferred cognitively with minimal linguistic experience (Gentner, 1982; Gentner & Boroditsky, 2001, p. 215). However, verbs follow linguistic dominance because they do not exist in the environment on their own, independent of language (Gentner, 1982, 2001). The linguistic distinction between nouns and verbs is based on "the pre-existing perceptual-conceptual distinction between concrete concepts, namely nouns and predicative concepts of activity, namely verbs" (Gentner, 1982, p. 324). Within this framework, Gentner (1982) proposed two strictly interwoven hypotheses; the natural partitions hypothesis and the relational relativity hypothesis. The natural partitions hypothesis claims that nouns are acquired early because the referents are readily available in the environment; the relational relativity hypothesis, on the other hand, claims that verb meanings do not naturally emerge from the structure of the word but by hearing the verbs in use.

Gentner's view on noun dominance has both been confirmed and challenged in a number of studies. Crosslinguistic evidence has been provided by studies on different languages such as Italian, Korean, Hebrew, French, Spanish, Dutch, as well as in English (Bornstein, Cote, Maital, Painter, Park, &

Pascual, 2004; Goldfield, 1993; Caselli, Bates, Casadio, Fenson, Fenson, Sanderl, & Weir, 1995; Dromi, 1987; Maital, Dromi, Sagi, & Bornstein, 2000; Jackson-Maldonado, Thal, Marchman, Bates, & Guitierrez-Clellen, 1993; Gillis & Verlinden, 1988).

Studies challenging noun dominance can be classified into two groups. One group of studies questions the universal nature of noun dominance. Tardif (1996), Gopnik and Choi (1995), Gopnik, Choi, and Baumberger (1996) stated that children use more verbs than nouns in Mandarin and Korean. On the other hand, Tardif, Shatz, and Naigles (1997) and Kauschke, Hae-Wook, and Soyeong (2007) stated that there are other reasons underlying noun dominance, such as word order of the language, child directed speech and interactional requirements. These challenges have led the argument to the view that language-specific characteristics have been neglected so far.

As the target languages of this research are Turkish and Dutch, first, we would like to raise attention on the results of studies on Turkish and Dutch monolingual children. The “nouns-before-verbs” pattern in acquisition has been handled in longitudinal, contextual and crosslinguistic studies in Turkish and Dutch. Türkay (2005) observed five Turkish speaking children and their mothers longitudinally and found out that the children in her study used nouns and verbs in equal measure, showing no privileged use of any category over the other. Kern and Türkay (2006) compared Turkish and French speaking children’s longitudinal data and found out that both groups of children used more nouns than verbs before and after the vocabulary burst period (around 20 months) but the gap between noun and verb categories in Turkish was always lower than the difference in French. Bornstein et al. (2004) made a cross linguistic analysis of vocabulary in young children and mentioned the noun primacy over verbs in Dutch speaking children’s lexical growth. Gillis (1984) observed a Flemish boy between the ages of 0;11 and 1;11 and found that the child's early lexicon was predominantly made up of nouns.

A rapid accumulation of studies with a focus on the acquisition of nouns and verbs has led to follow-up studies on bilingual children. Research in this domain with bilingual children presents a good arena to understand the interaction of language-general and language-specific processes in the early lexical development. In line with this objective, Xuan and Dollaghan (2012) conducted a study with 50 English-Mandarin bilingual children. The parents were supposed to report their children's lexicons using the English and Mandarin version of CDI. They found that the mean percentage of Mandarin nouns (38%) was significantly lower than the percentage of English nouns (54 %). In addition, the researchers examined the characteristics of the top 50 words and analysed these words to see if these early acquired words fitted into the four distinctive features called as SICI continuum (Maguire et al., 2006). The SICI continuum is based on four features: distinctive shape (S), easy individuation (I), concreteness (C), and imageability (I). In the related studies, these features have been argued to judge the typicality of objects. Xuan and Dollaghan (2012) concluded that the words in the top 50 list did not completely match the SICI features. Then, they inferred that only perceptual-cognitive factors were not satisfactory enough to explain the nature of the bilingual children's word learning. Following them, Lucas and Bernardo (2010) carried out a research with 60 Filipino-English bilingual children. Different from Xuan and Dollaghan, they audio-recorded the child-mother interactions and coded the data to see the frequency of nouns and verbs in the child's and the mother's utterances. They found out that Filipino-English bilingual children showed a noun bias in their early vocabularies but this noun dominance was in their English lexicon, not in their Filipino lexicon. They further added that this noun over verb dominance observed in English monolingual children is also available in a bilingual context. In line with this background, we aim here to investigate the noun-before-verb pattern in Turkish-Dutch bilingual children. There are two main reasons to focus on these two languages. Firstly, Turkish and Dutch are languages with very different language characteristics (see Table 1). Secondly, Turkish and Dutch are languages lying at the heart of this debate in the related literature. The research questions of this study are as follows:

Research Questions

1- What is the nature of Turkish-Flemish bilingual children’s early lexicon with respect to the noun bias phenomenon?
2- Do Turkish-Flemish bilingual children produce more nouns than verbs in Turkish?

3- Do Turkish-Flemish bilingual children produce more nouns than verbs in Dutch?

Table 1. Language-specific characteristics of Turkish and Dutch

	Turkish	Dutch
Language family	Ural-Altaic/Altaic	Indo-European/Germanic
Morphology	Agglutinated	Inflected
Word order	The canonical word order is SOV but it is very flexible. Five different word orders (OSV, SVO, OVS, VSO, VOS) are quite common in Turkish speakers' talks for pragmatic preferences.	The main clause word order is SVO but subordinate clause word order is SOV.
Noun morphology	Nouns are inflected for number, case and possession.	Nouns are inflected for number.
Verb morphology	Verbs are marked for person, number, tense, aspect, modality, voice, negation and interrogation	Verbs are marked for number and tense.
Subject drop	Subject is not obligatory.	Subject is obligatory
Noun-friendliness/verb-friendliness	Similar structural properties with languages known as verb-friendly such as Korean and Japanese	Similar structural properties with languages known as noun-friendly such as English, German and French

Methodology

Participants

Selection procedures: The study was advertised in the Turkish community in Flanders through social networks and personal announcements. Turkish-Flemish bilingual families whose children were in the target age group of the study were invited to contact the researcher. After their contact with the researcher, the families were visited at home and given more detailed information about the study. After the first encounter, the data was collected.

Characteristics:

Table 2. Sample Characteristics (N=19)

	CDI-I	*CDI-II*
Sex		
Female	*6*	*5*
Male	*3*	*5*
Birth Order		
First born	*2*	*3*
Later born	*7*	*7*
Primary caregiver		
Mother at home	*50 %*	*40%*
Daycare or non-parent	*50%*	*60%*

Mean (SD) age in months Mean (SD) daily exposure to Turkish:	*11.86 (2.39)* *60(%) (23%)*	*27.2 (7.74)* *56% (21%)*
Mean (SD) daily exposure to Dutch:	*40(%) (23%)*	*44% (21%)*

Data Collection Procedures

Instrument: Following the suggestions of Marchman and Martinez-Sussman (2002), Marchman, Kuan, Yoshuda, and Xuan (2005), and Xuan and Dollaghan (2012) about the productive use of counterparts of the CDI with bilingual populations, the Turkish and Dutch adaptations of CDI were used to measure the participant children's lexicon. There are two parts in the CDI: one for children aged between 8 and 16 months, and one for children aged between 16 and 36 months. In both parts, only the vocabulary section was considered in this study.

Vocabulary measures: To be consistent with the research findings in our reference study (Xuan & Dollaghan, 2012) in the noun category, we considered 'animals', 'vehicles', 'toys', 'food and drink', 'clothing', 'body parts', 'small household items, and 'furniture', while in the verb category we only took into account 'action words'. Before starting our analysis, we calculated the noun and verb percentage in the Turkish and Dutch CDI, following the parallel coding in our study (Table 3). As seen here, verbs are represented more in the Turkish CDI-I (22%) than they are in the Dutch CDI-I (13%) whereas Dutch nouns are more than Turkish nouns.

Table 3. Description of categories in the original version of CDI-I (T-CDI and D-CDI)

Language	Total Lexicon	Nouns	%	Verbs	%
Turkish-CDI-I	418	159	38	95	22
Dutch-CDI-I	434	213	49	57	13

Table 4 shows the noun and verb percentage in the Turkish and Dutch CDI-II. The pattern observed here in terms of noun-verb balance is similar to the CDI-I, verbs superiority over nouns in Turkish (20.5% vs. 15%) and noun superiority over verbs in Dutch (36.5% vs. 42.4%).

Table 4. Description of categories in the original version of CDI-II (T-CDI-II and D-CDI-II)

Language	Total Lexicon	Nouns	%	Verbs	%
Turkish-CDI-II	711	260	36.5	146	20.5
Dutch-CDI-II	702	298	42.4	106	15

Data Collection

Data collection took place in the children's home. As stated, the families were given detailed information about the study and some necessary tips about how to fill in the inventory given. Then, parents were instructed to fill in the screening questionnaire, which was about their child's age, sex, past and present medical status, birth order and primary caregiver. Then, a brief language exposure form was given to reveal the child's bilingual language exposure. After these steps, parents, in most cases mothers were instructed to complete the Turkish and Dutch version of CDI in a random order.

When needed, they communicated with the father or other family members about words they were not sure of their children's use.

Data Analysis

First of all, we have analyzed the frequency of words and nouns. Mean values and median of nouns and verbs were calculated for each age group. Calculating the median (the average number) for each data group is necessary when the data shows a wide range of scores. The median shows the midpoint, that is, the average number of the entity being analyzed. In order to reveal any significant differences among the occurrences of nouns and verbs in each age group, we used t-test for paired samples, a non-parametric statistical test which is used to calculate statistically significant differences between two groups of data. Likewise, in order to reveal any possible statistically significant difference between age groups, t-test for independent samples was used. When there is no normal distribution of the data, nonparametric tests, Mann-Whitney U for independent samples and Wilcoxon for paired samples, were used to calculate the difference between two sets of data. Statistical analysis was conducted through SPSS 18.

Results

We first looked at the total number of words, nouns and verbs in Turkish and Dutch in two age groups cumulatively: between 08-16 (CDI-I) and 16-36 (CDI-II) months.

Table 5. Mean (M), standard deviation (SD) and range for total number of words, total number of nouns, and total number of verbs in Turkish and Dutch in both age groups

	TURKISH					DUTCH				
	M	SD	Median	Min-Max	RANGE	M	SD	Median	Min-Max	RANGE
TOTAL WORDS	76.4	103.4	33	0-382	382	66.9	102.1	12	0-358	358
TOTAL NOUNS	50.8	66.9	21	0-245	245	53.4	79.4	12	0-245	279
TOTAL VERBS	25.6	37.7	12	0-137	137	13.4	23.1	1	0-79	79

The vocabulary development range is wide in both languages. Total Turkish vocabulary of our children changes from 0 to 382 with a mean value 76.4. The median 33 is showing that the average number of words is 33. That is, 9 children have a vocabulary of below 33 words and 9 children have a vocabulary above it. Total Dutch vocabulary ranges from 0 to 358 with the mean value of 66.9. The median is 12, meaning that 9 children have a vocabulary below 12 words and 9 children have a vocabulary above 12 words. In Dutch, the midpoint is 12 words. There is no statistically significant difference between Turkish and Dutch in terms of the total number of words ($t=-0.337$; $df= 18$; $p=0.793$).

Total nouns in Turkish range from 0 to 245 words with a median of 21, showing that 9 children have fewer than 21 nouns in their Turkish vocabulary, and 9 children have more than 21 nouns in their vocabulary. The mean number of nouns is 50,8. Nouns in Dutch also fall within a wide range, from 0 to 279. The mean value is 53,4. Although the mean number of nouns in Dutch is higher than the mean number of nouns in Turkish, the median in Dutch is less than the median in Turkish (12 and 21, respectively) indicating that the average number of nouns is 12 words, and 9 children have fewer than 12 words in their vocabulary, while the other 9 children have more than 12 words in their vocabulary. There is no statistically significant difference between the total number of nouns in Turkish and in Dutch ($t=-0.308$; $df=1$; $p=0.761$).

Total verbs in Turkish range from 0 to 137 with the mean value of 25.6. The median is 12 indicating that the midpoint is 12, and verbs in the vocabulary of 9 children fall below this midpoint while verbs in the vocabulary of 9 children fall above this midpoint. The mean value of verbs is 25.6. Total verbs in Dutch range from 0 to 79 and the mean number of verbs is 13. The median is 1 meaning that 9 children do not have any verbs in their Dutch vocabulary. There is a statistically significant difference between Turkish verbs and Dutch verbs (t=2.662; df=18; p=0.01).

The distribution of total number of words, verbs and nouns, in Dutch is illustrated in Figure 1.

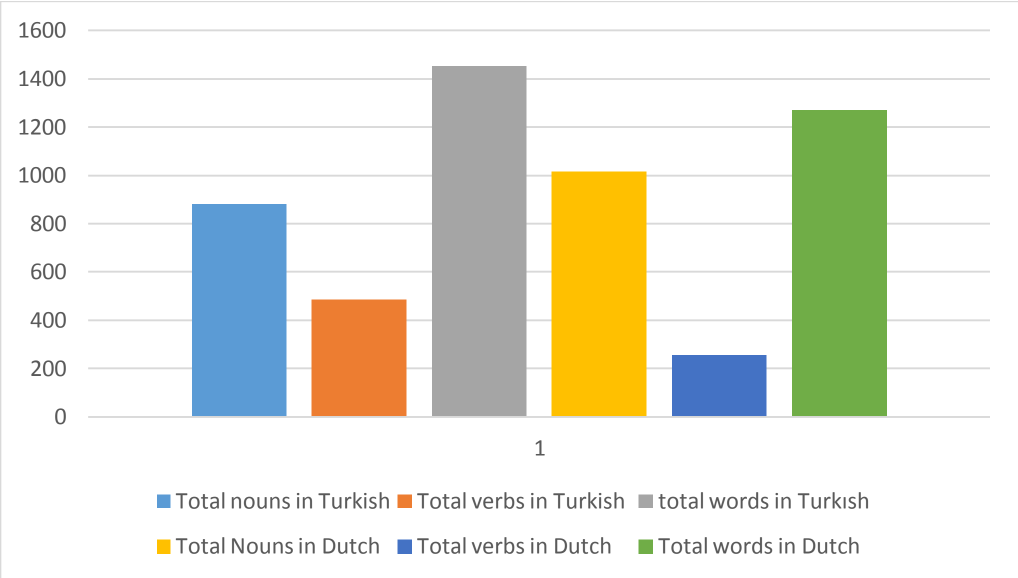


Figure 1. Distribution of total words, nouns, verbs in Turkish and in Dutch

We, then, compared the data from two perspectives; in terms of age and in terms of comprehension. Table 6 illustrates the mean, standard deviation and total number of words, nouns and verbs both in Dutch and in Turkish in CDI-I.

Table 6. Comprehension data between 08-16 months in Turkish and Dutch (Turkish CDI-I and Dutch CDI-I)

	TURKISH					DUTCH				
	M	SD	Median	Min-Max	Range	M	SD	Median	Min-Max	RANGE
TOTAL WORDS	31.6	34.2	12	2-91	89	9.4	10.1	7	0-29	**29**
TOTAL NOUNS	20.7	20.7	10	2-57	55	7.4	7.6	7	0-19	**19**
TOTAL VERBS	10.8	15.7	4	0-46	46	2.0	3.2	1	0-10	**10**

Comprehension data in Turkish show a profile as follows: The total number of words comprehended by children between 08-16 ranges from 2 words to 91 words with a mean of 31.6. The midpoint (median) is 12 indicating that 4 children can comprehend fewer than 12 words while 4 children can comprehend more than 12 words. Total number of nouns ranges from 2 to 57 with the mean number of 20.7. The median is 10, that is, 4 children comprehend fewer than 10 nouns, and 4 children comprehend more than 10 nouns. The mean number of verbs is 10.8 and the number of verbs comprehended range from 0 to 46. The median is 4, meaning that 4 children comprehended fewer than 4 verbs and others comprehended more than 4 verbs. When we look at the difference between nouns and verbs comprehended, we see that there is a statistically significant difference (z=-0.35;

p=0.04).

Comprehension data in Dutch shows a profile as follows: The total number of Dutch words comprehended by the 08-16 age group children ranges from 0 to 29 and the mean number is 9.4. The median is 7 indicating the midpoint of the number of words. 4 children comprehended fewer than 7 words and 4 children comprehended more than 7 words. The mean number of nouns is 7.4 and the number of total nouns comprehended ranges from 0 to 19. The median is 7 meaning that the average number comprehended is 7 nouns and 4 children comprehended fewer than seven, while the other 4 comprehended more than 7 nouns. The mean number of verbs comprehended in Dutch is 2.0 and the total number of verbs ranges from 0 to 10. The average number (median) of comprehended verbs is 1 indicating that 4 children comprehended no verbs in Dutch at all. There is a statistically significant difference between nouns and verbs comprehended in Dutch (t=-2.810; df=8; p=0.023).

The distribution of comprehension data in Turkish and Dutch by 08-16 months is illustrated in Figure 2.

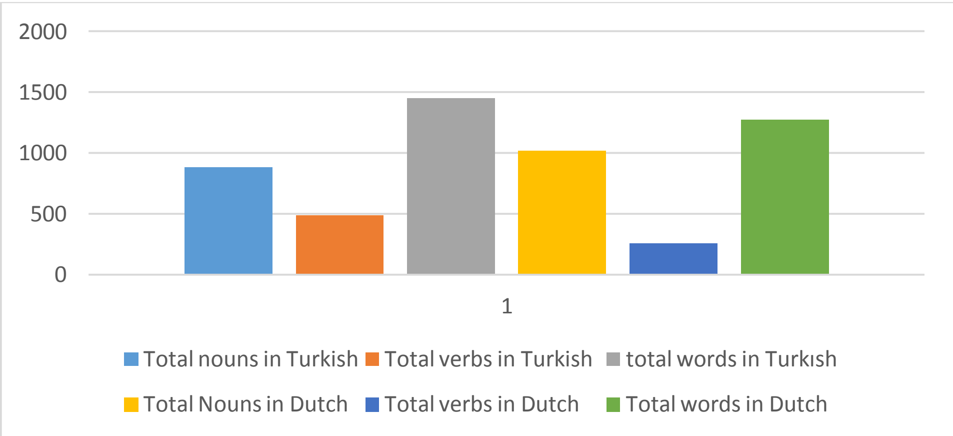


Figure 2. The distribution of comprehended nouns and verb in Turkish and Dutch in 08-16 months.

When we look at the comprehension data in terms of difference across languages, we see that there is not a statistically significant difference in total words comprehended between Dutch and Turkish (t=1.813; df= 16; p=0.08); neither in nouns (t=1.813; df=16; p=0.08); nor in verbs (t=1.657; df=16; p=0.117).

Table 7. Production data between 08-16 months in Turkish and Dutch (Turkish CDI-I and Dutch CDI-I)

	TURKISH					DUTCH				
	M	SD	Median	Min-Max	RANGE	M	SD	Median	Min-Max	RANGE
TOTAL WORDS	3.2	5.0	0.0	0-12	12	5.8	17.6	0.0	0-53	**53**
TOTAL NOUNS	1.6	2.5	0.0	0-6	6	4.7	14.3	0.0	0-43	**43**
TOTAL VERBS	2.0	4.0	0.0	0-12	12	1.1	3.3	0.0	0-10	**10**

The production data profile of children within the age span of 08-16 months in Turkish is as follows: The mean number of total words produced is 3.2. The number of total words produced changes from 0 to 12. The median is 0 because 6 out of 9 children in the group did not produce any word at all at the time of data collection. The mean number for produced nouns is 1.6 and the median is again 0 because 7 of the children were not able to produce any nouns at the time of data collection. The mean number for verbs is 2.0 and the average number of words (median) is 0 because of children who did not produce any verb. One child was able to produce 12 verbs and the other two did 1 and 5 verbs, respectively. As seen in Table 7, children produced more nouns than verbs in Turkish (10 nouns and 18 verbs). There is no statistically significant difference between total nouns and total verbs in Turkish (t=-0.529; df=8; p=0.611).

The production data profile of children within the age span of 08-16 months in Dutch is as follows: The mean number of total produced words is 5.8 and the total number ranges from 0 to 53. The average number of the total produced words (median) is 0 because there is only one child who was able to produce 53 words in total and the others did not produce any words at all at the time of data collection. In terms of the production of nouns, there is only one child who was reported to produce 43 nouns at the time of data collection in Dutch. The other children did not produce any nouns in Dutch. In terms of verbs, again there is only one child who was able to produce verbs in Dutch (N=10). There is no significant difference between nouns and verbs in Dutch (t=1.000; df=8; p=0.347).

Figure 3 illustrates the distribution of nouns and verbs produced across languages.

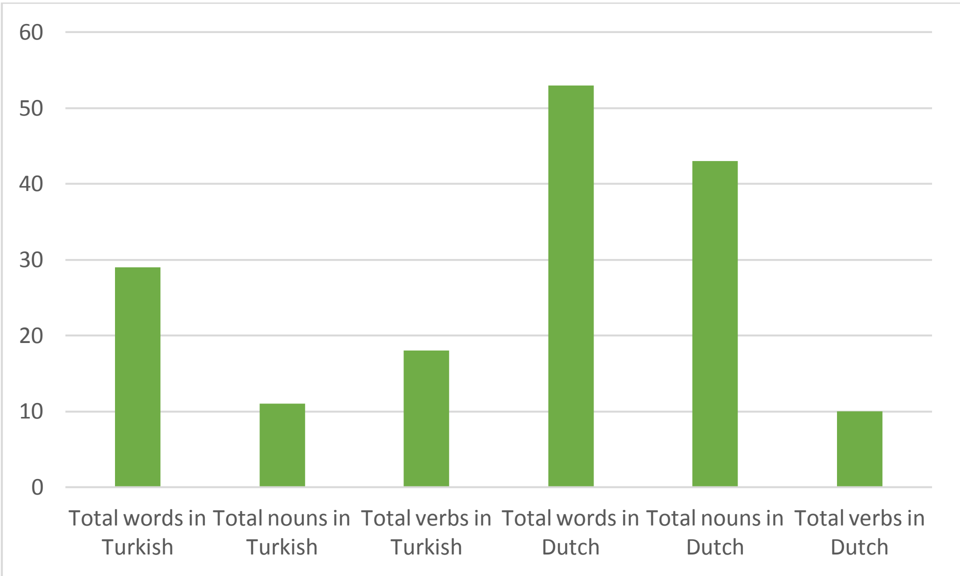


Figure 3. The distribution of produced nouns and verb in Turkish and Dutch in 08-16 months.

Then, we look at production differences across languages. There is no significant difference in terms of the total number of words produced between Turkish and Dutch (t=-0.435; df=16; p=0.669); neither in terms of nouns (t=-0.734; df=16; p=0.474), nor in terms of verbs (t=0.505; df=16; p=0.620).

Table 7 illustrates the production data between 16-36 months in Turkish and Dutch (Turkish CDI-II and Dutch CDI-II).

Table 7. Production data between 16-36 months in Turkish and Dutch (Turkish CDI-II and Dutch CDI-II)

	TURKISH					DUTCH				
	M	SD	Median	Min-Max	RANGE	M	SD	Median	Min-Max	RANGE
TOTAL WORDS	121.4	132.6	86.5	0-382	**382**	113.7	123	96.5	0-358	**358**
TOTAL NOUNS	76.9	83.4	46.5	0-245	**245**	90.6	95.2	81	0-279	**279**
TOTAL VERBS	44.5	52.6	23	0-137	**137**	23.1	28.8	15	0-79	**79**

The production profile of children aged between 16-36 months in Turkish is as follows: The total number of words produced by this age group ranges between 0 and 382. The mean number is 121.4 and the average number of words (the median) is 86.5; that is, there are 5 children that produced fewer verbs than 86.5 and 5 children that produced more verbs than 86.5. The number of nouns changes between 0 to 245, the mean is 76.9 and the median is 46.5; this is the midpoint indicating that 5 children produced fewer verbs, and 5 children produced more verbs than that number. The total number of verbs ranges from 0 to 137 with the mean number of 44.5. There is statistically significant difference between nouns and verbs in Turkish (t=2.375; df=9; p=0.04).

The production of Dutch words draws a similar profile in the sense of more nouns and fewer verbs. The total number of words produced in Dutch ranges from 0 to 358 with the mean of 113.7. The median 96,5 indicates that the average number of words produced is relatively high. The mean number of nouns is 90.6 and the average number (median) is 81. The verbs in Dutch are relatively fewer than nouns. The mean number is 23,1 and the median is 15. The average point is relatively lower than nouns in Dutch. The difference between nouns and the verbs produced in Dutch is statistically significant (t=3.159; df=9; p=-0.012).

Figure 4 illustrates the distribution of production data in Turkish and Dutch in 16-36 months (Turkish CDI-II and Dutch CDI-II).

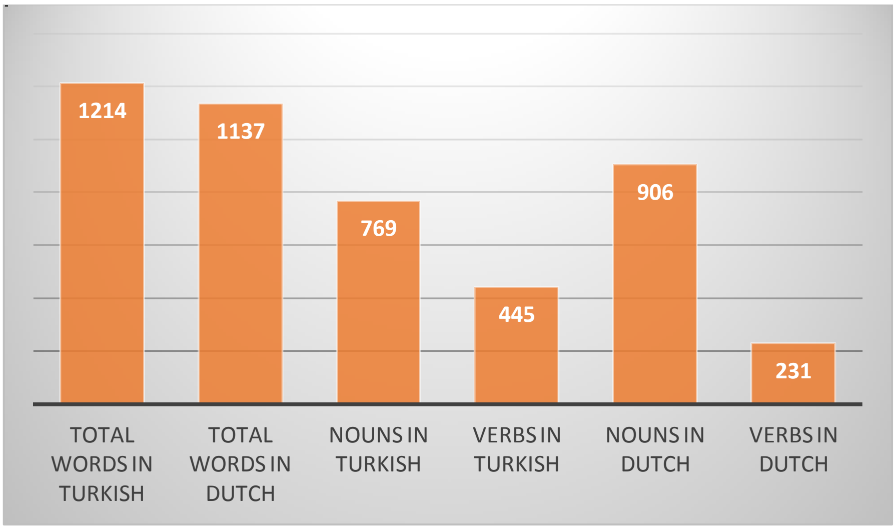


Figure 4. The distribution of produced nouns and verb in Turkish and Dutch in 16-36 months.

We, then, apply statistical test to see whether there is a statistically significant difference across languages. There is no statistically significant difference between the total number of produced words in Turkish and in Dutch (t=0.024; df=16; p=0.981). Likewise, noun production does not show a significant difference between Turkish and Dutch (t=-0.466; df=16; p=0.647). Although there are fewer verbs produced in Dutch, the difference is not statistically significant (t=1.030; df=16; p=0.318).

Next, we calculated the percentage of nouns and verbs in the children's total lexicons. Figures 5 and 6 display percentages of nouns and verbs as a function of total vocabulary size between ages 08-16. As seen in the Figure 5 for both languages, nouns occupy a bigger place than verbs, but in Dutch the percentage of nouns is higher while the percentage of verbs is lower compared to Turkish.

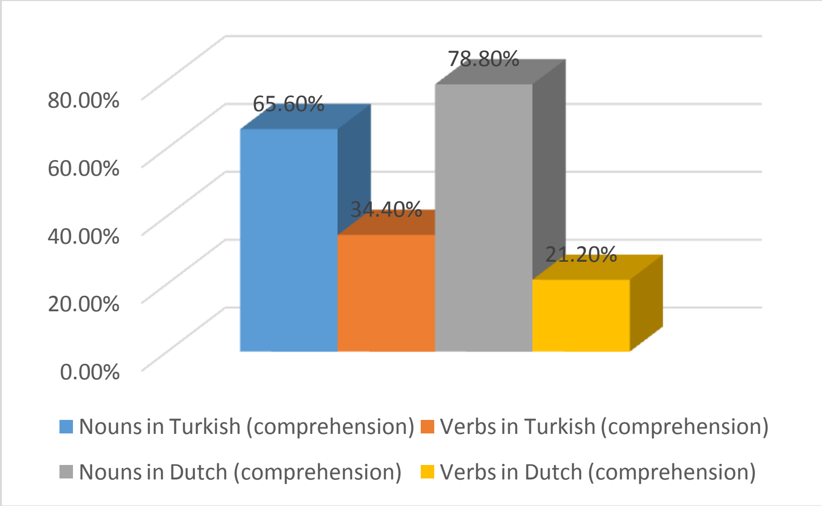


Figure 5. Frequency of nouns/verbs in the Turkish and Dutch CDI-I (comprehension)

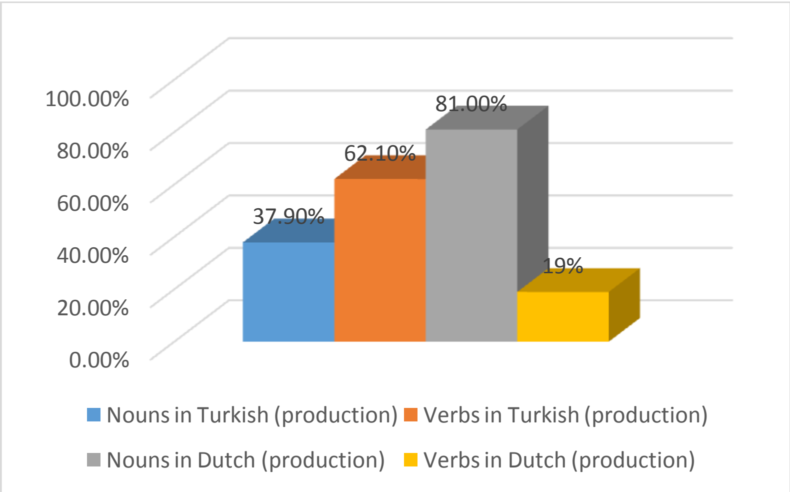


Figure 6. Frequency of nouns/verbs in the Turkish and Dutch CDI-I (production)

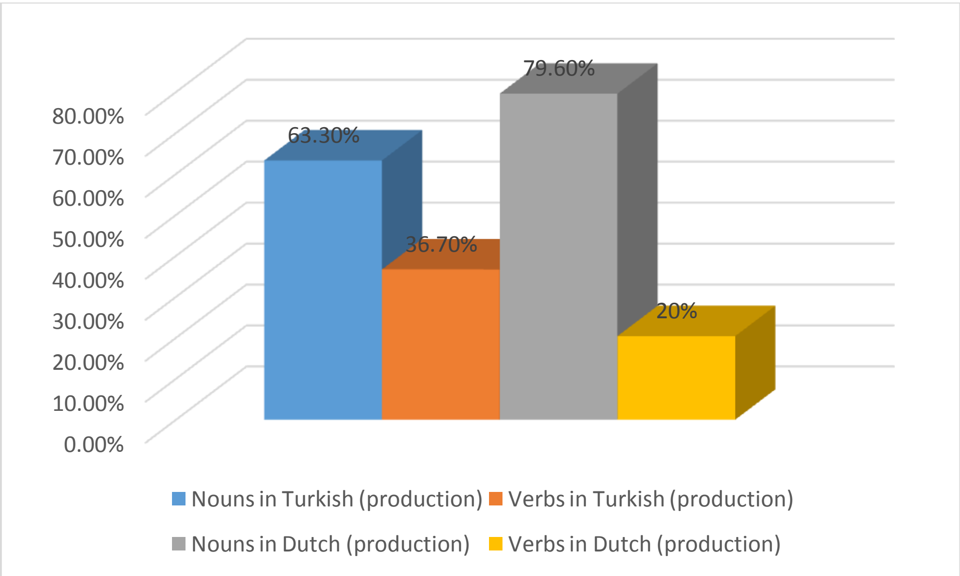


Figure 7. Frequency of nouns/verbs in the Turkish and Dutch CDI-II (production)

The production data displays a different picture. As illustrated in Figure 6, in Turkish, the percentage of verbs produced is higher than the nouns. That is to say that children produce more verbs than nouns in Turkish. However, in Dutch, more nouns but fewer verbs are produced.

The production data from CDI-II with older age group again indicates that total productive lexicon both in Turkish and Dutch consists of verbs to a great extent. There still is a different tendency between Turkish and Dutch. The frequency of verbs produced in Turkish is higher than the verbs produced in Dutch (36.7% and 20.3% respectively).

Discussion and Conclusion

We conducted this study to determine whether a noun bias would be found in Turkish-Dutch bilingual children's early lexicons. The general pattern we observed with Turkish-Dutch bilingual children in this study in terms of noun-verb dominance looks similar to the patterns observed in Turkish- and Dutch-speaking monolingual children's noun and verb development. As the results are preliminary, it may be too strong at this point to conclude that language-general mechanisms play a more crucial role in bilingual children's language development. Though this is the case, the interaction between language-general and language-specific characteristics, or in other words, cognitivism versus linguistic relativity, is remarkable in the results. The noun superiority over verbs in the children's Turkish and Flemish lexicons in both age groups (except productive vocabulary in the early age group, CDI-I), may be considered as a signal for a cognition-based approach into children's language trajectory but the difference in the noun-verb interplay in the children's Turkish and Dutch vocabularies may be attributable to language-specific processes employed to explain the nature of bilingual children's language growth. Similarly to studies on monolingual children, the preliminary results of this study should be supported by longitudinal and naturalistic data to reveal a complete picture of noun bias in Turkish-Dutch bilingual children's early language development. Another important point to mention in this study is individual variation. The high values in the standard deviations in the analysis indicate the need for a closer inspection of inter-individual differences. Various factors have been listed in the related studies that may be influential on young children's language development such as psychological, educational, social, and cultural-political factors (Verhoeven, 1999). In terms of psychological factors in the studies, there is a well-accepted approach that the mental storage of two languages is largely separated but at some points there are shared parts, especially on the level of general knowledge and skills (Leseman, 2010), so in following Leseman's perspective, we may propose that the children's Turkish and Flemish early lexicons are separate in terms of the gap between noun and verb categories but shared in terms of the noun dominance over verbs.

Acknowledgements: This research is being supported by an international cooperation project between FWO (Belgian) and TUBITAK (Turkey), 2014-2016. Many special thanks to: Prof. Dr. Ayhan Aksu-Koç for providing Turkish CDI-I and CDI-II and her invaluable help at every stage of the project, Prof. Steven Gillis for providing Dutch CDI-I and CDI-II, and Aysu Mutlutürk for providing help on how to use Turkish CDIs.

References

Bornstein, M. H.,Cote, L.R., Maital, S., Painter, K., Park, S.-Y., & Pascual, L. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75*, 115-1139.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*, 159-199.

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language*, *22*, 497-529.

David, A., & Wei, L. (2003). The composition of the bilingual lexicon. In J. Cohen, K. T. McAlister, K. Rolstad & J. MacSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 594-607).

Somerville, MA: Cascadilla Press.
Dromi, E. (1987). *Early lexical development*. Cambridge, UK: Cambridge University Press.
Gentner, D. (1981). Some interesting differences between nouns and verbs. *Cognition and Brain Theory*, *4*, 161-178.
Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (ed.), *Language development: Vol. 2. Language, thought, and culture* (pp. 301-333). Hillsdale, NJ: Erlbaum.
Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek & R. Golinkoff (eds.), *Action meets words: How children learn verbs* (pp. 544-564). Oxford, UK: Oxford University Press.
Gentner, D., & Borodistky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (eds.), *Language acquisition and conceptual development* (pp. 215-256). Cambridge, UK: Cambridge University Press.
Gillis, S. (1984). *De verwerving van talige referentie.* Unpublished Ph.D. Antwerp. Antwerp University .
Gillis, S., & Verlinden, A. (1988). *Nouns and verbs in early lexical development: Effects of input frequency*? Antwerp: University of Antwerp.
Goldfield, B. (1993). Noun bias in maternal speech to one-year-olds. *Journal of Child Language*, *17*, 171-183.
Gopnik, A., & Choi, S. (1995). Names, relational words, and cognitive development in English and Korean speakers: Nouns are not always learned before verbs. In M. Tomasello & W. E. Merriman (eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 83-90). Hillsdale, NJ: Erlbaum.
Gopnik, A., Choi, S., & Baumberger, T. (1996). Formalism or functionalism? Evidence from the study of language development. In M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan, & K. Wheatley (eds.) *Functionalism and formalism in linguistics: Vol. II. Case Studies* (pp. 317-340). Amsterdam: John Benjamins.
Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Guitierrez-Clellen, V. (1993). Early lexical development of Spanish-speaking infants and toddlers. *Journal of Child Language*, *20*, 523-549.
Kauschke, C., Lee, H.-W., Pae, S. (2007). Similarities and variation in noun and verb acquisition a crosslinguistic study of children learning German, Korean, and Turkish. *Language and Cognitive Processes*, *22*, 1-28.
Kern, S., & Türkay, F. A. (2006). Influence cognitive générale et ou langagiéré spécifique sur les premiers mots des enfants: Etude pilote chez des enfants français et des enfants turcs. *Poster presentend at ARCo. Colloque de l'Association pour la Recherche Cognitive*. France: Bordeaux.
Lucas, I. G., & Bernardo, A. B. I. (2008). Exploring noun bias in Filipino-English bilingual children. *The Journal of Genetic Psychology: Research and Theory in Human Development*, *169*(2), 149-164.
Leseman, P. M. (2000). Bilingual vocabulary development of Turkish preschoolers in the Netherlands, *Journal of Multilingual and Multicultural Development*, *21*(2), 93-112.
Maguire, M. J., Hirsy-Pasek, K., & Golinkoff, R. M. (2006). A unified theory of word learning: Putting verb acquisition in context. In K. Hirsh-Pasek & R. M. Golinkoff (eds.), *Action meets words: How children learn verbs* (pp. 364-391). New York: Oxford University Press.
Maital, S. L., Dromi, E., Sagi, A., & Bornstein, M. H. (2000). The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language*, *27*(1), 43-67.
Marchman, V., & Martinez-Sussman, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language and Hearing Research*, *45*, 983-997.
Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, *32*, 492-504.
Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian and Mandarin. *Journal of Child Language*, *24*, 535-565.
Türkay, N. F. (2005). *Children's early lexicon in terms of noun/verb dominance*. Unpublished PhD Dissertation. Adana: Çukurova University.
Verhoeven, L. (1999). Literacy and schooling in a multilingual society. In L. Eldering & P. P. M. Leseman (eds.), *Effective early education: Cross-cultural perspectives* (pp. 213-232). New York: Falmer Press.
Xuan, L., & Dollaghan, C. (2012). Language-specific noun bias: evidence from bilingual children. *Journal of Child Language*, *40*(5), 1057-1075.

Language acquisition in childhood obstructive sleep apnea syndrome

Georgia Andreou, Matina Tasioudi
andreou@uth.gr, tasioudim@hotmail.com

University of Thessaly

Abstract. Obstructive Sleep Apnea Syndrome (OSAS) is a condition in which the upper airway becomes blocked repeatedly during sleep, resulting in increased respiratory effort and snoring, recurrent hypoxia, and frequent arousals from sleep. Sleep disorders in children are associated with many neurocognitive problems, like reduced attention problems, hyperactivity, impulsivity, insufficient working memory, and learning difficulties. The prevalence of OSAS in children ranges from 1% to 6% in different epidemiological studies and the incidence peak was found in pre-school children, aged 2-8 years old, during the most critical age for language acquisition. Moreover, few studies have shown that language acquisition is also affected by the occurrence of OSAS but until now research has mostly examined the general language ability without focusing on specific language areas. Therefore, our goal is to investigate which language areas are mostly affected by this syndrome during early childhood. In the present study, the participants were 25 children with OSAS, according to a polysomnography examination performed at the University Hospital of Larissa, Greece, aged 4;1 to 6;11 (Mean Chronological Age - MCA: 5,6) years old, and 25 typically developing children (TD) of the same age, without breathing disturbances during sleep. Language performance was tested using a standardized psychometric language test (L-a-T-o I) created for the Greek language that focuses on all the basic linguistic areas (phonology, morphosyntax and semantics). Our findings indicated that children with OSAS had lower overall performance in general language ability compared to the TD children. More specific OSAS children showed significant difficulties in coping with complex phonological, morphosyntactic and semantic tasks. The lower performance of the OSAS group was found in morphosyntax and in phonology tasks, where the participants had scores near the borderline of normal performance. Our findings are consistent with literature findings indicating that OSAS affects language acquisition in early childhood, as a result of breathing disorders during sleep, nocturnal intermittent hypoxia and frequent arousals. Moreover, our results extend previous findings by focusing on language areas that seem most problematic in children with OSAS and suggest that early language intervention is necessary, along with medical treatment.

Keywords: obstructive sleep apnea syndrome, children, phonology, morphosyntax, semantics

Introduction

Obstructive Sleep Apnea Syndrome (OSAS) in children was first reported almost 40 years ago (Guilleminault, Eldrige, Simmons, & Dement, 1967) as a rare condition, focusing mainly on the severity and comorbidities of the syndrome, e.g., cor pulmonale. In 2002, the American Academy of Pediatrics (AAP) published a technical report on the diagnosis and management of childhood OSAS, using the following definition of the syndrome:

> "OSAS in children is a disorder of breathing during sleep characterized by prolonged partial upper airway obstruction and/or intermittent complete obstruction (obstructive apnea) that disrupts normal ventilation during sleep and normal sleep patterns. It is associated with symptoms including habitual (nightly) snoring, sleep difficulties, and/or daytime neurobehavioral problems. Complications may include growth abnormalities, neurologic disorders, and cor pulmonale, especially in severe cases."

This definition is still considered valid. According to the American Academy of Sleep Medicine (AASM, 2012) manual, sleep and associated events in obstructive apnea in children is scored when there is an absence (or >90% reduction) of airflow that lasts for 2 or more missed breaths and is associated with maintained inspiratory effort, while an obstructive hypopnea is described as having a

>30% reduction in airflow associated with a ≥4% decrease in oxygen desaturation and/or arousal and also lasts for 2 or more missed breaths with preservation of respiratory effort. OSAS is considered the most important manifestation of the spectrum of Sleep-Disordered Breathing (SDB), which includes several disorders as primary snoring and upper airway resistance syndrome, among others (Nespoli, Caprioglio, Brunetti, & Nosetti, 2013). The syndrome's prevalence during childhood varies in different studies between 1% - 6% and it peaks between 2-8 years of age (Balbani, Weber, Silke, & Montovani, 2005; Lumeng & Chervin, 2008; Marcus, Brooks, Ward, Draper, Gozal, Halbower, Jones, Lehmann, Schechter, Sheldon, Shiffman, & Spruyt, 2012).

When it comes to the pathophysiology of the syndrome, nasopharyngeal obstruction due to excess of lymphoid tissues (adenoids and tonsils) is considered the main cause since it alters the head's posture in order to facilitate breathing, and influences the craniofacial structure and the dentofacial development accordingly (Ellingsen, Vandevanter, Shapiro, & Shapiro, 1995). Childhood obesity is also considered an important predisposing factor for the occurrence of OSAS (Narang & Mathew, 2012).

In order to diagnose OSAS polysomnography's (PSG) sleep, recordings are characterized as the gold standard since clinical symptoms alone are not reliable enough for most patients (Marcus et al., 2012). The measurement of apneas and hypopneas during PSG allows quantifying the severity of OSAS in children using the combined number of apneas and hypopneas per hour, most known as the Apnea/Hypopnea Index (AHI), along with other measurements as severity of gas exchange abnormalities, and the amount of sleep disruption. An AHI>1 is considered abnormal according to the American Academy of Sleep Medicine, while immediate treatment is recommended for any child with an AHI >5. Severe childhood OSAS is defined as having an AHI >10. Nevertheless, we should not oversee that detailed questionnaires on day and night symptoms and a complete physical examination are also necessary for a proper diagnosis, classification of the severity and treatment of the syndrome (Urquhart, 2013).

There are several morbidities associated with the presence of OSAS in children, most of which result from the local and systemic inflammation, the intermittent hypoxemia and sleep fragmentation events. These events are strongly linked to delayed growth and development (Capdevila & Gozal, 2008). Among the most important consequences are also the cardiovascular complications (altered autonomic nervous system on the right and left ventricles), which thanks to early diagnosis and proper treatment present low frequency during the last years (Nespoli et al., 2013).

Adenotonsillectomy (AT) is the most common and effective way of treatment and helps at least 2/3 of the OSAS children, while orthodontic treatment and orofacial muscle reinforcing physiotherapy should also be considered in certain cases, along with treatment of obesity in case of presence. For patients who do not improve significantly with AT, a continuous positive airways pressure (CPAP) via a face-mask or nasal mask during sleep also has good results and is mostly recommended for children with OSAS and obesity (Marcus et al., 2012).

Besides the pathological morbidities of childhood OSAS, neurocognitive deficits are also a major consequence of the syndrome and are presented as neurobehavioral problems (aggressive/ depressive behavior, emotional lability), learning difficulties, poor school performance, ADHD, and even lower IQ compared to the healthy population (Bass, Corwin, Gozal, Moore, Nishida, Parker, Schonwald, Wilker, Stehle, & Kinane, 2004; Gottlieb, Chase, Vezina, Heeren, Corwin, Auerbach, Weese-Mayer, & Lesko, 2004; Kohler, Lushington, van den Heuvel, Martin, Pamula, & Kennedy, 2009; Kurnatowski, Putynski, Lapienis, & Kowalska, 2006; Rosen, 2004). Most of the research so far has focused only on cognitive functions such as attention and memory, and there are few studies that focused on language abilities (Andreou & Agapitou, 2007; Honaker, Gozal, Bennett, Capdevila, & Spruyt, 2009) which found significant semantic and phonemic language deficits in children and adolescents with OSAS or SDB.

What is quite promising, on the other hand, is that preliminary data suggest that some of these cognitive deficits may be reversible following treatment of mild sleep apnea in children. However, factors such as age at treatment time, duration of sleep disordered breathing, pre-morbid intellectual level, socioeconomic status, or the effectiveness of treatment may adversely affect long-term outcome

(Landau, Bar-Yishay, Greenberg-Dotan, Goldbart, Tarasiuk, & Tal, 2012; Montgomery-Downs, Crabtree, & Gozal, 2005).

The cause of cognitive decline in childhood OSAS seems to be complicated. Significant correlation has been found between cognitive impairment and daytime sleepiness, and there are also studies blaming mainly the nocturnal hypoxia for the same effects. Memory consolidation is postulated to occur during REM sleep (Diekelmann & Born, 2010), so sleep fragmentation may be one of the reasons affecting cognitive functions in children with OSAS. It is also claimed that OSAS results in damage to certain brain structures if left untreated. More specifically, it causes a decrease in gray matter in the hippocampus, in the anterior cingulate, in the cerebellum, and in the frontal, parietal, and temporal lobes and also in the hippocampal volume. In addition, the frontal and parietal cortices become abnormal and it is also considered the cause for decreased mean neuronal metabolite ratio of N-acetyl aspartate to choline in the left hippocampus and right frontal cortex. Pediatric researches have also found that there is a shared pathogenetic mechanism between endothelial dysfunction caused by OSAS and neurocognitive impairment. Another model proposed is the microvascular theory suggesting that a vascular compromise might exist in the small vessels of the brain. The most recent theory suggests that sleep defragmentation and hypoxia effects are synergistic and they interact with vulnerable brain regions such as hippocampus, prefrontal cortex, subcortical gray and white matter (for a more detailed review see Andreou, Vlachos, & Makanikas, 2014; Lal, Strange, & Bachman, 2012).

Based on the above facts we see that several studies have found that general cognitive ability and language are affected by the occurrence of OSAS, but until now research has mostly examined the general language ability without focusing on specific language areas. Therefore, our goal is to investigate the main language areas that are quite challenging for Greek children with developmental language disorders, namely morphosyntax, phonology and semantics (Stavrakaki, 2005), in order to examine which of these areas are also challenging for children with OSAS.

Method

In the present study, the participants were 25 children with OSAS, according to a polysomnography examination performed at the University Hospital of Larissa, Greece, (Apnea/Hypopnea Index/hour (AHI) = 5.25, SD = 3.26) aged 4;1-6;11 years old (Mean Chronological Age - MCA = 5.6, SD = 1.02) and 25 typically developing (TD) children of the same age (MCA = 5.7, SD = 1.16/ AHI = 0).

After the diagnosis, we informed the parents about the study and, following their consent, we performed individual examinations at the children’s homes that lasted 1-2 hours over 2 days. During the examination, language tests were administered to our participants. The same tests were administered to a control group of children of with the same age span that did not have breathing disorders during sleep.

To assess the language ability of children with OSAS, we administered to all participants a standardized language ability test for the Greek language that examines all levels of language and speech forms (L-a-T-o I: Psychometric test for language ability - Tzouriadou, Singolitou, Anagnostopoulou, & Vakola, 2008). It is also suitable for children 4;0-7;11 years old. It is based on the developmental approach on language acquisition and combines the developmental psychological and cognitive approaches. It consists of 10 subtests and evaluates both written and spoken language. It leads to a general language ability ratio (mean performance = 100, SD = 15) and evaluates also the adequacy of the phonological, morphosyntactic and semantic elements of speech (mean performance = 10, SD = 3).

In order to analyze our results we performed independent sample t-tests to compare the performance of the two groups of participants on the subtests administered using the statistical program SPSS18.

Results

Our results are presented in Table 1 where we can see the mean performance of each group in general language ability, semantics, phonology and morphosyntax.

Table 1. Mean performance of the groups in the language test

	T.D.		OSAS		T value	P value
	Mean	SD	Mean	SD		
Semantics	**10.9**	**(1.76)**	**8.4**	**(2.44)**	**3.05**	**0.004***
Phonology	**10.7**	**(1.95)**	**7**	**(1.91)**	**5.34**	**0.000***
Morphosyntax	**10.4**	**(2.11)**	**6.7**	**(1.70)**	**5.79**	**0.000***
General Language Ability ratio	**107**	**(13.48)**	**86**	**(11.18)**	**4.94**	**0.000***
***(Statistically significant result, p<.005)**						

As we can see by comparing the Mean Performance of the two groups of participants, the OSAS group performed worse than the typically developing group of children (TD) in all language levels tested. Especially in morphosyntax, the mean performance of the OSAS group was lower than 1 SD below the mean score for this age, as described in L-a-T-o I (Mean = 10, SD = 3). The scoring in phonology was also quite low, but still within the normal range, while the OSAS group performed better in semantics, but still worse than the TD group. The General Language Ability ratio was also lower for OSAS children, but also still within the normal range of results (Mean = 100, SD = 15). The between group analysis performed, using independent sample t-tests, indicated that the OSAS children scored significantly below the TD controls in each language domain tested and in general language ability.

Discussion

In this research, we considered it very important to administer a test with language tasks that focused on all basic language areas since literature so far hasn't shed enough light on the effects of childhood OSAS on language development, while there is a lot of research showing its effects on cognition (Gozal, 2008). We believe that the administration of these tasks will give new directions in research and will help focus on any language problems these children have. Based on our results, participants with OSAS had significantly lower General Language Ability scores in comparison to typically developing children and our findings are in accordance with previous literature findings that revealed deficits in language and general cognitive problems (Gozal, 2008; Halbower, & Mahone, 2006; Kohler et al., 2009; Kurnatowski et al., 2006). We also suggest that the reduced language ability that we found even in early childhood is one of the main factors leading to poor academic performance in older children with OSAS, or SDB (Andreou & Agapitou, 2007; Gozal & Pope, 2001) since it can lead to difficulties on processing complex verbal instructions and in poor expressive language at both the oral and written level.

Besides the General Language Ability in childhood OSAS, we also examined the performance of our participants in the three main linguistic areas of phonology, morphosyntax and semantics. These areas are so far the most examined areas in language disorders and in typical development (Guasti, 2007; Stavrakaki, 2005). As we mentioned earlier, there is only sporadic evidence on the effects of OSAS in these areas specifically and our purpose was to approach the language effects of the syndrome in a more systematic way and to a full extent. We found statistically significant results that reveal the effect of OSAS on all language areas examined, as shown in Table 1. More specifically, the phonology tasks revealed effects of OSAS on phonological development, since the OSAS group performed significantly worse than the TD group, in the borderline of normal performance. Our results in this part of the test are also consistent with previous results that revealed problems in phonemic fluency (Andreou & Agapitou, 2007) and reduced verbal abilities in children with OSAS (Honaker et al., 2009). These findings could be attributed to reduced auditory processing (Ziliotto, dos Santos, Monteiro, Pradella-Hallinan, Moreira, Pereira, Weckx, Fujita, & Pizarro, 2006) that is found in dichotic digits test in children with OSAS. Moreover, the problems that we found in phonemic production could also be attributed to early changes in facial bone growth that are found in children with OSAS (e.g., anterior and inferior position of the hyoid bone) (Vieira, Itikawa, de Almeida, Fernandes, Anselmo-Lima, & Valera, 2011).

Through the examination of the semantic development in childhood OSAS, we also found a significant difference in the performance between the two groups of participants. Semantics is so far the most studied linguistic area for adults with OSAS, and seems to be affected by its presence (see Andreou & Makanikas, 2014 for a review). Nevertheless, there are a few studies in children and adolescents with OSAS that have shown impairments in semantic development in the form of dysfunction in semantic fluency or receptive and expressive vocabulary (Andreou & Agapitou, 2007; see also Halbower & Mahone, 2006 for a review). Therefore, our results seem to be consistent with literature findings indicating lower performance in the semantic area, and we also suggest that possible reasons for these results are slower information processing and insufficient encoding, along with verbal memory problems that are found in children with OSAS (Spruyt, Capdevila, Kheirandish-Gozal, & Gozal, 2009).

With regard to morphosyntax, we also found statistically significant lower performance of the OSAS group of participants. This is the only area in which the OSAS participants performed slightly lower than the normal range of results as described in our standardized test, and seems to be the most affected in our study. However, there aren't any studies in childhood OSAS addressing the effects on morphosyntax, besides studies that have found insufficient encoding and difficulties in processing verbal instructions of increasing linguistic complexity (Honaker et al., 2009, Spruyt et al., 2009), that may explain our results. Nevertheless, morphosyntax is the most studied area in children with language disorders in Greek and seems to be the most challenging for them (Stavrakaki, 2005). According to our results, it is also challenging for children with OSAS.

Based on our findings, we see that OSAS affects language acquisition in early childhood and we could also suggest, based on the findings of the literature, that there is a complex mechanism affecting several brain areas that may explain these results. Our findings could be attributed to sleep defragmentation and hypoxemia, since these pathologic situations affect the function of hippocampus, prefrontal cortex, frontal lobes, and subcortical white matter, and these brain regions are highly associated in the literature with language functions and memory (Nagy, Westerberg, & Klingberg, 2004; Vigneau, Beaucousin, Hervé, Duffau, Crivello, Houdé, Mazoyer, & Tzourio-Mazoyer, 2006; among others). Nevertheless, in order to confirm this hypothesis, more research should be conducted combining language testing, preferably using online tests, and simultaneous brain activity monitoring. Therefore, regarding childhood OSAS, more research should be conducted based on a multidisciplinary approach in order to examine different aspects of the effects of the syndrome. Moreover, medical treatment should also be combined with early language intervention in order to avoid the risk of poor language development and consequent failure at school.

Conclusion

We studied the effects of OSAS in early childhood because previous studies on the language abilities of OSAS patients do not clearly indicate which are the language domains mostly affected and to what extent. Our findings reveal a significant effect of OSAS in all the administered tasks (general language ability, phonology, semantics, and morphosyntax). Therefore, we suggest that more research is needed in the field taking into account factors other than the severity of symptoms, such as the age of the participants, because it has been suggested that OSAS occurrence during critical ages of brain growth and development, such as childhood and adolescence, may cause notable language decline. Multidisciplinary approaches and brain monitoring techniques could also give new perspectives in finding the causes, in diagnosing and treating OSAS during childhood.

References

American Academy of Pediatrics (AAP). (2002). Clinical practice guideline: diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics*, *109*(4), 704-712.

American Academy of Sleep Medicine. (2012). *AASM Manual for the Scoring of Sleep and Associated Events Rules*: *Terminology and Technical Specifications,* Version 2.0.

Andreou, G., & Agapitou, P. (2007). Reduced language abilities in adolescents who snore. *Archives of Clinical Neuropsychology*, *22*(2), 225-229.

Andreou, G., & Makanikas, K. (2014). Obstructive Sleep Apnea Syndrome and its impact on verbal abilities from childhood to adulthood. In *Language and special education: A psycholinguisitc approach* (pp. 185-213). Thessaloniki: Giachoudis & Co. Publications.

Andreou, G., Vlachos, F., & Makanikas, K. (2014). Effects of chronic Obstructive Pulmonary Disease and Obstructive Sleep Apnea on cognitive functions: Evidence for a common Nature. *Sleep Disorders*, 1-18.

Balbani, A. P., Weber, S.A., Silke, A.T., & Montovani J. C. (2005). Update in obstructive sleep apnea syndrome in children. *Revista Brasileira de Otorrinolaringologia*, *71*(1), 74-80.

Bass, J. L., Corwin, M., Gozal, D., Moore, C., Nishida, H., Parker, S., Schonwald, A., Wilker, R. E., Stehle, S., & Kinane, T. B. (2004). The effect of chronic or intermittent hypoxia on cognition in childhood: a review of the evidence. *Pediatrics*, *114*(3), 805-816.

Capdevila, O. S., & Gozal, D. (2008). Neurobiological consequences of sleep apnea syndrome in children. *Revista de Neurologia*, *47*(12), 659-664.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*, 114-126.

Ellingsen, R., Vandevanter, C., Shapiro, P., & Shapiro, G. (1995). Temporal variation in nasal and oral breathing in children. *American Journal of Orthodontics and Dentofacial Orthopedics*, *107*, 411-417.

Gottlieb, D. J., Chase, C., Vezina, R. M., Heeren, T. C., Corwin, M. J., Auerbach, S. H., Weese-Mayer, D.E., & Lesko, S. M. (2004). Sleep-disordered breathing symptoms are associated with poorer cognitive function in 5-year-old children. *Journal of Pediatrics*, *145(4)*, 458-464.

Gozal, D. (2008). Obstructive Sleep Apnea in children: Implications for the developing of central nervous system. *Seminars in Pediatric Neurology*, *15*(2), 100-106.

Gozal, D., & Pope, D. (2001). Snoring during early childhood and academic performance at ages thirteen to fourteen years. *Pediatrics*, *107(6)*, 1394-1399.

Guasti, M. (2007). *L'acquisizione del linguaggio: Un' introduzione*. Milano: Cortina.

Guilleminault, C., Eldrige, F., Simmons, F.B., & Dement, W.C. (1976). Sleep apnea in eight children. *Pediatrics*, *58*, 23-30.

Halbower, A., & Mahone, M. (2006). Neuropsychological morbidity linked to childhood sleep-disordered breathing. *Sleep Medicine Reviews*, *10*(2), 97-107.

Honaker, S. M., Gozal, D., Bennett, J., Capdevila, O. S., & Spruyt, K. (2009). Sleep-disordered breathing and verbal skills in school-aged community children. *Developmental Neuropsychology*, *34*(5), 588-600.

Kohler, M. J., Lushington, K., van den Heuvel, C. J., Martin, J., Pamula, Y., & Kennedy, D. (2009). Adenotonsillectomy and neurocognitive deficits in children with Sleep Disordered Breathing. *PLoS ONE*, *4*(10), e7343.

Kurnatowski, P., Putynski, L., Lapienis, M., & Kowalska, B. (2006). Neurocognitive abilities in children with adenotonsillar hyperthrophy. *International Journal of Pediatric Otorhinolaryngology*, *70*, 419-424.

Lal, C., Strange, & C., Bachman, D. (2012). Neurocognitive impairment in Obstructive Sleep Apnea. *Chest*, *141*(6), 1601-1610.

Landau, Y. E., Bar-Yishay, O., Greenberg-Dotan, S., Goldbart, A. D., Tarasiuk, A., & Tal, A. (2012). Impaired behavioral and neurocognitive function in preschool children with obstructive sleep apnea. *Pediatric Pulmonology*, *47*, 180-188.

Lumeng, J.C., & Chervin, R.D. (2008). Epidemiology of pediatric obstructive sleep apnea. *Proceedings of the American Thoracic Society*, *5*, 242-252.

Marcus, C. L., Brooks, L. J., Ward, S. D., Draper, K. A., Gozal, D., Halbower, A. C., Jones, J., Lehmann, C., Schechter, M. S., Sheldon, S., Shiffman, R. N., & Spruyt, K. (2012). Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics*, *130*, 714-755.

Montgomery-Downs, H., Crabtree, V. M., & Gozal, D. (2005). Cognition, sleep and respiration in at-risk children treated for obstructive sleep apnea. *European Respiratory Journal*, *25*(2), 336-342.

Nagy, Z., Westerberg, H., & Klingberg, T. (2004). Maturation of white matter is associated with the development of cognitive functions during childhood. *Journal of Cognitive Neuroscience*, *16*(7), 1227-1233.

Narang, I., & Mathew J. (2012). Childhood obesity and Obstructive Sleep Apnea. *Journal of Nutrition and Metabolism*, vol. 2012, Article ID 134202, 8 pages.

Nespoli, L., Caprioglio, A., Brunetti, L., & Nosetti, L. (2013). Obstructive sleep apnea syndrome in childhood. *Early Human Development*, *89*(3), S33-S37.

Rosen, C. L. (2004). Obstructive sleep apnea syndrome in children: controversies in diagnosis and treatment. *Pediatric Clinics of North America*, *51*(1), 153-167.

Spruyt, K., Capdevila, O. S., Kheirandish-Gozal, L., & Gozal, D. (2009). Inefficient or insufficient encoding as potential primary deficit in neurodevelopmental performance among children with OSA. *Developmental Neuropsychology*, *34*(5), 601-614.

Stavrakaki, S. (2005). Greek neurolinguistics: The state-of-the-Art. *Journal of Greek Linguistics*, *5*, 181-228.

Tzouriadou, M., Singolitou, E., Anagnostopoulou, E., & Vakola, I. (2008). *Psychometric criterion of language adequacy (L-a-T-o-1)*. Thessaloniki, Aristotle University of Thessaloniki and Ministry of Education.

Urquhart, D. (2013). Investigation and management of childhood sleep apnoea. *Hippokratia*, *17*(3), 196-202.

Vieira, B. B., Itikawa, C. E., de Almeida, L. A., Fernandes, R. M., Anselmo-Lima, W. T., & Valera. F. C. (2011). Cephalometric evaluation of facial pattern and hyoid bone position in children with obstructive sleep apnoea syndrome. *International Journal of Paediatric Otorhinolaryngology*, *75*, 383-386.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*(4), 1414-1432.

Ziliotto, K. N., dos Santos, M. F., Monteiro, V. G., Pradella-Hallinan, M., Moreira, G. A., Pereira, L. D., Weckx, L. L., Fujita, R. R., & Pizarro, G. U. (2006). Auditory processing assessment in children with obstructive sleep apnea syndrome. *Brazilian Journal of Otorhinolaryngology*, *72*(3), 321-327.

A Greek/English bilingual child's acquisition of /fl/ and /vl/

Elena Babatsouli
ebabatsouli@ismbs.eu

Institute of Monolingual and Bilingual Speech

Abstract. The length of stages and the phonological processes involved in consonant cluster development are of interest in assessing child speech and in guiding intervention techniques for children with speech sound disorders. Of particular interest is the acquisition of /Cl/ clusters, as most children acquire /l/ much later than the consonant in first member position. This paper addresses the acquisition of /fl/ and /vl/ clusters in a Greek/English bilingual child's speech utilizing dense longitudinal data obtained on a daily basis. The results provide a new perspective in cluster development as the child has acquired all cluster members as singletons and reduces clusters at the start of the study at age 2;7. The second stage is absent in /fl/ and rare in /vl/, with the reduction and acquisition stages overlapping for about ten months. /fl/ reduces to [f] while the reduction of /vl/ is lexical dependent and not faithful to the sonority-based onset selection, with [l] persisting as long as [v]. Last, /vl/ is fully acquired three months after /fl/.

Keywords: consonant clusters, child, bilingual development, Greek, English

Introduction

Consonant cluster acquisition in child speech is of particular interest in phonological acquisition research since clusters are difficult to master in the language learning process. Markedness (Jakobson, 1941/68) underpins ease or difficulty of phonological learning, it determines acquisition paths, and is subject to universal, language-specific and learner-specific tendencies. Research on consonant cluster acquisition cross-linguistically accounts for this interplay of factors. A longitudinal case-study in child bilingual acquisition is a small-scale mirror of such research that adds to the cross-linguistic data pool and may inform typical and atypical language acquisition.

Greenlee (1974) first suggested stages in the acquisition of stop+liquid clusters in ten children's developmental data between 0;10-0;4 in Czech, English, Estonian, French, Serbian, and Slovenian. Similar temporal and phonological processes control cluster production that is found to progress through three main stages: *Stage I*: reduction to a single element; *Stage II*: two-element production involving assimilatory substitutions; *Stage III*: correct productions. Despite considerable overlapping between stages, there is cohesiveness across languages and children. Reductions to [stop] dominate *stage I*, and intermittent reductions to [liquid] are rare. Two-element clusters occur from 10 months to age 4;0. Idiosyncratic variation characterizes cluster production across children and languages: i) not all phonological processes apply to all clusters, ii) the length of overlapping between stages varies; iii) the effect of phonological processes varies, iv) element substitutions vary even in individual children.

Work since provides further evidence of Greenlee's stages (Smit, 1993; McLeod, van Doorn, & Reed, 2002; Yavaş & Babatsouli, 2016). Cluster types follow individual developmental paths being at different stages even within individual children (Ingram, 1976). Investigations also focus on what drives reduction patterns, which element is being substituted (C_1 or C_2), and what constraints drive the substitutions. The 'sonority-based onset selection' (Pater & Barlow, 2003) drives reduction to obstruents as a preference for non-sonorants in accordance with the *sonority scale* (Selkirk, 1984): *vowels>glides>liquids >nasals>fricatives>stops*, whereby vowels are highly sonorant. During acquisition, rising sonority clusters with a small sonority distance are more marked than large distance ones (Gierut, 1999). Regarding substitutions, C_1 is substituted if C_1[fricative] and C_2 is substituted if C_1[stop]. Mostly assimilations and fewer dissimilations guide substitution processes (Kirk, 2008).

A large cross-sectional study (1,049 children aged 2;0-9;0) on word-initial two-member clusters in English (Smit, 1993) validates the following: negligible whole-cluster deletion, reduction to a single element, and two-element production substituting either both elements or one. The last three processes are present in all ages. Element substitutions are overwhelmingly predicted by substitutions in singleton contexts. Reduction may be irrespective of markedness considerations: e.g. in /tw/, the unmarked /w/ is deleted. In their overwhelming majority, obstruent+lateral clusters (e.g. pl-, fl-) reduce to a targeted or substituted obstruent. A single exception is intermittent /fl-/ reduction to [l, w], where [w] substitutes /l/ in obstruent+lateral clusters. Notably, /l/ is marked in monolingual English acquisition. The stimulus word for /fl-/ is *flag* produced correctly at 13% by 3;0, and 80% after 4;6. Though only /fl/ is allowed in English, Greek also permits /vl/. Sparse work on Greek obstruent+lateral cluster acquisition (PAL, 1995; Kappa, 2002) supports arguments on obstruent retention.

The present study utilizes a Greek/English bilingual child's dense data on word initial and medial /fl/ (English, Greek) and /vl/ (Greek) from ages 2;7-4;0, also reporting productions in non-targeted contexts. The female child's spontaneous utterances, transcribed by the author orthographically and in IPA in CLAN (MacWhinney, 2000), are time-aligned to digital recordings; recordings were made on an average of one hour daily, 4 days a week. Acoustic analysis verifies transcription reliability. In spite of claimed single-subject limitations, the purpose is to examine whether the child's longitudinal and uninterrupted data verify current knowledge on stages and related phonological processes in cluster acquisition. The study enriches the cross-linguistic data pool (especially with regard to Greek) and builds on existing gaps by examining two similar clusters (fl, vl) with different phonotactic distribution in the languages involved, that develop alongside one another longitudinally in the same person's bilingual acquisition. Results are also of significance to speech sound disorder (SSD) intervention on clusters cross-linguistically.

The acquisition of /fl/.

Between ages 2;7-3;5, the child targeted /fl/ 124 times in 14 words in English and 3 words in Greek. The words in English are: *butterfly(ies), flag, flash(ed), flat, flip, flippers, floor, flower(s, flush, fly(ies)*. The words in Greek (targeted ADULT PRODUCTION is shown) are: *φλούδι* /fluði/ 'fruit skin', *φλούδια* /fluðja/ 'fruit skins', παντόφλες /padofles/ 'slippers'. The realizations of /fl/ are shown monthly in Table 1 together with non-contextual [fl] productions, meaning [fl] productions in non-targeted /fl/ words. English words are shown in italics, while for Greek words the targeted ADULT PRODUCTION is given.

Ages 2;7 and 2;8 are evidence of Greenlee's (1974) cluster reduction stage, where /fl/ is reduced to [f]. The lateral is deleted following the 'sonority-based onset selection' (Pater and Barlow, 2003), whereby the least sonorant element is retained. The child remains faithful to this reduction pattern even after 2;8 whenever there is a reduction. It is noted that the child has fully acquired (>90%) singletons /f/ and /l/ by 2;7, which explains why these are rarely substituted when targeting /fl/. In fact, only /f/ is substituted in 5 out 124 /fl/ attempts, in all of which /l/ is deleted. It is observed that Greenlee's reported *stage II*, where there is substitution of cluster element(s), is *absent* in this child's acquisition of /fl/, and this is not due to sampling deficiencies. There is a single occurrence where the child deletes /l/ producing /fl/ as [sf], [s] being a substitution of singleton /f/. It may be that the child realized her error in producing /f/→[s] and immediately corrected herself producing [sf]. This, however, happened while repeating *flat* three times consecutively in the same utterance in a form of practice.

At 2;9, [fl] occurs for the first time in 1 out 4 attempts in *flower(s)*, though /fl/ is reduced the other times. Also in all other words, /fl/ is reduced to [f]. That is, during 2;9, there is instantaneous overlapping of Greenlee's reduction stage and final stage of correct realizations of the targeted cluster, skipping the substitution stage. Until age 3;1, the only correct instance of /fl/ occurs at 3;0 in *flat* in 1 out 3 consecutive attempts in the same utterance; repetition is known to produce different outcomes, in general (Ingram, 1989).

Greek /fl/ words are targeted for the first time at age 3;0. This is rather expected as there are fewer /fl/ words in the Greek language than in English. A simple dictionary search on words with word-initial /fl/ produces some 1,223 types in English and some 100 types in Greek. At age 3;0, /fl/ was reduced to [f] all 3 times in Greek *φλούδι(α)* [fluði, fluðja] ‘fruit-skin(s)’, as was the case for English *flashed*, *floor*, *flower*, *flush*. As mentioned earlier, the only exception is *flat* where realizations varied because of repetition.

Table 1. The child’s realizations of /fl/

age

words	2;7	2;8	2;9	2;10	2;11	3;0	3;1	3;2	3;3	3;4	3;5
butterfly			3f	3f				f	2f	2f	f
butteflies									f		
flag							f				
flash			f					fl			
flashed						f					
flat						f, sf, fl					
flip		f			s						
flippers					s						
floor	f	4f	5f	5f, s	2f	6f	11f	f			3fl
flower	f	f	f, ʃ	4f		4f	3f		2fl	fl	2fl
flowers		f	f, fl	f				fl	fl	9fl, f	
fluði						2f				fl	
fluðja						f					
flush			2f			f					
flies					f						
fly		f	6f	f	3f				3fl	fl	vl, fl
padofles									fl		
non-contextual [fl]											
dolphin											fl
fraula									fl		
fresh										fl	
Friday											fl
friendly										fl	
fruit								fl			
further										fl	
pretty											fl
swap										fl	

At age 3;1, there is only evidence of the reduction stage, as /fl/ is reduced to [f] in every attempt of targeted *flag*, *floor* and *flower*. At age 3;2, there is evidence of substantial overlapping between the first and final stage in the acquisition of /fl/. Correct realizations of /fl/ occur in *flash* and *flowers*, and reductions to [f] in *butterfly* and *floor*. It is interesting that non-contextual [fl] also starts appearing, overgeneralizing the cluster in the wrong context. This happened in ***fr**uit* produced with [fl], possibly because [l] is the child's substitution of the Greek rhotic that also interferes in productions with the rhotic in English.

Age 3;3 shows clear evidence of the final stage of correct production. The only reductions to [f] occur in the compound word *butterfly(ies)* that persist until age 4;0. It is noted, however, that compound word *dragonfly* was produced correctly at age 3;6, the first time it was targeted. This suggests that the child’s difficulty comes from the rhotic preceding /fl/, whose main substitution is [l]. Correct productions of /fl/ occur in English *flower(s)*, *fly* and in tri-syllabic Greek *παντόφλες* /padofles/ ‘slippers’ that was targeted for the first time. Here, there is another non-contextual [fl] in Greek *φράουλα* /**fr**aula/ ‘strawberry’.

At ages 3;4-3;5, the patterns are reminiscent of age 3;3. /fl/ is reduced to [f] in the compound word *butterfly*, though preserved in all the other words: *floor*, *flower*, *fly*. At 3;5, however, an exception is found to the child’s realization patterns longitudinally. In 1 out of 2 times, /fl/ is produced as [vl] in *fly*. The child voices /f/ by assimilation to a preceding interlocutor's utterance that included the words *love* and *fly*. It is likely that voicing in *fly* was influenced by the adult's preceding production of [v] and [fl] in *love* and *fly*, respectively. The process of perseveration within the utterance is known to occur in child developmental speech (Stemberger, 1989). What we have here is a generalized perseveration whereby the child is priming productions from the interlocutor's preceding utterance.

During the same period, non-contextual [fl] is produced at an increasing rate in *dolphin*, *fresh*, *Friday*, *friendly*, *further*, *pretty* and *swap*. When targeting *dolphin*, metathesis of /l/ and /f/ occurs, producing the heterosyllabic cluster /lf/ as a tautosyllabic cluster [fl]; this repeats at age 3;8. In *pretty*, both cluster members are substituted: /p/ becomes [f] and /ɹ/ is substituted by [l], the child's substitution for the Greek rhotic. Lastly, /sw/ in *swap* becomes [sl] through lateralization of /w/. Notably, this is the reverse phonological process observed in monolingual English children (e.g. Smit, 1993), where unmarked [w] substitutes the later-acquired /l/. Babatsouli (2015) showed late acquisition of /w/ in this bilingual child's English, though /l/ was acquired early in both languages: /w/ is not phonemically targeted in standard Greek, which explains the delay here in terms of enacting factors in bilingualism.

The acquisition of /vl/

Targeted /vl/ is permitted in Greek but not in English. From 2;7-3;6, the child targeted /vl/ 155 times in the following 15 words (targeted ADULT PRODUCTION): *αυλή(ές)* /avli, avles/ 'yard(s)', *έβλεπα* /evlepa/ 'I was seeing', *σουβλάκι* /suvlaci/ 'skewer', *τουβλάκια* /tuvlaca/ 'small bricks', *βιβλίο(α)* /vivlio, vivlia/ 'book(s)', *βιβλιαράκι* /vivliaɾaci/ 'small book', *βιβλιοθήκη* /vivlioθici/ 'bookcase', *βλέπε* /vlepe/ 'see!', *βλέπω* /vlepo/ 'I see', *βλέπεις* /vlepis/ 'you see', *βλέπει* /vlepi/ 'he sees', *βλέπουμε* /vlepume/ 'we see', *βλέπετε* /vlepete/ 'you (plural) see'. The realizations of /vl/ are presented monthly in Table 2, where non-contextual [vl] productions and their corresponding targeted words are also shown.

As was also the case for /fl/, 2;7 clearly marks the reduction stage in /vl/ becoming [v] both word-initially and word-medially; this supports previous findings in the literature (Kappa, 2002; PAL, 1995). Reductions dominate the child's /vl/ realizations until 3;3, showing some overlapping with the final stage of correct production between 2;8-3;3. The second stage is evidenced to be instantaneous and very weak, appearing at 2;8 (also overlapping with correct productions), with only 3 occurrences out of 155 targeted /vl/ longitudinally: /vl/ becomes [tl] twice at 2;8 and [vɾ] once at 3;4 (a time when both first and third stage dominate). /l/ is substituted by a rhotic that is overgeneralized in the wrong context; note that [l] is the substitution of the targeted rhotic all along. This overgeneralization also occurs in /vl/ reduction between 3;3-3;4. Similarly, [ð] is overgeneralized substituting /l/ during /vl/. [tl] occurs twice in /vlepete/ assimilating /v/ to /t/, even though reduction to [t] at 3;0 is also evidenced in /vlepis/; in both cases coronal assimilation dominates.

This reduction pattern is striking in that, from age 2;8 until full acquisition at 3;6, /vl/ in /vlepo/, its conjugations /vlepis, vlepi, vlepoume, vlepete/ and its past progressive tense /evlepa/ is pre-dominantly reduced to [l]. In all other words by contrast, /vl/ is reduced to [v]. Greenlee (1974) and Smit (1993) reports /obstruent+lateral/ reduction to [l] as exceptional, only occurring for short spells. Here, /vl/ consistently reduces to [l] for ten months! A possible explanation is that the child anticipates [labial] in /p/, which inhibits her production of labiodental /v/ in the cluster.

Further, there are non-contextual productions of [vl] starting at 2;10. These come about via three occurrences of epenthetic [v] either next to a targeted /l/ in *let* or next to a targeted /ð/ in *the* and *this*. There is also an instance of epenthetic [l] next to a targeted /v/ in Greek /vapso/ 'to paint'. Other occurrences involve Greek /vɾ, vɣ, vð/, where [l] substitutes the second member, as also in singleton contexts. Lastly, /b/ in Greek /ble/ 'blue' is fricated, and heterosyllabic members /ɫ,v/ in *solve* are shifted in metathesis producing [vl].

Table 2. The child's realizations of /vl/

age

words	2;7	2;8	2;9	2;10	2;11	3;0	3;1	3;2	3;3	3;4	3;5	3;6
avli		v	vl	3v,vl		3vl	vl					
avles						vl						
evlepa								l		l		
suvlaci				2v,l								
tuvlaca	3v		v	5v			v	3vl				
vivlia				3v		4v	v			3v,3vl		
vivliaraci								v				
vivlio	v					2v	v	v		4v,3vl,vr	2v	
vivlioθici											vl	
vlepe										vl		
vlepo	v				2l,2vl	3l	l	4l	2l,vl	2l,ð	2vl	8vl,m
vlepis		2l	4l,vl		2l	6l,t	l	l	5l,r	2ð,ϕ,r	l,ð	4vl,l
vlepi			3l		l	2l			ð		vl,ð	vl
vlepume			2l						2l	2l	l	
vlepete		4l,2tl,vl		l		v				3l		
non-contextual [vl]												
every				vl								
ble											vl	
fly											vl	
let											vl	
solve										2vl		
the								vl				
this				vl								
ravði									vl			
vapso												vl
vγali												vl
vγalis											2vl	
vγalo											vl	
vraci				vl								
vrika										vl		
vro												vl

A comparison of /fl/ and /vl/ in acquisition shows that both clusters enter the stage of correct production between 2;8-2;9. However, even though /vl/ is realized correctly more frequently than /fl/, it is *fully* acquired three months after /fl/ at age 3;6; even though Greek /vivlio/ is not targeted at 3;6 but is targeted twice at 3;7, both times /vl/ is produced correctly. The delay of /vl/ may be explained by the fact that /v/ has a slightly smaller sonority distance from /l/ than /f/, making it more difficult to correctly produce /vl/ than /fl/. This agrees with the general observation that children acquire clusters with a smaller sonority distance between their members at a later stage (Gierut, 1999).

Summary

Analysis of the dense longitudinal speech data of a bilingual child's development of /fl/ in English and Greek, and /vl/ in Greek revealed several interesting results. Even though all cluster members are fully acquired in singleton context by the start of data collection (2;7), there is a clear reduction stage lasting one month for /vl/ and two months for /fl/. The substitution stage is non-existent in /fl/, while rare and overlapping with the other two stages in /vl/. The reduction stage is dominant for several months with correct productions overlapping it. Cluster reduction shows two patterns: /fl/ reduces to [f], while /vl/ reduces to [f] or [l], showing lexical dependence. To the author's knowledge, reduction of /vl/ to [l] has not been reported in the literature. Reduction of other /Cl/ clusters to [l] has been reported before but shown to be rare and not stretching for extended time periods, like the ten months reported here for /vl/. There are occurrences of overgeneralizations of /ɾ, ð/ in reduced /vl/ to [l]. Full acquisition of /vl/ is accomplished three months later than full /fl/ acquisition, supporting the view that clusters with a shorter sonority distance between their members are acquired later. Non-contextual productions of [fl] and [vl] start occurring at about the same age as contextual ones, evidencing rule overgeneralization in the child's phonetic repertoire. Several phonological processes cause non-contextual [fl], [vl]: metathesis of both members in heterosyllabic /lf/ and /lv/ clusters; [v] epenthesis next to [l] either for a targeted /l/ or a substituted / ð/ or /ɾ/; and [l] substitution of non-acquired /C/ in /fC/ and /vC/ clusters.

Conclusion

The results of the present study may be used as a guide for assessing consonant clusters in child speech and for intervention techniques in children with speech sound disorders where cluster acquisition is often a major problem. The results provide a different perspective and may be particularly useful for monolingual English children, who acquire /l/ as a singleton much later than the child of the present study, and thus, developing their /Cl/ clusters differently, acquiring them much later. Moreover, even though /vl/ is not permitted in English, its developmental perspective given here, which is first in the literature at least for Greek, may also prove useful in helping children with SSD to improve production of permitted clusters in English.

References

Babatsouli, E. (2015). Technologies for the study of speech: Review and an application [Special Issue on Language Disorders and ICT]. *Themes in Science and Technology Education*, *8*(1), 17-32.

Gierut, J. (1999). Syllable onsets: Clusters and adjuncts in acquisition. *Journal of Speech, Language and Hearing Research*, *42*, 708-726.

Greenlee, M. (1974). Interacting processes in the child's acquisition of stop-liquid clusters. *Papers and Reports on Child Language Development* (Stanford University), *7*, 85-100.

Jakobson, R. (1941/1968). *Child language, phonological universals and aphasia*. (Keiler, A. Trans.). The Hague: Mouton. Original work published in 1941 as *Kindersprache, aphasie und allgemeine Lautgesetze*.

Ingram, D. (1976). *Phonological disability in children*. New York: Elsevier.

Ingram, D. (1989). *Phonological disability in children* (2nd ed.). London: Cole and Whurr Umited.

Kappa, I. (2002). On the acquisition of syllabic structure in Greek. *Journal of Greek Linguistics*, *3*, 1-52.

Kirk, C. (2008). Substitution errors in the production of word-initial and word-final consonant clusters. *Journal of Speech, Language, and Hearing Research*, *51*, 35-48.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum.

McLeod, S., van Doorn, J., & Reed, V. A. (2002). Typological description of the normal acquisition of consonant clusters. In F. Windsor, L. Kelly, & N. Hewlett (Eds.), Themes in Clinical Phonetics and Linguistics (pp. 185-200). Hillsdale, NJ: Lawrence Erlbaum.

PAL (Panhellenic Association of Logopaedics) (1995). *Assessment of phonetic and phonological development*. Athens: PAL (in Greek).

Pater, J., & Barlow, J. (2003). Constraint conflict in cluster reduction. *Journal of Child Language*, *30*, 487-526.

Selkirk, E. (1984). On the major class features and syllable theory. In: M. Aronoff & R. T. Oehrle (eds.), *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students* (pp. 107-136. Cambridge, MA: MIT Press.

Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska Articulation Norms Project: Word-initial consonant clusters, *Journal of Speech and Hearing Research*, 36, 931-947.

Smit, A. B., Hand, L, Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Joumal of Speech and Hearing Disorders*, *55*, 779-798.

Yavas, M., & Babatsouli, E. (2016). Acquisition of /s/-clusters in a Greek-English bilingual child: Sonority or OCP? In M. J. Ball (ed.), *Sonority across languages* (forthcoming). Equinox Publishing.

How much should phones weigh in computing phonological word proximity?

Elena Babatsouli[1], David Ingram[2], Dimitrios Sotiropoulos[3]
ebabatsouli@ismbs.eu, david.ingram@asu.edu, dimsotirop@yahoo.com

[1]Institute of Monolingual and Bilingual Speech, [2]Arizona State University, [3]Technical University of Crete

Abstract. Phonological word proximity, PWP, was introduced by Ingram and Ingram (2001) and Ingram (2002) to evaluate performance in child speech *per word* by weighing correctly produced in context consonants *twice* as much as produced vowels and substituted consonants. Babatsouli, Ingram, and Sotiropoulos (2011, 2014) obtained an explicit formula for PWP *cumulatively* for all words in a speech sample, in terms of the proportion of consonants correct (PCC), the proportion of phonemes deleted (PPD), and the proportion of targeted consonants (PC). In the present study, the relative weight of phones is taken as an arbitrary number *n*, in order to compare the advantages and disadvantages of such a PWP to Babatsouli et al.'s (2011, 2014) PWP of n=2, in assessing child speech. The derived expression for PWP is similar to Babatsouli et al.'s (2011, 2014); however, the weights of PCC and PPD are now dependent on *n* as well as on PV. As the product nPC increases, the weight of PCC increases and that of PPD decreases; when *n* is greater than 2, the weight of PCC is greater than that of PPD; for n=2 the weights are equal, while for *n* smaller than 2, the weight of PPD is larger. However, the difference between PCC (or PPD) weights of different n's, which generally increases for increasing PC, remains effectively constant for PC larger than 40%. These results have implications on how to compute phonological word proximity (PWP) for assessment purposes. Smaller relative weights of phones guarantee larger phonological word proximities when the proportion of vowels produced is larger than the proportion of consonants correct, which is generally the case. In comparing phonological word proximity between two such samples with their difference in the proportion of phonemes deleted (PPD) being larger than their difference in the proportion of consonants correct (PCC), it is advantageous to use smaller relative weights of phones if larger differences in PWP are sought. When, however, changes in PCC are larger than changes in PPD, phonological word proximity (PWP) becomes more sensitive for larger relative weights of phones. Last, independent of the relative weight of phones, phonological word proximity is more sensitive than PCC when changes in PPD are larger than changes in PCC; otherwise, it is not. These results may guide the establishment of speech performance norms for normal children, as well as assessing children with speech sound disorders (SSP) whose PCC values vary little across categories of word complexity, such as across monosyllabic or multisyllabic words with singleton consonants and monosyllabic or multisyllabic words with consonant clusters.

Keywords: phonological word proximity, measure, assessment, child, normal speech, disordered

Introduction

The proportion of consonants correct (PCC) (e.g., Shriberg, Austin, Lewis, McSweeney, & Wilson (1997)) has been widely used in the literature since the mid-1980s, and in practice for assessing typical and atypical children's speech in development, as well as children's disordered speech in terms of consonants productions. However, it was not until the early 2000s that a phonological measure was proposed to evaluate whole word productions. Ingram and Ingram (2001) and Ingram (2002) introduced the phonological mean length of utterance (PMLU) as the arithmetic mean of the PMLU of individual words, which is defined as the sum of the produced vowels and the substituted consonants plus twice the correctly produced (as targeted) consonants. Furthermore, the same authors introduced the proportion of word proximity (PWP) *per word*, hereon referred to as *phonological word proximity*, as the proportion of the produced PMLU to the targeted PMLU, with the PWP for a number of words in a speech sample being the arithmetic average of the PWP of individual words.

Since the introduction of PMLU and PWP, researchers have used these measures to evaluate speech performance in monolingual and bilingual child speech. Taelman, Durieux, and Gillis (2005) discussed how to use CLAN (MacWhinney, 2000) to compute PMLU and PWP using large speech data. Bunta, Fabiano-Smith, Goldstein, & Ingram (2009) compared 3-year old Spanish-English bilingual children to their monolingual peers to compute, among other quantities, PWP and the proportion of consonants correct, PCC. They found that while PWP and PCC differ in general, bilinguals only differ on PCC from their monolingual peers in Spanish and that when comparing the Spanish and English of the bilingual participants, PCC was significantly different but PWP was similar. Burrows and Goldstein (2010) compared PWP and PCC accuracy in Spanish-English bilinguals with SSD to age-matched monolingual peers. Macleod, Laukys, & Rvachew (2011) compared the change in PWP to that in PCC for two samples of twenty children each, both taken at the age of 18 months and at 36 months. One of the samples involved monolingual English children while the other involved bilingual French-English children. Their results showed that the PWP change was larger than the PCC change.

Babatsouli, Ingram, and Sotiropoulos (2011, 2014) took another look at the proportion of phonological word proximity (PWP). Instead of defining PWP *per word*, they defined it *cumulatively* for all the words in a speech sample. This enabled them to express PWP for the whole speech sample analytically in terms of the proportion of consonants correct (PCC), the proportion of consonants deleted (PCD) and the proportion of vowels (PV) in the targeted speech, and obtain upper and lower PWP bounds in general.

In the present paper, yet another look is taken at PWP in order to question why the correctly produced (as targeted) consonants should weigh twice as much as vowels and substituted consonants. This weighing factor, 2, was decided arbitrarily by Ingram and Ingram (2001) and Ingram (2002) and the effect of its choice on the sensitivity of PWP to changes of PCC and PCD has not been examined to date. The analytical expression derived by Babatsouli et al. (2011, 2014) provides the starting point for such an examination which will be done in the present paper.

The present study is motivated by the need to provide a proper measure for practitioners to evaluate children’s speech performance, as far as a phonological word measure is concerned. Ingram (2015) points out that when comparing typically developing children to children with SSD, PCC changes are dramatically different across categories of word complexity: monosyllabic words without consonant clusters, monosyllabic words with at least one consonant cluster, multisyllabic words without consonant clusters, multisyllabic words with at least one consonant cluster. For example, for children with SSD, PCC will likely remain unchanged when comparing performance between words without consonant clusters and words with consonant clusters. While for typically developing children or children with speech delay, this is not the case.

This and other cases will be examined here, in general, in light of how to compute PWP with respect to the value of the relative weight between correct consonants on the one hand and vowels and substituted consonants on the other hand. Therefore, the results of the present paper will provide guidelines for assessing speech performance not only for all the words in a speech sample but also for different categories of word complexity in the sample. The results obtained here are applicable to samples of running speech as well as to speech samples obtained from picture naming tests.

Phonological word proximity (PWP) for general weight of correct consonants

Ingram and Ingram (2001) and Ingram (2002) introduced the phonological word proximity (PWP) *per word* as follows:

$$\mathrm{PWP} = (\mathrm{CCP} + \mathrm{PH})/(2\mathrm{CCT}+\mathrm{VT}) \qquad (1)$$

where CCP is the number of correctly produced (as targeted) consonants, PH is the number of consonants and vowels produced whether correctly or not (vowels are assumed to be produced correctly as targeted), CCT is the number of targeted consonants in the word, and VT is the number of targeted vowels in the word. Therefore, in computing PWP per word using equation (1), correctly produced (as targeted) consonants (CCP) are weighed *twice* as much as substituted consonants and produced vowels. PWP for a number of words in a speech sample was subsequently obtained as the arithmetic average of the PWPs per word. However, such a cumulative PWP could not be analyzed in general.

Babatsouli et al. (2011, 2014) expressed the PWP in (1) in terms of the proportion of correctly produced (as targeted) consonants to the targeted consonants (PCC), the proportion of deleted segments to the targeted segments (PPD), and the proportion of targeted vowels to all targeted segments (PV), as follows:

$$PWP = pPCC + (1-p)\,(1-PPD), \quad p = (1-PV)/(2-PV) \tag{2}$$

Then, by taking the weighted average of the PWPs per word given by (2), Babatsouli et al. (2011, 2014) obtained a cumulative PWP for all the words in exactly the same form as (2), with the three phonological parameter components PCC, PPD, and PV now computed as the weighted averages of their corresponding values per word. For example, the cumulative PCC is now the proportion of correctly produced (as targeted) consonants in the whole speech sample to the targeted consonants in the whole speech sample as well. The cumulative PWP as expressed by equation (2), made it possible to obtain, in general, its upper and lower bounds.

Here, in order to analyze the effect of the weighing factor for correctly produced (as targeted) consonants on the cumulative PWP, a general weight equal to n+1 is considered, where n is any real number greater than zero, as it would be senseless not to weigh correctly produced (as targeted) consonants more than substituted consonants. The weight which was taken by Ingram and Ingram (2001) and Ingram (2002) and adopted by Babatsouli et al. (2011, 2014) as equal to 2 (n=1), is a special case of the general n>0 considered here. Following a similar derivation as in Babatsouli et al. (2011, 2014), the cumulative PWP for a general n>0 now becomes

$$PWP = pPCC + (1-p)\,(1-PPD), \quad p = nPC/(1+ nPC) \tag{3}$$

where PC=1-PV is the proportion of consonants to all segments (consonants and vowels) in the targeted speech sample. It is seen that when n=1, (3) reduces to (2). Further, the weight of PCC, p, is an increasing function of nPC while the weight of PPD, 1-p, is a decreasing function of nPC. The numerical values of the two weights are depicted in Figure 1 for different values of nPC. It is seen that the weight of PCC is smaller than the weight of PPD for nPC values smaller than 1, the two weights are equal for nPC equal to 1, while the weight of PCC is larger than the weight of PPD for nPC values larger than 1. In Ingram's proposition, n is equal to 1 and, therefore, the weight of PCC is always smaller than the weight of PPD, independent of the speech sample, as the proportion of targeted consonants, PC, to all targeted segments is smaller than 1.

Now, p, the weight of PCC, will be compared to the weight of the proportion of consonants deleted to the targeted consonants, for different values of n. To do this, PPD is written in terms of its two components, the proportion of consonants deleted to the targeted consonants, PCD, and the proportion of vowels deleted to the targeted vowels, PVD, as the sum of the following two products:

$$PPD = PCD\,(PC) + PVD\,(PV) \tag{4}$$

Comparing the weight of PCC, p, to the weight of PCD, (1-p)PC, gives:

$$(1-p)PC/p = 1/n \tag{5}$$

so that the weight of PCC is larger than the weight of PCD for any n larger than 1, it is equal to it for n equal to 1 (Ingram's proposition), and it is smaller than it for any n smaller than 1. Therefore, the value of n affects the relative contributions of PCC and PCD in PWP as given by (5).

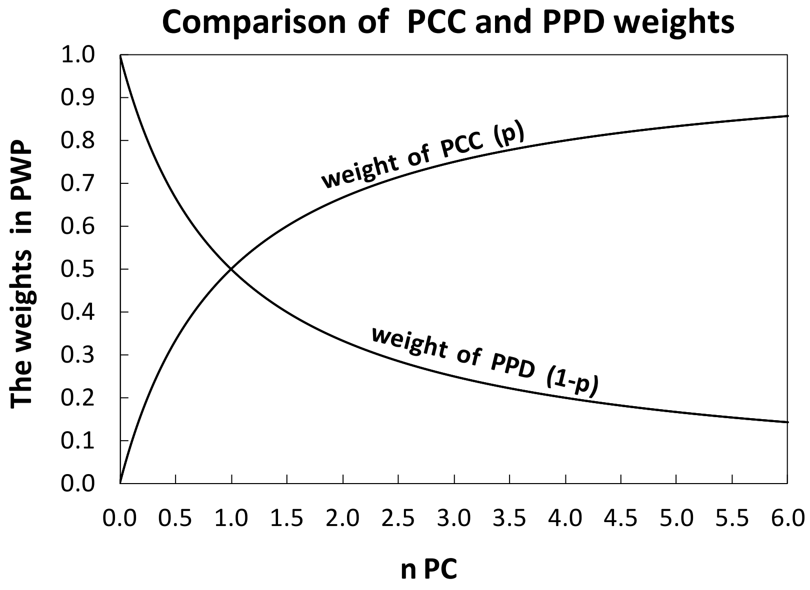


Figure 1. The weights of the proportion of consonants correct (PCC) and the proportion of phonemes deleted (PPD) versus nPC; n is the relative weight between correctly produced consonants and all phones, and PC is the proportion of consonants in targeted speech.

Applications

Hereon, the analysis will be such as to find practical applications directly. Three cases will be studied:

First case

In the first case, different speech performances on the same speech sample will be compared depending on the value of n chosen, i.e. the sensitivity of PWP on PCC and PPD changes will be examined in view of n, the relative weight of PCC to PCD. This is the case, for example, when comparing a child's performance at two different ages in development or when comparing two different children's (or groups of children's) performance at the same age. Here, for a given n, p is the same for both performances. To obtain an analytical expression for the change of PWP in terms of the changes of PCC and PCD, two cases are considered: a) the absolute value of the change of PPD is smaller than the absolute value of the change of PCC, and b) the absolute value of the change of PCC is smaller than the absolute value of the change of PPD. For the former case, without loss of generality, the change of PPD, ΔPPD, may be written as

$$\Delta PPD = -\kappa\, \Delta PCC, \quad 0 \le \kappa < 1 \tag{6}$$

where ΔPCC is the change of PCC across the two performances. Using (3) to compute PWP for each speech performance and then subtracting the two PWPs gives the change of PWP as:

$$\left| \Delta PWP \right| = \left| \Delta PCC \right| [\kappa + (1-\kappa)\, p] \tag{7}$$

Two remarks are made on this result: as the quantity in the bracket is smaller than 1 (its upper limit being equal to 1 when κ=1), the change of PWP is smaller than the change of PCC, and since p is an increasing function of nPC, the change of PWP gets closer to the change of PCC as nPC increases.

Case b can be obtained by setting κ=1/λ in (6) and (7) resulting in

$$\left| \Delta PWP \right| = \left| \Delta PPD \right| [1- (1-\lambda)\, p] \tag{8}$$

The quantity in the bracket is a decreasing function of nPC and smaller than 1, implying that the change of PWP is smaller than the change of PPD and it becomes even smaller as nPC increases. The conclusion drawn from equations (7) and (8) is that for small changes in PCC, the smaller the n is, the more sensitive PWP is across performances. On the other hand, for small changes in PPD, the larger the n is, the more sensitive PWP is across performances. When the absolute values of PCC and PPD changes are comparable ($\kappa=\lambda=1$), meaning that there is no change in the number of substituted consonants across performances, the change of PWP is comparable to the change of PPC and PPD.

Second case

In the second case, phonological word proximity, PWP, computed from performances across different speech samples will be examined. This case includes comparisons of a child's performance across speech samples that differ in the categories of word complexity that they include, i.e. monosyllabic words without consonant clusters, monosyllabic words with at least one consonant cluster, multisyllabic words without a consonant cluster, multisyllabic words with at least one consonant cluster. Here, having picked n, the weight p given by (3) changes across the speech samples as the proportion of consonants in targeted speech, PC, changes. How big is this change? By how much does it affect the value of the computed PWP? Suppose the PWP corresponding to the performance of a speech sample is computed using the weight p of the other speech sample. How different would it be from the actual PWP? Using (3) for two different weights, p, and then subtracting yields

$$\Delta PWP = - \Delta p\ (1\text{-}PCC\text{-}PPD) \tag{9}$$

where $\Delta PWP = PWP_2\text{-}PWP_1$ and $\Delta p = p_2\text{-}p_1$, with the subscript indicating the different p. It is noted that the quantity in the parenthesis in equation (9) is always smaller than 1, so that the change of PWP is smaller than the change of p. The change of p may be seen in Figure 2, where it is plotted for different values of PC for n=1 and n=2.

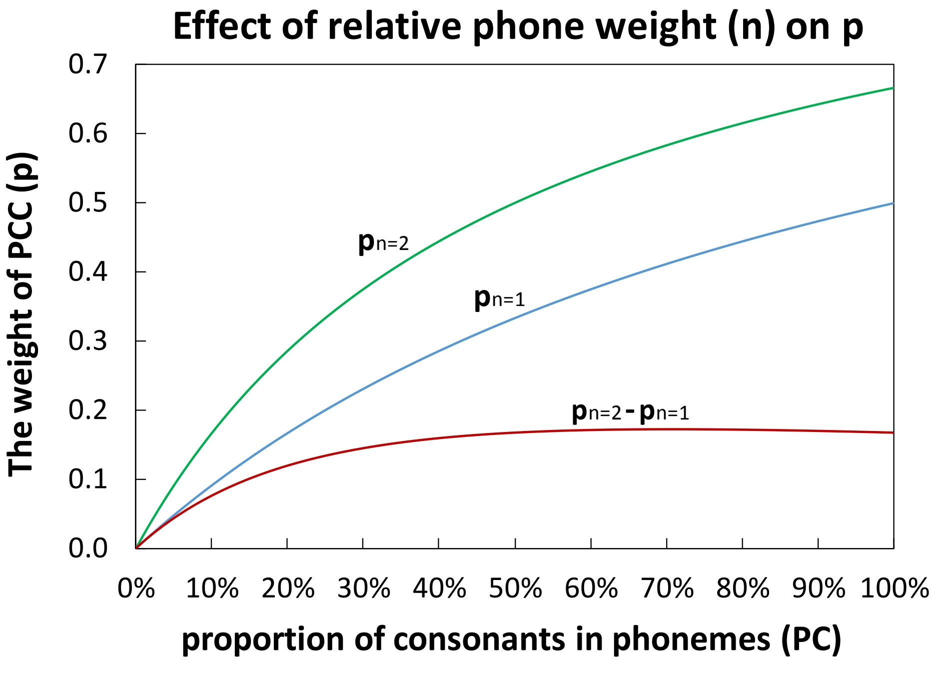


Figure 2. The weight, p, of the proportion of consonants correct (PCC) versus the proportion of consonants in targeted speech, PC, for two different values of the relative weight, n, between correctly produced consonants and all phones. The difference in the weight p versus PC is also shown.

PC values for typical words in speech samples exceed 55% for stress-timed languages, like English, and 50% for syllabic languages, like Greek and Spanish. Typical monosyllabic word samples that are used by one of the authors (D. I.) to differentiate normal from disordered child speech in English have the following PC values: 64% for monosyllabic without consonant clusters, 76% for monosyllabic words with a consonant cluster. The corresponding p values computed from (3) with n=1 are respectively 0.39 and 0.43. Therefore, if PWP for the consonant cluster words is computed using the p corresponding to the words with only singleton consonants, it will only differ from the true PWP by less than 4%. If n=2 instead of n=1 is used in computing p, then its values for the words with only consonant singletons and the words with a consonant cluster are respectively 0.56 and 0.60, resulting again in a very small error for PWP when using the p of the other word category.

Another example is given here using the data obtained by one of the authors (E. B.) from a child's English speech at the age of 3 years. The child's monosyllabic words with only singleton consonants have a PC equal to 59% while the monosyllabic words with a consonant cluster have a PC equal to 71%. The corresponding p values for n=1 are respectively 0.37 and 0.415. For n=2, they are 0.541 and 0.587. Again, the PWP computed using the p of the other word category would differ by an amount from the true PWP that can be neglected. Therefore, for most practical purposes, the conclusions drawn in the first case above, where p was invariant between speech samples, hold true here as well and they will not be repeated.

Third case

In the third case, the change in p will not be ignored across word categories. Ingram (2015) notes that there are cases of children's disordered speech where PCC changes across words with clusters and words without clusters are negligible. For such cases, it will be useful to use such n as to increase the PWP change across the word categories. This PWP change is compared for two arbitrarily chosen values of n. Without going through the algebraic details, use of equation (3) four times, twice for each n to compute the PWP for each word category, results in

$$\Delta PWP_1 - \Delta PWP_2 = -(p_2-p_1)\,\Delta PPD \tag{10}$$

where Δ is the change of the quantity of interest (PWP or PPD) across word categories and the subscript refers to the first or second n used in computing p and PWP. In (10), p may be computed for either category as it will yield the same result. This is because the difference (p_2-p_1) changes negligibly for any changes in PC values larger than about 50%. This may be observed in Figure 2 where p_1 (n=1) and p_2 (n=2) and their difference is plotted for all possible PC values. Comparing (p_2-p_1) values at different PC values larger than about 50%, it is seen that they are practically the same. For example, for PC=50%, p_1=1/3 and p_2=0.5 and, thus, p_2-p_1= 0.167. For PC=75%, p_1= 0.429 and p_2=0.6 and, thus, p_2-p_1= 0.171. The change in the difference (p_2-p_1) is indeed negligible. Derivation of equation (10) is based on this observation.

What does equation (10) imply for practical applications? Without loss of generality, let the subscript 1 refer to the smaller of the two n values chosen. Then p_2-p_1 is positive and for negative ΔPPD (for example PPD for monosyllabic words with only singleton consonants minus PPD for words with a consonant cluster), the left hand side becomes positive. For negative ΔPPD, ΔPWP is positive independent of the n chosen as ΔPCC=0, giving that ΔPWP is larger for the smaller n. Distinguishing PWP between categories of word complexity is sought in practice and, therefore, it is better in such cases, as the one considered here, to use as small an n as possible. Ingram's proposition of n=1 is the smallest integer n that can used for optimal results. Furthermore, equation (10) gives the difference in the change of PWP for two arbitrary values of n, for a given ΔPPD. To get a feeling on the amount that this difference changes for different values of n, it is now computed for PC=60% and n=0.5, n=1, and n=2. ΔPWP for n=0.5 is larger than ΔPWP for n=1 by the amount 0.14 times the absolute value of ΔPPD. In turn, ΔPWP for n=1 is larger than ΔPWP for n=2 by the amount 0.17 times the absolute value of ΔPPD.

Conclusion

Obtaining a formula for phonological word proximity (PWP) for a whole speech sample in terms of the proportion of consonants correct and the proportion of phonemes deleted, made it possible to examine the effect of the relative weight of phones and of the proportion of consonants in the phonemes on: a) the weight of each PWP component individually and in relation to each other, and b) the computed PWP and its sensitivity to measurements across different speech samples, in general, and across categories of word complexity in disordered child speech, in particular. The analysis and formulae given here provide guidelines to practitioners for child speech assessment. It is pointed out, however, that the present work applies mostly to normal or disordered child speech since targeted vowels are considered to be produced correctly when they are produced in context. The present work is being extended to *also* differentiate correct from incorrect vowels when they are produced in context. This will find applications in assessing second language speech where vowel mispronunciation occurs even in L2 learners at advanced levels.

References

Babatsouli, E., Ingram., D., & Sotiropoulos, D. (2011). Phonological word proximity in child speech development. *Manuscript*. December.

Babatsouli, E., Ingram., D., & Sotiropoulos, D. (2014). Phonological word proximity in child speech development. *Chaotic Modeling and Simulation*, *4(3)*, 295-313.

Bunta, F., Fabiano-Smith, L., Goldstein, B. A., & Ingram, D. (2009). Phonological whole- word measures in three-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics and Phonetics*, *23*, 156-175.

Burrows, L., & Goldstein, B. A. (2010). Whole word measures in bilingual children with speech sound disorders. *Clinical Linguistics and Phonetics*, *24*, 357-368.

Ingram, D. (2002). The measurement of whole-word production. *Journal of Child Language*, *29*, 713-733.

Ingram, D. (2015). Whole-word measures: Using the pCC-PWP intersect to distinguish speech delay from speech disorder. In C. Bowen (ed.), *Children's speech sound disorders* (2nd ed., pp. 100-104). Oxford, UK: John Willey & Sons.

Ingram, D., & Ingram, K. (2001). A whole-word approach to phonological analysis and intervention. *Language, Speech and Hearing Services in Schools*, *32*, 271-283.

Macleod, A. A., Laukys, K., & Rvachew, S. (2011). The impact of bilingual language learning on whole-word complexity and segmental accuracy among children aged 18 and 36 months. *International Journal of Speech and Language Pathology*, *13*, 490-499.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Shriberg, L., Austin, D., Lewis, B., McSweeney, J., & Wilson, D. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research*, *3*, 708-722.

Taelman, H., Durieux, G., & Gillis, S. (2005). Notes on Ingram's whole-word measures for phonological development. *Journal of Child Language*, *32*, 391-400.

Multilingualism and acquired neurogenic speech disorders

Martin J. Ball
martin.j.ball@liu.se

Linköping University

Abstract. Acquired neurogenic communication disorders can affect language, speech, or both. Although neurogenic speech disorders have been researched for a considerable time, much of this work has been restricted to a few languages (mainly English, with German, French, Japanese and Chinese also represented). Further, the work has concentrated on monolingual speakers. In this account, I aim to outline the main acquired speech disorders, and give examples of research into multilingual aspects of this topic. The various types of acquired neurogenic speech disorders support a tripartite analysis of normal speech production. Dysarthria (of varying sub-types) is a disorder of the neural pathways and muscle activity: the implementation of the motor plans for speech. Apraxia of speech on the other hand is a disorder of compilation of those motor plans (seen through the fact that novel utterances are disordered, while often formulaic utterances are not). Aphasia (at least when it affects speech rather than just language) manifests as a disorder at the phonological level; for example, paraphasias disrupt the normal ordering of segments, and jargon aphasias affect both speech sound inventories and the link between sound and meaning. I will illustrate examples of various acquired neurogenic speech disorders in multilingual speakers drawn from recent literature. We will conclude by considering an example of jargon aphasia produced by a previously bilingual speaker (that is, bilingual before the acquired neurological damage). This example consists of non-perseverative non-word jargon, produced by a Louisiana French-English bilingual woman with aphasia. The client’s jargon has internal systematicity and these systematic properties show overlaps with both the French and English phonological system and structure. Therefore, while she does not have access to the lexicon of either language, it would seem that she accesses both the French and English phonological systems.

Keywords: multilingualism, acquired neurogenic disorders, aphasia, apraxia of speech, dysarthria

Introduction

Speech disorders (as opposed to language disorders) are generally deemed to fall into several categories, for example: developmental, acquired neurogenic, genetic, results of surgery and other. Ball (2016) describes these types in detail but we will look briefly at each one here in turn. In each case, Ball (2016) provides further details and references.

Developmental

Various sub-types of speech disorder are found under this heading: articulation disorder (e.g., sibilant and rhotic problems, among others), motor speech disorders in children, childhood dysarthria, childhood apraxia of speech, phonological disorder (consisting of phonological delay, phonological deviancy-consistent, and phonological deviancy-inconsistent). See Howard (2013) and Bowen (2015) for further details, and Ball (2016) for discussion of different classifications for child speech disorders.

Genetic

Speech sound disorders with a genetic origin fall into two broad groups: cleft lip and palate and genetic syndromes. Cleft lip and palate can be subdivided into various types depending on which parts of the lip and palate are affected.

Genetic syndromes include Down, Williams, Fragile-X, Noonan, and Cri-du-chat (see Stojanovik, 2013, for references to these and other syndromes).

Results of Surgery

Surgical intervention to treat, for example, cancer can have effects on speech. In particular, we can note laryngectomy – leading to the adoption of esophageal or tracheo-esophageal speech or the use of external devices to produce a noise source, and glossectomy – the partial or total removal of the tongue (see Bressmann, 2013, for further details).

Other

Speech disorders also occur with the following other disorders of communication: hearing impairment (this primarily affects prosody but eventually also has an effect at the segmental level); voice disorders (primarily affects phonation, but may also exhibit problems with resonance and supralaryngeal articulatory settings); disorders of fluency (stuttering and cluttering have a primary effect on prosody, but often also result in problems at the segmental level).

Acquired Neurogenic Disorders

We have left to last this category as it is the main focus of this account, and thus will be described in greater detail than the previous varieties. The main types of acquired neurogenic disorders are: aphasia, apraxia of speech (AoS), and dysarthria. We will look at each of these in turn.

Aphasia

As non-fluent aphasia often co-occurs with Apraxia of speech, we look here at fluent aphasia. Fluent aphasia may show phonemic paraphasias, that is, incorrect phoneme use, or incorrect phoneme placement. So, the disorder is at the phonological level. Examples include ‘pat’ for *cat*; ‘tevilision’ for *television*, ‘fafter’ for *after*. Extreme forms may result in *jargonaphasia,* that is, the production of fluent, connected, but apparently unmonitored speech that is non-comprehensible and often characterized by the use of nonwords (Marshall, 2006).

Apraxia of Speech

This disorder is at the phonetic planning level, and people with AoS may be able to produce formulaic speech with little problem, but novel utterances demonstrate errors. Childhood Apraxia of Speech (CAS) includes a developmental variety, where no discernible neural insult can be found, although the symptoms are similar to the acquired variety.

The impairments include: slow speech rate, distortions to consonants and vowels, prosodic impairments, and inconsistency in errors (Jacks & Robin, 2010). Other features often noted are: articulatory groping, perseverative errors, increasing errors with increasing word length and increasing articulatory complexity, difficulties initiating speech.

Dysarthria

Various sub-types of dysarthria are recognized: flaccid, spastic, hypokinetic, hyperkinetic and ataxic (see Ackermann, Hertrich, & Ziegler, 2013). Dysarthria is a neuromuscular disorder at the level of motor implementation. The different types of dysarthria have differing effects on respiration (commonly short breaths only are possible); phonation (harsh, strained or breathy voice qualities), resonance (hypernasality found in several types), articulation (general imprecision), and prosody (rate may be slow, pauses may be excessive).

Acquired neurogenic speech disorders and multilingualism

Paraphasias and Multilingualism

We consider here only acquired neurogenic speech disorders in bi- and multilingual speakers, rather than cross-linguistic studies. Although there has been considerable research into language impairments in bilinguals with aphasia, there is much less known about acquired speech impairments in such speakers. However, there is work on phonemic paraphasias in bilinguals from South Africa, for example, Odendaal and Van Zyl (2009); Theron, Van der Merwe, Robin and Groenewald (2009); Kendall, Edmonds, Van Zyl, Odendaal, Stein and Van der Merwe (2015).

Odendaal and Van Zyl (2009) collected phonemic paraphasias from three bilingual English/Afrikaans speakers with aphasia. They found similar examples of errors in both languages. Error types were mostly substitutions, then deletions, then additions in both languages, and most errors occurred on high frequency words, again in both languages. Word length played no part in predicting errors, but there were more errors in the speakers' L2 in complex linguistic tasks, but not in simple ones.

Theron et al. (2009) reported that the English-Afrikaans bilinguals using phonemic paraphasias in their study had more difficulty in L2 than in L1. They were more interested in investigating durational features than in comparing the types of paraphasia between languages, however.

Kendall et al. (2015) looked at four Afrikaans/English bilinguals with aphasia and analysed errors in confrontational naming tasks in the two languages. This study is only peripherally relevant for our purposes as many of the errors were semantic (rather than phonological) and little detail is provided on the types of phonological error. Three of the four speakers performed significantly worse in their L2, but that there was little difference in proportion of error types between the speakers' languages.

Bhan and Chitnis (2010) report on a Telugu-English bilingual with subcortical aphasia. Their client produced phonemic paraphasias in both languages, as well as neologisms, semantic paraphasias, circumlocutions, etc. Typical phonemic paraphasias were found in both languages, though the authors do not compare or contrast the aphasic features between the languages.

AoS and Multilingualism

Laganaro and Overton Venet (2014) review the handful of studies into AoS and bi/ multilingual speakers. They describe work on Afrikaans-English bilinguals (Van der Merwe & Tesner, 2000; Theron et al., 2009). These confirm parallel impairments, and increased consequences for the lesser used language noted above. Laganaro and Overton Venet (2014) also report their own study on a Swedish-French bilingual with similar results.

The authors constructed pseudowords of three types: syllable types common to both French and Swedish, syllable types common to French, and syllable types common to Swedish. Further, all categories contained both high and low frequency occurring types. Accuracy was best on high frequency type 1 words; it was worst on low frequency type 1 words.

The observation that frequency of use summed across languages influence accuracy suggests both shared motor plans (i.e., used for both L1 and L2), and common gestural scores used in late bilinguals for common/similar phonological patterns across the speaker's languages. It also supports the importance of frequency of use as expounded in Bybee's (2001) model of a usage-based phonology.

Dysarthria and Multilingualism

Lee and McCann (2009) is one of very few studies on bilinguals with dysarthria. The authors examined the use of phonation therapy with two Mandarin-English bilinguals with flaccid dysarthria. Phonation therapy concentrates on establishing breathing patterns to improve the amount of air flow for speech and, as Mandarin is a tone language, improved phonation is needed to signal tones and thus improve the intelligibility of their spoken output. Indeed, Lee and McCann reported that their clients' intelligibility in Mandarin improved after phonation therapy, and that accuracy of tone production also improved. Intelligibility improvement was minimal in the speakers' English.

The studies reviewed above have highlighted the urgent need for more research in bi- and multilingual speakers with acquired neurogenic disorders of speech. In the next section, we turn to an example of bilingual jargonaphasia as one step towards this goal.

A case of bilingual jargonaphasia

We describe here phonological aspects of a case of bilingual jargonaphasia. This section is closely based on Müller and Mok (2012); the case was also described in Ball and Müller (2015).

Introduction

Perecman and Brown (1981) present a case study of phonemic jargon produced by KS, a man aged 74 at the time of the study, whose first language was German, who had acquired Argentinian Spanish as a second language in his early twenties, and had been a US resident and speaker of American English since his mid-twenties. While the sound inventory of KS's jargon represented "virtually every phoneme of standard English and German" (Perecman & Brown, 1981, p. 185), the frequency distributions in the jargon differed markedly from those in German and English norms. According to Perecman and Brown, KS's vowel inventory and distribution would suggest a German rather than an English vowel system (and, we may note, is suggestive of a Spanish vowel system, as well; Perecman and Brown do not discuss possible Spanish influence on KS's jargon).

Ms H, on whose speech output we report here, was 78 years old at time of data collection, and had experienced a left hemisphere CVA approximately nine months previously. Her first and second languages are French and English, respectively. English was her dominant language premorbidly, as regards frequency and domains of language use. She used French mainly with relatives and close friends of her own generation. Her education, as well as premorbid literacy practices, had been exclusively in English. Her husband is also a French speaker; their four children do not speak French, but have good conversational comprehension.

According to the Western Aphasia Battery (Kertesz, 1982) criteria, Ms H's scores are consistent with a classification of Wernicke's aphasia. Assessment was only carried out in English (as was language therapy), since no speech-language pathologist with sufficient fluency in French and experience in French-language assessment was available.

Sound inventories

The data analysed here represent an opportunistic sample, in that they consist of recordings of language therapy sessions made available to the authors by the clinician working with Ms H (all necessary permissions for the recording and use of the data for research purposes were obtained). Data were transcribed phonetically and below is an example of three attempts by Ms H to repeat the utterance 'hand me the nail polish':

(a) [tyː iː kɑːm *(whispered:* tʰy *)* fœlŭʰ ˈdɛn deˈde *(taps nail polish bottle)* fɒːˈse

(b) [njuː lɪt tʃu feɪˈfjo edeˈfɔʁ n̩fəˈbɛn frenˈdjuʰ ɛtˈfeʰ pəˈselʲ *(2 syllables)* dendru ɛ̃ˌsedeˈdruʰ ˈmijɛlˠ]

(c) [ˈwːʌn ˈtʃu dædˈjudĭ ˈhɔɹt ˌdɹɔfːɑˈbɛs]

Listener impressions recorded anecdotally were that (a) and (b) sounded more French than English, whereas (c) sounded more English. This would seem to derive from the differential use of both specific vowels and consonants in the three utterances, for example, front rounded vowel in (a), a nasal vowel in (b), and examples of approximant [ɹ] in (c).

In order to compare Ms H's speech output with her two premorbid languages, inventories of consonants and vowels were drawn up, as well as lists of syllable types, phonotactic possibilities, and stress patterns. Ms H's mainly used oral monophthongs and, although she did in fact use a number of nasal vowels ([ɛ̃, æ̃, ɔ̃, õ, ũ, ʊ̃]), these occurred much more rarely than their oral counterparts. Diphthongs occurred very rarely too: [aɪ, aʊ] were used twice each, and [eɪ, eə, oʊ] once each. As can be seen in Table 1, Ms H's vowel inventory overlaps with both the French and English vowel inventories, and shows significant rank correlations with both. Of the 29 vowels in Ms H's jargon, 15 are shared with

English, and 14 with French. The Spearman rank correlations between H's vowel inventory and English and French are r_s=0.386 (p=0.0265), and 0.581 (p<0.01) respectively (N=32 for both).

Table 1. Monophthong inventory

Front vowels			Mid	Back vowels	
i	y	ɪ		ʊ	u
e	ø				o
ɛ	œ				ɔ
			ɚ		
			ə		
æ			ʌ		ɒ
					ɑ
			a		

Ms H's consonant inventory is shown in Table 2. This inventory incorporates sounds that map onto virtually all contrastive consonants of English and French, the exception being the French /ɥ/ (also a rare consonant in French). Segments that have no counterpart in either French or English are of marginal occurrence in H's jargon ([ɟ] occurs twice, and [x] and [ç] once). Of the five most frequently occurring categories in H's jargon, [t, s, d, "r", m], three, [t, "r", s] are also in that group in English, and the same three [t, "r", s] in French. There is of course a significant overlap between the French and English consonant systems both as regards the segment inventories and in terms of their relative frequency of occurrence (r_s=0.701; N=25; p<0.01). The Spearman rank correlations between Ms H's consonants and English and French norms are r_s=0.721, and r_s=0.748, respectively (N=28, p<0.01 for both).

Table 2. Consonant inventory

Labial		Apical/Laminal			Dorsal			Glottal
p, b			t, d		ɟ	k, g		
m			n			ŋ		
			r					
			ɾ					
				tʃ, dʒ				
w	f, v	θ, ð	s, z	ʃ, ʒ	ç	x	ʁ	h
			ɹ		j			
			l					

Syllable types were classified as 'string-initial' and 'string-final' (where the term 'string' is employed rather than 'word' as, being jargon, it was not obvious where word boundaries occurred or whether one could sensibly use the term 'word' anyway). String-initially, #C- was used 65.9% of the time, #V- 22.9%, and #CC- 11.2%. String-finally, the percentages were: -V# 63.5%, -C# 33.8%, -CC# 2.7%. The restrictions on numbers of consonants in clusters are reminiscent more of French than of English, nevertheless the actual clusters used (see Müller & Mok, 2012, for details) do cover many English and French possible types.

Syllable counts and stress assignments are shown in Table 3. One- and two-syllable strings together greatly outnumber strings of three or more syllables in the data, with a slight preference for di- over monosyllabic strings (χ^2=12.24; df=2; p<0.05). String-final stress predominates, in both two- and three-syllable strings (strings with four or more syllables do not occur in sufficient numbers to show a clear pattern) and, again, this suggests perhaps a more French than English patterning, and it could well be these prosodic characteristics that contribute to listeners' perceptions of a French accent.

Table 3. Syllable counts and stress assignment

N syllables	**1**	**2**	**3**	**4+**
	39.51%	43.48%	13.85%	3.51%
Stress patterns:				
Two syllables	∑ˈ∑	∑∑ˈ	other*	
	77.36%	14.47%	8.18%	
Three syllables	∑∑ˈ∑	∑ˈ∑∑	ˈ∑∑∑	other*
	66.67%	17.65%	7.84%	7.84%

Comparative notes

The ratio of consonants to vowels in Ms H's output is 56.79% to 43.21%. These figures are very close to those given for French by Wioland (1985), 56.5% consonants and 43.5% vowels, respectively. For English, a somewhat higher proportion of consonant use is reported, 61.44% for consonants, and 30.56% for vowels (Ball & Müller, 2005). The almost complete absence of string-initial three element clusters from our data is interesting, which in English occur in some high-frequency words. Ms H's use of C+[j] clusters intersects with both French and English phonotactics. With the exception of a single string-initial [df], we have found no sound combinations that would violate both English and French phonotactic constraints.

The clear preference for string-final stress is reminiscent of French rhythmical patterns, in which phrase-final stress predominates. Therefore, it is tempting to assume that Ms H's jargon is constructed within a French rhythmical structure. However, phrase-final stress is also encountered in English, most typically in sequences of monosyllabic content words preceded by one or more unstressed function words ('at night', 'in the dark'), so this is not conclusive.

Conclusion

Ms H's jargon shows no systematic relationship between sound and meaning; and some of the few instances of possible real words are doubtful. In addition, no recognizable segmental or syllable-by-syllable relationship can be identified between Ms H's jargon and target words or sentences.

Ms H's jargon does, however, have internal systematicity: There are clear preference patterns in terms of segment frequencies, and sequential properties. The systematic properties of Ms H's output show overlaps with both the French and English phonological system and structure. While Ms H does produce some segments that are not part of either contrastive system, these are marginal in occurrence. Therefore, while she does not have access to the lexicon of either language, she accesses both the French and English phonological systems; her jargon shows a greater overlap with both French and English at the segmental level than French and English with each other.

It is of course quite usual that bilinguals access two language systems simultaneously and produce bilingual speech. It is therefore not surprising that the global difficulties with lexical retrieval of persons producing nonword jargon would lead to output that intersects with both contributing phonologies, but is not separable into discrete chunks of, in our case, “French” or “English” jargon.

Conclusion

The small number of studies available for review in the earlier part of this paper, together with the fascinating results from the case of bilingual jargonaphasia described later, highlight both the need for much more research in multilingualism and acquired neurogenic speech disorders, and the rewards that such research will bring.

References

Ackermann, H., Hertrich, I., & Ziegler, W. (2013). Dysarthria. In J. S. Damico, N. Müller & M. J. Ball (eds.), *The handbook of language and speech disorders* (pp. 362-390). Chichester, UK: Wiley-Blackwell.

Ball, M. J. (2016). *Principles of clinical phonology.* New York: Psychology Press.

Ball, M. J. & Müller, N. (2005). *Phonetics for communication disorders.* Mahwah, NJ: Erlbaum.

Ball, M. J. & Müller, N. (2015). Is there phonology without meaning? Poster presented at Manchester Phonology Conference, 23. University of Manchester, May 2015.

Bhan, S. & Chitnis, S. (2010). Lexical errors in narrative discourse of a bilingual subcortical aphasic. *International Symposium on Bilingual Aphasia at AIISH*, Mysore India 2010.

Bowen, C. (2015). *Children speech sound disorders (2nd edition).* Oxford: Wiley-Blackwell.

Bressmann, T. (2013). Speech disorders related to head and neck cancer: laryngectomy, glossectomy, and velopharyngeal and maxillofacial deficits. In J. S. Damico, N. Müller & M. J. Ball (eds.), *The handbook of language and speech disorders* (pp. 497-526). Chichester: Wiley-Blackwell.

Bybee, J (2001). *Phonology and language use.* Cambridge, UK: Cambridge University Press.

Howard, S. (2013). Children with speech sound disorders. In J. S. Damico, N. Müller & M. J. Ball (eds.), *The handbook of language and speech disorders* (pp. 339-361). Chichester: Wiley-Blackwell.

Jacks, A. & Robin, D. (2010). Apraxia of speech. In J. S. Damico, N. Müller & M. J. Ball (eds.), *The handbook of language and speech disorders* (pp. 391-409). Chichester: Wiley-Blackwell.

Kendall, D., Edmonds, L., Van Zyl, A., Odendaal, I., Stein, M., & Van der Merwe, A. (2015). What can speech production errors tells us about cross-linguistic processing in bilingual aphasia? Evidence from four English/Afrikaans individuals with aphasia. *South African Journal of Communication Disorders, 62,* article #111, 10 pages.

Kertesz, A. (1982). *Western aphasia battery*. New York: Grune & Stratton.

Laganaro, M. & Overton Venet, M. (2014). Apraxia of speech in bilingual speakers as a window into the study of bilingual motor speech control. In N. Miller & A. Lowit, (eds.), *Motor speech disorders: A cross-language perspective* (pp. 93-104). Bristol: Multilingual Matters.

Lee, T. & McCann, C. (2009). A phonation therapy approach for Mandarin-English bilingual clients with dysarthria. *Clinical Linguistics and Phonetics, 23,* 762-779.

Marshall, J. (2006). Jargon aphasia: What have we learned? *Aphasiology, 20,* 387-410.

Müller, N. and Mok, Z. (2012). Non-word jargon produced by a French-English bilingual. In M. Gitterman, M. Goral & L. Obler (eds.), *Aspects of multilingual aphasia* (pp. 224-241). Bristol: Multilingual Matters.

Odendaal, I. & Van Zyl, A. (2009). *Phonemic paraphasia in bilingual speakers with aphasia*. Unpublished dissertation, University of Pretoria, South Africa.

Perecman, E., & Brown, J.W. (1981). Phonemic jargon: A case report. In J. W. Brown (ed.), *Jargonaphasia.* (pp. 177-258). New York: Academic Press.

Stojanovik, V. (2013). Genetic syndromes and communication disorders. In J. S. Damico, N. Müller & M. J. Ball (eds.), *The handbook of language and speech disorders* (pp. 115-130). Chichester: Wiley-Blackwell.

Theron, K., Van der Merwe, A., Robin, D., & Groenewald, E. (2009). Temporal parameters of speech production in bilingual speakers with apraxic or phonemic paraphasic errors. *Aphasiology*, *23*, 557-583.

Van der Merwe, A. & Tesner, H. (2000). Apraxia of speech in a bilingual speaker: Perceptual characteristics and generalisation of non-language specific treatment. *South African Journal of Communication Disorders, 47,* 79-89.

Wioland, F. (1985). *Les structures syllabiques du français.* Genève: Slatkine-Champion.

Same challenges, diverse solutions: Outcomes of a crosslinguistic project in phonological development

Barbara May Bernhardt[1], Joseph Paul Stemberger[2]
bernharb@mail.ubc.ca, joseph.stemberger@ubc.ca

[1]School of Audiology and Speech Sciences, University of British Columbia, [2]Department of Linguistics, University of British Columbia

Abstract. An international crosslinguistic study of phonological development has been in progress since 2006. The study has included 12 languages in 13 countries: Romance (Manitoba French, Granada Spanish, Mexican Spanish, Chilean Spanish, European Portuguese); Germanic (German, Canadian English, Swedish, Icelandic); Semitic (Kuwaiti Arabic); Asian (Mandarin, Japanese); South Slavic (Slovene, Bulgarian). Assessment tools have been created for five additional languages or dialects: Anishinaabemowin (Algonquian, First Nations, Canada); Brazilian Portuguese, European French (Romance); Punjabi (Indo-Aryan); Tagalog (Austronesian). The objectives of the overall study are: (1) to document and account for similarities and differences between languages in typical and protracted phonological development in terms of word structure, consonants, vowels and features; and (2) to develop assessment tools for speech elicitation and analysis plus treatment activity suggestions across languages for speech-language therapists within the various countries and in Canada, with its highly diverse multilingual society. The current paper focuses primarily on the first objective, discussing aspects of diversity in methodology and outcomes from both practical and theoretical perspectives. Methodological diversity concerns primarily sampling characteristics (word list creation, data collection, age groups, degree of protracted phonological development) and transcription conventions. To enhance data comparability across languages, word lists are created (approximately 100 words) that reflect the major phonological characteristics of the language in words familiar to children (from monosyllabic to multisyllabic words). Native speakers collect the data after basic instruction from the project leaders, using digital audio recordings with the same or similar recording devices. All data are entered into Phon, a phonological analysis program, and double-checked for accuracy. To resolve between-language sampling differences, groups are matched on age and severity in terms of relevant parameters for analysis. To enhance reliability of transcription, native speakers and project leaders build transcription convention documents through discussion and consensus-building transcription practice. Relative to diversity in results, two aspects of within-language diversity are discussed: (a) the importance of considering dialectal influences when examining variable pronunciations across children (Shanghai Mandarin, Granada Spanish); and (b) the relevance of considering interactions between word structure features in apparently variable phonological patterns (Manitoba French). Relative to between-language diversity, differences in phonological inventories and frequencies across languages are discussed, focusing on comparisons of fricative production in German, English and Icelandic. The paper concludes with a review of considerations for future crosslinguistic research, both for the current and other projects. For clinical and multilingual applications, links to project resources are provided.

Keywords: crosslinguistic, methodology, transcription, multilingual, word structure, features

Introduction

Canada, although officially only a bilingual English-French society, has over 56 native languages (one being Inuktitut, an official language in the territory of Nunavut) and at last 90 immigrant languages. Speech-language pathologists are increasingly challenged to provide relevant speech assessments and treatment as the linguistic diversity expands. Assessment tools are not readily available, nor is there much information on phonological development in languages other than English. Thus, a crosslinguistic project in phonological development (both typical and protracted) was undertaken, starting in 2006. Canadian federal funding was obtained (Social Sciences and Humanities Research Council of Canada, grant numbers 410-2009-0348 and 611-2012-0164) in order to conduct the international project and begin to address the identified gaps.

The project thus had an initial practical and clinical motivation. However, other major objectives concerning the project were both theoretical and data-oriented. Over the past century, language researchers have analysed patterns across and within languages in order to determine the degree of similarity/universality of language systems. For the developing human being, this question concerns the relative influence of environment versus innateness in language acquisition. Regarding many aspects of language, typically developing children appear to have a similar maturational timetable across languages (e.g., Berko-Gleason & Bernstein Ratner, 2009; Jakobson, 1941/1968). However, researchers have also noted differences that reflect the language being learned (e.g., Ingram, 2012; Pye, Ingram, & List, 1987). What appears to be complex and late-acquired in one language may be early-acquired in another, depending on the frequency of the element and its functionality in the language. Crosslinguistic data are crucial for addressing this fundamental question about language, yet very few studies concerning one aspect of language development, phonological development, have been conducted using equivalent methodology across languages. The overarching objective of the crosslinguistic study is therefore to inform linguistic theories further on the questions of universality and diversity in phonological development, through analyses of child speech data from a variety of languages.

For ‘adult’ phonology, researchers have investigated both the structure of words (sequences of consonants and vowels, syllable formation, word stress) and segments (consonants/vowels) and their composite features across a wide range of languages (e.g., de Lacy, 2007). However, a comprehensive crosslinguistic comparison of word structure, segments and their interactions remains to be done for developing phonological systems, specifically taking into account the relative levels of complexity within each language. Within-language analyses in the project have revealed strong interactions between the stress and structure of the syllable and the mismatch patterns observed (e.g., Chávez-Peón et al., 2012), e.g., unstressed syllables showing many more and different mismatch patterns than stressed syllables. However, to address the main question of the proposed study concerning universality, additional within- and between-language comparisons need to be done that can take into account language type and relative complexities. Thus, a second major objective of the proposed study has been to conduct within- and between-language comparisons of segments, word structures and their interactions, both to learn more about various aspects of developing phonological systems and to address the question of universality.

While expecting both similarities and differences in developmental data across languages, diversity in sampling and transcription conventions during the project necessitated considerations of a number of methodological variables in addition to interpretation of differences in results. The following paper provides an overview of the diversity observed in methodology and results, and the solutions and interpretations applied. The major purposes of the current paper are thus to provide an overview of the study and initial results, to suggest strategies that may assist in future crosslinguistic research and to provide a link to resources that may be useful both clinically and for research. The paper begins with general methodology for the project, including solutions to the similar challenges in diversity of sampling and transcription. The subsequent section addresses within-language diversity, focusing on considerations of dialectal influence (Mandarin, Granada Spanish) on children's pronunciations, and the interaction of word structure and features in interpreting variable developmental patterns. The third section outlines similarities and differences in German, English and Icelandic fricative production by children with protracted phonological development, reflecting an influence of inventories and frequency within the languages. Discussion is woven throughout the paper, rather than included in a separate section. The paper explores implications of the current study for future research, and provides links to project resources for clinical and research purposes.

Methodology

The following section outlines the general methodology for the study in terms of data collection and analysis. Subsequent sections outline two of the common ways that methodology diverges across languages: sampling differences and transcription conventions. Finally, the project's solutions to those differences are presented and discussed.

General methods

The objective across languages is to collect data from 20-30 monolingual children aged 3-6 years and designated locally as having protracted phonological development (PPD). Exclusion criteria include sensorineural hearing loss, severe chronic otitis media, major language comprehension or cognitive delays and major orofacial anomalies. Limited sentence production is not an exclusionary criterion. In addition, if local funds permit, data are also collected from typically developing (TD) children matched for age and dialect area to the PPD cohort. A native speaker tests the children in a quiet room, usually in a preschool centre, in a 45-60 minute session. The data are digitally recorded with a high quality audio-recording device (most often with a M-Audio Microtrack II digital recorder and a Sennheiser remote system, i.e. transmitter EK 100 G2 and receiver SK 100 G2, with Countryman remote lapel microphones). If possible, video recordings are also made. For each language, an approximately 100-word list is developed for photo elicitation; project leaders and local investigators choose words familiar to children that cover all segments of the language (both consonants and vowels, at least twice across word positions) and the major word structures (stress patterns, CV sequences). Ten words are elicited three times each using either pictures or objects as a warm-up and to assess within-word variability later. Additional testing includes a hearing screening and a language comprehension test (where available), a short spontaneous language sample, and a one-page parental questionnaire about the child’s development and language use. Following development of a transcription-conventions document in conjunction with the local team and project leaders, a native speaker of the dialect area transcribes the sample, with reliability of transcription confirmed with a second native speaker or project leader experienced with transcription. (Further discussion of transcription follows below.) Data are then entered into Phon, a free phonological program for entry and analysis of phonological development data (http://childes.psy.cmu.edu/phon/), either locally or in the first author's laboratory at the University of British Columbia. Data are double-checked for accuracy of adult targets and alignment. Exportation of the data into spreadsheets allows data analysis in addition to what is provided by the phonological analysis program.

As can be seen above, the objective is to provide a common methodology across languages with similar (1) sample characteristics, (2) word list types, (3) data collection procedures, (4) transcription conventions and (5) data entry and analysis procedures. While the objective is generally achieved for items (2), (3) and (5), items (1) and (4) are more challenging. They are addressed in more detail below.

Sampling characteristics

As with all human-subject research, research is constrained by the availability of participants and by the time and resources available. The recruitment process can be slow for any number of legitimate reasons. In the end, we celebrate the participants that do agree to be part of our project, whether they match the numbers we desire or are at the preferred ages or developmental levels. Many language teams were able to find the target number of participants in the designated age range, including matched control groups of TD children, but not all. Furthermore, there were small variations in the number of children by age, reflecting the recruitment challenges. Relative to degree of PPD, preliminary data analysis showed that degree of PPD might not be equivalent across all languages. A Whole Word Match (WWM) was calculated for the various samples, showing ranges from less than 12% average across the cohort with PPD (English, Icelandic) to about 40% for Mandarin and Granada Spanish. (Whole Word Match means that the child’s pronunciation of a word matches the adult target, or is considered “close enough” by adult native speakers, with slight deviations in place or voicing ignored; Schmitt, Howard, & Schmitt, 1983.) The first question was whether the word lists from different languages were too different to compare, because the Germanic languages are more complex in word structure than Mandarin and Spanish, even for words familiar to children (e.g., more clusters, codas). However, looking at data for TD 4-year-olds, the WWM was 80% for Kuwaiti Arabic (with complex phonology) and Mandarin, and 85% for the Spanish-speaking children, i.e. virtually equivalent. Differences across languages in sampling characteristics for the cohorts with PPD more likely reflect variation in who was designated as having PPD. For the researcher, the question is how to resolve this kind of discrepancy. The discrepancy is not particularly important for within-language analyses, except when comparing experimental and control groups. The WWM scores did allow evaluation of the PPD classification, when there were matched groups of control participants included. For example, in the Granada Spanish cohort, one child originally designated as PPD had a WWM similar to the age-matched controls, and thus was reassigned to that cohort. Overall, however, the original classification as TD vs PPD was consistent with the WWM scores. Regarding between-language comparisons, discrepancies within groups are important. Thus, before analysis, matched samples are selected by age, gender (where possible) and various global phonological measures, e.g., WWM or Percent Consonants Correct (PCC) scores. If global phonological variables match, then validity of more specific between-language comparison is enhanced. Participant matching reduces the sample sizes and power, but provides more assurance that interpretation is based on phonological differences rather than sampling differences.

Transcription conventions

As is well-documented, achieving transcription agreement is difficult even for transcribers who speak the same language and have the same amount and type of training. This is particularly true for child speech or more 'disordered' speech, and if the transcribers use narrower transcription (Shriberg & Lof, 1991). Consensus-building activities and practice can enhance reliability, as can acoustic analysis, e.g., examining VOT, frication, formants, nasality, or duration (Bernhardt & Stemberger, 2012). Some of the differences and solutions for broad transcription conventions for reliability have been as follows:

1. Transcribers differ on whether to include or exclude a predictable word-initial glottal stop when not phonemic in the language. The convention of the local transcription team was followed. For analysis, even transcribed word-initial glottal stops were counted as deletions if a target consonant was missing, but identified as glottal stops for future considerations.
2. Standard local transcription preferences are generally followed, but are taken into account for cross-linguistic comparisons. For example, in Spanish, what is transcribed as palatoalveolar [t͡ʃ] (implying retracted tongue tip) is actually alveolopalatal [t͡ɕ] (with advanced tongue tip; Kochetov & Colantoni, 2011), but is nevertheless transcribed as [t͡ʃ]. In Icelandic, what is transcribed as palatal [c] is actually a fronted velar [k̟] (Árnason, 2011), but is transcribed as [c]. For Swedish, consonant length in adult speech is predictable from vowel length (Schaeffler & Wretling, 2003), and is therefore not transcribed.

Regarding narrow transcription, a number of differences arose during transcription comparisons and were resolved as follows:

1. For Granada Spanish, a number of conventions apply for allophonic variation. Because lenition of voiced stops (e.g., /b/ as [β]) is expected intervocalically, and the fricative/ approximant is not phonemically different from a stop that the child might use instead, the native speaker transcribers did not always note this allophonic difference in the transcript. A non-native transcriber verified stop vs continuant allophones spectrographically, and adjusted the adult target appropriately. That way, the data were maintained for future use, but no mismatch was indicated for the child, in agreement with the initial transcription. Similarly, when coda /s/ deletes in Granada Spanish, a small [h] or short [h]-like element can replace it (Martínez Celdrán & Fernández Planas, 2007). The second transcriber verified presence of aspiration on the spectrogram, and included the [h] in both the adult and child forms, but the difference is not considered clinically relevant. For both Granada Spanish and Icelandic, voiced “fricatives” are generally more approximant-like (Martínez Celdrán & Fernández Planas, 2007; Árnason, 2011). However, they ware transcribed using fricative symbols, in accordance with local transcription conventions.
2. Sometimes children appear to epenthesize a small vocalic element either after a coda consonant or between cluster elements. Native adult speakers may also have short transitional elements in certain clusters that are ignored as irrelevant. Generally, epenthetic “vowels” shorter than 40 milliseconds are examined acoustically by a second transcriber and written as superscripts (e.g., [ˈgᵊris]), with vowels of 60 msec or longer written on the line as vowels

(e.g., [gəˈris]). This arbitrary convention is currently being verified in an ongoing study of clusters.

3. A child's vocal tract anatomy generally includes a high flattish tongue body in a proportionally small (by adult standards) oral cavity. Thus, sibilants such as /s/ are often de-grooved or flattened, and the tongue body is high enough to cause frication along a longer axis than in adult speech. For the study, an attempt was made to distinguish such productions from [s]/[z] by transcribing them as [s$^{\theta}$]/[z^{ð}] (if dentalized with a flattish (but not fully flat) tongue and a low-enough tongue body), or alveolopalatal [ɕ]/[ʑ] (if ungrooved with a too-high tongue body).

Although full agreement is unlikely across languages, the reliability of transcription has been enhanced by this method. For the Granada Spanish, after consensus-building, agreement for segments was high: 96% for TD samples and 93.6% for PPD samples. What has become clear during the consensus-building process is that most transcribers hear and agree on most aspects of the child’s pronunciation, when brought to their attention. One difference across languages is that different things are considered irrelevant for transcription, because of differing local conventions or assumptions about importance for adult or clinical populations. The team learns together how to agree on what matters, what symbols to use, and how to interpret the symbols. As a final note, the development of transcription conventions includes identification of multiple acceptable adult targets in the dialect area, so that children are not penalized for “nonstandard” pronunciations learned from adult input. We return to such variation in adult speech in the next section.

Within-language results: Diversity and methods of analysis

Above we alluded to one of the key considerations for within-language analysis: what the adult target might be. This has proven to be relatively challenging for all of the languages. Even well-trained native speakers can be unaware of how their idiolect reflects or does not reflect the local dialect vs the standard. For the study, speech samples have been collected from at least one adult from the dialect area, but with too few speakers to adequately reflect adult variation. In addition, the teams have consulted literature on the dialect. Through meetings with the local teams, time is spent in determining what the range of possibilities for the adult targets might be. For Granada Spanish, data were collected from three adults from the same area. Not only did use of [s] (*seseo*) vs [θ] (*ceceo*) differ across speakers, but also within speakers. Other coda consonants can also be fully present, reduced, or absent. As noted, coda [s] may be replaced by [h] or [ʰ], with a laxing of the vowel quality: e.g., *dos* ‘two’ [dos]~[dɔh]~[dɔʰ]. Coda [l] and [r] can interchange, or delete, in the case of the medial consonant, with gemination of the following consonant: e.g., *alma* 'soul' [alma/aɾma/amːa] or *sarten* 'frying pan' [saɾˈten/salˈtẽ̞/saˈtːẽ̞] (note also /n/ deletion with vowel nasalization). The child’s pronunciation may be based on different adult targets, and all variants must be noted in the adult target set.

A more complex example is given below concerning Mandarin (also known as Putonghua). Mandarin is the most widespread language in China. Children start learning Mandarin at preschool around age 3, if they are not already speaking it natively. The phonetic inventory of standard Mandarin is presented in Table 1.

Coronal fricatives and affricates are a major component of the Mandarin consonant inventory, with contrasts in tongue-tip placement and height of the tongue body: dentals (with advanced tongue tip and low tongue body), alveolopalatals (with advanced tongue tip and high tongue body), and retroflexes (with retracted tongue tip). The laryngeal contrast is between voiceless unaspirated vs voiceless aspirated stops and affricates.

Data were collected from a group of 30 TD 4-year-olds in Shanghai, with an average of WWM of 80%. Several of the low frequency mismatch patterns provide insight into the dialect question (see Table 2 for the number of children showing various mismatches).

Table 1. Standard Mandarin phonetic inventory.

	Labial	Dental	Retroflex	Alveopalatal	Palatal	Dorsal
Stops	p pʰ	t tʰ				k kʰ
Affricates		ts tsʰ	tʂ tʂʰ	tɕ tɕʰ		
Fricatives	f	s	ʂ	ɕ		x
Nasals	m	n				(ŋ)
Approximants	w	l	ɻ		j	

Because Mandarin has a large number of articulatorily complex coronal fricatives and affricates, it is tempting to say that these may just be late-developing. Some children may still be learning the rhotic consonant and vowel, as well as other coronals; there are some dorsal-coronal interchanges, suggesting further that the retroflex issue is developmental. However, Shanghainese (the local language spoken in Shanghai before the introduction of Mandarin) does not have retroflex sibilants, and Shanghainese-influenced Mandarin may lack retroflexes. Some Shanghainese Mandarin speakers, particularly young adult women (potentially the mothers of the children in this centre) (Starr & Juraksy, 2004) are showing a hyper-correction phenomenon, using retroflexes where they are not in the adult Mandarin target. Some children may be exposed to adult pronunciations with these nonstandard sibilant productions, and this must be taken into account when evaluating their phonological development.

Table 2. Numbers of typically developing Mandarin-speaking 4-year-olds showing one or more mismatches by category.

	Vowel/tone mismatches		Consonant mismatch categories					
	Vowels	Tones	[-]>[+] retroflex	[+]>[-] retroflex	/ɹ/, /ɚ/ (various)	tɕ < > ɕ tɕ,ɕ < > s	Dorsal < > Coronal	Other
# Ss/30	4	13	13	20	19	8	3	11

Further to Table 2, there are a few children showing tone or vowel mismatches, even at age 4. Acoustic analysis of the tone trajectories for Shanghainese Mandarin shows a smaller pitch range for the tones in comparison with the Beijing standard (Lai et al., 2011). Shanghainese has less extreme pitch changes in tones than Mandarin. The children's Mandarin may be Shanghainese-influenced, which may lead to mistranscription of tones, if transcription is based on standard Mandarin. Additionally, the Shanghainese adult vowel space is smaller than that of adult Beijing Mandarin; the children's vowel space is smaller yet, and most likely is influenced by Shanghainese in this regard.

The point of this discussion is to stress the importance of always having normative data from adults of the dialect area, especially of the ages, genders, and occupations that tend to have close contact with preschool children. In this particular case, Mandarin may be a second language for many adults, who are likely to speak with a regional variant.

Word-structure segment interactions

Another source of variability in within-language data is the interaction between word structure and features. In data collected for Canadian (Manitoba) French, there were clear interactions between syllable stress and segmental production. In disyllabic words, /l/ and fricatives were significantly more accurate in stressed than unstressed syllables, as were nasal vowels. In multisyllabic words, all segments were more accurate in stressed than unstressed syllables (Bérubé et al., 2012). Thus, data analysis must necessarily evaluate variability in terms of syllable stress, word length, and syllable position.

Between-language diversity: Fricatives in German, English and Icelandic

Moving from within-language analysis to between-language analysis brings a whole new set of challenges in addition to the ones identified above. Assuming that all the transcription conventions are in place and that the data are reliable in terms of the transcription and of the identification of the adult target variant, the first thing to consider is whether the languages are similar enough to be compared (for the particular characteristics being focused on) and whether the samples are sufficiently matched to allow for comparison.

The first comparison in our crosslinguistic study was of German and English, two closely related languages (West Germanic) from the same family (Indo-European). A collaboration with Angela Ullrich in 2006 led to the launching of the whole crosslinguistic project. In comparing the data from German-speaking TD and PPD children (Ullrich, 2004) with English-speaking children's data, differences were noted suggesting that the language inventory and frequency of phonemes were relevant both in the timeline for mastery of phonemes and in the types of mismatch patterns that occur.

Because the German data for this crosslinguistic study were gathered in Cologne over a 2-year period, sufficient subjects were found to match the composition of the English sample relative to proportion of 3-, 4- and 5-year-olds, of gender and, to a certain extent, of the severity of PPD (English mean WWM = 12%, German mean WWM = 19%). Thus, comparisons between the whole German and English sample of children with PPD (30 children each) are considered valid (Bernhardt, Romonath, & Stemberger, 2014). In contrast, for the Icelandic-English comparison, because the population of Iceland is so much smaller than that of Canada, it was impossible to match the two samples for WWM. It was necessary to find a matched subset of subjects, resulting in a sample of thirteen 3-year-olds and ten 4-year-olds from each language (Bernhardt et al., 2015). The following analysis highlights some commonalities between these two studies, providing additional information relative to fricative production in these three Germanic languages, as an example of similar challenges and diverse solutions.

The original motivation for examining fricatives was the observation that, in terms of segmental development, preschoolers with PPD often show a fairly high match proportion for stops, nasals and glides, but lower match proportions for and later mastery of fricatives. Thus, the fricatives, with their more complex articulatory characteristics, provide sufficient data for comparison across languages, especially in terms of mismatch types. However, the three languages do not have the same fricative inventories, and we needed to consider what could be compared (and what could not). The solution was to pick fricatives common to the languages being compared, *matched for position in the word*, and only in stress-initial words (avoiding any effects of word prominence, as noted in the previous section, and taking into account that all words in Icelandic are stress-initial). Here we focus only on word-initial position, because all three languages showed an accuracy level of about 38% for fricatives in that position.

Based on complexity and universal patterns of development, certain similarities among languages were predicted: (1) manner would be most affected, with stops as common substitutions across the languages; (2) /f/ would show the highest match levels (due to clearly observable visual cues); (3) grooving for sibilants would show the lowest match levels (with ungrooved substitutions), and the relatively infrequent interdental /θ/ would be less advanced; (4) voiceless fricatives would have higher match levels than voiced fricatives (for German and English, which have voiced fricatives, but not for Icelandic, which has approximants rather than voiced fricatives).

There were also a number of differences predicted for the different languages: (1) German voiced fricatives and /ʃ/ would show higher match levels word-initially than the English ones, due to their higher frequency in German; (2) for mismatch patterns, more palatal substitutions were expected in

German and Icelandic, and more [θ] substitutions in English and Icelandic, reflecting their phonetic inventory; (3) affricates were considered more likely substitutions in German ([ts]) and English ([tʃ] and [dʒ]) than in Icelandic (which has no affricates). Icelandic's rich inventory of [+spread glottis] segments ([h], pre-aspirated stops, post-aspirated stops, voiceless fricatives, voiceless sonorants) were expected to show up in substitutions in fricative mismatch patterns, but only [h], post-aspirated stops and voiceless fricatives were expected in German and English substitution patterns. Relative match levels for the various fricatives are shown in Table 3.

Table 3. Match levels for word-initial fricatives in German, English and Icelandic preschoolers with protracted phonological development.

% match levels	50-65%	35-49%	20-34%	11-19%	6-10%
German	f v		z ʃ		
English		f z	v	θ	ʃ s
Icelandic		f s	θ		

Results support some but not all of the predictions about relative match levels. The highest match was shown for /f/ across languages, but the sibilants varied more than expected across languages. As predicted, German /ʃ/ was more advanced than its English counterpart. Contrary to expectation, English /z/ had a higher match level than the more frequent German /z/, and Icelandic /s/ had a match level equivalent to that of /f/. Simplicity and transparency were thus relevant for /f/ across languages, while frequency played a role for /v/ and /ʃ/ in German (higher feature frequency of [+labiodental] for both /f/ and /v/), but not for /z/. The interdental was slightly better than the sibilants in English, but not in Icelandic. Voiceless fricatives fared better in German and Icelandic as predicted, but not in English.

In terms of mismatch patterns, manner substitutions were relatively common, as predicted, but place substitutions were also common in German (which showed more other-fricative substitutions than in English). Other confirmed similarities were the appearance of ungrooved coronal fricatives for grooved ones, with the German children showing a relatively high proportion of lateralized fricatives (one solution to the 'grooving' challenge for tongue control, but less common in English and Icelandic). Inventory effects were noted: (1) palatal fricatives appeared more frequently in Icelandic and German than in English; (2) affricates appeared as substitutions in English and German but rarely in Icelandic; (3) although preaspirated stops and voiceless sonorants were rare substitutions in Icelandic, such substitutions did not appear at all in German and English, and there was a higher proportion of [h] substitutions than in German and English. Unexpectedly, [θ] often appeared as a substitution for /f/ in Icelandic, which is quite rare in English, but [f] appeared as a common substitution for /θ/ in Icelandic as in English (where it has often been reported; e.g., Bernhardt and Stemberger, 1998). For the three languages, 80% of the substitutions in Icelandic were within-inventory substitutions, 71% in English, and only 58% in German (due to the high frequency of lateralized fricatives). Solutions to grooving can lead to non-inventory substitutions, but there is a bias towards within-inventory substitutions.

Conclusion

The purpose of this paper is to demonstrate some of the challenges facing a crosslinguistic project in children's phonological development, to illustrate some of the solutions that we have developed given the complexities of international collaborations and differences among children, cultures, methods and languages, and to give a taste of the results coming from the project. The more languages and countries, the greater the diversity, but the richer the data when the challenges of methodology can be at least tempered by discussion, consensus-building, adult dialect definition and sample matching. Even the early results of the project as presented here speak to the complex interaction of articulatory complexity and language frequency as influences in phonological development. Some expectations are met, e.g., that unstressed syllables are weak prosodic domains where features may fail to surface, that the bulk of substitutions come from within the language's phonetic inventory, and that /f/ is an early fricative across at least the Germanic languages. But there are also creative solutions. If the tongue just cannot be grooved, a variety of fricatives or affricates might provide at least a partial match with the target.

Following the defined path of comparing languages from the same language family (as was done for fricatives in the Germanic language), a study of word-initial rhotic and lateral clusters is underway with seven languages that have apical taps and trills. In the near future, we plan to produce a volume of unusual case profiles in each of the languages in the project. A long-term project is analysis of word-structure development across the languages, because such data are relatively rare in the literature. Clearly, there are more data in this project than a team of over 60 people can possibly analyse. Thus, as much as possible, the data will be shared with Phonbank, in order that others might delve ever more extensively into questions about phonological development.

As a final note, another major objective of the project has been to provide materials for speech-language therapists and researchers. The website phonodevelopment.sites.olt.ubc.ca has free assessment materials in the various languages, tutorials about transcription, phonology and scan analysis for intervention planning and, last but not least, a set of fun-ological activities for clinical intervention as examples of what can be done within the world of word structure and speech sounds (so far in English, French, Slovenian, and Spanish).

References

Árnason, K. (2011). *The phonology of Icelandic and Faroese*. Oxford, UK: Oxford University Press.

Berko Gleason, J., & Bernstein Ratner, N. (2009). *The development of language* (7th ed.) Boston: Pearson/Allyn & Bacon.

Bernhardt, B. H., & Stemberger, J. P. (1998). *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. San Diego, CA: Academic Press.

Bernhardt, B.M. & Stemberger, J.P. (2012). Translation to practice: Transcription of the speech of multilingual children. In B. Goldstein & S. McLeod (eds.), *Multilingual aspects of speech sound disorders in children* (pp.182-190). Bristol, UK. Multilingual Matters.

Bernhardt, B. M., Másdóttir, T., Stemberger, J. P., Leonhardt, L., & Hansson, G. Ó. (2015). Fricative acquisition in English- and Icelandic-speaking preschoolers with protracted phonological development. *Clinical Linguistics and Phonetics*, *29*, 642-665.

Bernhardt, B. M., Romonath, R., & Stemberger, J. P. (2014). A comparison of fricative acquisition in German and Canadian English-speaking children with protracted phonological development. In M. Yavaş, (Ed.), *Unusual productions in phonology: Universals and language-specific considerations* (pp. 102–127). New York: Psychology Press.

Bérubé, D., Bernhardt, B.M., Stemberger, J.P. & Bacsfalvi, P. (2012). Nonlinear phonological analysis of the speech of Manitoba French-speaking preschoolers with protracted phonological development. Paper presented at *the International Conference of Clinical Phonetics and Linguistics*, June 27-30, Cork, Ireland.

Chávez-Peón, M. E., Bernhardt, B. M., Adler-Bock, M., Ávila, C., Carballo, G., Fresneda, D., Lleó, C., Mendoza, E., Perez, D., & J Stemberger, J. P. (2012). A Spanish pilot investigation for a crosslinguistic study in protracted phonological development. *Clinical Linguistics and Phonetics*, *26*, 255-2722.

de Lacy, P. (2007). *The Cambridge handbook of phonology*. Cambridge, UK: Cambridge University Press.

Ingram, D. (2012). Prologue: Cross-linguistic and multilingual aspects of speech sound disorders in children. In S. McLeod & B. Goldstein (eds.), *Multilingual aspects of speech sound disorders in children* (pp. 3-12). Bristol, UK: Multilingual Matters.

Jakobson, R. (1941/1968). *Child language, phonological universals and aphasia*. (Keiler, A. Trans.). The Hague: Mouton. Original (1941): *Kindersprache, aphasie und allgemeine Lautgesetze*.

Kochetov, A., & Colantoni, L. (2011). Coronal place contrasts in Argentine and Cuban Spanish: An electropalatographic study. *Journal of the International Phonetic Association*, *41*, 313-342.

Lai, Y., Bernhardt, B. M., Zhao, J. & Stemberger, J. P. (2011). Acoustic analysis of Mandarin tone in 4-year-old typically developing children in Shanghai. Poster presented at the York *Child Phonology Conference* June 15-17, 2011, York, UK.

Martínez Celdrán, E., & Fernández Planas, A. M. (2007). *Manual de fonética española: Articulaciones y sonidos del español*. Barcelona: Ariel.

Pye, C., Ingram, D., & List, H. (1987). A comparison of initial consonant acquisition in English and Quiché. In K. Nelson & A. van Kleeck (eds.), *Children's language*, vol. 6 (pp. 175-190). Hillsdale, NJ: Erlbaum.

Schaeffler, F. & Wretling, P. (2003). Towards a typology of phonological quantity in the dialects of Modern Swedish. In D. Recasens (ed.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS;* pp. 2697-2700). Barcelona, Spain: Universita autònoma de Barcelona.

Schmitt, L. S., Howard, B. H., & Schmitt, J. F. (1983). Conversational speech sampling in the assessment of articulation proficiency. *Language, Speech, and Hearing Services in Schools*, *14*, 210-214.

Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics*, *5*, 225-279.

Starr, R. & Jurafsky, D. (2004). Phonological variation in Shanghai Mandarin. Paper presented at the New Ways of Evaluating Variation (NWAV)-33 conference, October 3, 2004, Ann Arbor, Michigan.

Ullrich, A. (2004). *Nichtlineare phonologische Analyse deutschsprachiger Kinder*. [Nonlinear phonological analysis of German-speaking children]. Unpublished Master's thesis, Würzburg, Germany: Julius Maximilians Universität.

Non-native perception of English voiceless stops

Angelica Carlet[1], Anabela Rato[2]
angelica.carlet@uab.cat, asrato@ilch.uminho.pt

[1]Universitat Autònoma de Barcelona, [2]Universidade do Minho

Abstract. Previous research has suggested that Romance language learners of English are presented with a challenge when perceiving English stop consonant sounds in a native-like manner (e.g., Aliaga-García & Mora, 2009; Flege & Eefting, 1988). This cross-language difficulty may be related to the interplay of several factors, such as L1 attunement and/or L2 experience (Flege, Munro, & MacKay, 1995). The present study further investigated how the first language (L1) and language experience affect non-native perception of the English voiceless stops /p, t, k/. Specifically, the study aimed at investigating whether Catalan and Portuguese learners of English could identify and discriminate between aspirated and unaspirated voiceless stops, since in both Romance languages aspiration is a non-existent phonetic property (Ladefoged, 1972). Two groups of upper-intermediate learners of English (L1 Catalan (*n*=22) and L1 European Portuguese (*n*=19)) and a group of advanced Portuguese learners of English (*n*=22) performed two perception tasks: forced-choice identification and AX categorical discrimination. The testing stimuli consisted of word-initial voiceless aspirated stops (e.g., *pot*, *tool*, *key*) and word-initial consonant clusters after /s/ (e.g., *spot*, *stool*, *ski*) naturally produced by three British English talkers. The target (C)CVC words were embedded in the carrier phrase "This (target word)" so that the aspirated and unaspirated stop allophones would be presented in similar contexts (e.g., *This pot-This spot*). Findings show that the more advanced learners obtained significantly higher scores than the upper-intermediate learners, which seems to indicate an effect of language experience, in line with previous studies (e.g., Flege, Munro & Skelton, 1992). Moreover, despite the comparable VOT patterns in their L1s and the similar language proficiency in the target language, the upper-intermediate Portuguese group outperformed the upper-intermediate Catalan learners in both identification and discrimination tasks. This might be a result of greater daily exposure to the target language through outside classroom input (e.g., non-dubbing on TV), which may contribute to their FL learning success (Rubio & Lirola, 2010). All these findings taken together seem to indicate that quantity and possibly quality of L2 input in different L1 linguistic environments influence non-native speech perception, and L2 language experience promotes accurate L2 allophonic speech perception.

Keywords: L2 speech perception, English voiceless stops, language experience, L1 attunement

Introduction

Acquiring novel L2 speech sounds at an adult age tends to be a challenging task due to perceptual and production difficulties which result from the interplay of different factors such as L1 attunement and L2 experience (Flege, Munro, & MacKay, 1995). Second language perceivers tend to identify and discriminate non-native speech sounds with reference to the linguistic categories of their first language (Pisoni, 1982) and, according to L2 speech learning models (Best & Tyler 2007; Flege, 1995), degree of cross-language phonetic (dis)similarity tends to predict perceptual ease or difficulty. Specifically, the *Speech Learning Model* (SLM, Flege, 1995/2003) predicts that the more dissimilar a sound is in comparison to the learner's first language (L1), the easier the acquisition and category formation will be. Conversely, if the L2 sound is an allophone of an L1 sound (i.e. perceptually equivalent to an L1 sound) the less likely the establishment of a new category will occur. The *Perceptual Assimilation Model* (PAM, Best, 1995; PAM-L2, Best & Tyler, 2007) accounts for different patterns in the perceptual assimilation of non-native speech contrasts. If two non-native phones are perceived as exemplars of two different native phonemes (TC – "two category"), discrimination is expected to be excellent; conversely, poor discrimination is predicted if two non-native sounds are perceived as equally good or poor instances of the same native phoneme (SG – "single category"). Another case occurs when two non-native phones are heard as instances of the same native phoneme but one fits the L1 category better than the other (CG –"category-goodness"). Moderate to fairly good discrimination is predicted for the latter case, but not as good as in the TC assimilation. The perception of L2 allophonic variation may be therefore a case of CG assimilation since two L2 phones are assimilated to a single L1 phonological category, but one is considered a deviant examplar.

Another important factor in non-native speech acquisition is L2 experience, which refers to the amount of use and exposure to the target language. Acording to Flege (1991), such experience with the L2 is essential to improve one´s ability to differentiate native from non-native sounds, and the greater the amount of exposure to the target language, a more native-like acquisition may take place.

In the stop consonant domain, several studies have previously reported the difficulty to perceive English stop consonants in a native-like manner on the part of Romance language learners of English (e.g., Aliaga-García & Mora, 2009; Flege & Eefting, 1988). One of the reasons may lie in the fact that the perception of an allophonic contrast in complementary distribution such as the English aspirated-unaspirated voiceless stops tends to be more difficult and therefore less accurate than the perception of a phonemic contrast (e.g., Boomershine, Hall, Hume, & Johnson, 2008; Celata, 2009; Whalen, Best, & Irwin, 1997).

This study investigates how a first language (L1) and language experience affect non-native perception of the English voiceless stops (/p, t, k/). Specifically, the study examined whether Catalan and Portuguese learners of English were able to identify and discriminate between aspirated and unaspirated English voiceless stop contrasts, which consist of two context-dependent phones of one and the same phonological category (Celata, 2009).

In Standard Southern British English (SSBE), which is the target language of the present study, the voiceless stops /p, t, k/ are allophones in complementary distribution, being realized either as aspirated stops (in the onset of a stressed syllable) or as unaspirated stops (following the phoneme /s/, among other cases). The aspirated stops are produced with a long-lag mode, in which the voicing onset occurs substantially after the release. This voicing delay results in a VOT (voice onset time) of 30 ms or longer, corresponding to the aspiration interval (Cho & Ladefoged, 1999; Lisker & Abramson, 1964).

In Romance languages, on the other hand, the phonemes /p, t, k/ are always realized as unaspirated voiceless stops. They are produced with a short-lag mode, in which the onset of voicing coincides with the release of the stop closure, resulting in VOT values that are nearly zero. According to Lisker and Abramson (1964), the voiceless stops in Spanish tend to be produced with VOTs between zero and 10 ms. Similarly, Andrade (1980) reported VOTs ranging from zero to 30 ms for Portuguese voiceless stops produced in isolated words. Aspiration is thus a non-existent phonetic property in their L1s (Ladefoged, 1971) and learning the English long-lag VOT patterns is challenging for these learners (Alves & Zimmerman, 2015; Fullana & MacKay, 2008).

Taking into account the issues discussed above and the differences between the consonant inventories of the participants' L1 and the target language, this study sought to answer two questions and attest the corresponding hypotheses:

Q1. Does the **first language (L1)** affect the perception of non-native English voiceless stops /p, t, k/?

H1. The **L1** will not significantly affect the perception of the non-native voiceless stops due to the high degree of similarity between the consonant sound systems of the learners' L1s (Portuguese (EP) and Catalan (Cat)). Due to the fact that the allophonic contrast between aspirated and unaspirated voiceless stops does not exist in the learners' L1s, both EP and Cat perceivers will have difficulty distinguishing the target sounds.

Q2. Does **language experience** influence the perception of non-native English voiceless stops /p, t, k/?

H2. **Language experience** will play a role on the perception of the allophonic contrast (aspirated-unaspirated voiceless stops). Advanced learners of L2 English will discriminate and identify the target sounds better than the upper-intermediate learners.

Method

Participants

Sixty-three learners of English took part in the present study and were divided into three experimental groups: (i) 22 native speakers of Catalan with an upper-intermediate level of English; (ii) 19 native speakers of European Portuguese (EP) with an upper-intermediate level of English; and (iii) 22 native speakers of European Portuguese (EP) with an advanced level of English. The upper-intermediate Catalan and upper-intermediate Portuguese subjects were first-year English majors at the *Universitat Autònoma de Barcelona* and at *Universidade do Minho*, respectively. The advanced Portuguese learners were second-year English majors at the latter institution. The characteristics of the participants of each group can be seen in Table 1.

Table 1. Participants' characteristics by group

Group	Catalan (Upper-int.)	Portuguese (Upper-int.)	Portuguese (Advanced)
Age Mean (SD)	20.05 yrs. (4.03)	19.11yrs. (1.04)	21.91 yrs. (5.75)
Sex (F=female, M= male)	16F, 6M	15F, 4M	14F, 8M
Self-reported daily usage of English Mean % (SD)	28.82 (8.93)	14.21 (10.17)	27.73 (11.52)

Moreover, three English native speakers took part in the study by validating the stimuli and providing baseline data.

Stimuli

The testing audio stimuli were natural recordings of (C)CVC words with word-initial voiceless aspirated stops (/p, t, k/) and word-initial consonant clusters (/sp, st, sk/) embedded in the carrier phrase "This (target word)", so that the aspirated and unaspirated stop allophones would be aurally presented in similar contexts. The target stimuli included nine paired noun-phrases contrasting the aspirated-unaspirated English voiceless stop /p/ (e.g., *This pan-This span; This pot-This spot*), nine paired noun-phrases contrasting /t/ (e.g., *This tool-This stool; This table-This stable*), and six paired noun-phrases contrasting /k/ (e.g., *This can-This scan; This key-This ski).* Testing stimuli were elicited from three native British-English speakers by means of a phrase-reading task recorded with a Sony PCM-D50 portable digital recorder in a quiet room. Each speaker read each phrase twice so that the best tokens could be chosen for the perception test. All instances were closely monitored by one of the researchers.

Procedure and tasks

The perception tests were administered in quiet computer laboratory rooms with individual computers and headphones. L1 Catalan participants were tested at the *Universidad Autònoma de Barcelona* and the L1 Portuguese participants were tested at *Universidade do Minho*. After completing the language background questionnaire, the participants performed two perception tasks which are described in detail below. The overall duration of the testing session was approximately 25-30 minutes and the learners were given course credit for their participation.

Three native English speakers performed both tasks and obtained very high percentages of correct identification and discrimination (>95%), indicating that the testing stimuli were appropriately representative of each category tested.

Each participant performed two different tests, namely a categorical AX discrimination task and a 2AFC (alternative forced-choice) identification task. The perception tasks were set up in *TP* v. 3.1. (Rato, Rauber, Kluge, & Santos, 2015) and the order of both tasks and stimulus presentation was randomized.

The categorical discrimination task (CDT, Flege, Munro, & Fox, 1994) adopted in the present study was an AX type, having *same* and *different* trials and two different talkers within each trial. Subjects were presented with two subsequent stimuli (e.g., *This pot-This spot)* and had to decide whether they were being presented with two different allophonic realizations of the voiceless stop consonant or if the two stimuli consisted of the same allophonic realization of the stop consonant sound. Participants responded by clicking on the answers "same" or "different" (and they could listen to the same trial twice). There were a total of 108 trials, being 54 "same" trials and 54 "different" trials counterbalanced for each target allophonic contrast. Figure 1 exemplifies the AX discrimination task.

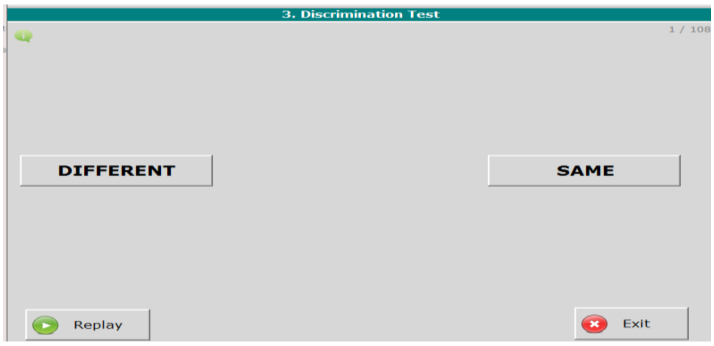


Figure 1. The AX discrimination task

In the two-alternative forced-choice identification tasks, subjects heard one single stimulus (e.g., *This pot*) and were asked to answer by labelling the noun-phase they heard. The response options were *"This + k"* and *"This + sk"* for the stimuli containing the velar voiceless stops; *This + t"* and *"This +St"* for the stimuli containing the alveolar voiceless stops; and *"This + p"* and *"This + sp"* for the stimuli containing the bilabial voiceless stops. There were a total of 162 trials, 54 per each voiceless stop consonant contrast (aspirated-unaspirated). Figure 2 exemplifies the 2 AFC identification task.

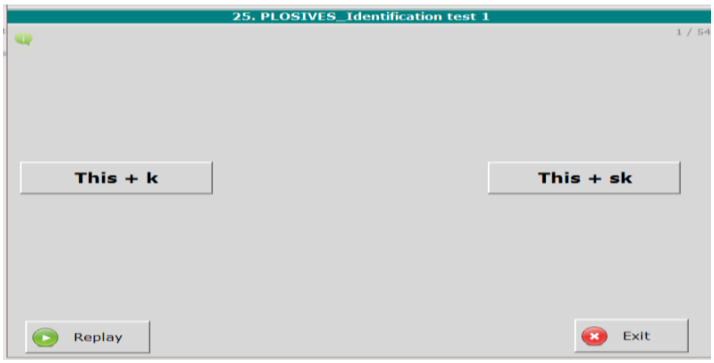


Figure 2. The 2AFC Identification task

Results and Discussion

The participants' perception of the target stop consonant sounds was assessed by calculating the correct percentage obtained in the two perception tests, namely the identification task (ID) and the categorical discrimination task (CDT). The results concerning the effect of L1 in consonant perception will be presented first, followed by the results on the influence of L2 experience.

First language (L1)

First language effect was assessed by comparing the two upper-intermediate groups (Portuguese (EP) and Catalan (Cat)). We had initially hypothesized that L1 would not be a significant predictor affecting the perception of the non-native voiceless stops due to the high degree of similarity between the consonant sound systems of the learners' L1s. Thus, both EP and Cat perceivers were expected to have similar difficulties distinguishing the target allophonic sounds.

A mixed-design 2X2 ANOVA exploring the effect of group as between-subject factor and task as a within-subject factor yielded a significant effect of task, $F(1, 39) = 134.061$, $p<.001$, no group per task interaction, $F(1, 39) = .054$, $p>.05$, and a significant main effect of group, $F(1,39) = 8.050$, $p<.01$. The effect of task and no interaction can be explained by the fact that the scores on the identification test were higher than the discrimination scores for both groups. The effect of group corresponds to the fact that the Portuguese learners outperformed the Catalan learners in both tasks.

A follow up one-way ANOVA revealed a significant effect of L1 on the discrimination ($F(1,39)=5.642$ $p<.05$) and identification of the target sounds ($F(1,39)=6.276$, $p<.05$). The Portuguese learners had a significantly better perceptual performance than the Catalan group in the discrimination (M=65.25, *SD*=5.06; M=60.99; *SD*=6.26) and identification (M=73.32, *SD*=5.56; M=71.61, *SD*=6.34) of the voiceless stop contrasts. Figure 3 shows the participants' performance in each task.

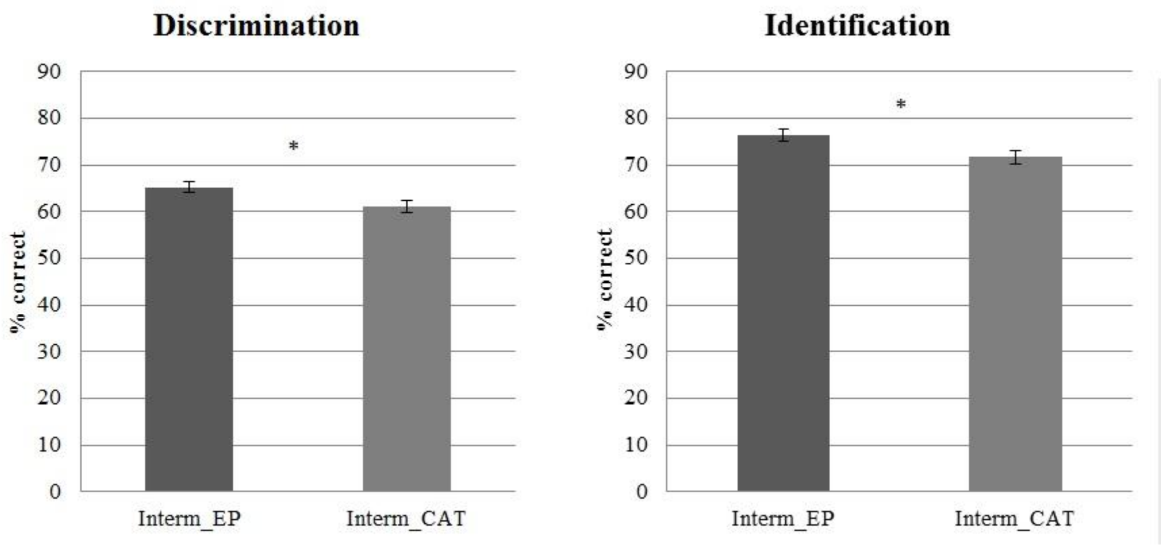


Figure 3. Mean percentage of correct scores in the discrimination and identification tasks

Overall, as observed in Figure 3, identifying the target sounds was less difficult than discriminating them for both groups of L2 learners. This may be explained by the fact that different tasks involve different mechanisms of short-memory. In the categorical discrimination task, which allows listeners to compare two stimuli in the auditory sensory memory, allophones were more difficult to distinguish than in the identification task, in which perceivers can rely on pre-existing mental representations of the target sounds. Contrary to Celata (2009), the allophonic effect seemed to emerge only in the discrimination paradigm impeding L2 listeners from successfully distinguishing the two phones of the English allophonic contrasts. One of the reasons that may explain this difficulty is the fact that the target stimuli were presented in noun-phrase tokens whose length may have negatively interfered on listeners' attention to focus on the between-category acoustic differences.

To further examine the learners' perceptual performance in the identification of both aspirated and unaspirated voiceless stops, a one-way ANOVA was run with a between-subjects design. The results showed a significant effect of the learners' L1 on the identification of the target aspirated voiceless stops ($F(1,39)=11.090$, $p<.01$), and no effect on the unaspirated voiceless stops ($F(1,39)=.424$, $p>.05$), as presented in Figure 4.

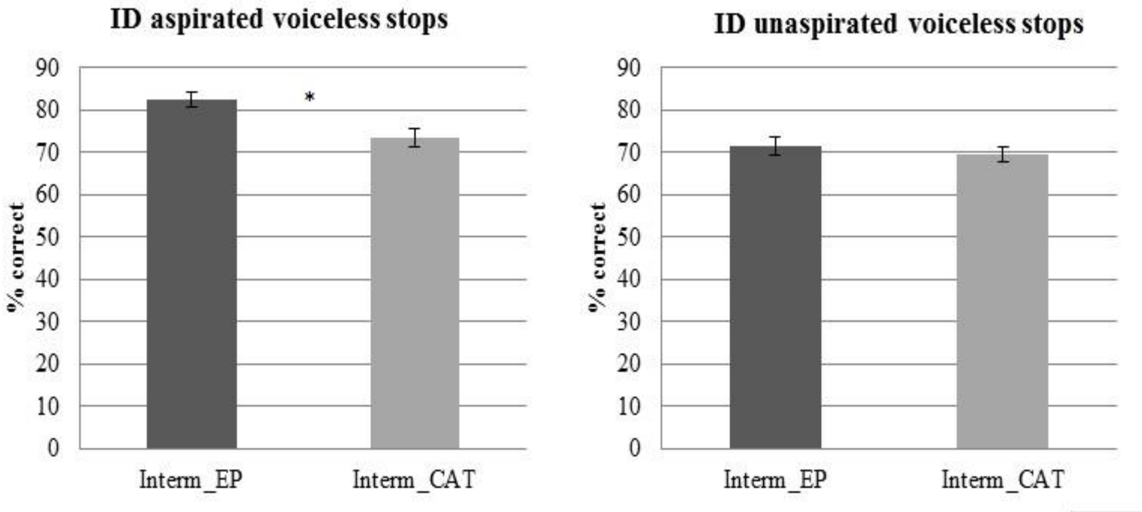


Figure 4. Mean percentage of correct scores in the identification of the aspirated and unaspirated voiceless stops

Although both groups had higher correct scores on the identification of the non-existent aspirated allophones [p^h, t^h, k^h] in their L1s than on the identification of the existent L1 unaspirated phonemes [p, t, k], the Portuguese learners performed significantly better (M=82.65, SD=1.88) than the Catalan learners (M =73.51, SD=1.75). The intergroup difference arose in the case of the aspirated L2 allophones because both Portuguese and Catalan phonological systems only include the unaspirated voiceless stops and thus no differences were expected to be found for this group of L1 consonants. Taking into account PAM-L2, the aspirated allophones may be considered the deviant phones, for which new L2 categories were predicted to be established. The better performance on the deviant phones suggest that these allophonic variants already coexist in the learners' phonological system.

Further follow-up statistical analysis of the performance in the categorization of each aspirated stop revealed a significant L1 effect on the identification of two aspirated voiceless stops: [t^h] ($F(1,39)=7.227$ p<05, and [k^h], ($F(1,39)=10.504$ p<01). Portuguese learners identified the alveolar aspirated voiceless stop (M=84.02, SD=8.91) and the velar aspirated voiceless stop (M=82.46, SD=9.06) significantly better than the Catalan speakers did (M=74.24, SD=13.50; M=71.21, SD=12.47, respectively). Regarding the bilabial aspirated stop, the perceptual performance of Catalan learners (M=75.08, SD=12.27) did not differ significantly from the performance of the Portuguese participants (M=81.48, SD=11.25). This may be due to the fact that the perceptual scores obtained on the identification of the bilabial voiceless stop were higher than the ones obtained for the other two target sounds, which made the intergroup difference smaller.

As observed in Figure 5, Portuguese learners outperformed Catalan learners in the identification of two of the target allophones. This may be explained by the fact that other variables, such as language exposure may have also played a role. Portuguese learners have greater daily exposure to the target language (outside classroom input, e.g., non-dubbing on television allows Portuguese EFL learners to watch films and programs in the original version), which promotes FL learning success (Rubio & Lirola, 2010).

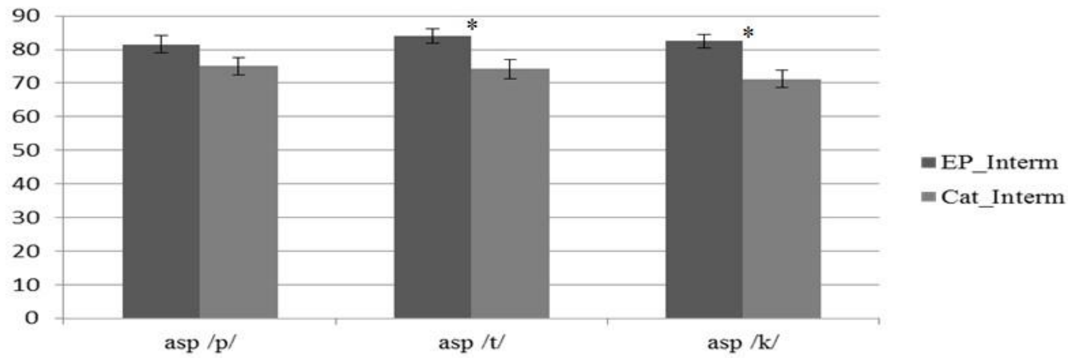


Figure 5. Mean percentage of correct scores in the identification of the aspirated voiceless stops

Language experience

The effect of language experience was examined by comparing the two groups of Portuguese learners: the upper-intermediate and the advanced groups. We hypothesized that language experience would have a positive effect on the perception of the English allophonic contrast. Due to the larger amount of L2 exposure, the advanced learners of English were expected to both discriminate and identify the target sounds better than the upper-intermediate learners.

A mixed-design 2X2 ANOVA exploring the effect of group as between-subject factor and task as a within-subject factor yielded a significant effect of task, $F(1,39) = 131.878$, $p<.001$, no group per task interaction, $F(1,39) = .288$ $p>.05$, and a significant main effect of group, $F(1,39) = 6.900$, $p<.05$. The interaction of group per task revealed no significant effect due to higher identification scores in comparison with the discrimination scores for both groups. The main effect of group is explained by the outperformance of the advanced learners in both tasks. Follow-up one-way ANOVAs revealed a significant effect of language experience on the discrimination ($F(1,39) = 5.650$ $p<.05$) and identification of the target sounds (($F(1,39) = 5.687$, $p<.05$). Figure 6 shows the learners' performance in each task.

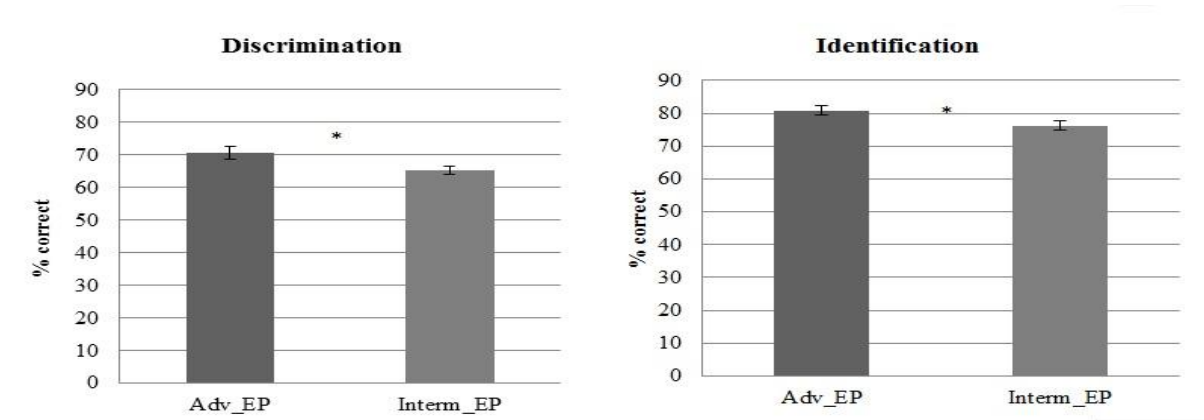


Figure 6. Mean percentage of correct scores in the identification and discrimination tasks

In order to further assess the effect of language experience on the identification of the stop consonants, follow-up one-way ANOVAs were conducted, having segment as the within variable and group as the between variable. The results yielded a significant effect of group on the identification of the aspirated voiceless bilabial stop only, ($F(1,39)=4.923$ $p<.05$, as presented in Figure 7.

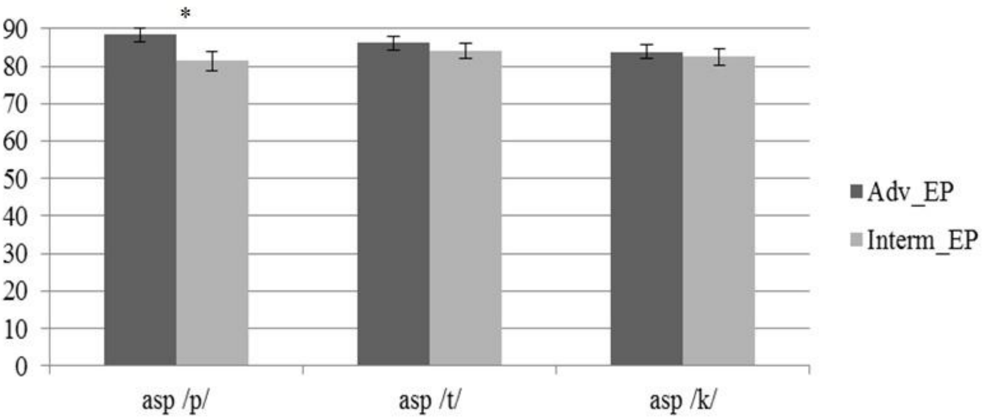


Figure 7. Mean percentage of correct scores in the identification of the three aspirated voiceless stops

Although the group of advanced learners obtained slightly higher numerical scores (*M*=88.39, *SD*=8.65 for [pʰ], *M*=86.21, *SD*=8.78 for [tʰ], *M*=83.86, *SD*=8.81 for [kʰ]) than the upper-intermediate group (*M*=81.48, *SD*=11.25 for [pʰ], *M*=84.02, *SD*=8.91 for [tʰ], *M*=82.46, *SD*=9.06 for [kʰ]) in the identification of the three target allophonic consonants, a significant difference was only found in the identification of the bilabial L2 allophone, with the advanced group outperforming the upper-intermediate learners.

Language experience positively affected the perceptual performance of the Portuguese learners with an advanced proficiency level of English, which suggest that both quantity and quality of L2 input and L2 use are factors that contribute to an improvement in L2 speech learning, as shown in previous studies (Flege, Munro & Skelton, 1992; Flege et al.,1995). However, since these two groups only differed in one-year L2 experience, the advanced learners' outperformance was not pervasive. Nonetheless, the high accurate identification scores reported for the L2 allophones (ranging from 84 to 88%) seem to suggest that the group with more language experience was successfully able to perceive the English allophonic contrast (aspirated-unaspirated), leading to the establishment of new phonetic categories, according to PAM-L2 (Best & Tyler, 2007).

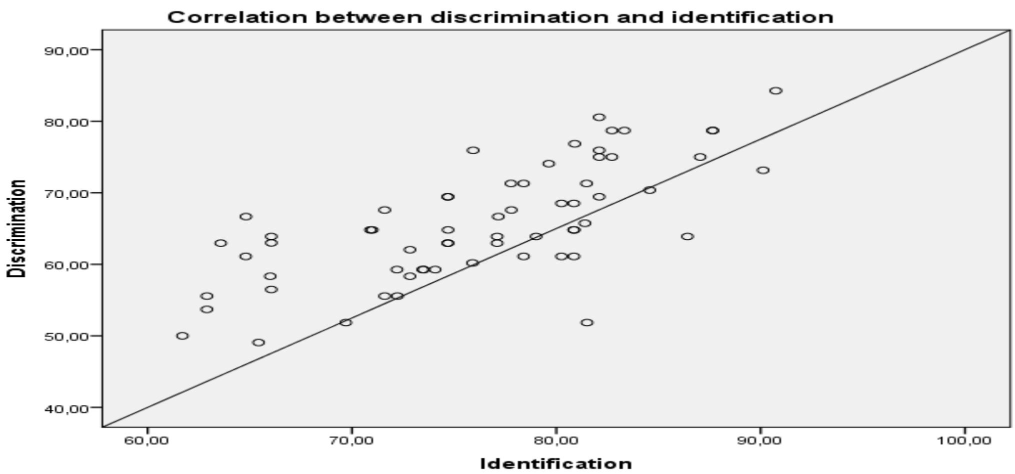


Figure 8. Correlation between identification and discrimination scores

Finally, in order to verify whether there was a relation between the performance of the three experimental groups in the identification and discrimination tasks, a *Pearson* correlation test was run. The result revealed a significant positive correlation between the perceptual performance of the three groups in both the AX discrimination task and the 2AFC identification task, (*r*=.706, *p* <.01). This indicates that the ability to identify and discriminate stop consonant sounds was positively correlated, as it can be observed in Figure 8. Moreover, the correlation increased with language experience. The two less experienced groups (i.e. upper-intermediate Portuguese and Catalan learners) data resulted in moderate *r* values (r=.430, p< .05; r=.513, p< .01, respectively) and the more experienced group (i.e. advanced Portuguese learners) resulted in a stronger *r* value (r=.741, p <.01).

Conclusion

The findings of this experimental study partially confirm the effect of the two investigated variables - L1 influence and language experience - on L2 speech perception. More specifically, the results have shown that the more advanced learners obtained significantly higher scores than the upper-intermediate learners which seems to indicate an effect of language experience, in line with previous studies (Flege et al., 1992). Moreover, despite the comparable VOT patterns in their L1s and the similar amount of exposure to the target language, the upper-intermediate Portuguese group outperformed the upper-intermediate Catalan learners in both tasks, and particularly in the identification of two new allophones (aspirated /t/ and /k/) which may be a result of greater exposure to the target language.

These findings seem to indicate that quantity and possibly quality of L2 input in different L1 linguistic environments influence non-native speech perception and L2 language experience promotes accurate L2 allophonic speech perception.

Limitations and future directions of study

The current study is not without limitations and they suggest interesting directions for future research. First, language experience was assessed by comparing a group of upper-intermediate Portuguese leaners and a group of Portuguese advanced learners. Further investigation should be carried out with additional low proficiency groups and an added advanced level Catalan group of learners, in order to provide a complete picture regarding the effect of language experience and L1 on the perception of non-native stop consonants.

Second, despite the similarities of the stop consonant inventories of the two populations tested, a difference in quantity and quality of target language input seem to have arisen between the two groups and influenced the outcome of the present study, therefore a more controlled language proficiency assessment (e.g., a vocabulary size test) would be ideally administered in a future study. Having a more fine-grained proficiency measure for each group could shed some light on some of the questions that remain unanswered.

This study was not set out to investigate production but due to the differences in the rhythmic patterns of the L1s involved we predict L1 may significantly affect the production of the target contrast. From classroom observation, and by comparing the Portuguese and Catalan groups of learners, we realised that on one hand, the intermediate learners of both languages have difficulty in producing the non-native contrast aspirated-unaspirated (they do not tend to produce aspirated stop consonants), so no effect of L1 initially seems to play a role. Moreover, the advanced learners (after explicit phonetic (articulatory) instruction) tend to improve the production of the target contrast. However, the Catalan learners tend to insert an epenthetic vowel preceding the s-clusters (unaspirated stops) whereas the Portuguese learners do not insert such epenthetic vowel, due to the different rhythmic patterns of the L1s. Therefore, further studies investigating L2 production of voiceless stops by Catalan and Portuguese learners are needed.

Acknowledgments: This research was supported by a research grant from the Spanish ministry of economy and competitiveness (FFI2013-46354-P) and by a grant from the Catalan government (2014SGR61) to the first author**.**

References

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In A. M. A. Watkins, A. S. Rauber, & B. O. Baptista (eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 2-31). Newcastle upon Tyne: Cambridge Scholars Publishing.

Alves, U., & Zimmer, M. (2015). Perception and production of English VOT patterns by Brazilian learners - the role of multiple acoustic cues in a DST perspective. *Alfa*, *59*, 155-175.

Andrade, A. (1980). *Estudos experimentais aerodinâmicos, acústicos e palatográficos do vozeamento nas consoantes*. Lisbon: CLUL

Best, C. (1995). A direct realist view of cross-language speech perception. In A W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. J. Munro (eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13-34). Amsterdam: John Benjamins Publishing.

Boomershine, A., Currie Hall, K., Hume, E., & Johnson, K. (2008). The influence of allophony vs. contrast on perception: The case of Spanish and English. In P. Avery, B. Dresher, & K. Rice (eds.), *Contrast in phonology: Perception and acquisition* (pp. 145-171). Berlin: Mouton.

Celata, C. (2009). The impact of allophonic variation on L2 speech perception. In A. M. A. Watkins, A. S. Rauber, & B. O. Baptista (eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 64-80). Newcastle upon Tyne: Cambridge Scholars Publishing.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of phonetics*, *27*(2), 207-229.

Flege, J. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*(1), 395-411.

Flege, J.E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233-277). Timonium: York Press.

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, *83*(2), 729-740.

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English/t/–/d/contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America*, *92*(1), 128-143.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *The Journal of the Acoustical Society of America*, *95*(6), 3623-3641.

Flege, J., Munro, M., & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*(5), 3125-3134.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437-470.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (eds.), *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

Fullana, N., & MacKay, I. R. (2008). I said made (mate?): Catalan/Spanish bilinguals' production of English word-final obstruents. *Canadian Acoustics*, *36*(3), 118-119.

Fullana, N. & Mora, J. C. (2009). Production and perception of voicing contrasts in English word-final obstruents: Assessing the effects of experience and starting age. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 97-117). Newcastle upon Tyne: Cambridge Scholars Publishing.

Ladefoged, P. (1972). Phonological features and their phonetic correlates. *Journal of the International Phonetic Association*, *2*(01), 2-12.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.

Pisoni, D. B. (1982). Perception of speech: The human listener as a cognitive interface. *Speech Technology*, *1*(2), 10-23.

Rato, A., Rauber, A. S., Kluge, D. C., & Santos, G. R. (2015). Designing speech perception tasks with TP. In J. A. Mompean & J. Fouz-González (eds.), *Investigating English pronunciation: Trends and directions* (pp. 295-313). Basingstoke, UK/New York, NY: Palgrave Macmillan.

Rubio, F., & Lirola, M. M. (2010). English as a foreign language in the EU: Preliminary analysis of the difference in proficiency levels among the member states. *European Journal of Language Policy*, *2*(1), 23-39.

Whalen, D. H., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English/p/allophones. *Journal of Phonetics*, *25*(4), 501-528.

Exploring the voice onset time of Spanish learners of Mandarin

Man-ni Chu, Yu-duo Lin
mannichu@gmail.com, dodo1014x@gmail.com

Graduate Institute of Cross-Cultural studies, Fu Jen Catholic University

Abstract. Voice onset time (VOT) has been examined across languages in plenty of studies (e.g., Lisker & Abramson, 1964), especially its process of acquisition. Several models, Contrastive Analysis Hypothesis (CAH, Lado, 1957), Perceptual Assimilation Model (PAM, Best, 1995), and Speech Learning Model (SLM, Flege, 1995, 1999, 2002) were reviewed to examine whether the VOTs' performance in L1 could influence acquisition in L2. Thus, fifteen Spanish natives learning Mandarin in Taiwan were recruited to participate in 2 perception and 1 production experiments. Several univariate (one way ANOVAs) and ANCOVAs were performed to examine the correct rate (perception) and the mean VOT values (production). The results showed that "Vowel" and "Place of Articulation" play important roles, and most of these Spanish-Mandarin learners made significant progress after several months of immersion in a Mandarin society. That the adjacent high/low vowels affect our subjects' perceptual discrimination can be explained by the vocal-fold tension of the high vowel /i/ (Higgins, Netsell, & Schulte, 1988; Cheng, 2013). The aspirated voiceless stops were perceived and produced significantly differently from unaspirated ones because they were easy for Spanish-Mandarin learners to learn, supporting CAH and PAM. To conclude, the environment where L2 learners stay would influence their progression in L2. The characteristics of preceding vowels and the nature of the stops could also be significant in the acquisition of aspirated/ unaspirated stops.

Keywords: voice onset time (VOT), perception, production

Introduction

Standard Chinese (SC), spoken as a lingua franca in Mainland China and Taiwan, has been known to have five vowels (/a/, /ə/, /i/, /u/ and /y/) and nineteen consonants (/p/, /pʰ/, /t/, /tʰ/, /k/, /kʰ/, /ts/, /tsʰ/, /tʂ/, /tʂʰ/, /m/, /n/, (/ŋ/), /f/, /s/, /ʂ/, /ʐ/, /x/ and /l/) (Duanmu, 2007). The syllable structure is $(C_1)^{(G)}V(C_2)$[1], where only /ŋ/ is not allowed to allot the $C_{1,}$ and C_2 only allows /n/ and /ŋ/. In this study, SC spoken in Taiwan is represented as Taiwan Mandarin (TM). On the other hand, Spanish spoken in most Southern American countries is noted as Caribbean Spanish (CS). CS has five vowels (/a/, /e/, /i/, /o/ and /u/) and twenty consonants (/p/, /b/, /t/, /d/, /k/, /g/, /x/, /f/, /θ/, /s/, /tʃ/, /ʝ/, /m/, /n/, /ɲ/, /r/, /ɾ/, /l/, /ʎ/ and (/ʃ/)) (Salcedo, 2010). The maximum Spanish syllable structure is $(C_1)(C_2)(S_1)V(S_2)(C_3)(C_4)$, where S stands for semi-vowels. The difference between the CS and TM is that the latter has two extra vowels, /y/ and /ə/. As for the stops, CS has voiceless/voiced counterparts, while TM has aspirated/unaspirated voiceless ones. The difference between aspirated and unaspirated stops lies on the duration of the voice onset time (VOT), which signifies a short postponement, measured from the beginning of a released burst (explosion) to the point that the vocal folds start to vibrate. Table 1 shows the mean VOT values of stops in TM and CS.

Lado (1957) proposes Contrastive Analysis Hypothesis (CAH) to claim that L2 learners are inclined to transfer meanings or forms from the primary language to the subsequent language. In other words, if the L2 patterns were similar to those of L1, learners might experience positive transfer, which meant that learning would be facilitated leading to a relatively easy acquisition. In addition, the more different the patterns of L2 are, the more difficult it would be for learners to acquire L2, due to negative transfer. More specifically in phonetic and phonological learning, the Perceptual Assimilation Model (PAM) proposed by Best (1995) claims that the similarity of the sounds between

[1] However, there are some other opinions, e.g., Lin (2007) believes that the syllable structure should be C_1GVC_2, where glide (G) occupies a time slot.

L1 and L2 could assist learning. The higher the degree of gestural similarity between native and non-native sounds, the better the mapping of the native phoneme categories and non-native phones. On the other hand, the Speech Learning Model (SLM) proposed by Flege (1995, 1999, 2002) claims that L2 learners would find it more difficult to acquire L2 if the L2 sounds or phonological structures are more similar to L1. If an L2 sound is similar to a particular L1 sound, learners' pre-existing L1 phonology will impede their accurate perception and production of that L2 sound. That is, phonological knowledge of L1 hinders an acquisition of similar L2 sounds.

Table 1. The Mean VOT Values of Stops in Mandarin (reprinted from Lin & Wang, 2005) and Spanish (reprinted from Lisker & Abramson, 1964)

	Stops	VOT values (ms)	VOT values (ms)
Unaspirated voiceless	/p/	6	4
	/t/	9	9
	/k/	24	29
	Mean	13	14
Aspirated voiceless	/p^h/	88	
	/t^h/	86	
	/k^h/	82	
	mean	85	
Voiced	/b/		-138
	/d/		-110
	/g/		-108
	mean		-118

Taken together, CAH could predict that aspirated stops not allowed in CS-TM leaners' phonological inventory impedes the learning of stops in TM. PAM would also focus on the simliarity of (un)aspirated voiceless stops between CS and TM, resulting in easier acquisition. SLM would claim that CS unaspirated stops prevent the establishment of a new distinct category for TM aspirated ones, resulting in interference. This raises a question: what happens if an L2 contrast involves some dimension, for example, aspiration, not being used contrastive in the learners' L1? Thus, several perceptual and production experiments were conducted to examine whether the CS-TM learners encounter any difficulty in producing initial aspirated stops in onset position, and whether any improvement can be observed after several months living in a TM society. More sepcifically, several factors (such as context, articulators, and so on) will be examined to understand what obstacles impedes their performance when learning TM.

Method

Participants

Fifteen participants[2] (6F, 9M, 18-25; mean age=24;) were recruited to participate in both perception and production experiments. All are citizens of the Dominican Republic and TM beginning learners. Most of them are Spanish-English bilinguals, and some of them also speak French, Italian, Russian. Mandarin classes are taken six hours a day, meaning they have an opportunity to explore Mandarin step by step. None of them reports any hearing or reading disability.

Perception experiment

Stimuli

A total of 288 Mandarin characters (the onset varies from 6 consonants [p, p^h, t, t^h, k and k^h] X 12 rhymes [a, i, u, o, in, iŋ, un, uŋ, ən, əŋ, ɑn and ɑŋ] X four tones [H (Tone 1), R (Tone 2), L (Tone 3)

[2] Originally, there were seventeen participants. However, because of the heavy Chinese learning program, only fifteen people were available.

and F (Tone 4)[3]]) were read by a native speaker of TM. All items were presented in Hanyu pinyin and the tones were presented in numbers (1 for H, 2 for R, 3 for L, and 4 for F), shown in Appendix.

Procedure and analysis

Participants were asked to choose one of three items represented in Hanyu Pinyin, based on the voice files that they just heard. If they were not sure, [other] was recommended. The first identification task took place after the participants had been learning TM for less than one month. The interval between 1st and 2nd experiment was about 3 months. The correct rate was served as the dependent variable when performing ANOVA for the 1st experiment and ANCOVA for the 2nd experiment. The independent variables factors suspected to be factors that affect learners' performance were: "Vowel", "Aspiration", "Place of Articulation (POA) of stops" and "CVC2".

Production experiment

Stimuli

The wordlist, shown in the Appendix, served as stimuli in the perception experiment.

Procedure and analysis

Praat (5.3.34) at a sampling frequency of 44100 Hz, and headphones (logitech) were used to record the stimuli read by the CS-TM leaners three times. The best stimuli was chosen and annotated by the second author. The production task took place after the second perceptual experiment was done. VOT values were measured as the dependent variable when performing ANOVA. The same factors were considered as indepedent variables in the perception experiment.

Results

Perception results (1st exp)

The ANOVA was performed and the results showed that there was a significant difference among vowels, $F(4, 191) = 6.179$, $p < .05$; the *post hoc* multiple comparison (Bonferroni) indicated that participants perceived stops adjacent to vowel /e/ and /i/ better than those adjacent to vowel /u/, and also perceived stops adjacent to vowel /i/ better than those adjacent to vowel /o/. There was also a significant difference in perceiving stops by means of their "POA", $F(2, 191) = 22.691$, $p < .05$. The bilabial and velar stops were better perceived than alveolar ones. Perception of aspirated stops was significantly better than that of unaspirated ones, $F(1, 191) = 34.36$, $p < .05$. Finally, 'tone' was one of the factors found to significantly affect perception of stops in TM, $F(3, 191) = 2.845$, $p < .05$.

Perception results (2nd exp)

The ANCOVA was performed and the results showed that there was no difference with "Vowel" ($F(4, 192) = 0.954$, $p > .05$), "Aspiration" ($F(1, 192) = 0.963$, $p > .05$), "Tone" ($F(3, 192) = 1.665$, $p > .05$), "POA" ($F(2, 192) = 1.033$, $p > .05$) and "CVC2" ($F(1, 192) = 1.773$, $p > .05$).

Comparisons between two perception results

A pair-samples *t* test was performed to examine the relationship between the 1st and 2nd experiment. The results showed no significant difference between the first ($M = -30.2$, $SD = 67.9$) and the second ($M = -29.01$, $SD = 85.1$); $t(674) = -0.54$, $p > .05$. This means that our participants did not significantly change their way of perceiving TM when the correct rates were measured.

[3] To describe the pitch contour: they are high-level (Tone 1), mid-rising (Tone 2), low-dipping (Tone 3) and high-falling (Tone 4) (Duanmu, 2007).

Production results

The ANOVA was performed and the results showed there was a significant difference among "Vowels" $F(4, 191) = 10.724$, $p < .05$. The *post hoc* multiple comparison (Bonferroni) indicated that our participants produced stops adjacent to vowel /e/ and /i/ longer than those adjacent to vowel /a/, and produced stops adjacnet to /u/ longer than those adjacent to vowel /a/ and /o/. There was also a significant difference in producing stops by means of their "POA", $F(2, 191) = 52.034$, $p < .05$. Velar stops were produced significantly longer than alveolars, which were significantly longer than bilabials. Production of aspirated stops was significantly longer than unaspirated ones, $F(1, 191) = 3302.011$, $p < .05$. In addition, the production of stops was significantly affected by "tones", $F(3, 191) = 10.215$, $p < .05$. The *post hoc* multiple comparison (Bonferroni) indicated that the production of stops with T3 was longer than those with T1 and T4. Finally, there was also a significant difference when the stimuli had a coda, $F(1, 191) = 24.027$, $p < .05$. Closed codas affect participants to produce longer VOT stops in onset position.

To summarize perception and production results in Table 2, 'Vowel', 'Aspiration or not', 'POA of the stop' affect perception and production of initial stops in TM. It is of interest that all these effects diminished in the second perceptual experiment.

Table 2. Results on perception and production

		vowel	tone	aspiration	POA	CVC2
perception	1st	*/e,i/>/u/; /i/>/o/		*asp>unasp	* bil,vel> alv	
	2nd					
production		*/e,i/ > /a/; /u/ > /a,o/	*T3>T1,T4	*asp>unasp	* vel > alv>bil	*with C2>without C2

Discussion

Perception

Comparing the two results on perception, CS-TM learners have acquired aspirated stops since they have been learning TM for less than a month, meaning that aspiration stops are easy for them to perceive. This result is in line with PAM; the more similar L2 sounds are to L1 sounds, the more accurately they are perceived. However, this result could still be partially explained in the light of SLM.

The longer learners are immersed in L2, the greater the influence they will experience. Flege (1987) studied the VOT values of both French and American English (AE) bilinguals finding that both their VOT values were affected significantly by the length of stay in the L2 environment. Compared to Sancier and Fowler's (1997) findings, whose participants had only stayed around three months, Flege's finding of the VOT shift was more observable. Thus, our results are adequately explained by SLM by means of the learning process. Comparing our perceptual data with those in Flege (1987) and Sancier and Fowler (1997), both the ability to produce and perceive a native sound is affected by immersion in the L2 environment for just a few months. The non-native sounds, aspirated stops, are emphasized after a short time learning. At the same time, the ability to perceive unaspirated stops is reduced because of lack of exposure to the native CS environment. In this sense, our results provide further evidence to support the claim in SLM, i.e. the influence of the perceptual ability of the native sound.

As concerns which factors affect stop identification, stops adjacent to front vowels are perceived better than those adjacent to back vowels. Morris, McCrea, and Herring (2008) and Higgins et al. (1988) explain that when a speaker produces vowel [i], the vocal folds tense and delay vibration. This may explain why VOT values are longer and more recognizable for stops adjacent to vowel [i] (Kondaurova & Francis, 2008). However, the opposite observation was made by Peng (2009), and Rochet & Fei (1991) that [p] in TM has longer VOT values before [u] than before [i]. Could this vocal-fold tension of vowel [i] be generalized to other front vowels causing longer duration? We will tentatively look into this possibility as further acoustic analysis research is needed.

Alternatively, Wu (2004) states that vowel /i/ has a lower F1[4] value compared to vowel /u/ (290 Hz vs. 380 Hz). Therefore, onset stops might have larger movement space when adjacent to vowel /i/ than to vowel /u/. In other words, because stops adjacent to vowel /i/ have longer VOT values, participants might perceive it better. Again, could this be generalized to other vowels, such as the mid-front vowel /e/. But such explanations need further investigation.

On the other hand, bilabial and velar stops are perceived better than alveolars. This result matches Winters' (2000) finding that labials and dorsals were more salient than coronal stops; as a result, they are easier to be perceived.

Production

Stops adjacent to vowels /e/ and /i/ are produced longer than those adjacent to vowel /a/. Those adjacent to vowel /u/ are produced longer than those adjacent to vowel /a/ and /o/. The distinction between those conditions is the height of tongue position. High vowels inferring a condition of longer VOT values (Cho & Ladefoged, 1999) could benefit CS-TM learners' production. The only similarity between perception and production in our result is that stops adjacent to vowel /i/ get longer VOT than those adjacent to vowel /a/ (Higgins et al., 1988).

The CS-TM learners produce velar stops longer than alveolar ones, which are longer than bilabial onsets. The VOT values of glottal and velar sounds are the longest (Cho & Ladefoged 1999; Kent & Read, 2002; Wu, 2004), similar to native speakers of TM as shown in Table 1. This means that, after several months of TM training our participants acquired aspirated stops well enough to distinguish the subtle difference like native speakers do. It is not surprising to find that the VOT duration of aspirated stops (long lag) is longer than that of unaspirated onset (short lag), since participants produced them well. Whether tone or nasal codas play any role in the production of initial stops remains unknown.

Conclusion

Our experiments, involving 15 CS-TM learners, have revealed a speedy process of learning initial aspirated stops in both perception and production. These results are consistent with the PAM model proposition that the closer the L2 sound to the L1 sound, the easier such a sound is acquired. We suspect that the same orthography of Hanyu Pinyin to represent the difference between voiced-voiceless in CS and unaspirated-aspirated voiceless stops is facilitatory: i.e. /b/ and /p/ (Pinyin) matches /p/ and /p^h/ (TM phoneme). Even though several factors, such as the vowel condition and the POA of the stop are found to affect the production and perception of initial stops, they are discussed in terms of articulation, or they conform to a pattern observed in most language, velars > alveolars > bilabials. How exactly tone and coda play a role remains to be further investigated.

References

Best, C. T. (1995). A direct-realist view of cross-language speech perception. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research* (pp. 171 - 206). York: Timonium.

Best, C. T., McRoberts, G., & Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology*, *14*, pp. 345-360.

Cheng, M.-C. (2013). Voice onset time of syllable-initial stops in Sixian hakka: Isolated syllables. *Journal of National Taiwan Normal University: Linguistics and Literature*, *58*(2), 193-227.

[4] The F_1 is mentioned here because it reflects the formant (position of the tongue) of the sound. The higher the tongue, the lower the F_1; the lower the tongue, the higher the F_1 will be.

Cho, T., & Ladefoged, P. (1999). Variation and universal in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*, 207-229.

Duanmu, S. (2007). *The phonology of Standard Chinese*. Oxford, UK: Oxford University Press.

Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47-65.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America, 89,* 395-411.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*, 1589-1608.

Flege, J. E. (1995). Second-language speech learning: Theory, findings and problems. In W. Strange (ed), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 233-272). Timonium, MD: York Press.

Flege, J. E. (1999). The relation between L2 production and perception. In J. Ohala, Y. Hasegawa, M. Granveille, & A. Bailey (eds.) *Proceedings of the XIVth International Congress of Phonetics Sciences* (pp. 11273-1276). Berkely, United States.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217–244). Trier, Wissenschaftlicher Verlag.

Higgins, M. B., Netsell, R., & Schulte, L. (1998). Vowel-related differences in laryngeal articulatory and phonatory function. *Journal of Speech, Language, and Hearing Research, 41*, 712-724.

Hu, G. (2012). Chinese and Swedish stops in contrast. In A. Eriksson, Å. Abelin, P. Nordgren, & K. Lundholm Fors (eds.), *Proceedings of Fonetik 2012* (pp. 77-80). Gothenburg Department of Philosophy, Linguistics and Theroy of Science, University of Gothenburg.

Kent, R. D., & Read, C. (2002). *The Acoustic Analysis of Speech* (2nd ed.). CA: Singlular.

Klatt, D. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*, 686 -706.

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, *124*(6), 3959-3971.

Lado, R. (1957). *Linguistics Across Cultures*. Ann Arbor: The University of Michigan Press.

Lin, Yen-Hwei (2007). *The Sounds of Chinese*. Cambridge, UK: Cambridge University Press.

Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics*, *36*(2), 308-317.

Peng, J.-F. (2009). *Factors fo voice onset time: stops in Mandarin and Hakka*. Unpublished master thesis, National ChengKung University.

Rochet, B. L., & Fei, Y. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: Production and perception. *Canadian Acoustic*s, *19(4),* 105-106.

Salcedo, C. (2010). The phonological system of Spanish. *Revista de Lingüística y Lenguas Aplicadas*, *5*, 195-209.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421-436.

Winters, S. (2000). Turning phonology inside out: Testing the relative salience of audio and visual cues for place of articulation. In R. Levine, A. Miller-Ockhuizen, & T. Gonsalvez (eds.), *Ohio State Working Papers in Linguistics 53,* (pp. 168-199). Ohio State University.

Wu, T.-J. (1998). Experimental study on Mandarin unaspirated / aspirated consonants. *Chinese Language*, *3,* 256 -283.

Wu, T.-J. (2004). *Wu Tsung-Ji Linugistic Papers*. Beijing: Commercial Press.

Zheng J. (2011). *Phonetics: Science of speech*. Taipei: Psychology Press.

Appendix

The wordlist in Hanyupinyin

ba1	ba2	ba3	ba4	da1	da2	da3	da4	ga1	ga2	ga3	ga4
pa1	pa2	pa3	pa4	ta1	ta2	ta3	ta4	ka1	ka2	ka3	ka4
bi1	bi2	bi3	bi4	di1	di2	di3	di4	gi1	gi2	gi3	gi4
pi1	pi2	pi3	pi4	ti1	ti2	ti3	ti4	ki1	ki2	ki3	ki4
bu1	bu2	bu3	bu4	du1	du2	du3	du4	gu1	gu2	gu3	gu4
pu1	pu2	pu3	pu4	tu1	tu2	tu3	tu4	ku1	ku2	ku3	ku4
bo1	bo2	bo3	bo4	do1	do2	do3	do4	go1	go2	go3	go4
po1	po2	po3	po4	to1	to2	to3	to4	ko1	ko2	ko3	ko4
ban1	ban2	ban3	ban4	dan1	dan2	dan3	dan4	gan1	gan2	gan3	gan4
pan1	pan2	pan3	pan4	tan1	tan2	tan3	tan4	kan1	kan2	kan3	kan4
bin1	bin2	bin3	bin4	din1	din2	din3	din4	gin1	gin2	gin3	gin4
pin1	pin2	pin3	pin4	tin1	tin2	tin3	tin4	kin1	kin2	kin3	kin4
bun1	bun2	bun3	bun4	dun1	dun2	dun3	dun4	gun1	gun2	gun3	gun4
pun1	pun2	pun3	pun4	tun1	tun2	tun3	tun4	kun1	kun2	kun3	kun4
ben1	ben2	ben3	ben4	den1	den2	den3	den4	gen1	gen2	gen3	gen4
pen1	pen2	pen3	pen4	ten1	ten2	ten3	ten4	ken1	ken2	ken3	ken4
beng1	beng2	beng3	beng4	deng1	deng2	deng3	deng4	geng1	geng2	geng3	geng4
peng1	peng2	peng3	peng4	teng1	teng2	teng3	teng4	keng1	keng2	keng3	keng4
bang1	bang2	bang3	bang4	dang1	dang2	dang3	dang4	gang1	gang2	gang3	gang4
pang1	pang2	pang3	pang4	tang1	tang2	tang3	tang4	kang1	kang2	kang3	kang4
bing1	bing2	bing3	bing4	ding1	ding2	ding3	ding4	ging1	ging2	ging3	ging4
ping1	ping2	ping3	ping4	ting1	ting2	ting3	ting4	king1	king2	king3	king4
bong1	bong2	bong3	bong4	dong1	dong2	dong3	dong4	gong1	gong2	gong3	gong4
pong1	pong2	pong3	pong4	tong1	tong2	tong3	tong4	kong1	kong2	kong3	kong4

The development and standardisation of the bilingual Maltese-English speech assessment (MESA)

Helen Grech[1], Barbara Dodd[2], Sue Franklin[3]
helen.grech@um.edu.mt

[1] Faculty of Health Sciences, University of Malta, [2] School of Community and Health Sciences, City University London, [3] School of Health Sciences, University of Limerick

Abstract. Speech language pathologists working with Maltese-English bilingual children often assess and diagnose speech disorders using assessment protocols standardised on monolingual, English-speaking populations. Such tests are considered inappropriate for the Maltese bilingual children since they are not linguistically or culturally oriented. An innovative speech assessment protocol which is bilingual in nature, was developed and standardised. Children were tested in Maltese and/or English depending on their language (or language mix) exposure. A novel feature of this assessment battery was that for all of the items, children were able to respond in either language, reflecting the reality of language mixing in a bilingual population. Trends of speech development for monolingual and bilingual children aged between 2;0-6;0 years are reported, differentiating between the emergence of the ability to produce speech sounds (articulation) and typical developmental error patterns (phonology). This assessment gives clinicians a more objective view of the discrepancy between typical development, delay and deviancy for children acquiring speech in Malta. The research findings are novel and have both theoretical and clinical implications.

Keywords: bilingual assessment, Maltese bilingual assessment, bilingual speech test, Maltese-English speech test

Introduction

Research related to children's speech and language development comes mainly from studies of monolingual English-speaking children (Hua & Dodd, 2006). However, there has been increased interest in children acquiring other languages (e.g., Fox, 2000 for German; Ballard & Farao, 2008 for Samoan). Research suggests that children acquiring different languages have some language specific developmental error patterns indicating that findings for one language are not applicable to other languages (e.g., So & Dodd, 1994 for Cantonese; Amayreh & Dyson**,** 1998 for Arabic; Grech, 1998 for Maltese; Zhu & Dodd, 2000 for Putonghua; Macleod, Sutton, Trudeau, & Thordardottir, 2010 for Québécois French). During the past decade, research on bilingual acquisition has become of more interest.

Studies on bilingual acquisition include those of Lleó, Kuchenbrandt, Kehoe, and Trujillo (2003): German-Spanish; Salameh, Nettlebladt, and Norlin (2003): Arabic-Swedish; Fabiano and Goldstein (2005): Spanish-English; Munro, Ball, Müller, Duckworth, and Lyddy (2005): Welsh-English; De Houwer, Bornstein, and De Coster (2006): Dutch-French; Holm and Dodd (2006): Cantonese-English; Sundara, Polka, and Genesee (2006): French-English. There are indications that children exposed to early sequential bilingualism show different patterns of phonological acquisition to those of monolingual children of the respective languages (e.g., Holm & Dodd, 1999; Grech & Dodd, 2008). Further, sequential bilinguals may exhibit differences in type and amount of errors from simultaneous bilinguals (De Houwer, 2009). When Wright and Gildersleeve (2005) compared 11 monolingual English-speaking children with five Russian-English bilinguals (two of whom learned Russian and English simultaneously and three who acquired English once their Russian was established), they found that the sequential bilinguals made more consonant errors than the simultaneous bilinguals and that overall the bilingual children made more errors than monolingual subjects.

The finding that bilingual children's phonological acquisition differs from that of monolinguals of either of the languages spoken indicates that having two phonologies affects the course of acquisition. This is in line with the *Interactional Dual Systems Model* for the mental organization of more than one language. The model asserts that bilingual children have two separate phonological systems, but that those two systems can influence one another. Paradis (2001) reported such cross-linguistic features in the productions of bilingual children. The model fits with data from other studies of bilingual children (e.g., Johnson & Lancaster, 1998 (Norwegian-English); Holm & Dodd 1999a, b, c (Cantonese-English, Italian-English and Punjabi-English, respectively); Keshavarz & Ingram, 2002 (Farsi-English); Salameh, Nettlebladt, & Norlin, 2003 (Swedish-Arabic)). On the other hand, Navarro, Pearson, Cobo-Lewis, and Oller (1995) found no atypical phonological error patterns in the speech of 11 successive bilingual Hispanic-English pre-school children in the US.

Hua and Dodd (2006) reviewed the varying reports concerned with the phonological development of bilinguals. Some studies (Burling, 1959/1978; Leopold, 1939-1949; Schnitzer & Krasinski, 1994, 1996; and Johnson & Lancaster, 1998) claim initial periods of a single phonological system for simultaneous bilinguals. Other studies (Wode, 1980; Fantini, 1985; Watson, 1991) report that successive bilinguals tend to superimpose an unknown system on the more stable one, using one system as a base, and differentiating the second system by altering or adding to the first system. Hua and Dodd concluded that apparently conflicting findings may reflect differences between the different language pairs learned, or the comparative length of exposure to a child's two languages.

Research describing bilingual language acquisition is limited in terms of the language pairs studied and the language learning contexts investigated. Data are often reported from studies where the children's first language is that of their immigrant parents in a country where the dominant language is English (e.g., Goldstein & Washington, 2001 for Hispanic children in the US, Stow & Dodd, 2003 for Pakistani Heritage languages in the UK). They are in a community where one language is spoken apart from the home language. The child becomes bilingual as a result of the shift of linguistic environment. However, there also exist simultaneous bilinguals where acquisition of two or more languages occurs very early in their lives. De Houwer (2009) refers to two sub-groups of such children, i.e. those who have bilingual first language acquisition (BFLA) when there is no existing chronological difference in the exposure of both languages; and those children who are early second language learners (ESLLs) where they are exposed to a second language on a regular basis between 18 and 48 months of age. De Houwer also refers to formal second language acquisition, whereby children are introduced to a second language and literacy at about 5 years of age. Second language acquisition (SLA), and English language learners (ELLs), or equivalent terms used in non-English contexts, are other terms used often referring to learning the second language at school. Sequential acquisition can also refer to learning subsequent languages at any time during life.

Another important limitation of the available data on bilingual children's acquisition of language is that, very often, the two languages studied come from the same language family that share similar phonological characteristics. There is evidence that research findings from two Indo-European languages (e.g., English-French; Spanish-English) differ from those for other language pairs (Hua & Dodd, 2006) where English is learned in addition to Cantonese (Yip & Matthews, 2007) or Maltese, which is Semitic in origin. Further, the number of children involved in these studies has been extremely small (many are case studies) with the consequence that there are no normative data for many language-pairs, thus limiting the assessment and diagnosis of speech and language disorders. Studies in English-speaking countries have established and standardized assessments to identify children with speech and language difficulties but no such protocols are available for the Maltese population. The authors attempted to address this gap in the knowledge base concerned with speech and language acquisition in the Maltese context by developing a bilingual speech assessment, and administering it on a large sample of Maltese children. Data were analysed for trends of acquisition for children reported to be 'monolingual' by their parents, and those exposed to Maltese and English at home.

The phonologies of Maltese and English have their origins in two different language groups (Semitic and Indo-European). The consonant phonetic inventory of Maltese is similar to that of English (with /ʔ/ and /ts/ being 'additional' Maltese phonemes while English /θ/, /ʒ/ and /ð/ are not part of the

Maltese inventory. However, the two languages differ in their phonotactics. Maltese has a greater range of possible consonantal clusters and consonantal sequences and is characterised by multi-syllabic lexemes (Borg & Azzopardi-Alexander, 1997). A speech test standardised on English-speaking children, such as the *Diagnostic Evaluation of Articulation and Phonology (DEAP)* (Dodd, Zhu, Crosbie, Holm, & Ozanne, 2002) is therefore not applicable since it does not cater for Maltese phonotactics and should not be used to assess Maltese-speaking children. The use of English-only standardized tests, when clinicians evaluate non-native English speakers has often been reported (e.g., Skahan & Lof, 2007) and should be avoided as it may lead to misdiagnosis of a speech disorder.

Clinicians need language specific tools to identify children with speech and language difficulties, since it is well known that children with an early history of language impairment may be at risk for continuing communication difficulties particularly related to written language development (Shriberg & Kwiatkowski, 1994). Educational achievement is also related to early speech and language abilities (Bickford-Smith, Wijayatilake, & Woods, 2005) as well as social, emotional, or behavioral challenges (Rome-Flanders & Cronk, 1998). It was therefore considered crucial to develop a speech and language assessment that can identify children who have speech and language disorder in Malta. The objectives of this study were to construct and standardize a speech assessment battery appropriate for children acquiring language in the bilingual language learning context of Malta. Traditionally assessments for children in bilingual contexts have consisted of two separate tests, one for each language. This does not reflect the reality of the way that bilingual children use language in terms of language mixing and word borrowing. Indeed to respond appropriately in the required language may require an added degree of metalinguistic control. Uniquely in this speech and language assessment battery, children are able to use Maltese or English in response to each stimulus item. Scoring and analysis also cater for language mixing (for details see Manual in Grech, Franklin, & Dodd, 2011).

The Maltese socio-linguistic context

The Maltese Islands have a complex language learning context. There are two official languages (Maltese and English), most children are bilingual in that they have some knowledge of both languages but one of the languages may be dominant. Reports from parents indicate that in some homes one of the languages may be used exclusively while other families use both languages (Grech & Dodd, 2008) so that the child is exposed to two languages at home soon after birth (simultaneous acquisition). In comparison, some children are exposed to only Maltese or English at home, followed by exposure to their second language in the community, usually by 3 years of age when they start attending pre-school (early sequential acquisition). This context was used in this large scale project to study the effects of language exposure at home on the rate and course of speech and language acquisition. Data from children reported by the carers to be simultaneously bilingual (as referred above) were analysed and reported separately from data of children who were reported to be exposed to Maltese or English at home (referred as monolingual in this study). The term 'monolingual' in the Maltese context has to be treated with caution and refers to home exposure. The language learning context of Malta, where most people have some knowledge of two languages reflects emerging patterns of language use in the European Union, due to population shifts, where many people have some knowledge and functional use of at least two languages, although one language may be dominant.

Research questions

The purpose of the study was to develop a bilingual Maltese-English speech and language assessment and to identify trends of acquisition for children reported to be 'monolingual' by their parents and those exposed to simultaneous bilingualism at home. This paper reports data related to the speech assessment.

The Maltese-English Speech Assessment (MESA) (Grech, Dodd, & Franklin, 2011) was constructed and evaluated to address the following research questions:

- Does the MESA demonstrate that a single battery can effectively assess monolingual and bilingual children?

- Do the speech acquisition patterns for these two populations differ?
- Is the MESA a reliable and valid test of speech development?
- Does the MESA distinguish between typically developing children and those with delayed or disordered speech (making the assessment a useful tool for clinicians working with Maltese-speaking children?

Methodology

The sample

The public registry of births for the Maltese Islands was accessed to draw a random sample of 1,000 Maltese children aged 2;0 to 6;0 years. All children whose parents consented to participate in the project (a total of 241 children) were assessed on a picture naming task to evaluate phone articulation, phonology and consistency of word production. The children were also assessed for oro-motor skills and the ability to repeat phonotactically complex words. The sample included a total of 134 girls and 107 boys. Twenty-two participants were aged 24-35 months; 35 were 36-41 months; 45 were 42-47 months; 40 were 48-53 months. Information was collected from the carers related to whether the children had an underlying sensory, cognitive or anatomical/physiological condition, family history of communication difficulties, and other factors such as socio-economic status that could reflect on their speech and language acquisition. However, this information did not result in exclusion of children from the study unless the assessment distressed them. The rationale for this decision was to avoid over-diagnosis of impairment, since data identifying typical performance must be based on a representative sample of the total population.

Table 1. Maltese sample by age and gender

Age in months	Total no. of age cohort	No. of girls	No. of boys
24-35	20	9	11
36-41	36	23	13
42-47	45	27	18
48-53	40	19	21
54-59	34	11	23
60-65	37	25	12
66-72	29	20	9
Total % of sample	241 100%	134 56.6%	107 44.4%

Table 2. Maltese sample by language learning context

Age in months	Maltese	English	Maltese-English	Total per cohort
24-35	10	1	11	22
36-41	18	2	15	35
42-47	23	2	20	45
48-53	28	1	11	40
54-59	22	1	11	34
60-65	22	1	14	37
66-72	15	3	10	28
Total % of sample	138 57.26	11 4.56	92 38.17	241 100

Other information related to the primary language of the child and language/s used at home was collected. The children were allowed to use the language they chose (either Maltese or English). Ninety-two children (38.17%) were reported by parents to speak both Maltese and English at home, 138 (57.26%) were reported to speak Maltese and 11 (4.56%) only English at home (see Tables 1 & 2 for details of the sample).

The Assessment Battery (MESA)

The MESA is based on the *DEAP* (Dodd et al., 2002) and consists of four tests that assess articulation, phonology, consistency of production and oro-motor skills.

The Articulation Assessment is meant to identify perceptually any phonemes that cannot be produced by the child. The assessment includes 42 pictures depicting all consonant and vowel sounds in English and Maltese. If a picture is not named spontaneously by the child the administrator attempts to elicit it through imitation in syllable context or in isolation.

The Phonology Assessment is meant to determine the use of surface speech error patterns (developmental phonological processes) that are produced by the child. These may include the language-specific ones (e.g., compensatory vowel lengthening), universal ones (e.g., fronting) and in some instances atypical patterns. Children are asked to name the same 42 pictures and in the same order as in the articulation sub-test, though these have a different coloured background.

The Inconsistency Assessment allows the administrator to evaluate the consistency of production (stability) of the child's contrastive phones. When considered part of the test battery, this assessment enables the identification of those children whose speech is inconsistent but who have no oro-motor difficulties. Children are required to name 17 pictures on three separate trials within one session.

The Oro-motor Assessment evaluates the child's oro-motor function in relation to his/her diadochokinetic (DDK) skills for sequencing and intelligibility. Imitation of isolated and sequenced movements involving speech musculature is also assessed via a separate sub-test. Another sub-test involves the repetition of a list of 11 words some of which are multi-syllabic some include syllable initial consonantal clusters. For this sub-test, the child is asked to repeat the word uttered by the administrator, 3 times consecutively. This word repetition test was included specifically because of the syllabic structure of Maltese and the wide range of multiple combinations of consonantal cluster possibilities as well as multi-syllabic utterances. Examples of words in this sub-test include: /tpɪndʒɪ/ meaning 'colouring; /hwɛɪɛʧ/ meaning 'clothes'; /sʊfɐrɪnɒɐ/ meaning 'match'.

The MESA portfolio includes the *Manual,* which provides clear instructions for its administration. The *Stimulus book* contains pictures that are culturally appropriate, age appropriate and colourful. This was checked by piloting the test on Maltese children of varying ages. Clinicians were also approached for feedback before confirming the list of pictures to be used. The pictures are visually attractive to children between 2;0-6;0 years of age on whom this test should be administered. The *Score sheets* are colour-coded for ease of reference and allow for entry of raw scores for each section. Different sub-tests can carried out on separate sessions (but close in time), particularly if the test is being used for review purposes. The articulation test is easy and quick to score whereby the clinician is only expected to circle any phones that the child does not produce in the adult form. Phonetic transcription according to the International Phonetic Alphabet (IPA) is required for the phonology test. This would allow for the identification of error patterns and idiosyncratic phoneme production. Quantitative analysis is recommended to calculate percent consonants correct (PCC) and percent vowels correct (PVC) for the different language codes (e.g., Maltese; Maltese-English). PCC and PVC measures are used regularly to index the phonological skills of children. PCC measures are reported to be linguistically and psychometrically valid (Shriberg, Austin, & Lewis, 1997).

The Inconsistency test score is calculated as a percentage of the number of words produced differently in 3 trials in relation to the total number of words produced 3 times. The other sub-tests are easy to score whereby accurate production is given a score and the total score per sub-test is noted. Speech-language pathologists (SLPs) are expected to administer the test, score, analyse the data, and compare them to 'typical' data.

Procedure

Most of the children were assessed at home in one or two sessions. During each 1-hour session short breaks were given as often as was considered necessary. A few children were assessed in the

University Communication Therapy Teaching and Research Clinic following parental request. The children completed the MESA and additional language related tasks that assessed narrative comprehension, expressive language, sentence imitation, and phonological awareness skills. The carers also completed checklists related to the child's voice quality, fluency, and pragmatic skills. This paper reports results of the MESA only. Pre-assessment criteria were set in relation to the test-administrators' language use for instruction. Maltese was used to give assessment instructions unless the child was English-speaking. If unsure, the examiner used the language chosen by the child. When carers reported that the child was bilingual, Maltese was used. A novel feature was that the children had the choice to respond in either language. Ideally, bilingual children should be tested in both languages. For the MESA study this was not done since data collection already involved considerable time commitment due to the administration of the MESA and a language assessment battery, which extended to 2 home visits for most children. This decision was also supported by there being only 3 additional English phonemes which do not exist in Maltese phonology, i.e. /θ, ð, ʒ/. It has been reported (e.g., Grech, 1998) that the Maltese use /t/ and /d/ for /θ, ð/, respectively when speaking English (generalisation of Maltese phonology). Meanwhile, /ʒ/ is the least frequently used phoneme of English (http://www.instructables.com/answers/What-are-the-most-commonly-used-to-least-comonly-u/). The MESA scores sheets allow for code-switching, the data showing that most children produced some words in both languages.

Reliability of the MESA

The accuracy and consistency of the MESA was measured by test-retest reliability and inter-rater reliability. Test-retest reliability was estimated by testing 5% of the sample twice (mean age: 51.6 months). The between test interval was less than 5 weeks. Inter-rater reliability was measured in relation to the degree of consistency between persons scoring, transcribing, and analysing the children's speech. The audio recordings of 12 children (5 % of Maltese normative sample; mean age: 46.3 months) were transcribed and analysed by 2 independent examiners.

Validity

The content and concurrent validity of the MESA were established in different ways. The data from the typically developing children using MESA were compared with those in Azzopardi (1997) and Grech (1998). These studies presented data from typically developing children. Azzopardi's (1997) phonological study investigated the development of Maltese consonants and some consonant clusters in 4-year old Maltese-speaking children. A cross-sectional study of 10 children was carried out. Parents were interviewed and relevant screening measures were applied before including children in the study. A phonological sample was collected at each child's home using picture elicitation materials designed specifically for this study. The sample was transcribed and analysed using the Phonological Assessment of Child's Speech (PACS) (Grunwell, 1985). The results indicated that: (a) fricatives and liquids were most likely to be misproduced; (b) only 5 developmental processes (error patterns) were observed, thus indicating that the children had eliminated most developmental phonological processes; and (c) many of the clusters were produced consistently.

Grech's (1998) exploratory study was related to the phonological development of 21 normally developing Maltese-speaking children. The children were recorded in their natural settings at four different stages between ages 2;0 and 3;6. The data collected were transcribed narrowly and analysed. Each child's phonetic/phonological inventory was identified; various developmental phonological processes were also recorded throughout the period of study. A developmental profile was collated for the group, indicating trends of stages of phonological development. This profile was compared cross-linguistically. The data fits in with current theories highlighting universal phonological acquisition particularly in the early years. As predicted some language-specific behaviour was also observed. The usefulness of the MESA was also validated by data from a clinical population (not part of the larger cohort of the study). It was hypothesized that data from children who had been clinically identified with speech sound disorder would differ from those of the *normative* sample and from those of children with 'other' communication impairments. Differential diagnosis of these children with impairments was made by a clinician using various speech and language assessment tools that are not 'standardised' on the local population because the latter are unavailable to date. The same criteria as for the *normative* sample were applied with regards to the decision as to whether these children were considered monolingual or bilingual.

Results

The analyses completed on the speech data included the following quantitative measures: percent consonants correct (PCC), percent vowels correct (PVC), percent inconsistency score, diadochokinetic score (DDK), single and sequenced oral movements (SSM), scores and word repetition (WR) score. Z-scores, standard scores and percentiles were calculated for each age band, for monolingual and bilingual children aged between 3;0 and 6;0 years of age allowing the detection of children performing below the typical range for this cohort. Data of the children who were younger than 3 years of age were not converted to standard scores because of the limited number of subjects.

Discussion

- *Does the MESA demonstrate that a single battery can effectively assess mono- and bilingual children?*

The results from the MESA were consistent with a developmental trajectory and it was possible to develop standard scores for test administration since the population tested in this study represent 2% of the total population in question. This applies for both monolingual and bilingual children. This calculation is based on the average number of annual births in Malta which is around 4,000 (National Statistics office & Public Registry (personal communication). The assessment battery worked particularly well with respect to the children's use of both languages, which was quite common. In an entirely monolingual test, it is problematic to decide how to deal with items where another language is used; since in this test either English or Maltese was acceptable, all responses could be used in the analysis.

- *Do the speech acquisition pattern for these two populations differ?*

The data indicate that children reported to be monolingual differed from children reported to be bilingual in Maltese and English. There is a clear pattern of faster phonological acquisition for bilingual children as from 3;6 years of age when compared to the monolingual cohort. This is in line with Paradis and Genesee's (1996) hypothesis of faster rate of acquisition of bilinguals when compared to monolinguals. However, these findings are not in line with those reported by Fabiano-Smith and Goldstein (2010) for Spanish-English speaking children who did not exhibit acceleration, a faster rate of acquisition when compared with monolingual peers on overall phonological accuracy, though these skills in the bilingual children were within the normal range of their monolingual counterparts in both English and Spanish. The data collected in this study indicate that early bilingual exposure might enhance phonological acquisition. The claim that children in a bilingual learning context may be at an advantage for spoken phonological acquisition, is supported by other researchers who looked at children exposed to more than one European language (e.g., Bialystok, Luk, & Kwan, 2005 for English-Spanish or Hebrew; Yavaş & Goldstein, 2006 for Spanish-English). Children who are regularly exposed to more than one spoken language would need to discriminate between languages using phonological cues and consequently become aware of the constraints specific to each language's phonology and increase their phonological knowledge. Phonological knowledge is considered to be a marker of phonological ability (Gierut, 2004).

- *Is the Mesa a reliable and valid test of speech development?*

This study also addressed the question as to whether the MESA is a reliable and valid clinical tool that distinguishes between typically developing children and those with delay or disordered speech. A high correlation between test and re-test for quantitative measures was noted. A high percentage test-re-test agreement was reached in relation to the children's production of consonants and error patterns. Similarly, a high correlation was obtained for inter-rater quantitative measures whereby a high percentage agreement was observed for the children's production of consonants and error patterns when rated by different assessors. The MESA is therefore a reliable tool to measure aspects of speech of monolingual and bilingual Maltese children. The error patterns are consistent with those found in the DEAP (Dodd et al., 2002) for children who chose to do the test mainly in English and with Azzopardi (1997) and Grech (1998) for the Maltese-speaking children, This contributes to the validity of the MESA. Its validity is further supported by the clinical sample data as indicated below.

- *Does the MESA distinguish between typically developing children and those with delayed or disordered speech (making the assessment a useful tool for clinicians working with Maltese-speaking children?*

The quantitative severity measures of the clinical sample show that the speech impaired group produced more consonant errors, are more inconsistent and do not produce more vowel errors than those with no speech difficulties. This points towards the validity of the MESA as a clinical tool for the diagnosis of speech impairment. There is a trend towards significance for DDK scores; the difference is probably due to fronting (/k/>/t/) rather than sequencing, fluency or precision of articulation. The fact that there is no significant difference for other oro-motor measures between the 2 clinical groups replicates other findings for speech disordered children and normally speaking controls. Only children with motor speech disorder do poorly on these tasks, as opposed to children with phonological disorder. Percentage correct word repetition just failed to reach significance (p=.07), but the mean scores were 60.5% versus 90% correct. The speech-impaired group showed higher mean scores for all the error types. Therefore, the MESA proved to be a clinically discriminatory and a valid tool for the assessment of speech disorders, since the two groups differed on key measures specific to speech, but not as could be predicted on the oro-motor measures. The MESA will aid clinicians to differentiate between 'typical' language development patterns and language disorder and to direct the most effective intervention to children who struggle with developing phonetic and phonological skills.

Conclusion

The MESA is an innovative protocol where the sub-tests devised are truly bilingual in nature. Hence, a child living in Malta would be tested in Maltese and/or English depending on which language/s (or language mix) s/he would be exposed to. This innovation is time-cost-efficient in that bilingual children need not have to go through 2 different tests for checking proficiency of speech skills since, as indicated above, the Maltese use mainly Maltese phonemes when speaking English. However, if the clinician has time, it would be ideal to administer the test in both English and Maltese to the bilingual child.

It also reflects the reality of the way that children use language in a bilingual situation. The MESA has been shown to be a clinically useful tool for assessing children differentiating between sub-types of speech disorder. Administration of the complete battery should enable the tester to differentiate between disorders of articulation (organic and functional), delayed phonological development, consistent and inconsistent phonological disorder and childhood apraxia of speech. Clinicians using the MESA will be able to reach a differential diagnosis that determines choice of evidence-based treatment approach. Therefore the MESA leads to the improvement of the quality of life of the communication disordered population. Moreover, as was hypothesised, the data collected clearly shows that children reported by parents to be monolingual differ in terms of phonological acquisition patterns from children reported to speak both Maltese and English at home. From the point of view of the test battery itself, it is clear that the standard scores for bilingual and monolingual children need to be given separately.

The results have implications for education, speech-language pathology, psychology and linguistics. For education, teachers of Maltese-speaking children currently have little information about the language competence of typically developing children at school entry, since 'test-book' knowledge is derived from studies of monolingual English speakers in the UK and US. The study's results will allow curriculum modification to better suit children's competence and improve learning outcomes.

SLPs currently have no normative data on the rate and course of language development in Maltese, making choice of intervention targets difficult. Educational psychologists' assessment of verbal cognitive ability is hampered by the dearth of information on Maltese speech and language development. It is hoped that other researchers would use the same framework to develop similar assessments for other bilingual groups in European Member States and elsewhere.

References

Amayreh, A., & Dyson, A. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research*, *4*, 642-653.

Azzopardi, S. (1997). *Phonological development of consonants in 4 year old Maltese children*. Unpublished dissertation, University of Malta.

Ballard, E., & Farao, S. (2008). The phonological skills of Samoan speaking 4-year-olds. *International Journal of Speech-Language Pathology, 10*(6), 379-391.

Bialystok, E., Luk, G. & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading, 9*, 43-61.

Bickford-Smith, A., Wijayatilake, L., & Woods, G. (2005). Evaluating the effectiveness of an early years language intervention. E*ducational Psychology in Practice, 21(3),* 161-173.

Borg, A. J., & Azzopardi-Alexander, M. (1997). *Maltese*. London: Routledge.

Burling, R. (1959/1978). Language development of a Garo and English child. *Word*, *15,* 45-68.Reprinted in E. Hatch (ed.), (1978). *Second language acquisition: A book f readings* (pp.54-75). Rowley, MA: Newbury House.

De Houwer, A. (2009). *Bilingual first language acquisition*. Clevedon, UK: Multilingual Matters.

De Houwer, A., Bornstein, M. H., & De Coster, S. (2006). Early understanding of two words for the same thing: A CDI study of lexical comprehension in infant bilinguals. *International Journal of Bilingualism, 10*(3), 331-347.

Dodd, B., Crosbie, S., Zhu, H., Holm, A., & Ozanne, A. (2002). *The diagnostic evaluation of articulation and phonology.* London: Psych-Corp.

European Commission. (2008). *Speaking for Europe languages in the European union*. Luxembourg: Office for Official Publications of the European Communities. Also available on line: ec.europa.eu/publications

Fabiano-Smith, L., & Goldstein, B. (2010). Phonological acquisition in bilingual Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research*, 53, 160-178.

Fabiano, L., & Goldstein, B. (2005). Phonological cross-linguistic effects in bilingual Spanish-English speaking children. *Attachment and Human Development*, *3*, 56-63.

Fantini, A. F. (1985). *The language acquisition of a bilingual child.* Clevedon: Multilingual Matters.

Fox, A. (2000). *The acquisition of phonology and the classification of speech disorders in German-speaking children*. Unpublished PhD thesis, UK: University of Newcastle-Upon-Tyne.

Gierut, J. (2004). Enhancement of learning for children with phonological disorders. *Sound to Sense June 11 - 13,* MIT, B164-B172.

Goldstein, B., & Washington, P. (2001). An initial investigation of phonological patterns in 4-year-old typically developing Spanish-English bilingual children. *Language, Speech, and Hearing Services in Schools, 32,* 153-164.

Grech, H. (1998) *Phonological development of normal Maltese speaking children*. Unpublished PhD thesis, UK: University of Manchester.

Grech, H., & Dodd, B. (2008). Phonological acquisition in Malta: A bilingual learning context. *International Journal of Bilingualism*, *12*, 155-171.

Grech, H., Dodd, B., & Franklin, S. (2011). *Maltese-English Speech Assessment (MESA)*. Malta: University of Malta. (ISBN: 978-99957-0-027-0).

Grunwell, P. (1985). *Phonological assessment of child speech* (PACS). Windsor:NFER-Nelson

Holm, A., & Dodd, B. (2006). Phonological development and disorder of bilingual children acquiring Cantonese and English. In Z. Hua & B. Dodd (eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 286-325). Clevedon, UK: Multilingual Matters.

Holm, A., & Dodd, B. (1999a). Differential diagnosis of phonological disorder in two bilingual children acquiring Italian and English. *Clinical Phonetics and Linguistics*, *13*, 113-129.

Holm, A., & Dodd, B. (1999b). An intervention case study of a bilingual child with phonological disorder. *Child Language Teaching and Therapy, 15*, 139-158.

Holm, A., & Dodd, B. (1999c). A longitudinal study of the phonological development of two Cantonese-English bilingual children. *Applied Psycholinguistics*, *20*, 349-376.

Hua, Z., & Dodd, B. (eds.). (2006). *Phonological development and disorders in children: A multilingual perspective*. Clevedon: Multilingual Matters.

Johnson, C., & Lancaster, P. (1998). The development of more than one phonology: A case-study of a Norwegian-English bilingual child. *International Journal of Bilingualism, 2,* 265-300.

Kheshavarz, M., & Ingram, D. (2002). The early phonological development of a Farsi-English bilingual child. *International Journal of Bilingualism, 6,* 265-300.

Leopold, W. F. (1939-1949). *Speech development of a bilingual child: A linguist's record.* Evanston, IL: Northwestern University Press.

Lleó, C., Kuchenbrandt, I., Kehoe, M., & Trujillo, C. (2003). Syllable final consonants in Spanish and German monolingual and bilingual acquisition. In N. Müller (ed.), *(In)vulnerable domains in multilingualism* (pp. 191-220). Amsterdam: John Benjamins.

Macleod, A., Sutton, A., Trudeau, N., & Thordardottir, E. (2010). The acquisition of consonants in Québécois French: A cross-sectional study of pre-school aged children. *International Journal of Speech-Language Pathology*, Early Online, 1-17.

Munro, S., Ball, M. J., Müller, N., Duckworth, M., & Lyddy, F. (2005). The acquisition of Welsh and English phonology in bilingual Welsh-English children. *Journal of Multilingual Communication Disorders, 3,* 24-49.

Navarro, A. Pearson, B. Cobo-Lewis, A. and Oller, D. (1995). Early phonological development in young bilinguals: Comparison to monolinguals. Paper presented to the *American Speech, Language and Hearing Association Conferenc1995.*

Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *International Journal of Bilingualism, 5,* 19-38.

Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition*, *18,* 1-25.

Rome-Flanders, T., & Cronk, C. (1998). Stability and usefulness of language test results under two years of age. *Journal of Speech and Language Pathology and Audiology, 2*(2), 74-80.

Salameh, E-K., Nettlebladt, U., & Norlin, K. (2003). Assessing phonologies in bilingual Swedish-Arabic children with and without language impairment. *Child Language Teaching and Therapy, 19*, 338-364.

Schnitzer, M. L., & Krasinski, E. (1994). The development of segmental phonological production in a bilingual child". *Journal of Child Language* 21. 585-622**.**

Schnitzer, M. L., & E. Krasinski (1996). The development of segmental phonological production in a bilingual child: A contrasting second case". *Journal of Child Language*, *23,* 547-571**.**

Shriberg, L., & Kwiatkowski, J. (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech and Hearing Disorders, 51,* 140-161.

Shriberg, D., Austin, D., & Lewis, B.A. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40,* 708-722.

Skahan, S., & Lof, G. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology, 16,* 246-259.

Slavik, A. (2001). Language maintenance and language shift among Maltese migrants in Ontario. *International Journal of the Sociology of Language*, *152*, 131-152.

So, L., & Dodd, B. (1994). Phonologically disordered Cantonese-speaking children. *Journal of Clinical Linguistics and Phonetics*, *8*, 235-255.

Stow, C. & Dodd, B. (2003). Providing an equitable service to bilingual children in the UK. *The International Journal of Language and Communication Disorders, 38*(4), 351-377.

Sundara, M. Polka, L., & Genesee, F. (2006). Language experience facilitates discrimination of /d–_/in monolingual and bilingual acquisition of English. *Cognition, 100*, 369-388.

Watson, I. (1991). Phonological processing in two languages. In E. Bialystok (ed.), *Language processing in bilingual children* (pp. 25-48.). Cambridge, UK: Cambridge University Press.

Wode, Henning. (1980). *Language Acquisitional Universals L1, L2, Pidgins, and FLT*. [Washington, D.C.] : Distributed by ERIC Clearinghouse.

Wright, K., & Gildersleeve-Neumann, C. (2005). Speech-sound development in preschoolers from bilingual Russian-English language environments. Poster presented at the *Oregon Speech-Language-Hearing Association Conference*, US (October).

Yavaş, M., & Goldstein, B. (2006). Aspects of bilingual phonology: the case of Spanish-English bilingual children. In B. Dodd & Z. Hua (eds.), *Phonological Development and Disorders: A Cross-linguistic Perspective* (pp. 265-285). Multilingual Matters.

Yip, V., & Matthews, S. (2007). *The bilingual child: Early development and language contact*. Cambridge, UK: Cambridge University Press.

Zhu, H., & Dodd, B. (2006). *Phonological development and disorders: A multilingual perspective.* Clevedon: Multilingual Matters.

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, *27*, 3-42.

Gradience in multilingualism and the study of comparative bilingualism: A view from Cyprus

Kleanthes K. Grohmann[1,3]*, Maria Kambanaros[2,3]
kleanthi@ucy.ac.cy, maria.kambanaros@cut.ac.cy

[1]University of Cyprus, [2]Cyprus University of Technology, [3]Cyprus Acquisition Team

Abstract. A multitude of factors characterises multilingual compared to monolingual language acquisition. Two of the most prominent factors have recently been put in perspective and enriched by a third: age of onset of children's exposure to their native languages, the role of the input they receive, and the timing in monolingual first language development of the phenomena examined in bi- or multilingual children's performance. We suggest a fourth factor: language proximity, that is, the closeness or distance between the two or more grammars a multilingual child acquires. This paper reports on two types of data: (i) the acquisition and subsequent development of object clitics in two closely related varieties of Greek by monolingual, bilingual, and multilingual children, all of whom are also bilectal, and (ii) performance on executive control in monolingual, bilectal, and multilingual children. The populations tested come from several groups of children: monolingual speakers of Standard Modern Greek from Greece, multilingual children from Cyprus who speak the local variety (Cypriot Greek), the official language (Standard Modern Greek), and Russian or English (and some children even an additional language) - and what we call monolingual bilectal children, native acquirers of Cypriot Greek in the diglossic environment of Cyprus who also speak the official language but have not been exposed to any other languages. In addition, there are Hellenic Greek children (with two parents from Greece) and Hellenic Cypriot children (with one parent Hellenic Greek, the other Greek Cypriot) residing in Cyprus. On the basis of the measures mentioned, we want to establish a gradience of bilingualism which takes into account two very closely related varieties, in this case: Cypriot Greek and the standard language; the larger picture, however, is one that applies this approach to other countries and contexts in which two or more closely related varieties are acquired by children. The experimental findings suggest that bilectal children do indeed pattern somewhere in between monolingual and multilingual children in terms of vocabulary and executive control, yet at the same time none of the three groups exhibit significant differences in their pragmatic abilities; the often raised 'cognitive advantage' of bilingualism must thus have to be further distinguished and refined. The analysis of object clitic placement is more complex, however, crucially involving sociolinguistic aspects of language development, most importantly schooling.

Keywords: acquisition, clitic placement, Cypriot Greek, dialect, executive control, socio-syntax

Introduction

This paper is a shortened version of Grohmann and Kambanaros (to appear), which is an attempt to bring together different aspects of language development in order to make the case for 'comparative linguality'. By that, we mean that language abilities can be compared across populations that differ on a range of properties: different languages (e.g., English vs. Greek), different lingualism (e.g., mono- vs. bilingualism), different modality (e.g., spoken vs. signed), different age (e.g., child vs. adult), different development (typical vs. impaired), different health (normal vs. pathological), different genes (regular vs. implicated), and so on. Here we would like to present a subset of that research agenda, one that tackles the notion of *comparative bilingualism*, first introduced in this context by Grohmann (2014b). This constitutes a more focused line of research that aims at comparing different groups of bilingual speakers so as to discern what role particular language combinations may play in a child's language development. Of particular interest is the language proximity, for example, if one of the languages is a close relative if not even dialect of the other. But once one looks at the issues closer, it turns out that the picture points more in the direction of gradience of *multi*lingualism. For presentational purposes, we limit ourselves here to a discussion of typical bilectal and bi-/multilingual language development.

From the earliest studies of language development, it has become very clear that, despite fundamental similarities, monolingual language acquisition differs greatly from bi- and multilingual acquisition. Depending on where one sets the boundaries, it might even be held that monolingualism does not really exist, but that depends on how we classify sociolects, idiolects, and others that speakers command. The multilingual child faces a number of obstacles that do not factor into monolingual mother tongue acquisition. Two obvious and well studied factors are the age of onset of children's exposure to each of their two or more native languages and the role of the input they receive in each in terms of quantity and quality (e.g., Meisel, 2009; Genesee, Paradis, & Crago, 2011; Unsworth, Argyri, Cornips, Hulk, Sorace, & Tsimpli, 2014). In addition, the timing in monolingual first language development of the phenomena examined in bi- and multilingual children's performance has been argued to influence whether a particular linguistic phenomenon is acquired early, late, or very late (Tsimpli, 2014). The present paper addresses a fourth factor (Grohmann, 2014b), namely the closeness between the two or more grammars a multilingual child acquire, or language proximity.

Greek in Cyprus: Setting the stage

Considering the linguistic closeness or distance between the grammars of the two or more languages a multilingual child acquires allows us to further entertain the above-mentioned notion of 'comparative bilingualism'. The larger research agenda is one in which comparable phenomena are systematically investigated across bi- and multilingual populations with different language combinations, ideally arranged according to typological or perhaps even areal proximity. Our present contribution pursues a much more graspable goal, however, namely to compare different populations of Greek speakers on the same linguistic and cognitive tools. These include lexical and morphosyntactic tasks as well as measures on language proficiency, pragmatics, and especially executive control. The populations tested range from monolingual children growing up in Greece to multilingual children growing up in Cyprus, with several 'shades' in between, all centred around the closeness between the language of Greece (Demotic Greek, typically referred to by linguists as Standard Modern Greek) and the native variety of Greek spoken in Cyprus (Cypriot Greek, which itself comes in different flavours ranging from basi- to acrolect). Detailed family and language history background information was also collected for all participants.

The official language of Greek-speaking Cyprus is Standard Modern Greek (henceforth, SMG), while the everyday language, hence the variety acquired natively by Greek Cypriots, is Cypriot Greek (CG). Calling CG a dialect of SMG as opposed to treating it as a different language is largely a political question; the proximity between the two is very high, and obviously so: The two modern varieties largely share a common lexicon, sound structure, morphological rule system, and syntactic grammar. According to *Ethnologue* (Lewis, Simons, & Fennig, 2015), lexical similarity between CG and SMG lies in the range of 84%–93% (http://www.ethnologue.com/ethno_docs/introduction.asp): "Lexical similarity can be used to evaluate the degree of genetic relationship between two languages. Percentages higher than 85% usually indicate that the two languages being compared are likely to be related dialects." In turn, if at or below the 85% mark, it is not immediately clear that one must be a dialect of the other, which leaves more room for ambiguities such as the much debated fate of CG.

But CG and SMG slightly differ at all levels of linguistic analysis as well. To briefly illustrate, there are naturally numerous lexical differences, as expected in any pair of closely related varieties, such as the CG feminine-marked *korua* instead of SMG neuter *koritzi* 'girl'. Phonetically, CG possesses palato-alveolar consonants, in contrast to SMG, so SMG [cɛˈɾɔs] becomes CG [tʃɛˈɾɔs] for *keros* 'weather'. The two varieties use a different morpheme to mark 3rd person plural in present and past tenses, such as CG *pezusin* and *pezasin* instead of SMG *pezun* 'they play' and *pezan* 'they were playing'. On the syntactic level, SMG expresses focus by fronting to the clausal left periphery, while CG employs a cleft-like structure, which it also extensively uses in the formation of *wh*-questions. And there are even pragmatic differences such as in politeness strategies: For example, the extensive use of diminutives in SMG is considered exaggerated by CG speakers. See, among many others, Grohmann, Panagiotidis, & Tsiplakou (2006), Terkourafi (2007), Grohmann (2009), Arvaniti (2010), and Tsiplakou (2014) for recent discussions and further references.

Traditionally, Greek-speaking Cyprus is characterised by diglossia between the sociolinguistic L(ow)-variety CG and the H(igh)-variety SMG (Newton, 1972 and much work since, building on Ferguson, 1959; see e.g., Arvaniti, 2010; Hadjioannou, Tsiplakou, & Kappler, 2011; Rowe & Grohmann, 2013). Moreover, while there is a clear basilect (‘village Cypriot’), there are arguably further mesolects ranging all the way up to a widely assumed acrolect (‘urban Cypriot’); Arvaniti (2010) labelled the latter Cypriot Standard Greek (CSG), a high version of CG which is closest to SMG among all CG lects. In fact, such CSG may be the real H-variety on the island, on the assumption that without native acquirers of SMG proper, the only Demotic Greek-like variety that could be taught in schools is a ‘Cyprified Greek’, possibly this ostensible yet elusive CSG. However, SMG can be widely heard and read in all kinds of media outlets, especially those coming from the Hellenic Republic of Greece. Note also that there is still no grammar of CSG available, no compiled list of properties, not even a term, or even existence, agreed upon; the official language is SMG.

With respect to child language acquisition, it should come as no surprise that to date no studies exist that investigate the nature, quality, and quantity of linguistic input children growing in Cyprus receive. There are simply no data available that would tell us about the proportion of basi- vs. acrolectal CG, purported CSG, and SMG in a young child’s life, and whether there are differences between rural and urban upbringing or across different geographical locations. At this time, such information can only be estimated anecdotally. We follow recent work from our research group, the Cyprus Acquisition Team (CAT), and adopt Rowe and Grohmann’s (2013) term *(discrete) bilectalism* to characterise Greek Cypriot speakers. We further assume that Greek Cypriots are sequential bilectal, first acquiring CG and then SMG (or something akin, such as CSG), where the onset of SMG may set in with exposure to Greek television, for example (clearly within the critical period) but most prominently with formal schooling (around first grade, possibly before, where the relation to the critical period is more blurred). What is more, due to the close relations between Cyprus and Greece (beyond language for historical, religious, political, and economic reasons), we are able to tap into two further interesting populations, all residing in Cyprus (Leivada, Mavroudi, & Epistithiou, 2010): Hellenic Cypriot children, who are binational having one parent from Cyprus (Greek Cypriot) and one from Greece (Hellenic Greek), and Hellenic Greek children with both parents from Greece. Anecdotally, we could then say that binational Hellenic Cypriot children are presumably simultaneous bilectals (strong input in SMG and CG from birth), while Hellenic Greek children are arguably the closest to monolingual Greek speakers in Cyprus (SMG-only input from birth), though with considerable exposure to the local variety (CG) certainly, once they start formal schooling.

Report of case study I: Clitic placement and the socio-syntax of language development

Just as language development in bilingual children should be compared to that of monolinguals, different language combinations in bi- and multilingual children should be taken into consideration as well. Looking at the four purported dynamic metrics of assessment, we do not yet know how much Greek *input* the bilingual children in Cyprus receive, and how SMG-like it is (which also holds for the bilectals, as noted above). The same goes for the age of *onset* of SMG, if indeed prior to formal schooling, or the exact role of CSG in this respect. However, we do know for *timing* that object clitics appear very early in Greek, both SMG (Marinis, 2000) and CG (Petinou & Terzi, 2002). And lastly, with respect to language *proximity*, CG as a ‘dialect’ of Greek is by definition very close to SMG.

One of the best studied grammatical differences between the two varieties pertains to clitic placement (see Agouraki, 1997 and a host of research since): Pronominal object clitics appear postverbally in CG indicative declarative clauses, with a number of syntactic environments triggering proclisis, while SMG is a preverbal clitic placement language in which certain syntactic environments trigger enclisis. The acquisition of object clitics is arguably a “(very) early phenomenon”, as Tsimpli (2014) calls it, since clitics represent a core aspect of grammar and are fully acquired at around two years of age. Using a sentence completion task that aimed at eliciting a verb with an object clitic in an indicative declarative clause (Varlokosta, Belletti, Costa, Friedmann, Gavarró, Grohmann, Guasti, Tuller et al., 2015), we counted children’s responses to the 12 target structures in CG, which should consist of verb-clitic sequences (as opposed to clitic-verb in SMG).

For the purpose of this research, the COST Action A33 Clitics-in-Islands testing tool (Varlokosta et al., 2015) - originally designed to elicit clitic production even in languages that allow object drop, such as European Portuguese (Costa & Lobo, 2007) - was adapted to CG (from Grohmann, 2011). This tool is a production task for a 3rd person singular accusative object clitic within a syntactic island in each target structure, in which the target-elicited clitic was embedded within a *because*-clause (where the expected child response is provided in brackets and the clitic boldfaced):

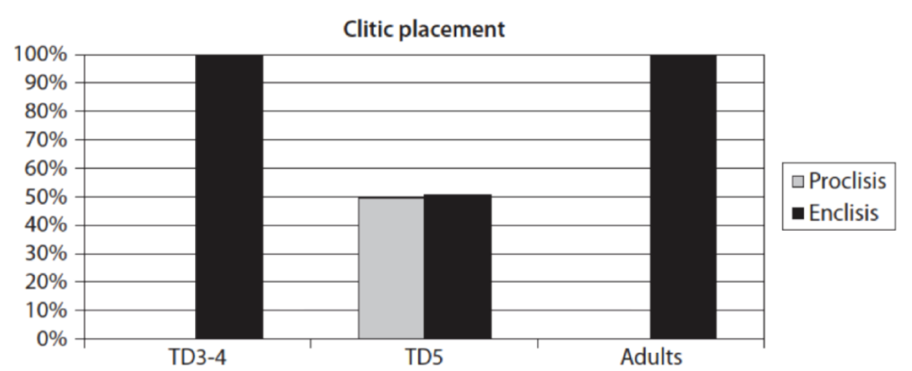
(1) To aγori vreʃi ti γata tʃe i γata e vremeni. Jati i γata e vremeni?
the boy wets the cat and the cat is wet why the cat is wet
I γata e vremeni jati to aγori… [vreʃi **tin**].
the cat is wet because the boy wet.PRES.3SG CL.ACC.3SG.FEM
'The boy is spraying the cat and the cat is wet. Why is the cat so wet?
The cat is wet because the boy… [is spraying it].'

The task involved a total of 19 items; 12 target structures (i.e. test items) after 2 warm-ups, plus 5 fillers. All target structures were indicative declarative clauses formed around a transitive verb, with half of them in present tense and the other half in past tense. Children were shown a coloured sketch picture on a laptop screen, depicting the situation described by the experimenter. The scene depicted in Figure 1 corresponds to the story and sentence completion in (1), for example.

(from Varlokosta et al., 2015)

Figure 1. Sample test item (clitics-in-islands task)

To anticipate the presentation and discussion of later results, the main pattern is consistent with the one originally reported for our first pilot study (Grohmann, 2011), which was confirmed and extended to many more participants in subsequent work (summarised in Grohmann, 2014a). This main pattern is provided in Figure 2.



(from Grohmann, 2011: 196)

Figure 2. Clitic placement in clitics-in-islands task (all tested groups)

With very high production rates in all groups (over 92%), the pilot study showed that the 24 three- and four-year-old children behaved like the 8 adult controls: 100% enclisis in the relevant context. In contrast, the group of 10 five-year-olds showed mixed placements, where that group is split further into three consistent sub-groups. This will be discussed in detail below.

All tests with Greek Cypriot bilectal children were carried out by native speakers of CG; those tests that were administered in SMG were done by a native SMG speaker. Testing was conducted in a quiet room individually (child and experimenter). Most children were tested in their schools or in speech-language therapy clinics, but a few were tested at their homes. It is well known that Greek Cypriots tend to code-switch to SMG or some hyper-corrected form of 'high CG' when talking to strangers or in formal contexts, as mentioned by Arvaniti (2010), Rowe and Grohmann (2013), and references cited there. For this reason, in an attempt to avoid a formal setting as much as possible (and thus obtain some kind of familiarity between experimenter and child), a brief conversation about a familiar topic took place before the testing started, such as the child's favourite cartoons.

All participants received the task in one session, some in combination with other tasks (such as those tested in Theodorou and Grohmann, 2015; see Theodorou, 2013). The particular task lasted no longer than 10 minutes. The pictures were displayed on a laptop screen which both the experimenter and the participant could see. The child participant heard the description of each picture that the researcher provided and then had to complete the *because*-clause in which the use of a clitic was expected; some participants started with *because* on their own, others filled in right after the experimenter's prompt of *because*, and yet others completed the sentence after the experimenter continued with the subject (the bracketed part in the example above).

No verbal reinforcement was provided other than encouragement with head nods and fillers. Self-correction was not registered; only the first response was recorded and used for data collection and analysis purposes. Regardless of a child's full response, what was counted were verb-clitic sequences only (for clitic production) and the position of the clitic with respect to the verb (for clitic placement). Testing was usually not audio- or video-taped, but answers were recorded by the researcher or the researcher's assistant on a score sheet during the session; many testing sessions involved two student researchers, with one carrying out the task and the other recording the responses (in alternating order). In those studies in which different clitic tasks were administered (Karpava & Grohmann, 2014) - not reported here - or where the same tool was tested in CG and SMG (Leivada et al., 2010), participants were tested with at least one week interval in between.

All these different studies with different populations and different age groups but the same tool show the following. First, the production rate of clitics in this task is very high from an early age on, safely around the 90% mark from the tested age of 2;8 onwards (lowest production at around 75%), over 95% at age 4;6 (lowest production at around 88%), and close to ceiling for 5-year-olds and beyond. The sub-group of 117 children from Grohmann, Theodorou, Pavlou, Leivada, Papadopoulou, and Martínez-Ferreiro (2012) performed as shown in table 1 (from Grohmann, 2014a, p. 17):

Table 1. Clitic production (adapted from Grohmann et al., 2012)

Age range (Number)	Overall clitic production	Target postverbal clitic placement
2;8–3;11 (N=26)	89.4%	89.2%
4;0–4;11 (N=21)	88.5%	88.0%
5;0–5;11 (N=50)	94.3%	68.0%
6;0–6;11 (N=20)	87.3%	47.0%
adult controls (N=8)	100%	100%

This said, Leivada et al. (2010) found considerably higher productions for the younger Hellenic Greek and Hellenic Cypriot children tested compared to their Greek Cypriot peers. However, just considering the 623 bilectal children analysed so far, we can confirm that the task was understood and elicited responses appropriate; in the widely tested age group of 5-year-olds, the production numbers are among the highest of all languages tested (Varlokosta et al., 2015), which means reliable data points for all 12 target structures; statistical analysis confirms that there were neither item effects nor test effects, that is, the productions for the 'long' (reported here) and 'short' version of the clitics tool (not reported here) are fully comparable (Grohmann, 2014a).

Second, and most importantly, the analysis of the 431 datasets of the bilectal children presented by Grohmann, Papadopoulou, and Themistocleous (submitted) are consistent with the findings of the much smaller pilot study. In other words, Figure 2 can be used as a general indicator: Up to around age 4, children reliably produce enclisis in this task at just shy of 90%, as expected (and confirmed by adult speakers), while we find considerable variation in clitic placement in the 5- to 7-year-olds.

To illustrate with the subset of 117 children again, when their non-target preverbal clitic placement productions were plotted according to chronological age, the resulting curve looks as in Figure 3 (from Grohmann & Leivada, 2011), where the x-axis indicates participants according to their chronological age and the y-axis non-target preverbal clitic placement in the participants' responses (percentage):

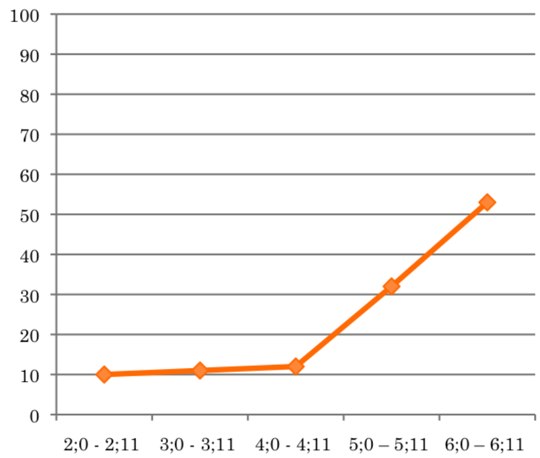


Figure 3. Non-target preverbal clitic placement (by chronological age)

However, what we can observe are apparent inconsistencies in terms of clitic placement, in particular by comparing younger with older children according to their schooling level. While for nursery children (mean age 3;3), target postverbal clitic placement lies at 93%, it decreases systematically for each additional year of formal schooling: kindergarten (4;3) at 82%, pre-school (5;5) at 73%, and first-grade (6;7) at 47% - from grade 2 onwards, the rates quickly shoot up towards 100% again (Grohmann, 2014a). This analysis is extended in Grohmann et al. (submitted). But using the same sub-group of 117 children again, compare Figure 3 above with Figure 4 (from Grohmann & Leivada, 2011), where the x-axis indicates participants according to their chronological age and the y-axis non-target preverbal clitic placement in the participants' responses (percentage).

The most striking result is that, while at the youngest ages, prior to formal schooling, the CG-target enclisis is produced predominantly, if not exclusively, once Greek Cypriot children start getting instructed in the standard language (SMG or some equivalent like CSG), their non-target productions of proclisis rise dramatically—all the way to second grade (not shown here; full analysis provided in Grohmann et al., submitted).

We suggest that these findings are best captured by the *Socio-Syntax of Development Hypothesis* (Grohmann, 2011), namely that an explicit 'schooling factor' is involved in the development of the children's grammar. Note that this grammatical development takes place past the critical period and does so possibly in combination with 'competing motivations' (Grohmann & Leivada, 2011; Leivada & Grohmann, in press). These arguably stem from the (at least) two grammars in the bilectal child's linguistic development that compete with each other. In other words, the Socio-Syntax of

Development Hypothesis can be seen as the specific trigger for the competing grammars of CG and SMG (and possibly CSG) in the development of clitic placement by young children speaking CG.

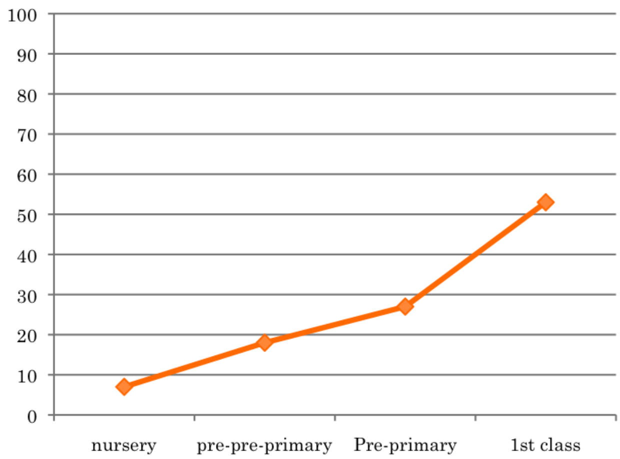


Figure 4. Non-target preverbal clitic placement (by schooling level)

Case study II: Cognitive advantage of bilectalism?

We will now turn to a first study on the purported bilingual status of Greek Cypriot bilectal children and its relevance for a more gradient, comparative bilingualism. The results from a range of executive control tasks administered to monolingual SMG-speaking children (in Greece) as well as CG–SMG bilectal and Greek–English bi-/multilingual children (in Cyprus) suggest that bilectal children behave more like their multilingual rather than their monolingual peers (Antoniou, Kambanaros, Grohmann, & Katsos, 2014) - that is, on a scale in between. A refined statistical analysis and additional discussion of this study can be found in Antoniou, Grohmann, Kambanaros, and Katsos (in press).

It has frequently been suggested that bilingualism bears an impact on children’s linguistic and cognitive abilities (e.g., Barac, Bialystok, Castro, & Sanchez, 2014). For example, as mentioned above in the context of Tsimpli (2014), bilingual children arguably have smaller vocabularies in each of their spoken languages as a result of input deficit. On the other hand, bilingual children seem to exhibit earlier development of pragmatic abilities, presumably compensating for their lower lexical knowledge by paying more attention to contextual information. And then there is the long-standing claim that bilingualism enhances children’s development of executive control, the set of cognitive processes that underlie flexible and goal-directed behaviour, commonly referred to as the ‘bilingual advantage’ or ‘cognitive advantage of bilingualism’ (Bialystok, 2009; Costa & Sebastián-Gallés, 2014). Taking a particular influential one of the many approaches to executive control, there is a tripartite distinction into working memory, task-switching, and inhibition (e.g., Miyake, Friedman, Emerson, Witzki, Howerter, & Wager, 2000).

This composite approach to executive control is arguably superior to an earlier suggestion that the bilingual advantage can be traced exclusively to more advanced inhibition alone (e.g., Bialystok, 2001). Here the idea was that, because both linguistic systems are activated when a bilingual speaks in one language, fluent use requires the inhibition of the other language. This constant experience in managing two active conflicting linguistic systems via inhibition enhances bilinguals’ inhibitory control mechanisms. This early view, however, has been challenged on several grounds (e.g., Bialystok, Craik, & Luk, 2012). One line of argument would be that the advantageous effects of bilingualism have been observed for the very first years of life, even for 7-month-old infants (Kovacs & Mehler, 2009). Since for bilingual infants language production has not yet started, there would be no need to suppress a non-target language. We are not sure that this argument goes through, though: After all, even bilingual infants are fully aware of the different languages they are acquiring, and while they may not need to inhibit one to produce the other, they presumably process the two (or more) languages and should therefore regularly inhibit one to process the other. However, there are a number of further arguments to take a more differentiated view on executive control as the measuring stick for the bilingual advantage, as put forth in many of the references cited above; see also Antoniou et al. (2014) for further discussion.

All in all, an advantage in executive control may be the result of constantly having to manage two different linguistic systems. So, one aspect of continued research on the topic would be to disentangle the different sub-components of executive control and determine which aspect(s) of executive control really relates to a bilingual advantage. Regarding performance on executive control in monolingual, bilectal, and bi- or multilingual children, our research question is then (Antoniou et al., 2014): What is the effect of bilectalism on children's vocabulary, pragmatic, and executive control skills?

A total of 136 children with a mean age of just above seven-and-a-half years of age participated in the study (Antoniou et al., 2014): 64 Greek Cypriots, bilectal in CG and SMG, aged 5-12 (mean age: 7;8); 47 residents of Cyprus, multilingual in CG, SMG, and English (plus in some cases an additional language), aged 5-12 (mean age: 7;8); and 25 Hellenic Greeks, monolingual speakers of SMG, aged 6-9 (mean age 7;4). Socio-economic status measures included the Family Affluence Scale (Currie, Elton, Todd, & Platt, 1997) and level of maternal and paternal education obtained through questionnaires. Since the multilingual children all attended a private English-medium school in Nicosia, their socio-economic was higher than the mean of all other participants.

A range of language proficiency measures were administered for expressive and receptive vocabulary, including the Greek versions of the Word Finding Vocabulary Test for expressive vocabulary and the revised Peabody Picture Vocabulary Test (SMG) as well as the Greek Comprehension Test (for either variety). For pragmatic performance, a total of 6 tools were used, tapping into relevance, manner implicatures, metaphors, and scalar implicatures; the bilectal and multilingual children received the test in CG, 17 bilectals took the test in both CG and SMG, and the monolinguals were tested in SMG only. As for non-linguistic performance, the WASI Matrix Reasoning Test was used to assess participants' non-verbal intelligence. The executive control tasks administered included a wide range of batteries. For verbal working memory, the Backward Digit Span Task was employed, and for visuo-spatial working memory, an online version of the Corsi Blocks Task. Inhibition was assessed through Stop-Signal and the Simon Task, and switching through the Colour-Shape Task. (For more details and references, see Antoniou et al., 2014.)

The preliminary results from this study can be presented across four types of group comparisons (Antoniou et al., 2014, building on Antoniou et al., 2013 but preliminary compared to Antoniou et al., in press). The first concerns background measures. The relevant subsets of the three participant groups of bilectal (n=44), multilingual (n=26), and monolingual children (n=25) were intended to be matched for age and gender; they did not statistically differ on age ($F(2, 92) = .696$, $p > .05$) or gender ($F(2, 92) = .587$, $p > .05$). However, they did differ on socio-economic status ($F(2, 89) = 9.622$, $p < .0001$), with the private-schooled multilingual children as a group coming from a higher socio-economic family background than the monolingual ones, and the bilectals from the lowest. The three groups also differed on non-verbal IQ ($F(2, 92) = 3.492$, $p < .01$), with the multilingual children higher than the two other groups, which did not differ significantly.

Next we compared the three participant groups' performance on the vocabulary measures. The multilingual children had a significantly lower vocabulary score than the bilectals, who in turn had a significantly lower vocabulary than the monolinguals, with both *p*s>.005 ($F(2, 89)=35.531$, $p<0.001$). From what is known about vocabulary growth in bilingual contexts (see references above), it was expected that the monolingual children would outperform the multilinguals; the fact that the bilectals fall in between fits nicely with our hypothesis that, on a gradient scale, bilectalism lies somewhere in between mono- and multilingualism.

The third group comparison concerns performance in the pragmatic tasks. Analyses of covariance (ANCOVAs), with vocabulary and SES and IQ as covariates, showed no significant differences between the three groups across all pragmatics tasks ($F(2, 87)=4.081$, $p<.05$), as shown in the six graphs, one for each task administered. No differences in the pragmatic tasks suggest that even those children who exhibit some sort of lower language (multilinguals, perhaps bilectals), they still show comparable pragmatic performance at the same age. With an eye on the Greek Cypriot bilectal children, this again suggests that they pattern somewhere in between; given the lower vocabulary scores compared to their monolingual peers from Greece (see second group comparison right above), they do perform the same in the six pragmatic tasks.

Lastly, and for the purposes of our research question perhaps most importantly, the child participants' performance on the executive control tasks was analyzed and submitted to principal component analysis. All three global executive control scores (working memory, inhibition, and switching) positively correlated with IQ. ANCOVAs on the three composite scores for executive control, with Group as a between-subjects factor and IQ, linguistic knowledge (Greek), age, and SES as covariates, revealed a significant effect of group only for the overall executive control score: a significant multilingual advantage over monolinguals, with a trend for a bilectal advantage. We illustrate this finding here with switch cost: Bilectals performed better than monolinguals in the congruent switch trials, with no other significant comparisons ($F(2, 87)=4.081$, $p<.05$); in the incongruent switch trials, bilectals also performed better than monolinguals ($F(2, 87)=5.805$, $p<.005$), with multilinguals almost better than monolinguals ($p=.108$).

These results can be summarised as showing that the bilectal children performed better than the monolinguals in overall executive control ability and slightly worse than multilinguals. With respect to the lack of a clear effect for switching, as opposed to vocabulary, for example, we would like to suggest that there is interference from language proximity: The more similar the two varieties, the more difficult it is to switch—or rather, the less there is a need to switch. For example, in a given group of individuals of whom all but one speak Greek and English, with one knowing no Greek, a Greek-language discussion would be translated or summarised in English for that individual (switching by the bilingual speaker(s)). In contrast, in a group of Greek speakers of whom only one does not speak Cypriot Greek, a CG-at large discussion would arguably not be translated or summarised in SMG for that individual (no switching by the bilectal speaker(s)). As noted in a different context by Runnqvist, FitzPatrick, Strijkers, & Costa (2012), this may in fact tie in with the reverse of a bilingual advantage, what they call the 'bilingual disadvantage'. Beyond the cases they examine (see e.g., Ivanova, & Costa 2008; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009), it has also been suggested that the cognitive advantage only surfaces in bilingual individuals who actually switch between their languages frequently (Prior & Gollan, 2011).

In terms of a larger discussion, we hasten to add that there is recent work that casts some doubt on the purported relation between bilingualism and executive control abilities (e.g., Paap & Greenberg, 2013; Paap & Sawi, 2014). Just like the above-mentioned modifications to the 'right' kind of model of executive control, there are a number of factors that make more careful investigations even more important. In the study reported here (Antoniou et al., 2014), for example, we compared group performances. However, the groups were composed of children of a considerable age range (almost three years) and, for obvious reasons for the populations chosen, there were significant differences in socio-economic status and non-verbal intelligence. Likewise, it is not yet clear in how much, if at all, the cognitive advantage observed in bilingualism pertains or increases in multilingualism. These are some of the considerations that our future work will aim to improve on in order to assess the purported bilingual advantage in executive control abilities in bilectal speakers as well as finer grained and better selected multilingual groups for comparison. In this context, we also hope to get a better idea on the possible role played by language proximity in executive control.

Overall discussion and outlook

We take the grammar of multilingualism to be a highly complex area of research that by definition needs to include a lot of different measurements—by which we mean, ideally, the investigation of different measures, different sets of data, different populations, all carried out by interdisciplinary research teams. There is a need for thorough sociolinguistic work, putting the languages under investigation into their social and communicative context, for example. There is a need for thorough theoretical linguistic work, identifying the relevant structures and patterns to be investigated. There is a need for thorough psycholinguistic work, designing and carrying out the best possible experimental methodology. There is a need for cognitive psychological work, probing executive control abilities. And there is a need for clinical linguistic work, assessing and treating language impairment.

This list can be added to and enriched in many ways. The bottom line is that the notion of *comparative bilingualism* can be quite useful and instructive for future research activities, especially when carried out across different countries and languages. The narrow goal of this article was thus to draw attention to this state of affairs and elaborate the research path of comparative bilingualism (Grohmann, 2014b), with a focus on Cyprus (Grohmann & Leivada, 2012, 2013; Kambanaros, Grohmann, & Michaelides, 2013b; Rowe & Grohmann, 2013, 2014; Karpava & Grohmann, 2014). One such intriguing path would be the role of comparative bilingualism for children with developmental language impairment, something we pointed to as well (Kambanaros et al., 2013a, 2014, 2015), even for therapy strategies (Kambanaros et al., to appear).

Putting all of this together, though, there is an even more general issue. Comparing cognitive and linguistic abilities across different populations and different groups of speakers may ask for a further 'specialised' area of research. The intention is to compare linguistic and cognitive abilities of monolingual, bidialectal, bilectal, bilingual, and multilingual speakers (comparative bilingualism, with more room for gradience, especially in combination such as Russian-Greek bilinguals in Cyprus) and different language-impaired populations (comparative biolinguistics, unearthing phenotypal variation), who themselves may be on different scales in the gradient spectrum of multilingualism. That is, among the future research participants, there will be vast variation and combinations of 'lingual' features, ranging from mono- to multilingualism, from simultaneous to sequential acquisition, from local to heritage language status, from typical development to impairment, from healthy to disorders of various degrees. We tentatively suggest a(nother) new term for this and are excited about what future research may bring: *comparative linguality*.

References

Agouraki, Y. (1997). On the enclisis/proclisis alternation. In: G. Drachman, A. Malikouti-Drachman, J. Fykias, & C. Klidi (eds.), *Proceedings of the Second International Conference on Greek Linguistics*, vol. II (pp. 93-404). Salzburg: University of Salzburg.

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (in press). The effect of childhood bilectalism and multilingualism on executive control. *Cognition*.

Antoniou, K., Kambanaros, M., Grohmann, K. K., & Katsos, N. (2013). *The effect of bilectalism on children's vocabulary, pragmatic and executive control abilities*. Talk presented at the *38th Boston University Conference on Language Development (BUCLD 38)*, Boston University (1-3 November 2013).

Antoniou, K., Kambanaros, M., Grohmann, K. K., & Katsos, N. (2014). Is bilectalism similar to bilingualism? An investigation into children's vocabulary and executive control skills. In: W. Orman & M. J. Valleau (eds.), *BUCLD 38: Proceedings of the 38th annual Boston University Conference on Language Development*, vol. 1, (pp. 12-24). Somerville, MA: Cascadilla Press.

Arvaniti, A. (2010). Linguistic practices in Cyprus and the emergence of Cypriot Standard Greek. *Mediterranean Language Review, 17*, 15-45.

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly, 29*, 699-714.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge, UK: Cambridge University Press.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*, 3-11.

Bialystok, E., Craik, F., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences, 16*, 240-250.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*, 135-149.

Costa, J., & Lobo, M. (2007). Clitic omission, null object or both in the acquisition of European Portuguese?. In S. Baauw, F. Drijkoningen, & M. Pinto (eds.), *Romance languages and linguistic theory 2005* (pp. 59-72). Amsterdam: John Benjamins.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience, 15*, 336-345.

Currie, C. E., Elton, R. A., Todd, J., & Platt, S. (1997). Indicators of socioeconomic status for adolescents: The WHO Health Behaviour in School-Aged Children Survey. *Health Education Research, 12*, 385-397.

Ferguson, C. A. (1959). Diglossia. *Word, 15*, 325-340.

Genesee, F., Paradis, J., & Crago, M. B. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning*, 2nd ed. Baltimore, MD: Paul Brookes.

Grohmann, K. K. (2009). Focus on clefts: A perspective from Cypriot Greek. In A. Tsangalides (ed.), *Selected papers from the 18th International Symposium on Theoretical and Applied Linguistics - Thessaloniki, 4-6 May 2007* (pp. 157-165). Thessaloniki: Monochromia.

Grohmann, K. K. (2011). Some directions for the systematic investigation of the acquisition of Cypriot Greek: A new perspective on production abilities from object clitic placement. In E. Rinke & T. Kupisch (eds.), *The development of grammar: Language acquisition and diachronic change* (pp. 179-203). Amsterdam: John Benjamins.

Grohmann, K.K. (2014a). CAT research on object clitic placement: Where we are now. In K. K. Grohmann & T. Neokleous (eds.), *Developments in the acquisition of clitics* (pp. 1-40). Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Grohmann, K. K. (2014b). Towards comparative bilingualism. *Linguistic Approaches to Bilingualism, 4*, 336-341.

Grohmann, K. K., & Kambanaros, M. (to appear). The gradience of multilingualism in language development: Positioning bilectalism within comparative bilingualism. *Frontiers in Psychology: Language Sciences*.

Grohmann, K.K., & Leivada, E. (2011). *Towards disentangling bi-x: Metalinguistic awareness and competing motivations*. Paper presented at the *2011 UIC Bilingualism Forum*, University of Illinois at Chicago (14-15 April 2011).

Grohmann, K. K., & Leivada, E. (2012). Interface ingredients of dialect design: Bi-*x*, socio-syntax of development, and the grammar of Cypriot Greek. In A. M. Di Sciullo (ed.) *Towards a biolinguistic understanding of grammar: Essays on interfaces* (pp. 239-262). Amsterdam: John Benjamins.

Grohmann, K. K., & Leivada, E. (2013). Lightverbhood in child language: Evidence from Cypriot Greek. Ine E. Blom, J. Verhagen, & I. van de Craats (eds.), *Dummy auxiliaries in first and second language acquisition* (pp. 141-170). Berlin: De Gruyter Mouton.

Grohmann, K. K., Panagiotidis, P., & Tsiplakou, S. (2006). Properties of *wh*-question formation in Cypriot Greek. In M. Janse, B. D. Joseph, & A. Ralli (eds.), *Proceedings of the 2nd International Conference on Modern Greek Dialects and Linguistic Theory (Mytilene, Greece: 30 September-3 October 2004)* (pp. 83-98). Patras: University of Patras.

Grohmann, K. K., Papadopoulou, E., & Themistocleous, C. (submitted). Acquiring clitic placement in bilectal settings: Interactions between social factors.

Grohmann, K. K., Theodorou, E., Pavlou, N., Leivada, E., Papadopoulou, E., & Martínez-Ferreiro, S. (2012). The development of object clitic placement in Cypriot Greek and the Romance connection. In S. Ferré, P. Prévost, L. Tuller, & R. Zebib (eds.), *Selected proceedings of the Romance Turn IV workshop on the acquisition of Romance languages* (pp. 128-152). Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Hadjioannou, X., Tsiplakou, S., & Kappler, M. (2011). Language policy and language planning in Cyprus. *Current Issues in Language Planning, 12*, 503-569.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica, 127*, 277-288.

Kambanaros, M., Grohmann, K. K., & Michaelides, M. (2013) [= Kambanaros et al. (2013b)]. Lexical retrieval for nouns and verbs in typically developing bilectal children. *First Language, 33*, 182-199.

Kambanaros, M., Grohmann, K. K., Michaelides, M., & Theodorou, E. (2013) [= Kambanaros et al. (2013a)]. Comparing multilingual children with SLI to their bilectal peers: Evidence from object and action picture naming. *International Journal of Multilingualism, 10*, 60-81.

Kambanaros, M., Grohmann, K. K., Michaelides, M., & Theodorou, E. (2014). On the nature of verb–noun dissociations in bilectal SLI: A psycholinguistic perspective from Greek. *Bilingualism: Language and Cognition, 17*, 169-188.

Kambanaros, M., Michaelides, M., & Grohmann, K. K. (2015). Measuring word retrieval deficits in a multilingual child with SLI: Is there a better language? *Journal of Neurolinguistics, 34*, 112-130.

Kambanaros, M., Michaelides, M., & Grohmann, K. K. (to appear). Cross-linguistic transfer effects after cognate-based therapy in a case of multilingual specific language impairment (SLI). *International Journal of Language & Communication Disorders*.

Karpava, S., & Grohmann, K. K. (2014). Bilingual acquisition of Cypriot Greek object clitic placement. In K. K. Grohmann & T. Neokleous (eds.), *Developments in the Acquisition of Clitics* (pp. 87-137). Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Leivada, E., & Grohmann, K. K. (in press). Language acquisition in bilectal environments: Competing motivations, metalinguistic awareness, and the socio-syntax of development hypothesis. In M. Katerbow & G. de Vogelaer (eds.), *Variation in language acquisition*. Amsterdam: John Benjamins.

Leivada, E., Mavroudi, P., & Epistithiou, A. (2010). Metalanguage or bidialectism? Acquisition of clitic placement by Hellenic Greeks, Greek Cypriots and binationals in the diglossic context of Cyprus. In A. Botinis (ed.), *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics 2010* (pp. 97-100). Athens: ISCA and the University of Athens.

Lewis, M. P., Simons, G. F., & Fennig, C. D. (eds.) (2015). *Ethnologue: Languages of the world* (18th ed.) Dallas, TX: SIL International. [online version: http://www.ethnologue.com]

Marinis, T. (2000). The acquisition of clitic objects in Modern Greek: Single clitics, clitic doubling, clitic left dislocation. *ZAS Working Papers*, *15*, 259-281.

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft, 28*, 5-34.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex ‘frontal lobe’ tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232-258.

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology: Language Sciences*, *5*, 962, doi:10.3389/fpsyg.2014.00962.

Petinou, K., & Terzi, A. (2002). Clitic misplacement among normally developing children and children with specific language impairment and the status of Infl heads. *Language Acquisition*, *10,* 1-28.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychology Society*, *17*, 682-691.

Rowe, C., & Grohmann, K. K. (2013). Discrete bilectalism: Towards co-overt prestige and diglossic shift in Cyprus. *International Journal of the Sociology of Language*, *224*, 119-142.

Rowe, C., & Grohmann, K. K. (2014). Canaries in a coal mine: Native speakerhood and other factors as predictors of moribundity, death, and diglossic shift in Cypriot Greek. *Mediterranean Language Review*, *21*, 121-142.

Runnqvist, E., FitzPatrick, I., Strijkers, K., & Costa, A. (2012). An appraisal of the bilingual language production system: Quantitatively or qualitatively different from monolinguals? In T. K. Bhatia & W. C. Ritchie (eds.), *The handbook of bilingualism and multilingualism*, (2nd ed., pp. 244-265). Malden, MA: Wiley-Blackwell).

Terkourafi, M. (2007). Perceptions of difference in the Greek sphere: The case of Cyprus. *Journal of Greek Linguistics*, *8*, 60-96.

Theodorou, E. (2013). *Specific language impairment in Cypriot Greek: Diagnostic and experimental investigations*. University of Cyprus, Nicosia: PhD dissertation.

Theodorou, E., & Grohmann, K. K. (2015). Object clitics in Cypriot Greek children with SLI. *Lingua*, *161*, 144-158.

Tsimpli, I. M. (2014). Early, late or very late: Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism*, *4,* 393-402.

Tsiplakou, S. (2014). How mixed is a ‘mixed’ system? The case of the Cypriot Greek *koiné*. *Linguistic Variation*, *14*, 161-178.

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. M. (2014). On the role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, *35,* 765-805.

Varlokosta, S., Belletti, A., Costa, J., Friedmann, N., Gavarró, A., Grohmann, K. K., Guasti, M. T., Tuller L., et al. (2015). A crosslinguistic study of the acquisition of pronoun and clitic production. *Language Acquisition*, doi:10.1080/10489223.2015.1028628.

Are speech sound disorders phonological or articulatory? A spectrum approach

David Ingram[1], Lynn Williams[2], Nancy J. Scherer[3]
david.ingram@asu.edu, williamsl@etsu.edu, nancy.scherer@asu.edu

[1]Arizona State University, [2]East Tennessee State University,
[3]Arizona State University

Abstract. An articulatory disorder is one that is solely articulatory, that is, the child can't produce the sound or sounds in question, e.g., lisping. A child with a phonological disorder also has an articulation problem, but the articulation is influenced by the acquisition of phonological representations. One way this may be seen involves a *substitution shift*, i.e. using a sound as a substitution but not as a correct consonant. An example would be a child producing /ʃ/ as [s], e.g., 'shoe' [su], but producing /s/ as [t], e.g., 'Sue' [tu]. Another way is when a child shows an interaction between consonant correctness and word complexity. For example, a child can produce a sound correctly in simple words, but has lower rates of correctness in more complex words, e.g., an /s/ in "Sue" versus an /s/ in "surprise". Substitution shifts and effects of word complexity are common in typical phonological acquisition. They are also found in a subtype of children with speech sound disorders, those who are primarily having a *phonological delay*. They are less common with other children who have speech sound disorders, who can be considered to be having a *phonological disorder*. The patterns of the latter group of children are not comparable to those of younger, typically developing children. These children are more apt to be unable to produce certain sounds regardless of word complexity. At the same time, they show patterns in their speech production that provide evidence that they are attempting to form and use phonological representations. That is, they do not solely have an articulation disorder. This article proposes that children with speech sound disorders fall along a spectrum, with articulation at one end and phonology at the other, e.g., lisp > disorder > delay. We demonstrate this with the case study of a child CS, and with a group study of children with cleft lip/palate. CS could not produce any fricatives other than /f/, nor affricates, and liquids. The substitution patterns were /s/ and /ʃ/ [θ], /tʃ/ [t] or [θ], /z/ [ð], /dʒ/ [d] or [ð], /θ/ [f], /v/, /r/ /l/ [w]. His errors show articulatory difficulties, yet he also used a substitution shift with [θ]. The second example is a study of 20 children with cleft lip/palate. The children showed a range of atypical phonological patterns triggered by the articulatory difficulty resulting from the repaired clefts. These had the look of children with phonological disorders. Samples taken later showed marked improvement, a shift from looking disordered to looking delayed. These cases suggest that assessment needs to identify the articulatory and phonological influences on a child's speech, and recognize that they fall along a spectrum from one to the other.

Keywords: articulation, phonology, speech disorders, markedness

Articulation versus phonology

It is the case that categorical perception, at least metaphorically, is a part of a number of areas of language acquisition. One such case is in the discussion of children's speech sound disorders (SSD). SSD are commonly discussed as being either articulatory or phonological in nature. This is also reflected in book titles on the topic over several years, e.g., *Phonological Disability in Children* (Ingram, 1976), *Articulation Disorders* (Sommers, 1983), *Normal and Disordered Phonology in Children* (Stoel Gammon & Dunn, 1983), etc. Very recently, the term SSD has been coined to reflect a neutral position as to which is being discussed as in *Children's Speech Sound Disorders* (Bowen, 2015).

At one end, there is some consensus that certain SSD are articulatory in nature. Some possible candidates for this account are problems with tongue thrusting referred as lisping, [r] problems, and possibly cleft lip/palate. These cases (at least the two first) involve a distorted phoneme, not the merger of one phoneme with another. Another example from children with clefts are backed compensatory substitutions. Both phonemic merger and phonological contrast are at the heart of the claims of a phonological influence on early speech. One of the earlier and most well-known proposals for the influence of phonology on speech development is that of Roman Jakobson in his seminal work *Child Language, Aphasia, and Phonological Universals* (Jakobson, 1941/1968). At the heart of Jakobson's view is the notion of maximal contrast. Suppose (contrary to Jakobson) that ease of articulation was the reason that children acquire consonants in the order that they do. One would anticipate that the earliest consonants would be unaspirated [p] and [t], then [k]. Jakobson, however, emphasized that such an account only works when surface productions are taken into consideration. Patterns of substitution, however, show the influence of phonology that underlies the child's productions. Early in his book, Jakobson discusses a Russian child who showed the patterns in (1).

(1) Maximal Contrast /g//t/: /t/ pronounced as [t]; /k/ pronounced as [t]; /g/ pronounced as [k]

In this example, the child does not produce a [k] for /k/, but does so for /g/. To say that the child could not articulate a [k] would be incorrect. He could, but only when it functioned to capture a maximal phonological difference within the stops, between /g/ and /t/.

These two simple cases are relatively clear and lead to a conclusion that children with speech sound disorders are articulatory if they show problems like lisping, and are phonological if they show patterns as in (1). Further, children can additionally distinguished by those who are typically developing, and those who are having a phonological delay. The latter group would be children showing instances of maximal contrast like Jakobson's Russian case, but who are older and thus slower developing. More in depth analyses show that the distinction is not always as easy to separate as suggested above. This was known even before Jakobson's work. A little cited study by Margaret Nice entitled "A child who would not talk" (Nice, 1925) demonstrated this. Nice studied the speech and language development of four of her children over several years. In this article, she contrasted the speech development of her fourth child with the other three. This child was a slow language learner, with a vocabulary of just around 50 words at age 3. Further, she demonstrated a very restricted inventory of speech sounds, suggestive of an articulation disorder. At the same time, she showed usage of reduplication, a phonological pattern that enabled her to expand her limited vocabulary despite her speech limitations. She showed both articulatory problems plus phonological development. She was delayed, but also could be claimed to be disordered. Without any speech intervention, however, she showed tremendous gains in both speech and language after 3,0 and caught up with her sisters by age four.

What we would like to suggest is that the distinction between articulatory and phonological development should not be treated categorically. We suggest alternatively that the two should be seen as end points on a 'spectrum'. A child whose speech is entirely explainable by articulation would be at one end, while other children showing both articulatory and phonological influences would fall along the spectrum toward a stronger influence of phonological factors. This will be done by demonstrating instances where a child's speech development shows a mix of both articulatory and phonological patterns.

There are two topics in particular that we will discuss. One is an examination of the spectrum proposal with a focus on 'whole word complexity'. It will be shown that some, but not all children show a correlation between consonant correctness and whole word complexity. That is, rates of consonant correctness are higher in simpler words than in more complex words. The second focus will be on the distinction between phonological delay and phonological disorder. We will demonstrate that this distinction can be made when children's whole word complexity is analyzed. Children who show a correlation between whole word and consonant correctness fall on the phonological end of the spectrum, while children who do not, fall towards the articulation side.

Whole word complexity

Recent work by the first author has examined whole word complexity in children by measuring consonant correctness in relation of word complexity (WC). WC is a relative term, with many possible levels. The initial research has focused on two aspects that influence word complexity, these being the occurrence of consonant clusters, and the occurrence of multiple syllables. This two dimensions lead to the four categories listed in (2).

(2) a. monosyllabic words with just single consonants (mono single)

b. monosyllabic words with consonant clusters (mono cluster)

c. multisyllabic words with just single consonants (multi single)

d. multisyllabic words with clusters (multi cluster)

The preliminary results with typically developing children has found that both syllabicity and clusters add to word complexity, as measured by the percentage of correct consonants for each (PCC). This is demonstrated in Table 1 where the results of two children, Ian and Jennika, are presented for the categories in (2). Note that there is a noticeable range of PCC scores, with the mean around 40%. These results have been replicated across a number of children.

The next phase of this line of research has explored the same analyses for children with SSD. It was found that one group of children with SSD showed a similar pattern to the typically developing children in that they showed a correlation between word complexity and PCC. These children were considered to be having a phonological delay. The results for two children from this group, Tim and Barry are given in Table 1. A second group, however, did not show a significant correlation. For these children, the lack of this effect was interpreted as an articulatory effect. That is, there were certain sounds that they could not make, regardless of a word’s complexity. These children were concluded to have an articulatory problem. The results for two children in this group, Alan and Danny, are given in Table 1. While preliminary, the range of PCC scores can be examined to see the difference. The range of scores for children with phonological delay are limited (around 10% or less), while those for children with delay are greater (means around 25%).

Table 1. PCC scores and word complexity

Children	**Age**	**Range**	**Mono Single**	**Multi Single**	**Mono Cluster**	**Multi Cluster**
Ian	1,9	33%	71%	53%	36%	38%
Jennika	1,11	30%	70%	56%	44%	40%
Tim	5,0	21%	26%	37%	16%	19%
Barry	8,9	31%	71%	68%	40%	40%
Alan	5,11	9%	18%	16%	20%	11%
Danny	5,6	14%	34%	26%	20%	30%

Evidence for a spectrum

Observe that these preliminary results followed the categorical interpretation discussed earlier. Here we discuss two further analyses that led to the spectrum proposal. One is a case study of a child CS who was diagnosed with a SSD. We demonstrate that an analysis of CS speech sample demonstrated both articulatory and phonological influences. The second study is a longitudinal study conducted on the speech of 20 two-year-old children with repaired child lip/palate. The first samples from the children showed a wide range of atypical speech patterns. Analyses found that each child used their own phonological strategies to cope with their clefts. Samples taken approximately ten months later showed significant improvements and typical phonological acquisition.

CS

CS was diagnosed as having a SSD with multiple sound substitutions. To determine his phonological system, CS was given a phonological assessment with approximately 300 words by the second author. It was determined that his consonant inventory consisted of two distinct consonant groups. There were 12 consonants that CS produced correctly with virtually 100% accuracy. These were /m, n, p, t, k, b, d, g, w, j, f, h/. At the other extreme, there were consonants that were produced with virtually 0% accuracy. He had no correct fricatives except /f/, no affricates, and no liquids. In other words, his consonants were either completely correct or completely incorrect. A preliminary conclusion was that he has a particular problem with articulation since his percentage of correct consonant production was not influenced by word complexity.

A relational analysis was then conducted to determine his patterns of substitution. Given that he was diagnosed as having primarily an articulation problem, it was predicted that his substitutions would follow well known *markedness* relations. That is, it was anticipated that his substitutions would consist of sounds he could produce for those more marked sounds that he could not. These predictions are shown in Table 2 along with the actual substitutions.

Table 2. CS substitution patterns

Class	Phonemes	Predicted	Actual
Fricatives	/s/, /ʃ/	[t]	[θ]
	/z/	[d]	[ð]
	/θ/	[f]	[f]
	/v/	[w]	[w]
Affricates	/tʃ/	[t]	[t] or [θ]
	/dʒ/	[d]	[d] or [ð]
Liquids	/r/, /l/	[w]	[w]

The results only partially supported the predictions of markedness. The phonemes that met the predictions were the liquids, and the fricatives /θ/ and /v/. The affricates varied between the predicted stop consonants and the more marked dental fricatives. The lingual fricatives /s/, /ʃ/ and /z/ did not meet the predictions. CS substituted the more marked dental fricatives.

The more marked substitutions can be accounted for using the phonological predictions of maximal contrast as discussed above. First, notice that CS can make dental fricatives, e.g., /s/ to [θ], but he does not correctly use it for /θ/, where he replaces the target with [f]. If he were to replace the lingual fricatives with stops, he would lose the underlying distinction between stops and continuants. By using the more marked dental fricatives, the distinction is maintained. This phonological influence competes with the unmarked options in the affricates. That is, the affricates vary between the predicted stops and the marked dental fricatives as with the lingual fricatives. The liquids met the predictions, but note that the underlying feature distinction between obstruents and sonorants is maintained with the unmarked option. We conclude that CS is a child who is on the articulatory side of the phonological spectrum, but still shows the underlying influence of acquiring the English phonological system.

Cleft Lip/Palate

The second study providing evidence for a phonological spectrum approach to SSD concerns a recent study on the speech of children with cleft lip/palate (CLP) (Ingram & Scherer, 2015). The study examined the speech development of 20 two-year old children with repaired CLP using the Profiles of Early Expressive Phonology Skills (PEEPS) assessment test (Stoel-Gammon & Williams, 2013). The speech samples were then analyzed using a multidimensional approach (Ingram & Dubasik, 2011). The analyses looked at whole word measures (Ingram, 2002), syllable shapes, and included relational and independent analyses.

Table 3. Atypical patterns in CLP

Productions	Percentage
No Fricatives	65%
Velar Fronting	50%
No Final Consonants	40%
Voice Ahead of Place, e,g,, /b,p,t,d/	30%
Glottal Compensation	25%
Gross Inclusion, e.g. [d] /d, s, dʒ, k, g /	10%
Other	75%

The children in the first samples showed a wide range in their PCC scores, 12% to 63% (x = 39%). The relational analyses revealed that each child showed a mix of typical and atypical phonological patterns. While no consensus exists concerning what is considered an atypical pattern, we identified six such potential categories. Table 3 presents them along with the percentage of children who showed instances of each. The first two categories identify an inability to produce fricatives and velars in any word position. The third identifies a lack of any or no more than one final consonant. Voice Ahead of Place is a pattern that emerges in some children who have a lack of velars. They proceed to acquire a voice distinction for lingual and/or coronal while this acquisition typically follows the acquisition of velars. Glottal compensation is an over reliance on glottals, a pattern discussed in the CLP literature. Gross Inclusion is a pattern discussed in Grunwell (1985) in which a single consonant, often [d], is used for a wide range of phonemes (at least four in our analyses). Other more idiosyncratic patterns were also noted and placed under the other category.

It was also the case that the children all had at least one atypical pattern, and some as many as 6. There were six children with 1 or 2, seven children with 3 or 4, and seven children with 5 or 6.

Ten children were selected for similar analyses taken approximately 10 months later. During this time, the children received 48 sessions of an early speech and language intervention that targeted lexical and speech sound production (primarily consonant production). There were great improvements for all the children at the second sample points. The mean PCC scores in sample 1 were 39%, and this doubled in the second samples to a mean of 80%. Relational analyses indicated that the atypical patterns of the earlier samples were no longer found. The children had by then acquired fricatives, velars and final consonants, and were only relying on gross inclusion. They has gone from showing patterns characteristic of phonological disorders to patterns more like those of younger typically developing children but delayed.

Summary

A child's acquisition of their language's phonological acquisition involves simultaneous acquisition of articulatory skills and phonological representations. During the process, there is an interaction between articulation limitations and the need to distinguish phonemes. A common way to cope with this problem is for the child to follow principles of markedness, that is, to express a phoneme not yet pronounceable by one that is less marked. This is a phonological phenomenon, but it can be seen as an articulatory limitation as well.

Children with speech sound disorders are children challenged with greater articulatory difficulties than typically developing children. In some cases, this appears to be a matter of delay. That is, the children appear to be following a typical course of acquisition, but they are older relative to their point of development. In other cases, their development is different looking, that is, atypical, and in these cases it is likely due to specifically more severe articulatory problems. Some of these are clearly identifiable as articulatory, such as in the case who lisps, or one who cannot make a /r/. Between these two types of development, there is a more complex situation, this being children who have a range of possible articulatory problems, and are compensating for them in phonological ways. In some instances we can see the interplay taking place between articulation and phonology in the form of substitution shifts, that is, instances where the child produces a sound, but as a substitution rather than for its phoneme equivalent.

We suggest that a way to better understand this more complex group showing what might be called a phonological disorder. The proposal is to distinguish the phonological patterns that appear to be more articulatory based (unmarked) from those that are phonological, and to see any particular child as following along a spectrum. Some may show primarily unmarked phonological processes that place them on the articulatory side. Others may make greater phonological compensations that can be identified as phonological. Children may also vary their place on this spectrum over time.

This proposal was demonstrated through the analysis of two different situations, one a case study of a child assessed as having a speech sound disorder, the other a group study of children with clefts. The consonants in the sample of the child in the case study divided into two clear groups, those that he could pronounce correctly and those that he could not. He appeared to be on the articulatory side of the spectrum, and showed predictable simplifications. It was also possible, however, to identify a clear instance where his substitutions were a substitution shift, driven by an effort to maintain maximal contrast. The group study on children with clefts also deals with an instance where articulation would be expected to be a major factor in assessment. These children at the same time would otherwise be expected to be typically developing, were they not having to deal with their clefts. The result of this circumstance was their showing temporarily very atypical phonological patterns in their early stage of phonological development. Later assessment showed noticeable improvements and the disappearance of the atypical patterns.

References

Babatsouli, E., Ingram, D., & Sotiropoulos, D. (2014). Phonological word proximity in child speech development. *Chaotic Modeling and Simulation*, *4*(3), 295-313.

Bowen, D. (2015). *Children's speech sound disorders* (2^{nd} ed.). Oxford, UK: Wiley-Blackwell.

Grunwell, P. (1985). *Phonological assessment of child speech (PACS)*. San Diego, CA: College-Hill Press.

Ingram, D. (1976). *Phonological disability in children*. London, UK: Edward Arnold.

Ingram, D., & Dubasik, V. (2011). Multidimensional assessment of phonological similarity within and between children. *Clinical Linguistics and Phonetics*, *25*(11-12), 962-967.

Ingram, D., & Scherer, N. J. (2015). Atypical patterns in the phonological systems of two-year-old children with clefts. Paper presented to the annual meeting of the American Speech-Language-Hearing Association, Denver, CO, November.

Jakobson, R. (1941/1968). *Child language, phonological universals and aphasia*. (Keiler, A. Trans.). The Hague: Mouton. Original (1941): *Kindersprache, aphasie und allgemeine Lautgesetze*.

Nice, M. (1925). A child who would not talk. *Pedagogical Seminary, 32*, 105-142.

Sommers, R. K. (1983). *Articulation disorders*. Engelwood Cliffs, NJ: Prentice-Hall.

Stoel-Gammon, C., & Dunn, C. (1985). *Normal and disordered phonology in children*. Baltimore, MD: University Park Press.

Voice onset time of the voiceless alveolar and velar stops in bilingual Hungarian-English children and their monolingual Hungarian peers

Ágnes Jordanidisz[1], Anita Auszmann[2], Judit Bóna[3]
ajordanidisz@gmail.com, auszmannanita@gmail.com, bona.judit@btk.elte.hu

[1]Association for Educational Needs/NILD, [2]Research Institute for Linguistics, Hungarian Academy of Sciences, [3]Eötvös Loránd University

Abstract. The VOT of Hungarian voiceless stops overlaps with that of the English voiced counterparts, which may present a challenge for Hungarian-English bilinguals. This study addresses this problem by investigating the VOTs of the two most common voiceless stops in Hungarian: /t/ and /k/, which are analyzed in the speech of bilingual Hungarian-English children and monolingual Hungarian children. The research question was whether bilingual Hungarian- and English-speaking children produce the voiceless dental and velar stop VOTs similarly to their monolingual Hungarian-speaking peers? We hypothesized that (1) VOT is longer in the speech of bilingual children when they speak Hungarian compared to their monolingual Hungarian-speaking peers' VOTs; (2) This difference is observed regardless of the type of speech in initial position in a picture naming test versus in spontaneous narratives; (3) The speech task highly influences VOTs. Ten bilingual Hungarian-English children (mean age: 6;6) and 10 monolingual Hungarian children (mean age: 6;6) participated in the study. A single-word picture naming task was used to elicit word-initial singleton stops in stressed position. Children were also asked to talk about school or free time, prompting narratives from which 10 /t/ and /k/ phonemes in CV position were selected for VOT analysis. PRAAT 5.0 was used to analyze the VOT values of word-initial, singleton, stressed /t/ and /k/ phonemes from both the picture elicitation task and the narrative. Results showed that bilingual children's VOTs are longer than those of their monolingual peers in the case of /k/, irrespective of the type of task; but there is no significant difference between the two groups in the case of /t/. There is also a statistically significant difference between the two types of speech tasks. Our findings indicate that bilingual Hungarian-English-speaking children produce their voiceless stops differently from their monolingual peers, and there is also a task effect.

Keywords: voiced onset time, voiceless stops, bilingual and monolingual children

Introduction

Voice onset time (VOT) is the duration between the burst and the onset of voicing of the next voiced segment (Lisker & Abramson, 1964; Lieberman & Blumstein, 1988). VOT values differ from one language to the other (Lisker & Abramson, 1967; Cho & Ladefoged, 1999). Languages of the world can be divided into several groups based on the VOTs of their stops (Lisker & Abramson, 1964). There are languages in which stops can be divided into three or four groups (Lisker & Abramson, 1964; Gandour & Dardarananda, 1984), while for example in Hungarian there are two categories of stops: voiced and voiceless ones (Lisker & Abramson, 1964; Gósy & Ringen, 2009). VOTs of voiceless stops can be aspirated or non-aspirated. In languages with aspirated stops like English (Torre & Barlow, 2009), VOTs are longer than in languages with non-aspirated stops like in Hungarian (Gósy & Ringen, 2009).

Voiced and voiceless stop consonant contrasts vary in terms of voice onset time (VOT) across different languages, which may pose challenges for bilinguals who learn languages that differ in this respect. Several studies examined VOT in bilingual children's speech production, and they found cross-language influence in bilingual phonological acquisition (e.g., Fabiano-Smith & Barlow, 2009; Fabiano-Smith & Goldstein, 2010; Fabiano-Smith & Bunta, 2012).

Hungarian-English bilinguals are faced with the problem of learning two languages that differ in how voiced and voiceless stops are distinguished based on VOT, because Hungarian has a voicing lead for voiced stops and a short lag for their voiceless counterparts, while English voiced stops tend to have a short lag while voiceless stops have a long lag (VOT). Consequently, the VOT of Hungarian voiceless stops overlaps with the English voiced counterparts, which may present a challenge for Hungarian-English bilinguals.

This study addresses the above-named problem by investigating the VOTs of the two most common voiceless stops in Hungarian: /t/ and /k/, which are analysed in the speech of bilingual Hungarian-English children and monolingual Hungarian children. In Hungarian, the acquisition of stops by monolingual children doesn't finish by the age of 13 years old, at which age /t/ VOTs differ from those in adults (Bóna & Auszmann, 2014). However, the VOTs of /k/ were similar to the values measured in adult speech at the age of 9 (Bóna & Auszmann, 2014). In this study, we investigate the VOTs of 6-7-year-olds whose language acquisition has not finished.

The research question was the following: Do bilingual Hungarian- and English-speaking children produce voiceless alveolar and velar stop VOTs in a similar fashion to their monolingual Hungarian-speaking peers? We hypothesize that (1) VOT will be longer in the speech of bilingual children when they speak Hungarian relative to their monolingual Hungarian-speaking peers' VOTs. (2) This difference will be observed regardless of the type of speech, whether the examined consonants are produced in initial position (in a picture naming test) or in spontaneous narratives. (3) However, the speech task will influence highly VOTs in both bilingual and monolingual speech. Our hypothesis concerning the influence of the speech task is based on the fact that the linguistic environment is more controlled in the case of a single-word picture naming test than in conversational speech, when children listen to the content more than on the form. Therefore, we hypothesize that VOT will be longer in the initial stressed positions of the picture naming task than in spontaneous speech.

Methods

Participants

Participants were chosen for the research following the Hungarian ethical regulations. 10 Hungarian-English bilingual children (mean age: 6;6) and 10 Hungarian-speaking monolingual children (mean age: 6;6) participated in this study. All participants have normal hearing and none of them show any cognitive, speech or language disorders based on school evaluation. All bilingual children live in Hungary and have studied at an American school in Budapest, Hungary for two years (K and grade 1). Their parents (or at least the mother) are Hungarian and their first language is Hungarian. All Hungarian-speaking monolingual children attend public elementary schools in the capital and none of them have learned foreign languages so far.

Material

All participants were tested individually. The recordings were made in a quiet room of the school using a Zoom H4 recorder. We investigated the speech of the children under two conditions. The first condition was a single-word picture naming test. It was only a part of a bigger picture naming test. The following words were used for this test:

words with /t/: *teknős* (turtle), *telefon* (telephone), *templom* (church), *toll* (feather)
words with /k/: *kabát* (coat), *kacsa* (duck), *kalap* (hat), *kalapács* (hammer), *kecske* (goat), *kés* (knife), *kéz* (hand), *könyv* (book), *kukac* (worm), *kút* (well), *kutya* (dog)

In this task singleton /t/ and /k/ phonemes in word-initial, stressed position were selected for VOT analysis.

The second condition was a spontaneous speech task. Children were asked to talk about school life or free time activities. In this case of conversational narratives, we investigated the plosives in CV positions.

Method

The segmentation of the VOT values of /t/ and /k/ phonemes and measurements were conducted using Praat 5.3 (Boersma & Weenink, 1998). The recordings were annotated by two of the authors while the third one controlled and checked annotations. In this study, voice onset time was defined as the time span between the beginning of the burst and the absolute onset of voicing as observed on the oscillogram and on the spectrogram in parallel (Beckman et al., 2011).

Statistical analyses (UNIANOVA, repeated-measures ANOVA, Mann-Whitney *U*-test, Wilcoxon-test) were carried out by SPSS 20.0. The dependent variables were (1) all data and (2) the average duration of /t/ and /k/ per child per task. Independent variables included language status and speech task.

Results

The picture naming task

First, we examined the average duration of /t/ and /k/ sounds per child, which meant 10 samples per group. Comparing the average VOT of these Hungarian voiceless stops produced by bilingual students to their monolingual peers' VOT, the statistical analysis did not show significant difference ($p>0,05$). However, data showed a tendency that bilingual children aspirate the examined stops to a greater extent than their monolingual peers (Figure 1).

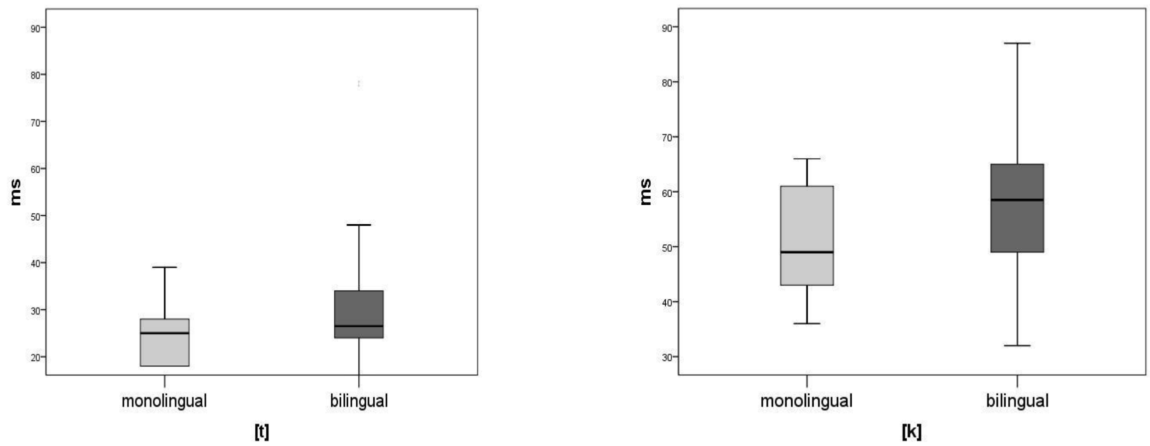


Figure 1. The comparison of average VOT of Hungarian /t/ and /k/ stops by monolingual and bilingual children

Secondly, we used all data in both groups: in the case of stop /t/ it meant 40, in the case of stop /k/ it meant 110 samples. This time, statistical difference was revealed between the bilingual and the monolingual groups in the duration of the voiceless stop /k/ (Z = -2,904; p = 0,004). However, in the case of sound /t/, statistical difference could not be detected. The tendency of the more aspirated pronunciation in the bilingual group is shown in Figure 2.

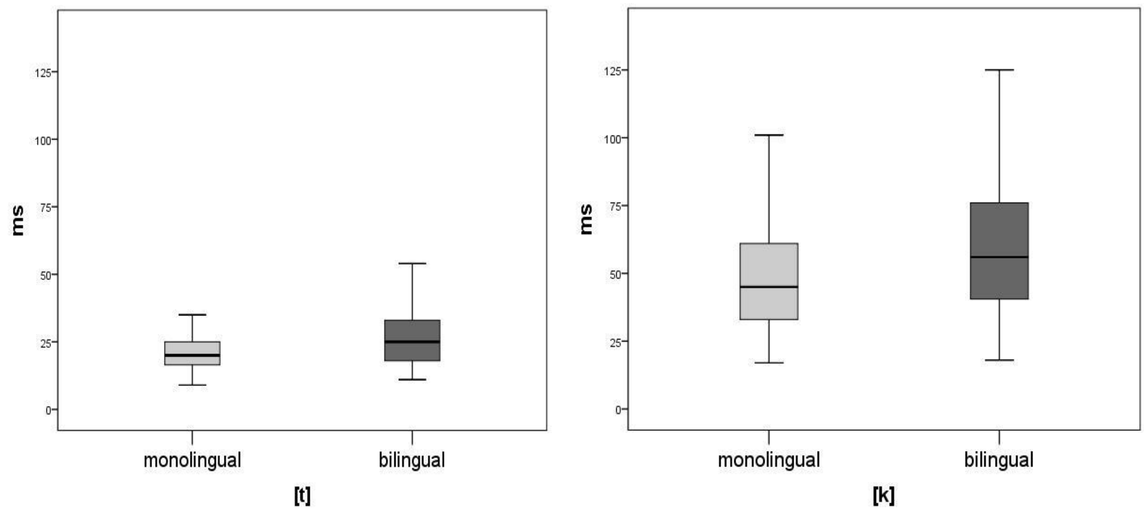


Figure 2. The comparison of all VOT data of Hungarian /t/ and /k/ stops by monolingual and bilingual children groups

Spontaneous speech

When we examined the average duration of /t/ and /k/ sounds per child in spontaneous speech gaining from 80 /t/ and 140 /k/ sound items, we did not find significant difference in the case of /t/ sound. On the other hand, the bilinguals' pronunciation of the Hungarian sound /k/ was significantly longer than that of their Hungarian monolingual peers: $F(1, 19) = 4{,}830$; $p = 0{,}041$; $\eta^2 = 0{,}212$ (figure 3).

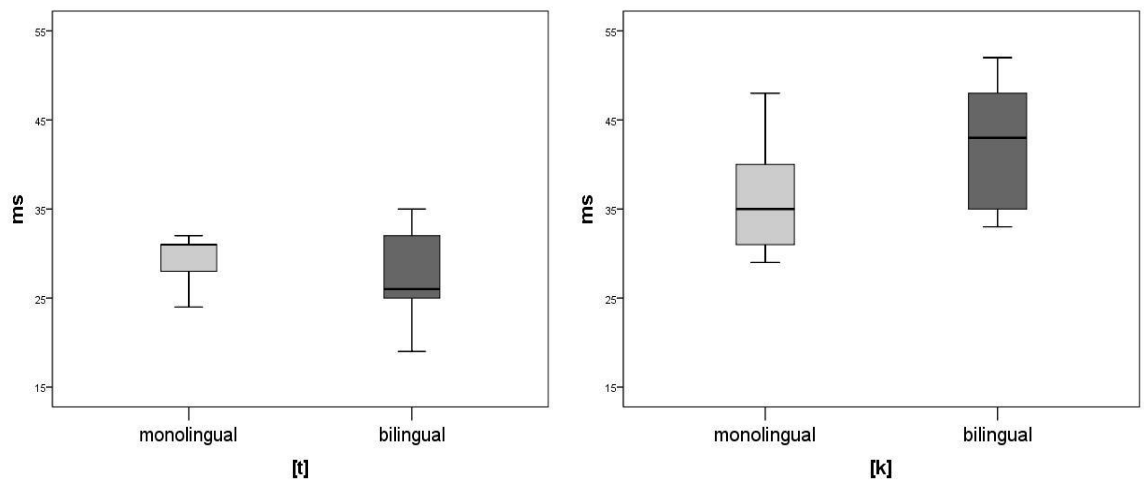


Figure 3. The comparison of average VOT of Hungarian /t/ and /k/ stops in spontaneous monolingual and bilingual speech

We further examined the two groups' spontaneous speech and we found that using all data in the statistical analysis confirmed the previous result (Figure 4): no significant difference was detected in the duration of /t/ sound between the bilingual and the monolingual pronunciation, while bilingual children pronounced the /k/ sound significantly longer, i.e. with aspiration, than their monolingual peers: $Z = -3{,}970$; $p < 0{,}001$.

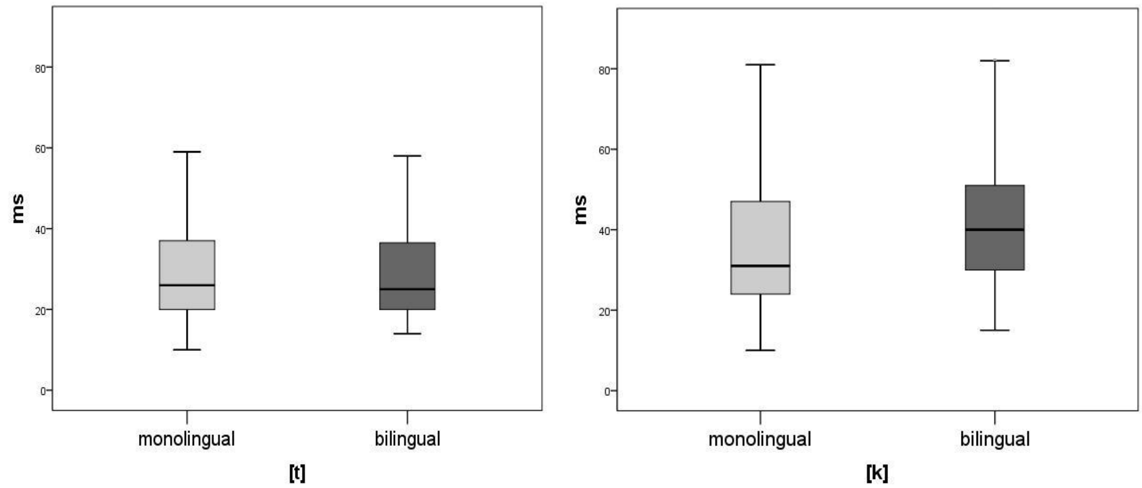


Figure 4. The comparison of all VOT data of Hungarian /t/ and /k/ stops in spontaneous monolingual and bilingual speech

The effect of the speech task

Comparing the length of the VOT of the examined voiceless stops in the two different speech tasks, namely in the picture naming test and the spontaneous speech, we found no significant difference in the case of /t/, but the VOT of the /k/ sound was significantly longer in children's utterances in the picture naming task than when they uttered this sound in spontaneous speech (Figures 5 and 6). It was true for both monolingual ($F(1, 19) = 17{,}895$; $p = 0{,}001$; $\eta^2 = 0{,}499$) and bilingual ($F(1, 19) = 6{,}418$; $p = 0{,}021$; $\eta^2 = 0{,}263$) groups. At the same time there were big individual differences.

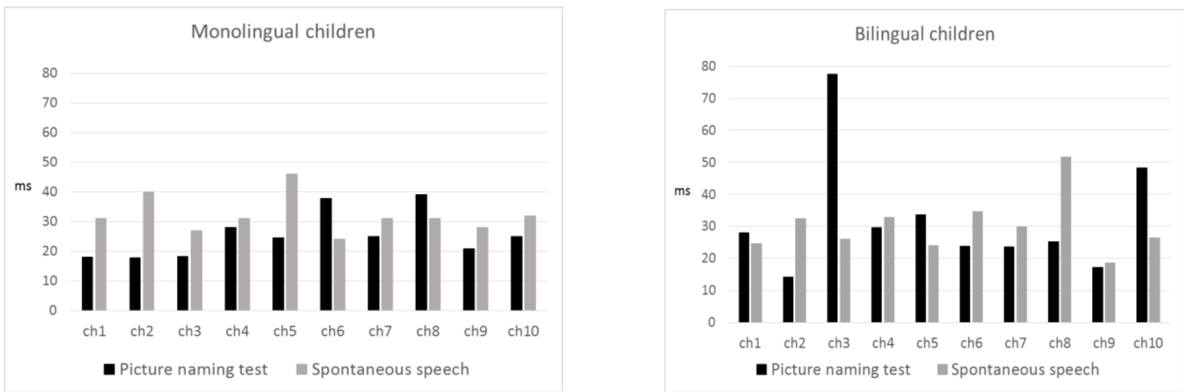


Figure 5. The effect of speech tasks: the average duration of /t/ per child in the monolingual and the bilingual groups

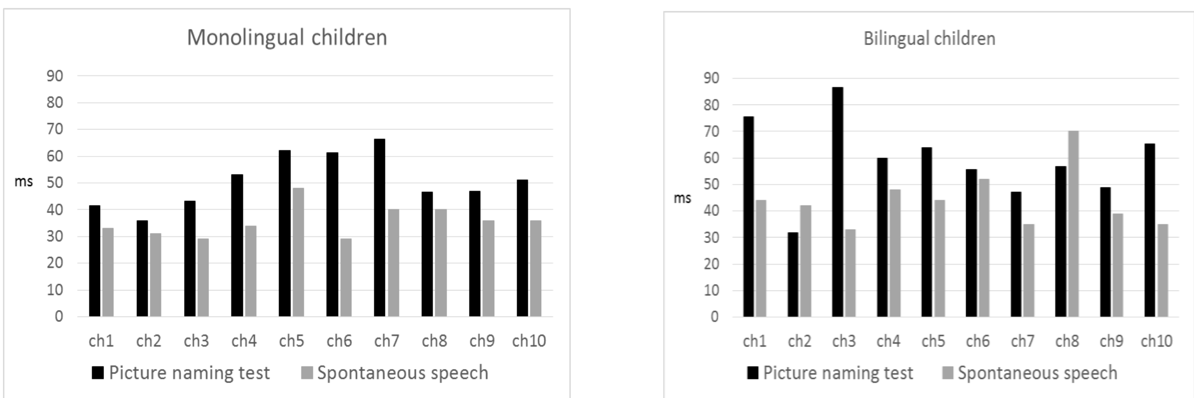


Figure 6. The effect of speech tasks: the average duration of /k/ per child in the monolingual and the bilingual groups

Interestingly, the alveolar stop showed the opposite tendency, namely children articulated the /t/ sound longer in the spontaneous speech, though the difference was not significant.

Discussion

Our hypotheses concerning the length of VOTs were confirmed partially. The VOTs of /k/ were significantly longer in the speech of bilingual children when they spoke Hungarian comparing them to their monolingual Hungarian-speaking peers' VOTs. However, the difference was not significant in the case of /t/ sound. We assume that our results are the effect of the following possible factors in our examination, namely:

a. It could be the result of the fewer /t/ VOT values in the sample. Almost three times more /k/ than /t/ sound could be gained from the picture naming test, and the ratio of /t/ and /k/ sounds was 4/7 in the spontaneous speech.
b. Children acquire the /t/ sound earlier, therefore they have used it in the Hungarian pronunciation for a longer period than the /k/ sound before their encounter with the native American articulation of these sounds.

Further, we expected that the difference between the bilingual and the monolingual pronunciation would be observed both in the case of the picture naming test and in the samples of spontaneous narratives. Our hypothesis was confirmed in the case of the voiceless stop /k/ when *all data* were considered. We did not find significant difference during the statistical analysis of the *average duration of the VOT per child* situation in the picture naming test which might be due to the fewer samples. In the case of the dental stop, significant difference in the VOT of the bilingual and the monolingual pronunciation could not be detected in either of the speech tasks. Our argument concerning this result was explained in the previous paragraph.

Our third hypothesis, that the speech task influences VOT was confirmed in the case of the velar stop /k/. Both bilingual and monolingual children articulated the VOT of these Hungarian voiceless stops longer in the case of clear initial position than in spontaneous speech, where the syllable which contained the examined sound was not stressed. We argue that the same result could not be detected in the case of the voiceless stop /t/ because its phonological features in the given picture naming test were less diverse than those of the /k/ sound. However, in the case of the /k/ sound, with an optimal variety of the phonological features, the results clearly showed that speech task highly influences VOT.

Conclusion

The present research provided first data concerning the effect of bilingualism on the VOT duration in the case of /k/ and /t/ voiceless stops in Hungarian. It clearly demonstrated that bilingualism exercises influence on the VOT in the case of the Hungarian pronunciation. Monolingual children at the age of 6 articulate the /k/ sound with less aspiration than their bilingual peers whose first language is Hungarian. In future research, there should be more control over the sample size to gain more reliable information about how bilingualism might affect the VOT of the examined sounds. Further research may focus on the analysis of the VOT of these stops in English, the bilingual children's other language and compare the results to their Hungarian pronunciation. Overall, the present research provides novel information for bilingual child language and serves as a foundation for further research.

References

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, *39*, 39-49.

Boersma, P., & Weenink, D. (1998). Praat: doing phonetics by computer (Version 5.0.1). http://www.fon.hum.uva.nl/praat/download_win.html.

Bóna, J., & Auszmann, A. (2014). Voice onset time in language acquisition: Data from Hungarian. In S. Fuchs, M. Grice, A. Hermes, L. Lancia, & D. Mücke (eds.): *Proceedings of the 10th International Seminar on Speech Production (ISSP)* (pp. 41-44). Köln, 5-8. May 2014. http://www.issp2014.uni-koeln.de/wp-content/uploads/2014/Proceedings_ISSP_revised.pdf

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*, 207-229.

Fabiano-Smith, L., & Barlow, J. A. (2009). Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism*, *1*, 1-17.

Fabiano-Smith, L., & Goldstein B. (2010) Phonological acquisition in bilingual Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research*, *53*, 1-19.

Fabiano-Smith, L., & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics and Phonetics*, *26*(2), 148-163.

Gandour, J., & Dardarananda, R. (1984). Voice onset time in aphasia: Thai II. Production. *Brain and Language*, *23*, 177-205.

Gósy, M., & Ringen, C. O. (2009). *Everything you always wanted to know about VOT* (in Hungarian). Paper presented at the Ninth International Conference on the Structure of Hungarian, Budapest, 1 September 2009. http://icsh9.unideb.hu/pph/handout/Ringen_Gosy_handout.pdf (downloaded 30 August 2012).

Lieberman, P., & Blumstein, S. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge, UK: Cambridge University Press.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*, 1-28.

Torre III, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, *42*, 324-333.

Structural language deficits in a child with DiGeorge syndrome: Evidence from Greek

Maria Kambanaros[1,3], Loukia Taxitari[1,3], Eleni Theodorou[3], Kleanthes K. Grohmann[2,3]
maria.kambanaros@cut.ac.cy, loukia.taxitari@cut.ac.cy, elenath@cytanet.com.cy, kleanthi@ucy.ac.cy

[1]Cyprus University of Technology, [2]University of Cyprus, [3]Cyprus Acquisition Team

Abstract. This study presents an investigation of language skills in a male child with DiGeorge syndrome (DGS), an autosomal dominant genetic disorder caused by a microdeletion on the long arm of chromosome 22. The syndrome is associated with an extensive and variable phenotype which includes mild differences in facial features, congenital heart disease, defects in the palate, recurrent ear infections, and learning problems as well as behavioural and social interaction difficulties. Delayed language onset and persistent language impairment in preschool ages have been described in the literature, though not much is known about language skills of DGS children across languages. The purpose of this study is to describe the language profile of a single child with DGS for Greek using case-based methodology, and to compare his performance on structured language tasks with those reported for preschool children with and without specific language impairment from our database on the same tools. The participant was born in Cyprus to Greek Cypriot parents. He was close to 6 years of age when testing began and enrolled in a normal mainstream preschool at the time of the study. He was receiving speech therapy on a weekly basis focused on his articulation and voice problems. He had fluent, overall intelligible speech and was social. There was no evidence of attention deficit/hyperactivity disorder on one-to-one testing. Non-verbal intelligence was measured with Raven's Coloured Progressive Matrices, while global language scores for receptive and expressive language abilities were derived from the Diagnostic Verbal IQ Test prior to the testing on structured language tasks. Receptive vocabulary knowledge was assessed on the Peabody Picture Vocabulary Test, expressive vocabulary on the Expressive Vocabulary Test. Structural language testing involved clitic production and a narrative retell task. The findings will be discussed in relation to two hypotheses, namely either that the profile of language impairment in children with DGS may be distinctive to the syndrome or that there is the possibility of co-morbidity of specific language impairment in DGS.

Keywords: clitics, co-morbidity, narrative, specific language impairment, testing, vocabulary

Introduction

Cross-linguistic research describing the language and cognitive abilities of children with rare syndromes is sparse. The aim of the present case study is to report on the language abilities of a school-aged boy genetically confirmed with DiGeorge syndrome (DGS). This syndrome follows an autosomal dominant inheritance pattern (a child only needs to get the abnormal gene from one parent in order to inherit the disease). However, only around 10% of cases are inherited; the majority of DGS occurrences are due to a random mutation (Shprintzen, 2008). Our participant falls into the latter category.

DGS results from a submicroscopic hemizygous deletion at chromosome 22q11.2 (Woodin, Wang, Aleman, McDonald-McGinn, Zackai, & Moss, 2001). The syndrome is also known as velocardiofacial syndrome (De Decker & Lawrenson, 2001). It is an increasingly common genetic disorder affecting at least 1 in 2,000-7,000 live births (Shprintzen, 2008). The phenotypic description of this microdeletion syndrome is quite varied, with close to 200 clinical features identified so far as related to abnormalities of the heart, palate, velopharyngeal mechanism, immune system, central nervous system, and brain morphology (see Woodin et al., 2001 and references within). However, each child is affected differently and the symptoms can vary widely, ranging from less severe to severely affected. Children with DGS tend to have similar facial features, including a long, narrow face; wide-set almond-shaped eyes; a broad nasal bridge and bulbous nose tip; a small mouth; small, low set ears that are folded over at the top; an irregular skull shape (www.nhs.uk/conditions/digeorge-syndrome). Furthermore, there is evidence that DGS remains undiagnosed in many children as an isolated speech and language disorder or developmental delays in the presence of few or no physical abnormalities (Niklasson, Rasmussen, Óskarsdóttir, & Gillberg, 2001).

There is a large literature on the behavioural and psychiatric profiles of individuals with 22q11 deletion syndrome (see Scandurra, Scordo, Canitano, & de Bruin, 2013 and references within); however, within the paediatric population information is limited. The large variation of the phenotype can make diagnosis more difficult. According to some researchers, the median age of diagnosis is 6.5 years (Solot, Handler, Gerdes, McDonald-McGinn, Moss, Wang, Cohen, Randall, Larossa, & Driscoll, 2000).

The majority of DGS individuals show relatively mild cognitive deficits, including sometimes mild mental retardation (MMR, i.e. IQ 51-0), with verbal IQ often significantly higher than performance IQ and/or non-verbal IQ. However, there are reports of individuals with low normal intelligence (IQ 71-85) and some with an IQ in excess of 85 (Niklasson et al., 2001). Individuals with DGS show relative strengths in verbal ability, rote processing, verbal memory, reading, and spelling. In addition, there are reported weaknesses in language abilities, attention, working memory, executive functions, visuospatial memory, and psychosocial functioning (see Woodin et al., 2001 and references within for both points).

In particular, research on the manifestations of speech and language disorders in DGS children is not prominent, despite communication impairment being one of the hallmark deficits of the syndrome. In DGS, speech and language delays have been reported in early childhood with persistent language impairment in preschool ages in the areas of word finding, vocabulary, syntax, and discourse (see Persson, Niklasson, Óskarsdóttir, Johansson, Jönsson, & Söderpalm, 2006 and references within). Preschool children with DGS often have a reduced vocabulary size, reduced sentence length, and delayed use of grammatical structures. Also, expressive language delays are more severe than receptive language delays (Persson et al., 2006), but this may not always be the case. Moreover, specific language impairment (SLI) has been reported for several individual children in large DGS cohorts (Solot et al., 2000) or smaller case studies (Goorhuis-Brouwer, Dikkers, Robinson, & Kerstjens-Frederikse, 2003). SLI is a term applied for children whose speech and language is substantially below age level for no apparent reason, that is, in the absence of neurological damage, impaired sensorimotor abilities, and so on (i.e. with normal intelligence levels, hearing, vision, etc.).

To our knowledge, our report is the first to describe in detail the linguistic manifestations of the language deficit associated with this particular genetic syndrome for Greek, and in the context of bilectalism (Rowe & Grohmann, 2013). Bilectalism is used here to characterise the situation in Cyprus in which children of Greek Cypriot parents, with Cypriot Greek-speaking family and friends, grow up, yet get exposed to Standard Modern Greek from an early age; first through media such as TV cartoons, later through public schooling starting in nursery and kindergarten, becoming gradually more systematic in primary school. In the absence of a separate Cypriot Greek orthographic system, Greek can only be taught through the medium of the standard variety in order to teach children how to read and write. We take this to be the standard path of language development by Greek Cypriot children, as relevant for our study.

Method

Aims

A core area of investigation will be the DGS participant's abilities in structural language, that is, his morphosyntactic abilities and performance in more complex language. Our testing battery contains several measures for structural language, ranging from morphosyntactic properties (agreement relations) and phenomena (object clitic production) to structurally complex clauses (including subordinates and relatives). But the entire testing battery goes well beyond structural language. As the first research on language abilities in DGS for (Cypriot) Greek, we take a broader angle.

The purpose of the present study is then to profile the language abilities of one male child (PI) with DGS and compare his performance to that reported for children of the same chronological age with typical language development and with SLI across a battery of linguistic tests. The aims of the study are three-fold:

1. to investigate PI's receptive and expressive language abilities, with an emphasis on structural language performance;
2. to compare PI's language performance across all measures with that of typically developing children of the same chronological age;
3. to compare PI's language performance across all measures with that of children of the same chronological age diagnosed with SLI.

Participant

Our participant, PI, was a preschool boy who was 5;11 (years;months) when the study began and enrolled in the preschool education program of a public school in Nicosia, Cyprus. He was not receiving special education services.

He was diagnosed with DiGeorge syndrome using the fluorescence in situ hybridisation test (FISH) by the Genetics Clinic of the Makarios Hospital in Nicosia. This test shows whether the region of chromosome 22 is present. If only one copy of chromosome 22 'lights up' with fluorescent DNA dye, rather than both copies, the test is positive for 22q11 deletion. Hearing was tested by the Audiology Clinic of the Makarios Hospital and reported to be within normal limits. Also, the hospital reported no positive assessment of autism spectrum disorder symptoms or any other psychiatric condition.

He was born from healthy, unrelated parents who are both highly educated with university degrees, in an allied health profession (mother) and information technology (father). PI has a healthy brother who is older by three years.

An oral-peripheral motor examination administered by a certified speech and language therapist (first author) revealed no structural abnormalities of the speech mechanism. At the time of the study, PI was receiving private speech therapy for voice quality (e.g., hypernasality) and mild misarticulations. He presented hoarseness and reduced vocal volume but generally intelligible speech during the time of the study. Testing across all measures was conducted over a three-month period.

Comparative groups

For our comparison with PI, a total of 19 bilectal Greek Cypriot preschool children participated in this study. For selection purposes, we considered as bilectal those children whose parents are both Greek Cypriots, who were born and raised in Cyprus, and who did not spend any large amount of time outside the island, including Greece. We did not control any more, specifically for balanced input or age of exposure to Cypriot Greek and Standard Modern Greek, but assumed the standard path of language development laid out above (see Grohmann & Kambanaros, to appear for more). The children were divided into two groups, one group including children with SLI of the same chronological age as PI and a control group, namely a chronological age-matched group.

All children came from the Limassol district and the majority attended public pre-primary or primary schools. Parental consent forms and an information letter that explained the purpose of the study were distributed, and only children whose parents gave written consent participated in the study. The consent form provided additional information such as demographics, the education level of each of the parents, and the parents' occupation (see Theodorou, 2013 for participant details). The criteria that restricted individual participation included: (i) a known history of neurological, emotional, developmental, or behavioural problems; (ii) hearing and vision not adequate for test purposes after school screening at the beginning of the school year; (iii) non-verbal performance not in the broad range of normal; (iv) gross motor difficulties; and (v) low socio-economic status. All the above information was obtained either from the speech-language therapists and teachers or from the children's parents. The children were divided into two comparative groups, a group with typical language development (TLD) and a group of children with SLI.

Children with SLI

Nine children with SLI (7 boys and 2 girls), aged between 4;11 and 5;11 with a mean age of 5;6 (SD 0;3), served as the language-impaired comparative group. Children were diagnosed with SLI by certified speech and language therapists based on case history information, informal testing (of comprehension and production abilities), analysis of spontaneous language samples, and clinical observation. Children with SLI included in the study were receiving speech and language therapy services by practitioners in private settings.

TLD group

Ten children with TLD (6 boys and 4 girls), aged between 4;5 and 6;6 with a mean age of 5;8 (SD 0;6), served as the chronological age-matched group. According to the classroom teacher and parent report, each participant in the control groups was typically developing in all respects. No child was previously referred to or had received treatment by a speech and language therapist.

Socio-economic status

All children came from families with a medium to high socio-economic status, as measured by mothers' education level using the European Social Survey (2010) database. We compared PI's mother's education level (undergraduate degree from university) to the education levels of the mothers of the other two groups. Her education level did not differ from the TLD group mothers' ($t_{(9)} = -.52$, $p = .62$), but it did so from the SLI group ($t_{(8)} = 2.47$, $p < .05$), whose mother's education levels had a lower mean than PI's mother's.

Nonverbal IQ

Prior to the study proper, all children were tested on the Raven's Coloured Progressive Matrices (Raven, Raven, & Court, 2000), following the Greek norms of Sideridis, Antoniou, Mouzaki, & Simos (2015). This requirement was satisfied for each child separately.

Materials and procedures

All language measures were administered to PI and the two groups of children, those with a clinical identification of language impairment and the typically developing ones. The tests are described in detail below.

A. Diagnostic Verbal IQ Test (DVIQ)

Children's global language abilities were measured using the Diagnostic Verbal IQ Test (Stavrakaki & Tsimpli, 2000), modified for Cypriot Greek (Theodorou, 2013). This test is used by language researchers and clinicians to assess language abilities for Greek-speaking children; while it is not yet standardised, though, it is in the process of undergoing standardisation in Greece. The DVIQ has five subtests: expressive vocabulary, comprehension of morphosyntax, production of morphosyntax, comprehension of metalinguistic concepts, and sentence repetition. The production of morphosyntax subtest includes such diverse grammatical properties and markers as nominal and verbal suffixes, object clitics, articles, agreement relations, and relative clauses among others.

Each child was tested individually on all subtests, which involved naming and showing pictures as well as completing and repeating sentences. Children's responses were recorded on the answer sheets, and later analysed and scored. Each correct response received 1 point, with the exception of the sentence repetition subtest, which was scored according to the number of errors in each repetition (maximum score of 3 points correct for each sentence). The original DVIQ has been used in published studies for the identification of children with SLI in Greece (Mastropavlou, Petinou, & Tsimpli, 2011) and Cyprus (Petinou & Okalidou, 2006; Mastropavlou et al., 2011); the Cypriot Greek-adapted version has also been tested widely in Cyprus by our research team for which published studies include Kambanaros, Grohmann, Michaelides, & Theodorou (2013; 2014), among others.

B. The Bus Story Test (BST)

A topic-centred narrative, the Bus Story Test (Renfrew, 1997) is used widely by speech and language therapists to assess narrative abilities in children ranging from 3 to 8 years of age. It is translated into

Greek and also used in Greece as a non-standardised measure. For this study, the Greek translation of the BST was used (with minor changes in phonology and morphology adapted to Cypriot Greek), since Greek Cypriot children are used to hearing stories in Standard Modern Greek rather than in Cypriot Greek from their preschool years.

The experimenter told each child the story individually while the child looked at the picture strips illustrating the story. Afterwards, the child was requested to retell the story as close to the original as possible. Test administration was around 10 minutes. The narrative samples were transcribed and divided into sentences (t-units). Children's narrative productions were each evaluated with respect to five descriptors, three from the BST manual and two additional ones developed for our research purposes:

1. *Information* (Renfrew, 1997): The number of relevant information pieces were tallied following the BST manual, where 'essential' information gets two points and 'subsidiary' information gets one point; the Information score is the total number of points accumulated.
2. *Subordinates Clauses* (Renfrew, 1997): All subordinate clauses were identified and counted for a total score, as per BST manual.
3. *A5LS*: The mean length of the five longest sentences was computed.
4. *MLU-word*: In the absence of normative data for mean length of utterance (MLU) in Cypriot Greek, it was calculated based on words for each narrative (MLU-word); all words were added up and the sum was divided by sentences produced (MLU-word was chosen, since there is no study to support the use of a morpheme-based MLU in any variety of Modern Greek).
5. *T-units* (Renfrew, 1997): The total number of sentences produced was added up, as suggested in the BST manual.

C. Expressive Vocabulary test (EVT)

In order to assess naming abilities, the Expressive Vocabulary Test (Vogindroukas, Protopapas, & Sideris, 2009) was administered. The EVT contains 50 concrete black-and-white pictures for naming. It is standardised in Greece and has norms for Standard Modern Greek. Because of the differences between the standard language and the dialect, 11 items have alternative words in Cypriot Greek (10 lexical alternatives and 1 phonological alternative); they were considered acceptable responses. Children were asked to name the object in the picture. Responses were recorded on the answer sheet and then scored as correct or incorrect on a word-by-word basis.

D. Peabody Picture Vocabulary Test (PPVT)

In order to assess receptive vocabulary skills, the Greek version of the Peabody Picture Vocabulary Test (Dunn & Dunn, 1981, adapted to Standard Modern Greek by Simos, Kasselimis, & Mouzaki, 2011), developed for research purposes, was used. The PPVT measures receptive vocabulary at the single word level.

The Greek version of the test consists of 228 items, equally distributed across 19 item-sets. Each set contains 12 items of increasing difficulty. The examiner presented a quadrant of four numbered pictures and asked the child to point to or say the number of the picture of the spoken word.

E. Clitics-in-Island Test (CIT)

The COST Action A33 Clitics-in-Islands Test (Varlokosta, Belletti, Costa, Friedmann, Gavarró, Grohmann, Guasti, Tuller, et al., 2015), a testing tool designed to elicit clitic production, was used. The CIT is a production task for 3rd person singular accusative object clitics in which the target-elicited clitic was embedded within a *because*-clause (a so-called syntactic island):

> I mama xtenizi ti korua tʃe i korua en omorfi. Jati i korua en omorfi? I korua en omorfi jati i mama tis… [xtenizi tin-CL].
>
> "Mommy is combing the girl and the girl is beautiful. Why is the girl beautiful? The girl is beautiful because her mommy… [combs her-CL]."

The CIT involved a total of 19 items; 12 target structures (i.e. test items) after two warm-ups, plus five unrelated fillers. All target structures were indicative declarative clauses formed around a transitive verb. All participating children were shown a coloured sketch picture on a laptop screen, depicting the situation that was described by the experimenter. Figure 1 illustrates for the above example.

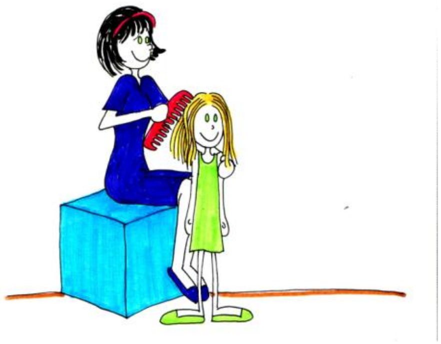


Figure 1. Test item #2 (example)

Participants heard the description of each picture that the experimenter provided and then had to complete the *because*-clause in which the use of a clitic was expected. The ideal response would be a verb-clitic sequence (such as *xtenizi tin* 'combs her' in the above example), but some participants started with *because* on their own, others filled in right after the experimenter's prompt of *because*. The experiments were not audio- or video-taped, but answers were recorded by the experimenter on a score sheet during the session.

Structural language probes

We consider structural language probes to be those that tap into morphosyntactic abilities and language complexity. For our purposes, the *comprehension and production of morphosyntax subtests* of the DVIQ, the *sentence repetition subtest* of the DVIQ, *number of subordinate clauses produced* on the BST, and *performance on the CIT* will serve as measures of structural language complexity for our analyses in this research.

Scoring and analysis

For all tests, an *accuracy score* was calculated by summing up the number of correct responses. For all sub-categories of the DVIQ, except sentence repetition, a single point was given for every correct response and no points for every incorrect one. For sentence repetition, 3 points were given for every correct response, 2 points for every response with one error, 1 point for every response with 2 errors, and no points for responses with 3 or more errors.

Results

The main statistical analysis used was the Crawford–Howell *t*-test (Crawford & Howell, 1998), a method developed in neuropsychology for the comparison of single cases with control groups (with small sample numbers). Using this method, PI's accuracy scores were compared to the TLD and SLI groups using a one-tailed *t*-test. The results are reported in Table 1 (where Table 1a includes correct scores as percentage points and Table 1b as standardised raw scores).

Table 1a. Percentage correct (DVIQ, PPVT, EVT, CIT) for PI (DGS) in comparison to the TLD group and the SLI group

Test	Sub-test	Score		
		DGS	TLD (SD)	SLI (SD)
DVIQ	Vocabulary	70.5%	84.8% (8.1)	62.1% (10.4)
	Production: Morphosyntax	33.3%	73.3% (7.8)	51.4% (10.1)
	Comprehension: Metalinguistic concepts	68.0%	79.6% (7.2)	72.0% (2.5)
	Comprehension: Morphosyntax	51.6%	81.9% (8.4)	79.2% (2.6)
	Sentence repetition	72.9%	94.8% (5.2)	85.2% (1.7)
	Total DVIQ Score	*61.9%*	*82.1% (4.9)*	*73.6% (1.6)*
PPVT		16%	30.1% (5.5)	25.8% (7.8)
EVT		50%	66.6% (10.3)	43.4% (5.5)
CIT		75%	96.7% (10.5)	91.7% (8.3)

Key: DVIQ=Diagnostic Verbal IQ Test, PPVT=Peabody Picture Vocabulary Test, EVT=Expressive Vocabulary Test, CIT=Clitics-in-Islands Test, DGS=DiGeorge syndrome, TLD=typical language development, SLI=specific language impairment, SD=standard deviation

Table 1b: Raw scores (BST) and standard scores (RCPM) for PI (DGS) in comparison to the TLD group and the SLI group

Test	Sub-test	Score		
		DGS	TLD (SD)	SLI (SD)
BST	Information	17	35.8 (11.5)	21.8 (8.9)
	A5SL	5.4	8.44 (2.1)	5.4 (.8)
	Subordinate clauses	1	7.8 (4.1)	1.7 (1.5)
	T-units	8	20.6 (3.9)	15.6 (3.8)
	MLU–word	4.6	4.7 (1.2)	3.4 (.7)
RCPM		69	90 (12.5)	100.6 (12.9)

Key: BST=Bus Story Test, RCPM=Raven's Coloured Progressive Matrices, DGS=DiGeorge syndrome, TLD=typical language development, SLI=specific language impairment, SD=standard deviation

Nonverbal IQ

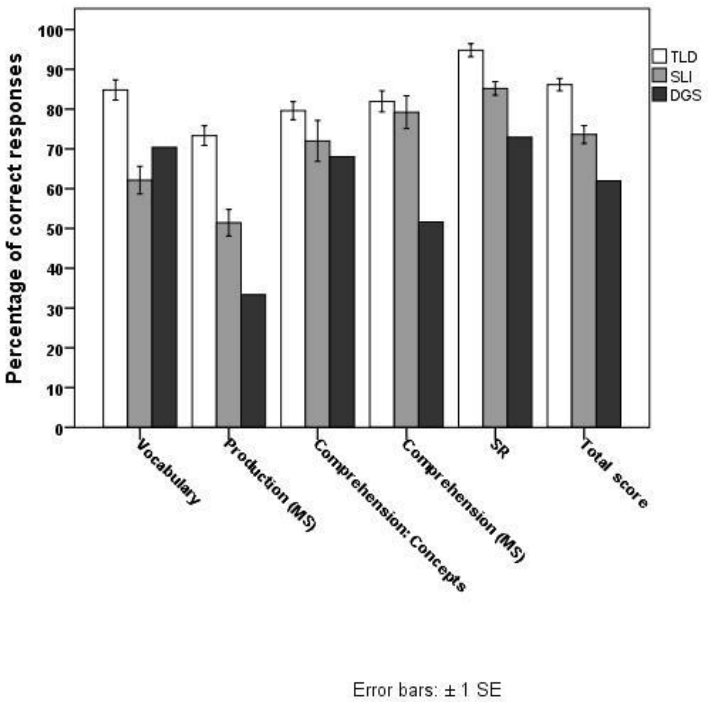
There were no statistically significant differences in nonverbal IQ between the SLI and TLD groups (Mann–Whitney U = 25.5, $n_1 = 10$, $n_2 = 9$, $p = .11$ two-tailed). When PI's was compared to the mean performance of each group, there was no significant difference from the TLD group ($t_{(9)} = –1.61$, $p = .07$), but his nonverbal IQ was significantly lower than that of the SLI group ($t_{(8)} = –2.29$, $p < .05$).

DVIQ

PI's comparison to the TLD group revealed that he performed significantly lower on three subtests of the DVIQ: (i) in the *production of morphosyntax* ($t_{(9)} = –4.9$, $p < .001$); (ii) in the *comprehension of morphosyntax* ($t_{(9)} = –3.46$, $p < .01$), and (iii) in the *sentence repetition* subtest ($t_{(9)} = –3.99$, $p < .01$). PI showed no significant difference from the TLD peers on the remaining two subtests of the DVIQ, namely comprehension of metalinguistic concepts ($t_{(9)} = –1.55$, $p = .08$) and expressive vocabulary

($t_{(9)} = -1.71$, $p = .06$). Overall, his total DVIQ score was significantly lower than that of his TLD peers ($t_{(9)} = -4.69$, $p < .01$).

When PI's performance was compared to that of the SLI group, he showed a significantly lower performance on two DVIQ subtests: (i) *comprehension of morphosyntax* ($t_{(8)} = -2.12$, $p < .05$) and (ii) *sentence repetition* ($t_{(8)} = -2.26$, $p < .05$). There was no significant difference between PI and the SLI group on the remaining DVIQ subtests, expressive vocabulary ($t_{(8)} = .75$, $p = .24$), comprehension of metalinguistic concepts ($t_{(8)} = -0.25$, $p = .41$), and production of morphosyntax ($t_{(8)} = -1.71$, $p = .06$). Overall, his total DVIQ score was not significantly different from the SLI group ($t_{(8)} = -1.64$, $p = .07$). Figure 2 shows PI's performance in comparison to the TLD and SLI groups on the subtests and total score on the DVIQ test.



Key: MS=morphosyntax, SR=sentence repetition,
TLD=typical language development, SLI=specific language impairment, DGS=DiGeorge syndrome

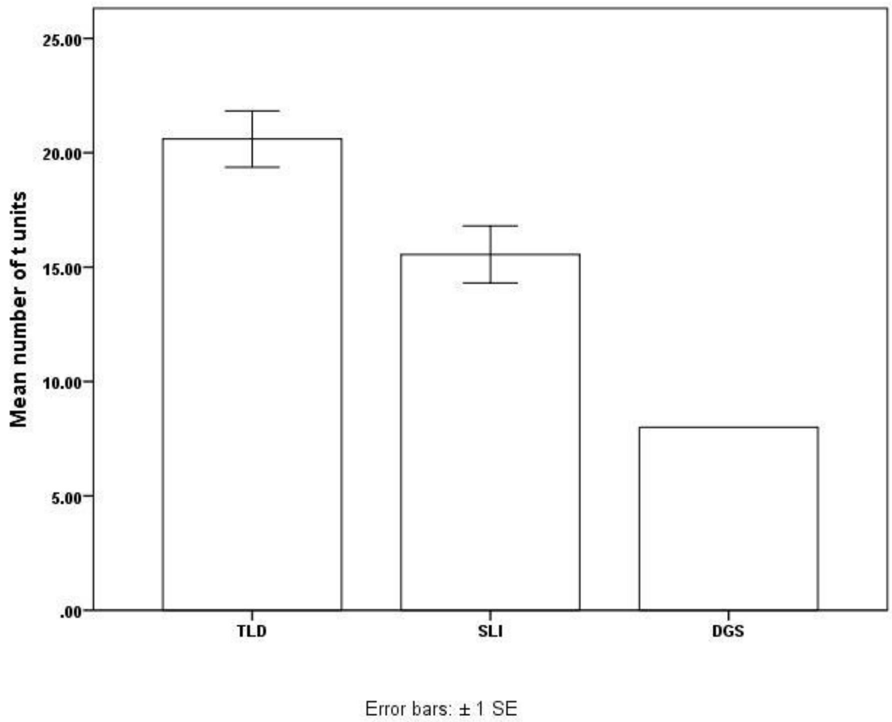
Figure 2. Performance of PI (DGS) in comparison to the TLD group and the SLI group on the five different subtests of the DVIQ

BST

When PI's performance was compared to the TLD group on the BST, he showed a significantly lower performance only on the *number of sentences (t-units) produced* ($t_{(9)} = -3.09$, $p < .01$). There was no significant difference between PI and the TLD group for Information ($t_{(9)} = -1.55$, $p = .08$), A5LS ($t_{(9)} = -1.37$, $p = .1$), number of subordinate clauses produced ($t_{(9)} = -1.58$, $p = .07$), and MLU-word ($t_{(9)} = -.06$, $p = .48$).

When compared to the SLI group, PI also only showed a significantly lower performance for *number of sentences (t-units) produced* ($t_{(8)} = -1.91$, $p < .05$). There were no significant differences between PI and the SLI group performance for Information ($t_{(8)} = -.51$, $p = .31$), A5LS ($t_{(8)} = 0$, $p = .5$), number of subordinate clauses produced ($t_{(8)} = -.42$, $p = .34$), and MLU–word ($t_{(8)} = 1.64$, $p = .07$).

Figure 3 shows PI's performance compared to the TLD and SLI groups on the number of sentences (t-units) produced.



Key: TLD=typical language development, SLI=specific language impairment, DGS=DiGeorge syndrome

Figure 3. Performance of PI (DGS) in comparison to the TLD group and the SLI group on the production of sentences (t-units) on the BST

PPVT

For receptive vocabulary at the single word level, PI performed significantly lower compared to the TLD group ($t_{(9)} = -2.42$, $p < .05$), but not when compared to the SLI group ($t_{(8)} = -1.19$, $p = .13$).

EVT

For expressive vocabulary, PI showed a similar performance on this task to both groups of children, those with TLD ($t_{(9)} = -1.54$, $p = .08$) and those with SLI ($t_{(8)} = 1.15$, $p = .14$).

CIT

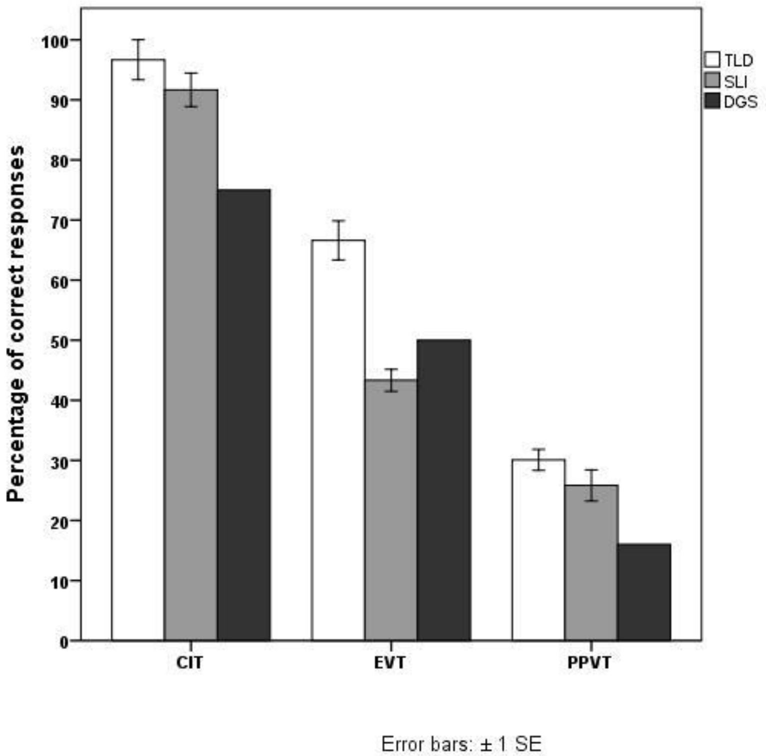
PI showed a significantly lower performance on this task than both groups of children, the children with TLD ($t_{(9)} = -1.95$, $p < .05$) and those with SLI ($t_{(8)} = -1.9$, $p < .05$).

Results from the PPVT, the EVT, and the CIT across PI and the two comparative groups of children are graphically displayed in Figure 4.

Discussion

The purpose of the present study was to profile, for the first time, the language abilities of a preschool Greek-speaking child with DGS and compare him to two groups of chronological age-matched children across a number of linguistic tools used for research purposes in Cyprus: one group of children with typical language development (TLD) and another with clinically diagnosed specific language impairment (SLI). Overall, the findings are relevant to clinical practice by demonstrating the value of language profiling in assisting in characterising the pattern of language impairment in a given child with DGS, with the ultimate aim of developing appropriate treatment plans. Also, by comparing our participant to a group of children with SLI, that is, with known profiles of speech and language difficulties, allows us to decipher whether the profile of DGS is distinctive to the syndrome or not.

With regards to our first aim, we will describe the findings based on the pertinent literature outlined in the introduction. Overall, global language ability, as probed by the DVIQ, is impaired. This finding supports what is reported so far (Persson et al., 2006), namely that language impairment is evident at the preschool stage in DGS. Similarly, our DGS participant's receptive and expressive language abilities are differentiated based on the single-word tools used: Expressive vocabulary (measured on the EVT) appeared intact for his chronological age, but *receptive vocabulary* (measured on the PPVT) lagged behind. This stark difference between receptive and expressive language abilities has not been described in the literature for large cohorts of DGS children neither in research in the US (Solot et al., 2000), nor in Sweden (Persson et al., 2006).



Key: TLD=typical language development, SLI=specific language impairment, DGS=DiGeorge syndrome
CIT=Clitics-in-Islands Test, EVT=Expressive Vocabulary Test, PPVT=Peabody Picture Vocabulary Test

Figure 4. Performance of PI in comparison to the TLD group and the SLI group on the CIT, the EVT and the PPVT

In relation to narrative retell production, PI neither showed difficulties retelling the information of the story nor using subordinate clauses or producing long sentences. This finding stands in stark contrast to what was reported for the Swedish cohort (Persson et al., 2006), where the majority of DGS children showed a low information score, lower number of subordinate clauses, and shorter sentence length than expected, according to the Swedish BST norms. However, PI also produced a significantly smaller *number of sentences* on the BST retell task. In our search of the literature, we have not found reported deficits in sentence productivity for DGS preschoolers, so we cannot classify PI with respect to a cross-linguistic profile for DGS on this aspect of language abilities.

Taking into consideration our second aim, PI was significantly more impaired in comparison to his typically developing, chronological age-matched peers on global language performance (total DVIQ score), and specifically on the two subtests *comprehension of morphosyntax* and *sentence repetition*. He was also significantly different on *receptive vocabulary* abilities, scoring way below his TLD peers. On the other hand, no such difference was evident for expressive vocabulary. On the narrative retell task, only the *number of sentences produced* was significantly lower for PI as compared to his TLD peers. In addition, he performed significantly worse than his TLD peers on the *CIT*. In other words, PI showed similar performance as the TLD group for comprehension of metalinguistic concepts (DVIQ), production of morphosyntax (DVIQ), retell information (BST), MLU-word (BST), A5LS (BST), number of subordinate clauses produced (BST), and expressive vocabulary (DVIQ and EVT).

Our third aim was to compare PI with a group of chronological age-matched children with SLI. Global language abilities (total DVIQ score) did not differentiate PI from the SLI group. In a similar vein, comprehension of metalinguistic concepts (DVIQ), production of morphosyntax (DVIQ), retell information (BST), MLU–word (BST), A5LS (BST), number of subordinate clauses produced (BST), expressive vocabulary (DVIQ and EVT), and receptive language abilities (PPVT) did not differentiate PI with DGS from the SLI group. He was significantly worse compared to the SLI children (i) on the *comprehension of morphosyntax* (DVIQ), (ii) on the *sentence repetition* (DVIQ), (iii) on the *Clitics-in-Islands Test* (CIT), and (iv) on the *total number of sentences produced* (BST).

Putting both group results together, those of the children with TLD and those of the children with SLI, PI was significantly worse on the number of sentences produced on the BST retell, on comprehension of morphosyntax, on sentence repetition, and on the clitic production task. The last three tasks probe structural language (morphosyntactic and language complexity). Unfortunately, we do not yet have a solid analytical knowledge base for the relevance of complex language stemming from the BST and the DVIQ as a marker of language difficulties. This is to say that we can describe the performance by individuals and groups, but we cannot yet pinpoint the source of deviations from the norm.

The relevance of clitic productions and their placement in the context of first language acquisition of Cypriot Greek has been highlighted in work from our research group since Grohmann (2011); see, for example, the recent summaries in Grohmann (2014) and Grohmann & Kambanaros (to appear). PI clearly behaves differently from both children with TLD and children with SLI by producing significantly fewer clitics than both. However, his clitic placements resembled more those of the typically developing peers, while he showed more omissions than either group, a phenomenon which is rare even for children with SLI (Theodorou & Grohmann, 2015). Clitic production vis-à-vis omission has been taken as a clinical marker for SLI in other languages, though it is unlikely to be a clinical marker for SLI in Cypriot Greek (Theodorou & Grohmann, 2015); see also Theodorou (2013) for further discussion and references.

Conclusion

The purpose of the present study was to provide evidence for the language profile of DGS. Based on our findings of a single case, we opt for a distinctive language profile of DGS for our participant, as he appeared significantly impaired on receptive and expressive measures of complex language, not evidenced by the chronological age-matched group with SLI in our study. While similar in terms of below-TLD performance, his inferior abilities in complex language also do not match those of SLI qualitatively. However, future work will have to decide on the final outcome. In that respect, we do hope that our findings provide awareness of DGS. They surely constitute a first contribution to the knowledge base of the behavioural language phenotype for (Cypriot) Greek, even if only based on a single case.

This study was a preliminary investigation of the language profile of DGS compared to children with SLI (as well as a typically developing control group). While the study presents data that support further research using a comparison group of children with SLI, several limitations were apparent based on the small number of participants. This precludes big generalisations for the different populations as a whole. However, the results of this study indicate the potential benefits of research with larger numbers of DGS and SLI children in order to tease apart the linguistic profiles of each group.

Seen from the perspective of a larger research agenda, further exploring the exact deficits in language and cognition presented by pathologies like DiGeorge syndrome contributes to the growing research interest in comparative biolinguistics (Boeckx, 2013; Boeckx & Grohmann, 2013; Benítez-Burraco &

Boeckx, 2014; Kambanaros & Grohmann, 2015). This research programme investigates similarities and, especially, differences in specific tasks and abilities across different pathologies, from developmental language impairment and acquired language disorders to apparently non-linguistic pathologies, that is, those that are not primarily connected to language. By so doing, we may be able to shed light on the assumed invariance of the human language faculty, perhaps even "uncover the locus of variation (and its constraints) across genotypes, pathologies, or across species" (Leivada, 2014: 54). The present paper contributes to this endeavour.

References

Benítez-Burraco, A., & Boeckx, C. (2014). Universal Grammar and biological variation: An EvoDevo agenda for comparative biolinguistics. *Biological Theory*, *9*, 122-134.

Boeckx, C. (2013). Biolinguistics: Forays into human cognitive biology. *Journal of Anthropological Sciences*, *91*, 1-28.

Boeckx, C., & Grohmann, K. K. (eds.). (2013). *The Cambridge handbook of biolinguistics*. Cambridge, UK: Cambridge University Press.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*, 482-486.

De Decker, H. P., & Lawrenson, J. B. (2001). The 22q11.2 deletion: From diversity to a single gene theory. *Genetics in Medicine*, *3*(1), 2-5.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

European Social Survey. (2010). *Round 5 source showcards*. London: Centre for Comparative Social Surveys, City University London.

Goorhuis-Brouwer, S. M., Dikkers, F. G., Robinson, P. H., & Kerstjens-Frederikse, W. S. (2003). Specific language impairment in children with velocardiofacial syndrome: Four case studies. *Cleft Palate-Craniofacial Journal*, *40*(2), 190-195.

Grohmann, K. K. (2011). Some directions for the systematic investigation of the acquisition of Cypriot Greek: A new perspective on production abilities from object clitic placement. In E. Rinke & T. Kupisch (eds.), *The development of grammar: Language acquisition and diachronic change* (pp. 179-203). Amsterdam: John Benjamins.

Grohmann, K. K. (2014). CAT research on object clitic placement: Where we are now. In K. K. Grohmann & T. Neokleous (eds.), *Developments in the acquisition of clitics* (pp. 1-40). Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Grohmann, K. K., & Kambanaros, M. (to appear). The gradience of multilingualism in language development: Positioning bilectalism within comparative bilingualism. *Frontiers in Psychology: Language Sciences*.

Kambanaros, M., & Grohmann, K. K. (2015). Grammatical class effects across impaired child and adult populations. *Frontiers in Psychology: Language Sciences*, doi:10.3389/fpsyg.2015.01670.

Kambanaros, M., Grohmann, K. K., Michaelides, M., & Theodorou, E. (2013). Comparing multilingual children with SLI to their bilectal peers: Evidence from object and action picture naming. *International Journal of Multilingualism*, *10*, 60-81.

Kambanaros, M., Grohmann, K. K., Michaelides, M., & Theodorou, E. (2014). On the nature of verb–noun dissociations in bilectal SLI: A psycholinguistic perspective from Greek. *Bilingualism: Language and Cognition*, *17*(1), 169-188.

Leivada, E. (2014). From comparative languistics to comparative (bio)linguistics: Reflections on variation. *Biolinguistics*, *8*, 53-66.

Mastropavlou, M., Petinou, K., & Tsimpli, I. M. (2011). The role of morpho-phonological salience in tense marking: A comparison between Greek and Cypriot-Greek SLI children. In A. Tsangalides (ed.), *Selected papers from the 19th International Symposium on Theoretical and Applied Linguistics* (Thessaloniki, 3-5 April 2009) (pp. 313-328). Thessaloniki: Monochromia.

Niklasson, L., Rasmussen, P., Óskarsdóttir, S., & Gillberg, C. (2001). Neuropsychiatric disorders in the 22q11 deletion syndrome. *Genetics in Medicine, 3*(1), 79-84.

Persson, C., Niklasson, L., Óskarsdóttir, S., Johansson, S., Jönsson, R., & Söderpalm, E. (2006). Language skills in 5-8 year old children with 22q11 deletion syndrome. *International Journal of Language and Communication Disorders*, *41*(3), 313-333.

Petinou, K., & Okalidou, A. (2006). Speech patterns in Cypriot-Greek late talkers. *Applied Psycholinguistics, 27*, 335-353.

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for Raven's progressive matrices and vocabulary scales.*

San Antonio, TX: Harcourt Assessment.
Renfrew, C. (1997). *The Renfrew language scales*, 4th ed. Milton Keynes: Speechmark.
Rowe, C., & Grohmann, K. K. (2013). Discrete bilectalism: Towards co-overt prestige and diglossic shift in Cyprus. *International Journal of the Sociology of Language*, *224*, 119-142.
Scandurra, V., Scordo, M. R., Canitano, R., & de Bruin, E. I. (2013). 22q11 Deletion syndrome and multiple complex developmental disorder: A case report. *Journal of Clinical Medical Research*, *5*(2), 135-139.
Shprintzen, R. J. (2008). Velocardiofacial syndrome: 30 years of study. *Developmental Disability Research Reviews*, *14*(1), 3-10.
Sideridis, G., Antoniou, F., Mouzaki, A., & Simos, P. (2015). *Raven's coloured progressive matrices and vocabulary* [in Greek]. Athens, Greece: Motivo.
Simos, P. G., Kasselimis, D., & Mouzaki, A. (2011). Age, gender, and education effects on vocabulary measures in Greek. *Aphasiology*, *25*, 475-491.
Solot, C. B., Handler, S. D., Gerdes, M., McDonald-McGinn, D. M., Moss, E., Wang, P., Cohen, M., Randall, P., Larossa, D., & Driscoll, D. A. (2000). Communication disorders in the 22q11.2 microdeletion syndrome. *Journal of Communication Disorders*, *33*, 187-204.
Stavrakaki, S., & Tsimpli, I. M. (2000). Διαγνωστικό Τεστ Γλωσσικής Νοημοσύνης για παιδιά σχολικής και προσχολικής ηλικίας: στάθμιση, στατιστική ανάλυση, ψυχομετρικές ιδιότητες [Diagnostic Verbal IQ Test for Greek preschool and school age children: Standardization, statistical analysis, psychometric properties]. In M. Glykas & G. Kalomiris (eds.), *Proceedings of the 8th Symposium of the Panhellenic Association of Logopedists* (pp. 95-106) [in Greek]. Athens: Ellinika Grammata.
Theodorou, E. (2013). *Specific language impairment in Cypriot Greek: Diagnostic and experimental investigations*. PhD dissertation, University of Cyprus, Nicosia.
Theodorou, E., & Grohmann, K. K. (2015). Object clitics in Cypriot Greek children with SLI. *Lingua*, *161*, 144-158.
Varlokosta, S., Belletti, A., Costa, J., Friedmann, N., Gavarró, A., Grohmann, K. K., Guasti, M. T., Tuller L., et al. (2015). A crosslinguistic study of the acquisition of pronoun and clitic production. *Language Acquisition*, doi:10.1080/10489223.2015.1028628.
Vogindroukas, I., Protopapas, A., & Sideris, G. (2009). *Word-finding test* [in Greek]. Chania, Greece: Glafki.
Woodin, M., Wang, P. P., Aleman, D., McDonald-McGinn, D., Zackai, E., & Moss, E. (2001). Neuropsychological profile of children and adolescents with the 22q11.2 microdeletion. *Genetics in Medicine*, *3*(1), 34-39.

The MAIN of narrative performance: Russian-Greek bilingual children in Cyprus

Sviatlana Karpava[1,4], Maria Kambanaros[2,4], Kleanthes K. Grohmann[3,4]
skarpava@uclan.ac.uk, maria.kambanaros@cut.ac.cy, kleanthi@ucy.ac.cy

[1]University of Central Lancashire, Cyprus, [2]Cyprus University of Technology,
[3]University of Cyprus, [4]Cyprus Acquisition Team

Abstract. Narratives can help identify linguistic, cognitive, semantic, and social abilities as well as the communicative competence and cultural awareness of a child; cultural communities, language environment, home language use, parental attitudes towards bilingual and bicultural learning, and level of language proficiency are some of the factors that influence the development of narrative abilities. The pilot study reported here investigates narrative performance by Russian-(Cypriot) Greek bilingual children in both of their languages, Russian and Cypriot Greek. A total of 23 simultaneous bilingual children across different age groups ranging from 3 to 11 years of age were tested with the MAIN, the Multilingual Assessment Instrument for Narratives, a narrative tool developed within COST Action IS0804. All participants were also tested on a battery of other tools, including proficiency tests for Russian and Cypriot Greek as well as several tasks assessing executive control. With regard to narrative abilities, the bilingual children performed similarly across both languages. Not surprisingly, their performance was higher on the retelling condition than on the telling condition. Also as expected, the bilingual children's narrative abilities improve with age, although the number of participants in each age group is too low to determine a concrete trajectory. A comparison of the bilingual participants' telling and retelling narrative productions with that of monolingual (standard) Greek- and monolingual Russian-speaking peers shows that these outperform the bilinguals mainly in story structure and internal state terms.

Keywords: bilingualism, narrative, communicative competence, macro-structure, telling, retelling

Introduction

The present study investigates the narrative performance of bilingual children with typical language development in both their languages, Russian and Greek. Concretely, as the research takes place in Cyprus, which is characterised by diglossia between the local variety and the standard language (see Rowe & Grohmann, 2013 for a recent overview), Cypriot Greek (henceforth, CG) was assessed where relevant. The relevance is three-fold. First, the local variety (CG) spoken in a linguistic environment where the official language is Standard Modern Greek (SMG) leads to children growing up to become '(discrete) bilectal' speakers (Rowe & Grohmann, 2013); for narrative abilities, it would be interesting to be able to distinguish between monolingual-mono(dia)lectal and monolingual-bilectal children. Second, it will be instructive to compare bilingual-mono(dia)lectal children with those participating in our research, bilingual-bilectal ones; they are arguably simultaneously bilingual, yet sequentially bilectal (for discussion, see Grohmann & Kambanaros, to appear). Third, the data of our particular group of Russian–Greek children, namely bilingual-bilectals, can be compared to that of their peers acquiring either language monolingually (regardless of dialectal issues); these would be children from Russia and Greece, respectively. (A fourth possible relevance is briefly presented right below.)

For the purposes of this research, narrative performance is measured by macro-structure in telling and retelling conditions, along the dimensions of story structure, structure complexity, and internal states terms. Other factors that have already been partially considered (and will be expanded in the future), include children's language competence, language of narration, executive control, chronological age, and schooling level, which have all been identified as relevant in the bilectal context (Grohmann & Kambanaros, to appear).

We choose narratives as a window into the bilingual children's communicative development because it has been argued that narrative performance can help identify linguistic, cognitive, semantic, and social abilities as well as a child's communicative competence and cultural awareness (e.g., Olley, 1989; Schneider, Hayward, & Dubé, 2006; Paradis, Genesee, & Crago, 2010). In fact, narrations are often employed in contexts of language assessment, for example, to probe for possible language delay or impairment (e.g., Leonard, 1998); as a fourth possible relevance of our research, then, data from language-impaired children could be used for additional comparison. In the long run, this becomes particularly relevant in the context of early diagnosis of speech/language and communication difficulties in bilingual children who often show language behaviour that is reminiscent of language-impaired children, even and especially in the absence of any developmental language problems (for a very recent up-to-date overview of these issues, see Armon-Lotem & de Jong, 2015).

It has also been suggested that narrative skills are important for children's scholastic achievements, since there is a close relationship between oral language skills and literacy (Snow, 2002). In turn, cultural communities, language environment, home language use, parental attitudes towards bilingual and bicultural learning, and the level of language proficiency are some of the additional factors that may affect children's development of narrative abilities (e.g., Jia, Yiu, Duncan, & Paradis, 2011).

Narratives are thus one measure to assess children's speech and language abilities (e.g., Hadley, 1998; Boudreau, 2008) and their communicative competence (e.g., Norbury & Bishop, 2003). Analysis of narrative productions can also be applied to different cultural and social populations such as bi- and multilingual children. These differ from their monolingual peers regarding narrative abilities, with some relevant dimensions of evaluation being dissimilar languages, possible variance in proficiency levels for the two or more languages, a different language environment, and more diverse cultural communities (Gutiérrez-Clellen, 2002; Fiestas & Peña, 2004). Narrative micro- and macro-structures also depend on non-linguistic factors such as executive control, which comes out as the effect of working memory, sequencing, and planning on the stories produced (Coelho, 2002). Note that there is a long-standing claim that bilingualism enhances children's development of executive functions, that is, the set of cognitive processes that underlie flexible and goal-directed behaviour, often referred to as the 'cognitive advantage of bilingualism' (for recent overview, see Costa & Sebastián-Gallés, 2014).

Turning to analyses of narrative productions, the data reported here have not yet been subjected to a detailed micro-structural analysis, which is sensitive to language-specific aspects; macro-structure is less language-specific and even language-dependent (e.g., Uccelli & Paez, 2007), which is why we concentrate on this level of analysis for the present study. Macro-structural story structure has been analysed to include story grammar, narrative quality, plotline, temporal-causal connection, episode structure completeness, and so-called Goal-Attempt-Outcome structure. Story grammar itself requires knowledge of both semantic-pragmatic information and a super-structural level of discourse organisation (see Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012; 2015 for details and references).

Method

Research questions

The main research questions of this study are the following:

1. With respect to narratives, do bilingual-bilectal children perform differently in each of their languages, Russian and (Cypriot) Greek?
2. Does mode of narration (telling/retelling) influence story structure, structural complexity, and the production of internal state terms by bilingual children, in either language or even both?
3. Are the bilingual children's narrative productions similar to or different from monolingual children, language-impaired children, and other bilinguals with different language pairs?
4. What role do variables such as age, schooling, proficiency level, cognitive abilities, and executive functions play in bilingual children's narrative performance?

What we report next is the result of a pilot study. It is meant to pave the way for a larger-scale cohort research project, which is why we chose (few) participants for all age groups ranging from as low as 3 to as old as 11 years of age. One rationale is, of course, to test the validity of the tool used, the MAIN

(Gagarina et al., 2015): Does it work for this particular bilingual population of children acquiring Russian and (Cypriot) Greek? Can it differentiate bilingualism (Russian, Greek) from bilectalism (Cypriot Greek, Standard Modern Greek)? Is it age-appropriate across a wide range? Does it produce sound data that can be analysed quantitatively and qualitatively? From this perspective, a different way to present the data would be as many different single-case studies, namely one for each of the 23 participating children. However, we believe that such a presentation would turn out even more complex, and confusing for the reader, so we decided to present it as if it were a bone fide cohort study - with the added wrinkle that, due to very low number of participant numbers for most of the age groups, the effect is arguably more cosmetic than methodological.

Participants

The participants of this study were 23 Russian-Cypriot Greek simultaneous bilinguals, 11 girls and 12 boys. Their age ranged from 3 to 11 years, though participants numbers for most of the age groups were very small indeed: 3;1 (N=1), 4;8 (N=2), 5;0–5;6 (N=5), 6;0–6;11 (N=9), 7;11 (N=2), 9;5 (N=2), 10;11 (N=1), and 11;4 (N=1). At the time of testing, they attended kindergarten, pre-primary, and primary school classes. All participating children came from mixed-marriage families, with a Greek Cypriot father and a Russian mother, in a middle-class setting. They were randomly recruited in urban and rural areas of the Larnaca and Nicosia districts of Cyprus.

Materials and procedure

The Multilingual Assessment Instrument for Narratives (MAIN), a narrative tool developed within COST Action IS0804, was used in order to elicit stories from bilingual children (Gagarina et al., 2012; 2015). The MAIN consists of four comparable six-picture stories. Two of the picture sequences were used for the telling condition and another two for retelling. Each sequence consists of six coloured pictures without text.

For the telling mode, the Baby Goats and the Hungry Cat stories were chosen; children were asked to come up with and tell the experimenter a story based on the six pictures. For the retelling mode, the Baby Birds and the Naughty Dog stories were chosen; children were asked to first listen to the story told by the experimenter and then retell it. There was mutual sharing of the visual context and stimuli between child and examiner. During the testing both the child and the examiner could see the pictures. Each child was tested individually in their home environment.

All participants were also tested on a battery of additional tests: the Diagnostic Verbal IQ Test (Stavrakaki & Tsimpli, 2000), adapted to CG from the Standard Modern Greek original (Theodorou, 2013), the Russian Proficiency Test for Multilingual Children (Gagarina, Klassert, & Topaj, 2010), and several tasks assessing executive functions (digit span test, word span test, fluency test, Raven's Coloured Progressive Matrices). A parental questionnaire focusing on participants' socio-economic and family language background was also used (Gagarina et al., 2010).

Data analysis

All data were recorded, transcribed, and analysed in terms of story structure, structural complexity per episode, and internal state terms. The analysis of story structure includes setting, mental state as initiating event, goal, attempt, outcome, and mental state as reaction (3 episodes in total). Structural complexity per episode (episode completeness) focused on whether children used Goal-Attempt-Outcome (GAO) in every episode. Internal state terms denote the types of mental state terms used by bilingual children in their narrative production.

Since Premack & Woodruff (1978), mental or internal state terms (ISTs) have been argued to relate to theory of mind and cognitive abilities. There are different types of ISTs, which can be classified, for example, into six categories (for further discussion, see Klop, 2011; Gagarina et al., 2012; 2015): perceptual verbs (such as *see*, *hear*, *feel*, *smell*), physiological adjectives (*thirsty*, *hungry*, *tired*, *sore*), predicates expressing consciousness (*alive*, *awake*, *asleep*), emotional adjectives (e.g., *sad*, *happy*, *angry*, *worried*, *disappointed*), mental predicates (e.g., *want*, *think*, *know*, *forget*, *decide*, *believe*, *wonder*, *have/make a plan*), and verbs of saying or 'linguistic verbs' (e.g., *say*, *call*, *shout*, *warn*, *ask*).

Results

Overall, it was found that with regard to narrative abilities (macro-structure: story structure, structural complexity, and ISTs), the bilingual children performed similarly across their two languages (slightly better for CG). Their performance was also higher on the retelling than the telling condition. With respect to story structure for the four stories, within and cross-language comparison showed that the bilingual Russian-CG children performed better in retelling than in telling. For the telling mode, there was no crucial difference between the Russian and CG productions of the two stories, Baby Goat and Hungry Cat. For retelling of the Baby Bird story, the bilingual children had a slightly better production in CG than in Russian, but the opposite held for the Naughty Dog story. Table 1 presents the overall scores (and percentages) for the quantitative measurement of story structure per story for the combined group of 23 participants (with the proviso that it includes productions from individuals aged 3 to 11 lumped together); the MAIN manual provides a maximum score of 17 for each story.

Table 1. Story structure (telling vs. retelling)

Story structure	CG	Russian
Baby Goat (telling)	141 (36.1%)	143 (36.6%)
Hungry Cat (telling)	144 (36.8%)	145 (37.1%)
Baby Bird (retelling)	188 (48.1%)	169 (43.2%)
Naughty Dog (retelling)	180 (46.0%)	186 (47.6%)

Table 2. Story complexity (telling vs. retelling)

Story/Mode	Struct. Complex. (TOTAL)	AO/AA	GA/GO	GAO
CG Baby Birds story structure	121	32	32	57
Russian Baby Birds story structure	99	37	38	24
CG Naughty Dog story structure	102	36	24	42
Russian Naughty Dog story structure	79	24	28	27
RETELLING TOTAL	**401**	**129**	**122**	**150**
CG Baby Goat story structure	76	49	18	9
Russian Baby Goat story structure	75	45	15	15
CG Hungry Cat story structure	62	18	20	24
Russian Hungry Cat story structure	61	13	18	30
TELLING TOTAL	**274**	**125**	**71**	**78**
CG TOTAL	**361**	**135**	**94**	**132**
Russian TOTAL	**314**	**119**	**99**	**96**

Table 3. Story structure, structural complexity, and internal state terms (means)

VARIABLES				TELLING			RETELLING		
	Language	Age	N	Story structure	Structural complexity	Internal state terms	Story structure	Structural complexity	Internal state terms
Mean	CG	80	23	6.2	3	3.8	8	4.8	5.5
SD		23.6		2.4	1.5	2.2	2.6	1.5	2.9
Mean	Russian	80	23	6.2	2.9	4.1	7.7	3.8	5.9
SD		23.6		2.6	1.4	1.9	2.6	1.5	2.1

Table 4. Internal state terms (telling vs. retelling)

INTERNAL STATE TERMS (TELLING/PRODUCTION)							
	Language	Perceptual state terms	Physiological state terms	Consciousness terms	Emotion terms	Mental verbs	Linguistic verbs
Total	Cypriot Greek	68	0	0	29	61	15
Mean		2.95	0	0	1.26	2.65	0.65
SD		2.28	0	0	1.83	2.08	1.43
Total	Russian	84	1	0	27	62	12
Mean		3.65	0.04	0	1.17	2.69	0.52
SD		2.65	0.20	0	1.33	1.86	0.89
INTERNAL STATE TERMS (RETELLING/PRODUCTION)							
	Language	Perceptual state terms	Physiological state terms	Consciousness terms	Emotion terms	Mental verbs	Linguistic verbs
Total	Cypriot Greek	79	4	12	67	58	33
Mean		3.43	0.17	0.52	2.91	2.52	1.43
SD		2.29	0.38	2.5	2.72	1.92	1.30
Total	Russian	89	9	0	55	73	46
Mean		3.86	0.39	0	2.39	3.17	2
SD		2.83	0.72	0	2.03	2.55	1.70

The analysis of structural complexity, qualitative organisation of episode structure, and macro-proposition that compose the plot (GAO) showed that the bilingual Russian-CG children had more structural complexity in retelling than in telling (total structural complexity, GA/GO, GAO); they also performed better in CG than in Russian. See Table 2 and Table 3 for the results, respectively.

The reader will recall the small participant numbers across too many age groups. Consequently, the results depicted in Tables 2 and 3 are obviously not very helpful, only indicative: They correspond to the mean performance of all children combined, from the single 3-year-old to the single 11-year-old. The above-mentioned different single-case studies approach might work better, thus presenting each child's scores individually. However, current space restrictions do not allow such a detailed analysis, which is why we restrict ourselves to reporting the data as if collected from a comparable cohort.

The analysis of the data further showed that the bilingual children used more ISTs in retelling than in telling. Specifically, they used more perceptual state terms, emotion terms, and mental verbs rather than psychological, consciousness terms, and linguistic verbs, as shown in table 4. According to a paired samples *t*-test, there is a statistically significant difference between ISTs in CG telling mode and CG retelling move (*$t_{(21)}$=4.577; p=.000*) as well as between ISTs in Russian telling and Russian retelling modes (*$t_{(21)}$=4.902; p=.000*).

These findings are somewhat more meaningful than the above. For starters, regardless of age (from 3 to 11), certain ISTs were never used, others very rarely, and yet others more frequently. The table also highlights similarities and potential differences between the two languages. In order to explore this further, however, an individual approach would have to be taken, which cannot be done here.

Please see Appendix 1 for the bilingual children's individual productions on story structure, Appendix 2 on structural complexity per episode, and Appendix 3 on total number of ISTs in tokens.

As expected, the bilingual children's narrative abilities in CG improve with age, although the numbers of participants in each age group are too low to generalise this (beyond 5- and 6-year-olds, perhaps); there is no such clear picture for Russian. Table 5 presents the raw scores for each age group and as such is more informative than Table 1 above. Note that all children were tested only on production and not comprehension. As said above, the maximum score for story structure is 17, while for internal state terms, the total number of IST tokens is counted. In terms of structural complexity per episode, within the Goal-Attempt-Outcome structured episodes, the scores represent how often a participant produces partial event sequences (AO, AA), incomplete episodes (GA/GO), and the targeted fully complete episodes (GAO); consequently, children get 1 point for each AO/AA, 2 points for each GA/GO and 3 points for each GAO (Gagarina et al., 2015).

Table 5. Story structure, structural complexity, and internal state terms (Age: telling vs. retelling)

VARIABLES				**TELLING**			**RETELLING**		
Age	Language	Age months	N	Story structure	Structural complexity	Internal state terms	Story structure	Structural complexity	Internal state terms
3-year-olds	CG	47	1	3.5	2	3.5	4	3.5	0.5
	Russian			3	2	2	4.5	3	5
4-year-olds	CG	56	2	6	2.5	4.2	7.5	5.5	6
	Russian			6.5	2.7	3	7.2	3.2	6.5
5-year-olds	CG	63.8	5	3.8	1.7	2.5	6.3	4.5	3.5
	Russian			2.8	1.4	3.3	5	3.1	4.5
6-year-olds	CG	75.4	9	6.7	3.2	3	8.2	4.6	5.3
	Russian			7.8	3.3	4.2	8.4	4	5.7
7-year-olds	CG	95	2	8	4.2	6	10.2	5.7	7
	Russian			6.7	2.7	6	8.25	4	6.7
9-year-olds	CG	113	2	8.5	4	7.7	10.2	5	11
	Russian			8	5.5	5.7	10.5	4.5	8
10-year-olds	CG	131	1	8	5.5	6	11	7	8
	Russian			8.5	5	6.5	12	7	9
11-year-olds	CG	136	1	6.5	2	3	7	4	4
	Russian			5	1.5	2	8	3.5	5.5

Since this table breaks the participants down into age groups, there is some comparability for each row. We thus yield a first indication of what age-related differences in performance could look like. Due to the low number of participants, it does not make sense, though, to dwell on this further; more data from more participants are needed for each age group, except perhaps the 5- and 6-year-olds:

There is a noticeable level of improvement from age 5 (N=5) to age 6 (N=9) for each level of macro-structural analysis as well as for retelling over telling, and for both languages.

Next, we compare our data from bilingual Russian-CG children with available data on monolingual Russian- and monolingual/bilectal CG-speaking children (Gagarina et al., 2012), both with typical language development (TLD) and with specific language impairment (SLI). Looking at the narrative productions in both modes (telling and retelling), monolinguals outperform their bilingual peers mainly in story structure, as shown in Table 6.

Table 6. Bilingual TLD vs. monolingual TLD and SLI (telling vs. retelling)

			TELLING/PRODUCTION			**RETELLING/PRODUCTION**		
Language	Age	N	Story structure	Structural complexity	Internal state terms	Story structure	Structural complexity	Internal state terms
Monolingual children with TLD (Gagarina et al. 2012: 96)								
Cypriot Greek	79.8	6	5.0	1.5	5.5	5.2	0.3	6.1
Russian	68	15	7.3	N/A	1.3	14.8	N/A	2.7
Greek	73.0	5	9.8	1.8	5.9	11.5	1.4	6.6
Bilingual children								
Cypriot Greek	63.8	5	3.8	1.7	2.5	6.3	4.5	3.5
Russian			2.8	1.4	3.3	5	3.1	4.5
Cypriot Greek	75.4	9	6.7	3.2	3	8.2	4.6	5.3
Russian			7.8	3.3	4.2	8.4	4	5.7
Cypriot Greek	95	2	8	4.2	6	10.2	5.7	7
Russian			6.7	2.7	6	8.25	4	6.7
Cypriot Greek	113	2	8.5	4	7.7	10.2	5	11
Russian			8	5.5	5.7	10.5	4.5	8
Monolingual children with SLI (Gagarina et al., 2012: 96)								
Russian	68	9	6.7	N/A	1.9	6.7	N/A	2.1
Greek	100.6	18	3.9	0.4	2.8	5.8	1.0	5.5

Note that we only use a subset of the bilingual participants (total N=18), somewhat matching the children's age from the studies compared to.

The comparison of our data from bilingual Russian-CG children with available data on monolingual Russian children (Gagarina et al., 2012), both with TLD and with SLI, showed that monolingual Russian children with TLD outperform their bilingual peers on story structure (telling and retelling), while the bilingual Russian-CG children scored higher on internal state terms (telling and retelling). Bilingual Russian-CG children were closer to monolingual Russian children with SLI in terms of story structure, but they were better on internal state terms. This is shown in Table 7. Again, we only employed a subset (total N=14) to match those children we have data for.

Table 7. Bilingual TLD vs. monolingual TLD and SLI (Russian: telling vs. retelling)

			TELLING/PRODUCTION			**RETELLING/PRODUCTION**		
Language	Age	N	Story structure	Structural complexity	Internal state terms	Story structure	Structural complexity	Internal state terms
Monolingual children with TLD (Gagarina et al. 2012: 96)								
Russian	68	15	7.3	N/A	1.3	14.8	N/A	2.7
Bilingual children								
Russian	63.8	5	2.8	1.4	3.3	5	3.1	4.5
Russian	75.4	9	7.8	3.3	4.2	8.4	4	5.7
Monolingual children with SLI (Gagarina et al., 2012: 96)								
Russian	68	9	6.7	N/A	1.9	6.7	N/A	2.1

The same comparison for Greek showed that the bilingual children were better than their monolingual and bilectal peers in story structure and structural complexity, but worse with respect to ISTs (telling and retelling). Monolingual Standard Modern Greek (SMG) children from Greece with TLD performed higher than the bilingual children on story structure and ISTs (telling and retelling), but the bilingual children performed better on structural complexity. Overall, the bilingual children with TLD performed better than the monolingual SMG-speaking children with SLI; see Table 8.

Table 8. Bilingual TLD vs. monolingual TLD and SLI (Greek: telling vs. retelling)

			TELLING/PRODUCTION			**RETELLING/PRODUCTION**		
Language	Age	N	Story structure	Structural complexity	Internal state terms	Story structure	Structural complexity	Internal state terms
Monolingual children with TLD (Gagarina et al. 2012: 96)								
CG	79.8	6	5.0	1.5	5.5	5.2	0.3	6.1
SMG	73.0	5	9.8	1.8	5.9	11.5	1.4	6.6
Bilingual children								
CG	75.4	9	6.7	3.2	3	8.2	4.6	5.3
CG	95	2	8	4.2	6	10.2	5.7	7
Monolingual children with SLI (Gagarina et al., 2012: 96)								
SMG	100.6	18	3.9	0.4	2.8	5.8	1.0	5.5

The subset used for this comparison is even smaller (N=11), to get closer to the target ages compared with (bearing in mind that on top of a chronological age match, children with SLI are typically compared to peers of around 2 years younger for approximate language age match).

Not surprisingly, the bilingual children's narrative abilities in CG improve with their school grade, while (perhaps also not surprisingly) the opposite effect can be observed for Russian. Note that the bilingual children get more CG input than Russian, certainly in the school environment, and they live in a CG-dominant society in which arguably Greek language input increases with more schooling as well. However, we have not yet analysed the data individually to test for the possible factor of age-(in)appropriate schooling levels for some of the child participants. This is shown in table 9.

Table 9. Story structure, structural complexity, and internal state terms (Schooling: telling vs. retelling)

			TELLING/PRODUCTION			RETELLING/PRODUCTION		
Schooling	N	Language	Story structure	Structural complexity	Internal state terms	Story structure	Structural complexity	Internal state terms
kindergarten	3	CG	5.1	2.3	4	6.3	4.8	4.1
		Russian	5.3	2.5	2.6	6.3	3.1	6
pre-primary	9	CG	5.3	2.7	3.2	7.5	4.5	5.2
		Russian	4.7	2.2	3.8	6.5	3.2	5
primary	11	CG	7.1	3.4	4.2	8.8	5	6
		Russian	1.5	1.3	1.6	1.6	1.1	2

It was found that, overall, the bilingual children's narrative abilities in CG and in Russian increase with their level of proficiency in each language. This was measured by Diagnostic Verbal IQ Test (DVIQ) scores for CG, using Theodorou's (2013) CG adaptation of the SMG original (Stavrakaki & Tsimpli, 1999) and the Russian Proficiency Test for Multilingual Children (RPTMC) scores for Russian (Gagarina et al., 2010). Language proficiency is thus a good predictor of bilingual children's narrative abilities; see Table 10 for CG and Table 11 for Russian.

Table 10. Story structure, structural complexity, and internal state terms (DVIQ: telling vs. retelling)

			TELLING/PRODUCTION			RETELLING/PRODUCTION		
DVIQ scores	N	Language	Story structure	Structural complexity	Internal state terms	Story structure	Structural complexity	Internal state terms
low scores	6	CG	3.4	1.8	2.7	4.9	3.5	3
		Russian	3.4	1.8	2.3	4.9	2.3	3.7
mid scores	10	CG	7	3.3	4.6	8.9	5.3	7
		Russian	6.8	3.3	5	8.4	4.3	6.9
high scores	7	CG	7.3	3.5	3.5	9.3	5.3	4.4
		Russian	7.9	3.4	4.4	9	4.5	6.3

Table 11. Story structure, structural complexity, and internal state terms (RPTMC: telling vs. retelling)

			TELLING/PRODUCTION			RETELLING/PRODUCTION		
RPTMC scores	N	Language	Story structure	Structural complexity	Internal state terms	Story structure	Structural complexity	Internal state terms
low scores	8	CG	4.6	2.3	3.5	6.5	4.3	3.9
		Russian	3.7	1.8	3.1	5.5	3.1	5
mid scores	8	CG	6.3	1.5	1	8	4.8	5.2
		Russian	6.5	2.8	3.9	8.3	3.8	5.8
high scores	7	CG	7.7	3.5	4.9	9.5	5.3	7.5
		Russian	8.8	4.3	5.3	9.5	4.6	7

According to a one-way ANOVA, the scores of the Raven's Coloured Progressive Matrices (Raven, 1938), which measures cognitive abilities, general human intelligence, and perceptual and analytic processes (Mackintosh & Bennett, 2005), are not significant for CG and Russian narrative production (telling/retelling). Before analysing the data obtained from the cognitive tasks administered, more children need to be tested across all age groups in order for the results to have any relevant meaning.

Discussion

The aim of this study was to present the macro-structural analysis of narrative discourse abilities by bilingual children and to address the four defined research questions. We pursue four main questions in this research; at this stage, however, only a small set of children have been tested, which makes a concrete interpretation of the findings difficult, to put it mildly. But we can discern tendencies, which is why we did test at least one child from each relevant age group.

The first question concerns the possible influence of language on narrative productions by bilingual children. The analysis of the data showed that there is no significant language effect with respect to macro-structure: The bilingual Russian-CG children performed nearly the same for story structure, episode complexity, and internal state terms across both languages, though perhaps with a slight advantage for CG.

Analysis of story grammar, with the episode being its central unit, reveals children's comprehension and production of logical relationships, how people and events are related temporally and causally, while story structure, the logical relationship among the episode components, is not language-specific (e.g., Coelho, Liles, & Duffy, 1994). This may be the explanation of the absence of a language effect in the current study.

Structural complexity per episode, whether children are able to generate a complete episode (i.e. GAO), is related to their ability to develop logical schemas or structured event complexes (Grafman, 2002), which are stored in the dorso-lateral prefrontal cortex (Wood, Knutson & Grafman, 2005). Macro-level of narrative organisation, what we refer to as macro-structure here, presupposes thematic coherence and semantic-pragmatic information. Children need to be able to realise the overall structure of the narrative, including adequate amount of information, and take the listener's point of view or knowledge into consideration (Berman & Slobin, 1994). Failure to narrate successfully can be explained by children's inefficiency in establishing logical relationships between events (temporal and causal) or because they have difficulties with episode structure, the central unit of story grammar, which includes both linguistic and non-linguistic processes.

Episode completeness is composed of macro-propositions (van Dijk & Kintsch, 1983). Each episode has three propositions: goal, attempt, and outcome. The task of a child is to recognise the propositions and to sequence them in a logical way. The structural complexity of narrative texts can depend on cognitive resources (Swets, Jacovina, & Gerrig, 2014); bilingual children might have a cognitive resource deficit and thus produce shorter, structurally simpler phrases.

The second question concerns the effect of narration mode on the bilingual children's narrative production (story structure, structural complexity, and ISTs). According to Boudreau (2008), narrative performance is influenced by task demands and elicitation frameworks. The results revealed that the mode of narration influences narrative production, both in Russian and in CG. This is in line with previous findings that retelling elicits longer and more detailed narratives with a more complex story structure than the telling mode (e.g., Schneider et al., 2006; Duinmeijer, 2010).

Story generation and story retelling narrative tasks are both cognitively and linguistically demanding. Duinmeijer (2010) suggests that story generation is linguistically more demanding than story retelling. In story retelling, there is a scaffolding effect, as the story is first told by an adult and then the child is asked to retell that story. In story telling, the child does not have an example of the story and has to formulate the plotline all by herself. This task is more demanding than retelling, both cognitively and linguistically (Norbury & Bishop, 2003). Story generation involves such executive functions as working memory and attention. It is important that children are able to interpret visual information from the pictures in a correct way (Trabasso & Rodkin, 1994). According to Purvis & Tannock (1997), retelling also requires such cognitive skills as attention and memory. Children first need to understand the story so that it can be retold in a correct way (e.g., Andreu, Sanz-Torrent, Olmos, & MacWhinney, 2011). Our hope for, and expectation from, further research is that accurate measures of cognitive abilities for each age group will allow a correlation analysis; ideally, advanced abilities in working memory and attention would influence story complexity in the telling mode.

The third research question focused on the difference between typically developing bilingual-bilectal Russian-CG children and monolingual children, both with typical language development and with SLI. The analysis of the data showed that monolingual children perform better than bilingual children, in particular with respect to story structure and structural complexity, but not on ISTs. Analysis of internal-state language in children's narratives arguably reflects their theory of mind abilities as well as understanding and awareness of intentionality and goal-directed behaviour of protagonists (see e.g., Nippold, Ward-Lonergan, & Fanning, 2005). Macro-structure is universal and language-general, reflecting general narrative discourse competence (Pearson, 2002). Bilingual children lag behind their monolingual peers in terms of structural complexity, as they are not able to produce complete and well-formed episodes; they also lack an understanding of narrative schemata, causality, perspective-taking, ability to plan, and meta-awareness (Westby, 2005).

Our fourth and final research question aimed at addressing the influence of such variables as age, schooling, proficiency level, cognitive abilities, and executive functions on narrative abilities of bilingual children. It was found that some of the variables are more important than others. There is an obvious effect of chronological age and schooling, as would be expected from any set of typically developing children. Arguably the same can be said for language proficiency level as a factor for narrative production by bilingual children in their respective languages.

It was found that the scores of the Raven's Coloured Progressive Matrices, which measures cognitive abilities, general human intelligence, and perceptual and analytic processes do not correlate with narrative production in the set of bilingual-bilectal children studied, though we left out the details due to an insufficient participant number per age group. Informally speaking, the tendency (not reported above) seems to be that the children's scores in the word span, fluency and digit span tests, which measure short-term verbal memory, executive and cognitive functions, and semantic and phonological memory all affect narrative skills of bilingual children; in the absence of larger participant numbers, however, it is not at all clear whether this is a real effect or (more likely) the simple result of age. After all, these measures are all expected to yield higher scores in older ages, and a single or two participants per age group are not enough to allow more specific inferences.

Conclusions

In this paper, we focused on narrative macro-structure in the narrations by bilingual children and how it may be influenced by such factors as language, task, age, schooling, proficiency, and cognitive abilities. We understand macro-structural analysis to deal with higher-order hierarchical organisation of the discourse which includes story structure, episode structure, and internal state terms.

It was found that the bilingual children performed similarly across their two languages (slightly better for Cypriot Greek), which can be explained by shared cognitive ability in the two languages. Their performance was higher in retelling than in telling. Cognitive abilities and executive functions tend to influence narrative macro-structure of bilingual children as well. They used more internal state terms in the retelling mode, which can be explained by a scaffolding effect, and also perceptual state terms, emotion terms, and mental verbs. The perceptive and productive lexicon in Russian is correlated with the production of ISTs (telling and retelling). Bilingual children had more structural complexity in retelling than in telling, and more so in CG than in Russian. Statistical analysis showed that age, schooling level, and language proficiency affect bilingual narrative ability.

Due to the increasing number of multilingual children in Cyprus, it is important to assess their linguistic and cognitive development and to distinguish early between typically developing and language-impaired children. The study of language acquisition norms for typical language development, language delay, and impairment can help prevent misdiagnosis of bilingual children.

The limitation of the study is the small number of participants. It is necessary to increase the number of children tested for each age group. This study presented only the macro-structure analysis, a further study is needed to compare micro- and macro-structure in narrative productions of bilingual children. Also, both typically developing and language-impaired bilingual children should be tested in order to detect possible language impairment in bilingual population as early as possible, to evaluate their language, and to provide treatment.

The measures of story grammar, structural complexity, and internal state terms tap into the narrative abilities of bilingual children. It is important to note that a combination of these measures provide a more accurate depiction of the discourse/narrative performance of bilingual children. Multiple narrative analysis is needed in order to assess how well bilingual children produce a story in each of their languages. More research is necessary in order to test the effect of shared vs. non-shared information; the absence of shared visual knowledge can trigger children to produce more complex stories as children's communicative behaviour is influenced by the knowledge state of their listeners (e.g., Short-Masterson, 2010). A process-based dynamic approach to narrative assessment, which deals with both process and result of narration, can be implemented (Gillam, Peña, & Miller, 1999).

It is important to assess bilingual children's narrative ability, their linguistic performance of discourse level in both of their languages. This assessment should be combined with a thorough evaluation of their cognitive skills, syntactic, morphological, and lexical (receptive, perceptive) abilities. Further research on narrative abilities of bilingual populations is important, as it can provide insight into their communicative competence, literacy, and academic success.

References

Andreu, L., Sanz-Torrent, M., Olmos, J., & MacWhinney, B. (2011). Narrative comprehension and production in children with SLI: An eye movement study. *Clinical Linguists and Phonetics, 25*(9), 767-783.

Armon-Lotem, S., & de Jong, J. (2015). Introduction. In S. Armon-Lotem, J. de Jong, & N. Meir (eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 1-22). Bristol, UK: Multilingual Matters.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Boudreau, D. (2008). Narrative abilities: Advances in research and implications for clinical practice. *Topics in Language Disorders, 28*(2), 99-114.

Coelho, C.A. (2002). Story narratives of adults with closed head injury and non-brain injured adults: Influence of socioeconomic status, elicitation task, and executive functioning. *Journal of Speech, Language, and Hearing Research, 45*(6), 1232-1248.

Coelho, C. A., Liles, B., & Duffy, R. (1994). Cognitive framework: A description of discourse abilities in traumatically brain-injured adults. In R. Bloom, L. Obler, S. De Santi, & J. Ehrlich (eds.), *Discourse analysis and applications: Studies in adult clinical populations* (pp. 96-110). Hillsdale, NJ: Lawrence Erlbaum Associates.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience, 15*, 336-345.

Duinmeijer, I. (2010). *Narrative ability, auditory attention and verbal memory in children with specific language impairment and typically developing peers*. Unpublished master thesis, University of Amsterdam.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools, 35*(2), 155-168.

Gagarina, N., Klassert, A., & Topaj, N. (2010). Sprachstandstest Russisch für mehrsprachige Kinder [Russian language proficiency test for multilingual children]. *ZAS Papers in Linguistics, 54*.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). Narrative assessment instrument for (multilingual) children. *ZAS Papers in Linguistics, 56*.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243-276). Bristol: Multilingual Matters.

Gillam, R., Peña, E. D., & Miller, L. (1999). Dynamic assessment of narrative and expository discourse. *Topics in Language Disorders, 20* (1), 33-47.

Grafman, J. (2002). The structured event complex and the human prefrontal cortex. In D. T. Stuss & R. T. Knight (eds.), *Principles of frontal lobe function* (pp. 292-310). New York: Oxford University Press.
Grohmann, K. K., & Kambanaros, M. (to appear). The gradience of multilingualism in language development: Positioning bilectalism within comparative bilingualism. *Frontiers in Psychology: Language Sciences*.
Gutiérrez-Clellen, V. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, *13*(2), 175-197.
Hadley, P. A. (1998). Language sampling protocols for eliciting text-level discourse. *Language, Speech, and Hearing Services in Schools*, *29*, 132-147.
Jia, R., Yiu, E., Duncan, T., & Paradis, J. (2011). *The Chinese–English children's dual language development: Focus on narrative abilities*. Paper presented at the 16th AILA World Congress, Harmony in Diversity: Language, Culture, Society, Beijing.
Klop, D. (2011). *The relationship between narrative skills and reading comprehension: When mainstream learners show signs of specific language impairment.* Unpublished doctoral dissertation, Stellenbosch University.
Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.
Mackintosh, N. J., & Bennett, E. S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, *33*, 663-674.
Nippold, M., Ward-Lonergan, J., & Fanning, J. (2005). Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic, and pragmatic development. *Language, Speech, and Hearing Services in Schools*, *36*, 125-138.
Norbury, C. F., & Bishop, D. (2003). Narrative skills of children with communication impairment. *International Journal of Language and Communication Disorders*, *38*, 287-313.
Olley, L. (1989). Oral narrative performance of normal and language impaired school aged children. *Australian Journal of Human Communication Disorders*, *17*, 43-65.
Paradis, J., Genesee, F., & Crago, M. (2010). *Dual language development and disorders: A handbook on bilingualism and second language learning* (2nd ed.). Baltimore, MD: Paul H. Brookes.
Pearson, B. (2002). Bilingual infants: What we know, what we need to know. In M. Suarez-Orozco & M. Paez (eds.), *Latinos: Remarking America* (pp. 306-320). Berkeley, CA: University of California Press.
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515-526.
Purvis, K. L, & Tannock, R. (1997). Language abilities in children with attention deficit hyperactivity disorder, reading disabilities and normal controls. *Journal of Abnormal Child Psychology*, *25*(2), 133-144.
Raven, J. C. (1938). *Progressive matrices*. London, UK: H. K. Lewis.
Rowe, C., & Grohmann, K. K. (2013). Discrete bilectalism: Towards co-overt prestige and diglossic shift in Cyprus. *International Journal of the Sociology of Language*, *224*, 119-142.
Schneider, P., Hayward, D., & Dubé, R. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech Pathology and Audiology*, *30*, 224-238.
Short-Masterson, K. J. (2010). Preschoolers' establishment of mutual knowledge during scripted interactions. *First Language*, *30*(2), 219-236.
Snow, C. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: RAND Corporation.
Stavrakaki, S., & Tsimpli, I. M. (2000). Διαγνωστικό Τεστ Γλωσσικής Νοημοσύνης για παιδιά σχολικής και προσχολικής ηλικίας: στάθμιση, στατιστική ανάλυση, ψυχομετρικές ιδιότητες [Diagnostic Verbal IQ Test for Greek preschool and school age children: Standardization, statistical analysis, psychometric properties]. In: M. Glykas & G. Kalomiris (eds.), *Proceedings of the 8th Symposium of the Panhellenic Association of Logopedists* (pp. 95-106) [in Greek]. Athens, Greece: Ellinika Grammata.
Swets, B., Jacovina, M., & Gerrig, R. (2014). Individual differences in the scope of speech planning: Evidence from eye-movements. *Language and Cognition*, *6*, 12-44.
Theodorou, E. (2013). *Specific language impairment in Cypriot Greek: Diagnostic and experimental investigations*. PhD dissertation, University of Cyprus, Nicosia.
Trabasso, T., & Rodkin, P.C. (1994). Knowledge of goal/plans: A conceptual basis for narrating 'Frog where are you?' In R. Berman & D. Slobin (eds.), *Relating events in narrative: A crosslinguistic developmental study* (pp. 85-106). Hillsdale, NJ: Lawrence Erlbaum Associates.
Uccelli, P., & Páez, M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, *38*, 225-236.
van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
Westby, C. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (eds.), *Language and reading disabilities* (pp. 157-232). Boston, MA: Allyn & Bacon.

Wood, J., Knutson, K., & Grafman, J. (2005). Psychological structure and neural correlates of event knowledge. *Cerebral Cortex, 15*, 1155-1161.

Appendix 1: Individual productions per child for story structure (telling vs. retelling)

Story structure (total/17 scores)									
Child	Age	TELLING				RETELLING			
		Baby Goats		Hungry Cat		Baby Birds		Naughty Dog	
		CG	Russian	CG	Russian	CG	Russian	CG	Russian
1	3;11	5	4	2	2	4	5	4	4
2	4;8	4	4	4	7	7	6	3	8
3	4;8	9	6	7	9	10	8	10	7
4	5;0	5	3	10	5	8	4	7	3
5	5;4	3	4	2	3	7	6	4	6
6	5;4	1	1	1	3	4	4	5	4
7	5;5	3	1	2	1	3	1	4	1
8	5;6	6	5	5	2	11	10	10	11
9	6;0	4	10	5	8	8	8	8	11
10	6;0	7	5	4	5	7	8	10	8
11	6;0	5	6	4	7	5	5	3	7
12	6;0	9	9	7	13	11	9	6	11
13	6;0	5	5	9	6	6	8	8	6
14	6;4	8	8	10	7	13	10	11	7
15	6;8	8	8	8	10	9	9	7	9
16	6;8	9	11	8	7	10	7	11	12
17	6;11	5	8	6	9	6	9	10	8
18	7;11	7	9	8	8	11	6	12	12
19	7;11	8	3	9	7	9	7	9	8
20	9;5	12	9	7	6	12	10	11	11
21	9;5	3	9	12	8	9	11	9	10
22	10;11	6	8	10	9	12	11	10	13
23	11;4	9	7	4	3	6	7	8	9

Appendix 2: Individual productions per child for story complexity (telling vs. retelling)

Structural complexity per episode (AO/AA: 1 point each, GA/GO: 2 points each, GAO: 3 points each)									
Child	Age	TELLING				RETELLING			
		Baby Goats		Hungry Cat		Baby Birds		Naughty Dog	
		CG	Russian	CG	Russian	CG	Russian	CG	Russian
1	3;11	4	4	0	0	3	4	4	2
2	4;8	4	3	3	3	7	5	6	5
3	4;8	3	3	0	2	7	3	2	0
4	5;0	3	2	4	2	5	2	5	2
5	5;4	2	2	0	2	6	4	3	3
6	5;4	1	1	0	2	3	3	3	4
7	5;5	3	1	0	0	3	1	3	0
8	5;6	3	2	1	0	7	7	7	5
9	6;0	3	3	2	4	4	4	3	4
10	6;0	4	3	0	2	3	3	5	3
11	6;0	2	4	0	2	5	4	2	4
12	6;0	4	4	3	7	6	5	2	5
13	6;0	2	3	7	2	2	4	3	0
14	6;4	4	3	8	2	8	5	8	3
15	6;8	3	3	3	4	6	5	3	3
16	6;8	4	3	5	2	8	6	5	4
17	6;11	3	5	2	5	4	6	7	5
18	7;11	5	1	6	4	3	5	6	3
19	7;11	3	4	3	2	7	3	7	5
20	9;5	7	5	2	5	8	5	3	7
21	9;5	2	8	5	4	6	5	3	1
22	10;11	4	5	7	5	7	6	7	8
23	11;4	3	3	1	0	3	4	5	3

Appendix 3: Individual productions per child for internal state terms (telling vs. retelling)

Total number of internal state terms in tokens (IST: 1 point each)									
Child	Age	TELLING				RETELLING			
		Baby Goats		Hungry Cat		Baby Birds		Naughty Dog	
		CG	Russian	CG	Russian	CG	Russian	CG	Russian
1	3;11	3	3	4	1	0	7	1	3
2	4;8	0	0	2	3	6	7	2	3
3	4;8	11	5	4	4	10	10	6	6
4	5;0	3	5	4	7	3	7	2	6
5	5;4	5	6	3	2	2	5	2	2
6	5;4	1	4	3	4	1	3	5	4
7	5;5	0	0	3	0	2	2	4	0
8	5;6	1	3	2	2	8	7	6	9
9	6;0	3	5	2	5	4	7	2	10
10	6;0	2	2	3	2	8	5	7	4
11	6;0	2	2	0	4	2	2	2	3
12	6;0	5	5	5	9	10	8	7	8
13	6;0	4	1	5	4	6	6	5	3
14	6;4	4	6	6	7	10	4	6	7
15	6;8	2	4	2	6	6	7	5	5
16	6;8	2	7	2	2	6	7	4	8
17	6;11	2	3	3	2	3	6	4	3
18	7;11	4	4	5	6	5	5	6	7
19	7;11	4	5	11	9	8	9	9	6
20	9;5	9	7	6	6	17	13	9	6
21	9;5	7	6	9	4	11	7	7	6
22	10;11	6	4	6	9	10	10	6	8
23	11;4	4	3	2	1	5	5	3	6

Cross-linguistic interaction: A retrospective and prospective view

Margaret Kehoe
margaret.winkler-kehoe@unige.ch

Université de Genève

Abstract. This paper provides a critical review of research on cross-linguistic interaction in the phonetic and phonological development of young bilingual children. After presenting some examples of cross-linguistic interaction (acceleration, delay, transfer) in German-Spanish bilingual children tested in Hamburg (i.e. Hamburg study), it examines whether other investigators have documented the same results as the Hamburg study. It investigates studies which have tested similar contact situations or which have looked at similar predictive factors such as frequency, complexity, structural ambiguity, or dominance. This survey indicates that very few generalizations can be gleaned across studies. The paper then explores possible reasons for the lack of generalizations, which include methodological limitations and the lack of an appropriate research model. Certain suggestions are made to improve the research model, which involve taking into account themes from both first and second language acquisition. Incorporating additional interaction patterns into the current framework, as well as considering the developing speech-motor and lexical abilities of young children might lead to a better explanatory model of phonological interaction in early bilingualism. The paper ends with an examination of new perspectives in studying cross-linguistic interaction.

Keywords: bilingual children, cross-linguistic interaction, phonetic and phonological acquisition, acceleration, delay, transfer

Introduction

During the last two to three decades, there has been considerable research investigating the phonetic and phonological abilities of young bilingual children. One of the research's main goals has been to determine whether the speech of bilingual children differ in qualitative and quantitative ways from the speech of monolingual children. The presence of systematic differences between monolingual and bilingual speech suggests that there is interaction between the two linguistic systems of the bilingual, a phenomena referred to as cross-linguistic interaction. The aim of the current article is to examine the findings on cross-linguistic interaction in the area of phonetic and phonological development. This study takes a "retro-" and "prospective" view because it looks back on what we have found out in past studies and looks forward to what we should find out in future studies. Indeed, we have collected a great deal of information on the speech of young bilinguals, but not all of it forms a coherent picture or is easily generalizable. Thus, this study aims to find coherency and generalizations by systematically examining similar contact situations across bilingual children and by looking at the effects of the same predictive factors across a range of studies. Referring to the English expression "I can't see the woods for the trees" whereby someone has difficulty seeing a situation clearly because they are viewing only the details, this article strives to "see the woods".

This article focuses on young bilinguals, who are either simultaneous bilinguals (acquiring two languages from birth) or early sequential bilinguals (acquiring a second language after the first language but before the age of five years). We do not distinguish between the two groups in this early period because several authors suggest that simultaneous and early sequential bilinguals display similar, although not identical, patterns of acquisition (Fabiano-Smith & Goldstein, 2010a, b; Splendido, 2014, in press). The article is divided into two main parts: retrospective and prospective. The retrospective part includes an overview and historical perspective of phonetic and phonological research in early bilingualism, followed by a closer examination of cross-linguistic interaction. It then attempts to seek generalizations across studies by examining similar contact situations and by testing similar predictive factors. The prospective part starts with a critical look at current research and the research model, and moves on to a presentation of new perspectives.

Retrospective

Overview and historical perspective

The field of early bilingualism is situated between the more prominent fields of first and second language acquisition and it shares with them many commonalities. Both first language acquisition and early bilingualism are concerned with children who are in the process of developing their phonological representations, articulatory, acoustic-perceptual, and higher-level, as well as their speech-motor control. Even if children are acquiring their second-language slightly later than at birth, they still have not obtained adult-like perceptual and articulatory knowledge of the sounds in their first language (Munson, 2004; Munson, Edwards, & Beckman, 2005; Nittrouer, 1992). Early bilingualism and second language acquisition deal in both cases with language contact within an individual, and as such similar outcomes that arise from this contact may be anticipated.

If we examine speech production studies in early bilingualism from a historical perspective, two important landmarks can be discerned. The first is the idea promoted by Volterra and Taeschner (1978) that bilingual children speak a mixed sort of language at the beginning. The second is the notion of cross-linguistic interaction introduced by Paradis and Genesee (1996) who argue that bilingual children operate with two systems from the beginning but with the possibility of interaction between the two systems. Since this time, we have not seen any major change in the orientation of the research, although some recent studies suggest new developments, notably a study by Lleó and Cortes (2013) which attempts to model cross-linguistic interaction and one by Vihman (2015), which takes a more critical stance towards the Paradis and Genesee (1996) position, arguing that it is “more programmatic than empirically testable”. Given an emergent view of phonology, Vihman (2015) points out that the question of whether there is one or two systems need not be asked.

This section does not focus on the first landmark since the view of a single system as proposed by Volterra and Taeschner (1978) has been severely criticized both empirically and methodologically (De Houwer, 1990; Genesee, 1989; Meisel, 1989). There is evidence that two-year-old children pragmatically separate their languages (they speak the language of their interlocutor most of the time), which presupposes that they also differentiate their language at other levels (Paradis & Genesee, 1996). In retrospect, we may wonder whether we have been too hasty in discarding certain aspects of a unitary system proposal, not the one proposed by Volterra and Taeschner (1978), but one of a more differentiated nature. Certain findings in phonology are consistent with this. Studies based on whole-word proximity show that bilingual children maintain the same distance between the target form and their own production in both of their languages (Bunta, Davidovich, & Ingram, 2006; Bunta, Fabiano-Smith, Goldstein, & Ingram, 2009), which Bunta et al. (2006) argue is consistent with an “underlying unitary hypothesis”. That is, the child’s two languages share the same underlying phonological properties but differ in terms of surface manifestations. Later, in this article we will present findings which indicate that, in many cases, the phonological systems of bilingual children resemble each other more than do those of their monolingual counterparts. These “merging” patterns may reflect a pooling of phonetic and phonological resources, rather than a lack of differentiation between the two phonetic systems. Finally, Vihman (2015) questions how separate the linguistic systems of older bilinguals are and provides both experimental and anecdotal evidence for non-selective language use and processing by older bilinguals. In short, the possibility of a unitary system at a certain level may reflect integrated and efficient language systems rather than a state of confusion.

This section focuses instead on the program of research stimulated by Paradis and Genesee’s (1996) article, in which the possibility of both separation and interaction was entertained. Over the last two decades, studies of both a clinical or linguistic nature have compared the phonetic and phonological abilities of monolingual and bilingual children. By “clinical”, I refer to those studies which have measured general aspects of phonological acquisition such as percentage consonants correct (PCC), whole-word proximity, phonetic inventory, and phonological processes (Holm & Dodd, 1999; Fabiano-Smith & Goldstein, 2010b; Gildersleeve-Neumann & Wright, 2010; Goldstein & Washington, 2001; Goldstein, Fabiano, & Washington, 2005; Grech & Dodd, 2008). These studies have examined whether bilingual children differ from monolingual children in terms of rate or style of acquisition, and in terms of the presence of error patterns. By “linguistic”, I refer to those studies which have focused on specific phonetic and phonological properties (e.g., coda consonants, clusters, rhythm, Voice Onset Time (VOT)) in one or both of the bilingual’s languages (Almeida, Rose, & Freitas, 2012; Bunta & Ingram, 2007; Kehoe, Lleó, & Rakow, 2004; Lleó, Kuchenbrandt, Kehoe, & Trujillo, 2003; Mayr, Howells, & Lewis, 2015; Mok, 2011, 2013). These studies have focused on cross-linguistic interaction, the nature and direction of it, and how to account for it by appealing to factors such as frequency or complexity. The separation between these two sets of studies is not sharp and clinical-type studies have also attempted to account for cross-linguistic interaction and linguistic-type studies have also explored the clinical implications of the findings.

To summarize the results of these two sets of studies, we refer the reader to a review article by Hamby, Wren, McLeod, & Roulstone (2013) on the influence of bilingualism on speech production. This study summarized the findings of 66 studies (63 individual articles) conducted during the last 50 years (1960 to 2010) on the speech production of bilingual infants and adults. There are two caveats which concern this review article: 1. it only includes studies in which one of the bilingual’s languages is English; 2. Over one third of the studies (24 out of 66) are based on Spanish-English children, indicating that much of what we know on bilingual speech acquisition pertains to this population. Keeping these limitations in mind, the conclusions of this review article are clear cut. In terms of global measures (typical of the clinical-type studies), there is no evidence that the speech of bilinguals differs greatly from that of monolinguals: bilinguals may do better than monolinguals (Goldstein & Bunta, 2012 ; Grech & Dodd, 2008; Johnson & Lancaster, 1998), less well than monolinguals (Gildersleeve-Neumann, Kester, Davis, & Peña, 2008; Law & So, 2006), or behave similarly to monolinguals (Goldstein et al., 2005; MacLeod, Laukys, & Rvachew, 2011). The authors stress, nevertheless, that there may be qualitative differences in acquisition and increased variation in speech production. More in-depth studies (typical of the linguistic-type studies) suggest that acquisition of sounds and sound structures may be accelerated or delayed depending upon interaction between the specific language structures under consideration. The fact that different types of interaction patterns take place at the same time may explain why no overall differences between monolingual and bilinguals are observed on global measures. There may indeed be a cancelling out effect. In the next section we explore cross-linguistic interaction in more detail starting first with a definition of it.

Cross-linguistic Interaction

Paradis and Genesee (1996) define cross-linguistic interaction (or interdependence) as “the systemic influence of the grammar of one language on the grammar of the other language during acquisition, causing differences in a bilingual’s patterns and rates of development in comparison with a monolinguals” (p. 3). By “systemic”, they mean that the influence is at the level of representation and it is sustained over a period of time. Paradis and Genesee (1996) consider three potential manifestations of cross-linguistic interaction, which are summarized below**:**

1. **Transfer**: the incorporation of a grammatical property into one language from the other;
2. **Acceleration**: the situation in which a certain property emerges in the grammar earlier than would be the norm in monolingual acquisition;
3. **Delay**: when the acquisition process is slowed down due to the burden of acquiring two languages.

In addition, the two grammars may not interact at all, in which case a bilingual’s grammatical development would resemble that of two monolinguals. This is referred to as *autonomous development.*

Before we consider some classic examples of cross-linguistic interaction, some clarification of terminology is called for. We prefer to characterize “delay” as the opposite of acceleration, that is, a certain property emerges in the grammar later than would be the norm in monolingual acquisition, rather than using the definition of Paradis and Genesee (1996), which according to Tamburelli, Sanoudaki, Jones, and Sowinska (2015), is an outdated interpretation of delay. Rather than the term "delay", Fabiano-Smith and Goldstein (2010b) recommend the term "deceleration" since the former may have pejorative connotations suggesting impairment; however, for the purposes of this article, we maintain the original term "delay". The use of "transfer" here refers specifically to the presence of a non-native sound or structure in one of the bilingual's languages which comes from its presence in the other bilingual's language. It should not be confused with a more general employment of "transfer" which is used synonymously with cross-linguistic interaction (e.g., positive and negative transfer).

Examples of cross-linguistic interaction

The examples of cross-linguistic interaction presented below stem from studies conducted at the Research Centre for Multilingualism in Hamburg Germany. Four simultaneous bilinguals were recorded from the onset of word production through to six years (Project B3/E3). Three other bilingual children were tested from word production through to two to three years of age (Project PEDSES). The bilinguals were children of Spanish-speaking mothers and German-speaking fathers. Each parent followed the "une personne, une langue" rule by addressing the child in his/her respective language. In addition, four monolingual Spanish children were recorded in Madrid, Spain, and five monolingual German children were recorded in Hamburg, Germany, from the onset of words through to about three years (Project PAIDUS; see Lleó, 2012, for a more detailed description of the monolingual and bilingual corpora). All children were audio-recorded in their homes, while interacting with a parent and an experimenter. Sessions were phonetically transcribed by native speakers of the respective languages and words were extracted from the sessions depending upon the phonological properties under analysis.

Acceleration

An example of acceleration comes from a study by Lleó et al. (2003) on the acquisition of syllable-final consonants or codas. An important phonological difference between German and Spanish is in the area of syllable structure. Spanish has less complex syllable structure than German. Spanish rhymes consist of a single vowel (e.g., yo [jo] "I"), a diphthong (e.g., ley [leɪ] "law") or a vowel plus consonant (e.g., sol [sol] "sun"). Only a restricted set of consonants appear in coda position, namely coronals such as /n/, /r/, /l/, /s/, /ð/, /θ/ (Harris, 1983). Complex codas do occur but they are rare. In contrast, the German rhyme allows many more possibilities. It consists minimally of two positions: a long vowel or diphthong (e.g., Tee [teː] "tea", Frau [fʁaʊ] "woman" or "wife") or a vowel with a consonant (e.g., Ball [bal] "ball"). There are no restrictions on consonants in coda position. They may be labial, coronal, or dorsal. The German rhyme may consist of more than two positions: a long vowel followed by a consonant (e.g., Hahn [haːn] "cock"), a short vowel followed by two consonants (e.g., Mund [mʊnt] "mouth"), or a long vowel followed by two consonants (e.g., Mond [moːnt] "moon"). Frequency data reveal that Spanish has 27% syllables with codas compared to German which has 67% (Meinhold & Stock, 1980). In sum, codas are more frequent in German and they are more complex.

Lleó et al. (2003) posited that two types of interaction effects may be observed in bilingual German-Spanish children acquiring codas. There may be acceleration of codas in Spanish due to their high frequency in German or delay of codas in German due to their low frequency in Spanish. They examined the structural presence of codas in the productions of three monolingual Spanish, three monolingual German, and five bilingual children (2 children from Project B3/E3; 3 children from Project PEDSES) from word onset to 2;4 years (1 child was only tested through to 1;9). Figures 1 and 2, adapted from Lleó et al.'s (2003) study, present the findings on coda production for the monolingual and bilingual children respectively. Figure 1 shows that coda production was higher in German than in Spanish monolingual children at all time points, reaching over 90% at the last time point in German compared to 30 to 40% in Spanish. Figure 2 shows that the bilingual children started to produce relatively high percentages of codas in Spanish as of 1;9 years. Coda production remained always higher in German than in Spanish but, importantly, coda production in Spanish was higher in the bilinguals than in the monolinguals. Thus, of the two possible interaction effects predicted by Lleó et al. (2003), only one was found, namely, acceleration of codas in Spanish. Lleó at al (2003) hypothesize that the high frequency of codas in German influenced the production of them in Spanish.

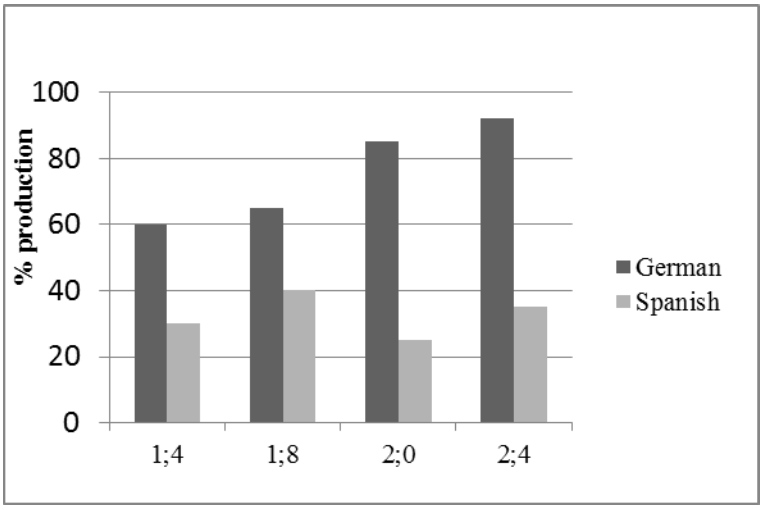


Figure 1. Percentages of codas produced by German and Spanish monolingual children (adapted from Lleó et al., 2003)

Delay

An example of delay comes from a study by Kehoe (2002) on the acquisition of vowel length in German-Spanish bilinguals. The German vowel system is more complex than the Spanish one. It not only has more vowels but it also has a phonological opposition that does not exist in Spanish, vowel length, which is characterized by both quantity (phonetic length) and quality (formant frequency) differences between long and short vowels. In contrast, Spanish has a classic five vowel system. Kehoe (2002) predicted that bilingual children might show a delay in their acquisition of vowel length. Her rationale was that vowel length is a marked phenomenon which requires a certain amount of positive evidence. In the bilingual situation, there is a dilution of this evidence leading to a possible delay in acquisition.

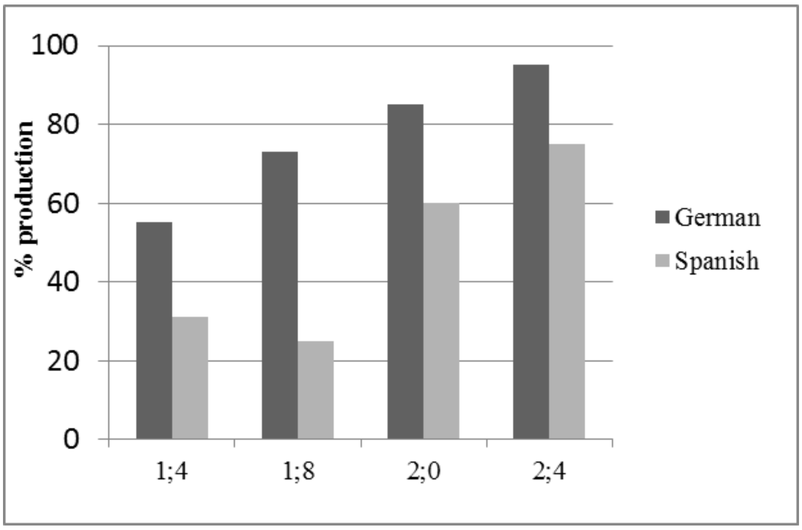


Figure 2. Percentages of codas produced by bilingual children in German and Spanish (adapted from Lleó et al., 2003)

Kehoe (2002) conducted both acoustic and transcriptional analyses of the word productions of three monolingual German and three bilingual German-Spanish children at two time periods: 1;10 - 2;0 and 2;3 - 2;6. The monolingual children produced long vowels significantly longer than short vowels in monosyllables and disyllables at both time periods. In contrast, the bilingual children did not produce long vowels significantly longer than short vowels in monosyllables and only some of the time in disyllables. Furthermore, the magnitude of the duration difference between long and short vowels was reduced compared to the one produced by the monolinguals (average ratio was 1.3 for bilinguals vs.

1.9 for monolinguals). This pattern is evident in Figure 3, which displays the duration values of long and short vowels in disyllables at 2;3 - 2;6 for the monolingual and bilingual children.

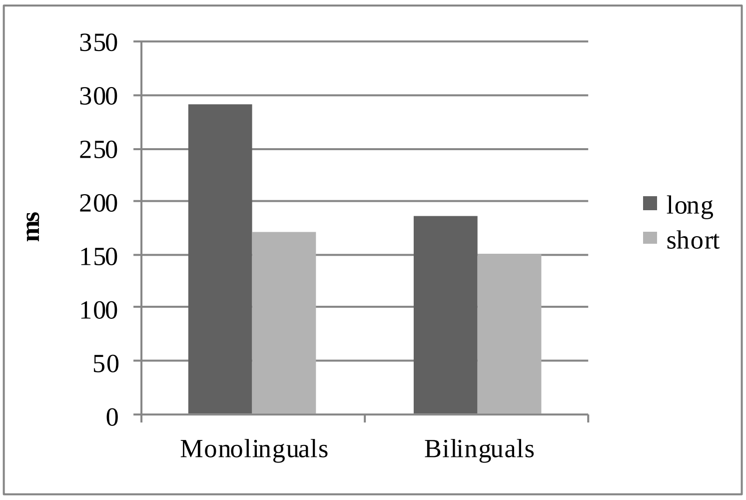


Figure 3. Mean duration values (ms) for target long and short vowels in disyllables spoken by monolingual and bilingual German-speaking children (adapted from Kehoe, 2002)

The transcriptional analyses supported the acoustic analyses in showing that bilingual children experienced more difficulty producing long and short vowels. Whereas monolinguals achieved accuracy rates of 80 to 90% for target long and short vowels at 2;3 to 2;6 in disyllables, the bilinguals achieved accuracy rates of only 50 to 70%. In sum, the findings confirmed Kehoe's (2002) predictions: bilingual children experienced difficulty acquiring the marked system of German vowels. Importantly, when Kehoe (2002) examined the children's acquisition of the five vowel system in Spanish, differences between monolinguals and bilinguals were not observed.

Transfer

The final example, that of Transfer, stems from a study of VOT by Kehoe et al. (2004). Both German and Spanish have voiced /b, d, g/ and voiceless /p, t, k/ stops but the phonetic basis underlying the voicing distinction is different in the two languages. In German, the opposition is between long and short lag whereas, while in Spanish it is between short lag and lead voicing.

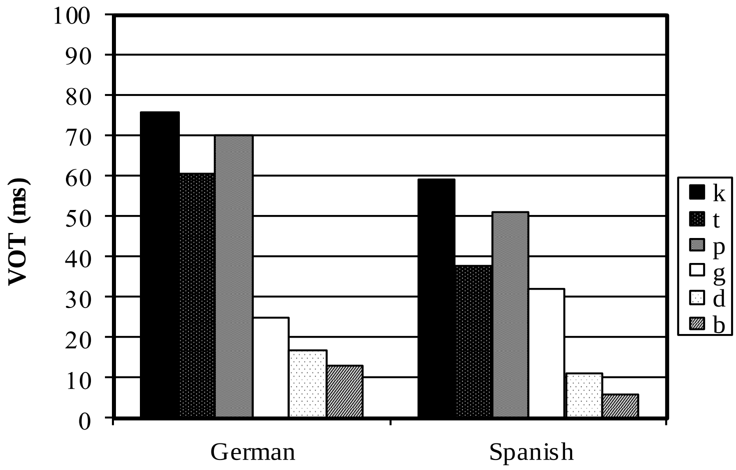


Figure 4. Mean VOT values of voiced and voiceless stops of Nils, a Spanish-German bilingual child at age 2;3 to 2;6 (adapted from Kehoe et al., 2004).

Kehoe et al. (2004) examined the acquisition of VOT in four bilingual children, aged 2;0 to 3;0 years. The bilinguals displayed several different patterns of acquisition; however, for the purposes of this section we concentrate on Nils, who at 2;3 to 2;6 produced not only his German voiceless stops in the long lag region but his Spanish ones were also produced with high VOT values (mean = 50 ms) (see Figure 4). Kehoe et al. (2004) interpreted this pattern as transfer of long lag voicing from German into Spanish. One of the possible reasons for this transfer was that Nils was becoming dominant in German due to his participation in a German kindergarten.

In sum, findings on the same group of German-Spanish bilingual children revealed different patterns of cross-linguistic interaction: acceleration of codas in Spanish, delay of the vowel length distinction in German, and transfer of long lag voicing into Spanish.

Seeking generalizations across studies

Having examined some classic examples of cross-linguistic interaction in the Hamburg data, we widen the literature base to look more closely at cross-linguistic interaction in other studies. We are interested in determining whether they have obtained similar results to Lleó, Kehoe and colleagues. We consider two points of comparisons: a) similar contact situations; and b) similar predictive factors.

Similar contact situations

By "similar contact situation" we refer to the situation in which the linguistic properties of the bilingual's two languages are similar to those under examination in another study. For example, a similar contact situation to Lleó et al.'s (2003) study on codas would be a situation in which a language with a high frequency of codas comes into contact with a language with a low frequency of codas. In the following sections, we examine studies on coda acquisition and VOT.

Acquisition of Codas

An analogous study to Lleó et al. (2003) would be that of Keffala, Barlow, & Rose (submitted) which examined syllable structure acquisition in Spanish-English bilinguals; English, like German, is a language which has many closed syllables. Gildersleeve-Neumann et al. (2008) also examined Spanish-English bilinguals, although their study is limited by the fact that only the English of the bilinguals was examined. Nevertheless, we include it in the current list to see if their results were consistent with those of Lleó et al. (2003) at least for one of the languages. Apart from these two studies, we found few other studies, which have examined coda production in children who are acquiring languages which are characterized by high and low frequencies of codas. There are several other studies which have measured coda production in bilinguals but they are not included because the percentages of coda production were too high to allow a good differentiation between the bilingual's two languages (e.g., Goldstein & Washington, 2001). Thus, we considered also complexity or the distributional features of the codas. As mentioned above, codas are more restricted in Spanish, being only coronal, whereas in German (also English) they are more varied, allowing all places of articulation. German (and English) may also contain coda clusters whereas they are infrequent in Spanish. Using a criterion of restricted/low complexity versus less restricted/high complexity codas, studies by Almeida et al. (2012) on Portuguese-French, and Ezeizabarrena and Alegria (2015) on Basque-Spanish can be included. Portuguese has more restricted use of codas, analogous to Spanish. In contrast, French and Basque allow more segmental diversity in their codas, analogous to German and English. In terms of frequency, however, closed syllables are not frequent in any of these languages. Almeida et al. (2012) focus on word-medial codas whereas the discussion of the other studies concern word-final codas. Table 1 lists the above-mentioned studies.

Table 1. Studies examining cross-linguistic interaction in coda production

Investigators	Languages	Results
Keffala et al. (submitted)	**Spanish – low frequency/restricted**	**Acceleration in Spanish**
	English- high frequency/ unrestricted	
Gildersleeve-Neumann et al. (2008)	**Spanish – low frequency/restricted**	**Delay in English**
	English- high frequency/ unrestricted	**(Spanish not tested)**
Almeida et al. (2012) (word-medial codas)	**Portuguese – restricted**	**Delay in French**
	French - unrestricted	
Ezeizabarrena & Alegria (2015)	**Spanish - restricted**	**No difference**
	Basque - unrestricted	

This comparison does not take into consideration differences in methodology. Rather, we are concerned with whether a similar contact situation: high frequency/unrestricted codas versus low frequency/restricted codas leads to similar outcomes in terms of cross-linguistic interaction across studies. The findings appear to be equivocal. The most similar study to Lleó et al. (2003), namely Keffala et al. (submitted), did find support for acceleration: bilinguals were more accurate than monolinguals in their production of codas (both in terms of structural and segmental accuracy) in Spanish. They also observed reduced coda production in the English of the bilinguals in comparison to the monolinguals; however, the differences were not significant. In contrast, Gildersleeve-Neumann et al (2008) did find significantly reduced coda production in the English of their bilinguals in comparison to the monolinguals, suggesting a delay effect. As mentioned, the results of the Spanish of their bilinguals are unavailable. The other two studies in which frequency played a lesser role did not find an acceleration effect. Almeida et al. (2012) found delay in the language with the less restricted use of codas (i.e. French) similar to Gildersleeve-Neumann et al.'s (2008) findings for English. In contrast, Ezeizabarrena and Alegria (2015) found that their subject produced more codas in Basque than in Spanish, which led them to conservatively interpret their findings in terms of language-specific development, meaning that there was no interaction between the two languages. In sum, three different interaction effects (acceleration, delay, and no differences) were documented in the current contact situation in which codas were examined.

VOT

Table 2 lists studies which have examined VOT in children acquiring languages with a long lag-short lag distinction and a short lag-lead distinction. We focus here on the development of long and short lag stops and not on lead voicing as many sources suggest that it is acquired late (Allen, 1985; Khattab, 2000; Macken & Barton, 1980). Table 2 makes clear that several different patterns of interaction have been documented in bilingual children acquiring languages with both long lag and lead voicing. Indeed, in the study by Kehoe et al. (2004), three different patterns amongst four children were observed. Two children exhibited a delay in the acquisition of long lag stops; one child acquired long lag stops similarly to monolinguals and one child displayed transfer of long lag stops into Spanish. These patterns have also been observed in the other studies on VOT. Fabiano-Smith and Bunta (2012) report VOT values in the short lag region for English /p/ by their bilingual Spanish-English three-year-olds, values lower than the monolingual English-speaking children. This result is consistent with delay. In contrast, Deuchar and Clark's (1996) study of a Spanish-English child suggests similar findings to that of monolinguals. Their subject, Manuela, produced voiced stops as short lag and voiceless stops as long lag in English, and made a type of contrast in the short lag region for Spanish. Khattab's (2000) study of VOT acquisition in slightly older Arabic-English children (ages 5, 7, and 10 years) also revealed few differences that could be directly related to bilingualism.

The bilingual children produced VOT values for short and long lag stops in English that were similar to those of monolinguals. Johnson and Wilson (2002) report findings consistent with transfer of long lag voicing from English into Japanese. They found that their English-Japanese bilingual children produced voiceless stops with long lag values in English but they also did so in Japanese instead of producing them in the short lag region. A similar pattern was reported for older French-English bilinguals (ages 6-, 8-, & 10 years) by Watson (1990). The bilingual French-English children produced voiceless stops with similar VOTs in both languages.

In conclusion, a review of studies on the acquisition of VOT in early and slightly older child bilinguals reveals varied findings. This is particularly apparent in the Hamburg study in which different patterns were observed across children despite similar experimental conditions.

Table 2. Studies examining cross-linguistic interaction in VOT

Investigators	**Languages**	**Results**
Kehoe et al. (2004)	**German**	**1. No differences; 2. Delay of long lag; 3. Transfer of long lag**
	Spanish	
Deuchar & Clark (1996)	**English**	**No difference**
	Spanish	
Fabiano-Smith & Bunta (2012)	**English**	**Delay of long lag**
	Spanish	
Khattab (2000)	**English**	**No difference**
	Arabic	
Johnson & Wilson (2002)	**English**	**Transfer of long lag**
	Japanese	
Watson (1990)	**English**	**Transfer of long lag**
	French	

Examining predictive factors

Given that an analysis of two contact situations related to coda presence and VOT were characterized by variable patterns amongst bilinguals, the current section takes a different perspective. We now seek generalizations by examining predictive factors which have been used to explain cross-linguistic interaction. We consider four main factors: frequency, complexity/markedness, structural ambiguity, and dominance. The first three are considered language-internal factors and the last, a language-external factor; although we acknowledge that frequency may arguably be placed as a language-external factor.

Frequency

"Frequency" here refers to the high or low presence of a segment or a phonological structure in a given language. As Lleó and Cortes (2013) note, frequency is a "gradual notion"; it is difficult to make a clear distinction between what is frequent and infrequent. Nevertheless, using phoneme or syllable-type counts, some kind of frequency grouping can be made. Using these measures, numerous studies in monolingual acquisition show that frequency is important in accounting for order of acquisition within a language: frequent structures are acquired before less frequent ones (Kirk & Demuth, 2003; Levelt, Schiller, & Levelt, 1999/2000; Stites, Demuth, & Kirk, 2004; Zamuner, Gerken, & Hammond, 2005). Frequency is also important in explaining order of acquisition cross-linguistically: a frequent phoneme or structure in one language is mastered earlier than the same less frequent phoneme or structure in another language (Lleó et al., 2003; Pye, Ingram, & List, 1987; So & Dodd, 1995).

How does "frequency" work in the bilingual situation? In fact, we know very little. To illustrate this, we wish the reader to imagine a table which contains those studies which have examined cross-linguistic interaction due to frequency effects. The frequency of a phonetic property in language A would be shown along the X-dimension, and the frequency of the same phonetic property in language B would be shown on the Y-dimension. To simplify matters, frequency would be divided up into discrete categories: high, moderate, low, and absent, the latter meaning that the phonetic property is not present. For the moment, whether a phonological property is present or absent is subsumed under frequency, but a finer distinction between these two factors may be necessary (see Lleó and Cortés, 2013, who distinguishes between frequency and "additive", the latter being concerned with whether the phonological property is present in the two languages of the bilingual). The intersection of "Language A - high frequency" and "Language B - low frequency" could be filled in with the studies of Lleó et al. (2003), Keffela et al. (submitted), and Gildersleeve-Neumann et al. (2008) which dealt with the situation of high frequency codas in German, respectively English, and low frequency codas in Spanish. Lleó et al. (2003) and Keffela et al. (submitted) reported acceleration of codas in Spanish whereas Gildersleeve-Neumann et al. (2008) reported results consistent with delay of codas in English. Unfortunately, we are unable to fill in many other squares in this table due to the lack of pertinent data. That is not to say that other studies on bilingual acquisition have not made reference to frequency. For example, Fabiano-Smith and Goldstein (2010b) found that bilingual Spanish-English children produced sounds that were shared amongst Spanish and English more accurately than sounds that were not shared. They then examined whether this effect could be due to the frequency of the phonemes rather than due to cross-linguistic interaction per se. They found that frequency did not predict the accuracy of shared sounds, suggesting that other factors were at play.

More recently, Tamburelli et al. (2015) found that frequency was not the decisive factor in explaining the acceleration effect that they observed in the acquisition of /s/ + obstruent clusters in the English of Polish-English bilinguals (aged 7 to 9 years). Despite the fact that /s/ + obstruent clusters were more frequent in Polish than in English in both word-initial and word-medial position (twice as frequent according to corpora based on the most frequent words in Polish and English), it was only in word-initial position that bilinguals performed significantly better than monolinguals in terms of cluster accuracy. The authors explained the differential effects between word-initial and word-medial position in terms of complexity rather than frequency (see further discussion under complexity).

The mechanism underlying the frequency effect in bilingual acquisition is also not well understood. Is there a pooling of the input such that a high frequency phenomenon in one language, when combined with a low frequency phenomenon in the other language, becomes moderately frequent across both languages? Is it then the moderate presence of this phenomenon in the overall input which is responsible for its faster acquisition rate? Or does the high frequency of a phonetic structure in one language leads to its faster acquisition in that language? The target structure is then transferred to the other language by mechanisms of cross-linguistic interaction, related to enhanced phonological representations and motor-speech practice. Applying MacWhinney's (2005) Unified Competition Theory to bilingual phonological acquisition, several authors suggest the second possibility is the most likely (Gildersleeve-Neumann & Wright, 2010; Goldstein & Bunta, 2012). According to these authors, structures and phonological properties common across languages will lead to frequent, strong, and reliable cues, which will result in the bilingual using knowledge in one language to aid acquisition in the other language. They use the term "positive transfer" to describe this process, whereby "transfer" is employed in the more general sense of "cross-linguistic interaction".

In reality, these two possibilities, pooling of the input or "positive transfer", may be difficult to distinguish; however, there is some evidence that the second is more likely. Kehoe and Lleó (2003) examined the order of onset and coda cluster acquisition in German and Spanish monolinguals and bilinguals. German has both onset and coda clusters whereas Spanish has only onsets clusters (and very infrequent coda clusters). They found that coda clusters were still acquired before onset clusters in the German of the bilinguals, similar to the monolingual situation, despite the fact that onset clusters were more frequent overall in the pooled input of German and Spanish than coda clusters (18.6% vs. 11.4%; see Table 4 in Kehoe & Lleó, 2003, based on Delattre & Olsen, 1969). In other words, Kehoe and Lleó (2003) did not find any evidence that the overall frequency of syllable types in the pooled input influenced the order of syllable type acquisition, although it had been shown to do so in the monolingual situation (Levelt et al., 1999/2000).

Complexity/markedness

Complexity, like frequency is a relative term which is difficult to define. In general, a phonetic /phonological property that contains more elements (e.g., features), more structure, or is more difficult to produce is more complex than a phonetic/phonological property that contains fewer elements, less structure, or is less difficult to produce. Complexity may be used synonymously with markedness or it may be distinguished from it. Lleó and Cortes (2013) use “markedness” in the sense of Jakobson (1968/1941): an unmarked entity is acquired earlier than a marked one and is more common in the languages of the world, whereas, complexity, in their approach, refers specifically to phonetic phenomena which involve allophony and allomorphy. In this section, we will use complexity and markedness similarly, although we acknowledge that a more detailed examination of this term should differentiate the two.

Several studies in bilingual phonological acquisition indicate that complex phonological entities may be associated with delay. Goldstein and Washington (2001) observed that complex sound classes such as fricatives and liquids were more likely to be delayed in Spanish-English bilingual children relative to monolingual controls than less complex sound classes such as stops and nasals. As mentioned previously, Kehoe (2002) found phonological vowel length to be delayed in the German of Spanish-German bilinguals but not the less complex system of Spanish vowels. Lleó (2002) found that bilingual Spanish-German children were delayed in the acquisition of unfooted syllables in Spanish; unfooted syllables being considered as marked. Studies focusing on rhythm have shown that stress-timed rhythm poses more difficulty for bilingual children than syllable-timed rhythm (Bunta & Ingram, 2007; Mok, 2011; 2013). In all of the above examples, a possible explanation is that the acquisition of a complex entity requires frequent exposure which is reduced in the bilingual situation, and consequently, delay ensues.

More recently, a number of studies have pointed to an alternative manifestation of complexity in bilingual acquisition, that of acceleration. Keffala et al. (submitted) argue that onset clusters in Spanish and English are complex, but in different ways. English onset clusters are structurally complex, containing both two- and three-element clusters (s-adjunct clusters), whereas Spanish onset clusters are segmentally complex, containing more marked or smaller sonority differences. They hypothesize that bilinguals will display accelerated acquisition of onset clusters because they are exposed to two different types of complexity across their languages. Their results confirmed their predictions: bilingual children were more accurate in onset cluster structure (i.e. cluster structure regardless of segmental accuracy) in Spanish and in onset cluster segments (i.e. segmentally accurate clusters) in both languages compared to monolingual controls. Similar findings have been reported by Tamburelli et al. (2015) for English onset cluster acquisition in Polish-English bilinguals. The authors argued that bilinguals displayed acceleration in word-initial /s/ + obstruent clusters because they were exposed to increased complexity (small sonority differences) in Polish clusters. Importantly no effect was found in word-medial position because structurally they are coda-onset sequences and not onset clusters. Mayr et al. (2015) also report results consistent with acceleration of word-final English clusters in their Welsh-English bilinguals. In comparison with norms on English monolingual children’s acquisition of word-final clusters (Templin, 1957), their bilinguals performed better.

In contrast to the above findings, not all studies on cluster acquisition have reported acceleration. Gildersleeve-Neumann et al. (2008) observed delay in the acquisition of English clusters by their bilinguals. This finding is more consistent with the earlier presented results in which marked structures may be associated with delay. For example, Gildersleeve-Neumann et al. (2008) observed delay in the acquisition of codas and clusters in their bilinguals but not in the acquisition of less marked aspects of phonology such as vowels.

In the area of phonological rhythm, Schmidt and Post (2015) document acceleration in one aspect of rhythm, namely consonantal variability. Their two-year-old English-Spanish bilinguals displayed lower consonantal variability in English, closer to the adult targets, than the English monolinguals. The authors propose that bilinguals have a developmental advantage over monolinguals because they are exposed to more varied structures, including different types of consonant intervals, across their two languages.

The second group of findings, on the acquisition of clusters (Keffela et al., submitted; Mayr et al.; Tamburelli et al., 2015) and on rhythm (Schmidt & Post, 2015) appeals to a different literature on complexity, one that shows that exposure to linguistic complexity may promote acquisition (Dinnsen & Elbert, 1984; Gierut, 1999, 2001, 2007). For example, Gierut (1999) found that children with phonological disorders when treated with a linguistically complex target (e.g., cluster with a smaller sonority difference) evidenced greater learning than those children treated with a less complex target. In the area of bilingual acquisition, the rationale appears to be that exposure to a complex structure in one language may lead to enhanced development of this same structure in the other language, although in the case of rhythm, Schmidt and Post (2015) found the acceleration effect in the structurally more complex language.

One important question is whether these two sets of findings, one showing delay and the other showing acceleration can be reconciled. In Tamburelli et al.'s (2015) study on onset clusters, the bilingual was acquiring similar types of structures, /s/ + obstruent clusters, albeit their differing complexity, in both languages. Thus, there may be a reinforcement effect. In studies which have shown delayed acquisition of segments, for example, liquids, the bilingual child was acquiring a complex target that occurred in only one of the languages (e.g., Spanish trill). Similarly, the spirant-stop alternation in Spanish (Goldstein & Washington, 2001; Lleó & Rakow, 2005) or the phonological vowel length distinction in German (Kehoe, 2002) are rules or structures that appear in one of the languages of the bilingual. The generalization could be that complex targets which appear in one of the bilingual's languages are associated with delay. This generalization cannot account for all the findings, however. Keffela et al.'s (submitted) and Gildersleeve et al.'s (2008) studies of clusters included structures which were present across both languages of the bilingual (e.g.,, obstruent + liquid clusters) and structures which were found only in English (e.g., /s/ + consonant clusters) and yet one study reported acceleration and one delay. In sum, additional research is needed to understand the outcome of "complexity" in bilingual phonological acquisition.

Structural ambiguity

Structural ambiguity has been posited as a possible explanation of cross-linguistic interaction (Döpke, 1999; Müller & Hulk, 2000; Paradis, 2000, 2001). The idea behind it is that cross-linguistic interaction occurs when there is partial structural overlap between two languages and one of the languages offers multiple options for analysis. This proposal has been mainly investigated in morpho-syntax and has led in recent years to the Interface Hypothesis, namely, that bilingual children have most difficulty acquiring language at interfaces: internal (syntax, semantic) or external (syntax, discourse) (Sorace, 2005; Sorace & Serratrice, 2009; White, 2011).

Recourse to structural ambiguity has not been frequent in studies on bilingual phonological acquisition, although two studies are noteworthy. Paradis (2000, 2001) accounted for the different truncation patterns of monolingual English and bilingual English (-French) children by appealing to structural ambiguity. The monolingual and bilingual children did not differ on the truncation patterns of SWSW words but they did on the truncation patterns of WSWS words. Paradis (2000, 2001) argued that WSWS words resemble the French WWWS words and, thus, are structurally ambiguous, whereas the SWSW words do not resemble any French pattern and thus are not structurally ambiguous and consequently are not affected.

Almeida et al. (2012) accounted for both delay in the acquisition of codas in French, and acceleration in the acquisition of onset clusters in Portuguese in terms of structural ambiguity. For matters of space, we will consider the case of onset clusters in Portuguese and French only. Vowel deletion is frequent in spoken Portuguese leading to many surface examples of consonant sequences. The Portuguese child, thus, has the difficult task of distinguishing between those sequences which are true clusters and those which are not true clusters but are due to surface elision. According to Almeida (2011), the fact that the same onset clusters exist in French and Portuguese (e.g., obstruent + liquid clusters) helps the bilingual child to identify true onset clusters in Portuguese and, thus, aids the child to acquire them more quickly.

Apart from these two studies, there are few other references to structural ambiguity in phonological studies of early bilinguals, possibly because phonology does not lend itself to ambiguity in the same way that morpho-syntax does. The extension of structural ambiguity, the Interface Hypothesis, has also not been extensively studied in early bilingual phonological research with the exception of Lleó (In press). Lleó (In press) explored interfaces in phonetics and phonology to determine if they are vulnerable domains for bilinguals as has been claimed in syntax. She identified numerous interfaces in the phonetic/phonology domains (e.g., segments and lexemes, segments and prosodic position, phonemes and phones, phonology and morphology, prosody and syntax, prosody and semantics, and prosody and pragmatics), but ultimately, discarded them as sources of difficulty for bilinguals. There are interface phenomena which do not appear vulnerable to cross-linguistic interaction (e.g., association of meaning and intonation in Prosody-Pragmatic interface) and other non-interface phenomena which are (see White, 2011).

Dominance

The language that the child hears and uses the most is typically his dominant language. Many studies label their bilingual children as being dominant in one language or the other (Ball, Müller, & Munro, 2001; Law & So, 2006; Mayr et al., 2015). These studies generally show that the dominant language of a bilingual is associated with faster phonological acquisition. For example, Law & So (2006) observed that Cantonese dominant bilinguals have faster Cantonese phonological development than Putonghua dominant bilinguals and vice versa. Mayr et al. (2015) found that Welsh-dominant bilinguals were more accurate on Welsh clusters than English-dominant bilinguals. However, another interesting finding from these studies has been that when complexity and dominance are pitted together, complexity often wins out. In the case of Cantonese and Putonghua, bilingual children acquired the segmental aspects of Cantonese phonology faster than the more complex Putonghua, regardless of their dominance. Similarly, bilingual children, regardless of whether they were English- or Welsh-dominant, produced English word-final clusters earlier than Welsh ones, presumably because they were less complex.

Dominance as an explanatory factor has been used to account for transfer in several studies on bilingual acquisition. Keshavarz and Ingram (2002) observed that their Farsi-English bilingual produced English two-syllable words with final stress (word stress in Farsi is predominantly on the final syllable) during a period in which he was dominant in Farsi. He then acquired the English stress patterns when he became dominant in English. As mentioned, Kehoe et al. (2004) observed transfer of long lag voicing from German into Spanish in Nils, a child who was becoming dominant in German. Other authors have also reported transfer of long lag voicing in the situation in which the ambient language also contains long lag stops, suggesting effects of dominance (Johnson & Wilson, 2002). One curious finding in these transfer cases is that complex structures may be transferred (e.g., iambic stress, long lag voicing), suggesting that the relationship between dominance and complexity is in itself complex.

While several studies make reference to dominance as an important explanatory factor, other studies do not find it useful. Almeida et al. (2012) point out that dominance cannot explain the patterns of their Portuguese-French bilingual, who displayed both acceleration of clusters and delay of codas during the same time period. It would be impossible for their child to be dominant in both languages at the same time. Rose and Champdoizeau (2007) document clear differences between the acoustic manifestations of stress in a French-English bilingual, consistent with the language-specific stress patterns, despite the fact that the child was dominant in English.

Seeking generalizations across studies: Summary

Our survey of four main predictive factors in bilingual phonological acquisition has yielded variable and unsatisfying results as did the previous analysis based on contact situations. Frequency has not been well studied in early bilingualism. Complexity appears to be better understood; however, there are opposing findings on complexity which remain to be reconciled. Structural ambiguity has not been fully explored in the phonological domain and may have limited application in this area. Dominance appears to matter but perhaps less so than other factors such as complexity.

One possible reason as to why few generalizations could be gleaned when examining similar contact situations or predictive factors is that we have ignored important methodological differences between studies. Some studies are based on small groups of bilinguals (very rarely, large groups) whereas others are single case studies. Some use word naming tasks to elicit productions whereas other are based on longitudinal naturalistic recordings. These factors may potentially lead to different outcomes in terms of cross-linguistic interaction.

Using a homogeneous data set: Lleó and Cortés' (2013) model

An alternative approach to seeking generalizations across studies is to seek generalizations within a single data-base, looking at several different phonological phenomena at the same time. This has been the approach of Lleó and Cortés (2013) who have developed a model of cross-linguistic interaction based on the Hamburg data. They have brought together findings on coda acquisition, vowel length in German and VOT (see above) as well as on place assimilation of nasals in Spanish, and on the spirantization rule in Spanish. In their model, four factors (frequency, unmarkedness, additive, and uniformity) account for the varied effects of acceleration, delay, and transfer; transfer being viewed as a more negative form of cross-linguistic interaction than delay. Their model is formalized as a table in which phonological phenomena receive plus or negative signs depending upon their values on the four factors: "Frequency" is marked positively if a phonological phenomenon has a high frequency (e.g., coda consonants in German). "Unmarkedness" receives a "+" if the phenomenon is unmarked. "Additive" is marked positively if the phenomenon under consideration occurs in both languages of the bilingual and "uniformity" is marked positively if the phonological phenomenon is not characterized by allophony or allomorphy. Tabulating across various phenomena, Lleó and Cortés (2013) conclude that frequency plays a greater role in accounting for the findings than markedness which in turn plays a greater role than uniformity and additiveness. Overall, they find the most positive effects manifest when a phenomenon occurs in both languages but is more frequent in one of the languages (e.g., codas in German and Spanish) and the most negative effects manifest when a phenomenon occurs in only one of the languages and violates uniformity (e.g., spirants in Spanish). In-between effects manifest when a phenomenon is low frequency in both languages (e.g., unfooted syllables in Spanish) or occurs in only one of the languages but doesn't violate uniformity (e.g., vowel length in German).

In sum, we may conclude that, at this stage in early bilingualism research, generalizations can only be obtained by looking at a homogenous set of data as Lleó and Cortés (2013) have done. Additional research is necessary before generalizations can be made across a broad range of studies. Before this research is conducted, however, it is worth reviewing what changes could be made before attempting new studies. We turn to the second section of this paper.

Prospective

Critical look at the research

Critical look at methodology

One striking limitation of research on bilingual phonological production is the lack of studies with large numbers of children. The review article by Hambly et al. (2013) illustrates the fact that single or multiple case studies (29 out of the 66 studies) predominate in this field, particularly in linguistic style studies which are pertinent to the topic of cross-linguistic interaction. Case studies are still very informative in the field of early bilingualism but they increase the risk that effects interpreted as cross-linguistic interaction may be due to individual differences. One clear reason for the low "n"s is the time consuming nature of bilingual research which often includes analyses of the two languages of the bilingual child as well as analyses of the two sets of monolingual controls, which multiples the number of subjects in a single study by four fold. Phonetic and phonological production research is also intrinsically time consuming when transcription and acoustic analyses are involved.

A second limitation of research on bilingualism is the fact that data on monolingual controls is not extensive for many languages of the world. Given that the current definition of cross-linguistic interaction refers to differences between bilinguals and monolinguals, a decision as to whether cross-linguistic interaction takes place can only be made when a study includes monolingual controls or when it refers to a solid base of monolingual data. Unfortunately, this cannot always be done.

There are other methodological limitations, including the fact that few are experimental, few are longitudinal, and few include extensive information on the language background of the children; however, the focus of this section is on another major limitation of current research, namely, the lack of a research model in early bilingualism.

Critical look at research model

One clear handicap of current approaches to early bilingual phonology is the lack of a research model specially designed to account for cross-linguistic interaction. In the overview to this article, we situated early bilingualism between the larger fields of first and second language acquisition. We noted that early bilingualism shares characteristics in common with both fields; however, it has not necessarily integrated these characteristics into a coherent model. If anything, early bilingualism has leant more towards the field of second language acquisition, but, even here, it has not adopted all aspects that could be useful to it. In the following sections, we expand upon findings in First and Second Language Acquisition research which could be incorporated into current approaches to early bilingualism.

Leaning on Second Language Acquisition

One of the most well-known models in second language research, the Speech Learning Model (SLM) of Flege (1995) has motivated some research in early bilingualism (Fabiano-Smith & Goldstein, 2010b; Gildersleeve-Neumann & Wright, 2010) although, strictly speaking, it is a model which is intended for children acquiring a second language after the age of five to six years (Flege, 1997), and, thus, does not concern the majority of studies presented here. Certain central tenets of the SLM are, nevertheless, implicit in speech production research in early bilingualism, in particular, that the two linguistic systems share a common phonological space in which bidirectional interaction occurs. It also provides a taxonomy for classifying the relationship between L1 and L2 sounds, which is sadly absent in early bilingualism.

In the SLM, an L2 sound is new (i.e. differs acoustically and perceptually from the L1 sound), identical (i.e. there is no significant acoustic difference between the L1 and L2 sound), or similar (i.e. there are significant and audible differences between the L1 and L2 sound, but both sounds can be transcribed with the same IPA symbol) with respect to the L1 system. It is the similar (but not identical) sounds which create the most difficulty for second language learners. Their acquisition often leads to two processes: perceptual assimilation or dissimilation. The acquisition of a similar L2 sound may result in equivalence classification which prevents a new L2 category from being formed and the categories of the L1 and L2 are merged together. “Merging” phenomena have been reported in acquisition of VOT, whereby second language learners produce stops in their L1 and L2 with similar VOT values (Flege, 1987). The acquisition of a similar L2 sound may lead to an opposite phenomenon in which the two categories move away from each other to avoid crowding the phonetic space. These “deflecting” phenomena have also been reported in the acquisition of VOT. Mack (1990) reports excessively high VOT values for English long lag stops (e.g., 108 ms) in a 10 year-old French-English bilingual child. Since the child produced French voiceless stops also in the long lag region (e.g., 66 ms), the long VOTs in English allowed the child to maintain phonetic contrast between his L1 and L2 systems. We return to a discussion of “merging” and “deflecting” phenomena below.

The main research model applied to speech production in early bilingualism has not been the SLM of Flege (1995) but Paradis and Genesee’s (1996) model, which, in reality, is not a model but a framework for describing patterns of cross-linguistic interaction. Furthermore, we may ask whether the three patterns outlined by Paradis and Genesee (1996): transfer, acceleration, and delay, are sufficient for accounting for all the findings in early bilingualism. We believe the answer is no (see Lleó, 2015, for similar views).

a. Transfer

First we consider "transfer". There are numerous studies which indicate that "transfer" is not frequent in young bilinguals. Goldstein and colleagues report percentages of below 1% for transfer effects in Spanish-English bilinguals (Goldstein & Washington, 2001; Goldstein, et al., 2005). Specifically, Fabiano and Goldstein (2005) report seven incidences out of a total of 1269 possible occasions in three bilingual Spanish-English children, aged 5;0 to 7;0. The seven instances included Spanish to English influence (e.g., /v/ →[ß] and English to Spanish influence (e.g., /s/ →[θ], /o/ →[ə], /r/ →[ɹ]). In another study, Fabiano-Smith and Goldstein (2010b) documented low percentages of transfer in two of their eight bilingual Spanish-English subjects. Examples included de-aspiration of stop consonants in English, which is consistent with difficulties in VOT. Apart from the findings with Spanish-English bilinguals, studies with other groups of bilinguals find equally low percentages of transfer. Law and So (2005) report few examples of transfer in Cantonese-Putonghua bilinguals; Salameh, Nettelbladt, and Nolan (2003) documented only occasional examples in their study of Swedish-Arabic bilinguals; and Mayr et al. (2015) also note few cases in word-final clusters in Welsh-English bilinguals.

Nevertheless, transfer has been documented on more than an occasional basis in certain phonetic and phonological domains, for example, VOT and /r/ acquisition. Not only did Kehoe et al. (2004) observe transfer of long lag voicing into Spanish from German but they also observed transfer of lead voicing from Spanish into German in one of the bilinguals in their study. Fabiano-Smith and Barlow (2010) report transfer of Spanish /r/ (i.e. [ɾ]) in five out of eight bilingual children's English phonetic inventories. We also observed high percentages of /r/ transfer in an unpublished case study of a trilingual child, acquiring French, Spanish, and Italian, in Geneva, Switzerland (Di Vietri, 2012). The bilingual child, aged two to three years, produced the uvular /r/ in French but also produced the uvular /r/ exclusively in Italian, and most of the time in Spanish as well. The /r/ in Italian is an alveolar trill and in Spanish, it is an alveolar trill and tap.

Given that transfer is not a frequent phenomenon in early bilingual speech production, we may wonder whether the examples of transfer in VOT and /r/ mentioned above are true examples of transfer or could be better captured under alternative classifications of cross-linguistic interaction (see below) or could be due to language external factors. VOT is a fragile phonetic domain in bilingual and second language acquisition and patterns of interaction may be specific to the temporal aspects of this domain. Lleó (2015) considers Nils' transfer of long lag voicing as an example of "fusion" (similar to "merging" described below). She advocates widening the definition of "transfer" to not only includes segments but rules and structures. In the case of /r/ transfer in the trilingual presented above, language-external factors related to dominance (the child was growing up in Geneva) or non-native input (the child was exposed to non-native speakers of Italian or Spanish) might explain the apparent high transfer rate in this child.

While "transfer" is an interaction pattern that is not frequent in early bilingualism, there are other interaction patterns which do occur and which are currently not included in Paradis and Genesee's (1996) list. We refer here to the "merging" and "deflecting" patterns, discussed above. They cannot be easily slotted into the categories of "acceleration" and "delay".

b. Merging patterns

Kehoe and Lleó (in press) document merging patterns in the vowel reduction processes of German-Spanish bilingual children. They measured the ratios of stressed-to-unstressed syllable durations in German and Spanish. Differences between ratios of stressed-to-unstressed syllable durations were significant in the German versus Spanish monolingual children. The ratios were greater than 1.0 in the German monolingual children (1.4 in phrase-final and 1.7 in phrase-medial position), but close to 1.0 in the Spanish monolingual children (.87 in phrase-final and 1.0 in phrase-medial position), reflecting the syllable-timed nature of Spanish and the fact that German unstressed syllables are schwa syllables, leading to greater acoustic distances between stress and unstress. Differences between the ratios of stressed-to-unstressed syllables durations in German and Spanish were not significant in the bilingual children. The ratios were reduced in the case of German (1.32 in phrase-final and 1.24 in phrase-medial) and similar or slightly increased in Spanish (.98 phrase-final and 1.04 in phrase-medial), resulting in a less extreme contrast between the German and Spanish systems, that is, in a merging effect. Other examples of merging patterns have been reported in acoustic measures of rhythm in Spanish-English bilinguals (Kehoe & Lleó, 2005; Kehoe, Lleó, & Rakow, 2011), intrinsic vowel duration in English-German bilinguals (Whitworth, 2000) and in VOT (Watson, 1990; Lleó, 2015b).

c. Deflecting patterns

Dodane and Bijeljic-Babic (in press) present findings on acquisition of the acoustic correlates of stress in French-English bilinguals which are consistent with deflecting patterns. They measured duration, F_0, and intensity in the disyllabic productions of French- and English-monolinguals and in French-English bilingual children, aged 4;0 years. Here we concentrate on their findings on duration and F_0 in the French productions of the monolingual and bilingual children. The monolingual children displayed a substantial final lengthening effect (the ratio of syllable 2 to syllable 1 was 1.72) and they produced no pitch accent on the first syllable, consistent with the language-specific stress pattern of French. The bilingual children also displayed a substantial final lengthening effect, significantly larger than the one made by the monolingual children (the ratio of syllable 2 to syllable 1 was 2.29) and larger than the one made in their English words (ratio of syllable 2 to syllable 1 was 1.3), although they also produced a pitch accent on the first syllable suggesting influence of English stress. What interests us in this section is the exaggeration of the final lengthening effect in French which allowed the bilinguals to make a maximal contrast between their two language systems.

Another example of a deflecting pattern has recently been reported by Yang, Fox, and Jacewicz (2015) in the area of vowel development. They documented acquisition of English vowels in a Mandarin child, aged 3;7. Although the child initially perceptually assimilated English vowels to L1 categories, after two months, the child drastically reduced English vowel space, producing his English back vowels as central variants. This abrupt restructuring of the vowel system was interpreted by the authors as an attempt by the child to maximize the contrast between his two languages; the reduction in L2 space was also accompanied by a mild expansion of the L1 Mandarin vowel space. Subsequent development of the L2 system was characterized by gradual enlargement of the reduced vowel space.

A different example of a "deflecting" effect comes from Paradis's (1996) reanalysis of Hildegard's German and English productions (Leopold, 1949/1971). She found that Hildegard used different prosodic structures in English compared to German (e.g., more disyllables in English; more closed monosyllables in German; reduplicated forms in English but not in German) despite the fact that the prosodic structures in the German and English words that she selected were very similar. In other words, Hildegard appeared to impose contrasts in her output which were not present in her input forms.

In sum, merging and deflecting effects do not only belong to the realm of second language acquisition but may be observed in young simultaneous and sequential bilinguals as well. It may be the case that merging patterns do not just reflect perceptual assimilation as in older bilinguals but may be the by-product of increased variability or of common speech-motor constraints which lead to similar output effects across both languages of the young bilingual. Deflecting patterns show consciousness of the need for contrast, either maximizing an existing contrast or imposing a new contrast, which appears already well developed in the young bilingual.

The previous section has highlighted the importance of looking towards second language acquisition for ways to expand upon the approach currently used in early bilingualism. We have recommended enlarging the set of interaction patterns, while at the same time reducing the importance of "transfer" as a possible manifestation of cross-linguistic interaction. We now turn to how approaches in first language acquisition may complement a possible research model.

Leaning on First Language Acquisition

Three themes central to first language acquisition will be discussed here: 1. speech motor control; 2. the developing lexicon; and 3. individual differences. The latter theme is linked to the first two. Phonological acquisition may be conceptualized as having two basic components: 1) a biologically based component associated with the development of speech-motor capacities; and 2) a cognitive-linguistic component associated with learning the phonological system of the ambient language (Stoel-Gammon, 2011). While the cognitive-linguistic component may vary between a bilingual child's two languages, the speech motor skills which underlie the two phonological systems may not necessarily vary. In actual fact, we know very little about the speech motor development of young bilinguals. There have been few studies that have separately studied phonological and motor speech aspects; however, studies on bilinguals with motor speech involvement (e.g., childhood apraxia of speech) show similar patterns across languages on motor-based tasks such as diadochokinetic tasks or token variability suggesting that aspects of motor control are language-neutral (Preston & Seki, 2011).

The fact that a bilingual child's phonological systems share a common speech-motor base as well as have many segments and phonological structures in common may explain findings in the literature which show that a bilingual child's two phonologies approximate each other at certain levels. These findings include the presence of common templates (Vihman, 2002, 2015; Kehoe, 2015); merging patterns (see above), and similar orders of acquisition of phonological structures (Almeida et al., 2012; Lleó, 2015b). That is, we may expect between-language correlations in the phonological domain of an order not typically observed in other domains of language such as semantics or morpho-syntax (Kehoe, 2011, 2015).

Acquiring a phonological system involves acquiring words. In emergent approaches to phonological development, learning phonological categories and acquiring words goes hand in hand (Edwards, Munson, & Beckman, 2011). Numerous studies have focused on the relationship between phonological and lexical development in monolingual children (see Stoel-Gammon, 2011 for a review). For example, studies on late talkers in English, Cypriot Greek, Italian and French consistently show that children with small vocabularies have less developed phonologies than children with large vocabularies (Bortolini & Leonard, 2000; Kehoe, Chaplin, Mudry, & Friend, 2015; Petinou & Okalidou, 2006; Paul & Jennings, 1992). These findings support the presence of a bidirectional relationship between phonology and the lexicon. Only recently have researchers started to examine the relationship between lexical and phonological development in bilingual children (Kehoe, 2011; 2015; Vihman, 2002; 2015), although many aspects of this relationship remain unstudied. For example, we do not know whether a language-specific or a combined vocabulary score is most predictive of a bilingual child's phonological ability in each language.

Developing articulatory and lexical abilities are important components of phonological acquisition which need to be controlled since they may lead to considerable individual differences amongst children. Interestingly, they are factors that are rarely controlled in studies on bilingual phonological acquisition, with some rare exceptions. Scarpino (2011) examined what factors were the best predictors of phonological production (as measured by PCC and whole word proximity) in a large group of Spanish-English children (n=199), aged 3;0 to 6;4 years. Important to the current discussion is that she found that language-specific vocabulary scores and the phonological accuracy in the other language were highly predictive of phonological proficiency in both the English and Spanish of the bilingual children. Other important factors were language use and age. Scarpino (2011) hypothesized that the predictive nature of the other language's phonological score reflected general developmental factors (articulatory maturation) and individual aptitude. Again, these findings emphasize the importance of general articulatory abilities and lexical development in understanding bilingual phonological development.

To illustrate the role that speech production differences may play in cross-linguistic interaction, we cite recent work by Kartushina and Frauenfelder (2013, 2014) in the area of second language acquisition. They examined the influence of individual L1 vowel production data on the perception and production of L2 vowels. To determine individual production ability, they measured the acoustic distance from the L1 and L2 vowel categories and the compactness of the L1 vowel category. We will focus on their findings on vowel compactness. They found that speakers who exhibit high variability in the way they produce the same L1 sound (high within-category variability), that is, sloppy speakers, experienced greater difficulty perceiving and producing L2 vowels. Conversely, speakers with compact L1 spaces established more precise L2 vowel categories. Translating these findings to early bilingualism, we may posit that young bilinguals, who display greater token variability, may be the ones to show more difficulties establishing native-like phonological categories and structures in both of their languages. In other words, our sloppy bilingual speakers may display the greatest degree of cross-linguistic interaction.

In sum, research in first language acquisition highlights the importance of considering the child's developing speech motor and lexical abilities as crucial factors to be controlled in studies on early bilingualism. These factors may lead to considerable individual differences among children which is inherent in the current state of research in the field. Individual production capacities may also be predictive of cross-linguistic interaction, as suggested by the research of Kartushina and Frauenfelder (2013, 2014).

Critical look: Summary

To conclude this section, we have argued that a new research model is sorely needed in early bilingualism. This is not only our conclusions but was one of the main recommendations of the review article by Hambly et al. (2013):

> "Developing models of cross-linguistic bilingual speech acquisition that take into account age of acquisition, length and type of L2 exposure, language proficiency, the development and capacity of perceptual and cognitive systems, individual variation alongside other phonological areas, such as rhythm and intonation is an enormous challenge but will assist practitioners as they assess the speech of bilingual children." (p.13)

This model needs to integrate important components from first and second language acquisition as well as develop its own unique aspects. We recommend that it expands upon existing manifestations of cross-linguistic interaction and considers the role of developing articulatory capacities and lexical abilities as a way of controlling for individual differences amongst bilingual children.

New Perspectives

One of the main conclusions of this review article is that it is time to examine more systematically what factors underlie cross-linguistic interaction in early bilingualism. Several researchers have already embarked upon this enterprise: Lleó and Cortés (2013) have developed a model of cross-linguistic interaction based on the Hamburg data; Keffela et al. (submitted) and Tamburelli et al. (2015) have examined the separate effects of frequency and complexity on syllable structure acquisition. In the remaining parts of the paper, we consider three other new perspectives in early bilingualism which should also enrich further research attempts: new methodologies, tracking cross-linguistic interaction over time, and the role of the input.

New methodologies

The majority of studies in early bilingual phonology are based on naturalistic language sampling procedures or word elicitation tasks. Many times, these methods bear witness to high production proficiency on the part of the young bilingual already by 3;0 years of age. The presence of high performance scores leads to ceiling effects which reduce the possibility of documenting important cross-linguistic effects.

Non-word repetition or sentence repetition tasks are useful procedures with bilingual populations as they reduce the effects of linguistic knowledge on test performance. Thus, they are very effective in diagnosing specific language impairment (Ferré, dos Santos, & Almeida, 2015; Thordardottir & Brandeker, 2013). They may also serve as alternative approaches for examining cross-linguistic interaction, since they augment the difficulty of the task thereby allowing more effective discrimination of monolingual-bilingual differences (Tamburelli et al. 2015; Marecka, Wrembel, Zembrzuski, & Otwinowska-Kasztelanic, 2015). Marecka et al. (2015) asked phonetically trained raters to assess the sentence repetitions of Polish monolingual and bilingual children aged 5;9 years. The bilinguals were simultaneous Polish-English bilinguals growing up in the United Kingdom. The raters were required to assess cross-linguistic interaction on several dimensions including vowel errors and stress change. Importantly the raters were blind as to whether they were assessing a bilingual or a monolingual speaker. The bilinguals made more speech errors than monolinguals and were judged as displaying more cross-linguistic interaction than their monolingual counterparts. These results underscore the usefulness of sentence repetition in combination with a rating scale as an alternative way of measuring cross-linguistic interaction.

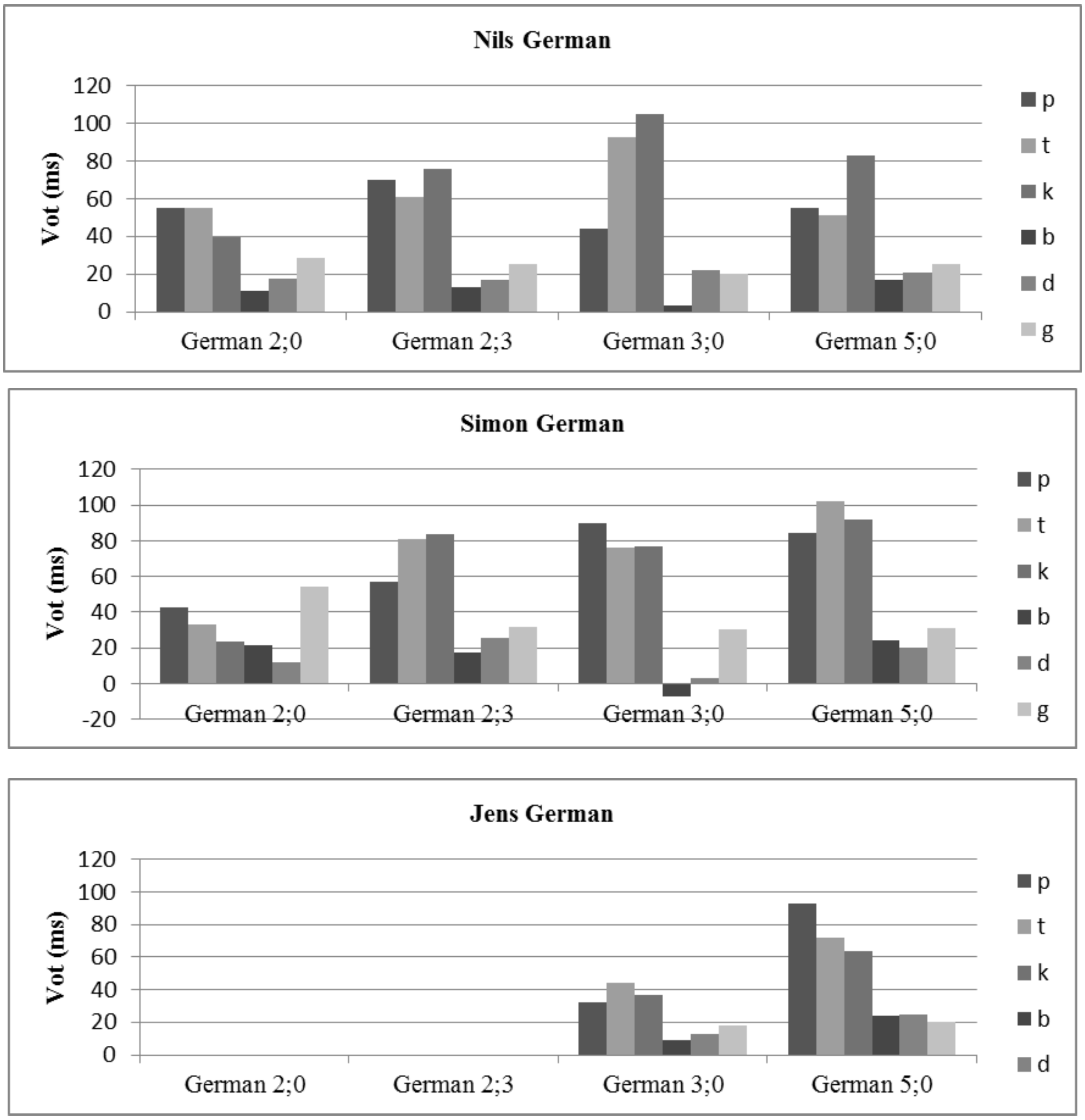


Figure 5. VOTs for target German voiced and voiceless stops in German-Spanish bilingual children at ages 2;0, 2;3, 3;0 and 5;0 (adapted from Rakow & Lleó, 2008).

Tracking Cross-linguistic Interaction

How momentary or persistent is cross-linguistic interaction? The current study has focused at one end of the life continuum, at two to five years of age. There is a growing collection of studies which have looked at the other end of the continuum, that is, at adult bilinguals who have acquired their languages from birth or very early on. These studies show that adult bilinguals may still differ from monolingual speakers in certain aspects of phonetics and phonology (e., VOT, foreign accent), particularly in their non-dominant language (Benmamoun, Montrul, & Polinsky, 2013; Kupisch, Barton, Lein, Schröder, Stangen, & Stöhr, 2014; Lein, Kupisch, & van de Weijer, 2015). There have been very few studies, however, that have looked in-between these two extremes and have tracked cross-linguistic interaction from childhood over an extended period of time. Such studies would provide useful information on the dynamic nature of phonological systems and their vulnerability.

One of the few studies that has been conducted stems from the Hamburg project. It tracked VOT development in three German-Spanish bilinguals, at 2;0, 2;3, 3;0 and 5;0 years (Rakow & Lleó, 2008). Two of the bilinguals were part of the group studied in Kehoe et al. (2004). At 5;0 years, all three bilinguals showed native-like distinctions between long and short lag stops in German but only one of the bilinguals, Simon, showed native-like distinctions between short lag and lead voicing in Spanish. These results are shown in Figures 5 and 6 for the German and Spanish voicing systems respectively. Of the two children who didn't acquire lead voicing, one of the children (Jens) made a distinction between voiced and voiceless stops in the short-lag region for two out of three places of articulation in Spanish. The other child (Nils) produced target voiceless stops with long lag values (> 40 ms) in Spanish. Why some children continue to display cross-linguistic interaction after a period and others do not remains mysterious; however, it should be noted that Simon, who achieved native-like voicing patterns in Spanish at 5;0 years, was not delayed in VOT acquisition (short vs. long lag distinction) at age 2;3. The other two children who did not acquire lead voicing in Spanish were already delayed in VOT acquisition at the earlier age ranges (for Nils at 2;0 and 2;3; for Jens at 3;0). This may suggest that factors contributing to cross-linguistic interaction are active over an extended period of time and may even be child-specific.

The role of the input

This review paper has mainly considered the role of language-internal factors (e.g., frequency, complexity, structural ambiguity) in influencing bilingual phonology. We have considered one language-external factor, dominance, which relates to quantitative aspects of the input. We have not considered qualitative aspects of the input such as whether the input the child is receiving is provided by native speakers or by multiple speakers. Indeed the source of language input for many bilingual children may be a single person or a small group of people.

Mayr and Montanari (2015) recently compared the role of input setting on the VOT acquisition of two trilingual English-, Italian-, and Spanish-speaking children, aged 6- and 8-years, growing up in the United States. The children's input in English was from their native English father and from the native speakers in the surrounding environment. The children's VOT values in English were essentially native-like and not susceptible to cross-linguistic interaction. The children's input in Italian was from their native Italian-speaking mother but also from English-accented input by the English dominant children in the Italian school the two children were attending. The children's VOTs in Italian were not native-like. They produced their target velar stops with high VOT values and they did not produce target voiced stops in Italian with consistent lead voicing. The children's input in Spanish came from a single person, the Spanish-speaking nanny. Surprisingly, the children's VOT values in Spanish were similar to the adult model and appeared unaffected by cross-linguistic interaction. The authors hypothesize that input from a single source may be conducive to phonological acquisition and may limit the effects of cross-linguistic interaction. This is an intriguing hypothesis and warrants further investigation. In sum, qualitative aspects of the input (e.g., presence of non-native input, single vs. multiple speakers) need to be given more attention in studies on early bilingualism.

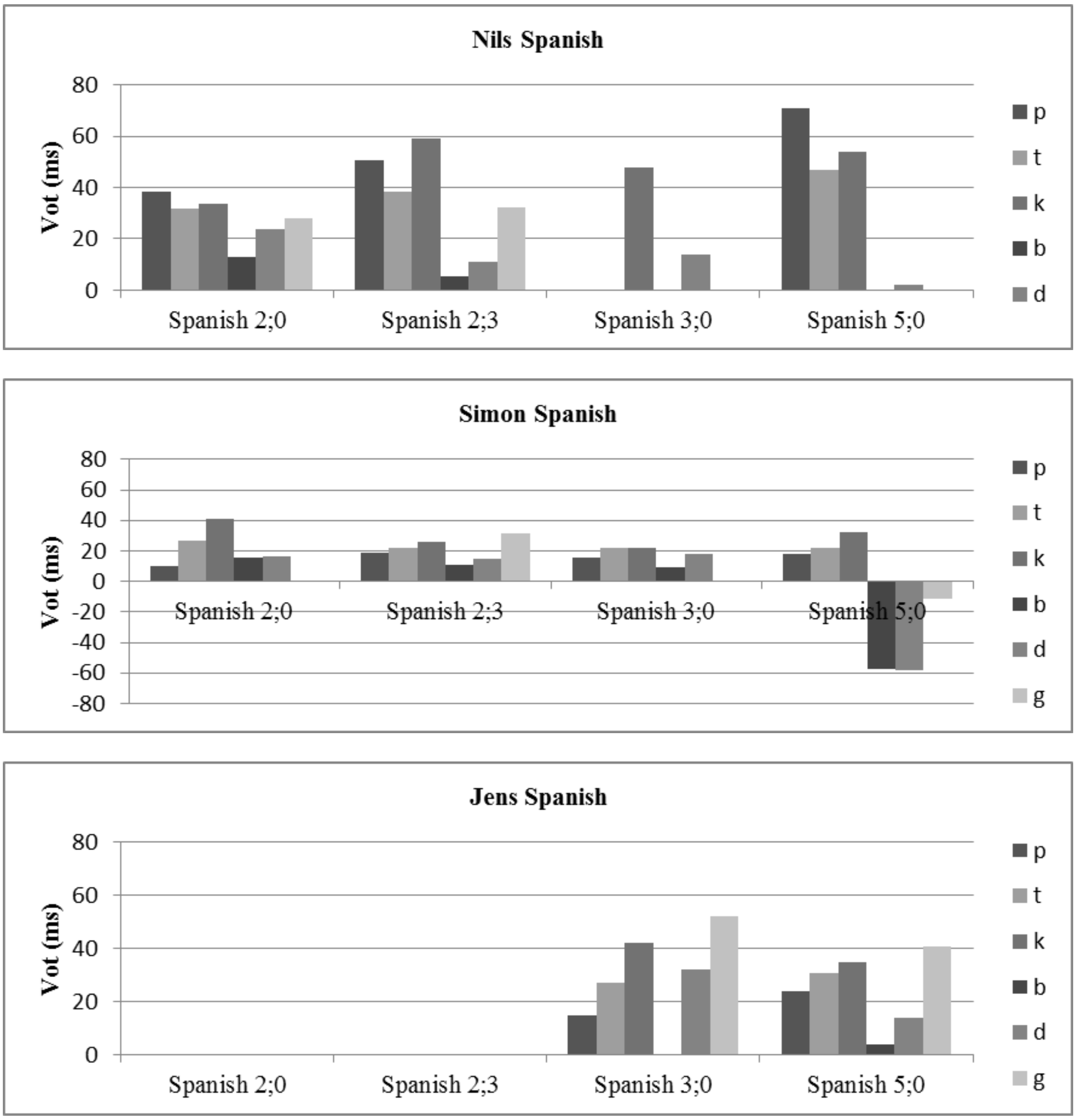


Figure 6. VOTs for target Spanish voiced and voiceless stops in German-Spanish bilingual children at ages 2;0, 2;3, 3;0 and 5;0 (adapted from Rakow & Lleó, 2008).

There are many possible new avenues of research in early bilingualism. We have focused on three areas: using new methodologies as a way of avoiding ceiling effects in phonological production research; tracking cross-linguistic interaction over time as a way of determining vulnerable domains in cross-linguistic interaction, and focusing on qualitative aspects of the input, a theme that may also give valuable information on what language-external factors influence the occurrence of cross-linguistic interaction.

Conclusion

Our review of the literature on cross-linguistic interaction leads to modest conclusions. Cross-linguistic interaction has been documented in multitudes of studies but it is still not well-understood. Our attempts to seek generalizations across similar contact situations or by examining common predictive factors have not yielded many salient outcomes. We observed that a similar contact situation (e.g., high frequency of codas in one language vs. low frequency in the other language) may result in diverse findings; the same predictive factor (e.g., high vs. low complexity) may also be subject to differing outcomes or interpretations. Methodological limitations related to small subject numbers and lack of extensive monolingual data contributes to the lack of generalizability of the findings. In addition, current research is hindered by an insufficient research model which does not take into account all the possible interaction patterns that may occur, or the importance of speech-motor or lexical factors in influencing early bilingual phonological development. We believe these factors should be incorporated into new models of early bilingualism as they may help to explain the striking individual differences that are evident in research on young bilinguals. Other goals of future research should be to track cross-linguistic interaction over time so as to understand which domains are particularly affected by bilingual input. It is hoped that future research which systematically examines predictive factors and which utilizes new methodologies may provide a clearer understanding of when and why cross-linguistic interaction occurs. It may be the case that seeking generalizations in this field will never be easy, however, due to the inherent uniqueness of each bilingual child. That is, the "woods" may remain always difficult to detect.

References

Allen, G. (1985). How the young French child avoids the prevoicing problem for word-initial stops. *Journal of Child Language*, *12*, 37-46.

Almeida, L. (2011). *Acquisition de la structure syllabique en contexte de bilinguisme simultané portugais-français*. Unpublished doctoral dissertation, Universidade of Lisboa.

Almeida, L., Rose, Y., & Freitas, J. (2012). Prosodic influence in bilingual phonological development: Evidence from a Portuguese-French first language learner. In A. Biller, E. Chung, & A. Kimball (eds*), Proceedings of the 36th Annual Boston University Conference on Language Development* (pp. 42-52). Somerville: Cascadilla Press.

Ball, M., Müller, N., & Munro, S. (2001). The acquisition of the rhotic consonants by Welsh-English bilingual children. *International Journal of Bilingualism*, *5*, 71-86.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, *39*, 129-181.

Bortolini, U., & Leonard, L. B. (2000). Phonology and children with specific language impairment : status of structural constraints in two languages. *Journal of Communication Disorders*, *33*, 131-150.

Bunta, F., & Ingram, D. (2007) The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4- and 5-year-old children. *Journal of Speech, Language, and Hearing*, *50*, 999-1014.

Bunta, F., Davidovich, I., & Ingram, D. (2006). The relationship between the phonological complexity of a bilingual child's words and those of the target languages. *International Journal of Bilingualism*, *10*, 71-88.

Bunta, F., Fabiano-Smith, L., Goldstein, B., & Ingram, D. (2009). Phonological whole-word measures in three-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics and Phonetics, 23*, 156-175.

De Houwer, A. (1990). *The acquisition of two languages from birth: A case study*. Cambridge: Cambridge University Press.

Delattre, P., & Olsen, C. (1969). Syllable features and phonic impression in English, German, French, and Spanish. *Lingua*, *22*, 160-175.

Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics*, *24,* 351-365.

Dinnsen, D., & Elbert, M. (1984). On the relationship between phonology and learning. *Journal of* Phonetics, 12, 28-37.

Di Vietri, V. (2012). *Le développement phonologique chez in enfant trilingue français, espagnol et italien: aspects segmentaux. Etude de cas*. Unpublished Masters thesis, University of Geneva, Geneva, Switzerland.

Dodane, C., & Bijeljic-Babic, R. (in press). Cross-language influences in the productions of bilingual children: Separation or interaction. In M. Yavaş, M. Kehoe, & W. Cardoso (eds.), *Romance-Germanic bilingual phonology*. South Yorkshire: Equinox Publishing Ltd.

Döpke, S. (1999). Cross-linguistic influences on the placement of negation and modal particles in simultaneous bilingualism. *Language Sciences*, *21*,143-175.

Edwards, J., Munson, B., & Beckman, M. E. (2011). Lexicon–phonology relationships and dynamics of early language development–a commentary on Stoel-Gammon's 'Relationships between lexical and phonological development in young children'. *Journal of Child Language*, *38*, 35-40.

Ezeizabarrena, M.-J., & Alegria, A. (2015). Early coda production in bilingual Spanish and Basque. In T. Judy & S. Perpiñán (eds.), *The acquisition of Spanish in understudied language pairings*. Amsterdam: John Benjamins.
Fabiano-Smith, L., & Barlow, J. (2010). Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism*, *13,* 1-17.
Fabiano-Smith, L., & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics & Phonetics*, *26*, 148-163.
Fabiano, L., & Goldstein, B. (2005). Phonological cross-linguistic effects in bilingual Spanish-English children. *Journal of Multilingual Communication Disorders*, *3*, 56-63.
Fabiano-Smith, L., & Goldstein, B. (2010a). Early, middle, and late developing sounds in monolingual and bilingual children: An exploratory study. *American Journal of Speech-Language Pathology, 19,* 1-12.
Fabiano-Smith, L., & Goldstein, B. (2010b). Phonological acquisition in bilingual Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research*, *53*, 1-19.
Ferré, S., dos Santos, C. & de Almeida, L. (2014). Potential clinical markers for SLI in bilingual children, In E. Grillo & K. Jepson (eds.). *Proceedings of the 39th Boston University Conference on Language Development* (pp.152-164). Somerville: Cascadilla press.
Flege, J. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, *8*, 162-177.
Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233-273). Timonium, MD: York Press.
Flege, J. (1997). English vowel productions by Dutch talkers: more evidence for the "similar" vs. "new" distinction. In A. James & J. Leather (eds.), Second-language speech (pp. 11-52). Berlin: Mouton de Gruyter.
Genesee, F. (1989). Early bilingual development: One language or two? *Journal of Child Language*, *6*, 161-179.
Gierut, J. (1999). Syllable onsets: clusters and adjuncts in acquisition. *Journal of Speech, Language, and Hearing Research*, *42*, 708-726.
Gierut, J. (2001). Complexity in phonological treatment: Clinical factors. *Language, Speech, and Hearing Sciences in Schools*, *32*, 229-241.
Gierut, J. (2007). Phonological complexity and language learnability. *American Journal of Speech-Language Pathology*, *16*, 6-17.
Gildersleeve-Neumann, C., Kester, E., Davis, B., & Peña, E. (2008). English speech sound development in preschool-aged children from bilingual English-Spanish environments. *Language, Speech, and Hearing Services in Schools*, *39*, 314-328.
Gildersleeve-Neumann, C., & Wright, K. (2010). English speech acquisition in 3- to 5-year-old children learning Russian and English. *Language, Speech, and Hearing Services in Schools*, *41*, 429-444.
Goldstein, B., & Bunta, F. (2012). Positive and negative transfer in the phonological systems of bilingual speakers. *International Journal of Bilingualism*, *16*, 388-401.
Goldstein, B., & Washington, P. (2001). An initial investigation of phonological patterns in typically developing 4-year-old Spanish English bilingual children. *Language, Speech, and Hearing Services in Schools*, *32*, 153-164.
Goldstein, B., Fabiano, L. & Washington, P. (2005). Phonological skills in predominantly English-speaking, predominantly Spanish-speaking, and Spanish-English bilingual children. *Language, Speech, and Hearing Services in Schools*, *36*, 201-218.
Grech, H., & Dodd, B. (2008). Phonological acquisition in Malta: A bilingual language learning context. *International Journal of Bilingualism*, *12*, 155-171.
Hambly, H., Wren, Y., McLeod, S., & Roulstone, S. (2013). The influence of bilingualism on speech production: A systematic review. *International Journal of Language and Communication Disorders*, *48*, 1-24.
Harris, J. (1983). *Syllable structure and stress in Spanish. A nonlinear analysis*. Cambridge, Massachusetts: MIT Press.
Holm, A., & Dodd, B. (1999). A longitudinal study of the phonological development of two Cantonese-English bilingual children. *Applied Psycholinguistics*, *20*, 349-376.
Jakobson, R. (1968). *Child language, aphasia, and phonological universals*. The Hague: Mouton (originally published as Kindersprache, aphasie und allgemeine lautgesetze, Uppsala, Sweden, 1941).
Johnson, C., & Lancaster, P. (1998). The development of more than one phonology: A case study of a Norwegian-English bilingual child. *The International Journal of Bilingualism*, *2*, 265-300.
Johnson, C., & Wilson, I.L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *The International Journal of Bilingualism*, *6*, 271-289.
Kartushina, N., & Frauenfelder, U. (2013). On the role of L1 speech production in L2 perception: evidence from Spanish learners of French. *Proceedings of Interspeech, 14,* 2118-2122.

Kartushina, N., & Frauenfelder, U. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology, 5*, Article 1246, 1-16.

Keffela, B., Barlow, J., & Rose, S. (submitted). Interaction in Spanish-English bilingual's acquisition of syllable structure.

Kehoe, M. (2002). Developing vowel systems as a window to bilingual phonology. *International Journal of Bilingualism, 6*, 315-334.

Kehoe, M. (2011). Relationships between lexical and phonological development : A look at bilingual children. *Journal of Child Language*, *38*, 75-81.

Kehoe, M. (2015). Lexical-phonological interactions in bilingual children. *First Language*, *35*, 93-125.

Kehoe, M., & Lleó, C. (2003). The acquisition of syllable types in monolingual and bilingual German and Spanish children. In B. Beachley, A. Brown, & F. Conlin (eds), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 402-413). Somerville, Ma: Cascadilla Press.

Kehoe, M., & Lleó, C. (2005). *The emergence of language specific rhythm in German-Spanish bilingual children.* Arbeiten zur Mehrsprachigkeit: Working Papers in Multilingualism 58. Hamburg: SFB 538.

Kehoe, M., & Lleó, C. (in press). Vowel reduction in German-Spanish bilinguals. In M. Yavas, M. Kehoe, & W. Cardoso (eds.), *Romance-Germanic bilingual phonology*. South Yorkshire: Equinox Publishing Ltd.

Kehoe, M., Chaplin, E., Mudry, P., & Friend, M. (in press). La relation entre le développement du lexique et de la phonologie chez les enfants francophones. *Rééducation Orthophonique*.

Kehoe, M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition, 7*, 71-88.

Kehoe, M., Lleó, C., & Rakow, M. (2011) Speech rhythm in the pronunciation of German and Spanish monolingual and German-Spanish bilingual 3-year-olds. *Linguistische Berichte*, *227*, 323-351.

Keshavarz, M. & Ingram, D. (2002). The early phonological development of a Farsi- English bilingual child. *The International Journal of Bilingualism, 6*, 255-269.

Khattab, G. (2000). VOT in English and Arabic bilingual and monolingual children. In D. Nelson & P. Foulkes (eds.), *Leeds Working Papers in Linguistics and Phonetics*, *8*, 95-122.

Kirk, C., & Demuth, K. (2003). Onset/coda asymmetries in the acquisition of clusters. In B. Beachley, A. Brown, & F. Conlin (eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 437-448). Somerville, MA: Cascadilla Press.

Kupisch, T., Barton, D., Lein, T., Schröder, J., Stangen, I., & Stöhr, A. (2014). Acquisition outcomes across domain in adult heritage speakers of French. *Journal of French Language Studies*, *24*, 347-376.

Law, N., & So, L. (2006). The relationship of phonological development and language dominance in bilingual Cantonese-Putonghua children. *International Journal of Bilingualism*, *10*, 405-428.

Lein, T., Kupisch, T., & van de Weijer, J. (2015).Voice onset time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism*, Early Online, 1367006915589424

Leopold, W. (1949/1971). *Speech development of a bilingual child. Vol I-II*. Evanston, Illinois: Northwestern University Press.

Levelt, C., Schiller, N., & Levelt, W. (1999/2000). The acquisition of syllable types. *Language Acquisition, 8*, 237-264

Lléo, C. (2002). The role of markedness in the acquisition of complex prosodic structures by German-Spanish bilinguals. *The International Journal of Bilingualism, 6*, 291-313.

Lleó, C. (2012). Monolingual and bilingual phonoprosodic corpora of child German and child Spanish. In T. Schmidt & K. Wörner (eds.), *Multilingual corpora and multilingual corpus analysis. Hamburger studies on multilingualism* 14 (pp. 107-122). Amsterdam/Philadelphia: John Benjamins.

Lleó, C. (In press). Acquiring multilingual phonologies (2L1, L2 and L3): Are the difficulties in the interfaces? In S. Fischer & C. Gabriel (eds.), *Grammatical interfaces. Manuals of Romance Linguistics*. De Gruyter.

Lleó, C. (2015). *On the permeability of German-Spanish bilinguals' phonological grammars*. Paper given at the International Symposium of Monolingual and Bilingual Speech (ISMBS), Crete, Greece.

Lleó, C., & Cortés, S. (2013). Modeling the outcome of language contact in the speech of German-Spanish and Catalan-Spanish bilingual children. *International Journal of the Sociology of Language*, *221*, 101–125.

Lleó, C., & Rakow, M. (2005). Markedness effects in voiced stop spirantization in bilingual German-Spanish children. In H. Cohen, K. MaAlister, K. Rolstad & J. MacSwain (eds.), *Proceedings of the 4th International Symposium on Bilingualism* (ISB4) (pp. 1353-1371). CD Rom: Cascadilla Press.

Lleó, C., Kuchenbrandt, I., Kehoe, M., & Trujillo, C. (2003). Syllable final consonants in Spanish and German monolingual and bilingual acquisition. In N. Müller (ed.), *(Non)Vulnerable Domains in Bilingualism* (pp. 191-220). Amsterdam: John Benjamins.

MacLeod, A., Laukys, K., & Rvachew, S. (2011). The impact of bilingual language learning on whole-word complexity and segmental accuracy among children aged 18 and 36 months. *International Journal of Speech-Language Pathology*, *13*, 490-499.

Mack, M. (1990). Phonetic transfer in a French-English bilingual child. In P. Nelde (Ed.), *Language attitides and language conflict* (pp. 117-124). Bonn, Germany: Dümmler.

Macken, M., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonological study of word-initial stop consonants. *Journal of Child Language*, *7*, 433-458.

MacWhinney, B. (2005). A unified model of language acquisition. In J. Kroll & A. De Groot (eds.), *Handbook of bilingualism. Psycholinguistic approaches* (pp. 49-67). Oxford and New York: Oxford University Press.

Mayr, R., Howells, G., & Lewis, R. (2015). Asymmetries in phonological development: the case of word-final cluster acquisition in Welsh-English bilingual children. *Journal of Child Language*, *42*, 146-179.

Mayr, R., & Montanari, S. (2015). Cross-linguistic interaction in trilingual phonological development: the role of the input in the acquisition of the voicing contrast. *Journal of Child Language*, *42*, 1006-1035.

Meinhold, G., & Stock, E. (1980). *Phonologie der deutschen Gegenwartssprache*. Leipzig: Bibliographisches Institut.

Meisel, J. (1989). Early differentiation of languages in bilingual children. In K. Hyltenstam & L. Obler (eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss* (pp. 13-40). Cambridge, UK: Cambridge University Press.

Mok, P. (2011) The acquisition of rhythm by three-year-old bilingual and monolingual children. *Bilingualism: Language and Cognition*, *14*, 458-472.

Mok, P. (2013) Speech rhythm of monolingual and bilingual children at age 2;6: Cantonese amd English. *Bilingualism: Language and Cognition*, *16*, 693703.

Müller, N., & Hulk, A. (2000). Crosslinguistic influence in bilingual children: Object omissions and root infinitives. In S. Howell, S. Fish & T. Keith-Lucas (eds.), *Proceedings of the 24th Annual Boston Conference on Language Development*. Volume 2 (pp. 546-557). Somerville, MA: Cascadilla Press.

Munson, B. (2004). Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean. *Journal of Speech, Language, and Hearing Research*, *47*, 58-69.

Munson, B., Edwards, J., & Beckman, M. E. (2005). Phonological knowledge in typical and atypical speech–sound development. *Topics in Language Disorders*, *25*, 190-206.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, *20*, 1-32.

Paradis, J. (1996). Phonological differentiation in a bilingual child: Hildegard revisited. In A. Stringfellow, D. Cahana-Amitay, E. Hughes & A. Zukowski, (eds.), *Proceedings of the 20th Annual Boston Conference on Language Development* (pp. 528-539).Somerville, MA: Cascadilla Press.

Paradis, J. (2000). Beyond "One System or Two?" Degrees of separation between the languages of French-English bilingual children. In S. Döpke (ed.), *Cross-linguistic structures in simultaneous bilingualism* (pp. 175-200). Amsterdam /Philadelphia: John Benjamins.

Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *International Journal of Bilingualism*, *5*, 19-38.

Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or inderdependent? *Studies in Second Language Acquisition*, *18*, 1-25.

Paul, R., & Jennings, P. (1992). Phonological behavior in toddlers with slow expressive language development. *Journal of Speech, Language, and Hearing Research*, *35*, 99-107.

Petinou, K., & Okalidou, A. (2006). Speech patterns in Cypriot-Greek late talkers. *Applied Psycholinguistics*, *27*, 335-353.

Preston, J., & Seki, A. (2011). Identifying residual speech sound disorders in bilingual children: A Japanese-English case study. *American Journal of Speech-Language Pathology*, *20*, 73-85.

Pye, C., Ingram, D., & List, H. (1987). A comparison of initial consonant acquisition of English and Quiché. In K. Nelson & A. van Kleeck (eds.), *Children's language*. Vol.6 (pp.175-190). Hillsdale, New Jersey: Erlbaum.

Rakow, M., & Lleó, C.(2008). *Bilingual acquisition of VOT across the first years of life and beyond*. Paper given at the Colloquium on Language Acquisition and Change. Across the Lifespan and across Generations. University of Hamburg, Germany.

Rose, Y., & Champdoizeau, C (2008). There is no trochaic bias: Acoustic evidence in favor of the neutral start hypothesis. In A. Gavarro & M. J. Frietas (eds.), *Language Acquisition and Development: Proceedings of GALA 2007* (pp. 359 - 369). Newcastle: Cambridge Scholars Publishing.

Salameh, E-K., Nettelbladt, U., & Norlin, K. (2003). Assessing phonologies in bilingual Swedish-Arabic children with and without language impairment. *Child Language Teaching and Therapy*, *19*, 338-364.

Scarpino, S. (2011). *The effects of language environment and oral language ability on phonological production proficiency in bilingual Spanish-English speaking children*. Unpublished doctoral dissertation, Pennsylvania State University.

Schmidt, E., & Post, B. (2015). Language interaction in the development of speech rhythm in simultaneous bilinguals. In E. Delais-Roussarie, M. Avanzi & S. Herment (eds.), *Prosody and Language in Contact* (pp. 271—291). New York, USA: Springer.

So, L., & Dodd, B. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, *22*, 473-495.

Sorace, A. (2005). Syntactic optionality at interfaces. In L. Cornips & K. Corrigan (eds.), *Syntax and variation: Reconciling the biological and the social* (pp. 46-111). Amsterdam: John Benjamins.

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, *13*, 195-210.

Splendido, F. (2014). *Le développement d'aspects phonético-phonologiques du français chez des enfants bilingues simultanés et successifs. Le VOT et la liaison dans une étude de cas multiples*. Etudes Romandes de Lund 92 (Doctoral dissertation).

Splendido, F. (in press). The initial development of VOT in early successive French-Swedish bilinguals. In M. Yavaş, M. Kehoe, & W. Cardoso (eds.), *Romance-Germanic bilingual phonology*. South Yorkshire: Equinox Publishing Ltd.

Stites, J., Demuth, K., & Kirk, C. (2004). Markedness vs. frequency effects in coda acquisition. In A. Brugos, L. Micciulla & C. Smith (eds.), *Proceedings of the 28th Annual Boston University Conference on Language Development*, Vol. 2 (pp. 565-576). Somerville, Massachusetts: Cascadilla Press.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*, 1-34.

Tamburelli, M., Sanoudaki, E., Jones, G., & Sowinska, M. (2015). Acceleration in the bilingual acquisition of phonological structure: Evidence from Polish-English children. *Bilingualism: Language and Cognition*, *18*, 713-725.

Templin, M. (1957). *Certain language skills in children: their development and interrelationships*. Minneapolis, Minnesota: University of Minnesota Press.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*, 426-445.

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*, 1-16.

Vihman, M. (2002). Getting started without a system: From phonetics to phonology in bilingual development. *The International Journal of Bilingualism*, *6*, 239-254.

Vihman, M. (2015). Prosodic structures and templates in bilingual phonological development. *Bilingualism: Language and Cognition*. http://dx.doi.org/10.1017/S1366728914000790

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of Child Language*, *5*, 311-326.

Watson, I. (1990). Acquiring the voicing contrast in French: A comparative study of monolingual and bilingual children. In J. N. Green & W. Ayres-Bennett (eds.), *Variation and change in French: Essays presented to Rebecca Posner on the occasion of her sixtieth birthday* (pp. 37-60). London: Routledge.

White, L. (2011). Second language acquisition at the interfaces. *Lingua*, *121*, 577-590.

Whitworth, N. (2000). Acquisition of VOT, and vowel length by English-German bilinguals: A pilot study. *Leeds Working Papers in Linguistics and Phonetics*, *8*.

Marecka, M., Wrembel, M., Zembrzuski, D., & Otwinowska-Kasztelanic, A. (2015). *Do early bilinguals speak differently than their monolingual peers? Predictors of phonological performance of Polish-English bilingual children. Proceedings of the International Symposium on Monolingual and Bilingual Speech* (ISMBS 2015) (this volume), Chania, Greece.

Yang, J., Fox, R., & Jacewicz, E. (2015). Vowel development in an emergent Mandarin-English bilingual child: a longitudinal study. *Journal of Child Language*, *42*, 1125-1143.

Zamuner, T., Gerken, L-A., & Hammond, M. (2005). The acquisition of phonology based on input: a closer look at the relation of cross-linguistic and child language data. *Lingua*, *115*, 1403-1426.

Testing hypotheses on frequency effects in first language acquisition - noun plural inflection in Danish children

Laila Kjærbæk, Hans Basbøll
kjarbak@sdu.dk, hba@sdu.dk

University of Southern Denmark

Abstract. On the basis of extensive literature studies, Ambridge, Kidd, Rowland and Theakston (2015) present five theses on frequency effects on language acquisition: i) the *Levels and Kinds Thesis* argues that frequency effects exist at all levels and are of many different kinds (e.g., type and token frequency effects as well as absolute and relative frequency effects); ii) the *Age of Acquisition Thesis* argues that all other things being equal, frequent forms will be acquired before less frequent forms. Since all other things are not equal, this claim does not entail a one-to-one relationship between frequency and age of acquisition; iii) the *Prevent Error Thesis* argues that high-frequency forms prevent (or reduce) errors in contexts in which they are the target; iv) the *Cause Error Thesis* argues that high-frequency forms also cause errors in contexts in which a competing, related lower-frequency form is the target; and v) the *Interaction Thesis* argues that frequency effects will interact with other effects. The acquisition of the Danish noun plural system is particularly interesting in this regard. The reason is that whereas English is characterized by having one default inflectional marker for a grammatical category (e.g., the plural suffix -*s*) and a minor number of exceptions to this default rule, Danish has several competing inflectional markers. Furthermore, there are important interactions between phonology and morphology in the Danish system (Kjærbæk, dePont Christensen, & Basbøll, 2014). In this study we will test the theses in a phonological perspective and explore the impact of phonetics on grammar. This we will do in three types of empirical data from children acquiring Danish as their first language: i) *Naturalistic data* consisting of spontaneous child language input and output from six children between the ages of 0;10-3;11, and their parents; ii) *Semi-naturalistic data* from structured interviews with 80 children between the ages of 3-9 years; iii) *Experimental data* from a picture based elicitation task with 160 children between the ages of 3-10 years. We present a scale with three degrees of transparency of the plural stem and of the plural suffix as well as a scale with six degrees of transparency of the Danish plural markers. We furthermore present a scale with three degrees of productivity. Productivity is here defined as the ability of an inflectional marker to occur on new words. For the plural system this means the ability to add the plural marker (stem change + suffix) to a new noun in order to form a new plural noun. We analyze the relation between acquisition rate and degree of transparency as well as degree of productivity.

Keywords: Danish, first language acquisition, frequency, morphology, noun plural, productivity, transparency

Introduction

A review article by Ambridge, Kidd, Rowland and Theakston (2015) presents evidence for the claim that input frequency effects are pervasive in children's first language acquisition. On the basis of extensive literature studies, they present five theses on frequency effects on language acquisition including considerable empirical support that exists for each of their theses across four domains, namely: 1) single words; 2) inflectional morphology; 3) simple syntactic constructions; and 4) more advanced constructions. The focus of the present study will be on inflectional morphology. Ambridge et al. (2015) argue for a learning mechanism that is frequency sensitive – a claim which is really not that controversial at all, since it is well known that frequency affects language acquisition one way or the other (e.g., Diessel, 2015; Rowe, 2015). Studies of frequency effects are still interesting though, because they reveal something about the learning mechanism and units used in language learning. The interesting question is, however, how frequency interacts with other factors, and this will be the topic of the present study.

Theses on frequency effects

The five theses presented by Ambridge et al. (2015) are:

1) *Levels and Kinds Thesis*. Frequency effects exist at all levels and are of many different kinds (e.g., type and token frequency effects as well as absolute and relative frequency effects).
2) *Age of Acquisition Thesis*. All other things being equal, frequent forms will be acquired before less-frequent forms. Since all other things are not equal, this claim does not entail a one-one relationship between frequency and age of acquisition.
3) *Prevent Error Thesis*. High-frequency forms prevent (or reduce) errors in contexts in which they are the target.
4) *Cause Error Thesis*. High-frequency forms, on the other hand, also cause errors in contexts in which a competing related lower-frequency form is the target.
5) *Interaction Thesis*. Frequency effects will interact with other effects.

We have studied whether Danish children's acquisition of the noun plural category supports these five theses, and we find the last thesis particularly interesting since the patterning of these interactions can bring us new information on the nature of the learning mechanism.

Factors affecting the acquisition of plural morphology

Earlier studies indicate that (among others) the following factors play a role in the acquisition of plural morphology, confer the *Interaction Thesis*:

Frequency

It is well known that frequency plays a role in many domains of human development – also in first language acquisition (e.g., Ambridge et al., 2015; Kjærbæk, 2013, 2015; Kjærbæk, dePont Christensen, & Basbøll, 2014). When we refer to frequency in this article, we mean input frequency, though studies indicate that children show effects of output frequency as well. We refer to both token frequency and type frequency (here defined as number of different lemmas).

Transparency

Transparency concerns the extent to which a plural pattern is parsable into its basic units. In English, for example, the plural form *hands* (*hand+s*) is more transparent than *feet* (e.g., Albirini, 2015). Basbøll, Kjærbæk, and Lambertsen (2011) and Laaha, Kjærbæk, Basbøll, and Dressler (2011) present a typologically relevant three degree gradation of stem changes. Laaha et al. (2011) found a correlation between the degree of transparency and Danish-speaking and German-speaking children's acquisition rate of plural stems – high transparency correlated with early acquisition of the stem change and lower transparency correlated with later acquisition.

Predictability

Predictability concerns the degree to which a specific plural form may be predicted based on the phonological, semantic, and structural features of its singular stem (e.g., Laaha & Dressler, 2012).

Productivity

Productivity concerns the extent to which a morpheme can be extended to foreign words, neologisms etc. (e.g., Dressler, 2003; Kjærbæk, 2013, 2015; Kjærbæk, dePont Christensen, & Basbøll, 2014).

The acquisition of the Danish noun plural system is interesting when testing the five frequency theses above: whereas for example English is characterized by having one default inflectional marker for a grammatical category (e.g., the plural suffix *-s*) and a minor number of exceptions to this default rule, Danish has several competing inflectional markers, and there are important interactions between phonology and morphology in the Danish system.

The Danish noun plural system

In Danish noun plural can be formed in four different ways, namely by:

a) adding a plural suffix to the singular stem as in *bil* [biːˀl] ‘car’ – *bil-er* [ˈbiːˀlɐ] ‘cars’
   - *Possible suffixes*: /ɐ/, /ə/, /Ø/ and the foreign suffixes /s/, /a/, /i/
b) stem change as in *mand* [manˀ] ‘man’ – *mænd* [mɛnˀ] ‘men’
c) adding a plural suffix to the singular stem combined with stem change as in *fod* [foðˀ] ‘foot’ – *fødd-er* [ˈføðˀɐ] ‘feet’
   - *Possible stem changes*: stød addition, stød drop, syncope, a-quality change, r-insertion, n-insertion and umlaut
d) no change (plural = singular) as in *mus* [muːˀs] ‘mouse’ – *mus* [muːˀs] ‘mice’

The Danish plural markers

Based on detailed analyses of the Danish noun plural system from a phonological perspective Basbøll et al. (2011) identified 23 different plural markers (suffix + stem change (incl. no change)). These plural markers are different from markers based on orthography.

Transparency scales

We here present a transparency scale for the Danish plural markers (stem + suffix). First we have developed a scale for the plural *stems* saying that a plural stem is:

1) *Transparent* when it equals the singular stem
   - No stem change (No Change)
     - *bil* [biːˀl] ‘car’ – *biler* [ˈbiːˀlɐ] ‘cars’
2) *Partly transparent* when it involves prosodic stem change
   - Stød drop, stød addition, syncope, a-quality/vowel length change (Prosodic Change)
     - *bord* [boɐ̯ˀ] ‘table’ – *borde* [ˈboːɐ] ‘tables’
     - *baby* [ˈbɛjbi] ’baby’ – *babyer* [ˈbɛjbiːˀɐ] ‘babies’
     - *gaffel* [ˈgɑfəl]/[ˈgɑfl̩] ‘fork’ – *gafler* [ˈgɑflɐ] ‘forks’
     - *blad* [blað] ‘leaf’ – *blade* [ˈblæːðə]/[ˈblæːð̩] ‘leafs’
3) *Not transparent* when it involves phonemic stem change
   - Umlaut, r-insertion, n-insertion (Phonemic Change)
     - *mand* [manˀ] ‘man’ – *mænd* [mɛnˀ] ‘men’
     - *søster* [ˈsøsdɐ] ‘sister’ – *søstre* [ˈsøsdʁɐ] ‘sisters’
     - *øje* [ˈʌjə] ‘eye’ – *øjne* [ˈʌjnə] ‘eyes’

The three degrees of transparency equals the three degrees of stem change presented in Basbøll et al. (2011) and Laaha et al. (2011).

We have, furthermore, developed a transparency scale for the Danish plural *suffixes*, saying that a plural suffix is:

1) *Transparent* when a suffix is simply added to the stem
   - /ɐ/-suffix
     - *bil* [biːˀl] ‘car’ – *biler* [ˈbiːˀlɐ] ‘cars’
2) *Partly transparent* when it is often reduced or assimilated with the stem
   - /ə/-suffix
     - *hus* [huːˀs] ‘house’ – *huse* [ˈhuːsə]/[ˈhuːːs] ‘houses’
3) *Not transparent* when it is phonologically non-existing
   - Ø-suffix
     - *mål* [mɔːˀl] ‘goal’ – *mål* [mɔːˀl] ‘goals’
     - *mand* [manˀ] ‘man’ – *mænd* [mɛnˀ] ‘men’

On the basis of the stem transparency scale and the suffix transparency scales above, we have developed a transparency scale for the whole plural *marker* where we combine the information of stem and suffix. When ranking the plural markers with respect to transparency versus opacity, we have considered Prosodic Change to subtract only little from transparency, resulting in the following four-step gradation (of Partly transparent) between Transparent and Not transparent:

1) *Transparent* when both the plural suffix and the plural stem are Transparent
   - /ɐ/-suffix + No Change
     - *bil* [biːˀl] 'car' – *biler* [ˈbiːˀlɐ] 'cars'
2) *Partly transparent 1* when the plural suffix is Transparent and the plural stem is Partly transparent
   - /ɐ/-suffix + Prosodic Change
     - *bord* [boɐ̯ˀ] 'table' – *borde* [ˈboːɐ] 'tables'
     - *baby* [ˈbɛjbi] *'baby'* – *babyer* [ˈbɛjbiːˀɐ] 'babies'
     - *sofa* [ˈsoːfa] 'sofa' – *sofaer* [ˈsoːfæːˀɐ] 'sofas'
     - *gaffel* [ˈgɑfəl]/[ˈgɑfl̩] 'fork' – *gafler* [ˈgɑflɐ] 'forks'
3) *Partly transparent 2* when the plural suffix is Partly transparent and the plural stem is Transparent
   - /ə/-suffix + No Change
     - *digt* [degd] 'poem' – *digte* [ˈdegdə] 'poems'
     - *tov* [tʌw] 'rope' – *tove* [ˈtʌwə]/[ˈtʌww̩] 'ropes'
4) *Partly transparent 3* when the plural suffix and the plural stem are both Partly transparent
   - /ə/-suffix + Prosodic Change
     - *hus* [huːˀs] 'house' – *huse* [ˈhuːsə]/[ˈhuːːs] 'houses'
     - *blad* [blað] 'leaf' – *blade* [ˈblæːðə]/[ˈblæːð̩] 'leafs'
5) *Partly transparent* 4 when the plural suffix is Transparent and the plural stem is Not transparent
   - /ɐ/-suffix + Phonemic Change
     - *bror* [bʁoɐ̯] 'brother' – *brødre* [ˈbʁœðʁɐ] 'brothers'
6) *Not transparent* when both the plural suffix and the plural stem are Not transparent
   - Ø-suffix + Phonemic Change
     - *mand* [manˀ] 'man' – *mænd* [mɛnˀ] 'men'
     - *søster* [ˈsøsdɐ] 'sister' – *søstre* [ˈsøsdʁɐ] 'sisters'
     - *øje* [ˈʌjə] 'eye' – *øjne* [ˈʌjnə] 'eyes'

We predict the acquisition to go from Transparent > Partly transparent > Not transparent. We furthermore predict the error direction to go from Not transparent > Partly transparent > Not transparent.

Noun plurals with insertion of /r/ as in *søster* [ˈsøsdɐ] 'sister' – *søstre* [ˈsøsdʁɐ] 'sisters' and /n/ as in *øje* [ˈʌjə] 'eye' - *øjne* [ˈʌjnə] 'eyes', have two possible analyses according to the principles we adopt: they can be considered as having a non-null plural suffix, i.e. /ɐ/-suffix and /ə/-suffix, respectively, combined with the phonemic stem change and syncope; this analysis is used in Laaha et al. (2011). Or they can be considered as having a Ø-suffix, and then the segmental stem change (insertion of /r/ or /n/) will be the only overt plural marker; this is the analysis chosen in the present paper (as in Basbøll et al., 2011; Kjærbæk 2015; Kjærbæk et al., 2014).

Productivity scale

Kjærbæk et al. (2014) presented a scale with three degrees of productivity. Productivity is here defined as the ability of the inflectional marker to occur on new words. For the plural system this means the ability to add the plural marker to a new singular noun in order to form a new plural noun. The productivity scale for the Danish plural markers is:

1) *Fully Productive* plural markers are plural markers taking the /ɐ/-suffix without phonemic stem change.
2) *Semi-productive* plural markers are plural markers taking the /ə/-suffix or Ø-suffix, in both cases without phonemic stem change.
3) *Unproductive* plural markers are plural markers with phonemic stem change (as well as plural markers with the foreign plural suffixes /s/, /a/ and /i/).

In this study we will test the five theses on frequency effects suggested by Ambridge et al. (2015) in a phonological perspective and explore the impact of phonology on morphology. This we will do in three types of empirical data from children acquiring Danish as their first language.

Empirical data

Naturalistic data

The naturalistic data consist of spontaneous child language input and output from:

a) the Odense Twin Corpus (OTC) (Basbøll et al., 2002). The subpart used here consists of data from two twin pairs: i) the girls Ingrid and Sara between the ages of 0;10 and 2;7; ii) the girl Cecilie and the boy Albert between the ages of 0;11 and 2;5
b) the Danish Plunkett Corpus (DPC) (Plunkett, 1985; 1986) which consists of data from two singletons: i) the girl Anne between the ages of 1;1 and 2;11; ii) the boy Jens between the ages of 1;0 and 3;11.

The corpus consists of video and audio recordings of children interacting with their families in naturalistic settings (playing and dining situations) in their own home. The input is a mixture of child directed and adult directed speech, though the child is always present. The data are transcribed orthographically using the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000a, b) and coded morphologically and phonologically (according to the standard pronunciation) in OLAM (Madsen, Basbøll, & Lambertsen, 2002). See Kjærbæk (2013) for a detailed description of the naturalistic data.

Table 1 shows the size of the corpus in raw numbers with regard to word tokens and word types (different lemmas) as well as noun tokens and noun types.

Table 1. Sample size of naturalistic spontaneous child language input and output

	Words		Nouns	
	Tokens	**Types**	**Tokens**	**Types**
Input	180,360	3,342	14,126	1,574
Output	40,987	1,399	5,743	607

Semi-naturalistic data

The semi-naturalistic data consist of structured interviews focusing on familiar routines. An investigator showed the child five pictures of, for example, a trip to the zoo and a birthday party while asking the child prepared questions for maximal elicitation of plural nouns (e.g., *Hvad ser du når du går i zoologisk have?* 'What do you see when you go to the zoo?'). All recordings are transcribed ortographically in CHILDES and coded morphologically and phonologically (according to the standard pronunciation) in OLAM. All nouns are furthermore transcribed phonetically according to the child's actual pronunciation.

80 monolingual Danish children (41 girls, 39 boys) in the age groups 3, 5, 7 and 9 years participated in this task, 20 children in each group. Children participating in the interviews also participated in the experiment (see just below).

Experimental data

The experimental data consist of data from a picture based elicitation task inspired by Jean Berko's study on both real words and pseudo-words (Berko, 1958). This experiment is only based on real words. The test material consists of 48 stimulus items. Only itmes with an overt plural marker were included in the test, i.e. Pure Zeroes (i.e. plural = singular, e.g., *mål* [mɔːˀl] 'goal' - *mål* [mɔːˀl] 'goals') were excluded because of the difficulty of distinguishing Pure Zero production from repetition of the singular form in the plural elicitation task. Since the plural suffixes /s/, /a/ and /i/ are very rare in child language, they were not included in the experiment.

Children were tested orally and individually. Each child was presented with a picture of an object whose name is a singular noun (e.g., *bil* 'car'), and the investigator said: *Her er en bil* 'Here is a car'. Then a second picture, of two instances of the same object, was shown to the child, and the investigator asked: *Her er to hvad?* 'Here are two what?', and the child's task was to provide the respective plural form. Test items were presented in different orders and were preceded by three training items.

160 monolingual Danish children between the ages of 3-10 years participated in the experiment.

Results

The results of the study are presented here.

Input frequency of the plural suffixes

Table 2 shows the input frequency of the Danish plural suffixes in our corpus of naturalistic child language input and output. We see that 64 % of the nouns (type frequency) take the /ɐ/-suffix, 20 % take the Ø-suffix whereas only 12% take the /ə/-suffix. The plural suffixes /s/, /a/, /i/ and nouns with only a plural form are excluded from the table – they sum up to a total of 4 %.

Table 2. Input frequency of the Danish plural suffixes

Suffix	Token	Type
/ɐ/	55 %	64 %
Ø	31 %	20 %
/ə/	10 %	12 %
Total	96 %	96 %

Correctly produced plural suffixes in the experiment

Figure 1 illustrates the proportion of correctly produced plural suffixes by age in the experiment. We see that the proportion of correctly produced plural suffixes increases with age. The /ɐ/-suffix constitutes the highest proportion of correctly produced plural suffixes followed rather closely by the /ə/-suffix, in fact they appear to coincide from the age of six. The proportion of correctly produced Ø-suffixes is rather low, compared to the other two suffixes. Please note that the only zero-plurals included in the experiment have phonemic stem change, that is, Pure Zeros (plural = singular) are not included.

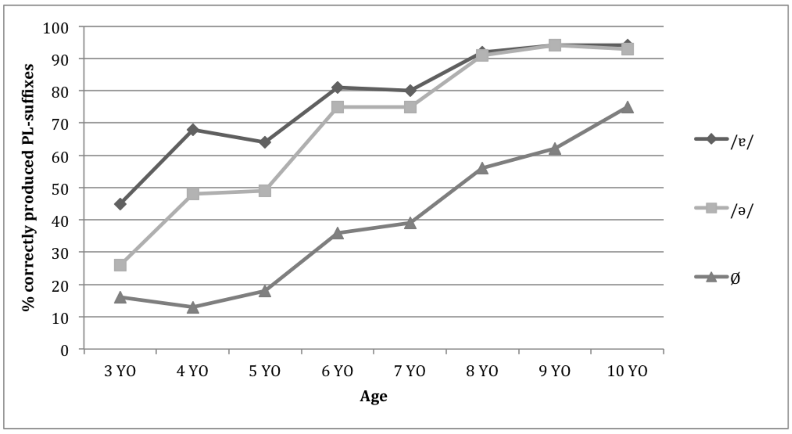


Figure 1. Proportion of correctly produced plural suffixes by age and type of suffix in the experiment
(Kjærbæk, dePont Christensen, & Basbøll, 2014, p. 62)

Input frequency of the plural stem changes

Table 3 shows the input frequency of the Danish plural stem changes (including No change) in our corpus of naturalistic child language input. Please note that a plural form can have more than one kind of stem change at the same time. We see that 71 % (type frequency) of the Danish nouns have No change of the plural stem compared to the singular stem. 14 % have Stød drop, 5 % have Stød addition, 4 % have Umlaut, 2 % have Syncope, 1 % have r-insertion, 0.6 % have a-quality change combined with change in vowel length and only 0.2 % have n-insertion (only one noun, namely *øje* [ˈʌjə] ‘eye’ – *øjne* [ˈʌjnə] ‘eyes’).

Table 3. Input frequency of the plural stem changes (including No change)

Stem change	**Tokens**	**Types**
No change	63 %	71 %
Stød drop	15 %	14 %
Stød addition	3 %	5 %
Umlaut	12 %	4 %
Syncope	1 %	2 %
r-insertion	4 %	1 %
a-quality change	0.2 %	0.6 %
n-insertion	2 %	0.2 %

Correctly produced plural stem changes in the experiment

Figure 2 illustrates the proportion of correctly produced plural stem changes by age and type of stem change in the experiment. The highest proportion we find with No change where the children only produce very few errors in all age groups. For all other stem changes we see that the proportion of correctly produced stem changes increases with age. It appears that the correctly produced plural stems fall into three categories:

1) No change

2) Syncope, a-quality change combined with change in vowel length, Stød drop and Stød addition (these are all prosodic stem changes)
3) Umlaut, r-insertion and n-insertion (which are all phonemic stem changes)

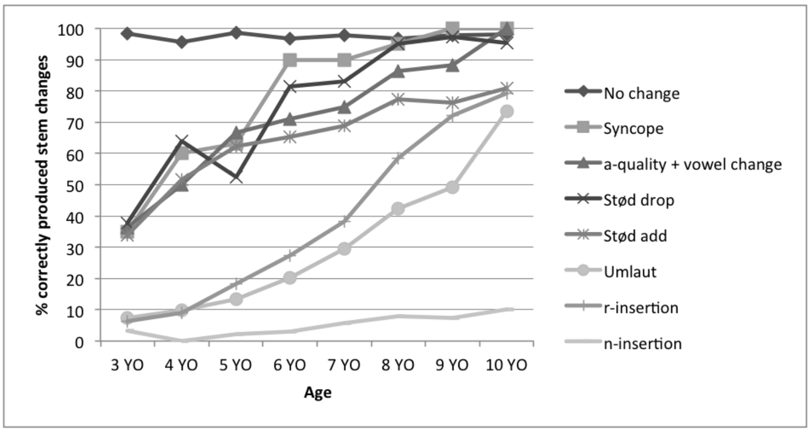


Figure 2. Proportion of correctly produced plural stem changes by age and type of stem change in the experiment (Kjærbæk, dePont Christensen, & Basbøll, 2014, p. 63)

Figure 3 illustrates the proportion of correctly produced plural stems by age and degree of stem transparency in the experiment. Again we see that the children produce very few errors in the No Change category (Transparent), followed by Prosodic Change (Partly transparent) and least correct in the Phonemic Change category (Not transparent).

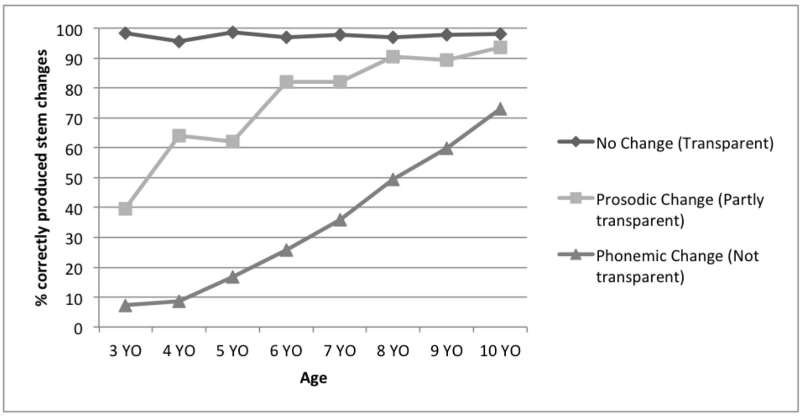


Figure 3. Proportion of correctly produced plural stems by age and degree of stem transparency in the experiment (Kjærbæk, dePont Christensen, & Basbøll, 2014, p. 64)

Input frequency of the plural marker

Table 4 shows the input frequency of the Danish plural markers in our corpus of naturalistic child language input. The plural markers are here divided according to their degree of productivity. We see that Fully Productive plural markers have an input frequency of 63 % (type frequency), Semi-productive 31% and Unproductive plural markers only have an input frequency of 6 %.

Table 4. Input frequency of the plural markers according to productivity

Degree of productivity	Tokens	Types
Fully productive	50 %	63 %
Semi-productive	32 %	31 %
Unproductive	18 %	6 %
Total	100 %	100 %

Correctly produced plural forms in the experiment

Figure 4 illustrates the proportion of correctly produced plural forms by age and degree of productivity in the experiment. In the younger age groups, children produce more correct plural forms of nouns taking a Fully Productive plural marker compared to nouns taking a Semi-Productive plural marker, but the difference between the two categories seems to vanish in the older age groups. Unproductive plural markers have much lower correctness rate in the experimental data compared to the other plural markers. Remember that there are no Pure Zeroes (plural = singular) included in the experiment, so the Semi-Productive plural markers here only include plural forms with the /ə/-suffix.

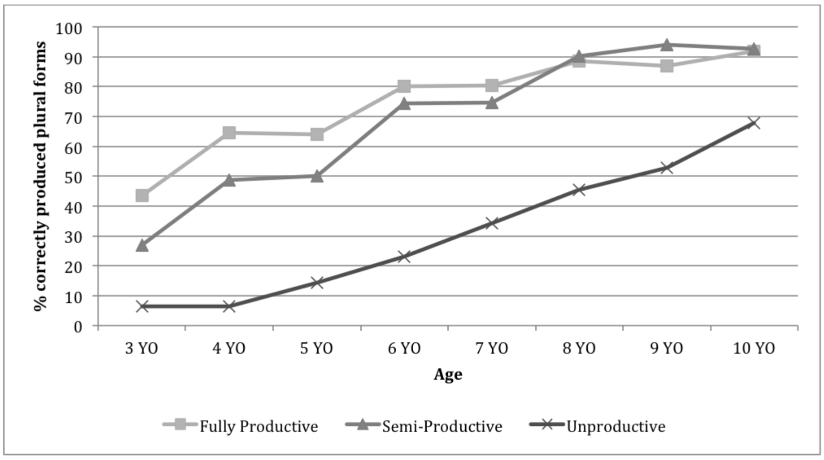


Figure 4. Proportion of correctly produced plural forms by age and degree of productivity in the experiment (Kjærbæk, dePont Christensen, & Basbøll, 2014, p. 65)

Classification of produced plural forms in the experiment

Figure 5 illustrates the produced plural forms in the experiment divided into four different categories: i) Correct plural forms; ii) Pure Zero (plural = singular); iii) Wrong stem and/or wrong suffix; iv) Other (when the child produced a completely different form, e.g., *piger* 'girls' instead of *døtre* 'daughters'). We see that Pure Zero (plural = singular) is clearly the most frequent error form in the experimental data. The children only produce very few error forms in the other two categories.

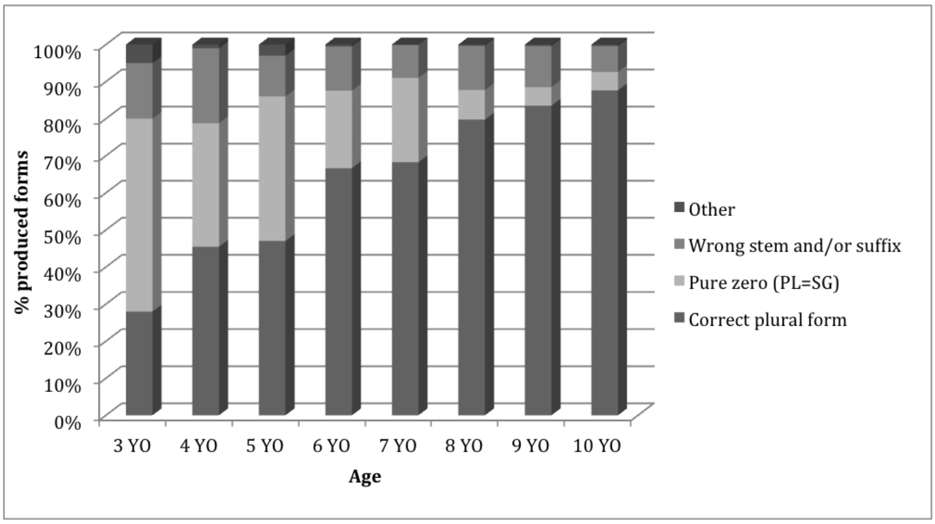


Figure 5. Produced plural forms in the experiment by age and type

Error direction in the structured interviews

Table 5 shows the error direction of the plural error forms produced by the children in the structured interviews. We see that 47 % of all error forms in the structured interviews are children producing a Semi-Productive plural marker instead of a Fully Productive plural marker (FP > SP). 19 % is children producing one Semi-Productive plural marker instead of another Semi-Productive plural marker (SP > SP), whereas 18 % of the plural error forms are children producing a Semi-Productive plural marker instead of a Fully Productive plural marker (SP > FP) and 8 % are children producing an Unproductive plural marker instead of a Semi-Productive plural marker (UP > SP). The remaining categories (UP > FP, FP > FP, UP > UP) are not very frequent (2 %, 2 % and 4 % respectively). In sum only 28 % of all plural error forms are in the expected direction, i.e. with increasing productivity (UP > SP > FP), whereas 47 % go in the opposite direction (FP > SP > UP).

Table 5. Plural error direction in the structured interviews

Error direction	Percentage of all errors
FP > SP	47 %
SP > SP	19 %
SP > FP	18 %
UP > SP	8 %
UP > UP	4 %
UP > FP	2 %
FP > FP	2 %
Total	100 %

Error pattern in the structured interviews

If we look further into our detailed analyses of the plural error forms produced by the children in the structured interviews we see that 47 % of all error forms are children producing the Semi-Productive plural marker Pure Zero (plural = singular) instead of the Fully Productive plural marker taking the

/ɐ/-suffix. 17 % of all error forms are children producing the Semi-Productive plural marker Pure Zero (plural = singular) instead of a Semi-Productive plural marker taking the /ə/-suffix. 10 % are a Fully Productive plural marker with /ɐ/-suffix instead of a Semi-Productive plural marker with /ə/-suffix. 9 % is with a Fully Productive plural marker with an /ɐ/-suffix instead of a Semi-Productive plural marker with Ø-suffix. And 6 % of the plural error forms are children producing a Semi-Productive plural marker with /ə/-suffix instead an Unproductive plural marker with Ø-suffix. These numbers are displayed in Table 6. The category *Others* includes errors outside the other categories – in total 6 out of the 11 categories.

Out of all error forms in the structured interviews 66 % are overgeneralizations of the Semi-Productive plural marker 'Ø' (Pure Zero, i.e. plural = singular). 70 % of these overgeneralizations of the Semi-Productive plural marker 'Ø' are changes from the Fully Productive plural marker /ɐ/ (i.e. in the opposite direction of what would be expected if productivity alone was a relevant factor). Out of the rest of the error forms (i.e. the 34 % that are not changes to the Semi-Productive marker 'Ø'), 62 % are overgeneralizations of the Fully Productive plural marker /ɐ/ (i.e. in the expected direction). There is only one single overgeneralization of an unproductive plural marker (UP Ø), and never of the unproductive foreign suffixes /s/, /a/ and /i/.

Table 6. Plural error patterns in the structured interviews

Error direction	Percentage of all errors
FP /ɐ/ > SP Ø	47 %
SP /ə/ > SP Ø	17 %
SP /ə/ > FP /ɐ/	10 %
SP Ø > FP /ɐ/	9 %
UP Ø > SP /ə/	6 %
Others	*11 %*
Total	100 %

Discussion

According to our naturalistic data the /ɐ/-suffix is the most frequent *plural suffix* in the language input to Danish children, then comes the Ø-suffix and last the /ə/-suffix (see Table 2). According to our experimental data, Danish children produce more correct plural suffixes of the /ɐ/-suffix than of the /ə/-suffix and the Ø-suffix (see Figure 1). This is in accordance with the *Age of Acquisition Thesis,* which claims that frequent forms are acquired before less frequent forms, as well as the *Prevent Error Thesis,* which claims that high-frequency forms prevent or reduce errors in contexts in which they are the target. The /ɐ/-suffix, however, is followed rather closely by the /ə/-suffix whereas the Ø-suffix seems to be acquired rather late compared to the other two suffixes, even though the Ø-suffix is almost twice as frequent (type frequency) as the /ə/-suffix in the children's language input. This indicates that type frequency is not the only factor playing a role in the acquisition of plural suffixes. Remember that the Ø-suffix only appears with nouns with phonemic stem change in this experiment.

Turning to *plural stem change* we see that according to our corpus of naturalistic child language input 71 % (type frequency) of the Danish nouns have No change of the plural stem compared to the singular stem. 14 % have Stød drop, 5 % have Stød addition, 4 % have Umlaut, 2 % have Syncope, 1 % have r-insertion, 0.6 % have a-quality change combined with change in vowel length and only 0.2 % have n-insertion. According to the *Age of Acquisition Thesis* as well as the *Prevent Error Thesis* we would therefore expect to see the highest correctness rate with stems of the No change-category, followed by Stød drop, Stød addition, Umlaut, Syncope, r-insertion, a-quality change combined with change in vowel length, and we expect to see the lowest correctness rate for stems with n-insertion. This is not quite what we see in our experimental data, though. Most importantly, Syncope and a-quality change combined with change in vowel length seem to be acquired relatively long before what could be expected on the basis of input frequency alone, i.e. other factors interact, confer the *Interaction Thesis*.

The correctly produced plural stems in the experiment fall into three categories: 1) No stem change; 2) Prosodic stem changes (Syncope, a-quality change, Stød drop and Stød addition); and 3) Phonemic stem changes (Umlaut, r-insertion, n-insertion) (see Figure 2). These categories are identical to the three degrees of transparency presented above: 1) No Change (Transparent); 2) Prosodic Change (Partly transparent); and 3) Phonemic Change (Not transparent). The Danish children produce very few errors in the No Change-category (Transparent), followed by Prosodic Change (Partly transparent) and least correct plural forms in the Phonemic Change-category (Not transparent) (see Figure 3). We therefore argue that degree of stem transparency affects the acquisition of the plural stem (e.g., confer the *Interaction Thesis*).

In the younger age groups, Danish children produce more correct *plural forms* of nouns taking a Fully Productive plural marker compared to nouns taking a Semi-Productive plural marker, but they appear to coincide in the older age groups. On the other hand, Unproductive plural markers have much lower correctness rate in the experimental data compared to the other plural markers. Remember that there are no Pure Zeroes (plural = singular) included in the experiment, i.e. the Semi-Productive plural markers here only include plural forms with the /ə/-suffix (see Figure 4).

According to the classification of produced plural forms in the experiment, it seems that Danish children produce the singular form of the noun when they don't know or recall the correct plural form (see Figure 5). This could be due to the type of experiment, where the child is meant to produce a plural form based on a singular form given by an investigator. The children may simply repeat the singular form given by the investigator. If this is the case, we should see the same pattern when completing the task with children acquiring other languages, but we don't. Gillis et al. (2008) compared Danish, German, Dutch, and Hebrew-speaking children who had completed the same task. Danish was significantly different for all age groups with Danish in the top. We believe that this is mainly due to the fact that Pure Zero (plural = singular) is a very important category in Danish, in German it occurs but it's less important, whereas it is not found in the two other languages. Furthermore, the dropping of the /ə/-suffix in Danish often results in a plural form which is almost identical with the singular form, as in *tov* [tʌw] 'rope' - *tove* [ˈtʌwə]/[ˈtʌw̩] 'ropes', and thereby plurals identical, or almost identical, with singular forms are even more frequent in Danish.

We have made detailed analyses of the plural error forms that the children produced in the structured interviews. Based on input frequency we expected the error direction to go from Unproductive to Semi-productive to Fully Productive – but as you see in Table 6, this is not the case. 47 % of all error form in the structured interviews goes from Fully Productive to Semi-productive, which is certainly not expected on the basis of either input frequency or transparency.

If we go further into the error forms we see that 47 % of all error forms are children producing the Semi-Productive plural marker 'Ø' (Pure Zero, i.e. plural = singular) instead of the Fully Productive plural marker with /ɐ/-suffix. 17% of all error forms are children producing the Semi-Productive plural marker Pure Zero (plural = singular) instead of a Semi-Productive plural marker with /ə/-suffix. 10 % are a Fully Productive plural marker with /ɐ/-suffix instead of a Semi-Productive plural marker with /ə/-suffix. 9 % with a Fully Productive plural marker with /ɐ/-suffix instead of a Semi-Productive plural marker with Ø-suffix. And last 6% of the error forms are children producing a Semi-Productive plural marker with /ə/-suffix instead of an Unproductive plural marker with Ø-suffix. Thus, Pure Zero (plural = singular) is not only the most frequent plural error form in the experimental data, where it could simply be a repetition of the singular form given by the investigator. It is also the most frequent plural error form in the semi-naturalistic data, where the child is not given a singular form. Furthermore, we see that 51 % of the Pure Zeroes (plural = singular) in the experimental data are produced in a plural context (e.g., **to bil* ‘two car’) and only 1 % in a singular context (e.g., *en bil* ‘a car’). 48 % are produced out of context (e.g., *bil* ‘car’) (Kjærbæk et al., 2014).

Conclusion

We can conclude that the Danish study supports the five theses on effects of input frequency presented by Ambridge and colleagues (2015). With regard to the *Interaction Thesis*, the present study points to the importance of frequency, transparency and productivity. The /ɐ/-suffix has a high input frequency and a stable phonology (Transparent), which results in early acquisition as well as in overgeneralization. The /ə/-suffix has a low input frequency and it is opaque since it is often reduced or assimilated with the stem (Partly transparent); this seems to result in later acquisition of the plural /ə/-suffix, but the difference between the /ə/-suffix and the /ɐ/-suffix seems to vanish at the age of six (start of pre-school). Furthermore, the /ə/-suffix is very seldom overgeneralized. The Ø-suffix has a low input frequency and it is not phonologically expressed (Not transparent); this results in late acquisition, but, interestingly, it is very often overgeneralized. The foreign suffixes /s/, /a/ and /i/ are not overgeneralized. The study furthermore indicates that transparency has an effect on the acquisition of the plural stem: No Change (Transparent) seems to be acquired early, then come Prosodic Change (Partly transparent) and last Phonemic Change (Not transparent).

Acknowledgement

We would like to thank the children who participated in this study, their families and the schools and day care centers where the data were collected. Thanks to Katja Rehfeldt for participating in the data collection, to Claus Lambertsen for his assistance with the OLAM-system and to René dePont Christensen for statistical assistance. This study was supported by the Carlsberg Foundation and the Department of Language and Communication, University of Southern Denmark.

References

Albirini, A. (2015). Factors affecting the acquisition of plural morphology in Jordanian Arabic. *Journal of Child Language*, *42*(2), 734-762.

Ambridge, B., Kidd, E., Rowland, C.F., & Theakston, A. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239-273.

Basbøll, H. (2005). *The phonology of Danish*. Oxford, UK: Oxford University Press.

Basbøll, H., Bleses, D., Cadierno, T., Jensen, A., Ladegaard, H.J., Madsen, T.O., Millar, S., Sinha, C., & Thomsen, P. (2002). The Odense language acquisition project. *Child Language Bulletin*, *22*, 11-12.

Basbøll, H., Kjærbæk, L., & Lambertsen, C. (2011). The Danish noun plural landscape. *Acta Linguistica Hafniensia*, *43*(2), 81-105.

Diessel, H. (2015). Frequency shapes syntactic structure. *Journal of Child Language*, *42*(2), 278-281.

Dressler, W.U. (2003). Degrees of grammatical productivity in inflectional morphology. *Italian Journal of Linguistics*, *15*(1), 31-62.

Gillis, S., Souman, A., Dhollander, S., Molemans, I., Kjærbæk, L., Rehfeldt, K., Basbøll, H., Lambertsen, C., Laaha, S., Bertl, J., Dressler, W.U., Lavie, N., Levie, R., & Ravid, D. (2008). The classical task: From singular to plural form in Dutch, Danish, Austrian German, and Hebrew. Presented at *the XIth Congress of the International Association for the Study of Child Language (IASCL)*, Edinburgh.

Kjærbæk, L. (2013). *Tilegnelse af bøjningsmorfologi: En undersøgelse af substantivernes pluralisbøjning hos normaltudviklede børn I alderen 0-10 år.* PhD Thesis. University of Southern Denmark.

Kjærbæk, L. (2015). Danske børns tilegnelse af bøjningsmorfologi – udviklingen af substantivernes pluralisbøjning i alderen 0-10 år. *Nydanske Sprogstudier (NyS)*, *48*, 106-138.

Kjærbæk, L., dePont Christensen, R., & Basbøll, H. (2014). Sound structure and input frequency impact on noun plural acquisition: Hypotheses tested on Danish children across different data types. *Nordic Journal of Linguistics*, *37*(1), 47-86.

Laaha, S., & Dressler, W.U. (2012). Suffix predictability and stem transparency in the acquisition of German noun plurals. In Kiefer, F., Ladánvi, M. & Siptár, P (eds.) *Current issues in morphological theory: (ir)regularity, analogy and frequency* (pp. 217-236). Amsterdam: John Benjamins.
MacWhinney, B. (2002a). *The Childes Project: Tool for Analyzing Talk. Volume 1: Transcription Format and Programs*. Mahwah, NJ: Lawrence Erlbaum Associates.
MacWhinney, B. (2002b). *The Childes Project: Tool for Analyzing Talk. Volume 2: The Database*. Mahwah, NJ: Lawrence Erlbaum Associates.
Madsen, T., Basbøll, H., & Lambertsen, C. (2002). OLAM – et semiautomatisk og lydstrukturelt kodningssystem for dansk. *Odense Working Papers in Language and Communication*, *24*, 43-56.
Plunkett, K. (1985). *Preliminary approaches to language Development*. Aarhus: Aarhus University Press.
Plunkett, K. (1986). Learning strategies in two Danish children's language development. *Scandinavian Journal of Psychology*, *27*, 64-73.
Rowe, M. (2015). Input versus intake – a commentary on Ambridge, Kidd, Rowland, and Theakston's 'The ubiquity of frequency effects in first language acquisition'. *Journal of Child Language*, *42*(2), 301-305.

Effects of English onset restrictions and universal markedness on listeners' perception of English onset sequences resulting from schwa deletion

Shinsook Lee
leesseng@korea.ac.kr

Korea University

Abstract. A considerable body of research on speech perception found that L1 phonotactic restrictions play a key role in the perception of not only L1 (Massaro & Cohen, 1983) but also L2 sound sequences (Depoux, Kakehi, Hirose, Pallier, & Mehler, 1999). However Berent, Steriade, Lennertz and Vaknin (2007) and Berent, Lennertz, Jun, Moreno, and Smolensky (2008) found that listeners' perception of onset clusters can be affected by the sonority-driven onset markedness in addition to L1 phonotactic restrictions. Specifically, they reported that onset clusters of sonority rises tended to be perceived more accurately than onsets of sonority levels, which were in turn perceived more accurately than onset clusters of sonority falls (e.g., *dlaf* vs. *tpif*. vs. *mdip*) across different L1 listener groups. Although English admits only onset sequences of a large sonority rise, like /bl/ and /gr/, certain prohibited onset clusters can emerge due to word-initial schwa deletion (e.g., *banana* [**bn**ǽnə], *potato* [**pt**éɪɾoʊ]). The current study investigated whether both L1 and L2 listeners were perceptually sensitive to the sonority-based onset markedness as well as to English legal vs. illegal onset clusters derived from word-initial schwa deletion in English. Native English, Korean, and Japanese listeners participated in identity judgment tests. The stimuli were made up of 28 bisyllablic and 28 trisyllabic English nonce words on the basis of Lee (2011). More specifically, 112 identical (e.g., *patoo*—*patoo, ptoo*—*ptoo;* nafamic—nafamic, nfamic—nfamic) and non-identical pairs each (e.g., *patoo*—*ptoo*, *ptoo*—*patoo;* nafamic—nfamic, nfamic—nafamic), resulting from initial schwa deletion were created from 56 nonce words. The stimuli were further divided into onsets of a sonority rise (e.g., *kl*, *dn*), flat (e.g., *pt*), and fall (e.g., *nf*). Participants identified, whether aurally presented two stimulus words were identical or not, by pressing a key on a keyboard. The results of accuracy indicated that English, Korean, and Japanese listeners were able to differentiate between well-formed and ill-formed English onset clusters, and reaction latency showed a similar trend. Importantly, the results of the sonority profiles were consistent with the findings of Berent et al. (2007; 2008), since all the listeners showed an illusionary vowel effect as a function of the onset markedness irrespective of their L1s. That is, the listeners tended to equate schwa-deleted forms with their corresponding vowel intact forms as the sonority-based onset markedness increases. The findings are further discussed in terms of L1 phonotactic restrictions, universal markedness, lexical representations, and L2 listeners' English proficiency.

Keywords: English onset restrictions, sonority-based markedness, initial-schwa deletion, perception

Introduction

Studies on phonological acquisition have documented that speakers' knowledge of their phonological system has a great impact on the perception of speech sounds. For instance, speakers' knowledge of an L1 sound system functions as a phonological filter, assimilating nonnative sounds to articulatorily and/or perceptually similar native sound categories (Best, 1995; Best & Tyler, 2007; Flege, 1995).

As for language-specific phonotactic constraints, Jusczyk, Luce, and Charles-Luce (1994) found that 9-month-old English infants, but not 6-month-old infants, showed preferences for nonce words with high-probability phonotactic patterns in English (e.g., [kæz] "kazz", [taɪs] "tyce") to those with low-probability phonotactic patterns (e.g., [guʃ] "gushe", [ʃaɪb] "shibe"). Messer (1967) reported that children acquiring English (mean age: 3;7 years) tended to discriminate and produce monosyllabic nonce words with a legitimate onset cluster (e.g., [frul], [trisk]) more accurately than those with an illicit onset cluster (e.g., [mrul], [ʃkib]). Messer attributed the result to the children's perceptual bias for well-formed speech sounds. Similarly, /tl/, /dl/ and /dn/ are illegitimate sound sequences in syllable- or word-initial position in English even though they can occur in word-medial position (e.g., *atlas*, *bedlam*, *kidney*, Hammond, 1999). Massaro and Cohen (1983) and Pitt (1998) found that English listeners' perception of illicit onset clusters was modulated by their L1 phonotactic restrictions in that ill-formed onset clusters tended to be misperceived as licit ones (e.g., /tl/ as [tr]). Hallé, Segui, Frauenfelder, and Meunier (1998) also reported that French listeners were more likely to misperceive word-initial /tl/ and /dl/ as [kl] and [gl], respectively (e.g., *tlabod*, *dlapot*). This was because /tl/ and /dl/ are illegitimate onsets whereas /kl/ and /gl/ are legitimate onsets in French and thus the result showed that French listeners' perception of onset clusters was influenced by the legitimacy of such onsets in French.

Similar effects of L1 phonotactic constraints were also reported by Depoux et al. (1999) who ran several experiments on native Japanese and French listeners using nonce words. Specifically, Dupoux et al. observed that Japanese listeners were more liable to misjudge nonce words with consonant clusters as their vowel inserted counterparts (e.g., *akmo-akumo*, *egdo-egudo*), as such consonant sequences deviate from the canonical syllable structures in Japanese. On the contrary, French listeners had much trouble distinguishing between nonce words with a vowel length contrast (e.g., *akumo-akuumo*, *egudo-eguudo*) since vowel length does not function as a contrastive feature in French. Spanish listeners were also found to misperceive English *sC* onset clusters as [ɛsC] due to the illegitimacy of *sC* onsets in Spanish. Accordingly, these results indicate that listeners are sensitive to their L1 phonotactic constraints and they tend to repair illegal sound sequences mostly by vowel epenthesis, making such illegitimate sequences fit with the canonical syllable structures of their L1s.

Moreover, many languages with onset clusters are known to show preferences for certain types of onsets (e.g., *pl*, *dr*, *gr*, etc.) to other types (e.g., *pt*, *bd*, *lb*, etc.), and such preferences have been attributed to the sonority contours of the onsets (Clements, 1990). Sonority is correlated with acoustic intensity (Ladefoged, 2006), sound audibility (Heffner, 1950), or articulatory properties (Yavaş, 2006). Because sonority is a relative property, sounds are arranged on the sonority scale, as demonstrated in (1) (adapted from Berent et al., 2008):

(1) A sonority scale: most sonorous ⟷ least sonorous
Glides = 4 > Liquids = 3> Nasals = 2 > Obstruents (Fricatives and Stops) =1

Languages, like English, have certain restrictions on possible onset cluster types such that there should be an abrupt sonority rise in the onset although onset sequences like /tl/ and /sr/ are ruled out in spite of the fact that they satisfy sonority requirements. Accordingly, English accepts words like *play*, *bring*, *grass*, *fry*, *shrine*, *cute*, and *twin* which all manifest a large sonority rise in the onset as the sound sequences consist of obstruents and liquids or glides (Clements, 1990; Selkirk, 1982). However, onsets with a small sonority rise such as oral stop/fricative plus nasal sequences or those with a sonority level are ruled out (e.g., *tm, km, fn, pn, pt, dg, ks*) although English admits *s*+nasal sequences like *small* and *sneak*, or *s+*stop sequences such as *spy*, *star*, and *sky*. Onsets with a sonority fall (e.g., *lb, rt, nt*) are also illegitimate in English.

In contrast, some languages attest other types of onset clusters with a small sonority rise, a sonority flat, or even a sonority fall. According to Berent et al. (2007, p. 594), Ancient Greek allows onsets of small rises (e.g., *pneuma*, "breath") while Hebrew manifests sonority flats (e.g., *ptil*, "wick"). Russian even accepts sonority falls (e.g., *rzhan*, "zealous", recited from Halle, 1971). Nonetheless, most languages do not tolerate onset clusters of a sonority fall whereas a great number of languages including English allow only onsets of a large sonority rise. Berent et al. (2007; 2008), and Berent, Lennertz, Smolensky, and Vaknin-Nusbaun (2009) investigated the role of sonority-based markedness in the perception of onset clusters by conducting several experiments including syllable count and identity judgment on native English, Russian, and Korean listeners. Specifically, Berent et al. (2007) employed monosyllabic nonce words with onset clusters of different sonority contours: Onset sequences with a sonority rise (e.g., *dlif*, *pnik*), those with a sonority level (e.g., *dbif*, *pkik*), and a sonority fall (e.g., *lbif*, *rtak*). They asked participants in their study to count the number of syllables (e.g., *dlif*: monosyllabic vs. *delif*: disyllabic) or to identify whether the orally presented items are identical or not (e.g., *dlif-dlif*: identical vs. *dlif-delif*: non-identical), using monosyllabic stimuli and their vowel-epenthetic disyllabic counterparts. Berent et al. found that native English and Korean listeners showed sensitivity to the marked nature of onset clusters, such that onset sequences of a sonority rise were better perceived than those of a sonority flat, which were in turn better perceived than those of a sonority fall, although Russian listeners showed somewhat different patterns due to their experience with onsets, like sonority falls. The results were the same regardless of the task types (i.e. syllable count or identity judgment task). Consequently, Berent et al. (2007) argued that the converging results from the experiments point to the existence of universal markedness of onset structures, such that marked onset clusters are more likely to be repaired relative to less marked ones. Sonority-based preferences for attested consonant clusters were reported by many researchers (Gierut, 1999; Ohala, 1999) but Berent et al.'s (2007; 2008; 2009) studies are important in that they provided evidence for the effects of sonority profiles even for unattested onset clusters.

In addition to phonotactic restrictions on onsets, English has an optional process of schwa deletion in word-initial position. According to Patterson, LoCasto, and Connine (2003), a schwa is more likely to be deleted in word-initial position when there is at least one preceding consonant and the following vowel is stressed, as in *police* [plí:s] and *tomato* [tméɪɾoʊ]. Interestingly, illegal onset sequences can result from an initial schwa deletion, as in *taboo* [tbú], *banana* [bnǽnə], and *magician* [mʤíʃən]. The onset clusters created by schwa deletion deserve our attention since they manifest a small sonority rise, a flat, or even a fall, in addition to a large rise (e.g., *polite* [pláɪt], *balloon* [blún]).

Recently, Lee (2011) investigated the perception of English onset clusters resulting from word-initial schwa deletion by English, Korean, and Japanese listeners, employing English nonce words. She examined whether different L1 listeners were perceptually sensitive to the difference between legal and illegal onsets created by initial schwa deletion in English (e.g., *klite-kolite*, *trilla-torilla* vs. *ptoo-patoo*, *nfamic-nafamic*) by running syllable count experiments. She also examined whether the listeners in her study showed perceptual sensitivity to illegal onsets of different sonority contours, resulting from initial schwa deletion. According to her, only English listeners were sensitive to the difference between legal and illegal onsets created by initial schwa deletion in terms of both response accuracy and latency. English listeners were also sensitive to the sonority-based onset markedness, but Korean and Japanese listeners showed only partial sensitivity to onsets of different sonority contours created by initial schwa deletion.

The present study explores whether native English, Korean, and Japanese listeners display sensitivity to English licit vs. illicit onset clusters, and to the different sonority contours of illegal onset sequences resulting from word-initial schwa deletion, by conducting an identity judgment task. Depoux, Pallier, Sebastián-Gallés, and Mehler (1997), and Strange and Shafer (2008) reported that task types can affect the results of speech perception experiments. Accordingly, it is of significance to investigate whether similar results to Lee's (2011) can be obtained when a different kind of experiment is conducted. Moreover, only a few studies have examined the interplay between English phonotactic constraints on onsets and word-initial schwa deletion. Further, the stimuli used in Berent et al. (2007) contained nonce words with /dl/ and /bw/ clusters, which crucially violate English-specific phonotactic constraints regardless of the sonority values of the sounds in the onset, causing a potential confusion from English phonotactic restrictions *per se*. More specifically, the paper aims to answer the following research questions: 1) Do English listeners show different perceptual patterns between attested English onsets of a large sonority rise and unattested onsets of a small sonority rise, a sonority flat, or a sonority fall created by word-initial schwa deletion?; 2) Do Korean and Japanese listeners, whose native languages do not have onset sequences, show similar patterns to native English listeners with respect to attested vs. unattested English onset sequences resulting from schwa deletion?; 3) Do native English, Korean, and Japanese listeners show perceptual sensitivity to the distinction between universally preferred and dispreferred English onsets, resulting from schwa deletion even when all the consonant clusters do not exist in English?

Method

Participants

Twenty-two native English listeners, exchange students or professors at Korea University, Kyonggi University, and Hoseo University in Korea participated in the experiment. Their ages ranged from 19 to 52 (mean: 28.2), and they were paid for their participation. Thirty-two native Korean listeners, undergraduate students majoring/double majoring in English language education at Korea University, also took part in the experiment in partial requirement of a course credit. They ranged in age between 19 and 29 (mean: 22.1). Their self-reported English proficiency was at an upper-intermediate or an advanced level. Further, twenty-four native Japanese listeners, recruited from Korea University and Hankuk University of Foreign Studies, participated in the experiment. They were, either short-term exchange students or studied Korean at the universities for one or two semesters. They ranged in age between 18 and 29 (mean: 23). They were paid for their participation. The Japanese students' self-reported English proficiency was at a low or low-intermediate level and most of them had difficulty communicating in English. Four students' data were excluded from the final analysis due to their high error rates (i.e. above chance level).

Stimuli

The materials consisted of 56 nonce words, 28 disyllablic and trisyllabic words respectively, and they were created based on English words. Two hundred and twenty four pairs, 112 identical (e.g., *kolite-kolite*, *klite-klite*) and 112 non-identical pairs (e.g., *kolite-klite*, *klite-kolite*) resulting from initial schwa deletion, were constructed from 56 nonce words based on Lee (2011). In addition, there were 42 English words, which consisted of 22 disyllabic and 20 trisyllabic words. Similar to nonce words, 84 identical (e.g., *police-police*, *plice-plice*) and non-identical pairs each (e.g., *police-plice*, *plice-police*) were created from 42 words, totaling 168 pairs. However, only nonce words were analyzed in the experiment and English words were used as fillers. Importantly, test materials were constructed in such a way that both attested and unattested onsets were created as a result of initial schwa deletion; half of the materials had attested onsets in English (e.g., *darole-<u>dr</u>ole*, *selester-<u>sl</u>ester*, *galimpic-<u>gl</u>impic*) whereas the other half had unattested onsets.

Unattested onset sequences created by schwa deletion were further divided into 3 categories based on the sonority profiles of the sounds in the onset: Onsets with small sonority rises (e.g., *tommand-<u>tm</u>and*, *donanza-<u>dn</u>anza*), onsets with sonority levels (e.g., *badelle-<u>bd</u>elle*, ketansic-<u>kt</u>ansic), and onsets with sonority falls (e.g., *ratoon-<u>rt</u>oon*, *nofetic-<u>nf</u>etic*).

The test materials were recorded by a phonetically-trained male American English speaker from Michigan, U.S. He produced the materials naturally three times in the carrier sentence "This is ____________" in a sound-proof laboratory booth. Schwa-deleted monosyllabic and disyllabic nonce words and words were created by excising a schwa from their matching disyllabic and trisyllabic counterparts at the zero-crossings (Berent et al., 2009; Lee, 2011). Because schwa is mainly recognized by its F1 and F2 (Flemming, 2006; Gay, 1978), the beginning and ending points of the schwa were inspected using both waveform and spectrogram. The stimuli were arranged in 3 blocks matched for the test condition (words/nonce words × attested/unattested × identity/non-identity × number of syllables) and either the identical or the non-identical test item appeared within the same block for each stimulus (see Appendix).

Procedure

The experiment was run using E-prime 2.0. Participants sat in front of a computer screen wearing over-the-ear headphones in a sound-attenuated room. Each trial started with a fixation (*) and participants were instructed to press the space bar to initiate the trial. Two auditory stimuli were presented in sequence with an onset-asynchrony of 1500ms, as in Berent et al. (2007). Participants were asked to determine whether the two stimuli were identical or not by pressing the 1 key for "identical" responses and the 2 key for "non-identical" responses. They were requested to press the corresponding key as quickly as possible. In order to help participants be familiarized with the task, 20 practice items were presented before the task. Response times were measured from the end-point of the second stimulus item and the inter-trial interval was 1000ms.

Results

Results of English licit and illicit onsets

One of the research questions posed in the present study was to find out whether native English, Korean, and Japanese listeners were able to distinguish between attested onset clusters (i.e. onset clusters with a large sonority rise and /s/+stop/nasal clusters) and unattested onset clusters (i.e. onset sequences with a small sonority rise, a sonority flat, or a sonority fall), resulting from schwa deletion in English.

Results of English listeners

Trials with identical pairs were more accurate (M=92.2%) and faster (419ms) than those with non-identical pairs (M=64.6%, M=429ms). However, the study is mainly concerned with responses to non-identical trials. Hence, only the results of non-identical items were analyzed and also correct answers deviating 2.5 SD from the mean were eliminated from the final analysis of RT (0.6% of the correct answers), similar to Berent et al. (2007). Mean accuracy (AC) and response time (RT) for non-identical items are provided in Table 1, as a function of legitimacy of onset clusters and the number of syllables.

Table 1. Mean AC and RT of non-identical trials for English listeners as a function of onset legitimacy and number of syllables

Input type	AC (% correct)		RT (ms)	
	Attested onsets	Unattested onsets	Attested onsets	Unattested onsets
Monosyllables	81.7	52.4	384	484
Disyllables	75.6	48.2	417	430
Total	78.8	50.3	401	457

Accuracy rates and latency were fit to a generalized linear mixed model with phonotactic constraints and syllable as fixed effects and participants as a random effect. The results from accuracy data revealed that there was a main effect of English phonotactic constraints ($F(1, 84)=48.213$, $p<.001$) but the effect of syllable or the interaction between the factors was not significant (all, $p>.05$). Specifically, the difference between attested and unattested onsets was significant for both schwa deleted monosyllabic ($t(84)=5.538$, $p<.001$) and disyllabic inputs ($t(84)=4.988$, $p<.001$). However, the results of response latency indicated that there were no main effects of the factors or an interaction (all, $p>.05$) although attested onsets elicited shorter response time than unattested onsets (401ms vs. 457ms). Consequently, the results revealed that native English listeners were able to distinguish between attested and unattested onset clusters resulting from schwa deletion, indicating that they have knowledge of English phonotactic constraints on onset sequences.

Table 2. Mean AC and RT of non-identical trials for English listeners as a function of onset markedness and number of syllables

Onset type	Monosyllables		Disyllables		Total	
	AC	RT	AC	RT	AC	RT
Rise	81.8	374	73.9	311	78.1	343
Level	59.1	455	60.8	310	59.9	382
Fall	36.4	463	30.7	435	33.5	449

Another point of interest in the present study was to find out whether native English listeners were sensitive to the sonority-based markedness of onset clusters. For this purpose, only unattested onset clusters created by schwa deletion were further analyzed. As mentioned earlier, the unattested onset clusters were divided into 3 categories based on the sonority profiles of the consonant sequences:

Onsets with a small sonority rise, a sonority level, and a sonority fall. Among the 112 pairs, 56 pairs were identical trials and another 56 were non-identical trials. Mean AC and RT for identical trials were much higher (M=94.8%) and faster (M=373ms) than non-identical trials (M=57.2%, M=391ms). However, only non-identical trials were analyzed in the present study. Mean AC and RT were analyzed in terms of onset markedness and the number of syllables, as given in Table 2 (1.6 % of correct answers were eliminated from the final analysis of RT).

The results of a mixed models analysis on response accuracy revealed that only the effect of onset sonority was significant ($F(2, 126)=38.060$, $p<.001$). Pairwise comparisons between onset clusters of different sonority values were all significant for schwa deleted monosyllabic stimuli (all $p<.001$). The difference between onsets with sonority rises and falls, and the difference between onsets with sonority levels and falls were significant for schwa deleted disyllabic stimuli (all $p<.001$). The results of response latency also yielded significant main effects of onset markedness ($F(2, 126)=5.423$, $p=.006$) and syllable ($F(1, 126)=8.766$, $p<.005$). As for the onset markedness, the differences between sonority rises and falls and between levels and falls were significant ($p=.001$, $p<.05$, respectively). The main effect of syllable was because schwa excised disyllabic stimuli were responded much faster than schwa excised monosyllabic stimuli ($t(126)=2.961$, $p=.004$).

Therefore, the results indicated that native English listeners tended to perceive schwa deleted forms as identical to their vowel intact counterparts as the sonority-based onset markedness increases, corroborating the findings of Berent et al. (2007; 2008; 2009).

Results of Korean listeners

Similar to the results of the English listeners, identical trials were more accurate (M=92%) and faster (M=345ms) than non-identical trials (M=60.9%, M=382ms). Table 3 provides mean AC and RT for non-identical trials (1.2% of correct answers were eliminated in the final RT analysis).

Table 3. Mean AC and RT of non-identical trials for Korean listeners as a function of onset legitimacy and number of syllables

Input type	AC (% correct)		RT (ms)	
	Attested onsets	Unattested onsets	Attested onsets	Unattested onsets
Monosyllables	70.5	55.0	360	422
Disyllables	64.7	53.0	372	372
Total	67.7	54.0	366	397

As shown in Table 3, items with attested onsets elicited more accurate and faster responses than those with unattested onsets. Statistical analyses on response accuracy revealed that only the simple effect of English phonotactics was significant ($F(1, 124)=17.362$, $p<.001$) for both schwa-excised monosyllabic ($t(124)=3.436$, $p=.0001$) and disyllabic ($t(124)=2.535$, $p<.05$) inputs. However, neither main effects nor an interaction was found significant for response latency (all, $p>.05$) although attested onset clusters elicited faster responses than unattested ones (366ms vs. 397ms). Thus, the results indicated that the Korean listeners seemed to have some knowledge of English phonotactic constraints on onset clusters, similar to the English listeners.

In order to find out whether the Korean listeners were also sensitive to the sonority-based onset markedness, the results on the unattested onsets were further analyzed in terms of the sonority contours of the onsets and the number of syllables. Mean response accuracy and latency for identical pairs elicited more accurate (M=92.7%) and faster (M=342ms) responses than non-identical pairs (M=60.1%, M=370ms). The results of non-identical pairs are presented in Table 4 (2.2% of correct answers were eliminated from the final analysis of RT).

Statistical analyses on mean AC and RT revealed that only the main effect of onset type was significant (AC ($F(2, 185)=53.441$, $p<.0001$); RT ($F(2, 186)=6.733$, $p=.002$)). Pairwise comparisons on the results of response accuracy indicated that the differences between onsets of different sonority values were all significant for both schwa deleted monosyllabic (all $p<.001$) and disyllabic items (all $p<.001$, or $p<.05$). As for the response latency, the differences between sonority rises and falls, and between levels and falls were significant ($p<.001$, $p<.05$, respectively). The same result was obtained for monosyllabic inputs while only the difference between rises and falls was significant for disyllabic inputs (all, $p<.05$). Thus, the results from the Korean listeners indicated that the Korean listeners' responses to marked onsets elicited more errors and slower responses than less marked onsets, similar to the English listeners.

Table 4. Mean AC and RT of non-identical trials for Korean listeners as a function of onset markedness and number of syllables

Onset type	Monosyllables		Disyllables		Total	
	AC	RT	AC	RT	AC	RT
Rise	82.0	325	80.9	303	81.5	314
Level	58.6	371	66.4	335	62.6	353
Fall	36.7	472	35.7	410	36.2	441

Results of Japanese listeners

Identical items elicited more accurate responses (M=87%) than non-identical items (M=49.4%) but RTs did not differ between identical (M=441ms) and non-identical trials (M=441ms). Response accuracy and latency for non-identical trials is presented in Table 5 (1.2% of correct answers were eliminated in the final analysis of RT).

Table 5. Mean AC and RT of non-identical trials for Japanese listeners as a function of onset legitimacy and number of syllables

Input type	AC (% correct)		RT (ms)	
	Attested onsets	Unattested onsets	Attested onsets	Unattested onsets
Monosyllables	59.3	46.8	432	466
Disyllables	47.7	43.8	427	441
Total	53.5	45.3	430	453

Statistical analyses on response accuracy indicated that only the main effects of English phonotactics and syllable were significant ($F(1, 76)=5.161$, $p<.05$; ($F(1, 76)=4.106$, $p<.05$, respectively). The effect of English phonotactics was because attested onsets elicited more accurate responses than unattested ones ($t(76)=2.282$, $p<.05$). Similarly, monosyllabic inputs were more accurately perceived than disyllabic inputs ($t(76)=2.034$, $p<.05$). However, the results on response time did not yield a significant main effect nor an interaction (all, $p>.05$) even though attested onsets educed faster responses than unattested ones (430ms vs. 453ms). Accordingly, Japanese listeners also seemed to have some knowledge of English phonotactic constraints on onset clusters as they displayed sensitivity to the distinction between attested and unattested onsets.

To examine whether Japanese listeners were sensitive to the sonority-based onset markedness, the results on unattested onsets were also analyzed with reference to the onset markedness and the number of syllables. Responses to identical pairs were more accurate (M=90.6%) and faster (M=437ms) than response to non-identical pairs (M=48.6%, M=456ms). The results of non-identical trials are presented in Table 6 (1.3% of correct answers were eliminated in the final RT analysis).

Statistical analyses on response accuracy and latency yielded a significant effect of onset type for response accuracy ($F(2,114)=27.034$, $p<.001$) but not for response latency ($p>.05$). Pairwise comparisons indicated that the differences between onsets of different sonority values were all statistically significant for monosyllabic (all $p<.05$ or $p<.001$) and disyllabic inputs (all $p<.001$ except the difference between rise and level). Unexpectedly, responses to monosyllabic onsets of falls were faster than responses to monosyllabic onsets of rises and levels.

Table 6. Mean AC and RT of non-identical trials for Japanese listeners as a function of onset markedness and number of syllables

Onset type	Monosyllables		Disyllables		Total	
	AC	RT	AC	RT	AC	RT
Rise	67.5	484	60.0	376	63.8	430
Level	51.7	539	52.5	411	52.1	475
Fall	32.9	456	27.1	468	29.9	462

The results of response accuracy revealed that the Japanese listeners also had a tendency to misperceive schwa-deleted forms as identical to their corresponding vowel-intact forms and this pattern becomes more salient as the sonority-based onset markedness increases. This suggests that the Japanese listeners were sensitive to the sonority contours of onset clusters, similar to the English and Korean listeners. As for response latency, the Japanese listeners were overall faster to onsets of sonority rises compared to more marked onsets, which implies that the results of response latency are partly in line with the results of response accuracy.

Effects of L1 and sonority contours of onsets in the perception of onset clusters

Effects of L1 on the perception of attested vs. unattested onset clusters in English

The results of previous section on attested vs. unattested English onsets, given in Figures 1 and 2, were further analyzed to examine the effect of the listeners' L1s on the perception of the onset clusters. A 3 language (English × Korean × Japanese) × 2 phonotactics (attested vs. unattested) × 2 syllable type (monosyllable vs. disyllable) mixed models analysis on response accuracy revealed that there were significant main effects of language ($F(2, 284)=17.502$, $p<.001$), phonotactics ($F(1, 284)=63.459$, $p<.001$), and syllable type ($F(1, 284)=6.980$, $p=.009$). A two-way interaction between language and phonotactics was also significant ($F(2, 284)=8.654$, $p<.001$). Pairwise comparisons indicated that the differences between the following listener groups were all significant for attested onsets: English vs. Korean, English vs. Japanese, Korean vs. Japanese (all $p<.001$). Interestingly, only the difference between the Korean and Japanese listeners was significant for unattested onsets ($p<.05$). The significant effect of language was because the English and Korean listeners outperformed the Japanese listeners (all $p<.001$). Accordingly, the results indicated that the listeners' perception of English attested and unattested onsets was modulated by their L1s to some extent. The results also seem to suggest that the listeners' English proficiency may have played a role in accounting for the difference between attested and unattested English onsets to some extent given that the Korean listeners were advanced level learners of English unlike the Japanese listeners whose English proficiency was rather low. Further, the simple effect of phonotactics was due to the fact that attested onset clusters were more accurately perceived than unattested ones across all the listener groups ($t(284)=8.166$, $p<.001$), indicating that all the listeners had some knowledge of English phonotactic constraints on onset sequences even though the degree of their knowledge varied depending on their linguistic backgrounds. In addition, the effect of syllable was significant because monosyllables were better identified than disyllables ($t(284)=2.649$, $p=.009$). However, a 3 language × 2 phonotactics × 2 syllable type mixed models analysis on response time indicated that none of the factors or interactions were statistically significant.

Effects of L1 and sonority contours on the perception of unattested onset clusters

The results of the unattested onset clusters were further analyzed in terms of the listeners' L1s and sonority profiles, as shown in Figures 3 and 4. A 3 language (English × Korean × Japanese) × 3 sonority contour (rise vs. level vs. fall) × 2 syllable type (monosyllable vs. disyllable) mixed models analysis on response accuracy yielded only significant main effects of language ($F(2, 425)=10.929$, $p<.001$) and sonority contour ($F(2, 425)=111.785$, $p<.001$). Planned comparisons on the effect of language revealed that the contrasts between the English and Japanese listeners and also between the Korean and Japanese listeners were significant ($p=.002$, $p<.001$, respectively) due to the English and Korean listeners' outperforming the Japanese listeners. As for the effect of sonority, the contrasts between onsets of different sonority values were all significant across the listener groups (all $p<.001$

except the difference between sonority rises and levels for the Japanese listeners (p<.05)). Accordingly, the results also corroborate the findings of Berent et al. (2007; 2008; 2009) on the sonority-based grammatical structure of onset clusters.

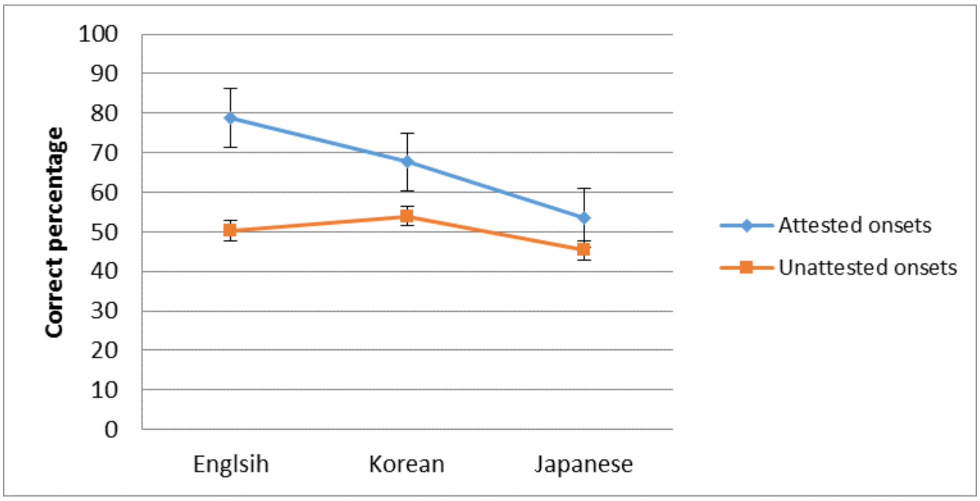


Figure 1. Mean percentage correct (%) of attested vs. unattested onset clusters by listener group

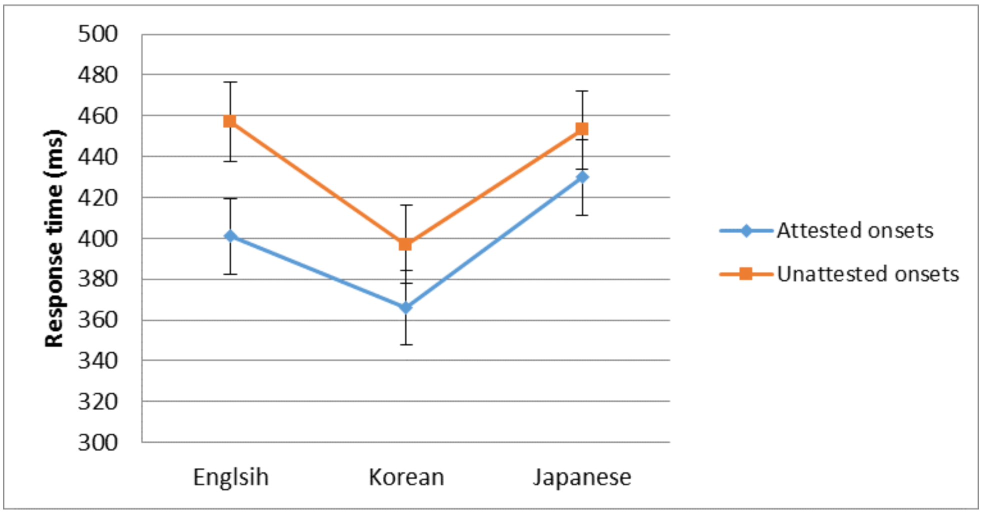


Figure 2. Mean reaction time (ms) of attested vs. unattested onset clusters by listener group

Further, statistical analyses on response latency revealed that the simple effects of language, sonority contour, and syllable type were all significant but the interactions were not. Specifically, the main effect of language (F(2, 426)=5.863, p=.003) was because the English and Korean listeners were much faster to respond than the Japanese listeners (p<.05, p=.001, respectively). The simple effect of sonority contours (F(2, 426)=5.780, p=.003) was because the difference between sonority rises and falls was significant (p=.001). The effect of syllable type was also significant (F(1, 426)=9.260, p=.002) as latency of disyllables was faster than that of monosyllables (373ms vs. 438ms).

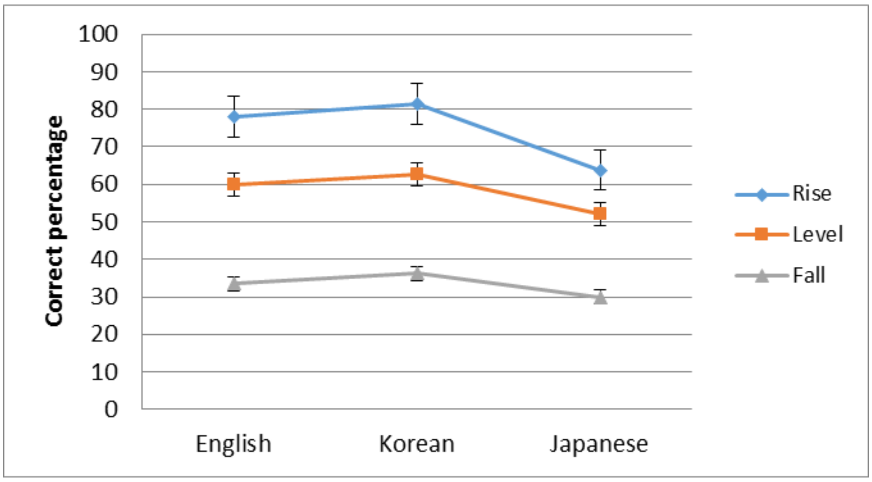


Figure 3. Mean percentage correct (%) of unattested onset clusters by sonority contour and listener group

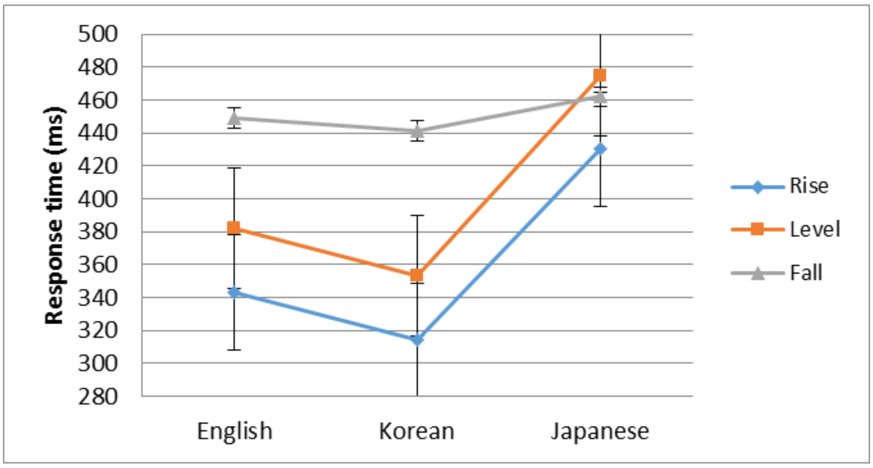


Figure 4. Mean reaction time (ms) of unattested onset clusters by sonority contour and listener group

To sum up, the results of unattested onset sequences seem to suggest that the listeners have some inherent or universal knowledge of the sonority-based markedness of onsets, because the listeners in the present study showed some sensitivity to the sonority contours of onsets regardless of their language backgrounds. Further, the Korean listeners were slightly more accurate and faster to respond than the English listeners, which was somewhat unexpected. This may be ascribable to the fact that the Korean listeners participated in the experiments as partial requirement of their course credit unlike the English listeners and thus they may have tried to do their best in the experiments.

General discussion

Berent et al. (2007; 2009) reported that English listeners were sensitive to the grammatical structures of onset clusters, such that they tended to show more perceptual illusions for universally disfavored onsets of a sonority fall (e.g., *lb*) than relatively favored onsets of a sonority rise (e.g., *bn*). Berent et al. (2008) also reported that Korean listeners showed similar patterns to English listeners with respect to universally dispreferred vs. preferred onsets, thus arguing for the universality of grammatical structures of onset clusters. However, Berent et al. (2007; 2009) admitted that listeners' experience with their L1 was partly ascribable to their bias towards preferred onsets because Russian listeners tended to perceive disfavored onsets such as *lb* correctly on most trials due to the existence of such clusters in Russian.

The paper explored whether native English, Korean, and Japanese listeners were sensitive to the attested vs. unattested onset clusters in English derived from word-initial schwa deletion, and whether they displayed perceptual sensitivity to the sonority profiles of unattested onsets. In order to seek for answers to the questions, the study conducted identity judgment tasks on 22 English, 32 Korean, and 20 Japanese listeners using English nonce word stimuli.

The results revealed that the listeners' perception of English attested and unattested onset sequences was modulated by their knowledge of English phonotactic constraints as well as by their L1s. More specifically, all the listener groups were sensitive to the difference between attested and unattested onsets in terms of response accuracy, which implies that the listeners had some knowledge of English phonotactic restrictions on onset clusters. But their perception of onsets was also influenced by their language backgrounds, as the English listeners outperformed the Korean and Japanese listeners. Further, the results seem to indicate that the listeners' English proficiency may have influenced the perception of attested and unattested onsets in English to some degree. This was because the Korean listeners' response accuracy was higher than that of the Japanese listeners, which is partly attributed to the Korean listeners' high English proficiency compared to the Japanese listeners'. Also, listeners were overall more accurate in perceiving monosyllabic inputs than disyllabic ones, thus showing the effect of the number of syllables of the stimuli.

Importantly, the results of the sonority profiles of onset sequences corroborated the findings of Berent et al. (2007; 2008; 2009) because all the listener groups had a propensity to misperceive schwa-deleted forms as their vowel intact counterparts as a function of the markedness of onset sequences regardless of their language backgrounds or their English proficiency. That is, all the listener groups' responses to onsets of sonority rises educed fewer errors than onsets of sonority levels, which in turn elicited fewer errors than onsets of sonority falls. The listeners' response latency was also in line with the grammatical structures of onset sequences, as less marked onsets overall tended to elicit faster responses compared to more marked onsets. Therefore, the results seem to add another segment to the findings of Berent et al. (2007; 2008; 2009) on the relationship between perceptual illusion and the sonority-based onset markedness.

Peperkamp (2007) argued that Berent et al.'s (2007) findings of the sonority-based markedness of onset sequences may be due to the phonetic properties of such clusters rather than universal structures of onset sequences. For example, Peperkamp claimed that schwa could be more strongly co-articulated with a following nasal, thus causing listeners to misjudge words such as *benif* as monosyllabic rather than disyllabic. In a similar vein, Wright (2008) casted some doubts on the Sonority Sequencing Principle (Selkirk, 1984) and claimed that it should be reformulated as a perceptually motivated constraint, in which the robustness of encoding of phonetic cues for speech perception is maximized through "redundancy of cues, the auditory impact of cues, the perceptual distance between cues, and the resistance of cues to environmental masking" (2008, p. 35). For instance, he contended that the most common syllable patterns like CGV and CLV can be derived on the basis of the robustness of perceptual cues such as auditory nerve fiber boost, increased redundancy and perceptual distance, etc. (Wright, 2008, p. 51).

However, the results of the present study also corroborate the findings of Berent et al (2007; 2008; 2009) on the universally favored vs. disfavored onset clusters, rooted in the markedness hierarchy since the Korean and Japanese listeners showed the same pattern as the native English listeners with respect to the sonority profiles of unattested onset sequences even, when the onsets were derived from initial schwa deletion in English. Accordingly, the listeners' sensitivity to the sonority profiles of onsets does not seem to lie entirely on the phonetic differences of the given stimuli.

Additionally, as mentioned earlier, Lee (2011) investigated the perception of English onsets created by initial schwa deletion among native English, Korean, and Japanese listeners, employing English nonce words. Her results showed that only English listeners (neither Korean nor Japanese listeners) were able to distinguish between English legal and illegal onset clusters. Korean and Japanese listeners were also only partly sensitive to the sonority-based onset markendess, as they showed high accuracy rates and faster latencies to onsets with sonority plateaus than to onsets with sonority rises, unlike English listeners who complied with the grammatical structures of onset sequences. The different results from Lee and the present study may be partly attributable to different task types employed in Lee's study and the present study. More specifically, Lee conducted syllable judgment tests whereas the present study conducted identity judgment tests.

The impact of task types on the perception of speech sounds has been discussed by many scholars. For example, Depoux et al. (1997) found that French listeners were able to identify the location of stress in an AX discrimination test but they were poor at spotting the position of stress in an ABX discrimination test, where three stimulus items were uttered by different talkers. This was because the ABX test is more challenging than the AX task in terms of phonetic variability as well as a short-term memory load (Depoux, Peperkamp, & Sebastián-Gallés, 2001). The effects of task types in experimental results were also much discussed in Strange and Shafer (2008).

Thus, the findings of the current study, that native English, Korean, and Japanese listeners displayed sensitivity to the sonority-based onset markedness as well as to the distinction between attested and unattested onsets in English, seem to suggest that the different task types employed in the present study and Lee's study are partly responsible for the different results obtained between the two studies. However, we cannot exclude the possibility that the different results may partly be due to the listener variation, which deserves further research.

Finally, the fact that the English listeners overall performed better than the Korean and Japanese listeners with respect to the distinction between licit and illicit onset clusters in English seems to indicate that the English listeners may have benefited from their experience with schwa-deleted variants as well as schwa-intact inputs. That is, English listeners often delete the initial schwa in casual speech and hence they are more likely to be familiar with schwa-deleted variants of lexical items due to their exposure to such forms in their daily life unlike Korean or Japanese listeners. Accordingly, the English listeners in the present study may have had less difficulty perceiving schwa-intact vs. schwa-deleted forms given that the nonce word stimuli were constructed based on English words. This seems to imply that variants of lexical items such as schwa-deleted forms should be included in listeners' lexical representation, along with the canonical forms of the lexical items, as argued by Patterson et al. (2003). This further supports the arguments made by exemplar theory (Pierrehumbert, 2001) concerning the lexical representation of words.

However, whether human beings possess universal structures of onset markedness or whether such universal structures are rooted in the robustness of perceptual cues and in some other factors, remains to be seen. Consequently, more research should be done to elucidate the role of the sonority-based markedness of onsets in speech perception.

References

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*, 591-630.

Berent, I., Lennertz, T., Jun, J., Moreno, M., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, *105*(14), 5321-5325.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaun, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. Phonology, *26*, 75-108.

Best, C. A direct realist view of cross-language speech perception. (1995). In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171-204). Baltimore: York Press.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. Munro (eds.), *Language experience in second language speech learning: In honor of J. Flege*, (pp. 13-34). Amsterdam/Philadelphia: John Benjamins.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (eds.), *Papers in laboratory phonology I* (pp. 283-333). Cambridge, UK: Cambridge University Press.

Depoux, E, Pallier, P., Sebastián-Gallés, N., & Mehler, J. (1997). A distressing 'deafness' in French? *Journal of Memory and Language, 36*, 406-421.
Depoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1568-1578.
Depoux, E, Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress deafness. *Journal of the Acoustical Society of America, 110,* 1606-1618.
Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-277). Baltimore: York Press.
Flemming, E. (2006). Contrast and the realization of schwa vowels in English. Slides from a presentation. http://web.edu/flemming/www/.
Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech, 21,* 347-353.
Gierut, J. (1999). Syllable onsets: Clusters and adjuncts in acquisition. *Journal of Speech, Language and Hearing Research, 42*, 708-726.
Halle, M. (1971). *The sound pattern of Russian*. The Hague: Mouton.
Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance, 24,* 592-608.
Heffner, R-M. S. (1950). *General phonetics*. Madison: The University of Wisconsin Press.
Hammond, M. (1999). *The phonology of English: A prosodic optimaltiy-theoretic approach.* Oxford, UK: Oxford University Press.
Jusczyk, P., Luce, P., & Charles-Luce, J. (1994). Infants' sensitivity of phonotactic patterns in the native language. *Journal of Memory and Language, 33,* 630-645.
Ladefoged, P. (2006). *A course in phonetics*. Boston, MA: Blackwell.
Lee, S. (2011). The perceptual interaction between schwa deletion and English phonotactic constrains. *Korean Journal of Linguistics, 36* (3), 699-733.
Massaro, D., & Cohen, M. (1983). Phonological context in speech perception. *Perception and Psychophysics, 34*(4), 338-348.
Messer, S. (1967). Implicit phonology in children. *Journal of Verbal Learning and Verbal Behavior, 6,* 609-613.
Ohala, D.K. (1999). The influence of sonority on children's cluster reduction. *Journal of Communication Disorders, 32,* 397-421.
Patterson, D., LoCasto, P., & Connine, C. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica, 60,* 45-69.
Peperkamp, S. (2007). Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin, *Cognition, 104,* 631-637.
Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and emergent grammar* (pp. 137-157). Amsterdam: John Benjamins.
Pitt, M. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics, 60*(6), 941-951.
Segui, J., Frauenfelder, U., & Hallé, P. (2001). Phonotactic constraints shape speech perception: Implications for sublexical and lexical processing. In E. Dupoux (ed.), Language, *brain, and cognitive development* (pp. 195-208). Cambridge, MA: MIT Press.
Selkirk, E. (1982). The syllable. In N. Smith (ed.), *The structure of phonological representations* (pp. 337-383). Dordrecht, Holland: Fortis P.
Selkirk, E. (1984). On major class features and syllable theory. In M. Aranoff & R.T. Oehrle (eds.), *Language and sound structure: Studies in phonology presented to Morris Halle by his teacher and students* (pp. 107-36). Cambridge, MA: MIT Press.
Strange, W., & Shafer, V. (2008). Speech perception in second language learners. In J. H. Edwards & M. Zampini (eds.), *Phonology and second language acquisition* (pp. 153-91). Amsterdam: John Benjamins.
Wright, R. (2008). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner & D. Steriade (eds.), *Phonetically based phonology* (pp. 34-57). Cambridge, UK: Cambridge University Press.
Yavaş, M. (2006). *Applied English* phonology. West Sussex, UK: Willey Blackwell.

Appendix

Nonce words which conform to English phonotactic constraints on onset clusters

Disyllabic words	Trisyllabic words
kolite	senocious
kolice	sonenzic
darole	perrific
darade	poronto
pollapse	terrelic
pollect	torilla
torrect	selester
tareer	selantic
galoon	karestic
gelieve	karula
samine	berocious
semect	berentic
sekkose	galunster
saccort	galimpic

Nonce words which violate English phonotactic constraints on onset clusters

tommand[tm]	konato[kn]
tommute	konorrow
panoe[pn]	denanna[dn]
panal	donanza
patoo[pt]	topater[tp]
patet	topastic
badelle[bd]	ketender[kt]
bagaar[bg]	ketansic
nekotic[nk]	nafamic[nf]
nekkempt	nofetic
mappell[mp]	mapestic[mp]
mapeck	mapecian
ratame[rt]	romina[rm]
ratoon	romelic

On the permeability of German-Spanish bilinguals' phonological grammars

Conxita Lleó
lleo@uni-hamburg.de

University of Hamburg

Abstract. This article focuses on the development of seven German-Spanish simultaneous bilingual children's phonological grammars, compared to two control groups: a Spanish group of three monolinguals, and a German group of five monolinguals. The bilingual children, growing up in Germany, were exposed to the two languages from birth, under similar conditions: the mothers, who in the first years were the main caretakers, were native speakers of Standard Peninsular Spanish (5), Mexican Spanish (2), or Chilean Spanish (1); the fathers were native speakers of Northern German. All children but one began attending a kindergarten at 3;0. The domains observed present phonological differences between German and Spanish. The hypotheses considered are based on Paradis and Genesee (1996)'s predictions. Here, we will report on results in Spanish, where we have found: acceleration (syllabic codas), a slight delay (VOT, unfooted syllables), transfer (spirants, place assimilated nasals), as well as a different order of acquisition (depending on types of prosodic words) and fusion (e.g., in the VOT of one child). The observed effects cannot be explained on the basis of interfaces because all phenomena analyzed involve various interfaces, and should be affected in a similar way. They cannot be explained by (in)balance, either, as they may appear both, in a balanced as well as a non-balanced condition. It is proposed that Optimality Theory offers the most explanatory analysis of the results: According to such analysis, markedness constraints are dominant at first, and are demoted soon thereafter due to frequent violations. The acceleration found in the bilinguals' coda production is thus due to the frequent violations of the NOCODA constraint that the child experiences in target German, i.e. due to the joint additive effect of the two languages. Transfer, on the other hand, is due to the outranking position that German attributes to the UNIFORMITY constraint, which prevents lexical items from alternating.

Keywords: bilingualism, cross-language interaction, hierarchy of constraints, optimality theory, phonological acquisition, prosodic structure

Introduction

Babatsouli and Ingram (2015, p. 173) refer to the enormous task involved in the study of bilingual acquisition, given the great number of variables to take into consideration: "the extensive number of possible language combinations...the [various] bilingual contexts... simultaneous or dual acquisition... if the latter, which language is acquired earlier... the contexts of the input... design issues concerning the number of necessary participants... whether to conduct case studies or cross sectional studies... the various components of language... and the range of research questions," and they add: "When the math is done, the number of possible bilingual studies approaches one million." Certainly this conclusion is discouraging, if taken on the pessimistic side, or it can be received optimistically, as a challenge to begin disentangling a minimum number of variables, which will allow us to look for some further variables, etc.

Some researchers focus on structural variables, whereas others prefer to struggle with sociolinguistic ones. Thomason and Kaufman (1988, p. 35), for instance, declare: "The starting point for our theory of linguistic interference is this: it is the sociolinguistic history of the speakers, and not the structure of their language that is the primary determinant of the linguistic outcome of language contact." Although sociolinguistic as well as other types of variables may be essential, the aim of this article is to consider the relevance of some structural and psycho-linguistic factors to be able to unify them with sociolinguistic ones. The ultimate goal of the article is to explore possible ways to predict the outcome of various bilingual constellations. For instance, if one of the languages is trochaic and the other one iambic, what will the simultaneous bilingual's stress systems of these languages look like, i.e. if there is interaction between the languages, in what direction will the influence go? Or if one of the languages is syllable-timed and the other one is stress-timed, how will the learner's rhythmic structures of these languages look like, i.e. will there be interaction, and in what direction?

There are not enough studies trying to systematize cross-language interaction of a bilingual child acquiring two languages simultaneously. The best-known proposal, often referred to in the field, is Paradis and Genesee (1996, p. 3). They proposed three potential manifestations of interaction, which they call “inter-dependence”: *Delay*, *Acceleration* and *Transfer*. In order to test the presence of these phenomena in their bilingual data, they observed three French-English bilinguals at 2;0, 2;6, 3;0, growing up in Montreal. In the Syntax (finiteness, negation and pronominal subjects), they found:

- No delay
- No acceleration
- No transfer

However, they did not observe other areas of grammar than syntactic ones, and in fact the question about the presence of these phenomena in Phonology is legitimate and overdue. Thus, the main research question of this article is whether the simultaneous acquisition of two languages by the child involves cross-language interaction; and if there is interaction, whether it manifests itself by delay, acceleration, or transfer.

Fabiano-Smith and Goldstein (2010) reviewed Paradis & Genesee's proposal and modified the term "delay", substituting *deceleration* for it, as the latter does not evoke negative associations, which the former does. Another aspect they deal with is the notion of Transfer, which is restricted to segmental Transfer: segments of one language are used in the other language (e.g., German /R/ for Spanish /ɾ/ and /r/). Further additions to the Paradis and Genesee (1996)'s types of interaction have been proposed in Lleó (2006) and Queen (2001). Lleó (2006) proposes a *Different Order of Acquisition*: Two categories that are acquired in the order A then B by monolinguals may be acquired as first B then A by bilinguals in that same language. Queen (2001) proposes the notion of *Fusion* for intonation, which implies a two-way influence between two languages. Her example indicates that bilingual children use two rises both in German and Turkish, emerging from the interaction between the two languages: L*HH% (similar to a German contour), and L%H% (similar to a Turkish contour). This is comparable to bi-directional Transfer.

With these additions to Paradis and Genesee in mind, we will consider five types of interaction, which will be exemplified with phenomena studied in a group of 7 German-Spanish bilingual children, and a control group of 3 Spanish monolinguals, as well as a control group of 5 German monolinguals (see Table 1).

Table 1. Children who participated in the project, with L1 Spanish and German, monolingual and bilingual

Spanish	Age	Acquisition	Environment	Children
L1	1;6-3;0	monolingual	Spain	3
L1	1;6-3;0	bilingual	Germany	7
German				
L1	1;6-3;0	monolingual	Germany	5
L1	1;6-3;0	bilingual	Germany	7

Thus, we will study five types of interaction, each one of them exemplified by a grammatical phenomenon of Spanish:

- Delay: Pre-tonic (unfooted) syllables in Spanish

- Order of acquisition: Prosodic Word structures
- Acceleration: Syllable codas in Spanish
- Transfer: voiced stops vs. spirants in Spanish
- Fusion: German and Spanish VOT from 2;0 to 2;6

Studies under consideration

All studies are based on the data collected within three projects, developed at the University of Hamburg: PAIDUS with data from 5 German monolingual children from Hamburg and 4 Spanish monolingual children from Madrid; PEDSES, with data from 3 German-Spanish bilingual children from Hamburg; and PhonBLA, with data from 4 German-Spanish bilingual children from Hamburg. The main difference between PEDSES and PhonBLA is that the former stopped as the children were 2;6 to 3;0, whereas PhonBLA continued until the latter were about 6 years old. For a detailed description of these and further related corpora, see Lleó (2012). Acoustic analyses were done with PRAAT (Boersma & Weenink, 2001).

Participants and Data

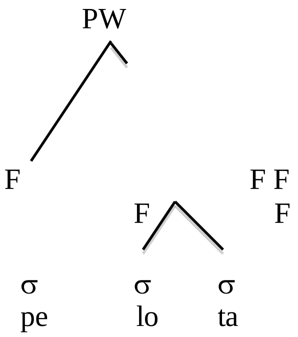
The main participants were the German-Spanish bilingual children and their monolingual controls introduced in Table 1. The bilinguals lived in North Germany (Hamburg), and had a native Spanish speaking mother (Peninsular Standard Spanish), and a native German speaking father (North Standard German). The German speaking controls were from the same German region as the bilinguals (North Germany), whereas the Spanish speaking controls were from the Madrid area (Spain), where they had been born and lived. The bilinguals were relatively balanced until about 3 years of age, with a slight dominance of German, the majority language, which became stronger at about the age of 3, as they began to attend a Kindergarten. All data had been collected in semi-spontaneous situations, playing with and talking to the child. Utterances were transcribed and introduced into a database: EXMARaLDA, developed by Thomas Schmidt and Kai Wörner, at the Research Center on Multilingualism of the University of Hamburg (SFB 538). Topics to study were selected on the basis of (similarities and) differences between German and Spanish.

Table 2. Examples of unfooted syllable production by monolingual children in Spanish

written word	meaning	Phon word	child form	child	child age
conejo	'rabbit'	/ko'nexo/	[no'nino]	José	(1;7,27)
			[to'leto]		(1;11,23)
pelota	'ball'	/pe'lota/	[ba'pɔta˘]		(1;9,2)
sombrero	'hut'	/som'bɾeɾo/	[ba'bɛlo]		(1;9,2)
zapato	'shoe'	/θa'pato/	[pa'papa]		(1;7,27)
			[pɐ'pɐto]		(1;9,2)
zapato	'shoe'	/θa'pato/	[ha'patʃa]	Miguel	(1;6,7)
bizcocho	'cookie'	/biθ'kotʃo/	[vi'kɔtʃɔ]		(1;7,26)
manzana	'appel'	/man'θana/	[pɑ'sanɑ]	María	(1;10,17)
			[pa'θanɑ]		(1;10,17)
trompeta	'trumpet'	/tɾom'peta/	[ha'badɑ]		(1;10,17)
			[bɔ'pitɑ]		(2;0,11)

Results

Delay: Percentages of unfooted syllable deletion in Spanish are much higher in bilinguals as compared to monolinguals.



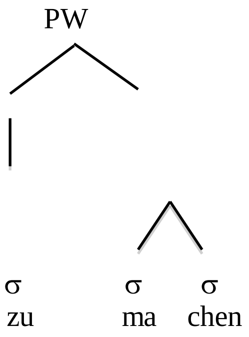


Figure 1a. Prosodic structure of Spanish pelota 'ball'

Figure 1b. Prosodic structure of German zumachen 'close'

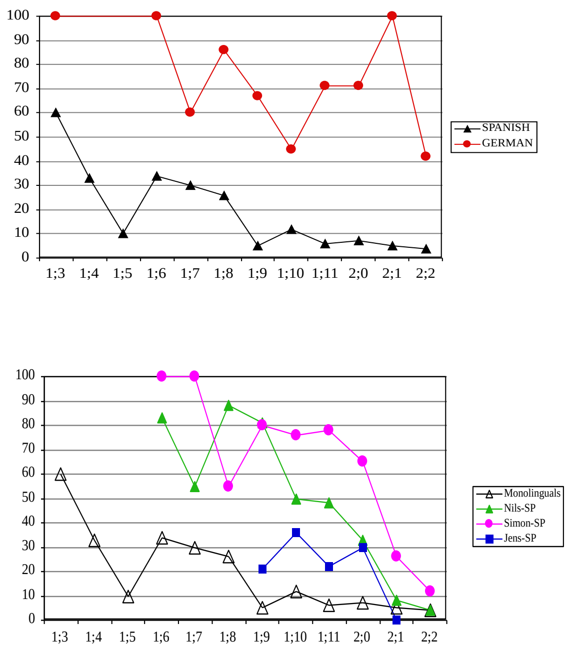


Figure 2. Percentages of unfooted syllable truncation by Spanish and German monolinguals, and by three bilinguals: Nils, Simon and Jens

Examples are shown in Table 2, as produced by three monolingual Spanish children. It has been argued (Lleó, 2002) that unfooted syllables fill a prosodically weak and thus vulnerable position, and are often deleted. Figure 1a shows the prosodic structure of the Spanish word *pelota* 'ball', comprised of the foot *lota* and the pretonic syllable *pe*, which is the unfooted syllable. Spanish has many trisyllabic words, also in child language, as evidenced by the examples of Table 2. German has also unfooted syllables, but far less in number than Spanish. The majority of trisyllabic words in German do not show the structure of 1a, but that of 1b, comprised of two feet. Spanish Monolinguals begin to produce unfooted syllables very early, whereas German children have much truncation until 2;0 and later. On the other hand, Figure 2 shows that whereas Spanish monolinguals (as a group) have reduced percentages of unfooted syllable truncation, German monolinguals' unfooted syllable truncation reaches very high percentages. Bilinguals on the other hand have about as much truncation of unfooted syllables in Spanish as German monolinguals until about 2;0.

These data have been described within the model of Optimality Theory, according to which, in order to produce one of these unfooted syllables, a constraint disallowing them must be violated. Such constraint is ALIGN LEFT, namely,

(1) ALIGN(PW, L, Ft, L): All PWs must have their left edge aligned with a Ft

On the other hand, in order to produce more than a foot (as the German children begin to do before producing unfooted syllables), a constraint disallowing prosodic structures longer than one foot must be violated. This constraint is ALIGN RIGHT, namely,

(2) ALIGN(Ft, R, PW, R): The right edge of all Ft must be aligned with the right edge of a PW

Taking into consideration that Align constraints are markedness constraints, ex-hypothesis they are both outranking at the initial stages of acquisition. In the following stage, these ALIGN constraints will be demoted: in Spanish ALIGN LEFT will be demoted, and in German ALIGN RIGHT. Both will be positioned below the faithfulness constraint, MAXIO. However, there will probably be some difference between monolingual and bilingual children, as for a while, the Spanish language, if weaker than the German language in the mind of the bilingual child, may miss demoting ALIGN LEFT and demote ALIGN RIGHT instead. That is, the child may behave in Spanish as if it were German.

Order of acquisition: Spanish Monolinguals produced trisyllables of the type as in Figure 1a (i.e. with a pre-tonic or unfooted syllable) before producing monosyllables, but German-Spanish bilinguals produced monosyllables in Spanish before producing trisyllables.

In Lleó (2006) it was shown that the order of acquisition in Spanish is related to the need to produce unfooted syllables, in the sense that the frequent production of multisyllabic words, especially the production of those words with unfooted syllables, leads to early demotion of LEFT ALIGNMENT, which makes the early production of unfooted syllables possible. Table 3 shows the development of Prosodic Word structures in Spanish by monolinguals and bilinguals, with indication of age at each different step. In these figures, feet are represented as sequences of syllables or as a single syllable (if heavy, i.e. if the rhyme is comprised of a long vowel or a vowel followed by a coda.

Table 3. Development of Prosodic Word Structures in Spanish by monolinguals and bilinguals

Age	Monolinguals
1;2-1;3	$['(\sigma\sigma)_{Ft}]_{PW}$
1;3-1;6	$[\sigma'(\sigma)_{Ft}]_{PW}$ $[\sigma'(\sigma\sigma)_{Ft}]_{PW}$
1;7	$['(\sigma)]_{PW}$
1;10-2;2	$[(\sigma\sigma)_{Ft}'(\sigma\sigma)_{Ft}]_{PW}$

Age	Bilinguals
1;2-1;4	$['(\sigma\sigma)_{Ft}]_{PW}$
1;5-1;6	$['(\sigma)]_{PW}$
1;8-1;10	$[\sigma'(\sigma)_{Ft}]_{PW}[\sigma'(\sigma\sigma)_{Ft}]_{PW}$
2;0-2;3	$[(\sigma\sigma)_{Ft}'(\sigma\sigma)_{Ft}]_{PW}$

The Tables show that in the case of monolinguals, the evolution of word structures goes from trochees, to amphibrachs (trisyllables consisting of a trochaic foot preceded by a pretonic or unfooted syllable) to monosyllables, whereas bilinguals also begin with trochees, and from there they go to monosyllables and later on to trisyllables. In fact, the order followed by the bilinguals is the same as the one followed by the German monolinguals (see Lleó, in press). There is a difference though between the preferred trisyllables in Spanish (as in Figure 1a), while trisyllables produced in German are preferably those with the structure in Figure 1b. For the child to produce unfooted syllables, ALIGNLEFT is relevant again, but bilinguals do not reach the threshold to demote ALIGNLEFT as soon as the monolinguals, given the reduced number of amphibrachs in child German. Quadrisyllables are the last prosodic structures to be acquired in the time span here observed (up to 2;2). Only Jens begins to produce some quadrisyllables at 2;0, whereas Nils and Simon begin at 2;3. (Note that the time spans separated by a hyphen in table 3 refer to word structures produced sooner by some child/children, and later by some other child/children).

One more aspect to note in relation to this form of interaction is that order of acquisition is a pure quantitative relation, which in fact is reducible to delay and acceleration: If monolinguals produce first A and then B, whereas bilinguals produce first B and then A, this also means that bilinguals show acceleration of B and delay of A.

Acceleration: Bilinguals acquire syllable codas in Spanish faster compared to Spanish monolinguals.

German has more than 60% of closed syllables, whereas Spanish has less than 30%. Closed syllables are marked, as they violate the NOCODA constraint, which bans coda production. The two factors, Markedness and Frequency, would seem to be in contradiction in German closed syllables, given that markedness is expected to cause delay, whereas highly frequent production (e.g., of codas) may lead to the opposite result. A look to the percentages of coda production brings following results (see Figure 3): Monolingual German children produce more than 80% of codas after 1;11, but monolingual Spanish do not yet produce 50% of codas at 2;4. Bilinguals produce more than 50% of codas in Spanish after 1;9. Because both languages have codas, an additive effect takes place. Besides, codas are very frequent in German, thus, a frequency effect takes place, too. It is also important to note that coda production by bilinguals in German does not differ from that of monolinguals. Whereas from the point of view of universals, marked categories are expected to be acquired later than unmarked ones, the effect can also be the opposite, for the following reason. CVC syllables are marked compared to CV syllables. However, in OT, Frequency in conjunction with Markedness lead to early demotion of the NOCODA constraint. That is, the NOCODA constraint is often violated, and in Bilingualism frequent violation has the side effect of acceleration, the result being that all bilinguals produce more codas than monolinguals.

Transfer: Bilinguals substitute voiced stops for spirants in Spanish, after a period in which both monolinguals and bilinguals produced many spirants.

Spanish has an obligatory process of spirantization, by which the voiced stops /b, d, g/ are produced as such in absolute initial position after nasal consonants, and after other non-continuant consonants; but after vowel or continuant consonants, they are produced as continuants or spirants: [β, ð, ɣ]. Figure 3 shows the percentages of target-like spirantization by monolinguals and bilinguals. Percentages by three monolinguals of the database PAIDUS (Jose, Maria and Miguel) are presented as a group, while bilingual values are shown individually for three children: Jens, Nils and Simon. Monolinguals produce high percentages of spirants from the earliest stage, whereas bilinguals have high percentages of continuants at first, but after 2;6 they mainly produce stops. Replacement of stops for spirants could be due to markedness (stops are unmarked) or to Transfer (from German). Given the regression that begins at about 2;6 in the case of bilinguals, the avoidance of spirants in Spanish is rather interpretable as due to Transfer from German into Spanish. The question then arises as to what is being transferred.

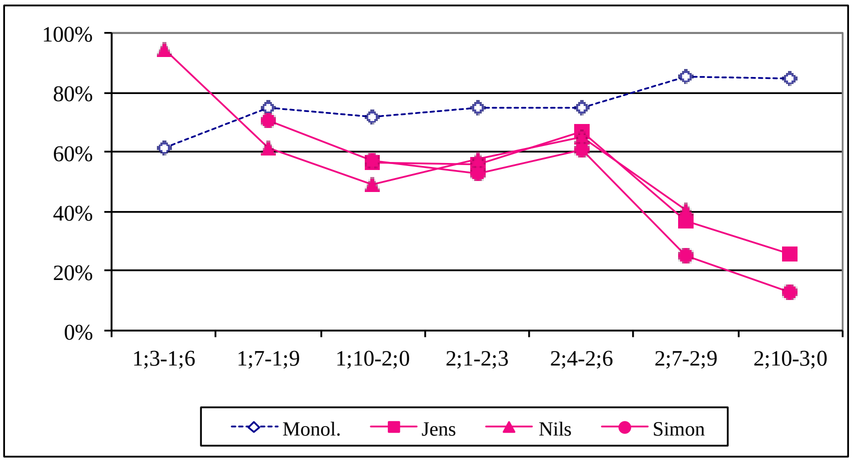


Figure 3. Percentages of target-like spirantization by three monolingual Spanish children as a group and three German-Spanish bilinguals, individually

The most straightforward answer to this question has been based on the classical proposal that considers phonological segments as the target for transfer. Obviously, here segments are not transferred, because Spanish already has voiced stops in its inventory. Is it then the occurrence of a certain type of segment (here, voiceless stops) in a certain type of environment (here, following a continuant segment, as e.g., a vowel, /s/, /ɾ/, etc.) that is being transferred? Some clarification about my use of the notion Transfer may be needed here. Any category or phenomenon going from language A to language B constitutes Transfer, if it enters language B anew. The notion of Transfer was born and was maximally used in the era of structuralism, which used to deal with tangible categories. That is why *segments* were the targets of transfer (Gass & Selinker 1983). Later on, with the advent of Generative Phonology, *rules* were transferred (e.g., the rule of Glottal Stop Insertion [GSI, Wiese 1996: 173; Lleó & Vogel 2004: 82] in the L2 Spanish produced by L1 speakers of German), and this meant making a large step into abstraction. More recently, in times of OT, abstraction is going even farther and hierarchies of constraints can be targeted by transfer, as well.

(3) GSI: Insert [+ constricted glottis] / $_F$[___ [- consonantal]

Both the Spanish spirantization as well as the lack of it in German can be accounted for by means of two constraints:

(4) UNIFORM EXPONENCE (UE): A lexical item is invariable for property P.

(5) AGREE$_{[cont]:}$ After a vowel (or a [+cont] segment, voiced obstruents are [+cont].

Both constraints are present in both languages, but whereas Spanish abides by AGREE[cont], German does not and keeps the voiced stops without modifying them independently of the environment. This means that the two languages have different hierarchies of constraints. The Spanish hierarchy has an outranking AGREE, which favors assimilation thus spirantization, whereas the German one has an outranking UE, which keeps the underlying form unchanged.

(6) Spanish hierarchy: AGREE $_{[cont]}$ >> UE

(7) German hierarchy: UE >> AGREE $_{[cont]}$

Bilinguals in Germany at about 2;6 adopt the German grammar in Spanish, i.e. the German hierarchy, with regard to lack of assimilation. Note that the notion of Transfer is adaptable to different theories, as mentioned above. Nowadays, within OT, constraints are the building stones of grammar, which builds a hierarchy in each language. Thus, it is this order of constraints, alias Grammar, that is being transferred, bringing with it an ever greater degree of abstraction for the concept of Transfer.

***Fusion* (or *Merger*)**: Nils at age 2;0-2;3 produces voiced stops with lead voicing in German (and Spanish), and at age 2;3-2;6 he produces voiceless stops with long lag in Spanish (and German); see

Deuchar and Clark (1996) for the VOT study of an English-Spanish bilingual child.

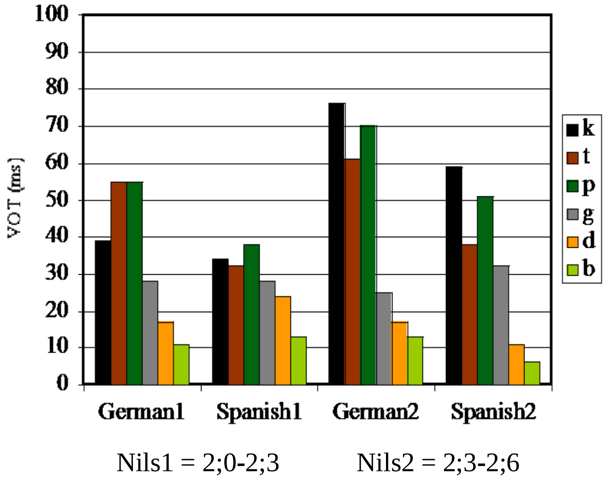


Figure 4. VOT values for the German and Spanish voiceless and voiced stops produced by the German-Spanish bilingual child Nils, during two time spans.

Figure 4 shows the VOT systems for the bilingual child Nils at two different time points. In time 1 (2;0-2;3), the German voiceless stops have a mean value of 50 ms., and the mean of the Spanish voiceless amounts to 35 ms. German voiced stops have a mean of 19 ms. and 31% of lead voicing (11/35). The mean of Spanish voiced stops amounts to 21 ms. and 19% of lead voicing (5/26). In time 2 (2;3-2;6), the VOT of German voiceless stops has notoriously increased to a mean of 70 ms., and the VOT of Spanish voiceless reaches a mean of 50 ms. German voiced stops have a mean of 18 ms. and 10% of lead voicing (3/30). The mean of Spanish voiced stops is similar: 16 ms. and they have 44% of lead voicing (4/9).

These results show an interesting phenomenon, namely the creation of a new category, a sort of "shortish" long-lag, which in Spanish has a mean of 50 ms., and in German a mean of 70 ms. Thus, for Spanish it is a bit too long, an in German falls in the short range. The values at time 2 show that short-lag is used to produce both, the German and also the Spanish voiced stops. This illustrates that Fusion involves bi-directional transfer plus some new category emerging from the joining of two categories, one from each language. The child differentiates voiceless and voiced in both languages, values approaching the German adult ones, but not totally.

Recapitulating OT grammar

The NOCODA constraint is violated both in Spanish (not too often) and very often in German, which means that additiveness (occurrence of codas in both languages) and frequency (in at least one of the languages) play an important role. Both factors, additiveness and frequency, accelerate the demotion of NOCODA in the bilingual Grammar of Spanish. NOCODA and also ALIGN, being Markedness constraints, are outranking at first, and must be demoted in order for the child to produce codas and pre-tonic syllables (Tesar & Smolensky, 1993). However, ALIGNLEFT shows additiveness, too, as there are unfooted syllables in both languages, but frequency is not high: in German child language unfooted syllables occur seldom, and in Spanish they appear more often, but not to be compared with the frequency of codas in German. Thus, both languages, Spanish and German, together, have a boosting effect on codas (which leads to acceleration), but do not reach the threshold to produce unfooted syllables and demote ALIGN LEFT, which brings some delay, when bilinguals are compared to monolinguals. Both types of constraints regulate prosodic structure, and their demotion is urgently needed, in order for the child to advance in the mastery of his/her phonological modules. Other markedness constraints are more restricted in the range that they regulate, as for instance those that ban specific marked segments, which we can consider unwanted segments, like the following:

- No /r/
- No Fricatives
- No long-lag Stops
- No Pre-voiced Stops
- No Long Vowels, etc.

Such marked segments are produced by the monolingual child sooner or later depending on the degree of difficulty, and frequency. For instance, it is well-known that the Spanish /r/ is a difficult segment, being one of the last to be acquired (not before age 3), followed by pre-voiced stops, which are not generally mastered before 4 or 5 years of age. Such marked segments are often substituted by other segments, less marked ones; e.g., [d] or [ð] generally substitute for [r] in child Spanish. In the case of bilinguals, on the one hand, they have a larger choice given the presence of the other language, which in the case of German offers [R] as a choice (also produced by the monolinguals as a substitute for [r], but not as often as by bilinguals). Clearly, though, such segments are not isolated but constitute classes of segments, identified by some specific feature. On the other hand, given that the child is able to decompose segments in (some of) their features, s/he may keep the features of the segment, except for the one banned by the specific constraint. But the bilingual child is capable of comparing the segments belonging to each language, and of choosing the one from the other language, for reasons of simplicity. For instance, if the constraint AGREE outranks UE, there is going to be a lot of form variability, which in German is not preferred. An outranking AGREE corresponds rather to the grammar of Spanish. Thus, Optimality Theory explicitly shows that in the present case *Transfer*, understood as transfer of the hierarchy of constraints, maintains the uniformity of lexical forms. This matches a characteristic of German phonology: it preserves the integrity (form and representation) of lexical units. It reflects the influence of German as a demarcating language (Trubetzkoy, 1939), vs. the grouping character of Spanish (Chen 1990). That is, words in German have clear edges, whereas in Spanish this is not the case, as often the ending of one word together with the beginning of next word constitute one single syllable (Colina, 1997).

Effects of the various forms of interaction

Delay is soon overcome, and acquisition takes place in the bilinguals as in the case of monolinguals. Acceleration is a temporary advantage, which at the end is counterbalanced, and acquisition takes place as in the case of monolinguals. While these two manifestations of interaction are temporary effects, without long-lasting consequences, *Transfer* may have long-lasting or even remaining effects. Order of acquisition is also compensated in the long run, so that acquisition is achieved as in the case of monolinguals. Fusion, as proposed by Queen (2001) is similar to transfer, as it introduces new categories that emerge under contact.

We can thus say that cross-language interaction shows quantitative and qualitative differences. Quantitative differences are: Delay, Acceleration and Variation in acquisition order. Qualitative differences are: Transfer and Fusion. It has been proposed elsewhere (Lleó & Cortés, 2012) that the crucial structural factors to predict type of cross-language Interaction are: Frequency, Additiveness (Presence in the other language), Uniformity (Complexity of the category), and Unmarkedness. The way that these four factors affect cross-language interaction can be observed in Figure 5, which on the left-hand side shows the constraint involved (demoted sooner or later in relation to monolingual development), the effects, from positive, to neutral or no-effect, to negative effect caused by demotion of the relevant constraint. On the right-hand side, factors are listed for each case, indicating by means of + whether the relevant factor is involved, and by means of — whether the relevant factor is not involved. If a factor is not relevant in a certain case, it appears in parentheses (e.g., Uniformity is not relevant for coda production).

	NoCoda is demoted below Faithfulness sooner	POSITIVE EFFECT (acceleration: SP codas)		Markedness + Frequency + Additiveness + (Uniformity)
↓	NoComplex is demoted below Faithfulness at same time	NO EFFECT (SP clusters)	↓	P is non-marked; if marked, similar frequency in A and B
	AlignLeft is demoted below MaxIO later	(SHORT) NEGATIVE EFFECT (short delay: SP unfooted syllables)		Markedness + Frequency — (Additiveness) (Uniformity)
	Agree[cont] is demoted below UE later	(LONG) NEGATIVE EFFECT (TRANSFER: SP spirants)		(Markedness) (Frequency) Additiveness — Uniformity —

Figure 5. Demotion of relevant constraints, factors affecting cross-language interaction and their effect.

As a result of the type of cross-language Interaction that emerges because of the presence or absence of certain factors, Interaction is not a discrete category, in a binary sense (of just being yes/no), but a gradient category, which may have a short lasting (AlignLeft in Spanish) or a long lasting (Spanish Agree[cont]) negative effect. Figure 4 mentions some examples affected by the corresponding factors, in the middle column. However, the crucial element affected by the presence or absence of factors is the constraint that bans or allows the forms, depending on whether the relevant constraint is outranking or not. Figure 4 shows the relevant constraint as the first element in each case, and indicates the time sequence, in which demotion in the bilingual grammar takes place, always in relation to the monolingual grammar.

By way of conclusion

We started asking about cross-language interaction in bilingual acquisition and about the outcomes of acquiring two languages with different properties, and we found grammars that after a certain initial time span enter a stage of strong restrictions, based on Markedness outranking Faithfulness. Markedness outranking Faithfulness characterizes child phonology, and thus Markedness must be demoted in order for the Grammar to converge with the target language. However, different languages do different things: In traditional Standard Generative Phonological terms, German has a rule of Glottal Stop Insertion (Wiese, 1996), which takes place in the case of a lexical item that does not have a consonantal Onset, even if that word is preceded by a consonantal coda. In Spanish, in such a situation, resyllabification is applied. That is, somehow, the coda fills the missing Onset of the following word (Colina, 1997; Harris, 1983; Hualde, 1992; Lleó, forthcoming). Moreover, in Spanish, there is assimilation of certain features **(**Spirantization and also assimilation of the PA of nasals to the following obstruent). These processes, which contribute to further confusion of word edges, are characterized as Grouping phenomena by Chen (1990), whereas something like Glottal Stop Insertion in German is considered demarcative (Trubetzkoy, 1939).

Why would one language prefer Uniformity over Agree, and another language would prefer the reverse, Agree outranking Uniformity? Certain languages preserve the integrity of lexical items (demarcative), while other languages maintain the flow of connected speech (grouping). These two different characteristics of the languages of the world make tendencies explicit, which if not treated within OT would remain unexplored and invisible.

Acknowledgments

Thanks go to

- The Research Center on Multilingualism (SFB 538), German Research Foundation (DFG) and University of Hamburg.
- The research assistants of the project along the last several years, especially to Dr. Margaret Kehoe, and to the student assistants.

- The children of the projects from Hamburg and Madrid, and their parents.

References

Babatsouli, E. & Ingram, D. (2015). What Bilingualism tells us about phonological acquisition. In R. H. Bahr & E. R. Silliman (eds.), *Routledge handbook of communication disorders* (pp. 173-182). Routledge: Taylor & Francis.

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5* (9/10), 341-345.

Chen, M.Y. (1990). What must phonology know about syntax? In S. Inkelas & D. Zec (eds.), *The Phonology-Syntax Connection* (pp. 19-46). Chicago and London: The University of Chicago Press.

Colina, S. (1997). Identity Constraints and Spanish Resyllabification. *Lingua*, *103*(1), 1-23.

Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics*, *24*, 351-365.

Fabiano-Smith, L. & Goldstein, B.A. (2010). Phonological acquisition in bilingual Spanish-English speaking children. *Journal of Speech, Language and Hearing Research*, *53*, 160-178.

Gass, S. M., & Selinker, L. (eds.) (1983). *Language transfer in language learning*. Cambridge, MA: Newbury House.

Harris, J.W. (1983). *Syllable structure and stress in Spanish. A nonlinear analysis*. Cambridge, MA: The MIT Press.

Harris, J.W. (1984). La espirantización en castellano y la representación fonológica autosegmental. *Estudis Gramaticals 1. Working Papers in Linguistics* (pp. 149-167). Bellaterra: Universitat Autònoma de Barcelona.

Hualde, J.I. (1992). On Spanish Syllabification. In H. Campos & F. Martínez-Gil (eds.), *Current Studies in Spanish Linguistics* (pp. 475-493). Washington D.C.: Georgetown University Press.

Lleó, C. (2002). The role of Markedness in the Acquisition of Complex Prosodic Structures by German-Spanish Bilinguals. *International Journal of Bilingualism*, *6* (3), 291-313.

Lleó, C. (2006). The Acquisition of Prosodic Word Structures in Spanish by Monolingual and Spanish-German Bilingual Children. *Language and Speech*, *49*(2), 207-231.

Lleó, C. (2012). Monolingual and bilingual phonoprosodic corpora of child German and child Spanish. In T. Schmidt & K. Wörner (eds.), *Multilingual corpora and multilingual corpus analysis. Hamburger Studies on Multilingualism 1*4 (pp. 107-122). Amsterdam/Philadelphia: John Benjamins.

Lleó, C. (forthcoming). Lexically empty onsets in L1 phonological acquisition of Spanish and German. In N. Cedeño & A. Rafael (eds.). *The Syllable in Romance Languages: Studies in Honor of James W. Harris*. Berlin & New York: Mouton de Gruyter.

Lleó, C. (in press). Acquisition of speech sound. In U. Domahs & B. Primus (eds.), *Laut, Gebärde, Buchstabe (Sound, sign, letter).* De Gruyter.

Lleó, C., & Cortés (2013). Modelling the Outcome of Language Contact in the Speech of Spanish-German and Spanish-Catalan Bilingual Children. In J. Kabatek & L. Loureido (eds.), Special Issue on *Language Competition and Linguistic Diffusion: Interdisciplinary Models and Case Studies. International Journal of the Sociology of Language*, *221*, 101-125.

Lleó, C., Kuchenbrandt, I., Kehoe, M., & Trujillo, C. (2003). Syllable final consonants in Spanish and German monolingual and bilingual acquisition. In N. Müller (ed.), *(In)vulnerable Domains in Multilingualism* (pp. 191-220). Amsterdam, Philadelphia: John Benjamins.

Lleó, C., & Rakow, M. (2005). Markedness Effects in Voiced Stop Spirantization in Bilingual German-Spanish Children. In J. Cohen, K. T. McAlister, K. Rolstad & J. MacSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism (ISB4)* (pp. 1353-1371). CD Rom: Cascadilla Press.

Lleó, C., & Vogel, I. (2004). Learning new segments and reducing domains in German L2 Phonology: The role of the Prosodic Hierarchy. *International Journal of Bilingualism*, *8* (1), 79-104.

Paradis, J. & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition*, *18*, 1-25.

Queen, R. M. (2001). Bilingual intonation patterns: Evidence of language change from Turkish-German bilingual children. *Language in Society*, *30*(1), 55-80.

Tesar, B., & Smolensky, P. (1993). *The learnability of Optimality Theory: An algorithm and some basic complexity results*. Rutgers Optimality Archive ROA-2, http://ruccs.rutgers. edu/roa.html.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, criolization and Genetic Linguistics*. University of California Press.

Trubetzkoy, N. (1939). *Grundzüge der Phonologie*. Prag: Travaux du Cercle Linguistique de Prague.

Wiese, R. (1996). *The phonology of German*. Oxford, UK: Clarendon Press.

Do early bilinguals speak differently than their monolingual peers? Predictors of phonological performance of Polish-English bilingual children

Marta Marecka[1], Magdalena Wrembel[1], Dariusz Zembrzuski[2], Agnieszka Otwinowska-Kasztelanic[2]
mmarecka@wa.amu.edu.pl, magdala@wa.amu.edu.pl, d.zembrzuski@gmail.com, a.otwinowska@uw.edu.pl

[1]Faculty of English, Adam Mickiewicz University in Poznan,
[2]Institute of English Studies, University of Warsaw

Abstract. It is a common belief that speech production of early bilinguals is similar to that of their monolingual peers and that these bilinguals speak both languages without a foreign accent. While some studies suggest that this is indeed the case and that bilinguals are similar in their phonological development to monolinguals (Holm & Dodd 1999), others show considerable differences between bilingual and monolingual children when it comes to speech production (e.g., En et al., 2014; Mayr et al., 2015). The study explores the phonological patters in the L1 speech of Polish-English bilingual children as compared against the speech of their monolingual peers. The participants were 59 bilingual children of Polish migrants to the UK, and 24 monolingual Polish children matched for age, gender and socio-economic status, who were recorded repeating a set of sentences in Polish. All bilingual children were exposed to Polish from birth and spoke this language at home with their families. Nevertheless, we hypothesised that bilingualism would affect their overall phonological performance in Polish, and that their speech will be characterised by phonological crosslinguistic influence (CLI) from English. Speech sample recordings came from a database collected by the Bi-SLI-Poland project within the European COST Action IS0804 with the use of the Polish *Sentence Repetition Task* (Banasik et al., 2011). The data collection procedure involved a sentence repetition task, in which the participants repeated 68 sentences that they heard through the headphones. For each child 14 preselected sentences from this task were subsequently analyzed auditorily by phonologically trained independent raters, who assessed the number of phonological alterations and the degree of cross-linguistic influence (CLI) from English in children's speech. Moreover, detailed background information on the bilingual children's language development, language input and output was collected. This information was used in regression analysis to establish the predictors of CLI in bilingual's Polish speech. The results of the study indicate that the L1 speech patterns of the bilingual children differed from the speech patterns of their monolingual peers since, in the case of the former group, Polish speech was affected by CLI from English. The analysis of the background factors revealed that the degree of CLI in the L1 speech of Polish-English bilinguals depended on the quantity and quality of the L2 input those children had received.

Keywords: early bilingualism, phonological development, cross-linguistic influence, Polish-English bilinguals

Introduction

Research shows conflicting results regarding the differences in phonological development between bilingual children and their monolingual peers. On the one hand, certain scholars indicate that bilingual children develop similarly to their monolingual peers in each language and that they distinguish between the two phonological systems (Holm & Dodd, 1999; Johnson & Wilson, 2002). On the other hand, the majority of researchers indicate that bilingual and monolingual children show distinct phonological patterns in a particular language at both the segmental and suprasegmental levels (Vihman, 1996). This would indicate that the bilingual children differ from the monolingual children and that the two languages in the bilingual mind might interact. This idea is quite widespread in the literature on bilingualism. For instance, the Speech Learning Model of second language acquisition (Flege, 2002) assumes that the phonetic categories from both languages in the bilingual mind occupy the same phonological space. Also, Dynamic Systems Theory, which has been gaining popularity in the current literature on bilingual and multilingual acquisition (de Bot, Lowie, & Thorne, 2013; de Bot, Lowie, & Verspoor, 2007; Herdina & Jessner, 2002), points to the existence of an interaction between the pertinent languages.

Assuming there is an interaction between the languages of a bilingual child, is cross-linguistic influence (CLI) bi-directional or uni-directional? What is the direction of influence? Such questions are frequently raised in studies investigating bilinguals, including minority group and migrant children, yet the results appear to be mixed. On the one hand, some studies show influence from the minority language, often the first language of the bilingual participants, to the community language. This is the case, for instance, in Spanish-English bilinguals in the USA, who demonstrated CLI from Spanish (the minority language) to English (the community language) in the production of L2 segments (Barlow, 2014). However, other researchers show a clear CLI from the community language to the minority language in bilingual speakers (Mayr, Howells, & Lewis, 2014) or even cases of the first language attrition in the bilingual speakers (Schmid, 2013).

In this paper, we investigate the phonological patterns in the minority language (i.e. Polish) of Polish-English bilingual children of the Polish migrants in the UK. The participants have been living in the UK for most of their lives, yet Polish was chronologically their first language. We examined whether the Polish speech of these bilingual children was different from the speech of their monolingual peers due to the special setting of acquisition. Moreover, we were interested in whether any environmental factors such as the quantity and quality of the Polish input were connected to the degree of CLI as perceived by native users of Polish, who were phonologically trained. On the basis of previous literature which points to the interaction between the languages in the bilingual mind, we hypothesised that there would be evidence of CLI from the community language (English) to the minority language (Polish) of the participants, despite the fact that Polish was chronologically the first language of the participants.

Method

The data for the current project come from a database collected by the Bi-SLI-Poland project carried out within the European COST Action IS0804 (see Acknowledgements).

Participants

The participants' pool comprised 59 bilingual children of Polish migrants to the UK, aged 4;5 to 6;11 (*M* = 5;9, *SD* = 9 months) and 24 monolingual Polish children, matched for aged, sex and socio-economic status. In both groups, females constituted around 60% of the sample. Prior to the experiment, the parents of the bilingual children filled in an extensive language development questionnaire containing questions about children's background, exposure to both languages and language output. The background data revealed no significant differences between the bilingual and the monolingual children in terms of their socio-economic status. The bilingual group was also assessed as fairly homogenous: all the participants had very frequent or exclusive contact with Polish from birth or from the first month of life, all had at least one Polish parent, and all children used both Polish and English on a regular basis. As many as 96% of the participants uttered their first word in Polish, and 70% of the children were also Polish dominant in terms of proficiency, as reported by the parents.

Procedure

The study constitutes part of a larger project, devoted to creating a linguistic profile of Polish-English bilingual children living in the UK on the basis of the COST Action data (see Acknowledgements). To answer the research question posed in the current study, we investigated the recordings of the Polish Sentence Repetition Task (Banasik, Haman, & Smoczyńska, unpublished) from the database. The task consisted of 68 sentences in Polish, recorded by two Polish native speakers. The sentences varied in grammatical complexity and length. Each sentence was played to the participant through the headphones and the child's task was to repeat it. The subsequent repetitions were audio recorded. The children were tested individually in a quiet room at home, or at school. This task was initially designed to test children's morpho-syntactic abilities, but it was chosen for this study since it offered consistent phonological output across the participants.

Data analysis

First of all, five randomly selected participants' recordings were transcribed phoneme by phoneme by three trained phoneticians. Those transcriptions constituted the basis for a diagnostic list, i.e. a list of possible speech patterns found in the speech of the children and deviating from the monolingual norm due to CLI (see also Marecka, Wrembel, Otwinowska-Kasztelanic, & Zembrzuski, 2015). The list contained 12 problem areas in which cross-linguistic influence occurred, including:

Vowel production

1. Vowel quality distorted
2. Vowel quantity distorted
3. Vowel reduction applied to Polish
4. Polish nasal vowels misarticulated

Consonant production

5. Production of non-native-like consonants
6. Reduction of consonantal clusters
7. Substitution of consonantal clusters (change of quality in the cluster, e.g., substitution of one consonant)
8. Lack of consonant palatalisation in appropriate context
9. Atypical VOT patterns in plosives
10. Voice assimilation process not applied

Suprasegmentals

10. Incorrect number of syllables
11. Incorrect stress pattern

Out of the set of 68 sentences in the original sentence repetition task, 14 diagnostic sentences were selected, as they offered the richest phonological contexts for further analysis.

Six phonetically trained Polish raters took part in the assessment procedure. Each set of 14 sentences was analyzed auditorily by two raters. The speech samples of the monolingual and bilingual children had been randomized, thus the raters were blind as to the linguistic background of the participants. Each rater received a card with 14 sentences transcribed in the International Phonetic Alphabet and they had to mark on the cards the articulatory alterations stemming from CLI made by the children. Then the raters were requested to classify these alterations into one of the 12 categories (problem areas) from the diagnostic list. On that basis, we could assess how many speech alterations occurred in children's articulations for each category from the list and overall. Further, the raters were to judge the degree of cross-linguistic influence in children's speech for each category from the diagnostic list on a three-point scale (0 -significant CLI from English, 0.5 - occasional CLI, 1 - no CLI). To assess the overall level of CLI in the speech samples, the total sum of those assessment points was calculated for each child. The raters' responses were cross-checked by two authors of the present study.

To address our research questions, we compared the overall number of speech alterations as well as the overall level of CLI between the monolingual and bilingual children groups. We also compared the number of alterations and assessments of CLI for each of the 12 categories (problem areas) from the diagnostic list. The analyses performed allowed us to evaluate the research hypothesis regarding differences in speech between monolingual and bilingual speakers.

We were also interested in exploring the predictors of such differences. To this end, we extracted a number of variables related to participants' language background from the parental questionnaires. These included:

- children's age (in months)

- the risk of developmental delay (based on questions about language disorders in the family, late onset of speech, etc.)
- the first contact with English (in months)
- the quality and quantity of early exposure to English (measured in the months of exposure times the reported frequency of exposure as measured on a five-point scale)
- the overall quality and quantity of the Polish input (reported frequency of input measured on a five-point scale times the number of people speaking in Polish to the child)
- the overall quality and quantity of the English input (reported frequency of input measured on a five-point scale times the number of people speaking in English to the child)
- the Polish output produced by the child (reported frequency of input measured on a five-point scale times the number of people child speaks to in Polish)
- the English output (reported frequency of input measured on a five-point scale times the number of people child speaks to in English)
- mother's education (in years)
- father's education (in years)

Further, we performed correlation and regression analyses to investigate if any of the above variables could predict the overall CLI assessment and of the number of alterations in bilingual children's productions.

Results

Monolingual vs. bilingual speakers

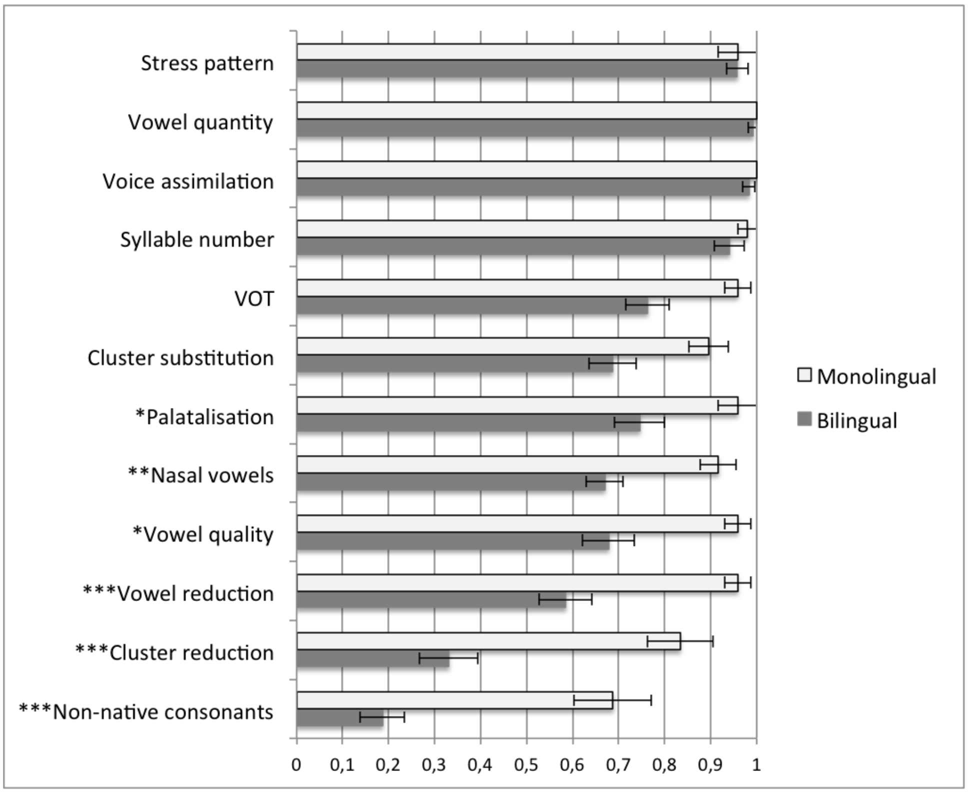
The differences in the number of alterations between particular categories from the diagnostic list are presented in Table 1. As predicted by our research hypothesis, the Polish speech of bilingual children differed significantly from the speech of their monolingual peers. The raters reported on average 7 alterations in the speech of the monolingual children (SD = 7.35), as opposed to 26.54 alterations in the speech of bilinguals (SD = 14.57). The difference is statistically significant, as indicated by the Mann-Whitney U test (U = 1248.00, p < .001). There were also differences in the overall CLI assessment between the two groups. Monolinguals scored on average 11.1 out of 12 points maximum (SD = 1.14), which indicated that the raters detected very little to no CLI in their speech, while bilinguals scored 8.52 out of 12 (SD = 1.88). The difference between the two groups was, again, significant, as indicated by the Mann-Whitney U test (U = 130.5, p = .001).

Additionally, we investigated in which areas, as enumerated in the diagnostic list, the differences between the monolinguals and bilinguals were most pronounced. The answer to this question can be gleaned from the bar plot in Figure 1, which shows the average CLI assessment scores for each diagnostic list area for the monolingual and bilingual groups (with standard error of the mean indicated). The problem areas at the top of the chart are the ones where the differences were less pronounced, the ones on the bottom are those where the differences were the greatest. The asterisks indicate for which problem areas the differences between the monolinguals and bilinguals were statistically significant (as measured with Mann-Whitney U test with Bonferroni corrections). As shown by the plot, the greatest differences were found in the production of consonants and in cluster reduction.

Table 1. The number of speech alterations in monolingual and bilingual group

Problem area	Speech alterations monolinguals		Speech alterations bilinguals		Mann-Whitney U test	
	M	SD	M	SD		
Vowel quality	0.33	0.64	2.31	3.40	U =311.0	***
Vowel quantity	0.13	0.45	0.15	0.49	U =674.0	
Vowel reduction	0.29	0,62	2.34	2.23	U =271.5	***
Nasal vowels	0.88	0.95	2.27	1.71	U =342.0	**
Non-native consonants	2.88	4.11	9.27	6.08	U =220.5	***
Cluster reduction	1.04	1.57	4.71	3.14	U =195.0	***
Cluster substitution	0.58	0.83	1.49	1.33	U =407.0	*
Palatalisation	0.25	0.68	1.63	2.15	U =403.0	**
VOT	0.21	0.72	1.47	1.84	U =341.5	***
Voice assimilation	0.08	0.28	0.15	0.48	U =681.0	
Syllable number	0.21	0.51	0.49	1.01	U =598.0	
Stress pattern	0.13	0.45	0.25	0.65	U =648.0	
IN TOTAL	**7.00**	**7.35**	**26.54**	**14.8**	**U =130.5**	*******

'***' $p < .001$, '**' $p < .01$, '*' $p < .05$



'***' $p < .001$, '**' $p < .01$, '*' $p < .05$

Figure 1. Average CLI assessment scores for bilingual and monolingual speakers in the problem areas from the diagnostic list

Predictors of speech alterations and CLI in bilingual speakers

Before running the regression analyses, we first created a correlation matrix with the data extracted from the parental questionnaires and with the overall number of speech alterations and the overall CLI assessment. The only variable extracted from the questionnaires that was correlated (negatively) with the overall CLI assessment was the overall input in English (r = -.28, p = .043, 95% CI -.51, -.01). This result shows that the more input in English the child received, the lower the overall CLI assessment, i.e. the more the child's speech was characterized by cross-linguistic influence from English. Marginally significant were also the negative correlations with the English output (r = -.23, p = .088, 95% CI -.46, .03) and the quality and quantity of early exposure to English (r = -.25, p = .0869, 95% CI -.48, .01). None of the questionnaire variables correlated significantly with the number of speech alteration in the children's Polish production.

Following the correlation analyses, a multiple regression analysis was conducted and the best-fitting model was selected using the all-subsets method (with the use of the leaps package in R: Lumley & Miller, 2004). The best model for the general CLI assessment is presented in Table 2. As can be seen, the overall input in English is the sole predictor of CLI. The regression model (F(1,51) = 4.306, p = .043) explains, however, merely 8% of the variance (R^2 = .08, R^2_{Adjusted} = 0.06). The best-model for the overall number of speech alteration contained maternal education as the sole predictor, this model, however, failed to reach statistical significance (F(1,51) = 2.354, p = .131, R^2 = .04, R^2_{Adjusted} = 0.03).

Table 2. The regression model for the overall assessment of CLI in bilinguals

	B	Std. Error	t value	ß	Pr(>|t|)
(Intercept)	10.08	0.79	12.73		<0 .001
Overall input in English	-0.04	0.02	-2.08	-0.28	0.043

Discussion and conclusions

Our data show clear differences between the Polish speech of bilingual and monolingual speakers. The Polish-English bilingual children showed more speech alterations in their productions of Polish and their speech was affected by CLI from English. The differences between monolingual and bilinguals manifested themselves especially in the more marked aspects of Polish phonology, namely the production of the consonants and the consonantal clusters, but not in the suprasegmental features of their speech. These results conform to the interactive theories of bilingualism, stating that the two languages in the bilingual mind do influence each other (de Bot et al., 2007; 2013). The findings also indicate that due to the interaction of the two languages, bilingual speech development differs from the monolingual development. Furthermore, they suggest that the minority language of the speakers might be affected by CLI, despite being chronologically the first acquired language.

The investigation of possible factors influencing the degree of CLI was not conclusive, as indicated by the small amount of variance explained by our regression models. This was possibly due to the fact that the participants' sample was fairly homogenous. However, our results do suggest that the degree of CLI in the minority language might be influenced by the quality and quantity of input the children receive in the second language, the community language. In the study, the children who received more input in English, were assessed as being more affected by cross-linguistic influence from English when speaking Polish.

Overall, the study indicates that the phonological development in the first language of the migrant children might be affected by the influence from the community language, especially if the children receive significant amounts of input from the community language.

Acknowledgements

The research presented here was supported by the Polish Ministry of Science and Higher Education grant (Decision nr 0094/NPRH3/H12/82/2014), *Phonological and Morpho-syntactic Features of Language and Discourse of Polish Children Raised Bilingually in Migrant Communities in Great Britain*" carried out at the Faculty of Modern Languages, University of Warsaw, Poland.

The data for this study come from the Bi-SLI-Poland project within the European COST Action IS0804. The project was supported by the Polish Ministry of Science and Higher Education /National Science Centre (Decision nr 809/N-COST/2010/0) and carried out in the years 2010-2014.

References

Banasik, N., Haman, E., & Smoczyńska, M. (2011). *Sentence Repetition Task.* Unpublished Material.

Barlow, J. A. (2014). Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology*, *5*, 1-14.

de Bot, K., Lowie, W., & Thorne, S. L. (2013). Dynamic systems theory as a comprehensive theory of second language development. In M. Mayo, M. Gutierrez-Mangado, & M. Adrián (eds.), *Contemporary Approaches to Second Language Acquisition* (pp. 199-220). Amsterdam: John Benjamins.

de Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(01), 7-21.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rhode (eds.), *An Integrated View of Language Development Papers in Honor of Henning Wode* (pp. 217-244). Trier: Wissenshaftlicher Verlag.

Herdina, P., & Jessner, U. (2002). *A Dynamic Model of Multilingualism: Perspectives of Change in Psycholinguistics*. Clevedon: Multilingual Matters Ltd.

Holm, A., & Dodd, B. (1999). A longitudinal study of the phonological development of two Cantonese-English bilingual children. *Applied Psycholinguistics*, *20*(03), 349-376.

Johnson, C. E., & Wilson, I. L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism*, *6*(3), 271-289.

Lumley, T., & Miller, A. (2004). *Leaps: regression subset selection*. R package version.

Marecka M., Wrembel M., Otwinowska-Kasztelanic, A., Zembrzuski D. (2015). "Phonological Development in the Home Language among Early Polish-English Bilinguals," In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, UK: The University of Glasgow.

Mayr, R., Howells, G., & Lewis, R. (2014). Asymmetries in phonological development: the case of word-final cluster acquisition in Welsh-English bilingual children. *Journal of Child Language*, *42*(01), 146-179.

Schmid, M. S. (2013). First language attrition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 117-123.

Vihman, M. M. (1996). *Phonological Development*. Cambridge, UK: Blackwell Publishing.

Third language acquisition: An experimental study of the pro-drop parameter

Stamatia Michalopoulou
stamatia.michalopoulou@gmail.com

Aristotle University of Thessaloniki

Abstract. The present paper attempts to investigate the interlanguage of Greeks (mother tongue, L1=Greek) who have already acquired English as a first foreign language (L2) when acquiring German as a second foreign language (L3). These adults are sequentially trilingual, i.e. they began acquiring their L1, L2 and L3 at a different point in time in a purely monolingual environment through organized instruction (for the L2 and the L3). The present study investigates the 'Pro-drop Parameter', a syntactic parameter in which the three examined languages have different values. In order to examine the interlanguage of the Non-Native-Speakers (NNS) of German, an experimental study was conducted, consisting of two tasks, a Grammaticality Judgement Task and a Preference Task (in the present paper only the former is presented). These tasks have measured the judgments and preferences respectively of the three groups of participants. Two groups consisted of NNS with different levels of proficiency in German, but the same in English. The third group comprised of native speakers of German and served as the control group. The results of both experimental tasks show that none of the languages that the NNS already know seem to play a more significant role than the other in shaping their interlanguage in both proficiency levels in German. Both languages seem to be equally important and available in order to provide an appropriate linguistic representation of the target language at any given time. According to these data, it seems that the theoretical model concerning acquisition of a third language which best describes the interlanguage of the NNS, is that of Flynn, Foley & Vinnitskaya (2004), namely, the 'Cumulative-Enhancement Model for Language Acquisition'. According to this model, each language already acquired is equally important and available to play a role in acquiring the target language and can contribute to the development of the syntactic structure of each subsequent language either in a positive way or in a neutral way. That is, there is only "positive language transfer" or no linguistic transfer at all to the target language.

Keywords: third language acquisition, interlanguage, pro-drop parameter

Introduction

The basic and dominant topic of discussion of theoretical and experimental approaches on the acquisition of a foreign language is to investigate and determine the source of linguistic transfer of syntactic structures and functional categories (Gass, 1996; Odlin, 1989, 2003) in the interlanguage of non-native speakers (NNSs) (Selinker, 1972; Sharwood-Smith, 1994; Han & Tarone, 2014). Until recently, research has studied foreign language as second language, and has ignored additional foreign languages previously acquired (Klein, 1995; Leung, 2007). This has probably led to errors regarding the identification of the source of language transfer during the acquisition of the target language, since there was not only one language that could be the source of it, but two (or even more).

Regarding the investigation of the interlanguage, since the NNS already know two languages, the source of language transfer cannot be determined unless the studied syntactic phenomenon and parameter are differently valued at their mother tongue (L1) and at their first foreign language (L2), and their second foreign language (L3) is similar to or different with one of the two.

The paper is structured as follows: Firstly, the most important theories and hypotheses about the complex phenomenon of L3 Acquisition (L3A) are presented. Then there is a short reference to the syntactical phenomenon studied with special reference to the reasons advocating its choice for research. Then, a description of the study's design and methodology is given. Finally, the most significant results are presented, accompanied by a discussion.

Theoretical approaches to third language acquisition

Research on L3A was initially based on theoretical hypotheses made about L2 Acquisition. However, these hypotheses may not always be sufficient for the analysis and interpretation of L3A. Mostly, in the last decade, research has led to new theoretical approaches adapted to describe in the most coherent way possible the multidimensional data of the new scientific field. Next, the four main theoretical approaches in L3A and their basic principles are presented.

Developmentally moderated transfer hypothesis (Håkansson, Pienemann, & Sayheli, 2002)

According to this hypothesis, the L1 still has a privileged role in L3A. The L1 is the exclusive source from which morphosyntactic features are transferred to the interlanguage of NNS. Linguistic influence from the L1 to the foreign language follows a concrete evolutionary process.

Second language status factor hypothesis (Williams & Hammarberg, 1998)

The basic principle of this hypothesis is that there is a separate mechanism that is activated by acquiring every foreign language and is not the same as that in L1 acquisition. All non-native languages are grouped in a separate area in the mind from that of L1. During L3A, there is faster and more direct access to L2 than to L1. The L2, rather than L1, has more influence on the interlanguage of NNS during L3A.

Cumulative-enhancement model for language acquisition (Flynn, et al., 2004)

According to this model, neither L1 nor another language plays a dominant role in the acquisition of the subsequent language. Each language already been acquired is as important, and perhaps available at the same degree, to play a role in acquiring the target language and can contribute to the development of the syntactic structure of each subsequent language only in a manner, that is either positive or neutral; that is, there is only "positive language transfer" or no linguistic transfer at all to the target language.

Typological primacy model (Rothman, 2011)

The basic principle of this model is that the linguistic transfer during foreign language acquisition does not always have a positive effect and does not always seem to facilitate L3A. The initial stage of the acquisition of a foreign language is determined selectively from the (psycho)typological distance or proximity that exists between any given pair of interacting languages. This is true either when this proximity is objective or a subjective perception of the NNS. It is also applicable even if it is not the most economical choice, or simply even when it actually hinders instead of facilitating L3 development.

The pro-drop parameter

In order to identify the source of linguistic transfer, a syntactic phenomenon with specific properties must be selected and studied. This syntactic phenomenon is realized differently in the NNS' L1 and L2, while their L3 resembles either one or the other language, as regards this phenomenon. After studying the syntactic properties of the three test languages, Greek (L1), English (L2) and German (L3), the syntactic phenomenon, which was chosen to be studied in this research is the 'Pro-drop Parameter' or the 'Null-Subject Parameter', a parameter in which the three examined languages have different values (White, 1989).

The existence or not of null subjects in one language, i.e. whether the subject (*pro*) of an inflected verb of a sentence can be dropped or not, is controlled by the Pro-drop Parameter (e.g., Chomsky, 1981a; 1981b; Jaeggli, 1982; Rizzi, 1982; 1986; Huang, 1984)

Linguistic typology

This parameter is so significant in linguistic typology, that its realization or not in a certain language is a basic factor in language classification. D' Alessandro (2014) makes the following categorization of languages:

- Canonical Null-Subject Languages
  e.g., Greek, Italian, Spanish
- Radical Null-Subject Languages
  e.g., Chinese, Japanese, Korean
- Partial Null-Subject Languages
  e.g., Finnish, Hebrew
- Expletive Null-Subject Languages
  e.g., German, Danish
- Non Null-Subject Languages
  e.g., English, French

According to this categorization, Greek is a Canonical Null-Subject Language and English is a Non Null-Subject Language. Among other properties of the parameter, this means that a pronoun in Greek does not necessarily have to be realized in subject position, i.e. overt grammatical subjects may be omitted (e.g., both *ego pezo* and *Ø pezo* are correct). On the contrary, the pronominal subject in English cannot be omitted in order to constitute a grammatically correct sentence (e.g., *I play* but not ** Ø play).*

The pro-drop parameter in German

German is classified as a Non Null-Subject Language by many researchers (e.g., Cabredo Hofherr, 1999, 2003; Holmberg, Nayudu, & Sheehan, 2009). Therefore, characteristics similar to those in French are ascribed to German as well, because in most cases German does not allow the omission of the overt grammatical subject (e.g., *ich spiele* but not ** Ø spiele).* In fact, there are some instances in German where omission of the overt grammatical subject is also permitted. Therefore, current theoretical approaches classify German among Expletive Null-Subject Languages (D' Alessandro, 2014). A case, where the expletive subject can also be omitted in German, is identified in the passive Voice of specific verb classes.

The passive voice in German

In German, it is possible for some verb classes only to appear without a subject but only if these verb classes are in the passive voice. In particular, these verb classes are: verbs that accept complement in dative (examples 1 and 4), verbs that accept a prepositional phrase as a complement (examples 2 and 5), and unergative verbs (examples 3 and 6). Next, verbs that belong to these verb classes are allowed to appear in the passive voice, either with the expletive subject 'es' (examples 1, 2, 3) or with no subject at all (examples 4, 5, 6).

(1)	Es	wurde	der Mutter	im Haushalt	nie	geholfen.
	ES	AUX $_{\text{3 SG PRES}}$	the mother $_{\text{DAT SG.}}$	with the household	never	helped $_{\text{PASS PART}}$

They never used to help the mother with the household.

(2)	Es	wurde	den ganzen Nachmittag	nach dem Schlüssel	gesucht.
	ES	AUX $_{\text{3 SG PRES}}$	the whole afternoon	for the key $_{\text{PREP}}$	looked $_{\text{PASS PART}}$

The whole afternoon they were looking for the key.

(3)	Es	wurde	in Frankreich	spontan	demonstriert.
	ES	AUX $_{\text{3 SG PRES}}$	in France	spontaneously	demonstrated $_{\text{PASS PART}}$

In France, they demonstrated spontaneously.

(4)	Der Mutter	wurde	im Haushalt	nie	geholfen.

the mother $_{\text{DAT SG.}}$ AUX $_{\text{3 SG IMPERF}}$ with the household never helped $_{\text{PASS PART}}$
They never used to help the mother with the household.

(5) Den ganzen Nachmittag wurde nach dem Schlüssel gesucht.
the whole afternoon AUX $_{\text{3 SG IMPERF}}$ for the key $_{\text{PREP}}$ looked $_{\text{PASS PART}}$
The whole afternoon they were looking for the key.

(6) In Frankreich wurde spontan demonstriert.
in France AUX $_{\text{3 SG IMPERF}}$ spontaneously demonstrated $_{\text{PASS PART}}$
In France, they demonstrated spontaneously.

It is obvious in the examples that the Null-Subject Parameter in German is realized in some cases in the same way as in English and in others as in Greek. For this reason, this parameter was chosen to be studied in the present research.

The experimental procedure

In order to investigate the interlanguage of the NNS, an experimental study consisting of two tasks, a Grammaticality Judgement Task and a Preference Task, was conducted. These tasks have measured the judgments and preferences respectively of the three groups of participants. Only the Grammaticality Judgement Task (GJT**)** is tackled in this paper.

The grammaticality judgement task

The GJT consisted of 144 experimental utterances. 72 of them were grammatically correct and in the passive voice, while 72 were grammatically incorrect and in the active voice. A total of 36 verbs were used four times each in four different sentences. Two of them were grammatically correct in the passive voice and the other two were grammatically wrong in the active voice. The verbs were derived from six verbal classes; 6 verbs were used from each verbal class. The verbs used are divided into two broad categories of verb classes, as far as subject omission is concerned:

A) those that do permit omission of the subject in the passive voice, that is:
i) verbs that accept complement in the dative,
ii) verbs that accept a prepositional phrase as a complement
iii) unergative verbs

B) those that do not permit the omission of the subject neither in the active nor in the passive voice, that is:
i) verbs that accept a complement in the accusative
ii) verbs that accept two complements both in the accusative and in the dative
iii) verbs that are allowed to build impersonal passive as well.

Sentences with verbs in the first category, that permit omission of the subject in the passive voice, appeared in the following versions: the two correct sentences in the passive voice had either the expletive subject '*es*' (experimental condition: [R./ -lex. sub./ +es]) (Example 7) or no subject at all (experimental condition: [R./ -lex. sub./ -es]) (Example 8). The two wrong sentences in the active voice had either no subject at all (∅) (experimental condition: [W./ -lex. sub./ -es]) (Example 9) or had the finite verb incorrectly placed in the third place of the sentence (V3) (Example 10) (this case is not examined in this paper). Examples 7-10 are:

(7) Es wird immer lange auf den Bus 100 gewartet.
ES AUX $_{\text{3 SG PRES}}$ always too long for the bus 100 $_{\text{PREP}}$ waited $_{\text{PASS PART}}$
They always wait too long for the bus 100.

(8) Auf die Braut wird immer lange gewartet.
for the bride $_{\text{PREP}}$ AUX $_{\text{3 SG IMPERF}}$ always too long waited $_{\text{PASS PART}}$
They always wait too long for the bride.

(9) * Mädchen, warum wartet nicht auf euren Bruder?

girls, why wait 2 PL PRES not for your brother PREP
Girls, why don't you wait for your brother?

(10) * Alle Kinder ungeduldig warten auf die Sommerferien.
all the kids NOM PL impatiently wait 3 PL PRES for the summer vacation PREP
All the kids wait impatiently for the summer vacation.

Sentences with verbs in the second category, that do not allow omission of the subject neither in the passive nor in the active voice appeared in the following versions: the two correct sentences in the passive voice had either a lexical subject as well as the expletive subject '*es*' (experimental condition: [R./ +lex. sub./ +es]) (Example 11) or only a lexical subject (experimental condition: [R./ +lex. sub./ -es]) (Example 12). The two wrong sentences in the active voice had either no subject at all (Ø) (experimental condition: [W./ -lex. sub./ -es]) (Example 13) or had the finite verb incorrectly placed in the third place of the sentence (V3) (Example 14). This case is not examined in this paper.

(11) Es werden alle Frauen einmal im Leben geliebt.
ES AUX 3 SG PRES all women NOM PL once in life loved PASS PART
All women are loved once in their lifetime.

(12) Die Schauspielerin wurde von allen Regisseuren geliebt.
the actress NOM SG AUX 3 SG IMPERF by all the directors loved PASS PART
The actress was loved by all the directors.

(13) * Klaus, liebst mich?
Klaus, love 2 SG PRES me ACC SG
Klaus, do you love me?

(14) * Sehr das Kind liebt seine Großeltern. (P3)
very the kid NOM SG loves 3 SG PRES his grandparents ACC PL
The kid loves very much his grandparents.

Also, 144 distractor sentences were used, half of which were grammatically correct while the other half were not.

The participants

Grammaticality judgements of 73 people that formed three groups were taken into consideration. 49 NNS constituted two homogeneous groups with different levels of proficiency in German (basic: B1 and advanced level: C1), but at the same level of proficiency in English (advanced level: C1) (according to the Common European Framework of Reference for Languages, Council of Europe, 2007). The third group consisted of 24 native speakers of German (who also had an advanced level in English) and served as a control group (CG). All NNS participated in placement tests for English and German language and completed a questionnaire on their demographic data. They were asked to characterize every experimental sentence choosing a mark from a five-grade *Likert* scale (Jamieson, 2004): *5* if they would say this sentence for sure, *1* if they would definitely not say this sentence, etc.

Research hypothesis

In order to investigate the source of language transfer in the interlanguage of the NNS, the tested experimental conditions were grouped in two main cases.

A) An experimental condition, where there is similarily in L1 (Greek) and L3 (German), but difference in English (L2). This is the case when the subject of the finite verb can be omitted (always in Greek, or under certain circumstances, i.e. only in the passive voice of certain verbal classes in German) (experimental condition: [R./ -lex. sub./ -es]).

B) An experimental condition, where a number of phenomena are being investigated, i.e. German (L3) has the same syntactic properties as English (L2), but at the same time it differs from Greek (L1). Such is the case where a subject is required, and especially where an expletive subject is allowed (experimental condition: [R./ +lex. sub./ +es]). Grammaticality judgements for these two experimental conditions were juxtaposed with common in all three languages experimental conditions, where there is at least one subject for the finite verb (experimental conditions: [R./ +lex. sub./ -es] and [R./ -lex. sub./ +es]).

The research hypothesis was: if grammaticality judgements of the NNS were more successful in case A, then we assume that L1 has more influence on their L3 interlanguage. Conversely, if grammaticality judgements of the NNS were more successful in case B, then we assume that L2 has greater influence when acquiring L3.

There is, of course, the possibility that grammaticality judgments of the NNS prove to be equivalent or approximately the same in both cases (A and B). This means that neither of the acquired languages has greater influence on L3 interlanguage. In this case, other differences must be investigated in order to reach a conclusion about the source of linguistic transfer in L3 interlanguage.

Results and discussion of the theoretical hypotheses

In this section, results of the GJT are discussed in conjunction with the four theoretical approaches about L3A mentioned above.

Developmentally moderated transfer hypothesis (Håkansson et al., 2002)

According to this hypothesis, the L1 still has a privileged role in L3A. In Figure 1, the results of experimental sentences containing verbs are presented, which allow subject omission in the passive voice (case A).

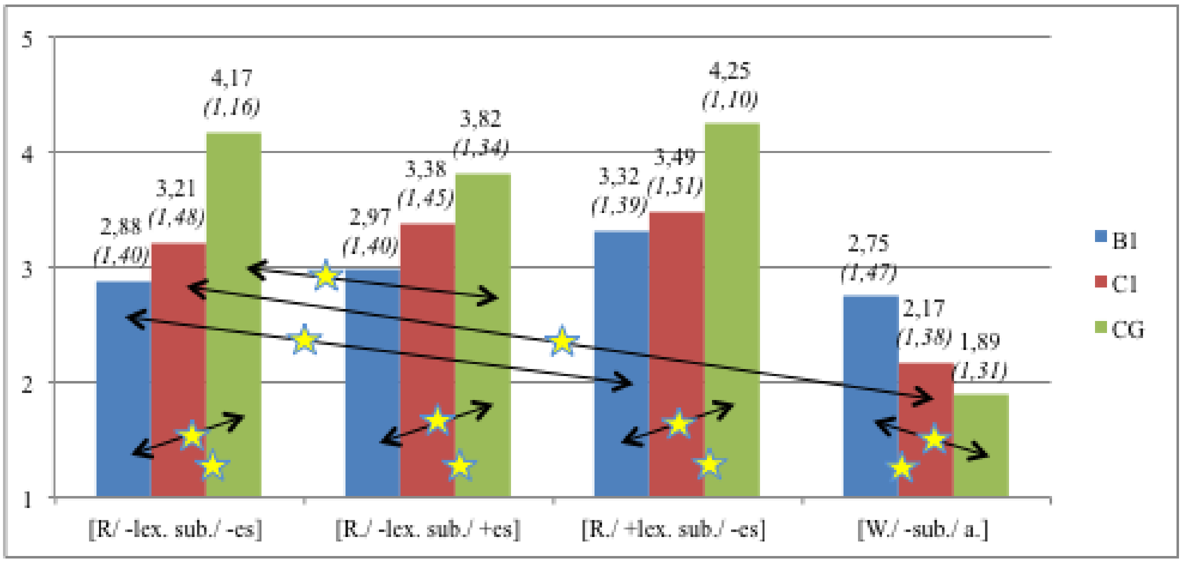


Figure 1. Results for sentences containing verbs that allow subject omission in passive voice

The results indicate that in all comparisons that engage structure without any subject (experimental condition: [R./ -lex. sub./ -es]), the NNSs of both experimental groups consider the other structure to which it is compared as grammatically better (i.e. where there is at least one subject that is either lexical or expletive (experimental conditions: [R./ +lex. sub./ -es] and [R./ -lex. sub./ +es]). It seems that although German has increased verbal morphology, this does not fulfill requirements for Agreement, since there is still a need for morphophonological realization of the subject of the finite verb.

Of particular interest is the comparison between grammaticality judgements on wrong sentences without a subject in the active voice (experimental condition: [W./ -lex. sub./ -es]) to sentences containing equivalent verbs in the passive voice, where subject omission is allowed (experimental condition: [R./ -lex. sub./ -es]). This allows comparison of the participants' judgements for sentences containing verbal classes that can be found without a subject forming correct sentences in the passive voice, but wrong ones in the active voice.

This comparison leads us to conclude that, even at a high level of proficiency in German as L3, the NNS are hesitant to accept as correct a structure without a subject in the passive voice, although they recognize it, scoring a statistically significant difference when they encounter it in a wrong sentence in the active voice. The results of this comparison indicate with relative certainty that the NNS do not transfer into their L3 interlanguage a structure without a subject from their native language where it is grammatically correct.

Second language status factor hypothesis (Williams & Hammarberg, 1998)

According to this hypothesis, L2 has more influence than L1 on the L3 interlanguage of NNSs. In order to verify this hypothesis, comparisons were made between acceptable structures in German and English (L3 and L2, respectively), but not grammatically correct in Greek (L1). In these structures, both a lexical and the expletive subject '*es*' coexist (experimental condition: [R./ +lex. sub./ +es]) with other experimental conditions where there is only one subject, either only a lexical one or only an expletive one '*es*' (experimental conditions: [R./ +lex. sub./ -es] and [R./ -lex. sub./ +es]).

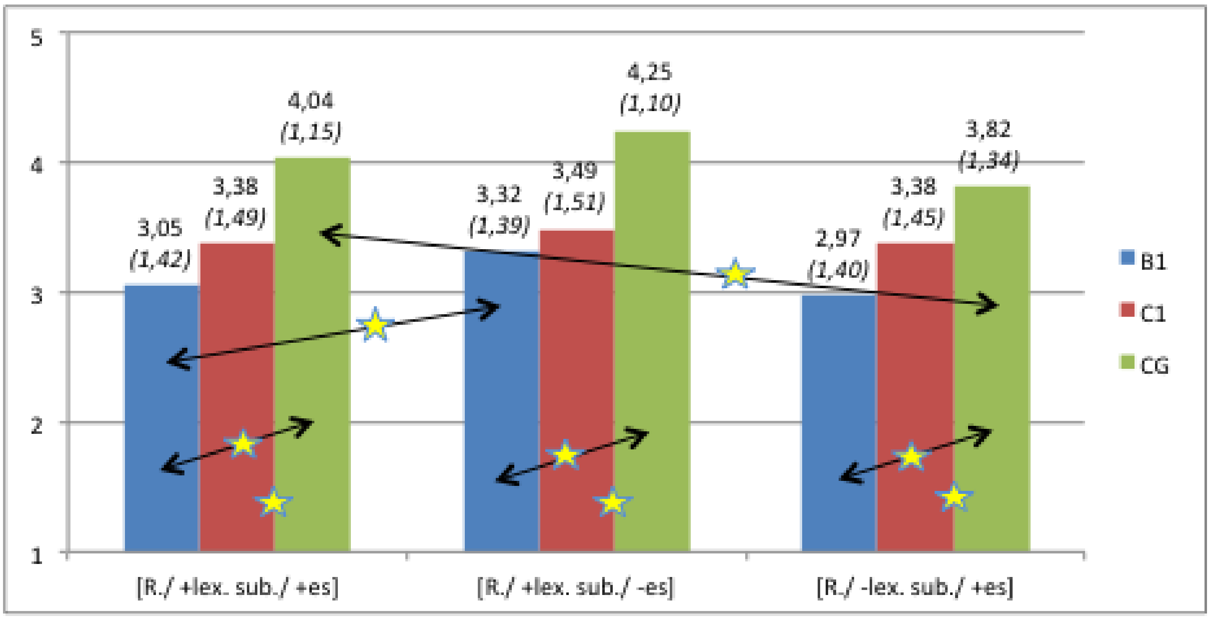


Figure 2. Results for sentences containing verbs that do not allow subject omission

The NNS seem to consider as more grammatical an experimental condition in which there is only one subject. There is actually a rising tendency between the performance of individuals I groups B1 and C1, with the first appearing less certain about their choices and having inconstant judgements. Conversely, people with very good knowledge of German made choices that were largely consistent.

The results of both groups of the NNS do not show that there is particular influence on their L3 interlanguage from their L2 (English).

The general conclusion regarding the two hypotheses is that the NNS do seem to transfer into their L3 interlanguage structures that exist in both their L1 and L2. For example, common structures in the three languages seem to be more transferable in contrast to structures that exist either in their L1 or L2. The NNSs resort to the choice of structure they are familiar with and which is acceptable in both languages they already know. An examination of the first two theoretical models shows that it is neither the L1 nor the L2 that play dominant roles in shaping the L3 interlanguage of the NNSs.

Next, the last two theoretical models proposed about the L3A are examined.

Cumulative-enhancement model for language acquisition (Flynn et al., 2004)

According to this model, every language already acquired is important and available to the same degree to play a role when acquiring an additional language. It contributes to the development of syntactic structure in any subsequent language in a positive or neutral manner. There is either "positive linguistic transfer" or no transfer at all to the targeted language.

If this model applies to the present data**,** then the structures that appear either only in the L1 or in the L2 are transferred more easily to their L3 interlanguage. However, according to the data presented above, the NNS are hesitant to choose structures that appear only in their L1 or in their L2. They feel more confident to choose structures that are acceptable in both languages they already know. However, differences in common structures in all three languages are not assessed as statistically significant.

In the hypothetical case that the NNSs had not previously acquired Greek (L1) that allows omission of the subject or English (L2) where expletive subjects are allowed, but in some other language that does not have these syntactic properties, NNSs would not formulate grammaticality judgements that are so target-like when acquiring L3 German. In this hypothetical case, statistically significant differences may be expected between grammaticality judgements for these particular structures compared to others that appear in both languages already acquired. In this study, no such differences were noted. Therefore, we assume that prior knowledge of the languages in which syntactic structures exist rather facilitates L3A, compared to the hypothetical case in which the NNS would face these particular structures for the first time in L3. Of course, in order to strengthen this supposition, results of this research should be compared with experimental data in other researches studying participants with different L1 and L2 backgrounds than those tested in the present study in order to compare the grammaticality judgements of the participants. If in such a comparison, we notice statistically significant differences between grammaticality judgements about these structures between two groups with different L1s and L2s, and the people in the present research having performed better than the group in the other study, then we could confidently admit that the Cumulative Enhancement Model adequately describes the present experimental data. Otherwise, we would have considerable evidence that the model is not sufficient for their interpretation.

Typological primacy model (Rothman, 2011)

The last theoretical model examined is the Typological Primacy Model proposed by Rothman (2011). According to this model, both formal linguistic typology and psychotypology play an important role in acquiring a new language. By psychotypology, we mean the speakers' subjective perception of about the distance or proximity between two languages (Kellerman, 1977; 1992). In this particular case, two of the languages studied - English (L2) and German (L3) - are connected genetically, as they belong to the same subgroup of the Indo-European language family, i.e. German languages. In addition, other factors, like a common alphabet could make English and German appear closer according to the NNSs' psychotypological perception, which however cannot be controlled with an objective criterion. Based on this, it is obvious that there is both objectively historical typological proximity but also subjectively psychotypological similarity between English and German (greater than the one between Greek and German).

Therefore, if this model is applied effectively to the present data, then the grammaticality judgments of the NNSs would be more target-like for the experimental conditions that exist only in English (as compared not only to the structures lacking in English or in Greek), but also to the common structures that exist in both languages (Greek and English). However, there were not such findings in the results of the statistical analyses applied to the data.

A general remark on methodology would be that, even if we accept that the NNSs' preferences show a slight advantage in structures that appear only in English (L2) as compared to other structures, it could be argued that it is not (psycho)typological proximity that plays a significant role. Instead, it is the L2 Factor Hypothesis that influences language transfer; this hypothesis is actually verified by the experimental data.

In order to be able to verify the Typological Primacy Model, it is necessary to have experimental data from another group of people with English as L1, Greek as L2, when acquiring L3 German. To argue that typological closeness plays the most important role, regardless of the chronological order according which the NNSs have acquired their languages before starting L3 acquisition.

Conclusion

According to this study's experimental data and resulting analysis, it appears that none of the theoretical models fully describes the NNSs' L3 interlanguage. However, it could be concluded that the Cumulative-Enhancement Model for Language Acquisition (Flynn et al., 2004) is the one that describes L3 interlanguage best, because it is in agreement with the experimental data here. Ideally, data here should also be compared with data from individuals with different linguistic backgrounds, as mentioned above.

Interest in L3A remains large and can only grow more. Given that L2 acquisition theory can contribute significantly to the development of linguistic theory, then obviously the study of L3A and multilingualism can contribute to this in an even greater extent. As the data show, L3A may be a rich source of information for linguistic theory and can reveal different kinds of language economy rules that could eventually help us understand better the function of the language system.

References

Cabredo Hofherr, P. (1999). Two german impersonal passives and expletive pro. *Catalan Working Papers in Linguistics*, *7*, 47-57.

Cabredo Hofherr, P. (2003). Arbitrary readings of third person plural pronominals. In M. Weisgerber (ed.), *Proceedings of the Conference Sinn und Bedeutung 7* (pp. 81-94). Universität Konstanz.

Chomsky, N. (1981a). *Lectures on government and binding: The Pisa lectures*. Dordrecht: Foris.

Chomsky, N. (1981b). Principles and parameters in syntactic theory. In N. Hornstein & D. Lightfoot (eds.), *Explanation in linguistics: The logical problem of language acquisition* (pp. 32-75). London, UK: Longman.

Council of Europe. (2007). *Common European framework of reference for languages*. Retrieved from http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf on 10.11.2015.

D' Alessandro, R. (2014). *The null-subject parameter: Where are we and where are we headed?* Leiden University Centre for Linguistics.

Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *The International Journal of Multilingualism*, *1*(1), 3-16.

Gass, S. (1996). Second language acquisition and linguistic theory: The role of language transfer. In W. Ritchie & T. Bhatia (ed.), *Handbook of second language acquisition* (pp. 317-345) San Diego: Academic Press.

Håkansson, G., Pienemann, M., & Sayheli, S. (2002). Transfer and typological proximity in the context of second language processing. *Second Language Research*, *18*, 250-273.

Han, Z. H., & Tarone, E. (2014). *Interlanguage: 40 years later.* Amsterdam: John Benjamins.

Holmberg, A., Nayudu, A., & Sheehan, M. (2009). Three partial null-subject languages: A comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica*, *63*(1), 59-97.

Huang, J. C.-T. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry*, *15*, 531-571.

Jaeggli, O. (1982). *Topics in Romance syntax*. Dordrecht: Foris.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*(12), 1,217-1,218.

Kellerman, E. (1977). Towards a characterization of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin*, *2*, 58-145.

Kellerman, E. (1983). Now you see it, now you don't. In S. Gass & L. Selinker (eds.), *Language transfer in language learning* (pp.112-134). Rowley: Newbury House.

Klein, E. (1995). Second vs. third language acquisition: Is there a difference? *Language Learning*, *45*(3), 419-465.

Leung, Y.-K. I. (2007). L3 acquisition: Why is it interesting to generative linguistics. *Second Language Research*, *23*(1), 95-114.

Odlin, T. (1989). *Language transfer: Crosslinguistics influence in language learning*. Cambridge, UK: Cambridge University Press.

Odlin, T. (2003). Crosslinguistic influence. In C. Long & M. Doughty (eds.), *The Handbook of Second Language Acquisition* (pp. 436-486). Malden, MA: Blackwell.

Rizzi, L. (1982). *Issues in Italian syntax*. Dordrecht: Foris.

Rizzi, L. (1986). Null objects in Italian and the theory of Pro. *Linguistic Inquiry*, *17*, 501-557.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, *27*, 107-127.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, *10*, 209-130.

Sharwood-Smith, M. (1994). *Second language learning: Theoretical foundations*. London, UK: Longman.

Williams, S., & Hammarberg, B. (1998). Language switches in L3 production: Implications of a polyglot speaking model. *Applied Linguistics*, *19*, 295-333.

White, L. (1989). *Universal grammar and second language acquisition*. Amsterdam: John Benjamins.

Vowel reduction in early Spanish-English bilinguals; how native is it?

Kelly Millard, Mehmet Yavaş
kmill072@fiu.edu, yavasm@fiu.edu

Florida International University

Abstract. This study acoustically investigates the duration of English reduced vowels in unstressed syllables produced by early Spanish-English bilinguals. The aim of the study is to determine whether the bilinguals' productions of English reduced vowels match the norm provided by monolingual English speakers. The vowels were analyzed in two different stress environments and the frequency in which the word containing the vowel occurs in everyday speech was also measured, to determine whether or not these two factors contribute to the duration of the vowel that is produced. The productions of these vowels by Spanish-English bilinguals were compared to a control group consisting of monolingual English speakers in order to determine the amount of deviation. The results confirm that there is in fact a statistically significant difference in the duration of the reduced vowels between Spanish-English bilinguals and monolingual English speakers, and support the view that even early exposure to L2 may not be enough for bilinguals to acquire native-like phonetic patterns in L2.

Keywords: bilingualism, vowel reduction, prosodic environment, frequency

Introduction

It is a well-known fact that vowel reduction, which is one of the typical characteristics of stress-timed languages, is a commonly occurring phenomenon in Standard American English (hereafter SAE) (Flemming, 2009). This vowel reduction is a result of contrasting vowel qualities becoming neutralized and it occurs in unstressed syllables (Chomsky & Halle, 1968). Spanish, as a typical syllable-timed language, does not have this feature. For example, if we consider the English word *probability* and its cognate in Spanish, *probabilidad*, we see the difference very clearly. The two words share sounds, the same meaning and the same number of syllables, but the similarities do not go beyond that. In Spanish, the stress is on the last syllable. Although the remaining syllables are unstressed, they all have full vowels. In English on the other hand, the word reveals a rather different picture: The third syllable receives the primary stress, and the first syllable has a secondary stress, and thus these two syllables have full vowels. The second and fourth syllables are unstressed and have reduced vowels (schwas). Consequently, such differences result in the different rhythms in the two languages.

Vowel reduction is very frequent in English; vowels can be reduced to a schwa /ə/ (an unstressed centralized mid vowel) when they are in an unstressed syllable. Unlike stressed vowels, vowels reduced to a schwa are not produced with their full phonetic value in English (Chreist, 1964). An example of this stress and reduction pattern can be seen in the English words *photograph* [fótəgræ̀f] → *photography* [fətɑ́grəfi]; in the first word, the first syllable is stressed and it has a full vowel, and the second syllable is unstressed and squeezed between syllables with the primary and secondary stresses, and thus has a reduced vowel. In the second word, the primary stress shifts to the second syllable and the vowel becomes full. Since the first syllable is unstressed right before the primary stress, its vowel is reduced.

Previous studies have shown an average value of schwa duration by native English speakers to be around 55 to 64 milliseconds, while full vowels, such as /i/ and /o/, in stressed syllables can reach up to 156 milliseconds (Yavaş, 2011; Flemming, 2009). According to Chreist (1964), this vowel reduction rule is an important feature of American English for L2 learners, which is not relevant in other languages, such as Spanish. Ignoring the rule of stress and vowel reduction will result in a foreign accent. Halle, Morris, and Vergnaud (1987) also describe this as a "striking phonetic property of English" (p. 239). Therefore, the differentiation between full vowels and reduced vowels in unstressed syllables is vital when learning English as a second language. Because vowels present more challenges than consonants in L2 acquisition, and because Spanish lacks the vowel reduction process altogether, it is reasonable to think that L1 Spanish speakers who learn English are likely to have difficulties in mastering the vowel reduction patterns of English.

Although there are studies supporting the claim that people who learn a second language before the end of the critical period (puberty) have a much better chance of achieving native-like pronunciation, as opposed to learners who learn a second language after the end of the critical period, several recent studies have shown that a lag of even a few years in acquiring an L2 tends to have dramatic consequences on both speech production and perception (Fowler, Sramko, Ostry, Rowland, & Halle 2008; Flege & MacKay, 2004; Sebastian-Galles & Soto-Faraco, 1999). Also, recent comprehensive and detailed linguistic analyses of early learners have revealed that even very low ages of acquisition (hereafter AOA) do not automatically result in completely native-like L2 proficiency (Abrahamsson, 2012; Stolten, Abrahamsson, & Hyltenstam, 2014). To what extent Spanish-English bilinguals' productions match the monolingual English patterns is the central question addressed in this paper.

As mentioned earlier, vowels in English can only be reduced to schwa when the syllable is unstressed. Although both syllables can be stressed in a disyllabic word, two stressed syllables are not generally found in a row in words that are 3 syllables or more (Yavaş, 2011). So, if a word has two stressed syllables, the unstressed syllable with the reduced vowel will be found between the two stressed syllables. Because Spanish does not have vowel reduction in unstressed syllables, a contrast in duration and overall intensity between syllables occurs which results in the stressed vowel being produced longer or with more intensity than the norm (Ortega-Lebaria & Prieto, 2009). Bearing in mind this relationship between stress and reduced vowels in English, it is relevant to consider the stress patterns when analyzing reduced vowels between the two languages. Therefore, the vowels will also be analyzed in two different stress environments for this study. These are: a) post-secondary and pre-primary stress, (hereafter Stress 1), as in *constitution* [kànstətúʃən], and b) post-primary and pre-secondary stress (hereafter Stress 2), as in *satisfied* [sǽtəsfaìd] In both environments, the vowel is located in between two stresses.

It is possible that the position of the vowel in reference to the primary stress may be a contributing factor to the reduction of the vowel with the expectation that the schwa, which occurs before primary stress (in the stress 1 position), will undergo more of a reduction since it is in a weaker position. By looking at the vowels in the two different stress environments, it can be determined whether, or not, the type of stress contributes to the length of the vowel.

Word frequency may also be a contributing factor in the accuracy of reduced vowel pronunciation. Frequency counts determine how frequently a word is said or used and, according to Fabiano-Smith and Goldstein (2010), a higher frequency is linked with a greater accuracy rate than a lower frequency. It may be possible that the production of reduced vowels by Spanish-English bilinguals in words with a higher frequency may be closer to the average norm of monolingual English speakers than the reduced vowel in words with a lower frequency.

The purpose of this research is to determine whether or not early Spanish-English bilinguals (having learned English before the age of 9) do in fact produce an unstressed vowel with a duration that is measurably different than that of the average native monolingual English speaker, despite the fact that they have learned English during the critical period, and appear to have fluency similar to that of a native speaker. This will be determined by acoustically analyzing the phonetic and temporal qualities of the Spanish L2 production of the unstressed vowel in the two stress environments, stress 1 and stress 2. Once the length of the vowel is measured, it will be compared to that of native English speakers.

Putting all the above together, we have the following main hypothesis:

H.1- Early Spanish-English bilinguals (with English dominant fluency) will produce English reduced vowels with longer duration than those of monolingual English speakers.

This hypothesis will be supplemented with the following ancillary one.

H.2- Different stress environments and word frequency will be influential factors in the duration of the reduced vowels.

Method

Participants

The participants of this study are 40 Spanish-English bilinguals and 40 monolingual English speakers. The first group consisted of early Spanish-English bilinguals who learned English before the age of 9, (mean age at the time of language exposure 3;8, and 27 was the mean age at the time of participation). They belong to the very typical local pattern, whereby the children are in a Spanish-speaking environment until they begin their education. Although they are typically Spanish monolinguals until they start kindergarten, the language dominance shifts to English through elementary education, and strengthens more thereafter. The second group consisted of monolingual English speakers; 44 was the mean age at the time of participation. All of the participants in the study were university level students or adults living in the United States and all were able to understand, speak, read, and write English.

Stimuli

Each participant performed a reading task in which they were instructed to read 20 English sentences while being audio recorded. Each sentence contained a target word strategically placed in the middle of the sentence to avoid putting emphasis on that particular word. The target word contained a schwa in one of the two stress environments; ten out of twenty words contained a schwa in stress 1 and the remaining ten out of twenty words contained a schwa in stress 2. Words with sonorant consonants neighboring the schwa were avoided to obtain a clearer reading and a more accurate measurement of the vowel duration. Table 1 displays examples of the stimuli used.

Table 1. Examples of stimuli used from each stress environment

Stress 1	Stress 2
The priest used an *invocation* to begin the service.	The guests were *satisfied* with the service.
I like to eat *avocado* in my salad.	I hope that I *recognize* everyone at the reunion.
I try to avoid *repetition* in my day to day life.	A library *database* is used to search for information.

Participants were given a Language Background Questionnaire at the start of the procedure. They were then instructed to read the twenty English sentences that were presented individually on a computer screen via PowerPoint while being audio recorded. Recordings were saved at 44100 Hz sampling rate and were segmented and analyzed using PRAAT speech analysis software version 5.4.10 (Boersma & Weenink, 2015).

Data yielded 800 tokens (20 targets x 80 participants). However, some of the data were unable to be used due to pronunciation errors. There were consistent pronunciation errors in the word “pedagogue” for both monolingual and bilingual groups. There were also some deletions among the monolinguals, particularly in the word “invocation” and “convocation”, due to the vowel production being longer than the normal range of schwa for monolingual English speakers. This may be a result of over pronunciation due to the spelling of those particular words. In general, monolingual schwa productions above 70ms were not included as part of the control measurements. A small number of words also underwent schwa deletion and were therefore unusable. The final number of tokens used in the study was 765 (388 for bilinguals and 377 for monolinguals).

Analysis

The reduced vowel targets produced by the Spanish-English bilingual group were compared to those produced by the English monolingual group. T-tests were conducted in order to make comparisons of the length of the reduced vowel between the two groups and within each stress environment. In other words, stress 1 of the monolinguals was compared to stress 1 of the bilinguals and stress 2 of the monolinguals was compared to stress 2 of the bilinguals. A pairwise t-test was also conducted in order to compare the stress environments within each group; stress 1 was compared with stress 2 within the bilingual group and another pairwise t-test compared stress 1 with stress 2 within the monolingual group.

The frequency of occurrence for each target word was recorded as well using the Corpus of Contemporary American English (COCA) (Davies, 2008) word frequency database that is based on a 450 million word-list. A t-test was conducted in order to compare the three most frequent words between the monolingual group and the bilingual group, as well as the three least frequent words between the monolingual group and the bilingual group. Finally, the three most frequent words were compared with the three least frequent words for stress 1 within the bilingual group, and also within the monolingual group, separately. The same was done for stress 2.

Results and Discussion

Item Level Averages

The mean duration of the schwa was first determined for each individual word for monolinguals and bilinguals, separately. In all words, the bilingual group resulted in a larger mean duration than the monolinguals with the largest difference being 18.3 ms. in the word *pedagogue* and the smallest difference being 3.6 ms. in the word *recognize.*

Durational T-test Results by Stress Environment

An independent sample t-test was conducted to compare the average time it took to produce a reduced vowel between bilinguals and monolinguals. This was done in both stress environments as shown in Table 2. In stress 1, the results suggest that bilinguals take significantly longer ($M=46.01$, $SD= 6.27$) than monolinguals ($M=36.02$, $SD= 5.07$); $t(78)= 7.833$, $p<.05$. In stress 2, the bilinguals again take significantly longer ($M=51.70$, $SD= 7.71$) than monolinguals ($M= 40.16$, $SD= 4.96$); $t(78)= 7.833$, $p<.05$.

Table 2. Mean duration of reduced vowels produced by monolinguals and bilinguals in each stress environment

	Monolinguals	**Bilinguals**	**P-value**
Stress 1	36.02	46.01	<.05
Stress 2	40.16	51.70	<.05

Figures 1 and 2 show the stress 1 and stress 2 results displayed in separate box plots. The graphs clearly show that the monolingual group produces a vowel that is shorter in duration than the bilingual group in both stress environments. There is also a larger standard deviation among the bilingual group than the monolingual group.

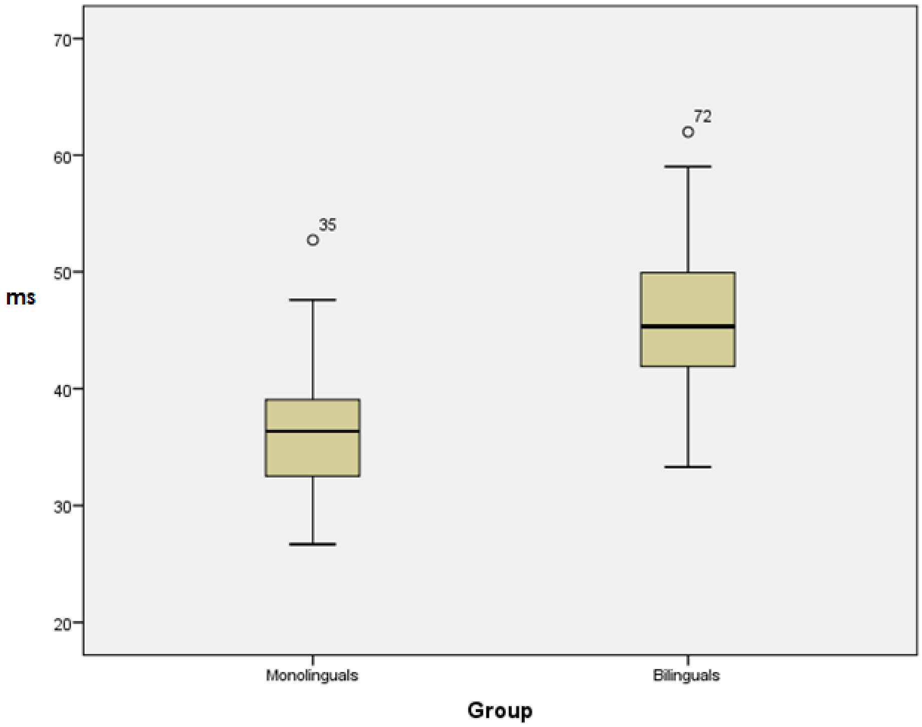


Figure 1. Reduced vowel duration in monolingual and bilingual productions within stress 1

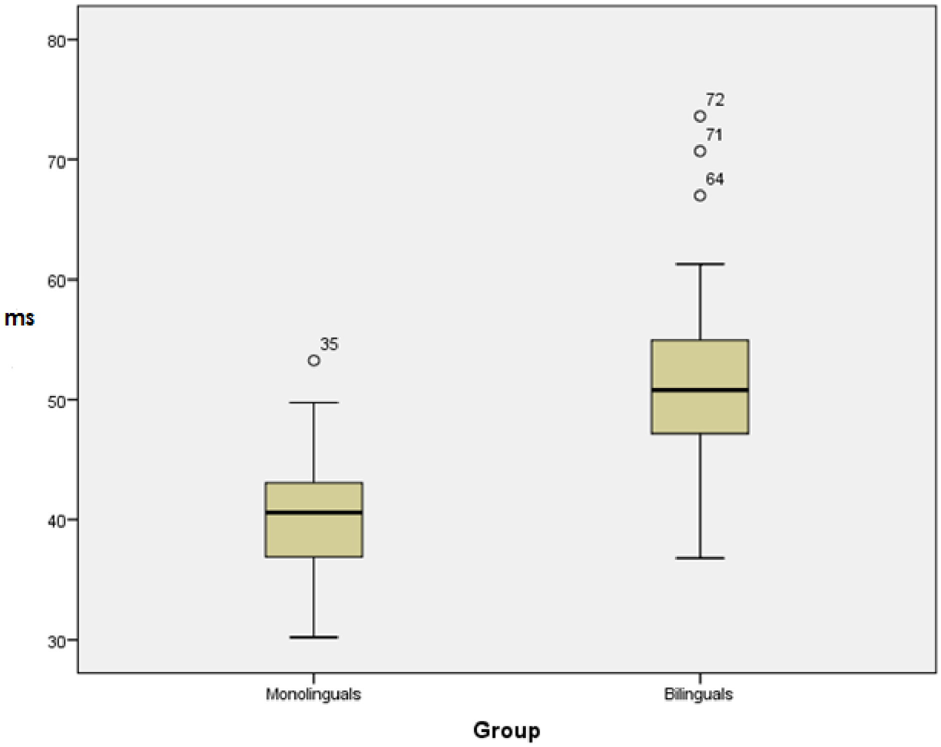


Figure 2. Reduced vowel duration in monolingual and bilingual productions within stress 2

A paired sample t-test was conducted to compare the average time it took to produce a reduced vowel between the two stress environments for each group, separately, to determine if there is a significant difference between the stress environments. The results shown in Table 3 suggest that, for monolinguals, the vowels produced in stress 2 are in fact significantly longer (M=40.16, SD= 4.96) than the vowels produced in stress 1 (M=36.02, SD=5.07); $t(39)= 5.939$, $p<.05$, and for bilinguals, the vowels in stress 2 position are significantly longer (M=51.70, SD=7.71) than the vowels in stress 1 position (M=46.68, SD=5.99); t(38)= 5.815, p<.05

Table 3. Mean duration of reduced vowels produced in each stress environment for each group

	Stress 1	Stress 2	P-value
Monolinguals	36.02	40.16	<.05
Bilinguals	45.68	51.70	<.05

Word Frequency T-test Results

A paired sample t-test was conducted in each stress environment for the monolingual and bilingual group, separately, to compare the average duration of the vowel in the three most frequent words with the three least frequent words. The results are displayed below in Table 4.

Table 4. Mean vowel duration in the most and least frequent words for each group and stress environment.

	Stress 1-Most Frequent	Stress 1-Least Frequent	P-value	Stress 2-Most Frequent	Stress 2-Least Frequent	P-value
Monolingual	29.45	40.08	<.05	41.38	44.08	.051
Bilingual	38.63	49.75	<.05	50.53	55.03	.007

In stress 1, there is a significant difference suggesting that the least frequent words (M=40.08, SD = 9.17) produced by monolinguals are longer than the most frequent words (M=29.45, SD= 6.54); t(39)= 5.873, p<.05; In bilingual production, the least frequent words (M= 49.75, SD= 11.96) are longer than the most frequent words (M= 38.63, SD= 9.61); t(39)= 6.236, p<.05.

In stress 2, no significant difference in the monolingual comparison was found, however longer vowels were observed in the most frequent words (M= 44.08, SD=7.27) than in the least frequent words (M=41.38, SD=7.20). It is expected that a larger sample size may produce significant results for this comparison. On the other hand, the vowels in the least frequent words produced by bilinguals (M= 55.03, SD= 9.24) were significantly longer than those in the most frequent words (M= 50.53, SD= 10.27); t(39)= 2.861, p=.007.

Table 5. Mean Monolingual and Bilingual productions in frequent and infrequent words within each stress environment.

	Monolingual	Bilingual	P-value
Most Frequent Stress 1	29.45	38.63	<.05
Most Infrequent Stress 1	40.08	49.75	<.05
Most Frequent Stress 2	41.38	50.53	<.05
Most Infrequent Stress 2	44.08	55.03	<.05

Presented in Figures 3-6 are the box plots corresponding to the above findings.

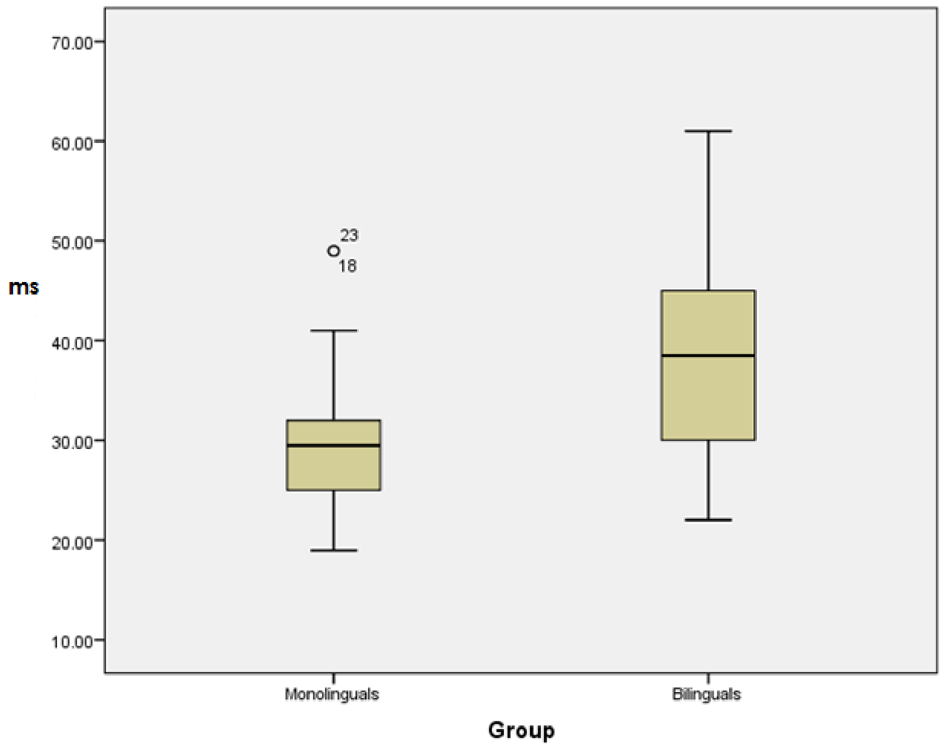


Figure 3. Monolingual and Bilingual production of most frequent words in stress 1

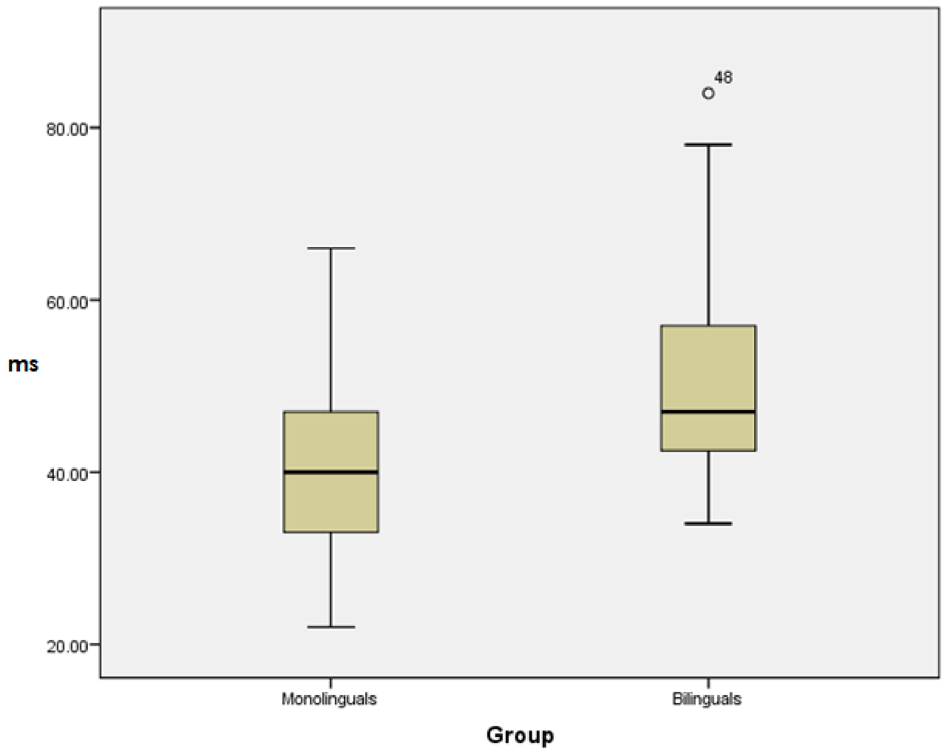


Figure 4. Monolingual and Bilingual production of least frequent words in stress 1

A t-test was then conducted to compare the duration of the reduced vowel between monolinguals and bilinguals in the frequent and infrequent words, separately, and within each stress environment, separately. In all comparisons, the bilinguals produced a longer reduced vowel. In stress 1, bilingual production was significantly longer in frequent (M= 38.63, SD= 9.61) and infrequent words (M=49.75, SD=11.96) than that of the monolinguals (M= 29.45, SD= 6.54); t(78)= 4.99, $p<.05$,

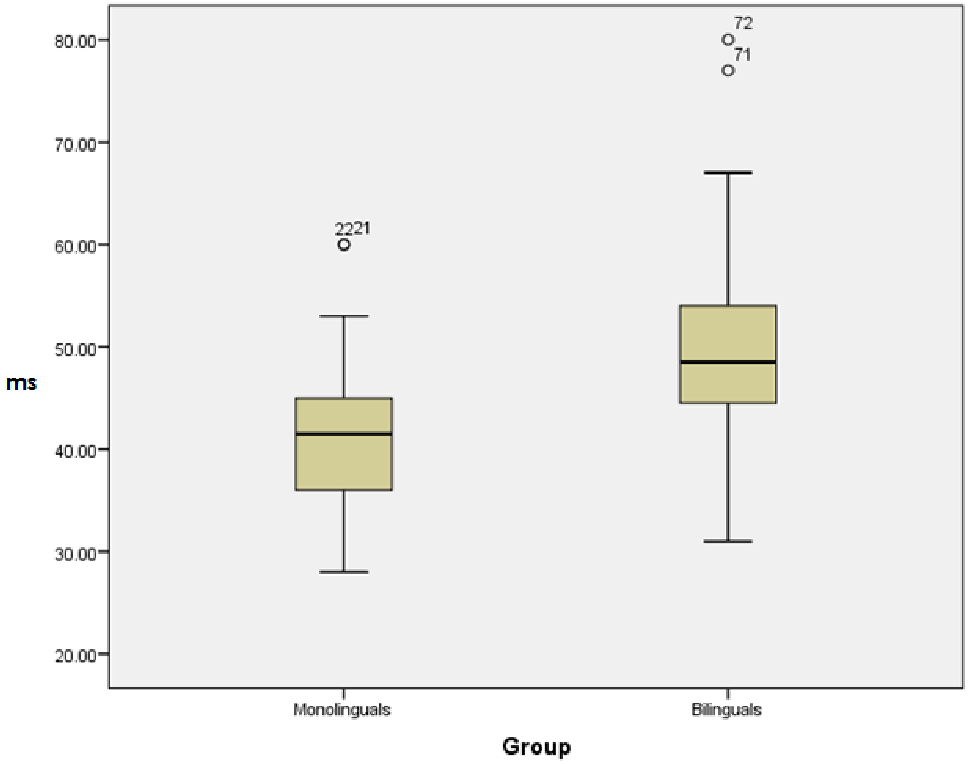


Figure 5. Monolingual and Bilingual production of most frequent words in stress 2

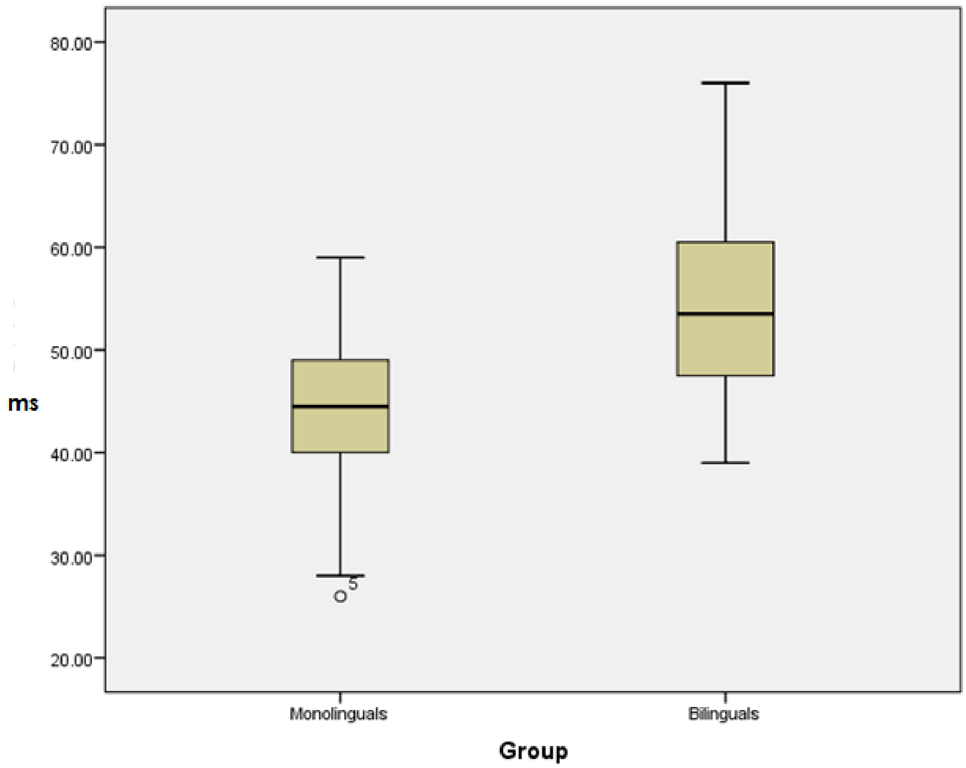


Figure 6. Monolingual and Bilingual production of least frequent words in stress 2.

(M=40.08, SD=9.17); t(78)=4.06, $p<.05$. In stress 2, the bilingual production was significantly longer in frequent (M= 50.53, SD= 10.27) and infrequent (M=55.03, SD=9.24) words than that of monolinguals (M= 41.38, SD= 7.20); t(78)= 4.61, $p<.05$, (M=44.08, SD=7.27); t(78)= 5.89, $p<.05$. Table 5 displays these results for the comparisons of monolingual and bilingual productions of the reduced vowel in frequent and infrequent words in each stress environment.

Conclusion

The results confirm the hypotheses that, despite having learned English before puberty and becoming English dominant regarding fluency, early Spanish-English bilinguals do not match the norms regarding American English reduced vowels provided by monolingual speakers; the productions of bilinguals were significantly longer than those produced by monolingual English speakers. The findings are in agreement with other studies (Bosch et. al., 2000; Dupoux, Peperkamp, & Sebastian-Galles 2010) in that even early exposure to L2 may not be enough for bilinguals to acquire native-like phonetic patterns in L2.

The stress environment in which the schwa occurred also seems to be a contributing factor to vowel length. There is a significant difference between monolinguals and bilinguals in both stress 1 and stress 2 environments, both of which resulted in the bilinguals producing a much longer vowel than monolinguals. Moreover, the vowels in stress 2 environment appear to be longer than those in stress 1 when each group was looked at individually. In other words, these results suggest that, for both monolinguals and bilinguals, there is more of a reduction in the stress 1 environment. This is a logical outcome considering that the vowel that is positioned before primary stress, rather than secondary stress, is in weaker position and is, therefore, expected to undergo a more severe reduction.

The frequency tests for both stress environments were as expected in accordance with Fabiano-Smith and Goldstein (2010) in that the words of lower frequency had a significantly longer schwa duration than the most frequent words. In other words, the more frequent words were produced with a schwa that was closer to the native production. This was the pattern for both the monolingual group and the bilingual group confirming the hypothesis that accuracy is linked with word frequency.

References

Abrahamsson, N. (2012). Age of onset and ultimate attainment of L2 phonetic and grammatical intuition. In N. Abrahamsson & K. Hyltenstam (eds), High-level L2 acquisition, learning and use [Special Issue]. *Studies in Second Language Acquisition*, *32*(2). Cambridge, UK: Cambridge University Press.

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [PRAAT]. Version 6.0.04, retrieved 1 November 2015 from http://www.praat.org/

Bosch, L., Costa, A., & Sebastian-Galles, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*, 189-221.

Chomsky, N., & Halle, M. (1968). *The sound age of onset pattern of English*. New York: Harper & Row.

Chreist, F. M., (1964). *Foreign accent*. Englewood Cliffs, NJ: Prentice-Hall.

Davies, M. (2008) *The corpus of contemporary American English: 450 million words, 1990 present*. Available online at http://corpus.byu.edu/coca/

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress 'deafness' in simultaneous French–Spanish bilinguals. *Cognition*, *114*, 266-275.

Fabiano-Smith, L., & Goldstein, B. A. (2010). Phonological acquisition in bilingual Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research*, *53*, (1), 160-178.

Flege, J. E., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1-34.

Flemming, E. (2009). The phonetics of schwa vowels. In D. Minkova (ed.), *Phonological Weakness in English.* Hampshire, England: Palgrave Macmillan Publishing.

Fowler, C., Sramko, V., Ostry, D., Rowland, S., & Halle, P. (2008). Cross-language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics*, *36*, 649-663.

Halle, M., & Vergnaud, J. (1987). *An essay on stress*. Cambridge, MA: MIT.

Ortega-lebaria, M., & Pilar, P. (2009). Perception of word stress in Castilian Spanish: The effects of sentence intonation and vowel type. In M. C. Vigario, S. Frota, J. Freitas (eds.), *Phonetics and Phonology: Interactions and Interrelations* (pp. 35-50). John Benjamins Co.

Sebastian-Galles, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*, 111-123.

Stolten, K., Abrahamsson, N., & Hyltenstam, K. (2014). Effects of age of learning on voice onset time: Categorical perception of Swedish stops by near-native L2 speakers. *Language and Speech*, *57*(4), 425-450.

Yavaş, M. (2011). *Applied English phonology*. Oxford, UK: Wiley-Blackwell.

Bilingual language and speech patterns: Evidence from English (L1) and Greek (L2)

Eleni Morfidi, Eleni Samsari
emorfidi@cc.uoi.gr

University of Ioannina

Abstract. The present study builds on theory and research highlighting the contributions of oral language to literacy development. The focus is on phonological development and overall language efficacy in relation to literacy skills. The paradigm of bilingualism is used to address issues of concern in the area of oral-written language development. More specifically, the present study aims to identify patterns of errors in phonological production during L2 speech and investigate the relationships between speech, language and literacy skills among bilinguals who learn both English (L1) and Greek (L2). The sample consisted of 13 children (8 boys, 5 girls) (ages 8.5 to 12 years). The length of their instruction in US Greek community schools ranged from 2 to 6 years. Measures targeted oral and written language skills in both English and Greek. Narrative procedures have been used to extract samples of words for analyzing patterns of speech errors, and derive percentages of vowels (PVC), consonants (PCC) and phonemes correct (PPC). Furthermore, mean length of utterance (MLU), fluency (words per minute, WPM), expressive vocabulary (number of different words, NDW) and narrative structure (NSS), have been calculated in both languages. Phonological awareness, vocabulary and literacy measures (word reading, reading comprehension) have also been included. Children's phonological errors and frequencies are reported. The errors are most likely due to interference evidence of L1-L2 interaction. Despite the small sample size, the emerging pattern of relations revealed connections between speech, language and literacy skills within and across languages. The results lend some support to the concept of universality and provide an indication for a unitary language system. The relationship of L2 PVC and PPC with L1 Phonological Awareness and the network of connections between L1 and L2 Phonological Awareness, language and literacy skills suggest a link between oral language and children's phonologies, phonological awareness and literacy. The findings of the present study draw attention to aspects of speech that should not be overlooked during instruction of L2 Greek, since they carry valuable information for the identification of bilingual children with speech and language difficulties. Individual production profiles are necessary to accompany individual educational plans in order to identify individual differences over and above the age at which L2 is acquired or the degree of exposure.

Keywords: phonology, bilingualism, oral language, literacy

Introduction

The relationship between speech and language skills attracts increasing research interest. Moreover, the interrelations of both speech and language skills with literacy, have relatively been under-researched. The investigation of connections between oral and written language informs normal development along a continuum and respective deficits with regard to the source of disordered speech, language and literacy.

Reading requires the development of mappings between speech sounds and letters and this depends on speech skills. Wider language skills are required to understand the meanings of words and sentences, to integrate these into text and to make inferences that go beyond printed words. Speech and language skills interfere with phonological awareness, a complex construct tapping underlying phonological representations. Phonological awareness is highly related to the acquisition of early literacy skills and individual differences in this critical ability may determine future learning outcomes. An adequate level of phonological awareness must be in place so that children can apply their knowledge of oral language structure to written language, whereas instruction should foster the development of oral language skills as a foundation for literacy development (Snowling & Stackhouse, 2006; Snowling & Hulme, 2012).

Different theories and studies have examined the complex relationship of early speech and language skills with literacy. The first one poses the direct influence of phonology, syntax, semantics and narration on the growth of reading (*comprehensive language approach, CLA*) while the second suggests an indirect relationship with phonological awareness as a mediator between them (*phonological sensitivity approach, PSA*) (Storch & Whitehurst 2002; Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg, & Poe, 2003; NICHD, 2005; Neuman & Dickinson, 2011, for reviews). However, more research is needed to inform theory and practice on their interrelations and contributions at different developmental stages, and their pathways to literacy.

Research on speech, language and literacy across languages presents a challenge for the investigation of underlying processes in the acquisition of language and literacy skills. It provides an insight into the understanding of complex relationships between oral and written language. Furthermore, it has implications for theory development on dual language coding and the potential to influence practice (Antoniou et al., 2015; Castro et al., 2014).

The transferability of oral and written language skills is well documented (e.g., Odlin, 1989; Catts & Kamhi, 2005). Can the study of English and Greek, two languages with differences in overall phonological complexity, be informative of the rate and route in the development of skills subsequently?

The present study has an exploratory purpose. It aims to identify patterns of errors in phonological production during L2 speech and investigate the relationships between speech, language and literacy skills among bilinguals who learn both English (L1) and Greek (L2).

Method

Sample

The participants were thirteen bilinguals (8 boys, 5 girls) from the Greek-American community in the United States, following grades 3 to 6. Their ages ranged from 8.5 to 12 years. The length of their instruction in Greek ranged from 2 to 6 years. All of them reported speaking both languages at home.

Measures

Children's retelling on the 'Frog where are you' story (Mayer, 1969) was used to extract a sample of words for analyzing patterns of speech errors, and derive percentages of vowels (PVC), consonants (PCC) and phonemes correct (PPC) following procedures described in Pascoe, Stackhouse, & Wells (2006). Using the Systematic Analysis of Language Transcripts (SALT) (Miller & Iglesias, 2003-4), mean length of utterance (MLU), fluency (words per minute, WPM), expressive vocabulary (number of different words, NDW) and narrative structure (NSS), were calculated in both languages.

Phonological awareness was measured using the Comprehensive Test of Phonological Processing (Wagner, Torgesen,, & Rashotte, 1999) in English and a parallel version was developed in Greek. The Elision, Word Blending, Word Segmenting, Nonword Blending and Nonword Segmenting subtests were used to test the ability to process sounds in both languages.

The Woodcock Picture Vocabulary test (Woodcock, 1991) was used in both languages to assess vocabulary knowledge in both L1 and L2. It includes fifty-eight picture items of increasing difficulty.

The sight word efficiency subtest of the Test of Word Reading Efficiency (TOWRE, Torgesen, Wagner, & Rashotte, 1999) was used to measure word reading fluency. The test contains 104 words of increasing difficulty. It provided a model for the Greek word reading test, developed with words and lemmas from the Hellenic National Corpus (HNC) (Hatzigeorgiu et al., 2000).

The Woodcock Passage Comprehension from the Woodcock Language Proficiency Battery-Revised (Woodcock, 1991) was administered to examine reading comprehension using a cloze task. The children read a sentence or short passage where individual words were omitted. To develop a parallel Greek version, the original items were translated, while adaptations were necessary for a few items.

Parents' Questionnaire: One of the parents was asked to answer forty three questions regarding: the child's abilities to speak, understand, read and write English and Greek; the language family members and friends use and the language the child is using when communicating with them; parents' competence in speaking, understanding, reading and writing English and Greek; habits at home (e.g., story telling in English and Greek); and parent expectations for achievement in both languages.

Results

The first aim of this small-scale study has been the investigation of children's L2 phonologies. The results are presented below separately for each child in chronological order along with examples and error frequencies.

TV (male, 12:00): depalatalization of palatal fricative [ʝ]→[γ] and voiceless velar stop (19.3%), gliding of mid vowels [o] and [e] (29.5%), distorted liquid: /l/ somewhat velarized (19.3%), distorted liquid: /ɾ/→[r] (13.6%), glotalization of /x/ (13.6%), consonant deletion (3.4%), cluster reduction [/t/ - /tr/] (1.1%).

DL (male, 11:39): depalatalization of voiceless velar stop (58.8%), cluster reduction (deletion) [/t/-/tr/, /t/-/nt/, /st/-/str/, /x/-/rx/] (17.6%), consonant deletion [/s/, /k/, /l/] (8.8%), syllable reduction in multisyllabic words (4-syllables) (2.9%) vowel reduction [/vatráça/-/vatraxia/] (2.9%), vowel change [/i/-/u/, /a/-/e/] (8.8%).

SE (male, 11:36): distorted liquid: /ɾ/→[r] (41.6%), distorted liquid: /l/ somewhat velarized (37.5%), depalatalization of palatal fricative [ʝ]→[γ] and voiceless velar stop (6.2%), gliding of mid vowel [e] (4.1%), vowel change [/a/-/o/, /e/-/a/] (10.4%).

DS (female, 11:36): distorted liquid: /ɾ/→[r] (30.7%), distorted liquid: /l/ somewhat velarized (25.6%), glotalization of /x/ (10.2%), gliding of mid vowel [o] (5.1%), depalatalization of palatal fricative [ʝ]→[γ] and voiceless velar stop (20.5%), stopping (2.5%), reduplication [γαγαβγιζε] (2.5%), cluster reduction [/g/-/gr/] (2.5%).

JG (male, 11:33): distorted liquid: /ɾ/→[r] (17.7%), cluster reduction (deletion and coalescence) [/t/-/tr/, /kx /-/ks/, /px/-/ps/, /sx/ -/sk/, / m/-/sm/, /t/-/tr/, /k/-/kr/] (25.8%), vowel change [/a/-/o/, /a/-/ u/] (6.4%), depalatalization of palatal fricative [ʝ]→[γ] and voiceless velar stop (35.4%), distorted liquid: /l/ somewhat velarized (11.2%), syllable reduction in multisyllabic words (4-syllables) (3.2%).

JK (female, 11:23): depalatalization of palatal fricative [ʝ]→[γ] and voiceless velar stop (40.8%) distorted liquid: /ɾ/→[r] (25.3%), distorted liquid: /l/ somewhat velarized (25.3%), cluster reduction [/pl/-/pn/] (1.4%), vowel change [/a/-/o/] (7%).

IM (male, 10:99): devoicing [/p/-/b/] (11.1%), stopping (3.7%), gliding of palatal fricative [ʝ] (11.1%), syllable reduction in multisyllabic words (4-syllables) (3.7%), depalatalization of voiceless velar stop (44.4%), distorted liquid: /ɾ/ (7.4%), distorted liquid:/l/ somewhat velarized (18.5%).

CK (female, 10:55): distorted liquid: /ɾ/→[r] (60%), distorted liquid: /l/ somewhat velarized (16%), vowel change [/a/-/o/, /a/-/e/] (16%), devoicing (8%).

NF (male: 10:03): glotalization of /x/(1.2%), gliding of palatal fricative [ʝ] as singleton (7.7%), distorted liquid: /ɾ/ distorted in clusters (sometimes →[ɹ]) (32.4%), distorted liquid: /l/ somewhat backed though not clearly velarized (41.5%), cluster reduction (coalescence & deletion) [/t/-/tr/, /rt/-/tr/] (7.7%), consonant deletion [/n/] (2.5%), syllable reduction in multisyllabic words (4-syllables) (2.5%), vowel change [/i/-/e/, /o/-/u/] (3.8%).

AM (female, 8:92): depalatalization of voiceless velar stop (25.8%), distorted liquid: /ɾ/→[r] (16.1%), cluster reduction (deletion and coalescence) [/p/-/mp/, /tz/-/ts/, /ts/-/tz/, /kθ/-/ks/, /θp/-/sp/ (25.8%), consonant change [/δ/-/z/, /θ/-/s/] (16.1%), syllable reduction in multisyllabic words (4-syllables) (6.4%), vowel change [/o/-/e/] (9.6%).

VH (male, 8:80): depalatalization of palatal fricative [ʝ]→[γ] and voiceless velar stop (23.5%), distorted liquid: /ɾ/→[r] (35.2%), vowel change [/o/-/e/] (35.2%), glotalization of /x/ (5.8%).

PM (male, 8:73): depalatalization of palatal fricative [ʝ]→[γ] and voiceless velar stop (28.5%), distorted liquid: /ɾ/→[r] (21.4%), distorted liquid:/l/ somewhat velarized (14.2%), /ɾ/ deletion [/aɣói/-/aɣóri/] (7.1%), cluster reduction (coalescence) [/θf/-/sf/] (21.4%), vowel change [/a/-/e/] (7.1%).

KK (female, 8:50): gliding of palatal fricative [ʝ] (10.7%), depalatalization of voiceless velars (stop & fricative) (17.8%), distorted liquid: /ɾ/ distorted in clusters (sometimes → [r], [ɹ]) (35.7%), cluster reduction (deletion) [/p/-/sp/, /spro/-/spa/] (21.4%), vowel reduction [/vɾáð/-/vɾáði/] (3.5%), voicing [/t/-/d/, /g/-/k/] (10.7%).

In order to explore the relationship between speech, language and literacy skills, inter-correlations and between language differences have been calculated. About half of the calculated correlations turned out to be significant which indicates great potential of links across the studied variables. The small sample size does not allow for a clear emerging pattern of relationships. We emphasize on those reflecting oral-written language connections within and across languages. Below are reported the most important findings, and statistics indicators of their significance.

L2 PVC, PCC and PPC was related to L2 vocabulary (r=.55, .57 and .62, $p \leq .05$ respectively). PPC was related to L2 NSS and NDW (r=.59 and .55 $p \leq .05$, respectively). L2 PVC and PPC were related to L1 Phonological Awareness (r=.55 and .57, $p \leq .05$, respectively). L2 Phonological Awareness was related to L2 Word Reading (r=.61, $p \leq .05$) and Reading Comprehension (r=.58, $p \leq .05$); L1 Word Reading (r=.74, $p \leq .01$) and Reading Comprehension (r=.63, $p \leq .05$), whereas L1 and L2 Phonological Awareness were highly related (r=.80, $p \leq .001$).

Other language and literacy measures showed a pattern of relations, within and across languages. More specifically, L2 NSS was related to L2 vocabulary (r=.77, $p \leq .01$) and L1 vocabulary (r=.55, $p \leq .05$). L2 NDW was related to L2 vocabulary (r=.72, $p \leq .01$), Phonological Awareness (r=.60, $p \leq .05$), Word Reading (r=.62, $p \leq .05$) and Comprehension (r=.73, $p \leq .01$); L1 vocabulary (r=.56, $p \leq .05$), Phonological Awareness (r=.64, $p \leq .05$), Word Reading (r=.64, $p \leq .05$) and Comprehension (r=.73, $p \leq .01$). L2 WPM was related to L2 vocabulary (r=.74, $p \leq .01$), Word Reading (r=.66, $p \leq .05$) and Comprehension (r=.81, $p \leq .001$); L1 Vocabulary (r=.55, $p \leq .05$), Word Reading (r=.58, $p \leq .05$) and Comprehension (r=.61, $p \leq .05$). Similarly, L1 MLU was related to L1 vocabulary (r=.69, $p \leq .01$), Word Reading (r=.69, $p \leq .01$) and Comprehension (r=.61, $p \leq .05$); L2 Word Reading (r=.59, $p \leq .05$) and Comprehension (r=.59, $p \leq .05$). L1 NDW was related L1 vocabulary (r=.57, $p \leq .05$), whereas, L1 NSS was related to L1 vocabulary (r=.62, $p \leq .05$), Reading Comprehension (r=.68, $p \leq .01$) and L2 Word Reading (r=.63, $p \leq .05$). L1 WPM was related to L1 vocabulary (r=.69, $p \leq .01$), Word Reading (r=.80, $p \leq .001$) and Comprehension (r=.79, $p \leq .001$); L2 Word Reading (r=.78, $p \leq .01$) and Comprehension (r=.73, $p \leq .01$).

Between language comparisons indicated that there was no significant difference in L1 and L2 phonological awareness (t=0.12, df=12, p=ns) and NDW (t=1.40, df=12, p=ns). All other measures showed significant differences NSS (t=3.84, df=12, $p \leq .01$), MLU (t=6.10, df=12, $p \leq .001$), WPM (t=7.82, df=12, $p \leq .001$), Word Reading (t=4.98, df=12, $p \leq .001$), Reading Comprehension (t=4.32, df=12, $p \leq .001$), to the expected L1>L2 direction.

Discussion

Despite the sample size, age or gender differences, there is some consistency of errors across children in the following: cluster reduction, syllable reduction in multisyllabic words, vowel reduction, depalatalization of palatal fricatives and voiceless velar stop, distortion of liquids. The errors are few rather than predominant and most likely this is due to interference rather than non-acquisition, evidence of L1-L2 interaction. Such errors are indicators of developmental processes and reveal the succession of phases of learning to master new structures (Beach, Burnham, & Kitamura, 2001). If interdependence in the acquisition of two languages is a possibility, transfer and delays (or accelerations) are characteristics of a complex process. However, the exact nature and degree of this interaction will remain elusive for some time (Babatsouli & Ingram, 2015).

Age may also be related to language interaction in that gradually L1 and L2 phonetic subsystems can be separated without influencing each other (Fledge, Schirru, & MacKay, 2003). In the present study age was related to PPC (r=-.55, $p \leq .05$) and marginally non-significantly related to PCC (r=-.53, ns). PVC was related to the language the child is using with other adults at home and book sharing experiences in L2 (not related to homework) (r=.73 and .71, $p \leq .01$, respectively) [data not presented in the results section].

The findings of the present study draw attention to aspects of speech that should not be overlooked during instruction of L2 Greek, since they carry valuable information for the identification of bilingual children with speech and language difficulties. Individual production profiles are necessary to accompany individual educational plans. There may be individual differences over and above age at which L2 is acquired or the degree of exposure. Therefore, it is important to determine if the child is acquiring language at the typical rate, identify the pattern of acquisition the child is following and the kind of phonological learning a given child is using (Ingram, 2001).

Some level of cross-linguistic transfer of language and literacy skills was evident across L1 and L2 in line with research which has implemented oral language assessment through narrative procedures (Miller et al. 2006). Moreover, the inter-relations between speech, language and literacy skills across languages lend support to the concept of universality and provide an indication for a unitary language system (Antoniou et al., 2015; Babatsouli & Ingram, 2015 for a relevant discussion).

The relationship of L2 PVC and PPC with L1 Phonological Awareness and the network of connections between L1 and L2 Phonological Awareness, language and literacy skills suggests a link between oral language and children's phonologies, phonological awareness and literacy in support of the phonological sensitivity approach (PSA) (Dickinson et al. 2003). However, the Comprehensive Language Approach (CLA) cannot be precluded.

References

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition*, *18*(4), 683-695.

Babatsouli, E., & Ingram, D. (2015). What bilingualism tells us about phonological acquisition. In R. H. Bahr & E. R. Silliman (eds.), *Routledge handbook of communication disorders* (pp. 173-182). Routledge: Taylor & Francis.

Beach, F., Burnham, D., & Kitamura, C. (2001). Bilingualism and the relationship between perception and production. *The International Journal of Bilingualism, 5*(2), 221-235.

Castro, D., Paez, M., Dickinson, D., & Frede, E. (2011). Promoting language and literacy in young dual language learners: Research, practice, and policy. *Child Development Perspectives*, *5*(1), 15-21.

Catts, H. W. & Kamhi, A. G. (eds.) (2015). *The connections between language and reading disabilities.* Mahwah, NJ: Erlbaum.

Dickinson, D., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E., & Poe, M. (2003). The comprehensive language approach to early literacy. *Journal of Educational Psychology*, *95*(3), 465-481.

Fledge, J., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*, 467-491.

Ingram, D. (2001). The assessment of phonological disorders in children: The state of the art. In M. Crary (ed.) *Phonological intervention, concepts and procedures* (pp. 1-12). San Diego, CA: College-Hill Press.

Mayer, M. (1969). *Frog, where are you*? New York, NY: Puffin Books.

Miller, J. F., & Iglesias, A. (2003-4). *Systematic analysis of English and Spanish Language transcripts*. Language Analysis laboratory, Waisman Center, University of Wisconsin-Madison, Madison, WI.

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice*, *21*(1), 30-43.

Neuman, S. & Dickinson, D. (2011). *Handbook of early literacy research*. New York, NY: Guilford.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning.* Cambridge, UK: Cambridge University Press.

Pascoe, M., Stackhouse J., & Wells, B. (2006). *Persisting speech difficulties in children*. Chichester, UK: Willey.

Snowling, M., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. *International Journal of Language and Communication Disorders*, *47*(1), 27-34.

Snowling, M., & Stackhouse, J. (2006). *Dyslexia, Speech and Language*. Chichester, UK: Whurr.

Storch, S., & Whitehurst, G. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*(6), 934-47.

Torgesen, H. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: PRO-ED.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: PRO ED.

Woodcock, R. (1991). *Woodcock language proficiency battery-revised*. Itasca, IL: Riverside Publishing.

NICHD Early Child Care Research Network (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology*, *41*(2), 428-442.

A developmental study of self-repairs in Spanish normal-speaking children and comparison with a case study of specific language impairment

Mª Isabel Navarro-Ruiz[1], Lucrecia Rallo Fabra[1 2]
isabel.navarro@uib.es, lucrecia.rallo@uib.es

[1]Departament de Filologia Espanyola, Moderna i Llatina, Universitat de les Illes Balears
[2]Institut de Recerca i Innovació Educativa-IRIE, Universitat de les Illes Balears

Abstract. Disfluencies are a common trait of every-day speech, with self-repairs being the most common examples of breaks of the speech flow (MacLurg, 2014). According to Levy (1999), the complexity of self-repairs cannot just be explained in terms of the monitor but also in terms of the metalinguistic function. The aim of the present study is two-fold. On the one hand, we intend to provide developmental data on self-corrections. This information will facilitate assigning a developmental age to children acquiring their first language and will be used subsequently for screening purposes among specific language impairment (SLI) populations. The data from the normal-speaking children (control group) are obtained through a cross-sectional study: forty children, whose ages range from 22 months to 10 years, were audio- and video-recorded during a 45-minute semi-structured interview and subsequently transcribed using the SALT system (Systematic Analysis of Language Transcript). The language samples were coded by the first author; two additional transcribers categorized 10% of the language sample for reliability purposes. Cohen's kappa coefficient shows a high inter-rate agreement (k = 0. 87 and k = 0.90) indicating that the coding of language categories is reliable. SLI is a language disorder characterized by an alteration of normal language development of receptive and productive skills. It may affect one or more linguistic levels. The SLI longitudinal data are obtained from a child with SLI who was recorded at 10 time intervals during the interval from 3;4 to 6;0 years. The child's linguistic profile is compared with the data from the control group. The results of the control group show that the first self-repairs affected the phonological level and emerged around 1;10 years. The maximum frequency of self-repairs among the TD children exhibits a U-shape pattern of reorganization (Bowerman, 1982). Pragmatic self-repairs are the most commonly used by TD children and they emerge around 1;10 years, just like syntactic self-repairs. As for the child with SLI, the first self-repairs emerge at 4;3 years and they affect morphology and syntax. The frequency of self-repairs is considerably lower in the utterances of the child with SLI relative to the control group and it increases as a function of age. It is hypothesized that an improvement of linguistic competence goes hand in hand with metalinguistic abilities, which leads the child with SLI to self-correct his/her deficiencies. This finding is in line with Berthood's proposal (2000) of language development and metalinguistic function.

Keywords: specific language impairment, self-repairs, metalinguistic abilities

Introduction

The term maze is used as a broad category of linguistic performances, which includes different linguistic performances related to fluency, such as communication breakdowns, speech disruptions, speech repairs, and disfluencies. (MacLurg, 2014). These productions are associated with different language levels, different levels of processing and also with the metalinguistic function, that is, a speaker's ability to revise the intended message after giving the message more thought. This paper focuses on fluency mazes, more specifically on self-repairs.

Self-repairs are related to monitoring processes and reflect a speaker's linguistic competence to fix his own errors (Levy, 1999; Postma, Kolk, & Povel, 1990). Levy (1999) claims that self-repairs are the result of a complex process, consisting of the speaker's ability to monitor her own speech, identify errors and fix them. This process can only be achieved if the speaker has acquired sufficient linguistic competence and metalinguistic abilities that allow him to revise the message. Context and degree of overall attention determine the success of this process. These variables facilitate the acquisition of different language uses as a function of context and give the child feedback, enriching his previous knowledge, extending and modifying it (Clark, 2014). This process allows the child to adapt and adjust socially to discourse (Laakso, 2010). These contexts boost the appearance of self-repairs. Children aged 18 to 24 months can monitor speech while they are delivering it (Clark, 1978). They may use repairs to interact with the recipient, that is, they do not just repair their own productions, but also the recipients' (Forrester & Cherington, 2009).

Children and adults use similar self-repair strategies. However, some differences can be found depending on the child's stage of development. De Ruiter (2013) reports that children stall their productions if they detect an error but, unlike adults, they do not stop if the message they deliver is not appropriate in a given situation. Two different approaches may account for this behaviour. The first one claims that children distinguish both strategies as a function of context. The other approach associates this behaviour to different processing levels. Apparently, the child's ability to detect an inappropriate production requires a more complex level of processing than detecting an error (De Ruiter, 2013).

Specific language impairment (SLI) is a language disorder characterized by an alteration of normal language development of receptive and productive skills. To date, the causes of this disorder have not been clearly established. Although in some cases the disorder has a genetic basis, it often involves an interaction of genetic and environmental factors (see Bishop, 2006 for a review). Children with SLI struggle to communicate and show difficulties understanding language if this requires complex vocabulary or syntax (Bishop, 2000). These difficulties affect various language levels, although individual variability in performance is often the rule. An child with SLI may also exhibit deficits in social interaction thus affecting their relationships with peers. From the pragmatics perspective, Bishop and Adams (1989) reported difficulties in sentence formulation, semantic selection, and use of stereotyped utterances with abnormal prosody.

The level of linguistic competence in children with SLI, their knowledge of the language and their "reflection and building abilities" are below the level of their normally developing peers (Navarro, 2001). Their processing and abstraction levels are comparable to children at a lower stage of development (Navarro, 1997). They also exhibit differential metalinguistic skills, such as a general delay in the emergence of self-repairs, which at onset affect the more preserved language levels. (Navarro-Ruiz & Rallo-Fabra, 2001).

Method

Given the methodological limitation of testing a large group of normal-speaking children, hereon typically developing (TD) children over a period of nine years, this study combined a cross-sectional design with a longitudinal design. The cross-sectional design allowed us to obtain data of typical developing children linguistic profiles at different stages of development, which served as baseline to compare the linguistic profile of the child SLI with his age-matching controls.

Participants

A total number of 40 typical developing children (TD) (20 male, 20 female) aged 2 to 10 years participated in the study. The choice of the age span was based on the findings of prior work, which document that the early emergence of self-repairs occurs around two years.

All the children lived in Catalonia, a bilingual region in North-Eastern Spain, and came from households of a mid-sociocultural level. Only the participants that were exposed to both Spanish and Catalan at home and school were selected.

Procedure

The children interacted with the experimenter in a playroom, which contained a wide selection of toys and story-books. Children were not prompted to make use of their metalinguistic abilities through specific questions posed by the experimenter for two reasons: (i) to preserve the ecological validity of the study (ii) to control the frequency of use of these abilities in a child-adult interaction. It was also deemed appropriate to control the discursive style of the interviewer. Therefore, the experimenter's role was that of an external participant in a natural playing situation. This role was preserved throughout all the interviews conducted with the TD children and the child with SLI.

Each interview lasted about 45 minutes approximately. Child and adult interacted playing and talking about topics of general interest to all children. The experimenter had an informal meeting with the children prior to the audio/video-recording sessions. A total of 20 time intervals were established, starting at 1;10 years. The time interval between each developmental time was 4 months for the 1;10 to 4;4 year children. Children up to 4 years are known to make significantly faster progress in terms of linguistic competence than older children. Older children (over 4;4 years) were recorded at 6-month intervals. The child with SLI was recorded at 11 time intervals: 3;4, 3;7, 3;10, 4;2, 4;6, 4;10, 5;2, 5;6, 5;9, 6;2, 6;10 years.

Transcription of language samples

The language elicitations were audio/video-recorded and transcribed using the SALT system (Systematic Anaylsis of Language Transcript) (Miller & Chapman, 1985). Utterances of two words or longer were taken as the unit of analysis. Following Crystal, Fletcher, and Garman (1976), falls in intonation were considered as utterance boundaries in case of ambiguity. The first 5 minutes of each sample were discarded. At the beginning of each recording session, children tend to be nervous and/or easily distracted. After the first five minutes elapsed, the next 100 utterances were considered for the analysis. The corpus was coded by the first author plus two additional transcribers, who coded 10% of the language sample to ensure reliability. Cohen's kappa coefficient showed a high inter-rate agreement (k = 0. 87 and k = 0.90) indicating that the coding of the language categories was reliable. The taxonomy of language categories used for the self-repairs was based on the five language levels: phonology, morphology, lexis, syntax, and pragmatics.

Results

Typical-developing children

Overall, in typical-developing children, self-repairs emerge as early as 1;10 years (see Table 1). Phonological self-repairs are the first to emerge and disappear around 4 years. The first language sample (1) corresponds to a phonological self-repair by a 22 month old. The target Spanish word *pequeñas* /peˈkeɲas/ is first replaced with [peˈtenas] showing a fronting of both the velar stop and the palatal nasal. Immediately after, the child self-repairs her first attempt and elicits a second word much closer to the target /peˈkjeɲas/:

(1)

{the child points to a pair of small trainers}
Child 1: son petenas pequieñas
Child 1: they are small, tiny

The next sample (2) shows an instance of pragmatic self-repair by another TD 22-month-old child. The child is describing a picture with various stars. She mentions *star*, and later realizes she has not talked about the *star* yet. Therefore, she substitutes the definite article for the indefinite article:

(2)

{22 month girl, describes a picture with various stars}
Child 1: l´estel (:) una estel
Child 1: the star (:) one star

Morphological self-repairs appear a bit later, around 2;2 years, and they reach the maximum frequency in the 3-4 years age span. Overall, morphological self-repairs are the most frequent followed by syntactic and lexical self-repairs. Sample (3) to (6) show instances of morphological and lexical self-repairs:

(3)

Niña: Si, pero yo tengo (un, un) una muñeca
Child: Yes, but I have (a, a) a doll (repairs morphology: genre)

(4)

Niño: (:2) (un) (:2) oto cochesito?
Child: (:2) (a) (:2) another small car? (repairs lexis)

(5)

Niño: Aquest es el cochecito (:) era
Child: This is the small car (:) was (repairs morphology: tense marking)

(6)

Niña: Una (bic una a, a) una moto*
Child: A (bik one, a, a) a moto* (repairs lexis)

From 4 years onwards, there is an increase in syntactic and morphological self-repairs. This is partially in line with an earlier study reporting self-repairs by Finnish-speaking children aged 4 (Salonen & Laakso, 2009), that found that syntactic and lexical self-repairs were the most frequent among four-year-olds.

(7)

Niño 8;6 años
Child 8;6 years
Niño: Si, ahora (van, va) siempre va empatando a cero
Child: Yes, now (they are, he is) he is always tying on cero (morphological self-repair)

As for the frequency of maximum use, phonological self-repairs reach the maximum frequency at 1;10 years. Pragmatic self-repairs reach the highest frequency of use at 3 years, followed by morphological self-repairs at 3;4 years, and lexical self-repairs at 7;6-8 years. Except for lexical self-repairs, all language levels exhibit a U-shape pattern of usage at different developmental stages (see Figure 1). Bowerman (1982) suggests that this pattern responds to a 'developmental process from onset to expertise in an ability', and it involves successive reorganizations when a category has its maximum frequency at more than one age.

Table 1: Age of emergence and frequency of maximum use of self-repairs in the typical developing children (TD) and the child with SLI as a function of the language level affected

Language level	Emergence TD	Emergence SLI	Maximum use TD	Maximum use SLI
Phonology	1;10	x	1;10	X
Morphology	2;2	4;2	3;4	4;6
Syntax	2;2	4;6	6;6	4;6
Lexis	2;2	5;2	7;6	5;2
Pragmatics	2;2	4;10	3	2;2

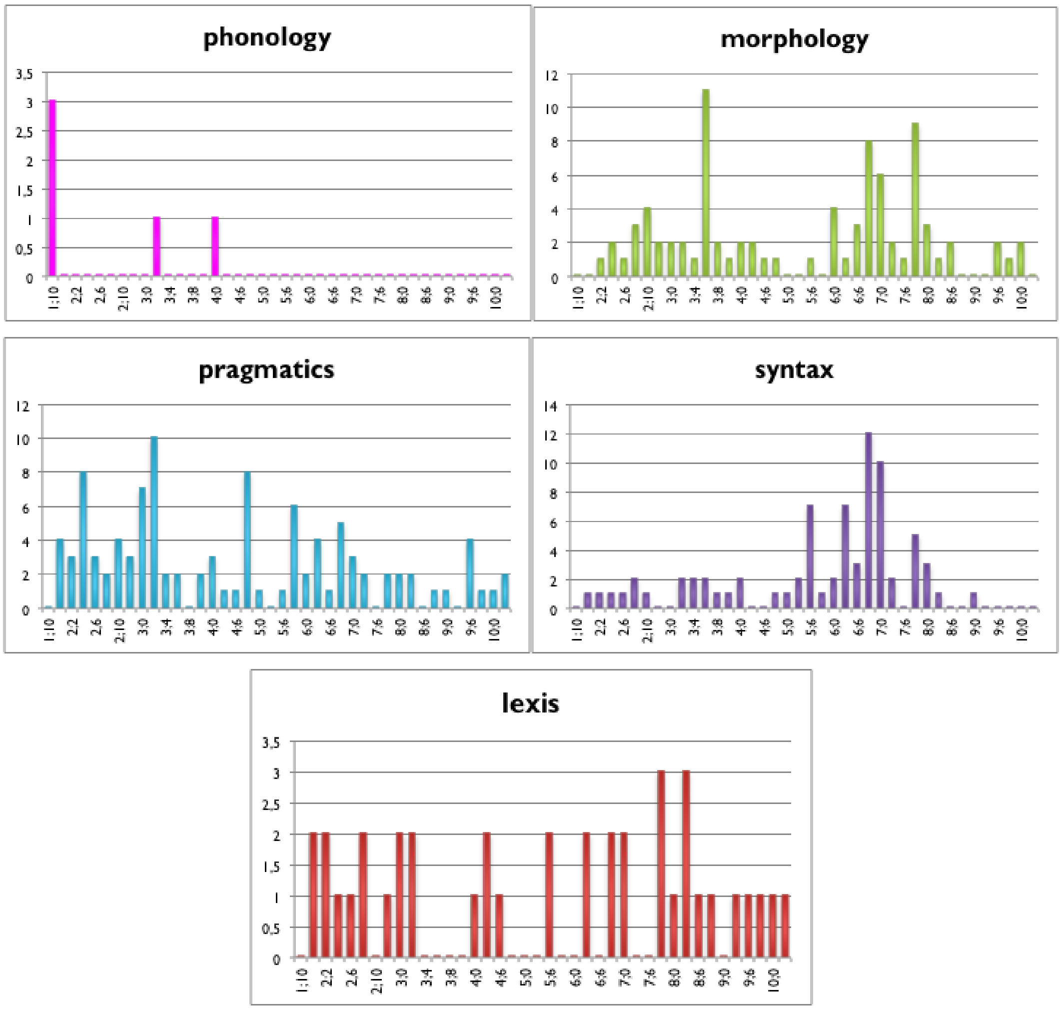


Figure 1. Frequency of maximum use of self-repairs by TD children as a function of language level

The child with SLI

The child with SLI does not exhibit phonological self-repairs. The first self-repairs to emerge affect morphology at 4;2 years. Later, the child starts to repair syntax and pragmatics at 4;6 and 4,10 years, respectively. Lexical self-repairs are the last to emerge at 5;2 years. The absence of phonological self-repairs could be traced down to the fact that the child with SLI was first recorded at the age of 3;4 years. We speculate that the child might have repaired phonology prior to this age but, unfortunately, these data are not available.

As for the frequency of maximum use, syntactic and morphologycal self-repairs appear at 4; 6 years. At 5;2 years, lexical and pragmatic self-repairs reach the frequency of máximum use.

The frequency of self-repairs is considerably lower in the utterances of the child with SLI relative to his TD peers. Early emergence of self-repairs in the child with SLI involves the better-preserved language levels. Self-repairs involving the more affected language levels emerge later. Reorganization processes (U-shape) are absent in the child with SLI. In line with Berthoud-Papandropoulou (2000), we speculate that the emergence of metalinguistic abilities is closely related to the child gaining knowledge of his own language. This strengthens his ability to think about the language and self-repair his own errors.

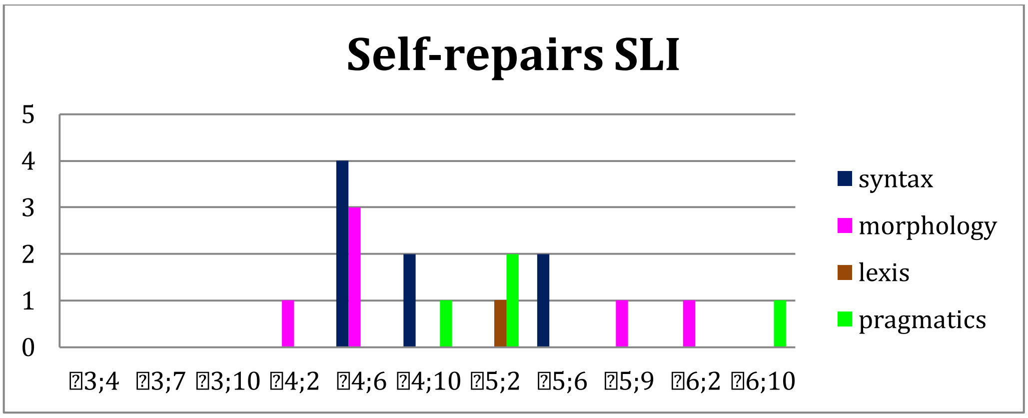


Figure 2. Frequency of maximum use of self-repairs by the child with SLI

A comparison of the self-repairs by the TD children and the child with SLI reveals that the child's behaviour corresponds to a child at a lower stage of development. Morphological self-repairs emerge at 2;2 years in TD children, whereas in the child with SLI do not emerge until 4;2 years. TD children start to repair syntax and lexis before they turn 2 years of age, specifically at 1;10 years. In contrast, the child with SLI does not start to repair syntax until two years later (see example below). Similarly, lexical and pragmatic self-repairs by the child with SLI are late in the stage of development. They do not emerge until past the 5 years of age, a considerable delay with respect to TD children, who use pragmatic repairs before 2 years of age.

(8)

The child with SLI 5;2 years
Niño: El caba no la cebra*
Child: The hor not the zebra (repairs lexis)*

(9)

The child with SLI 5;6 years
Niño: Igual es que >
Child: May be it's that
Niño: Igual>
Child: Maybe>
Niño: Igual el que viste.
Child: Same you saw (repairs syntax)

Frequencies of maximum use also exhibit a delay (see Figure 2). The child with SLI does not use phonological self-repairs. Among TD children, pragmatic self-repairs reach the maximum frequency at 3 years, whereas the maximum frequencies for the child with SLI are much later, at 5;2 years. The frequency of morphological self-repairs reaches the highest values at 3;4 years among TD children, but for the child with SLI a delay of more than one year is found (4;6 years). Surprisingly, most syntactic and lexical self-repairs were found earlier in the child's utterances, at 4;6 and 5;2 respectively, than in the TD children's utterances, over 6-8 years span.

Discussion and conclusions

The present study has provided cross-sectional data of self-repairs by 40 Spanish typical developing children. These can be used as baseline for screening purposes in cases of language-impairment. TD children have shown U-shape reorganizations, which vary chronologically depending on the language level affected.

There is no agreement in the literature as far as the metalinguistic nature of self-repairs is concerned. Some authors (Gombert, 1990; Kamiloff, 1986) consider that self-repairs that emerge at the initial stages of development cannot be labelled as metalinguistic, while others consider that they are (Clark, 1978). Following Levy (1999), we suggest that the relationship between self-repairs and metalinguistic abilities must take into account three issues*: (i)* young children base their self-repairs on overall mental representations of linguistic knowledge, (*ii)* self-repairs are a consequence of the monitoring system (Levelt, 1983) and *(iii)* the monitor receives input from communicative intentions, not pure linguistic acts. Given these circumstances, we assume that children at an early age can perform these meta-operations because these levels are more accessible to young children whose communicative intentions surpass their linguistic structures. Self-repairs reflect their own system along with their age-related limitations.

Children with SLI may exhibit less self-repairs than their TD peers. Nevertheless, at certain stages of development, children with SLI may outperform their TD peers. This might look as if their linguistic abilities are not improving, but it actually responds to the development of their own system. In children with SLI, better knowledge of their linguistic system triggers the use of self-repairs at later stages of development, in which recursive processes would be the rule. For instance, in the case of verb tense marking in Spanishs, children learn the rules and apply them on a regular basis, no matter if the verbs are regular or not, which implies making errors. At some point they learn that certain verbs do not follow the general rule. This allows the child to build up new, richer information that accounts for these exceptions to the general rule.

References

Berthoud-Papandropoulou, I. (2000). Lo sviluppo della competenza metalinguistica nel bambino. *Metis*, *1,* 7-25.

Bishop, D. V. M. (2000). Pragmatic language impairment: A correlate of SLI, a distinct group or part of the acoustic continuum? In D. V. M. Bishop & L. B. Leonard (eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome* (pp. 99-113). Hove: Psychology Press.

Bishop, D. V. M. (2006). What causes Specific Language Impairment in children? *Current Directions in Psychological Science*, *15*(5), 217-221.

Bishop, D. V. M., & Adams, C. (1989). Conversational characteristics of children with semantic-pragmatic disorder. II. What features lead to a judgement of inappropriacy? *British Journal of Disorders of Communication*, *24*, 241-263.

Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In E. Wanner & L. R. Gleitman (eds.), *Language acquisition: The state of the art* (pp. 319-346). New York: Cambridge University Press.

Clark, E. V. (1978). Awareness of language: Some evidence from what children say and do. In A. Sinclair, R. J. Jarvella & W. J. M. Levelt (eds.) *The child's conception of language*, (pp. 17-44). Berlin: Springer Verlag.

Clark, E. V. (2014). Pragmatics in acquisition. *Journal of Child Language, 41*(S1), 105-116.

Crystal, D., Fletcher, P. & Garman, M. (1983). *Análisis gramatical de los trastornos del lenguaje*. Barcelona: Editorial Médica y Técnica,

De Ruiter, L. (2013). Self-repairs in German children's peer interaction - initial explorations. In R. Eklund (ed.), *Proceedings of disfluency in spontaneous speech, DiSS 2013TMH-QPSR, 54*(1), 29-32.

Forrester, M. A., Cherington, S. M. (2009). The development of other-related conversational skills: A case study of conversational repair during the early years. *First Language, 29,* 166-191.

Laakso, M. (2010). Children's emerging and developing self-repair practices. In H. Gardner & M. Forrester (eds.) *Analysing interactions in childhood: Insights from conversation analysis* (pp. 74-99). Oxford, UK: Wiley-Blackwell.

Levy, Y. (1999). Early metalinguistic competence: Speech monitoring and repair behavior. *Developmental Psychology*, *35*(3), 822-834.

Miller, G., & Chapman, J. (1985). *Systematic Anaylsis of Language Transcript (SALT).* Madison: University of Wisconsin.

MacLurg, A. (2014). *Mapping mazes: Developing a taxonomy to investigate mazes in children's stories*. MA Thesis. Edmonton: University of Alberta.

Navarro-Ruiz, M. I., & Rallo-Fabra, L. (2001). Characteristics of mazes produced by SLI children. *Clinical Linguistics, and Phonetics, 15* (1-2), 63-66.

Postma, A., Kolk, H., & Povel, D. J. (1990). On the relation among speech errors, disfluencies, and self-repairs. *Language and Speech*, *33*(1), 19-29.

Salonen, T., & Laakso, M. (2009). Self-repair of speech by four-year-old Finnish children. *Journal of Child Language*, *36*(4), 855-882.

Vowel duration contrast in three long-short pairs by Hungarian 5-, 6-, and 7-year-olds

Tilda Neuberger[1], Judit Bóna[2], Alexandra Markó[2], Ágnes Jordanidisz[3], Ferenc Bunta[4]
neuberger.tilda@nytud.mta.hu , bona.judit@btk.elte.hu, marko.alexandra@btk.elte.hu, ajordanidisz@gmail.com, fbunta@uh.edu

[1]Research Institute for Linguistics, Hungarian Academy of Sciences, [2]Eötvös Loránd University, [3]Association for Educational Needs/NILD, [4]University of Houston

Abstract. *Purpose*. This study focuses on phonemic vowel quantity differentiation in 3 pairs of short versus long vowels (/i, iː, o, oː, u, uː/) by 5-, 6-, and 7-year-old monolingual Hungarian-speaking children. We hypothesized that there would be vowel quantity and vowel quality differences and also a trend toward greater vowel differentiation as children aged. *Method*. Participants included 3 groups of monolingual Hungarian-speaking children: 5-year-olds (n=6), 6-year-olds (n=7), and 7-year-olds (n=14) recruited from Hungarian public schools. The participants had typical cognitive skills, speech, language, and hearing within normal limits per parent and teacher report. Audio recordings were collected via conversational samples as the children interacted with an experimenter, discussing favorite pastimes, everyday lives, or favorite stories. Vowels were analyzed using PRAAT 5.0 (Boersma & Weenink, 2015) except final vowels and distorted productions or substitutions. ***Results***. There were statistically significant effects for vowel quality (F (2, 42) = 10.12 at p = 0.000, partial η^2 = 0.33) and vowel quantity (F (1, 21) = 67.49 at p = 0.000, partial η^2 = 0.76), but no age effect or age by quantity interaction were found. Our results provided support for our predictions regarding differences based on vowel quantity and quality; however, age and age by quantity interaction effects were not found. *Conclusions.* Overall, our participants did distinguish vowels based on duration and vowel quality, but those distinctions did not depend on age, possibly indicating that differentiation may occur as early as 5 years of age. Further research is needed to verify our findings using more participants and longitudinal data to track the development of the phonemic contrast between short and long vowels in Hungarian.

Keywords: Hungarian, child phonology, vowel quantity and duration, vowel quality, short/long vowels, conversational speech

Introduction

Phonemic use of vowel duration differs across languages, ranging from no functional vowel quantity discrimination (as in Spanish, cf. Malmberg, 1971), to duration used only as an acoustic cue but not a phonological feature (as in English, cf. Kassai, 1979), to vowel duration used as both an acoustic cue and a phonological feature (such as Hungarian or Estonian, cf. Lehiste, 1965). Languages that do use vowel quantity as a phonological feature (i.e. distinguish vowel duration on a phonemic level) have two distinct vowel duration categories, such as long versus short vowels as in Hungarian or even three phonemic levels based on vowel duration. as does Estonian (Lehiste, 1965). Note that in the relevant literature, both vowel quantity and vowel length are used to refer to phonemic differentiation of vowels based on duration. We opt for the former, because length can also imply measurement of distance, so we chose our course of action in the interest of precision.

Vowel quantity is a distinctive phonological feature in Hungarian, having 7 pairs of short versus long vowel phonemes in the language: /i – iː, y – yː, ø – øː, ɛ – eː, u – uː, o – oː, ɒ – aː/ (Nádasdy & Siptár, 2001). Nonetheless, from a purely phonetic point of view, only 5 of the phonemically short versus long pairs differ primarily on duration: /i – iː, y – yː, ø – øː, u – uː, o – oː/ (Gósy, 2004), the other two pairs (/ɛ – eː, ɒ – aː/) display qualitative as well as durational differences (see Figure 1 below based on Szende, 1994).

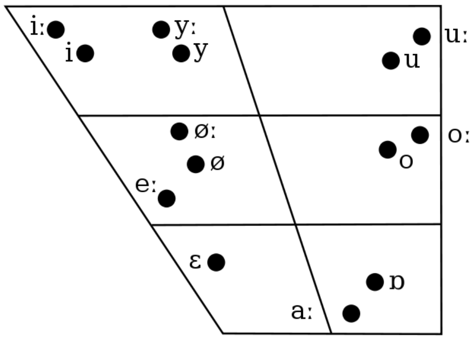


Figure 1. The Hungarian vowel system (Szende, 1994)

It must be noted that the phonemic feature of vowel quantity does not always manifest itself consistently in a physically measurable fashion; for example, the duration of phonemically short versus phonemically long vowels can have an overlap ranging from 28% (Thai) to as high as 90% (German), depending on the language (Lehiste, 1970). Factors such as vowel quality, tempo, position of the vowel, word length, and others affect vowel duration in Hungarian (Gósy & Beke, 2010). However, in Hungarian, despite variations in duration due to factors noted above, the durational differences distinguishing short from long vowels are maintained to the degree that the categorical distinctions persist. For example, irrespective of stress patterns, Hungarian vowels maintain the short versus long contrast in a measurable fashion pertaining to their duration (Gósy & Beke, 2010). Moreover, short and long vowels maintain their durational differences in isolated words read out loud as well as in spontaneous, conversational speech even though some overlap does exist between the vowel quantity categories (Bóna, 2012; Gósy & Beke, 2010). The drive to maintain the durational differences between short and long vowels in Hungarian is so strong that evidence can be found even in the speech patterns of elderly adults (over 70 years of age) where vowel quality differences become reduced, yet durational differences are maintained, albeit not as notably as in the case of younger adults (Bóna, 2012).

In terms of phonological acquisition, vowel duration is one of the last contrastive vowel features acquired by monolingual Hungarian-speaking children, and one to undergo gradual development starting with certain pairs (e.g., /i/ - /i:/) around 4 years of age and in limited contexts (Zajdó & Powell, 2008). However, differences in vowel duration are still not completely mastered in an adult-like fashion in all 7 short-long vowel pairs by 6 years of age (Bóna & Imre, 2010; Deme, 2012). Consequently, our study investigates vowel quantity differentiation in 3 pairs of vowels that only differ in duration (/i, i:, o, o:, u, u:/) between the ages of 5;0 and 7;11 to investigate the changes, if any, occurring in this age range. We focus on the vowel pairs noted above, because these contrasts differ primarily on duration and the pairs are qualitatively very similar, so this allows for investigating the discrimination of the durational contrast avoiding the confounding factor of possible qualitative differences.

Zajdó (2002) found that children at age 3 were able to produce adult-like short and long vowels with 90% mastery level while imitating their caregivers' productions of two-syllable items, such as *pipi* /*pipi*/ (= small chicken) and /*pi:pi:/* (non-word), using puppets that bore the target names. Children displayed more accurate and earlier pronunciations of unrounded vowels relative to rounded ones. The measurement involved perceptual judgements by an adult native Hungarian speaker. In a follow-up study, Zajdó (2015) found that Hungarian-speaking children between 2 and 4 years of age were able to modify vowel duration based on the caregiver's model. As in the previous study, the tokens were CV(:)CV(:) labels, such as /pipi/ versus /pi:pi:/, given to puppets, and the participants' caregivers provided the model in continuous speech. The caregivers were asked to elicit the names of the puppets during their interactions with the children.

In the absence of an adult model that young children could imitate, the separation of short versus long vowels based on duration appears to be less certain. Bóna and Imre (2010) found that in conversational speech samples, only children between 5 and 6 years of age began to separate short and long vowels in a systematic fashion in their productions. Furthermore, even by 6 years of age, not all vowel pairs displayed clear differentiation based on duration in the samples of Hungarian-speaking children. Specifically, only /o/ versus /o:/ and /u/ versus /u:/ showed statistically significant differences based on vowel duration. These results were later replicated by Deme (2012) who investigated the short and long vowels of 6- to 7-year-olds, demonstrating statistically significant vowel duration differences in the same pairs only (/o/ - /o:/ and /u/ - /u:/).

The lack of unequivocal differentiation between short and long vowels may also have perceptual underpinnings. Gósy (2006) found that Hungarian-speaking children could only differentiate vowel duration accurately in perception 28% of the time at 5 years of age, 65-70% at 7 years of age, and 75-80% of the time between the ages of 8 and 9 years. Based on these results, there appears to be a considerable increase in accuracy to perceive short versus long vowels between the ages of 5 and 7 years. Consequently, studying the duration of vowels produced by Hungarian-speaking children in this age range may reveal important insights into the development of vowel duration differentiation, specifically, and speech development, more generally. The main research question is how phonemic vowel quantity discrimination manifests itself in the conversational speech samples of 5- to 7-year-old monolingual Hungarian-speaking children. A related question is how and if vowel quality affects vowel durations and their differentiation. Finally, would there be an interaction between the different age groups and durational differences? In other words, does the differentiation of vowels based on duration depend on age? Based on existing research on and the research questions above, we posit the following hypotheses:

1) We expect to find differences in vowel duration both at ages 5 and 7.
2) As vowel quantity contrasts develop, we predict an age effect in that durational differences will be better expressed at age 7 than at age 5.
3) We expect there to be an effect of vowel quality on the duration of the vowels and also predict differential quality effects on the different short and long vowel pairs.

Method

Participants

The present study adheres to the ethical guidelines provided by the Hungarian review board that oversees the ethical treatment of human subjects in research. Written parental consent and child assent were obtained prior to the execution of the study from each of the children and their parents or legal guardians. There were 3 groups of monolingual Hungarian-speaking children: 5-year-olds (n=6), 6-year-olds (n=7), and 7-year-olds (n=14) recruited from Hungarian public schools in the Budapest metropolitan area. The participants had typical cognitive skills, speech, language, and hearing within normal limits per parent and teacher report. The socio-economic status of the participants was not controlled; however, all of the children were recruited from public schools from the same area and were typically from Hungarian middle-class families.

Materials and Procedure

Audio recordings were collected via conversational speech samples as the children interacted with an experimenter, discussing their favorite pastimes, everyday lives, or favorite stories. These interactions were quasi-naturalistic to prompt conversational samples from the participants and, at the same time, provide a somewhat controlled context so that the samples would be comparable. Recordings were conducted at school (kindergarten or elementary) in a quiet room to provide familiar environment for the children using a Zoom H4n portable recorder. Each recording session included a minimum of 5 minutes of conversation between the participant and the experimenter.

In order to control for factors affecting vowel duration, the following criteria were used for selecting vowels for analysis:

a.) Only allophones of /i/, /i:/, /o/, /o:/, /u/, and /u:/ were analyzed to control for vowel quality and to ensure that the pairs would only differ in duration. In addition, as vowel pairs, these are among the most frequently occurring ones in Hungarian (Gósy, 2004).
b.) Distorted productions of vowels (such as substitutions or hesitations) were excluded from the analyses.
c.) Terminal vowels (i.e. vowels in absolute final position) were not included in the analyses.

The final data set included a total of 2413 vowels whose durations were analyzed using PRAAT (Boersma & Weenink, 2015). Table 1 displays the number of vowels analyzed per each phonemic category. The segmentation of the vowels was based on their second formants supported by visual analysis of their respective wide-band spectrograms and waveforms. All of the vowel measurements were verified via interrater reliability by two of the authors of this paper. Items that were in disagreement were discarded from the analyses.

Table 1. Number of the analyzed vowels

Vowel	Short	Long
[i]	754	169
[o]	866	269
[u]	262	93
Total	1882	531

After obtaining and verifying the measurements, the durations of the vowels were analyzed using a repeated measures ANOVA. The independent variables were age with three levels (5-, 6-, and 7-year-olds; between-subjects variable), vowel quantity with two levels (short versus long; within-subjects factor), and vowel quality with three levels (/i, i:/, /o, o:/, and /u, u:/; within-subjects factor). The dependent variable was the duration of the vowel measured in milliseconds.

Results

Before conducting the analyses for testing our hypotheses, we verified that our data adhered to the assumption of sphericity. In order to test this assumption, we conducted Mauchly's tests of sphericity and found no statistically significant ones for the within-subjects effects and their interactions, suggesting that the variances of the differences between all pairs of related groups were equal. Consequently, the F ratios for our ANOVA were interpretable and valid.

Our first hypothesis predicted that we would find differences in vowel duration between phonemically short versus long pairs for all the participants. This hypothesis was supported by a main effect for vowel quantity [F (1, 21) = 67.49 at p = 0.000, partial η^2 = 0.76]. Figure 2 below displays the means for vowel duration in milliseconds and their respective standard deviations per vowel for each age group.

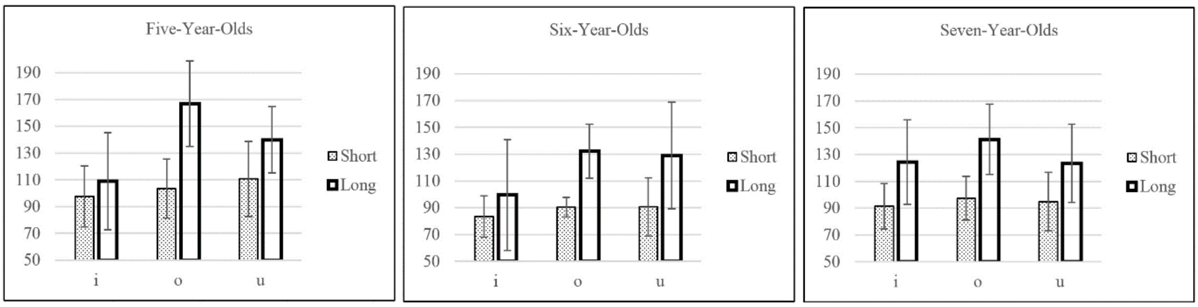


Figure 2. Mean vowel durations in milliseconds and their standard deviations

The second prediction was that as vowel quantity contrasts developed, age effects would be better expressed at age 7 than at age 5. In order to test this hypothesis, we investigated the main effect of age as well as the interaction effect of age by vowel quantity. Neither the main effect for age [F (2, 21) = 3.03 at p = 0.070, partial η^2 = 0.224], nor the interaction between age and vowel quantity were statistically significant [F (2, 21) = 0.18 at p = 0.838, partial η^2 = 0.017]. These findings suggest that age did not have an effect on vowel duration, and vowel quantity differentiation also did not depend on the age of the participants.

According to our third hypothesis, we expected to find an overall vowel quality effect as well as a dependence of that effect on the quantity of the vowel. This hypothesis was supported by our data in that there was both a main effect of vowel quality [F (2, 42) = 10.12 at p = 0.000, partial η^2 = 0.33] as well as an interaction effect for vowel quality by vowel quantity [F (2, 42) = 5.69 at p = 0.007, partial η^2 = 0.213]. These results indicate that vowel quality affects vowel duration in general and the effect of vowel quality depends on the quantity of the vowel (short versus long). Figure 2 above illustrates the differences between the vowels depending on their quality and quantity, separated by age.

Discussion

Overall, our findings suggest that Hungarian-speaking children do differentiate short and long vowels based on segmental duration in their conversational speech production, and vowel quantity is produced in a distinct fashion even at 5 years of age. Furthermore, contrary to our prediction that vowel quantity discrimination would depend on age, we did not find an age by vowel quantity interaction, suggesting that 5-year-olds, 6-year-olds, and 7-year-olds may not produce short versus long vowels in a unique fashion. It is also possible that such an interaction exists but the limitations of our data (discussed further below) did not allow us to find the effect, so further research is needed in this area.

Regarding the effects of vowel quality on vowel duration, we found support for both the idea that there is an overall quality effect, but perhaps more importantly, we also found evidence that vowel quality has a differential effect on vowel quantity. That is to say, the quality of the vowel interacts with vowel quantity, so short versus long vowels may be affected differently based on the quality of the vowel.

Our study contributes novel information to the literature, but it is not without its limitations. Future studies should employ a longitudinal design as well as a larger number of participants. Another limitation is that our data are based on conversational speech samples, so the linguistic environment could not be completely controlled. However, having conversational speech recordings does reflect naturalistic spontaneous speech, so in that respect our data are more representative of real speech than more controlled samples (such as single-word elicitation tasks). In the future, it would be desirable to include both conversational and controlled speech samples and compare the two to investigate the effects of the linguistic environment on vowel quantity and quality. In addition, our study did not compare the productions of children to the adult target, so studies should also incorporate comparisons of children's productions to the adult target to investigate the age at which vowel durations become adult-like.

The present study represents pilot work that added information to our knowledge base regarding the acquisition of phonemic vowel quantity contrasts in monolingual Hungarian-speaking children. Our data also generated new questions regarding how vowel quantity contrasts develop and at what age they become adult-like, prompting a need for further research in the area as well as providing specific direction for subsequent studies.

References

Boersma, P., & Weenink, D. (2015). Praat (Version 5.4.08) [Computer software]. Retrieved May 23, 2015, from http://www.praat.org/

Bóna, J. (2012). A rövid-hosszú magánhangzók realizációi idősek spontán beszédében [The production of long and short vowels in the spontaneous speech of the elderly]. *Beszédkutatás 2012*, 43-57.

Bóna, J., & Imre, A. (2010). A rövid/hosszú magánhangzók óvodás és kisiskolás gyermekek beszédprodukciójában [Long/short vowels in preschool- and school-age children's speech]. In J. Navracsics (ed.), *Nyelv - beszéd - írás* (pp. 49-56). Tinta Kiadó, Budapest.

Deme, A (2012). Óvodások magánhangzóinak akusztikai jellemzői [Acoustic properties of vowels produced by preschoolers]. In A. Markó (ed.) *Beszédtudomány. Az anyanyelv-elsajátítástól a zöngekezdési időig* (pp. 77-99). Nyelvtudományi Intézet, Budapest.

Gósy, M. (2004). *Fonetika, a beszéd tudománya* [Phonetics: The science of speech]. Osiris Kiadó, Budapest.

Gósy, M. (2006). A beszédhangok megkülönböztetésének fejlődése [The development of phonemic differentiation]. *Beszédkutatás 2006*, 147-159.

Gósy, M., & Beke, A. (2010). Magánhangzó-időtartamok a spontán beszédben [Vowel duration in spontaneous speech]. *Magyar Nyelvőr*, *134*, 140-165.

Kassai, I. (1979). Időtartam és kvantitás a magyar nyelvben [Duration and quantity in Hungarian]. *Nyelvtudományi Értekezések*, *102*. Akadémiai Kiadó, Budapest.

Lehiste, I. (1965). The function of quantity in Finnish and Estonian. *Language*, *41*(3), 447-456.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: Cambridge University Press.

Malmberg, B. (1971). *Svensk fonetik* [Swedish phonetics]. Gleerups, Lund.

Nádasdy, Á., & Siptár, P. (2001). A magánhangzók [Vowels]. In F. Kiefer (ed.), *Strukturális magyar nyelvtan 2: Fonológia* [Sturctural Hungarian grammar 2: Phonology] (pp. 42-94). Budapest, Akadémiai Kiadó.

Szende, T. (1994). Illustrations of the IPA: Hungarian. *Journal of the International Phonetic Alphabet*, *24*(2), 91-94.

Zajdó, K. (2002). *The acquisition of vowels by Hungarian-speaking children aged two to four years: A cross-sectional study*. Unpublished doctoral dissertation. University of Washington.

Zajdó, K., & Powell, S. (2008). The acquisition of phonological vowel length in children acquiring Hungarian. In R. Sock, S. Fuchs & Y. Laprie (eds.) *Proceedings of the 8th International Seminar on Speech Production* (pp. 173-176). Strasbourg, France.

Zajdó, K. (2015). Can caregivers facilitate the acquisition of language-specific segment-duration patterns in children's speech? Paper presented at the *International Symposium on Monolingual and Bilingual Speech 2015*, Chania, Greece.

Are they still Russian-speaking? Comparing the heritage learners of Russian in non-formal frameworks in Israel and Italy

Marina Niznik[1], Monica Perotto[2]
marinan@post.tau.ac.il, monicaperotto@alice.it

[1] Tel Aviv University, [2]University of Bologna

Abstract. The paper presents the results of a field study conducted in Israel and Italy in 2014. The aim of the study was to investigate the process of language acquisition by children from Russian-speaking families who enrolled in non-formal Russian educational networks. Fifty-seven adolescents from Israel and 45 adolescents from Italy aged 8-15 participated in the study. All of them may be defined as heritage language (HL) learners (Valdés, 2000; Polinsky & Kagan, 2007). The research utilized a specially designed test consisting of oral and written components. Although there are substantial differences between the Russian-speaking populations in the two countries, the present research indicated that the motivation for learning Russian as well as the difficulties encountered during the learning process were very similar in both samples. The case system, verb aspect, and verbs of motion were found to be the most challenging topics to be mastered by all the respondents. Oral proficiency in both groups was much higher than their written proficiency. For the majority, an integrative motivation prevailed as opposed to an instrumental motivation. The present findings constitute an important basis for developing teaching materials that can be used at least partially in many countries outside the Russian Federation.

Keywords: second language pronunciation, intelligibility, word stress, tonal word accent

Introduction

After the collapse of the USSR in the early 1990s, a large number of former Soviet citizens immigrated from Russia and other Soviet republics. Although the scientific and popular literature addresses these people as "Russian" or "Russian-speaking" and view them as a homogeneous group, they in fact constitute a varied conglomerate of people who differ in many aspects, such as their nationality, beliefs, cultural and educational backgrounds, and motives for emigration.

This pertains to people for whom Russian is their mother tongue as well as for those who are bilingual (usually those of Ukrainian or Belorussian origin). There are also those who use Russian as their *lingua franca,* usually coming from the indigenous populations of the non-Slavic former Soviet republics (Laitin, 1998, p. 29-31).

No unified pattern of absorption has been found for former Soviets in their new countries. In places with a relatively high concentration of Russian-speaking newcomers, they tend to form Russian enclaves with a diverse array of services in Russian, including shops, clinics, clubs, evening schools and more. Such enclaves are very typical in Israel and the USA (Gold, 1997; Niznik, 2005). Perotto (2009) found that in Italy, Russian-speaking immigrants are scattered across the country, which may partially explain why Russian services, such as Russian evening and summer schools are less available in Italy than in Israel, where there are many Russian enclaves. As a result, the preservation or loss of the Russian language is almost entirely dependent on the language policy within the family, i.e. how much time, effort and money adult family members are willing to spend on maintaining their child’s language of origin. Many Russian-speaking children in Italy come from mixed families, where the language of communication is usually Italian. Thus, the task of maintaining and improving children’s fluency in Russian is even more challenging.

Given this, it is not surprising that the representatives of 1.5 generation, i.e. those who immigrated with their parents during childhood (Garcia-Call & Nagnuson, 1997) and second generation Russian-speaking immigrants in Italy form a rather heterogeneous group. Their level of proficiency depends on their birthplace, where they began their formal education, in what language it was conducted, how long it lasted (when conducted in Russian), and their family's language policy.

Despite the different history and nature of the Russian-speaking immigration to Israel and Italy (returning Diasporas vs. immigration), those who identify themselves as "Russians" in both countries have much in common. First, Russian is the mother tongue for the vast majority of first-generation immigrants, and they share a strong affiliation with Russian culture, which they perceive as the most important part of their identity. Therefore, many immigrants attempt to at least partially transmit their native language and culture to the next generation.

In Italy, the increasing number of immigrants from Russia and the former Soviet Union is a relatively recent phenomenon. According to the *Dossier Caritas Migrantes* 2012, the total number of legal migrants in Italy by the end of 2011 amounted to 5.011 million people, including 223,782 Ukrainians, 147,519 Moldovans, and 7,090 Russians (Dossier Statistico, 2012).

The first wave of immigration dates from the beginning of the last century, although Russian writers and travelers came to Italy earlier as well. Russians have never formed enclaves in Italy, the Russian community was diverse, yet dominated by Russian nobility and the intelligentsia (Scandura, 1995). Today the third generation of this first wave of Russian-speaking migration lives in Italy, yet only few still speak Russian (Perotto, 2009, 2010, 2013).

Israel, unlike Italy, is a country of immigrants (Michaeli, 2007). However, immigrants from Russia undoubtedly played a unique role in the history of the country. The Jewish immigration from Russia to Palestine began in 1882 with the repatriation of young Zionist activists from the *BILU* group. Over time, the flow of immigrants increased, then decreased, but never ceased entirely. The largest wave of Russian-speaking immigrants arrived after the collapse of the Soviet Union in the early 1990s of the 20th century. They now account for approximately 15% of the total population in the country (Epstein & Khanin, 2007). Russian influences are noticeable in almost all spheres of Israeli life. For example, most of the classic literature included in the school curriculum stems from Russian roots (Tchernichovsky, Rachel, Bialik). Zeev Jabotinsky, the founder of the "Beitar" movement, which later became the basis for the “Likud” (now the ruling Party of Israel) was one of the most popular journalists in the Russian empire. Maxim Gorky, regretting the departure of Jabotinsky, argued that the Zionists gained what Russian literature lost (Wheatcroft, 2008).

Israeli theater also has significant Russians roots. In 1913, a group of Jewish actors created a studio at the Moscow Art Theatre headed by Konstantin Stanislavsky. Later, Stanislavsky gave “Habima” (the name of the studio) to one of his favorite students, Evgeniy Vakhtangov.

Habima (Hebrew for “The Scene”) became the first modern theater, presenting performances in Hebrew. This story repeated in a new version almost one hundred years later, when in the early 1990s, a group of actors led by former Moscow director Yevgeny Arye created *Gesher* theater in Tel-Aviv (Hebrew for “bridge”), which became one of the most popular theater companies in Israel.

In the 1960s-1970s the prestige of Russian culture in Israel waned, and it gave way to the American culture and lifestyle.

The mass immigration of the 1990s partially revived Russian culture in Israel. The number of newspapers in Russian dramatically increased, a new TV station was established, and radio channels broadcasting in Russian were launched. The advent of the Internet enabled various electronic resources. In light of these changes, the question of the place of the Russian language in the formal and informal education systems became more relevant than ever. Russian-speaking parents and the teachers of Russian began to lobby for the introduction of the Russian language to the list of officially recognized school subjects. They succeeded only partially. The Russian language was introduced as an optional subject, mainly in high school. At the same time, students in primary and even in secondary schools do not have the opportunity to study their language of origin as part of their regular curriculum.

Those who wish to learn the language at an early stage usually turn to the private sector, Russian evening schools or private tutors. A considerable number of Russian educational networks were established in the early nineties aiming to bridge the gap between the Israeli public education system and the aspirations of many Russian-speaking newcomers. The Russian language became an essential issue in their curriculum (Kopeliovich, 2011).

Profile of children who speak Russian as a heritage language

"One-and-a-half" and second generation young immigrants are defined as heritage speakers, that is, in the words of Maria Polinskaya (Polinskaya, 2010, p. 344; Polinsky & Kagan, 2007, p. 369), their first language became their home or family language.

Interest in language maintenance, language shifts and native language attrition has increased dramatically in recent years and has led to research of this phenomenon among diverse languages and age groups. The following definitions are widely quoted by researchers, who characterize heritage language as a cultural and linguistic phenomenon. According to Fishman (1992), "A heritage language is any language that a given group/individual has cultural, ethnic, or religious allegiance to (but does not have to have a speaking ability in)."

This population is often defined as unbalanced bilinguals. Their language skills depend on whether they were born in a country where they spoke their heritage language or in the country of their present residence (Orfitelli & Polinsky, 2013). It is also of importance in what language they acquired their primary literacy skills.

Valdés (2000) focuses on language use, stating that a heritage language is a language that an individual grew up overhearing or speaking at home. These definitions establish fundamental distinctions between heritage and non-heritage language learners. They are neither typical students of a foreign language, nor of a native language. Kagan and Dillon (2001), who explored the profile of heritage students of Russian in the USA and examined the distinctions between these students and traditional foreign language students, claim that to place heritage speakers together with students of Russian as a foreign language is to fail the needs of both groups. Researchers discovered that the heritage learners of Russian often have high aural proficiency, native-like pronunciation, and vocabulary that is adequate for the needs of family and possibly the community as well.

Similar conclusions were made in a survey of Russian-speaking adolescents in Israel (Niznik, 2005). These subjects have some grammatical intuition that will function effectively if supported by a declarative knowledge of grammar. They stagnate in classes that focus on rudimentary pronunciation, beginnings of grammar and vocabulary. Heritage learners need a macro, not a micro approach to grammar, a paradigm of declensions and conjugations rather than a case-at-a-time approach. They also need extensive work on orthography, unlike foreign language learners who can basically write anything they can say.

Since the family language is mainly used in everyday communication, it strongly interacts with the dominant language, to varying degrees, and is often heavily influenced by the latter. This leads to what experts call cross-language transfer (Dominguez 2009) or crosslinguistic interference (Paradis & Navarro, 2003). One of the most common problems for children who immigrated at an early age is the lack of regulatory support in their heritage language learning. As a result, they learn only the language their parents or close family members use. Dominguez calls it *incomplete acquisition* or *deviated input* (Dominguez, 2009).

The duration and intensity of the child's communications with his or her parents (or one of them) in their native language, and the compliance of this language with the accepted norms become crucial factors in heritage language acquisition (Genesee & Nicoladis, 2005). Otherwise, as Montrul stated:

> "When bilingual children are exposed to less than optimal input conditions during the age of primary linguistic development (birth-4 years) and/or the period of later language development that takes place during the pre-school and school years (4-13 years), many aspects of grammar may not reach full development and remain incompletely acquired" (Montrul, 2009, p. 241).

Heritage learners tend to be fairly fluent in the spoken language mainly within the family sphere (Valdés, 2000, 2001; Kagan & Dillon, 2001; Niznik, 2005). At the same time, their mastery in writing and their pragmatic competence are relatively low. Their weak skills in reading and writing often prevent heritage speakers from acquiring further mastery of their language. The only language sources available to them are limited to oral speech. This makes systematic tutoring in various after-school networks quite crucial for maintaining the language for the next generation.

Aims and objectives of the study

The present study aimed to compare the results of a field research conducted simultaneously in Italy and Israel on the basis of a unified written and oral test**.** The subjects, heritage speakers aged 8-15, were recruited from Russian-speaking and mixed families. Montrul (2009) defines this age as a period of later language development.

Special attention was paid to the identification of linguistic difficulties and possible signs of language shift in each sample. A comparative analysis of the results indicated common characteristics as well as differences in the oral and written speech of the subjects in the two groups.

Sample and research methodology

In Italy, 45 respondents (aged 8-15) participated in the survey. They attended Russian Centers for Continuing Education in Rome (25 respondents attended the Nikolay Gogol Education Center) or Milan (20 respondents attended the Harmony Educational Center).

Twenty-eight respondents were born in Italy; 17 in Russia, Moldova or Ukraine. Twenty-seven were raised in mixed families; 17 came from families that spoke only Russian.

In Israel the sample included 57 respondents aged 9 to 14 from across the country. Thirty-eight respondents were born in Israel, while the remaining respondents immigrated with their families from Russia or the former Soviet republics. Only two respondents were raised in mixed language families.

Despite the different nature of Russian-speaking immigration to Israel and to Italy, the Russian Centers for Continuing Education have much in common in these two countries.

Italian respondents attend so-called *Saturday schools*, paid centers for additional education, where the courses are taught entirely in Russian. They operate once a week (on Saturdays) and offer between 4 to 7 hours of tutoring in various subjects. Children and young people from kindergarten until the end of high school are involved in this educational network. The youngest children usually study 4 hours a day. The instruction is focused on language development through play, music and dance.

Primary school children can take test preparation classes and Russian language classes. In addition, they can study math, literature, environment and art.

The Nikolai Gogol Educational Center in Rome was founded in 2003 in the parish of the Orthodox Church of St. Nicholas, the Wonderworker of Myra. The center offers an option of obtaining an academic degree in Russian. Religious tutorship is also an integral part of the educational process. Russian is not only the language of instruction but also the language of communication in the center. The center also offers dancing lessons, piano lessons, as well as psychological counseling for children and parents.

The Harmony Educational Center was established in Milan in 2005. This is a school complex (including a pre-school, primary and secondary schools), which, in the founders' words, is "united by common goals: a uniform, continuous process of education, that comprehensively covers the development of the child at all stages of his growing up" (www.scuolarussamilano.com).

The activities of the center include a choir, an art studio, a chess studio, drama, and English language classes. There are also various tourist activities available for the Harmony students: city tours, trips and summer youth camps.

In Israel, as mentioned above, Russian is taught as an optional school subject in governmental schools as well as in various private centers for continuing education. The respondents in this research were recruited only from non-formal educational networks.

Eight Israeli respondents were students in the Jerusalem evening school IGUM – an acronym of *Irgun Morim Olim* – The Association of Immigrant Teachers, who consider themselves followers of the famous Russian psychologist Lev Vygotsky.

A lot of attention is given to the comprehensive development of each child. IGUM students are playing to learn – the teachers view games as the most important educational tool. The curriculum includes Russian language and literature, reading, math, logic, applied arts, and the English language.

Fifteen respondents attend the evening lyceum called “Impulse” in Beer-Sheva which was founded by the Ben-Gurion University staff. Lyceum students can take classes in Russian language and culture, natural science, and chess.

An additional twenty respondents attend the Herzliya lyceum “Erudite”, which has existed for over 40 years. The curriculum there consists of Russian and English, physics, math, art and logic. Russian culture is an integral part of the Russian language lessons.

The “Mofet” network (*model, sample* in Hebrew) numbers about 17 evening schools with intensive study of physics and math. Russian language is also part of the curriculum. Nine respondents attend the evening school “Mofet” in Haifa, while another four respondents study with private tutors. The majority of respondents study Russian two academic hours a week.

The study utilized a test developed especially for this research, which consisted of three parts. In the first part, the respondents were asked to present themselves in writing and to answer a few questions about themselves and their family. This was followed by a reading comprehension task. The short story by V. Dragnskiy “A Childhood Friend” was adapted specifically for this research. The tasks not only assessed the respondents’ writing and reading skills, but also their ability to identify stylistically marked forms of speech that are assumed to be challenging for this audience. In the second (oral) part of the test, respondents were asked to make up a picture sequence story “Peak Badaluk went to the Forest”.

Most of the children in the study qualified as subordinative (or late) bilinguals, because they did not acquire their two languages simultaneously during their first three years of life, when implicit memory is developed. Russian is not the first language to all of them; some began being exposed to Russian after the age of three, during development of “explicit memory”, which is why their automatic language skills are less accurate on the syntagmatic level (Paradis, 2005).

When Russian-speaking children enter primary school, they increasingly speak the dominant language, speaking less and less Russian. Table 1 shows where they use Russian.

Table 1. Use of Russian

Country (number of respondents in the sample)	Language spoken at home	Language spoken with friends	Language they reported as easy to speak
Italy (45)	Russian and Italian: **22** Russian: **17** Italian: **12** Russian and other languages: **2**	Italian: **29** Russian: **1** Russian and Italian: **13** Russian, Italian and Ukranian: **1** No answer: **1**	Italian: **35** Russian: **8** Russian and Italian: **1** No answer: **1**
Israel (57)	Hebrew: **47** Russian: **10**	Hebrew	Hebrew: **40** Russian and Hebrew: **15** No answer: **2**

This data demonstrates that not all the resondents speak Russian at home and very few speak Russian with their friends. The majority feel much more comfortable with the dominant language. Similar results were obtained by Kagan and Dillon (2001) among American respondents.

Findings

The written task presented a greater challenge for the respondents. Not everyone was familiar with the Russian written script. This fully supports the findings of Peeters-Podgaevskaya and Dorofeeva who claim that school-age immigrant children require constant training of their graphic memory to bridge the gap between them and monolingual children. In their words, "a monolingual child learns the basics of the grammatical structure of the Russian language at the age from five to seven. Bilinguals are usually far behind" (Peeters-Podgaevskaya & Dorofeeva, 2013).

In the Italian sample, all of the respondents coped with the task with greater or lesser success; in the Israeli sample two participants failed. A few respondents answered the questions in printed letters (8 in the Israeli sample and 6 in the Italian). Learning the Russian alphabet is not an easy task either for Italians or Israelis, but for Israelis it is a greater challenge due to the great difference between the two scripts.

Interestingly, recent studies suggest that difficulties in writing are typical not only of bilinguals, but also of Russian students.

Russian experts say that the level of competence in writing has dramatically decreased in Russia. "A school graduate does not go beyond the level of A2 in his writing assessment" (Evgrafova, 2011, p. 287-288). However, there is no doubt that for the aforementioned reasons, the written assignment for the current research sample was far more difficult than for their peers in Russia.

The analysis of the test results revealed the following types of errors and language difficulties:

Spelling mistakes

The writing of many respondents can be qualified as phonetic.

The most common mistakes in the written test were as follows:

-A confusion between the letters **И** [i] and **Й** [y]**:** *в Россий*, *в Италий* [v Rossiy, v Italiy] in the Italian sample;

- In both samples, many misused the soft sign **Ь** and hard sign **Ъ** or simply omitted them: попросил *купит* грушу, *семя, италианский, италянский* [poprosil *kupit* grushu, *semya, italianskiy, italyanskiy*] (*to buy a pear, family, Italian*);

- The use of **Ш** [sh] instead of Ч [ch] in both samples: *што, штобы, потомушто* [shto, shtoby, potomushto] (*what, that, because*);

- Various errors when writing vowels in the unstressed position: *чимпёнат* мира, в *телевизере, шыколадного, жорналистом* [chimponat mira, v televizere, shykoladnogo, zhornalistom] (*world Cup, on TV, from chocolate, a journalist*) in both samples.

There were more errors in writing unstressed vowels in the Israeli sample, and the Israelis also often omitted vowels, which may be explained by the fact that in Hebrew vowels are not often designated by the letters: *дволен*, из *двана,* [dvolen, iz dvana] (*is content, from the coach*).

Israeli children often spell names with a lowercase letter. This error appeared in approximately one-third of the sample. Some respondents wrote "Bear" in one place with an uppercase grapheme and in another case with a lowercase grapheme. This recurrent mistake likely stems from the fact that in Hebrew there are no capital letters.

Fused spelling of nouns and pronouns with a preposition, for example *влесу, унево, кнему* [vlesu, unevo, knemu] (*in the forest, he has, to him*) can also be explained by the influence of Hebrew, because this corresponds to the rules of spelling in Hebrew.

Morphological errors

There was great diversity in morphological mistakes in both samples. The most common in oral speech was prepositional-case mismanagement. For example: «Это его родители. Они живут в *одного дома*. Бабушка говорит что-то *мальчика*»; «живут в *домик*»; «говорю на *русскому языку* / по *русскому языке*»; «пошел со *свои* друзьями» [«Eto yego roditeli. Oni zhivut v odnogo doma. Babushka govorit chto-to mal'chika»; «zhivut v domik»; «govoryu na russkomu yazyku / po russkomu yazyke»; «poshel so svoi druz'yami»] (*These are his parents.They live in the same house. …speak Russian… went with his friends*).

The use of accusative of animate nouns was particularly challenging for children in both countries. For example: «я вижу *один мальчик*»; «пришел папа и звал *свои друзья*» [ya vizhu odin mal'chik»; «prishel papa i zval svoi druz'ya»] (*I saw a boy … Dad came and called his friends*). Many failed to use the right form if «лев - льва» - *lion* in accusative: «Увидел *лев* и испугался»; «Они убили *лев/лева*» [«Uvidel lev i ispugalsya», «Oni ubili lev/leva»] (*He saw a lion and got scared*. *They killed the lion*).

However, errors in the Italian and the Israeli samples did not show many signs of overgeneralization, i.e. prevalence of one case (Polinsky & Kagan, 2007). In the Italian sample, these errors were present to a greater extent in the speech of children who were not exposed much to Russian, did not receive any tutoring in Russian and lived in mixed families.

The greatest number of errors in the Israeli sample was observed in the use of the accusative and instrumental cases – the respondents tended to use the nominative instead, e.g., «Мальчик жил с *папа* и *мама»*; «Все друзья пошли выручать *Пик*»; «Мальчик хотел быть *астроном*» [«Mal'chik zhil s papa i mama»; «Vse druz'ya poshli vyruchat' Pik»; «Mal'chik khotel byt' astronom»] (*A boy lived with his mom and his dad. The friends came to rescue Pick*).

Some respondents used the genitive or accusative instead of the instrumental: «…побежал за *Пика*, начал бегать за *Пика*» [… pobezhal za Pika, nachal begat' za Pika] (*ran after Pick*).

Many mistakes were found in the use of nouns after numerals, such as «шесть *месяца*, 3 *лет*» [«shest' mesyatsa, 3 let»] (*in six months, three years*).

The use of possessive pronouns is also a complex task: «Мама плакала, потому что не было *его* сына»; «Я вижу мальчику, бабушку и муж *его*» [«Mama plakala, potomu chto ne bylo yego syna», «Ya vizhu mal'chiku, babushku i muzh yego»] (*Mom was crying because her son wasn't there. I see boys, granny and her husbund*).

Children from the Israeli sample also confused pronouns «свой» (which is used instead of possessive pronouns when referring back to the subject of the sentence) and «его, ее» (*his*, *her*): «Пик вернулся к *его* маме»; «его папа позвал всех, чтобы найти *его* сына» [«Pik vernulsya k yego mame», «yego papa pozval vsekh, chtoby nayti yego syna»] (*Pick went back to his mom. His father called everybody to find his son*). There is no equivalent to pronoun «свой» [«svoy»] in Hebrew. In the Italian sample, this error was less common and was more likely due to inter-language transfer: in Italian the pronouns 'suo' (m), 'sua' (f) (*his*, *her*) sound very similar to «свой», «своя» [«svoy», «svoya»]. However, in the Italian sample, the children accurately matched the noun with the pronoun in gender: «свой домик, своя мама» [«svoy domik, svoya mama»] (*his own house, his mother*).

In both samples, there were mistakes in the use of verbal aspects. A few examples from the Italian sample are: «Пришел папа и *звал* свои друзья» (*Dad came and called his friends*), «долго сын *потерялся*» (*his son was lost for a long time*), «они *убьют* его и Пик Бадалук идет к маме» (*they would kill him and Pick Badaluck would get back to his mother*). There were many similar examples in the Israeli sample, as well: «Вдруг вышел лев и *бегает* за Пика»; «Вдруг мальчик *видел* льва»; «Все *бегали* искать Пика» [«Vdrug vyshel lev i begayet za Pika», «Vdrug mal'chik videl l'va», «Vse begali iskat' Pika»] (*All of a sudden the lion came out and started chasing Pick. All of the sudden the boy saw the lion. Everybody started looking for Pick*). As noted in the relevant literature (e.g., Polinsky, 2008, p. 154-155), the Russian language of American children from Russian-speaking families is dominated by semantic rather than morphological expressions of the verb form.

> “The descriptive generalizations are clear. First, in the aspectual system of American Russian, just one member of the so-called aspectual pair is typically maintained. Second, and related to the first point, there are a large number of analytical expressions which replace synthetic aspectual verbs in Full Russian” (Polinsky, 2006, p. 230).

Regarding the first hypothesis, the examples (some listed above) confirm that the verbs denoting activities in process (atelic- Italian), e.g., “звать”, “плакать”, “жить” [“zvat”, “plakat”, “zhit"] (*call, cry, live*) are always used in an imperfective form, while the verbs denoting activities with results (telic), such as “потеряться”, “убить”, “дать” [“poteryat'sya”, “ubit'”, “dat'”] (*get lost, kill, appear*) are used in their perfective form.

Many respondents in both samples confused the use of the verbs *говорить-сказать* [govorit'-skazat'] (*speak*, *say*). For example: «Мама *говорила с мальчиком*, что нельзя выходить за забор», «Мама разрешила мальчику выйти из дома, но *говорила*, что нельзя идти в лес»; «Мама обняла Пика но *говорила*, что нужно слушаться родителей» [«Mama govorila s mal'chikom, chto nel'zya vykhodit' za zabor», «Mama razreshila mal'chiku vyyti iz doma, no govorila, chto nel'zya idti v les». «Mama obnyala Pika no govorila, chto nuzhno slushat'sya roditeley»] (*Mom told the boy not to get over the fence. Mom hugged the son and told him he should have listened to his parents*.)

The Israelis found it difficult to use perfective verbs, and they tended to use the imperfective rather than the perfective, and not vice versa, such as: «Вдруг мальчик *видел* льва» [«Vdrug mal'chik videl l'va»] (*All of a sudden a boy saw a lion*). An opposite pattern was found in the Italian sample where the perfective was used instead of the imperfective: «Мама *скажет*, что не надо идти в лес»; «она *скажет* папу Бадалук и папа *скажет* эти люди» [«Mama skazhet, chto ne nado idti v les», «ona skazhet papu Badaluk i papa skazhet eti lyudi»] (*Mom would say that he shouldn't go to the forest. She would tell Badaluck's father and he would tell other people*.) In this case, again, the children adopted a strategy of semantic identification between the two members of the verb pair: “*сказать*” [skazat'](*to say*) (*telic* verb) used in the sense of the Italian verb *dire*, while *talking* (*atelic* verb) was used with the meaning of *parlare*.

When analyzing the use of the verbs, special attention was given to difficulties associated with the prefixed verbs.

To indicate the start of an action, respondents often substituted the inchoative (ingressive) verb with a prefix по- [po-] or про- [pro-], the faze verb *начать* [nachat'] (*to begin*). The verb was used as a link-verb and was followed by an infinitive of the imperfective. For example: «он испугался и *начал* плакать»; «бабушка заметила и *начала* плакать»; «папа *начал* гудеть в трубу», «он начал бегать…, начал кричать» [«on ispugalsya i nachal plakat'», «babushka zametila i nachala plakat'», «papa nachal gudet' v trubu», «on nachal begat'…, nachal krichat'»] (*He got terrified and started to cry, granny payed attention to it and started to cry, dad started to blow the trumpet, he started to run around… and started to scream*).

These findings strongly confirm the second hypothesis (referred to in Polinsky, 2006, p. 225 and Polinsky, 2008, p. 156) as examples of the reorganization of the verb aspect system in the Russian language of Russian heritage speakers.

Similar structures can be found in the speech of native speakers, but not as one of the prevailing ways of perfectation.

In both samples, this can be explained by the influence of the dominant language: the ingressive verb cannot be formed with the help of a prefix in any of these two languages, but only by lexical means like *iniziare* (*start*) in Italian or *התחיל* (*start*) in Hebrew.

In the Israeli sample, mistakes were also observed in the formation and use of the reflexive verbs, such as: «Мама *ругалась* на Пика, когда он выходил за забор»; «однажды Пик *захотелся* пойти в лес»; «Мишка *был* очень *обрадовался*» [«Mama rugalas' na Pika, kogda on vykhodil za zabor», «odnazhdy Pik zakhotelsya poyti v les», «Mishka byl ochen' obradovalsya»] (*Mom scolded her son for jumping over fence; once Pick wanted to go to the forest; Mishka was very glad*).

Sometimes the respondents used reflexive verbs without the necessary postfix *-sya:* «лёв очень *разозлил* на Пика»; «мама говорит Пику, что нужно *слушать* старших» [«lov ochen' razozlil na Pika», «mama govorit Piku, chto nuzhno slushat' starshikh»] (*The lion got mad at Pick, Mom sad to Pick he should listen to adults*). There are reflexive verbs in Hebrew, but they do not always correspond with Russian reflexive verbs, which sometimes leads to errors. For the Italian sample, reflexive verbs do not present much difficulty, with the exception of «Мальчик *испугался лев*» [«Mal'chik ispugalsya lev»] (*A boy got terrified by the lion*), which was observed in several responses.

Many respondents had difficulties with verbs of motion. More than half of the Israeli respondents confused the verbs *принести* [prinesti] (*to bring*) and *привести* [privesti] (*to bring, to lead*): «Пика *принесли* домой» [«Pika prinesli domoy]» (*Pick was brought home*). Many confused the verbs *бегать* [begat'] (*to run*) и *(по)бежать* [pobezhat'] (*to start running*): «Вдруг вышел лев и *бегает* за Пика»; «Все *бегали* искать Пика» [«Vdrug vyshel lev i begayet za Pika», «Vse begali iskat' Pika»] (*All of a sudden the lion came out and started to chase Pick. All the people started to run looking for Pick*) Italian respondents frequently used the expression «*убегает* на дерево» [«ubegayet na derevo»] (*escaped to the tree*), «вместо «залезает на дерево» [«zalezayet na derevo»] (*climb the tree*). The respondents in both samples often misused expressions denoting the direction of motion: «он не должен идти *в лесу*»; «пошел *в лесу*», «залез *на дереве*» [«on ne dolzhen idti v lesu», «poshel v lesu», «zalez na dereve»] (*He shouldn't go to the forest, he went to the forest, he climbed the tree*). There was a lot of confusion in both samples in the use of the following combination: the prefixed verb of motion + preposition + noun. For example: «мальчик пришел *до мамы* и все хорошо» [«mal'chik prishel do mamy i vse khorosho»] (*The boy came back to his mom and everything ended well*). These mistakes were especially typical of children who lived in mixed families (in the Italian sample) or for those who were born and raised outside the language environment (in the Israeli sample). At the same time, children who speak Russian all the time at home found it easy to correctly use expressions, such as «залез на дерево», «убежал к маме» [«zalez na derevo», «ubezhal k mame»] (*climbed the tree, ran back to mom*). One of the youngest respondents (aged 8), who was born in Italy, lived in a Russian family and attended a Russian Saturday School since the age of 4, demonstrated the correct use of the verb pair *залезать-залезть* [zalezat'-zalezt'] (*to climb*): «Пик на дерево залезает, а лев, он не может залезть на дерево» [«Pik na derevo zalezayet, a lev, on ne mozhet zalezt' na derevo»] (*Pick climbed the tree and the lion couldn't climb the tree*).

Lexical errors

Many lexical mistakes in the Italian sample obviously originated in calquing Italian. For example: «*не можно*» [«ne mozhno»]) (*impossible*) (Italian: *non si può*) – a detachment of denial; «они все *африканы*» [«oni vse afrikany»] (*they all are Africans*) (Italian: *africani*); «папа с трубой *делает знак*, *делает звон*» [«papa s truboy delayet znak, delayet zvon»] (*Dad with a trumpet made a sign, he rang*) (Italian: *fa un segno, fa un suono*); «он не знал *где* (Italian: *dove*) идти» (*He didn't know where to go*) – in Italian the interrogative word for motion and state are the same. One respondent said: «Пик пошел в *джунглю*» [«Pik poshel v dzhunglyu»] (*Pick went to jungle*) calquing the Italian word *igiungla*, which is feminine. There are plenty of calques in the Israeli sample, as well: «Пик начал *кричать* людей» [«Pik nachal krichat' lyudey»] (*Pick started to cry for people*). לקרוא in Hebrew means both *позвать* [pozvat'] (English: *to call*) and и *кричать* [krichat'] (*to cry for*); «стали звать людей из *поселения*» [«stali zvat' lyudey iz poseleniya»] (התנחלות) (*They started to call the people from the settlement*) «мама рассердилась, но обрадовалась в *ту же меру*» [«mama rasserdilas', no obradovalas' v tu zhe meru»] (באותה המידה) (*Mom was equally happy and angry*); «груша *стоит много*» [«grusha stoit mnogo»] (עולה הרבה) (*The punching ball cost a lot*).

In Italian, there were a few cases of borrowings, mainly of nouns, like: «Жил-был мальчик и его *фамилия*» [«Zhil-byl mal'chik i yego familiya [Italian: *family*]»] (*Once upon a time there lived a boy and his family*), or of code mixing: «Он бегал, бегал и ... *arrampicato* [Italian: *climb*] на дерево» [«On begal, begal i ... arrampicato (Italian: *climb*) na derevo»] (*He ran, ran… and climbed the tree*). «Лев ... lo *aggredisce*...(attack- Italian) на него» [«Lev ... lo aggrediscena nego»] (*The lion… attacked him*). «Папа ходил, ходил и позвал друзей взять... *armi* (Italian: *arms*)[«Papa khodil,

khodil i pozval druzey vzyat'... armi]» (*The father walked, walked and then asked his friends to bring weapons*). Borrowings and code-switching are typical for those raised in mixed families. Code mixing was rare in the Israeli sample. One example: «Он взял רובה» (Hebrew: *a gun*) [«On vzyal רובה»] (*He took a gun*), «пошел в יער (Hebrew: *forest*) [«poshel v יער»] (*he went to the forest*).

The lack of practice in Russian language, especially for children born outside Russian-speaking regions, often leads to difficulties and doubts in the choice of vocabulary. An example from an Israeli respondent: «Папа в дудку ... *позвонил* ... людей» [«Papa v dudku ... *pozvonil* ... lyudey»] (*The father blew the trumpet to call people*), and from the Italian sample: «Пик стал… *плеваться* яблоками во льва» [«Pik stal… *plevat'sya* yablokami vo l'va»] (*He began to spit…apples*). Interestingly, in both samples there were respondents who used the word “tiger” instead of “lion”.

In the Israeli sample an interesting neologism was observed: *негрик* [negrik] (*a black person*) and in Italian: an old verbal form *воротился* [vorotilsya] (*to come back*).

The oral fluency of the respondents was often interrupted by pauses; they often needed additional incentive or assistance to continue speaking. In the Italian sample, this was mostly typical of the respondents from mixed families. During the oral test, prompting was avoided as much as possible but, in some cases, it was obvious that the child’s vocabulary was not adequate to express what he/she wished to convey, so the tester provided the word the respondent was seeking in Hebrew or Italian, respectively.

Syntactic and stylistic mistakes

The most common mistake, characteristic of many respondents in Italy and in Israel, was the use of the conjunction *если* [yesli] (*if*) instead of the interrogative participle *ли* [li] «Мальчик просит у мамы, *если* может выйти» [«Mal'chik prosit u mamy, yesli mozhet vyyti»] (*A boy asked his mother whether he can get out or not*).

In their daily lives, the respondents speak but do not write in Russian, which can explain the frequent use of colloquial and vernacular forms in both their written and oral tests. For example: «в *телике*», «лев хочет его *кушать/скушать*», «*штоб*», «*чтоб*» лев не убил его» [«v telike», «lev khochet yego kushat'/skushat'», «shtob», «chtob» lev ne ubil yego»] (*on TV, therefore, the lion wants to at him*). *Что* [chto] (*that*) is used instead of *который* [kotoryy] (*which*): «Все бегали искать Пика, *что* был в лесу» [«Vse begali iskat' Pika, chto byl v lesu»] (*Everybody started looking for Pick, who was in the forest*).

Colloquial and slang expressions from the short story used in the test (*приспичило, спятил, перебейся*) [prispichilo, spyatil, perebeysya] (*he has an itch to…, gone mad…, to do without*) were correctly understood by almost all the respondents in the Italian sample. This indicates the wide use of spoken language in the Italian participants.

By contrast, the children in the Israeli sample demonstrated less understanding of such stylistic subtleties.

Motivation for learning Russian

This study also aimed to assess why children learn Russian outside the school curriculum and to identify the main types of motivation for learning the language.

In determining the nature of motivation, the present study relied on the works of Robert Gardner and Wallace Lambert (e.g., Gardner & Lambert, 1972) that laid the foundations of socio-psychological approaches to the study of motivation. They suggested that one’s attitude towards people who speak the target language and one’s attitude towards the educational process affect one’s motivation when learning a foreign language. Gardner and Lambert proposed the most commonly used framework for understanding the different motivations that language learners typically have. They distinguish two types of language learning motivation: **integrative motivation** and **instrumental motivation**.

Integrative motivation reflects “high drive on the part of the individual to acquire the language of a valued second-language community in order to facilitate communication with that group” (Gardner, Smythe, Clement, & Gliksman, 1976, p. 199). It is associated with components such as “interest in foreign languages”, “desire to learn the target language”, “attitudes toward learning the target language”, “attitudes toward the learning situation”, “desire to interact with the target language community” and “attitudes toward the target language community” (Gardner, 1982). Integratively motivated learners want to learn the language so that they can better understand and get to know the people who speak that language. They are also interested in the culture associated with that language.

This motive is clearly distinct from a second drive, instrumental motivation, where the learner’s interest in learning the foreign language is driven by pragmatic, utilitarian benefits of proficiency in a language, such as fulfillment of an academic requirement, obtaining a certificate, a better job or a higher salary.

Several studies have found that language learners that are motivated integratively are more successful than those that are motivated instrumentally. This is probably due to the greater motivation of integratively motivated learners (Taylor, Meynard, & Rheault, 1977; Ellis, 1997; Crookes & Schmidt, 1991).

Today, many scholars criticize the assumption that these two types of motivation are mutually exclusive (Brown, 2000). In fact, integrative and instrumental motivation can exist at the same time, complementing each other. Motivation may include elements of both types, which highlights the complexity of the problem itself. However, some motives can prevail over others. Despite the obvious disadvantages that any categorization of this type may have, the proposed division enables researchers to assess the general direction.

Although there are more detailed scales of motivation available today, these did not seem appropriate for studying motivational analysis among younger children (under age 12), which is quite a challenging task in itself. Notably, 8 participants in this age group in the Italian sample and the 2 participants in the Israeli sample did not answer the question “Why did you decide to learn Russian?” Many additional respondents needed further discussion to help them articulate the answer to this question.

Integrative motivation prevailed in both samples. This is hardly surprising, since Russian is taught outside the school curriculum, thus eliminating the typical school instrumental motive: to achieve a good grade and to complete the school year.

However, most respondents were too young to explore their future career possibilities. This was even more relevant for the Israeli sample. Israeli teens have years of army service after they finish high school and their career choice is still quite a distant prospect.

It may, however, explain the fact that in the Italian sample, 5 out of 37 respondents demonstrated instrumental motivation (“I will need it in the future”, “I would like to become a journalist”, “I think I can use it in the future”) as compared to not even a single subject in the Israeli sample.

Therefore, as noted above, most of the responses may be categorized as showing integrative motivation. The respondents in the Italian sample, however, point to family reasons: “To speak with my grandmother”, “My parents are Russian”, “My parents insisted”, “My parents made me”, “I have relatives in Russia, I want to go there". They also show a generally positive attitude towards the Russian language and culture: “This is a rich and beautiful language”, “This is an important language”, etc. The Israeli sample was dominated by family reasons: “I have relatives in Russia”, “I have relatives in Israel who do not know Hebrew”.

Conclusions

Despite the fact that the study groups differ from each other, many of their reasons for studying Russian, as well as the challenges they face while learning are similar. In both samples, the respondents’ integrative motivation prevails over their instrumental motivation. The greatest challenge for the respondents in both samples is with the Russian verbal system and particularly with verbs of motion. However, in both samples the respondents demonstrated familiarity with the basic rules of Russian grammar. The best results in both samples were evidenced in respondents who were born and attended school in Russia or in one of the former Soviet republics. Good results were also obtained by children who live in a Russian family and have attended Russian educational networks from an early age. They speak with greater confidence, their speech is more morphologically accurate, and is lexically richer. They hardly resort to code mixing. These findings indicate the need to develop teaching materials that can serve various groups of children of Russian-speaking immigrant families that share a lot in common.

References

Brown, H. D. (2000). *Principles of language learning and teaching* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Caritas e Migrantes (2012). Dossier Statistico Immigrazione 22° Rapporto. http://www.dossierimmigrazione.it/docnews/file/2012_Dossier_Scheda.pdf (in Italian).

Crooke, G., & Schmidt, R. W. (1991). Motivation : Reopening the research agenda. *Language Learning*, *4*(4), 469-512.

Dominguez, L. (2009). Charting the route of bilingual development. *International Journal of Bilingualism*, *13*, 271-287.

Ellis, R. (1997). *The study of second language acquisition*. Oxford, UK: Oxford University Press.

Epstein, A., & Khanin, Z. (2007). Introduction. In A. D. Epstein & V. Z. Khanin (eds.), *Every seventh Israeli: The Jews of the former Soviet Union - patterns of social and cultural integration* (pp. 6-11). Tel Aviv, Ashdod: Bar Ilan University.

Gardner, R. C. (1982). Language attitudes and language learning. In E. Bouchard Ryan & H. Giles (eds.), *Attitudes towards language variation* (pp. 132-147). London, UK: Edward Arnold.

Gardner, R.C., & Lambert, W. E. (1972). *Attitudes and motivation*: *Second language learning*. Rowley, MA: Newbury House.

Gardner, R. C., Smythe, P. C., Clement, R., & Gliksman, L. (1976). Second language acquisition: A social psychological interpretation. *The Canadian Modern Language Review*, *32*, 198-213

Garcia-Call, C., & Nagnuson, K. (1997). The psychological experience of immigration: A development perspective. In A. Booth, A. C. Crouter & N. Landale (eds), *Immigration and the family: Research and policy on US immigrants* (pp. 91-131). Manwah, NJ: Erlbaum.

Genesee, F., & Nicoladis, E. (2005). *Bilingual first language acquisition*, HDBK BFLA, Online: www.psych.mcgill.ca/perpg/fac/genesee/HDBK BFLA FINAL.pdf

Kagan, O., & Dillon, K. A. (2001). New perspective on teaching Russian: Focus on the heritage learner. *Slavic and East European Journal*, *45*, 507-518.

Kopeliovich, S. (2011). How long is the 'Russian Street' in Israel? Prospects of maintaining the Russian language. *Israel Affairs*, *17*(1), 108-124.

Laitin, D. D. (1998). *Identity in formation: The Russian-speaking populations in the near abroad*. Ithaca, NY: Cornell University Press.

Michaeli, S. (2007). Preface. In A. D. Epstein & V. Z. Khanin (eds.), *Every Seventh Israeli: Patterns of social and cultural integration of the Russian speaking immigrants* (pp. 6-11). Jerusalem, Ramat Gan: Bar Ilan University.

Montrul, S. (2009). Knowledge of tense-aspect and mood in Spanish heritage speakers . *International Journal of Bilingualism*, *13* (2), 239-269.

Niznik, M. (2005). Searching for a new identity: The acculturation of Russian- born adolescents in Israel. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp.1703-1721). Somerville, MA: Cascadilla Press.

Orfitelli, R., & M. Polinsky (2013). Is it processing all the way down? *Linguistic Approaches to Bilingualism 3*(3), 335-340.

Paradis, M. (2005). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.

Paradis, J., & Navarro, S. (2003). Subject realization and crosslinguistic interference in bilingual acquisition of Spanish and English: What is the role of the input? *Child Language*, *30*, 371-393.

Perotto, M. (2009). *Lingua e identità dell'emigrazione russofona in Italia.* Napoli: Liguori (in Italian).

Perotto , M. (2010). Migranty iz postsovetskikh stran v Italii (rezul'taty empiricheskogo sotsiolingvisticheskogo issledovaniya . *Diaspory*, *1*(1), 82-100 (in Russian)

Perotto, M. (2013). Vital'nost' russkogo yazyka u vtorogo pokoleniya russkoyazychnykh migrantov v Italii. In M. Garzaniti, A. Alberti, M. Perotto, B. Sulpasso (eds.), *Contributi italiani al XV Congresso degli Slavisti* (pp. 229-247). Firenze: Firenze University Press (in Russian).

Polinsky, M. (2006). Incomplete acquisition: American Russian . *Journal of Slavic Linguistics*,*14*, 191-262.
Polinsky, M. (2008). Heritage language narratives. In D. M. Brinton, O. Kagan & S. Bauckus (eds.), *Heritage language education: A new field emerging* (pp. 149-164). New York, NY: Routledge.
Polinsky, M., & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom . *Language and Linguistics Compass*, *1*(5), 368-395.
Polinskaya, M. S. (2010). Russkiy yazyk pervogo i vtorogo pokoleniya emigrantov, zhivushchikh v SSHA. *Slavica Helsingiensia*, *40*, 336-352 (in Russian).
Scandura, C. (1995). L'emigrazione russa in Italia: 1917-1940. *Europa orientalis*, *14*(2), 341-366 (in Italian).
Taylor, D. M., Meynard, R., & Rheault, E. (1977). Threat to ethnic identity and second-language learning In H. Giles (ed.), *Language, ethnicity and intergroup relations* (pp. 99-118). London, UK: Academic Press.
Valdés, G. (2000). The teaching of heritage languages: An introduction for Slavic-teaching professionals. In O. Kagan & B. Rifkin (eds.), *The learning and teaching of Slavic languages and cultures* (pp. 375-403). Bloomington, IN: Slavica.
Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. Peyton, J. Ranard & S. McGinnis (eds.), *Heritage languages in America: Preserving a national resource* (pp. 37-80). McHenry, IL: The Center for Applied Linguistics and Delta Systems.
Wheatcroft, G. (2008). Review: 'Jabotinsky, Weizmann, and the roots of the most contentious communal struggle on earth today' .The *Times* Magazine, 2/20/2008, UK.

Tag questions used by Turkish-Danish bilinguals: A developmental profile

F. Hülya Özcan
fozcan@anadolu.edu.tr

Anadolu University

Abstract: Tag questions have attracted attention from both the syntactic and pragmatic points of view. In the acquisition process, children master not only the syntactic structure of the tag questions but the functions conveyed, as well. The acquisition process, therefore, is affected by the syntactic structure of a particular language, which, consequently, may have an impact on the acquisition process of bilinguals. For example, in Turkish, tag questions are formed by adding either `değil mi? (isn't it?)' or 'öyle mi? (is that so?)', both of which are tagged to affirmative and negative predicates either. verbal or nominal (Göksel & Kerslake, 2005, p. 289-290). On the other hand, in Danish, there are three basic types of tags, which are adverbial tags, sentential tags and tags of the wh-type (Heinemann, 2010, p. 2707). The fact that the bilingual child will choose the language requiring less effort (Mithun, 2012) and the syntactic differences between Turkish and Danish in terms tag questions have motivated this study. The aim of this study, therefore, is to study the use of tag questions in the conversations of bilingual Turkish-Danish speaking children in order to draw a developmental profile in the use of tag questions in the bilingual setting to see if acquiring another language and the transparent nature of Turkish tags makes a difference in the acquisition process. For this purpose, spontaneous conversations of bilingual Turkish-Danish grade school students from Grade 1 through Grade 9 have been studied. For the present study, all grades from the 1st to 8th grade were included in order to be able to see the developmental profile. In each grade, subgroups (a boys-only group and a girls-only group) were included in the analysis in order to reveal any possible gender differences in the data. Therefore, we have 10 groups, 34 participants, in total. The findings showed that there are linguistic, cultural and social rules affecting tag use.

Keywords: tag questions, bilingual conversational setting, Turkish-Danish bilinguals

Introduction

Tag questions, being no different than any other structure in any particular language, do not receive much attention in speech. In fact, tag questions are within our life so much that we are not even aware that we use them and why we use them. The first image of tag questions was myself trying to fill in the blanks with the correct auxiliary in English. This kind of activity put tag questions in the pages of grammar books. I was not, then, aware of the fact that we had such a structure in Turkish. It was only in 1980s the first time I heard that tag questions have a meaning and function and that they were associated with the women. Then, I revived the memory of my mother saying:

(1a) Shall I get the dinner ready?
(1b) I'll get the dinner ready, shall I?

I thought she sometimes used interrogatives and tag questions at other times to express the same event because mom was a woman. Then, in the UK, I frequently heard "you know what I mean, don't you?" from both males and females in daily speech, I once again realized that my mother didn't express the same action with a tag question not just because she was a woman and that there must be more into tag questions than being only a female language. While expressing the same action with two different question types, my mother was definitely trying to get a message across, which only dawned on me now, more than 30 years later.

This intellectual journey I have experienced is not just the result of my own academic growth but also of the change of views on tag questions.

Let's first look at how tag questions are defined. It is possible to find definitions of tag questions in different types of books. Dictionary definitions regard tag questions from a syntactic point of view:

> "A question converted from a statement by an appended interrogative formula, e.g., *it's nice out, isn't it?"* (Oxford Dictionary)

> "a reduced form such as *will you*? as in *You will be there, will you*? or *hasn't she*? as in *She has arrived, hasn't she*?" (Matthews, 1997, p. 371)

Linguistic definitions regard tag questions from both syntactic and pragmatic points of view, and therefore, we get the following definitions. Tag questions are:

> "*a midway between an outright statement and a yes-no question: it is less assertive than the former but more confident than the latter*" (Lakoff, 1973, p. 54)

> "*used when the speaker is stating a claim, but lacks full confidence in the truth of that claim*" http://itre.cis.upenn.edu/~myl/languagelog/archives/000873.html

These definitions tell us that a statement in speech indicates that the speaker is confident in the knowledge but a question, on the other hand, requires some kind of knowledge which the speaker does not know leaving a room for the hearer to comment and to share the commitment. These characteristics make tag questions polite expressions. This is why, tag questions are associated with women's speech since women's speech is more polite than men's speech (Lakoff, 1973, p. 56).

Speech functions of tag questions have been defined as well. Tag questions:

> "*express unassertive, polite comments*" (Lakoff, 1973, 1975)

> "*are suggestions for which the speaker seeks confirmation, which can be either rejected or confirmed*", "*call for an agreement or disagreement*" (Hudson, 1975)

> "*are conversational organizers to regulate turns and to convey certain social and psychological meanings*" (Levinson, 1970)

> "*indicate powerless identity in interaction because there is no challenge to the authority*" (Ochs, 1992)

The pragmatic force (illocution) of tag questions leads to pragmatic classification of tag questions according to the function for which it is used by the speaker. The fact that, by using a tag question, a speaker becomes unassertive or polite covering an hostile accusation, or seeks confirmation, organizes conversational turns, expresses powerless identity in interaction, seeks attention or signals an end of a discussion, tells us that tag questions do not show lack of self-confidence nor are they specific to women. Tag questions have functions in discourse; pragmatic factors and power may affect the use of them. (Lakoff, 1973; Hudson, 1975; Levinson, 1983; Togeby, 1992; Ochs, 1992; Nåsslin, 1984; Algeo, 1998, 1990, 2006; Holmes, 1995; Kimps, 2007).

Acquisition studies of tag questions focus on the phonological and syntactic productions of tag questions (Brown & Hanlon, 1970; Brown, 1973; de Villiers & de Villiers, 1978; Berninger & Garvey, 1982, Dennis, Sugar, & Whitaker, 1982; Fletcher, 1985; Wells, 1985; Richards, 1990). McTear (1988. P. 87) analyzed tags in terms of pragmatic force and states that tag questions were frequently used to initiate and reinitiate conversations. Children's initiations and reinitiations with tags received responses more than the declaratives. Therefore, children develop awareness of tag questions in the acquisition process with age.

These studies lead to a conclusion that acquisition of syntax and prosody is not enough for the productive use of tag questions and children should be able to grasp the function of tag questions in discourse. The acquisition of tag questions, therefore, requires the acquisition of prosody, syntax, meaning, and function. This complexity of both pragmatic functions and syntactic structures of tag questions may have an impact on the acquisition process of bilinguals. Apart from the linguistic factors, there are sociolinguistic factors such as age, gender and interactional roles and interpersonal relations. Tottie and Hoffman's results (2006, p. 306), indicating that there are nine times more questions in British English than in American English, also brings up the effect of culture/cultural differences in the use of tag questions. If different cultures have different tendency in the use of tag questions, bilingualism may have an effect on it, too.

Based on these issues, this study has been designed to study the use of tag questions in the conversations of Turkish-Danish speaking bilingual children. The aim is to study a developmental profile in the use of tag questions in the bilingual setting to see if acquiring another language in another culture and the transparent nature of Turkish tags makes a difference in the acquisition process. Within this framework, the following research questions will be delineated:

1. Do bilingual Turkish-Danish speaking children use tag questions in Turkish, in Danish or in both?
2. Which form of tag questions do children use in their conversations in both languages?
3. Which functions are expressed by the tag questions in a symmetrical discourse situation, in other words, in a no power situation in which all the speakers are equal?
4. Does language mixing affect use of tag questions?
5. Are there any sociolinguistic factors such as age, gender and interpersonal relations which effect the use tag questions?

As will be explained in the methodology section, conversational samples are taken in a situation where 3 to 4 school children of the same gender, from the same school and school year, and the same socio economic background. They all deal with the same task. In terms of these variables, the situation is expected to be a no power situation and the discourse is symmetrical. One of the participants can be identified as "powerful" when s/he is "institutionally responsible for the conduct of the talk", with greater social power and status in the context of the conversations, such as a doctor vs. patient, teacher vs. student.

Background information on tag questions in Turkish and in Danish

Tag questions in Turkish

In Turkish, there are two forms of tag questions and defined as "a question that is annexed to a statement and is used to seek confirmation of that statement" (Göksel & Kerslake, 2005, p. 289).

1. ***Değil mi***? is added to the end of the sentence, which is a combination of a negative particle with a question suffix. This is the unmarked tag questions form which expresses confirmation.
2. ***Öyle mi***? follows a statement, which is a combination of the demonstrative adverbial and a question suffix. This type of tag questions seeks for assertion for the newly acquired information.

Both forms are tagged to affirmative and negative predicates either verbal or nominal (Göksel & Kerslake, 2005, p. 289-290).

Kornflit (1997, p. 7) names these unmarked tag questions with *değil mi*? as leading Yes/No questions since the structure of the statement to which the adverbial particle is appended defines the answer.

Tag questions in Danish

There are three basic types of tags in Danish:

1. Adverbial tags ***ikke (not), vel (right)***
2. Sentential tags ***du ikke det*** (don't you), ***har du det*** (do you)
3. Wh-type tags ***hvad***? ***eller hvad***?

(Heinemann, 2010, p. 2707-8)

Adverbial tags are more commonly used than sentential tags. A speaker's choice of either form of tags depends on pragmatic purposes. Tag questions in Danish are not as transparent as Turkish tags as seen in the examples.

The study

Participants

This study is a part of a large-scale project, Køge Project, which was designed to collect spoken language data from bilingual Turkish-Danish grade school students from Grade 1 through Grade 9. The participants, who belong to the second generation of Turkish immigrants, were born and raised in Denmark. For the present study, 22 groups, each group containing 3 or 4 children, were included. All grades from the 1st to 8th grade, except the 5th grade, were included in order to be able to see the deelopmental profile. In each grade, subgroups, a boys-only group and a girls-only group were included in the analysis in order to reveal any possible gender differences in the data. Therefore, we have, 42 girls and 34 boys, 75 participants in total.

Table 1. Functional categories of tag questions

		Holmes	Algeo	Tottie & Hoffman	Mithuan
Epistemic Modal Tags					
1	Tags asking for information -express genuine speaker uncertainty -no presumption for the answer	Epistemic modal	Informational	Informational	
2	Tags seeking confirmation -speaker uncertainty	Epistemic modal	Confirmatory	Confirmatory	
Affective tags -Tags expressing an attitude to an addressee					
1	Tags attempting to involve the addressee	Facilitative	Facilitating	Facilitating	
2	Tags expressing confrontational strategies	Challenging	Peremptory	Peremptory	
3	Tags expressing an insult or provocation		Aggressive	Aggressive	
4	Tags emphasizing what the speaker says; stressing the speaker's point of view		Punctuational	Attitudinal	
5	Tags which softens the negatively affective utterances	Softening			
6	tags encouraging the addressee to agree with different attitudes expressed by the speaker	conducive			
5	Tags expressing an appeal for acknowledgement of shared knowledge to establish a bond between the speakers				Shared knowledge, experience & values
6	Tags which appear in proposal for joint actions				Joint plans

Data collection

The data consists of the conversations of the bilingual children with peers in groups. Each group was given a task in which they had to negotiate and fulfill a kind of problem-solving task. In grade one and three, they were given a cardboard picture of a house and furniture catalogues and were asked to furnish a house. The participants were encouraged to discuss and decide together. In grade five, the students were given travel catalogues to use and a world map. They were asked to identify and cut the pictures of the places in the travel catalogue and glue them on the correct place on the map. In grade 7 and 8, they were asked to create a cartoon strip or a collage with pictures from teenage magazines.

Data Analysis

The conversations were transcribed in compliance with the CHILDES conventions (MacWhinney, 1995). The data were analysed both qualitatively and quantitatively. Each occurrence of tag questions were counted and then frequencies were calculated. The functions are assigned according to the functional categories defined by Holmes (1983, 1995), Algeo (1990), Tottie and Hoffman (2006), and Mithuan (2012). Table 1 illustrates the functions of tag questions defined by different studies. To determine these categories, occurrences of tag questions with the previous and following sentences were given to 5 native speaker raters to categorize the tags according to the functions they were used. These raters, all linguists, categorized each occurrence of a tag question. As a researcher, I wasn't involved in categorizing. Afterwards, the categories which were given by the raters were evaluated; when the raters did not agree on a category, that is, when there are a variety of categories identified, I listened to the conversations and read the transcripts and gave the final verdict on the function. Then, frequencies of the occurrences and types were given.

Results

We first look at the frequency of tag questions used by all the participants across age groups.

Table 2. Distribution of tag questions across age groups

Grade	# of turns	# of utterances	# of questions	Tag questions
1. grade	1331	1751	65	22 **(7%)**
2.grade	1571	2057	237	43 **(18%)**
3.grade	1916	2028	310	15 **(5%)**
4.grade	2159	2240	214	25 **(12%)**
6.grade	1846	1872	110	18 **(16%)**
7.grade	1702	1754	163	15 **(1%)**
8.grade	164	176	31	5 **(16%)**

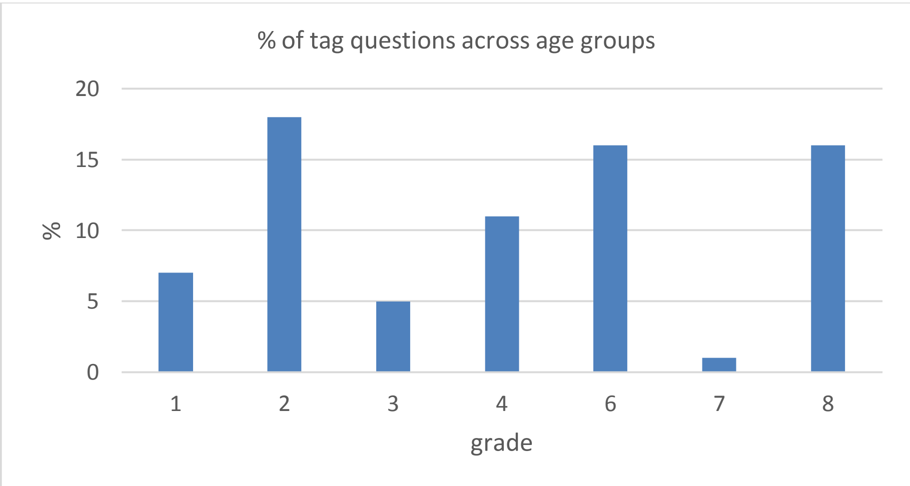


Figure 1. % of tag questions across age groups

As illustrated in Table 1, tag questions are used at every age group although fewer in number. 2nd graders are the most talkative group with 2057 utterances and with the highest frequency of tag questions. Although the number of the questions asked by 3rd graders is the highest of them all, tag questions do not take much of these questions. 7th graders do not produce tag questions although 9% of their utterances are questions.

Then, we looked into whether gender makes a difference in terms of the total number of tag questions.

Table 3. Distribution of tag questions across gender

	Total # utterances		Total # questions		Tag questions	
	Girls	Boys	Girls	Boys	Girls	Boys
1.grade	650	1101	129	165	12% (15)	4% (7)
2.grade	1029	1028	120	117	25% (30)	11% (13)
3.grade	956	1072	167	143	4% (7)	6% (8)
4.grade	1060	1180	117	97	10% (12)	13% (13)
6.grade	767	1105	44	66	32% (14)	6% (4)
7.grade	807	947	79	93	13% (9)	6% (6)
8.grade	176	-	31	-	16% (5)	-

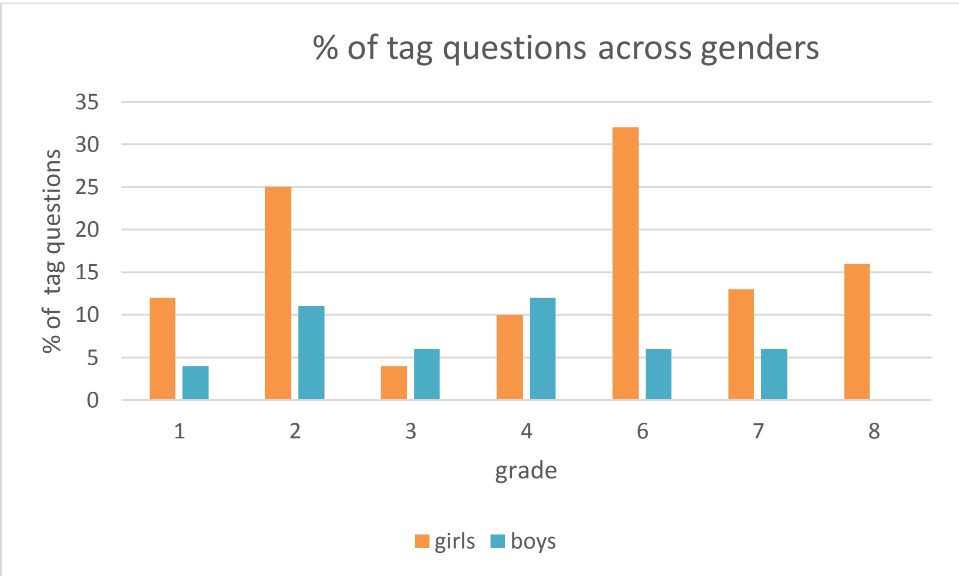


Figure 2. Frequency of tag question across genders

Bilingual boys have more utterances than girls. The total number of the questions asked by boys are more than the girls at 1st, 6th and 7th grades. Only in the 2nd grade, both the total number of utterances and the questions become equal between two genders. Although boys speak more and ask more question than girls, tag questions are more frequent in girls' speech than boys. In the 3rd and 4th grades, boys have more tag questions.

We then looked at whether tag questions are used in Turkish or in Danish or in both. As mentioned earlier, tag questions in Turkish is asked by adding an adverbial particle "değil mi" or "öyle mi" to any sentence without considering negative or positive polarity. In Danish, on the other hand, there are 3 kinds of tags, adverbial, sentential and polarity is important when an adverbial particle is used. Besides, polarity is important when using adverbial tags. This situation makes Turkish tags easy to access.

Table 4. Distribution of tag questions in Turkish and in Danish

	Total # tag questions		Danish	
	Turkish	Danish	Girls	Boys
1.grade	22	-	-	-
2.grade	41	2	-	2
3.grade	14	1	-	1
4.grade	18	7	3	4
6.grade	13	5	1	4
7.grade	4	13	8	5
8.grade	-	5	6	-

Danish tag questions, although few in number, start in the 2nd grade with boys. Girls start producing Danish tags at a later stage in the 4th grade. Yet, until the 7th grade, Turkish tag questions outnumber the Danish tag questions. In the 7th grade, Danish tag questions increase with the increasing number of tags used by girls.

Types of tag questions

Tag questions in Turkish were expressed by the adverbial "değil mi" at all ages by both genders. There is only one instance of "öyle mi?" among 112 tag questions.

Tag questions asked in Danish are adverbial tag questions which were expressed by adverbials '***ikke***' and '***vel***' (82%). Sentential tag questions form the 9 % of the tag questions and other 9% of the tags were expressed by wh-tags.

Functions of Tag questions

Functions are assigned to the questions using function categories defined by Holmes (1983, 1995), Algeo (1990), Tottie and Hoffman (2006), and Mithuan (2012) since there are no functional categories defined for Turkish tag questions reported by Göksel and Kerslake (2005) and Konflit, 1997), which are seeking confirmation and seeking assertion. Findings on functions will be presented in 3 stages; functional categories across age groups, across gender and across language.

Functional categories across age groups

Table 5 illustrates functional categories as epistemic modal and affective tags.

Table 5. Frequency of epistemic modal tags and affective tags

	1st gr.	2nd gr.	3rd gr.	4th gr.	6th gr.	7th gr.	8th gr.
Epistemic modal tags	55%	49%	33%	52%	50%.	33%	60%
Affective tags	45%	51%	67%	48%	50%	67%	40%

The distribution of the epistemic modal and affective tags do not show a consistent developmental trend. So there must be other factors playing a role in choosing a function to express with a tag. We, therefore, look at the functions girls and boys express with tag questions.

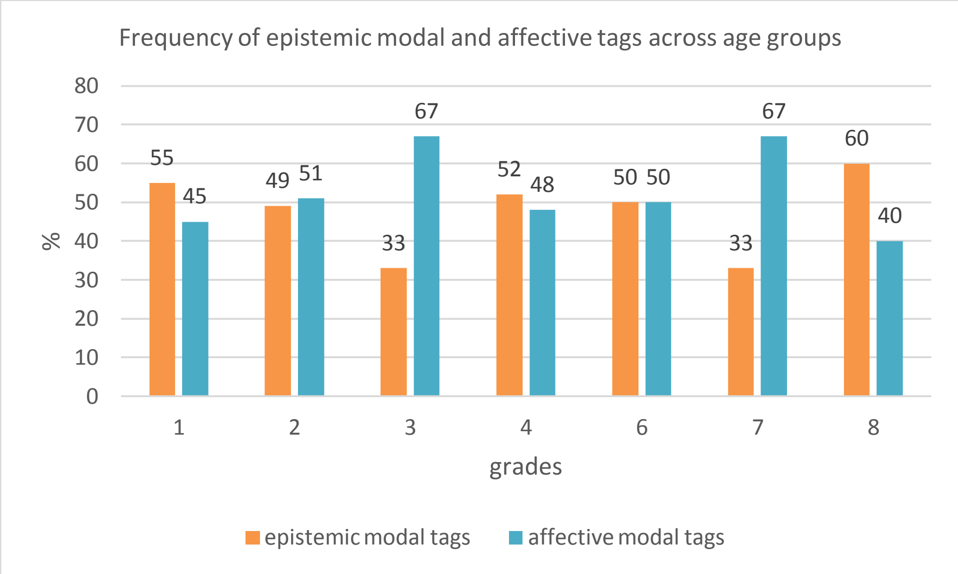


Figure 3. Frequency of epistemic modal tags and affective tags

We, then, look at whether epistemic and affective tags differ across genders first in total then across age groups.

Table 6. Distribution of functions of tag questions across genders

	Epistemic Modal Tags	Affective tags	total
Girls	40 (43%)	52 (57%)	92
Boys	27 (53%)	24 (47%)	51

Figure 4 illustrates the distribution of these functional types of tag questions across genders in general and Figure 5 illustrates these functions across age groups and genders.

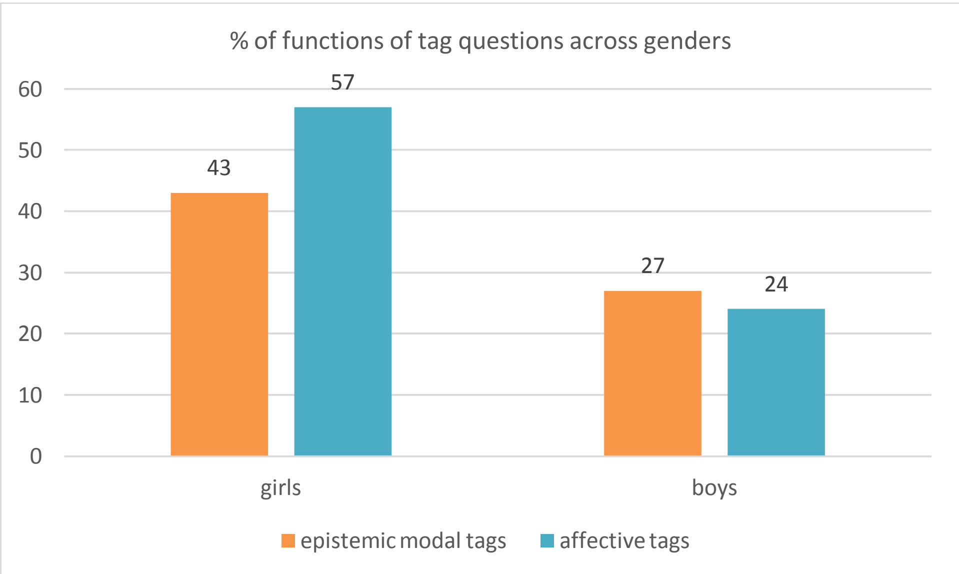


Figure 4. Distribution of epistemic and affective tag across genders

Gender makes a difference in terms of the functions of the tag questions. Boys prefer epistemic modal tags, in other words, when they ask for information and seek confirmation. Girls, on the other hand, apart from asking for information and seeking confirmation, use tag questions to express an attitude to an addressee more.

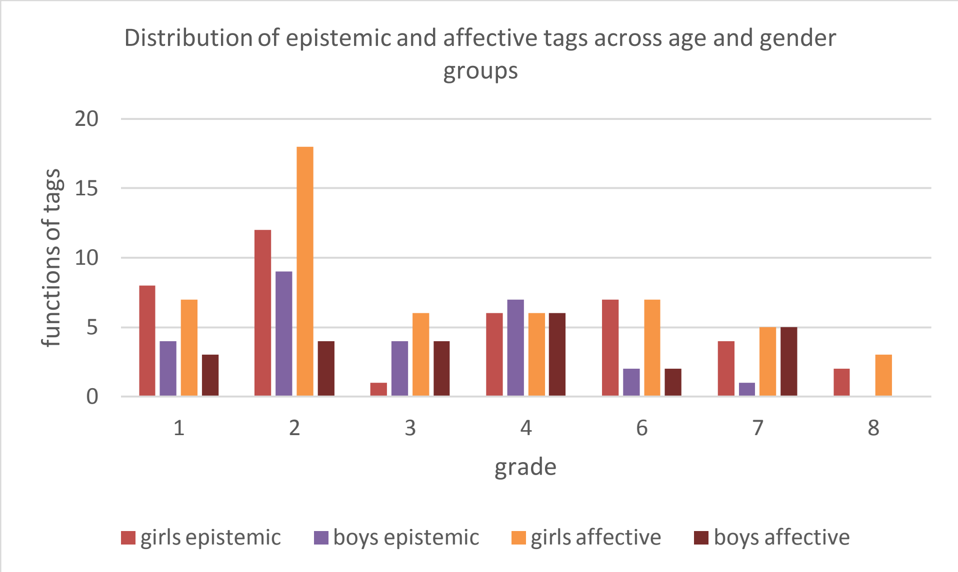


Figure 5. Distribution of epistemic and affective tags across age and gender groups

When we further analyze the affective functions, we see that the same functions are expressed both by the girls and the boys except two exceptions. Boys do not use tag questions to express shared knowledge and experience (Table 6).

Table 7. Affective tag functions across genders

Girls	Boys
Facilitative tags-tags attempting to involve the addressee	Facilitative tags-tags attempting to involve the addressee
Challenging tags-tags expressing confrontational strategies	Challenging tags-tags expressing confrontational strategies
Attitudinal tags-tags stressing the speaker's point of view	Attitudinal tags-tags stressing the speaker's point of view
Tags which appear in proposals for joint actions	Tags which appear in proposals for joint actions
Tags which softens the negatively affective utterances	Tags which softens the negatively affective utterances
Conducive-knowledgeable/evaluative/ experienced/authoritative	Conducive-informative/critical/ evaluative/knowledgeable/
Tags expressing shared knowledge, experience and values	
Girls	boys
Other-asking for approval -asking for agreement	Other-asking for agreement -asking for approval -self assurance

There are other categories added by the raters. They are asking for agreement and asking for approval, as illustrated in examples (1) to (4).

(1) *HAC: ben hepsi altina alta yazacagim kimin olursa buraya yazsin. (*I will write all of them down at the bottom (of the paper). Whoever has anything to write, they can write here?*)

(2) *HAV: olsun kimin olursa yazarsın değil mi Hacer kız. (*Fine. You will write whoever it belongs to, won't you HAC, girl*.)

Second speaker is expanding the first speaker's statement and seeking agreement to what she has just said using a tag question.

Another category identified by the raters is asking for approval.

(3) YUS: han sagde otte lad os se om der er et tal ell er et bogstav kom så bogstav nej tre hundrede tre hundrede kroner nej det vil vi ikke have okay videre hvad var by hvad var hovedby hvad øh. (*He said the eight. Let's see, there is a letter or number. Nej? 300 kroner. No we don't want that! What was town? Right?*)

(4) PEM: var der medlem der så var vi derovre der kommer rigtig mange tyrkere ikke.
(*We were there as a member there as we were over there there comes a great many Turks, don't they? (right?*))

(5) *PEM: synes du de der de skulle ikke vaere der nær du ser .
*HAT: ja det synes jeg også synes du ikke .
(*Yes it seems I also, don't you think?)*

Another function, which were used only by boys a number of times, is expressing self assurance (5) and (6).

(6) YUS: Mehmet dur jeg har en god en Mehmet saadan der armene skal ogsaa vaere fede ikke. (*I have a good one Mehmet such that the arms shall also be fat, right?*)

(7) YUS: jeg tager lige mikrofonen af okay det goer ikke noget vel.
(*It makes nothing, does it?)*

Table 8. Similarities and differences between monolinguals and bilinguals

Turkish speaking monolinguals (Özcan, 2014)	Turkish-Danish speaking bilinguals
More questions, more tag questions (1823 and 143 respectively)	Less questions, the same number of tag questions (1589 and 138 respectively)
The number of tag questions change according to gender. Girls ask more tag questions than boys.	Girls ask more tag questions. The number of tag questions change according to gender.
Boys never outperform girls in terms of tag questions.	In the 3rd and 4th grade, around the ages of 9-10, boys ask more tag questions than girls.
Epistemic modal tags are used more than the affective modal tags at all grades except for the 1st grade.	There is no explanatory trend in terms of the increasing age.
Girls' epistemic modals increase at 3rd and 5th grades (9-11 years)	Girls' epistemic modals are higher at the 1st and 3rd grade (7-9 years) and then decrease.
Boys use more epistemic modals and less affective modals	Boys use more epistemic modals and less affective modals
Boys express fewer functions through tags than girls.	Girls and boys express the same functions through tags. Boys are as expressive as girls.
Power relations affect both the use and the function of tag questions.	There is no observable power relations.

Monolingual Turkish-speaking and bilingual Turkish-Danish speaking children display similarities and differences while using tag questions. Monolinguals ask a lot of questions, among which are tag questions. However, bilingual children do not ask many questions but the number of tag questions is as high as in monolinguals. Girls ask more tag questions than boys in both groups. This may be related to the number of the questions asked since girls asked more questions. Bilingual boys show a similar tendency with girls, while monolingual boys do not perform the same way as the girls do. Monolingual children start using different functions with increasing age. Monolinguals establish power relations within the group although they are expected to be equal, and these power relations affect the use tag questions as well as the functions they are used for.

Discussion

Tags have plural functions and we can see that these plural functions are performed by school children regardless of the culture they are in and of the language they are speaking. Bilinguals' use of tag questions in both Turkish and Danish shows us that the function to be expressed is important, not the language used at the moment of speaking. Tag questions are not deferential and specific to women. The girls' use of more tag questions depends on who holds the power in the group in the monolingual group (Holmes, 1984, 1985, 1995). In the bilinguals, boys sometimes outperform the girls in using tag questions. Not any of the functions of tag questions subordinate women; boys and girls use tag questions alike.

Not any of the functions of tag questions works to subordinate women; that is, tags are employed by women and men alike. Variety in the functions of tag questions do not show any consistent profile with the bilinguals but school development seems to have an effect on the type of tag questions in the monolingual group. More reading, more listening or participating in social and school activities may raise awareness in pragmatic functions of the language and may affect language performance as well.

Conclusion

Tag questions are affected by linguistic, sociolinguistic, and cultural factors, all of which are interrelated at the same time. Tags are used to convey a certain attitude and are also adjusted according to the response expected to be given. Bilinguals' use of tags in Danish at later ages has led us to conclude that the pragmatic force of tag questions may require acquisition of enough linguistic competence. On the other hand, tags may be specific to certain groups and fulfil a social role within a community and speakers adopt a role considering the cultural factors. In order to see whether and how culture and, as a consequence, social roles affect the use of tag questions, more detailed analysis in the mixed gender groups is needed.

References

Algeo, J. (1988). The tag question in British English: it's different i'n'it? *English worldwide*, *9*, 171-191.

Algeo, J. (1990). It's a myth, innit? Politeness and the English tag question. In C. Ricks & L. Michaels (eds.) *The state of the language* (pp. 443-450). Berkeley: University of California.

Algeo, J. (2006). *British or American English?* Cambridge, UK: Cambridge University Press.

Berninger, G., & Garvey, C. (1982). Tag construction: structure and function in child speech. *Journal of Child Language*, *9*, 151-158.

Brown, R. (1973). *A first language: The early stages*. London: Allen & Unwin.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (ed.), *Cognition and the development of language* (pp. 11-54). New York: Wiley.

Dennis, M., Sugar, J., & Whitaker, H. A. (1982). The acquisition of tag questions. *Child Development*, *53*(5) 1254-1257.

De Villiers, J., & De Villiers, P. (1978). *Language acquisition*. Cambridge, MA: Harvard University Press.

Fletcher, P. (1985). *A child's learning English*. Oxford, UK: Blackwell.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London, UK: Routledge.

Heinemann, T. (2010). The question-response in Danish. *Journal of Pragmatics*, *42*, 2703-2725.
Holmes, J. (1983). The functions of tag questions. *English Language Research of Journal*, *3*, 40-65.
Holmes, J. (1995). *Women, men, and politeness*. New York: Longman. http://itre.cis.upenn.edu/~myl/languagelog/archives/000873.html
Hudson, R. A. (1975). The meaning of questions. *Language*, *51*(1), 1-31.
Kimps, D. (2007). Declarative constant polarity tag questions: A data-driven analysis of their form, meaning and attitudinal uses. *Journal of Pragmatics*, *39*, 270-291.
Kimps, D, Davidse, K., & Cornillie, B. (2014). A speech function analysis of tag questions in British English spontaneous dialogue. *Journal of Pragmatics*, *66,* 64-85.
Kornflit, J. (1997). *Turkish*. London, UK: Routledge.
Lakoff, R. (1973). *Language and woman's place*. New York: Harper and Row.
Levinson, S. C. (1983). *Pragmatics*. Cambridge, UK: Cambridge University Press.
MacWhinney, B. (1995). The CHILDES Project: Tools for analyzing talk (2nd ed.). Hillsdale, NJ: Erlbaum.
Matthews, P. H. (1997). *Concise dictionary of linguistics*. Oxford, UK: Oxford University Press.
McTear, M. (1985). *Children's conversations*. Oxford, UK: Blackwell.
Mithun, M. (2012). Tags: Cross-linguistic diversity and commonality. *Journal of Pragmatics*, *44*, 2165-2182.
Nåsslin, S. (1984). *The English tag question: A study of sentences containing tags of the type ISN'T IT?, IS IT?. Stockholm studies in English 60*. Stokholm: Almqvist and Wiksell.
Özcan, F. H. (2014). Functions of tag questions in the acquisition of Turkish: A developmental profile. Paper presented in the *XVII ICTL*, Rouen: University of Rouen, 3-5 September, 2014.
Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin (eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335-358). Cambridge: Cambridge University Press.
Richards, B. (1990). *Language development and individual differences*. Cambridge: Cambridge University Press.
Tottie, G., & Hoffman, S. (2006). Tag questions in British and American English. *Journal of English Linguistics*, *34*(4), 283-311.
Wells, C. G. (1985). *Language development in pre-school years*. Cambridge, UK: Cambridge University Press.

Three-year-old children acquiring South African English in Cape Town

Michelle Pascoe, Jane Le Roux, Olebeng Mahura, Emily Danvers, Aimée de Jager, Natania Esterhuizen, Chané Naidoo, Juliette Reynders, Savannah Senior, Amy van der Merwe
michelle.pascoe@uct.ac.za

Division of Communication Sciences and Disorders, University of Cape Town

Abstract: *Background*. South Africa has eleven official languages, yet speech acquisition in this context of rich linguistic diversity has not been well-researched. Studies have detailed development of isiXhosa phonology, but little work has focused on the typical acquisition of multiple languages in this context, and speech development in young children acquiring South African English has yet to be detailed. *Aims and Objectives*. We describe speech development in 3-year-old children acquiring South African English in Cape Town. The study objectives were (a) to describe the children's consonant and vowel inventories, and phonological processes by language background (monolingual, bilingual and trilingual); (b) to determine the prevalence of speech disorders in our sample of the population, and (c) To describe the diagnostic category of participants with speech disorders with reference to Dodd's (2005) diagnostic framework. *Participants*. One hundred and fifty children between the ages of 3;0–3;11 acquiring South African English were assessed. They were selected from a range of areas in Cape Town representing a variety of socio-economic backgrounds. Children were excluded from the study where languages other than English, Afrikaans or isiXhosa (the three main languages spoken in the region) were spoken. *Method*. Participants were individually assessed using the Articulation, Phonology and Inconsistency subtests of the Diagnostic Evaluation of Articulation and Phonology (DEAP, Dodd, Hua, Crosbie, Holm and Ozanne, 2002). Results. Children from the different language backgrounds (monolingual, bilingual and trilingual) obtained percentage consonant correct (PCC) scores that were not statistically different. Participants were still acquiring ten consonants, which fits with the literature on English consonantal acquisition in other countries. Percentage vowels correct (PVC) was found to be a more complex, less reliable index of speech accuracy. While adaptations were made for dialectal differences, bilingual isiXhosa children achieved significantly lower PVC scores than other groups, suggesting that our scoring adaptations were insufficient. Cluster reduction and backing were particularly frequent among the bilingual English/isiXhosa children showing the influence of L1. Approximately 7% of children in the sample were found to have speech difficulties, and of these children the greatest proportion exhibited phonological delays. *Discussion*. Results are discussed in relation to normative data collected for different populations. Three-year-olds acquiring South African English in Cape Town are broadly similar to three-year-olds acquiring English in other contexts, although L1 influences are clear and must be taken into account. Clinically, there is a lack of information about typical speech sound development in South African children. This information is urgently needed by Speech and Language Therapists practising in the region to assist in the identification and management of children with speech difficulties. Theoretically the study contributes to knowledge of typical speech development in multilingual contexts.

Keywords: South African English, isiXhosa, Afrikaans, acquisition, speech delay

Introduction

Linguistic and cultural diversity are enshrined in South Africa's progressive constitution, which recognizes eleven official languages. These include the West Germanic languages of English and Afrikaans, and nine indigenous languages from the Bantu language family, namely isiZulu, isiXhosa, Sepedi, Setswana, Sesotho, Xitsonga, siSwati, Tshivenda, and isiNdebele. Multilingualism is common with the exact combinations of languages and dialects spoken varying from region to region. In Cape Town, the main languages spoken are Afrikaans (spoken by 49.7% of the population), isiXhosa (spoken by 24.7%), and English (spoken by 20.3%) (Statistics South Africa, 2011). The constitution states that all official languages must enjoy parity of esteem and be treated equally. However, English, the fourth most spoken language in the country, is widely perceived as the language of power and social mobility. Most urban people understand English and it is widely used in the media and government.

Speech acquisition in this context of rich linguistic diversity has not been well-researched. Some of the Bantu languages now have small datasets and a growing collection of published papers detailing typical development of children's speech in these languages. For example, isiXhosa phonological development has been documented in children between the ages of two and six, by authors such as Gxilishe (2004) and Maphalala, Pascoe and Smouse (2014). Much of this work follows an assumption that the children are monolingual speakers, or if they are bi- or multilingual, the exposure and abilities in all their languages are not well documented. Little work has focused on the typical acquisition of multiple phonologies at the same time. In contrast, adult speakers of South African English, Afrikaans and isiXhosa have been fairly well-studied in the sociolinguistic literature, and dialectal features of their speech have been documented (Mesthrie, Schneider, Burridge, Kortmann, & Upton, 2004; Van Rooy, 2008). We know a considerable amount about the target phonologies of these languages, the factors influencing speech patterns in these dialects and the way in which these (and other local languages) interact. However, there is a lack of information about speech development in typical South African children, as well as the nature and prevalence of speech difficulties that may occur.

Around the world, speech sound difficulties make up a large proportion of speech-language therapists (SLT's) caseloads. They are known to affect more children than any other developmental communication disorder and, if left unmanaged, can result in long-term academic, literacy and social difficulties (Broomfield & Dodd, 2004; Fox & Dodd, 2001). According to the International Classification of Functioning, Disability and Health (World Health Organization, 2007), activity and participation are profoundly impacted due to speech impairments. Although the prevalence of speech difficulties in South Africa has not been documented, in the United States it is estimated at 7.5% of children between the ages of 3 and 11 years (Ruscello, 2008) and in the UK, approximately 48 000 children are referred for speech difficulties each year (Broomfield & Dodd, 2004).

For SLTs working in South Africa, the lack of knowledge about typical speech development presents a real challenge. It is the role of the clinician to identify children with speech sound difficulties and assist them and their families. However, at the same time, SLTs have to ensure that they do not 'pathologise' children who are in fact typical. Such decisions necessarily rely on normative data, collected from the same population as that of the child in question. In South Africa, there are few speech assessments appropriate for use with local indigenous languages and English assessments used will have been normed on different populations, e.g., monolingual British children. The Diagnostic Evaluation of Articulation and Phonology (DEAP) (Dodd et al., 2002), is a widely-used speech assessment for children aged 3;0 to 6;11 years who are acquiring English. It was standardised in 2001/2002, using a sample of 144 monolingual Australian children and 584 monolingual British children. Eighty-three bilingual Punjabi/Mirpuri/Urdu-English speaking children, aged 3;9-6;11, also participated in the standardisation project.

The DEAP aims to classify children into the categories of Dodd's (2005) differential diagnostic framework. As speech difficulties are heterogenous in nature, a theoretical framework is useful for understanding and classifying them. The four main categories of speech sound difficulties described in this framework are as follows:

- Articulation difficulties: Phonetic difficulties linked to one or a small set of speech sounds.
- Delayed speech: Speech development follows the typical course but the child's speech resembles that of a younger child.
- Speech disorder: Speech does not follow the typical course of development, e.g., unusual processes are used.
- Inconsistent speech disorder: Speech does not follow the typical course and in particular is characterised by unstable error forms.

Evidence indicates that children acquiring a range of different languages can be described using this framework and it appears to be valid for both monolingual and multilingual groups (Holm, Dodd, Stow, & Pert, 1999). A comparison of English, Cantonese, Putonghua, Spanish, and German children showed results were similar across these languages (Waring & Knight, 2013). Across languages the distribution of diagnostic categories is fairly constant and is estimated at approximately 50% of children with phonological delay; 25% are considered to be in the consistent atypical disorder category; and the remaining children split equally between articulation disorder (12.5%) and inconsistent phonological disorder (12.5%) (Broomfield & Dodd, 2004).

Of course, the challenge of carrying out reliable, valid, culturally appropriate assessment is not limited to South Africa, but is a worldwide issue (De Lamo White & Jin, 2011). In response to this need, the International Expert Panel on Multilingual Children's Speech produced a position paper which calls for us to:

> "… generate and share knowledge, resources, and evidence to facilitate the understanding of cultural and linguistic diversity that will support multilingual children's speech acquisition… acknowledge and respect [children's] existing competencies, cultural heritage, and histories… assessment and intervention should be based on the best available evidence." (International Expert Panel on Multilingual Children's Speech, 2012, p. 2).

The current project aimed to respond to this call, as well as the needs of SLTs working in South Africa. As clinicians, we strive to identify children with speech sound difficulties as early as possible, to offer evidence-based treatment and prevent negative sequelae, such as academic and psychosocial problems which have been shown to be linked to speech difficulties. Thus we aimed to describe the typical development of South African English speech by three-year-old children in Cape Town who have a range of different language backgrounds.

Method

Aims and objectives

The study aim was to describe the development of speech in 3-year-old children acquiring South African English in Cape Town. More specifically, the objectives were:

1) To describe the children's consonant and vowel inventories, and phonological processes by language background (monolingual, bilingual and trilingual);
2) To determine the prevalence of speech disorders in our sample of the population;
3) To describe the diagnostic category of participants with speech disorders with reference to Dodd's (2005) diagnostic framework.

Participants

The participants were 150 children aged between 3;0 and 3;11 years acquiring South African English. Children were recruited from a range of different areas in Cape Town representing a variety of socio-economic backgrounds. Children were excluded from the study where languages other than English, Afrikaans or isiXhosa were spoken, since these other languages may have impacted the results. Children were also excluded if information about their language background could not be obtained. Tables 1 and 2 provide an overview of the participants.

Table 1. Participants by language background

Language Background	Monolingual English	Bilingual Afrikaans	Bilingual isiXhosa	Trilingual
n (%)	69 (46%)	48 (32%)	25 (16.67%)	8 (5.33%)

Table 2. Number of participants by age and gender

	Younger group 3;0–3;5	**Older group 3;6–3;11**	**Total (n)**
Male	22	42	64
Female	35	51	86
Total (n)	57	93	150

Procedure

Approval was obtained from the University Of Cape Town's Human Research Ethics committee and written informed consent was obtained from the parents of the participants. In addition, the children gave their verbal assent to participate in the project. Children were individually assessed in quiet rooms at creches/schools. The Articulation, Phonology and Inconsistency subtests of the Diagnostic Evaluation of Articulation and Phonology (DEAP, Dodd et al., 2002) were administered. These subtests required children to name pictures and produce short strings of connected speech. For all picture-naming tasks, a hierarchy of cues was used when participants had difficulty producing an appropriate response. Semantic cues (e.g., "You use it to tell the time") were provided first, followed by forced choice within category (e.g., "Is it a watch or a phone?"), then imitation cues. Short breaks were taken to ensure maximum concentration levels of the participants. Two student researchers engaged with each participant, and real-time transcription took place, as well as audio-recordings, which could be used for later re-transcription. A short questionnaire was given to parents to obtain information about each child's language abilities and exposure to the different languages, as well as general developmental information.

Dialectal considerations for scoring

Normative data associated with the DEAP derives from British and Australian children, and Dodd, Holm, Hua, and Crosbie (2003) have cautioned that dialectal variation should be considered when interpreting the results of a speech sample. We consulted literature regarding adult production of varieties of South African English (e.g., Bowerman, 2008; De Klerk & Gough, 2004; Lass, 2004; Mesthrie et al., 2004; Van Rooy, 2008). Features that we expected to find in the children in our sample based on the typical speech production of adult models included:

- Alveolar trill /r/ and post-vocalic /r/ (Afrikaans and isiXhosa children)
- Word final devoicing (e.g., 'dok' for *dog*) (Afrikaans and isiXhosa children)
- /θ/ produced as /f/ word finally (e.g., 'teef' for *teeth*) (Afrikaans)
- Reduced contrasts between long and short vowels (*seat* / *sit*); fewer central vowels and avoidance of schwa (isiXhosa children)
- Vowel raising (English L1 speakers), e.g., 'yis' for *yes*; 'Efrica' for *Africa*.
- A 'kit / bit' split, i.e. the words *kit* [kɪt] and *bit* [bət] do not rhyme. [ɪ] is used when it occurs before or after velars, after /h/, before /ʃ/, and word initially. [ə] is used elsewhere (English L1 speakers).
- Production of *bath* with a low and fully back /a:/ (English L1 speakers).

Bearing in mind these typical adult productions we modified DEAP scoring so that children would not be considered atypical if they showed these features.

Data Analysis

Results from each child's individual assessment were analysed quantitatively and qualitatively, in accordance with the DEAP manual (Dodd et al., 2002). Quantitative analysis used the indices of percentage consonants correct (PCC) and percentage vowels correct (PVC) to capture the degree of accuracy that children exhibit in their production of speech segments in words. Inconsistency scores were calculated based on the proportion of words that children could not produce in the same way when provided with multiple opportunities to produce it (more than 40% is considered to indicate inconsistency). The descriptive analysis was in the form of phonetic inventory descriptions and descriptions of phonological processes. A 90% criterion was used to determine whether a consonant should have been acquired for participants of the specific age or language category. According to Dodd et al. (2003) a participant presents with a phonological process when a process occurs five or more times in the speech sample (or more than twice in the case of weak syllable deletion). The results for each subtest were analysed for each individual participant, and then compared to other participants by statistical comparison, using independent *t*-tests and ANOVA, with language, age, and gender categories in order to establish a set of normative data for the Cape Town population. Participants with difficulties were classified into diagnostic categories based on Dodd's diagnostic category framework (Dodd et al., 2003; Dodd, 2005).

Results

Objective 1: To describe the children's consonant and vowel inventories, and phonological processes by language background (monolingual, bilingual and trilingual)

In this section we start by focusing on the consonantal inventory of participants, then vowels and finally phonological processes. As our analysis showed no significant difference between males and females, we do not report further on gender differences.

Table 3 shows the mean PCC scores for each participant group by language background. Although the trilingual participants' PCC scores were slightly higher than participants in the other language categories, this difference did not reach statistical significance.

Table 3. Mean percentage consonants correct (PCC) for participant groups by language background

Language background of children	PCC
Monolingual (n=69)	83.82 (sd 8.6)
Afrikaans bilingual (n=48)	83.94 (sd 7.13)
isiXhosa bilingual (n=25)	83.5 (sd 5.4)
Trilingual (n=8)	87.26 (sd 6.4)

Articulation results showed ten consonants that appear to develop later in Cape Town English. Table 4 shows the consonants of English which were the most challenging for the children in our sample to acquire. A substantial proportion of monolingual participants have not yet acquired all of these consonants. A similar trend is evident for Afrikaans bilingual participants, many of whom still need to acquire seven of these ten consonants. Afrikaans bilinguals were the only language category to score <90% for the acquisition of /z/. IsiXhosa bilinguals show the most advanced consonant acquisition with only four consonants not meeting the 90% criterion level. The trilingual participants experienced the most challenges, with nine consonants still to be acquired. Alveolar trill /r/, was evident across all language categories and is considered dialectal for this population. Some differences were noted between the younger and older three year olds: /ð, ʤ, ʒ/ were acquired by more than 90% of the older age category, but not yet for the younger group.

In particular, it was noted how L1 (Afrikaans or isiXhosa) influenced South African English. Such influences included:

- Alveolar trill /r/ was evident in many of the participants' speech. This is a dialectal variant of /ɹ/.
- Trilingual participants used /r/ most frequently, followed by bilingual participants and few monolingual participants. The use of /r/ was consistent between the younger and the older age categories.
- /θ/ was produced as /f/ in the final word position, but also other word positions.
- Word-final devoicing was common (discussed further in phonological processes section).

Table 4. Percentage of children in each group who have acquired each consonant

Consonants	Monolingual	Afrikaans bilingual	isiXhosa bilingual	Trilingual
ŋ	100	100	96	87.5
v	98.55	95.83	100	87.5
z	97.1	89.58	96	100
ʃ	89.86	83.33	100	87.5
tʃ	91.3	77.08	96	75
ɵ	75.36	77.08	72	87.5
ð	79.71	68.75	64	75
ʒ	86.96	81.25	88	87.5
ʤ	82.61	91.67	100	87.5
ɹ	85.51	75	84	87.5

Note: Shaded cells indicate where groups did not achieve the 90% criterion for acquisition of a consonant.

Table 5. Mean percentage vowels correct (PVC) for participant groups by language background

Language background of children	PVC
Monolingual (n=69)	95.31 (sd 3.81)
Afrikaans bilingual (n=48)	95.31 (sd 3.1)
isiXhosa bilingual (n=25)	89.3 (sd 7.1)*
Trilingual (n=8)	94.66 (sd 3.75)

* indicates significantly lower score, $p<0.01$

Table 5 shows the mean PVC scores for each participant group by language background. Differences between the monolingual, Afrikaans bilingual and trilingual group were not significant. PVC scores for the isiXhosa bilingual group were significantly lower ($t(28.29) = 4.143$, $p<0.01$) than that of all other groups. When we considered this discrepancy in light of the very similar PCC scores, we questioned the validity of the scoring process for the vowels of this group. Although the literature about adult productions had prepared us to modify our criteria when scoring this group, the criteria were not always consistently applied and we were still not accounting sufficiently for all differences associated with this group. isiXhosa bilingual children showed reduced contrasts between long and short vowels, as predicted. The vowel system of South African English, with its twelve vowels and eight diphthongs, was typically reinterpreted by isiXhosa speakers as a five vowel system.

Phonological processes refer to any simplifications of typical speech development that are present in a child's speech. In this study, we considered a child to be using a process when it appeared five or more times in the speech sample, or more than twice for weak syllable deletion. Typical processes and unusual processes were identified using the guidelines in the DEAP guidelines. Our children may be different to the DEAP normative sample given the difference in language background. Tables 6 and 7 summarize the main typical and atypical processes used.

Table 6. Typical phonological processes

	Percentage of children from each group, using phonological process			
	Monolingual (n=69)	**Afrikaans bilingual (n=48)**	**isiXhosa bilingual (n=25)**	**Trilingual (n=8)**
Gliding	63.77	43.75	28	50
Cluster reduction	28.99	27.08	52	12.5
Fronting	24.64	20.83	4	12.5
Stopping	18.84	41.67	48	37.5
Weak syllable deletion	11.6	6.25	16	25
Deaffrication	4.35	2.08	0	0
Final consonant deletion	1.45	8.33	12	0

Gliding was the most widely used typical phonological process. It was the most frequently used process for all groups, with the exception of the isiXhosa bilingual category, who used cluster reduction and stopping more frequently. isiXhosa has no clusters in it, apart from borrowed/loan words from English and Afrikaans, so it may be that these children had more need of the simplification process because they had had less exposure to these words in their L1.

Backing and devoicing were the most commonly used phonological processes for the monolingual, Afrikaans bilingual and isiXhosa bilingual participants. Backing, considered an atypical phonological process by Dodd et al. (2003), was evident in three of the four groups, and used by 16% of the isiXhosa bilingual children suggesting that it should perhaps be considered as typical for this group. Devoicing was predominantly used by the isiXhosa bilingual participants (16%), although it was also used by monolingual (4.35%) and Afrikaans bilingual (8.33%) participants to a lesser extent. Adults often devoice final consonants in South African English and this feature was thus, not surprising. It should not be considered as atypical, or even a developmental process given its widespread occurrence in adult speech. Despite the low percentages, Afrikaans bilingual participants were the only group to use assimilation (2.08%) and intrusive consonants (2.08%).

Table 7. Atypical phonological processes

	Percentage of children from each group, using phonological process			
	Monolingual (n=69)	**Afrikaans bilingual (n=48)**	**isiXhosa bilingual (n=25)**	**Trilingual (n=8)**
Backing	5.8	4.17	16	0
Devoicing	4.35	8.33	16	0
Lateralisation	4.35	0	4	0
Initial consonant deletion	1.45	2.08	4	0
Intrusive consonants	0	2.08	0	0
Affrication	0	2.08	0	0

Objective 2: To describe the prevalence of 3-year-old children with speech disorders in our sample of the population

Of 150 children assessed, we diagnosed ten participants as having speech difficulties using our modified scoring of the DEAP. This equates to a prevalence of 6.66%.

Objective 3: To describe the diagnostic category of participants with speech disorders with reference to Dodd's (2005) diagnostic framework

Table 8 summarises the ten children found to have speech difficulties. Most of the children with difficulties were from the monolingual English group. As this was the largest group in the study this is expected. When considering prevalence by language group, it was noted that the trilingual group, the smallest group, has the highest prevalence at 12.5% (one of eight children), followed by the monolingual group (seven of 69 children); and then bilingual Afrikaans (two of 48 children) and bilingual IsiXhosa (no children). Delay was the most common difficulty with six of the children falling into this category.

Table 8. Children with Speech Difficulties

	Group	**Age** years; months	**Gender**	**Diagnostic Category**
Child 1	English	3;2	F	Inconsistent
Child 2	English	3;6	F	Delayed
Child 3	English	3;7	M	Articulation
Child 4	English	3;7	M	Disorder
Child 5	English	3;7	M	Inconsistent
Child 6	English	3;8	F	Delayed
Child 7	English	3;8	F	Delayed
Child 8	Bilingual Afrikaans	3;7	F	Delayed
Child 9	Bilingual Afrikaans	3;9	M	Delayed
Child 10	Trilingual	3;5	F	Delayed

Discussion

Children from different language backgrounds obtained PCC scores that were not statistically different. This gives clinicians a baseline against which children can be compared in determining the presence of speech difficulties. There are ten consonants that three-year-olds in Cape Town may still be acquiring. These fit with the international literature on English consonantal acquisition (Bleile, 2007; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1992).

PVC was found to be a more complex index of speech accuracy. While adaptations were made for dialectal differences, bilingual isiXhosa children achieved significantly lower PVC scores than other groups, suggesting that our scoring adaptations were insufficient. Although the literature about adult productions had prepared us to modify our criteria when scoring this group, the criteria were not always consistently applied and we were still not accounting sufficiently for all differences associated with this group. The vowel system of South African English was typically reinterpreted by isiXhosa speakers as a five vowel system, as has been described by Van Rooy (2008). Awareness of dialectal vowel variations is essential in the South African context to prevent over diagnosis.

Gliding, cluster reduction, stopping and fronting constituted the ‘Big Four’ phonological processes for three-year-olds in our sample. These are typical processes and expected in light of other studies of English phonological development (Cohen & Anderson, 2011; Dodd et al., 2003). Cluster reduction was frequently used by isiXhosa-speaking children, possibly indicating the influence of L1. isiXhosa has no clusters in it, so these children might have relied on the process to help them simplify words containing clusters. Backing, considered an atypical phonological process by Dodd et al. (2003), was evident in three of the four groups, and used by 16% of the isiXhosa bilingual children. This process might be typical for isiXhosa children. There is a high frequency of velar phonemes in isiXhosa (Niesler, Louw, & Roux, 2005) so these children may favour a posterior place of articulation. Devoicing was used, predominantly by the isiXhosa bilingual participants (16%), but also by monolingual (4.35%) and Afrikaans bilingual (8.33%) participants. Since adults often devoice final consonants in South African English this should be considered as typical, and not necessarily developmental.

Prevalence of speech difficulties was estimated at 6.66% for this three-year old sample. This fits broadly with prevalence data from other studies, e.g., 2.3-24.6% (Law, Boyle, Harris, Harkness, & Nye, 2000); and 1.5% (McKinnon, McLeod, & Reilly, 2007). Comparisons are difficult because some studies have looked at a range of ages, and others only at one age, and speech difficulties may peak or fall at particular ages. Nevertheless to date there has been no prevalence data on speech difficulties in South African children and so this figure provides a preliminary set of prevalence data for this context and age group.

Of the children diagnosed with speech sound disorders, 60% were found to have a phonological delay. Fox and Dodd (2001) indicated that 47–65% of children diagnosed with speech difficulties (across a range of languages), present with a phonological delay. Thus our data fits with previous research that suggests that delay is the most common difficulty. However, given that only ten children presented with speech difficulties, these figures should be interpreted with caution and may not be representative of the entire population. A larger prevalence study should be undertaken at a national level in South Africa.

Conclusion

Three-year-olds acquiring South African English in Cape Town are broadly similar to three-year-olds acquiring English in other contexts, although for bilingual speakers L1 influences must be taken into account. While adaptations were made for dialectal differences, bilingual isiXhosa children achieved significantly lower PVC scores than other groups, suggesting that our scoring adaptations for vowels were insufficient. Approximately 7% of children in the sample were found to have speech difficulties, and of these children the greatest proportion exhibited phonological delays. This study represents only a small fragment of the language complexity in South Africa. It will be strengthened by work that investigates other age groups using the same methodology. We have presented data about the English of bilingual and multilingual children, but need to further document the children's isiXhosa and/or Afrikaans speech in order to fully understand their abilities. Another line of research arising from the present study is an intervention project describing therapy (process and outcomes) for the bilingual and multilingual children diagnosed with speech sound difficulties in this project. Dodd's (2005) classification system is useful beyond assessment, as it allows for selection of an appropriate intervention approach. Children with speech difficulties in South Africa urgently need to benefit from this knowledge. Careful documentation of their speech difficulties and responses to intervention will add to our knowledge in this field, especially given the rich multilingual and multicultural diversity of the country.

References

Bleile, K. (2007). *The late eight*. San Diego: Plural Publishing.

Bowerman, S. (2008). White South African English: Morphology and syntax. In R. Mesthrie (ed), *Varieties of English* (pp. 472-487). Berlin and New York: Mouton de Gruyter.

Broomfield, J., & Dodd, B. (2004). Children with speech and language disability: Caseload characteristics. *International Journal of Language and Communication Disorders*, *39*(3), 303-324.

Cohen, W., & Anderson, C. (2011). Identification of phonological processes in preschool children's single word productions. *International Journal of Language and Communication Disorders*, *46*(4), 481-488.

Crosbie, S., Holm, A., & Dodd, B. (2005). Intervention for children with severe speech disorder: A comparison of two approaches. *International Journal of Language and Communication Disorders*, *40*(4), 467-491.

De Klerk, V., & Gough, D. (2004). Black South African English. In R. Mesthrie (ed), *Language in South Africa* (pp. 356-380). Cambridge University Press.

De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language and Communication Disorders*, *46*(6), 613-627.

Dodd, B. (2005). *Differential Diagnosis and Treatment of Children with Speech Disorder* (2nd ed.). London: Whurr Publishers.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic evaluation of articulation and phonology*. London: Psychological Corp.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics and Phonetics*, *17*(8), 617-643.

Fox, A. V., & Dodd, B. (2001). Phonologically disordered German-speaking children. *American Journal of Speech Language Pathology*, *10*(3), 291-307.

Gxilishe, S. (2004). The acquisition of clicks by isiXhosa-speaking children. *Per Linguam*, 20(2), 1–12.

Holm, A., Dodd, B., Stow, C., & Pert, S. (1999). Identification and differential diagnosis of phonological disorder in bilingual children. *Language Testing*, *16*(3), 271-292.

International Expert Panel on Multilingual Children's Speech (2012). Multilingual children with speech disorders: Position paper. Bathurst, NSW, Australia. August 29, 2014 Online: http://www.csu.edu.au/research/multilingual-speech/position-paper

Lass, R. (2004). South African English. In R. Mesthrie (ed), *Language in South Africa* (pp. 104-126). Cambridge, UK: Cambridge University Press.

Law, J., Boyle, J., Harris , F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of literature. *International Journal of Language and Communication Disorders, 35*, 165-88**.**

Maphalala, Z., Pascoe, M., & Smouse, M. R. (2014). Phonological development of first language isiXhosa-speaking children aged 3;0–6;0 years: A descriptive cross-sectional study. *Clinical Linguistics and Phonetics*, *28*(3), 174-194.

McKinnon, D., McLeod, S. & Reilly, S. (2007). The prevalence of stuttering, voice and speech sound disorders in Primary school students in Australia. *Language Speech and Hearing Services in Schools, 38*, 5-15.

Mesthrie, R., Schneider, E., Burridge, K., Kortmann, B., & Upton, C. (2004). *A Handbook of varieties of English*. Berlin: Mouton de Gruyter.

Niesler, T., Louw, P., & Roux, J. (2005). Phonetic Analysis of Afrikaans, English, Xhosa and Zulu using South African Speech databases. *Southern African Linguistics and Applied Language Studies*, *23*, 459-474.

Ruscello, D. M. (2008). *Treating articulation and phonology disorder in children.* St. Louis: Elsevier Health Sciences.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The Speech Disorders Classification System (SDCS) Extensions and lifespan reference data. *Journal of Speech Language and Hearing Research*, *40*(4), 723-740.

Statistics South Africa (2011). Pretoria: pp. 23-25. November 13, 2013 Online: http://www.statssa.gov.za/Census2011/Products/Census_2011_Census_in_brief.pdf

Van Rooy, B. (2008). Black South African English: Phonology. Chapter in R. Mesthrie (Ed). *A Handbook of varieties of English* (pp.177-187). Berlin: Mouton de Gruyter.

Waring, R., & Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Language and Communication Disorders, 48*(1), 25-40.

World Health Organisation (2007). *Classification of functioning, disability and health version for children and youth*. Geneva: World Health Organisation.

Second dialect imitation: The production of Ecuadorian Spanish assibilated rhotics by Andalusian speakers of Spanish

Esperanza Ruiz-Peña, Diego Sevilla, Yasaman Rafat
esperanza.rpsi@gmail.com, dsevilla@uwo.ca, yrafat@uwo.ca

Western University

Abstract. This is an acoustic study that examines whether equivalence classification in second dialect acquisition operates in a similar fashion to second language acquisition. Subsequently, a real word and a nonce word imitation task were conducted to examine whether native Andalusian Spanish speakers could accurately imitate Ecuadorian Spanish assibilated rhotics. Two distinct patterns of production emerged in the real and nonce word imitation tasks that echo the results previously reported for native English-speaking participants' productions of assibilated rhotics. Moreover, a comparison of an acoustic analysis of the Andalusian Spanish-speaking participants with the Ecuadorian Spanish speakers' productions showed that not all features of these rhotics were equally acquirable.

Keywords: second dialect acquisition, phonology, imitation, production, Spanish, English

Introduction

Despite the growing interest in second dialect (D2) phonetic and phonological acquisition (e.g., Babel, 2009; Nielson, 2011), not much is known about whether the same mechanisms that underlie second language (L2) acquisition also underlie D2 phonetic and phonological acquisition. This study has two aims: (1) to test whether equivalence classification (Flege, 1995) operates in the same way in D2 phonetic and phonological acquisition as in L2 phonetic and phonological acquisition? As such, it aims to determine whether assibilated rhotic production by Andalusian Spanish speakers pattern with English speakers' productions previously reported in Rafat (2015), and (2) whether knowledge of target words affects D2 production. Flege's Speech Learning Model predicts that the smaller the acoustic-phonetic distance between the L1 and the target language (TL) sounds, the higher the possibility of equivalence classification (mapping the TL sound on to the L1 sound). This hypothesis has also been formulated to say that ‘old’ sounds are not problematic for L2 learners, those that are 'new' will eventually be acquired by L2 learners, and those that are 'similar' will be mapped on to an existing first language (L1) phonetic category, and will be most difficult to acquire. The acoustic/phonetic distance between the L1 and the target language defines how L2 sounds may be mapped on to pre-existing L1 categories.

In order to answer our questions, we will report on native Andualusian Spanish-speaking participants' assibilated rhotic production of Ecuadorian Spanish and compare them to the assibilated rhotic production data for native English-speaking participants in Rafat (2015).

Assibilated rhotics

The unity of rhotics as a class is questionable and they have often been reported to alternate with fricatives (Solé, 1992, 1998, 2002). The fact that they have been diachronically grouped into the rhotic class has been attributed to their orthographic representation, such as the letter <r> by Ladefoged and Maddieson (1996). Because of their varied acoustic nature, assibilated/fricative rhotics provide an excellent opportunity for comparing D2 and L2 productions. Assibilated/fricative rhotics have been compared to palato-alveolar sibilants because of their articulatory and acoustic similarity (Colantoni, 2006; Hall, 1997; Maddieson 1984; Solé, 2002). On the other hand, they have traditionally been classified as rhotics (Ladefoged & Maddieson, 1996; Quilis & Carril, 1971). More recently, the data from Rafat (2015) have suggested that they might acoustically exhibit both assibilation and rhoticity.

Ecuadorian, Andalusian and English rhotics

Although the tap (e.g., *caro* ['ka.ɾo]) and the trill (e.g., *perro* - ['pe.ro]) are the two rhotic phonemes in Spanish, these phonemes can have other realizations. The assibilated rhotic in Spanish is characterized as a kind of strident fricative and (e.g., Bradley 1999; Colantoni, 2001; 2006; Colantoni & Rafat, 2013; Harris, 1969; Lipski, 1994; Quilis, 1999; Quilis, 1999; Rissel, 1989; Vásquez Carranza, 2006; Widdison, 1998) and exists in several varieties of Spanish (e.g., Bradley, 1999; Colantoni, 2001; 2006; Colantoni & Rafat, 2013; Harris, 1969; Lipski, 1994; Quilis, 1999; Navarro Tomás, 1971; Rissel, 1989; Vásquez Carranza, 2006; Widdison, 1998), including Ecuadorian Spanish (e.g., Bradley, 1999). Andalusian Spanish, on the other hand, is mainly characterized by a trill and a tap and their reduced forms (e.g., Blecua, 2001). Moreover, Andalusian Spanish is generally characterized by a weakened articulation (e.g., Hualde, 2005), and it includes sibilants such as [ʃ] as an allophonic variant of affricates, such as [t͡ʃ] (e.g., Carbonero, 2001, Jiménez, 1999). Neither English, nor Andalusian Spanish have been reported to have an assibilated rhotic. Furthermore, whereas Spanish has two phonemes, namely the tap and trill, English has an approximant phoneme with retroflexed and bunched variants (e.g., Delattre & Freeman, 1968; Westbury, Hashi, & Lindstrom, 1999).

This study

Rhotics in general have been categorized as difficult sounds for both L1 (Bosch, 1983; Carballo & Mendoza, 2000; Jiménez, 1999) and L2 learners (Colantoni & Steele, 2007, 2008; Face, 2006; Major, 1986; Rafat, 2015; Reeder, 1998; Waltmunson, 2005). With respect to assibilated rhotics, previously, Rafat (2015) examined the production of Mexican assibilated rhotics by English speakers who had never been exposed to Spanish prior to the experiment. She examined the effect of orthography in two different contexts: auditory-only and auditory-orthographic. Whereas the auditory only condition led to a higher rate of sibilant, in particular an [ʃ] production, the auditory-orthographic condition yielded a higher rate of rhotic production (both assibilated rhotics and approximant rhotics). Given the scarcity of literature on D2 phonetic and phonological acquisition, the acoustic similarity of assibilated rhotics to both sibilants and rhotics, and the fact that both English and Spanish include both rhotics and sibilants, we aim to compare the D2 production of these sounds by Andalusian Speakers with English-speakers previously reported in Rafat (2015). Therefore we asked following two questions:

- Do assibilated rhotic production by native Andalusian Spanish speakers pattern with native English speakers' productions, leading us to conclude that equivalence classification (Flege, 1995) operates in the same way in D2 phonetic and phonological acquisition?
- Does knowledge of the target words affect assibilated rhotic production patterns in native Andalusian Spanish speakers?

We predict the following. First, if equivalence classification operates in the same way as in L2 phonetic and phonological acquisition, because Andalusian Spanish, similarly to English, includes both rhotics and sibilants, Andalusian Spanish speakers' assibilated production patterns will be similar to those of the English speakers reported in Rafat (2015). That is, assibilated rhotics will be mostly categorized as 'similar' sounds and therefore be produced as other types of rhotics or sibilants. Second, knowledge of real words will make rhoticity more salient and result in a higher percentage production of assibilated rhotics in the real words task in comparison with the nonce word task. Furthermore, there will be a higher rate of sibilant production in the latter than the former.

Method

Participants

Participants were 10 native Spanish speakers from Seville, Spain, and were over 18 (mean age was 37:7). They had at least 12 years of education. None of the Spanish-speaking participants had previous contact or had listened the Ecuadorian variety of Spanish.

Procedures

Participants were required to do two repetitions of the five following tasks: a picture naming task, two imitation tasks (real words and nonce words) and two reading tasks (real words and nonce words). Participants repeated each task twice. The Spanish-speaking participants were also required to complete a linguistic background questionnaire. Here, we will report the two imitation tasks, based on real words and Spanish sounding nonce words produced by an Ecuadorian Spanish speaker. Spanish-speaking participants were instructed to imitate the words they were hearing as similarly as possible to what they were hearing. Both tasks were done via PowerPoint, where the participants would hear a word once and had to repeat it immediately. The inter-stimuli interval was three and a half seconds.

A digital audio recorder Olympus LS-7 Linear PCM was used with an unidirectional sound-canceling microphone Olympus ME-52W to record the Spanish-speaking participants. The recordings were recorded with a sampling rate of 44.1 KHz and 16 bits per sample; all audio were stored in wave format.

Stimuli

The data presented in this study were collected as part of a larger study with a total of 529 Spanish stimuli and 155 fillers for the picture naming task, the two imitation tasks and the two reading tasks. In this study we are focusing on a subset of the stimuli for the real word and nonce word imitation tasks. Each imitation task included a set of 110 target words and 36 fillers and controlled for position and stress. All words were bisyllabic and nonce words did not violate Spanish phonotactics. Here, we will report on 30 stimuli for each of the imitation tasks (see Appendices A and B).

The stimuli were produced by a 33 years old male Ecuadorian (Quito) Spanish native speaker. The speaker exaggerated the assibilation of the rhotics in order to make them more salient.

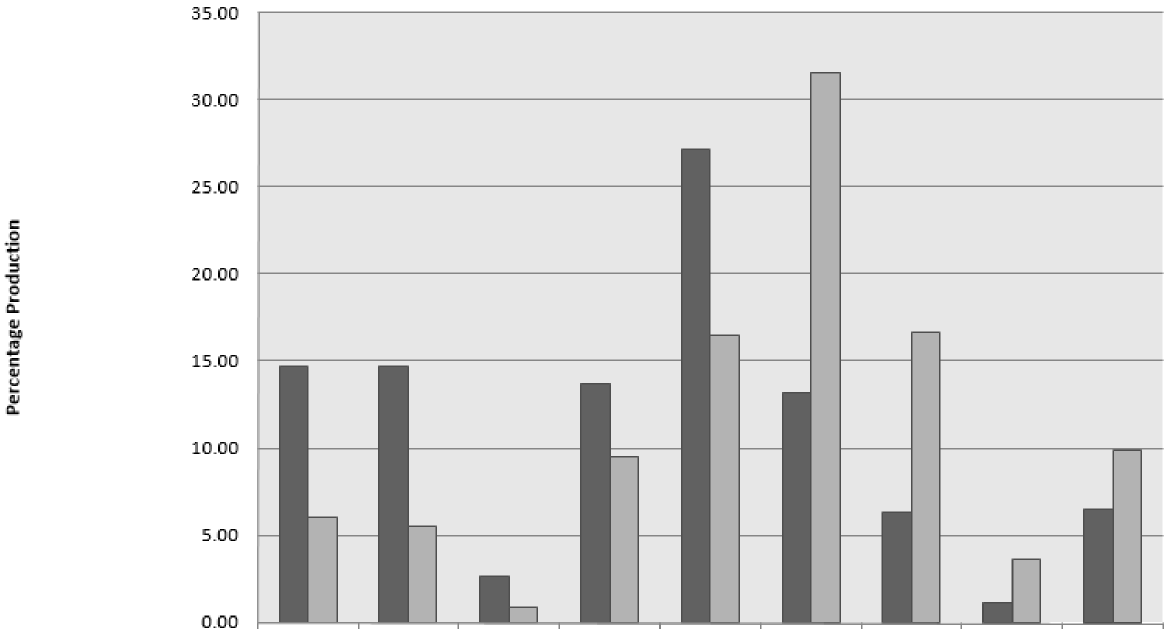
Data analysis

A total of 1,200 tokens were analyzed. Both the Spanish-speaking and Ecuadorian Spanish productions were transcribed by two of the authors. One of the authors is a native speaker of Andalusian Spanish and the other is a speaker of Ecuadorian Spanish. The stimuli produced by the Ecuadorian Spanish speaker and the productions of the Andalusian speakers were also analyzed acoustically in PRAAT (Boersma & Winnink, 2012). Segmentation and labeling of the 300 words considered here were done manually based on spectrograms and waveforms. The following measurements were considered in order to determine the degree of assibilation: duration of the rhotic (ms), the F2 (Hz), the COG (HZ), intensity (dB) and relative intensity.

The duration of assibilated rhotics was measured from the beginning to the end of the frication noise. Relative duration was defined as the duration of the segment divided by the duration of the entire word. COG's measurements were measured by capturing a 40ms. Hamming window and passing the sounds to a low filter under 15000 Hz for the whole duration of the assibilated rhotics. Relative intensity was measured as the preceding vowel's intensity minus the assibilated rhotic's intensity. The data were introduced in the programming language Python.

The results provided in this section are based on the two imitation tasks (real and nonce words). For the real words task (Figure 1), 88 tokens (14,67%) were produced as [r]; 88 tokens (14,67%) as /ɾ/; 16 tokens (2,67%) as /ʁ/; 82 tokens (13,67%) as /ɹ/; 163 tokens (27,17%) as /ř/; 79 tokens (13,17%) as

/ʒ/; 38 tokens (6,33%) as /ʃ/; 7 tokens (1,17%) as /s/; and 30 tokens (6,50%) as other. In the nonce words, 36 tokens (6,00%) were produced as [r], 33 tokens (5,50%) as [ɾ], 5 tokens (0,83%) as [ʁ], 57 tokens (9,50%) as [ɹ], 99 tokens (16,50%) as [ř], 189 tokens (31,50%) as [ʒ], 4 tokens (16,67%) as [ʃ], 22 tokens (3,67%) as [s] and 43 tokens (9,83%) as other.



	/r/	/ɾ/	/ʁ/	/ɹ/	/ř/	/ʒ/	/ʃ/	/s/	other
real word imitation task	14.67	14.67	2.67	13.67	27.17	13.17	6.33	1.17	6.50
nonce word imitation task	6.00	5.50	0.83	9.50	16.50	31.50	16.67	3.67	9.83

Figure 1. Percentage type of sound production for [ř] in imitation tasks

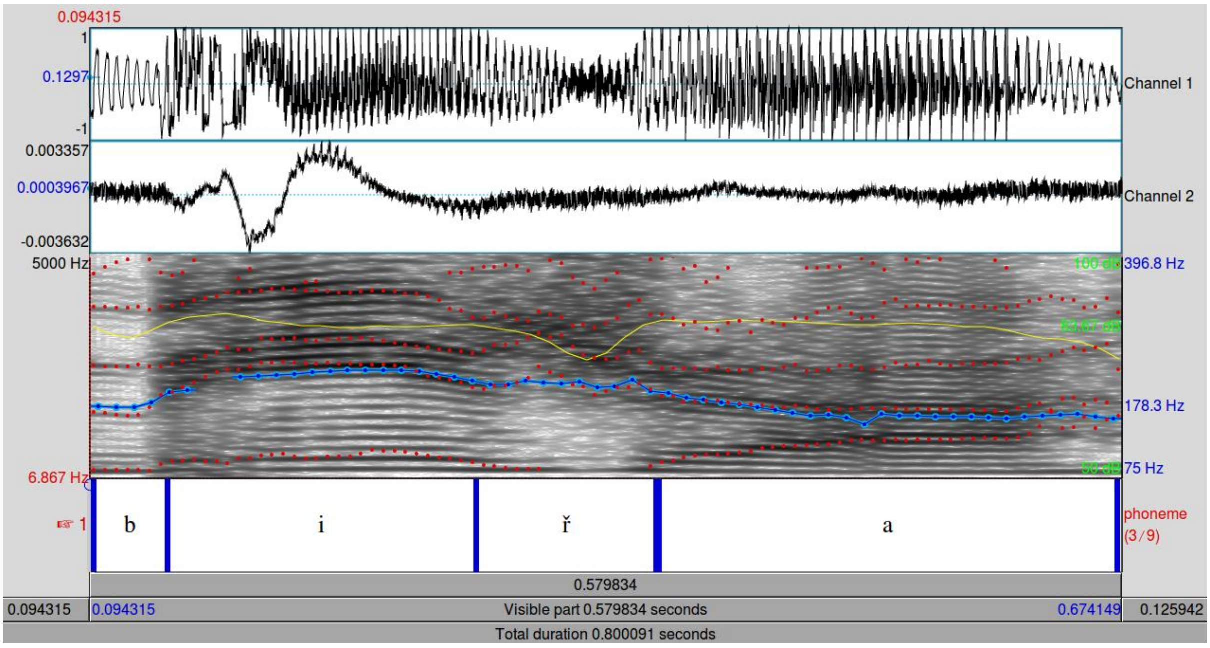


Figure 2. Spectrogram of a real word *birra* ['bi.řa] (beer) by an Andalusian Spanish speaker

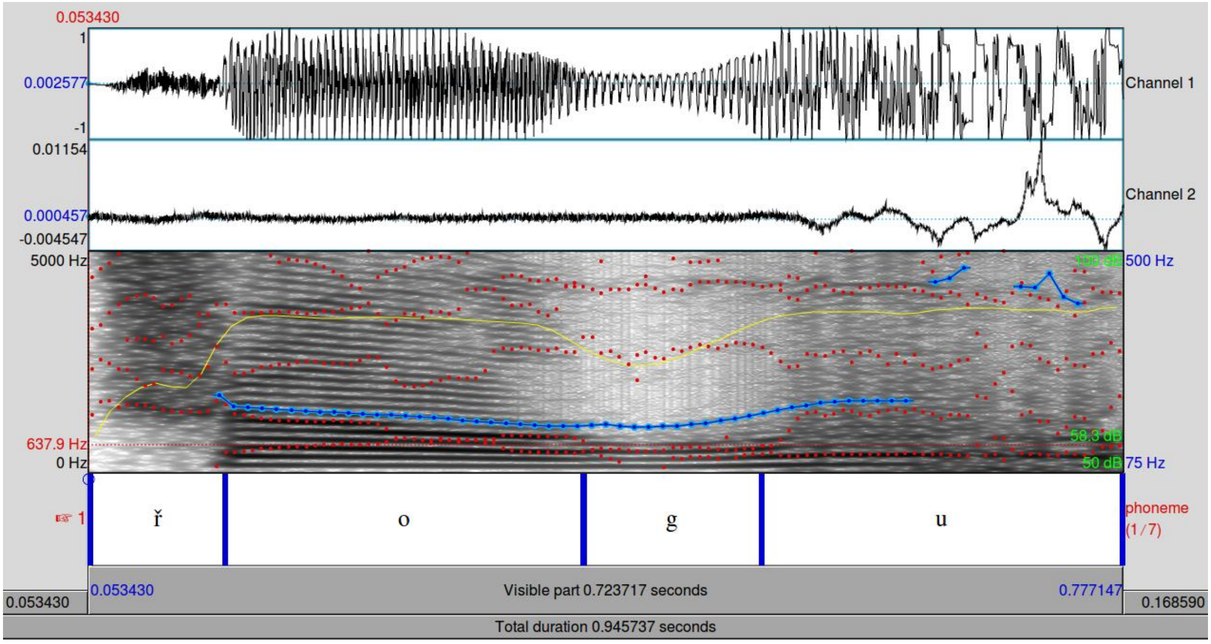


Figure 3. Spectrogram of a nonce word *rogú* [řo.'gu] by an Andalusian Spanish speaker

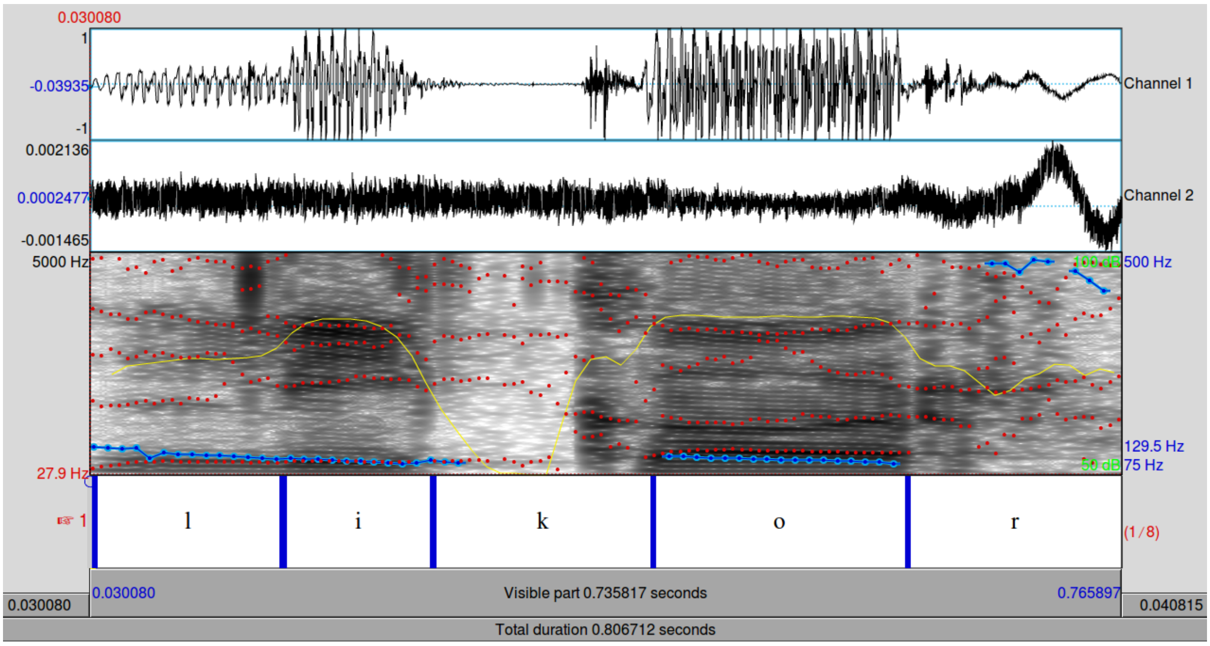


Figure 4. Spectrogram of a real word *licor* [li.'kor] (liqueur) by an Andalusian Spanish speaker

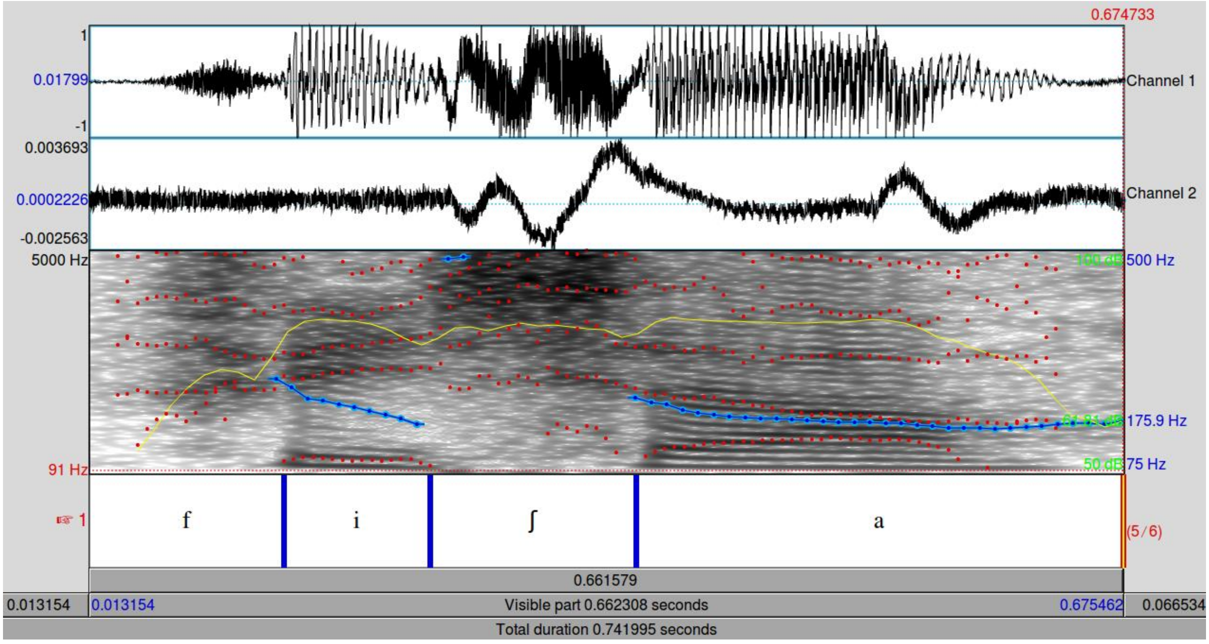


Figure 5. Spectrogram of a nonce word *firrá* [fi.'ʃa] by an Andalusian Spanish speaker

Figures 2 and 3 are spectrograms for *birra* ['bi.ra] (beer) and *rogú* [ro.'gu] respectively, which show the production of an assibilated rhotic by our Spanish-speaking participants. Figures 4 and 5, on the other hand, are spectrograms for the word *licor* [li.'kor] (liqueur), which exemplify a trill and a sibillant production *firrá* [fi.'ʃa], representative of the real word imitation, and the nonce word imitation tasks, respectively.

For the contingency tables, Chi-square and Fisher's exact tests were run. Whereas for the distributions of the [ř] features a two-sample Kolmogorov-Smirnov Test was used. The results obtained in Figure 1 show that the imitation task based on real words yielded a significantly higher rate of [ř] (e.g., 27.17% vs. 16.5 in the imitation task and nonce word imitation task, respectively) (*p*=0,000). Moreover, whereas there was a significantly higher rate of other rhotics (*p*=0,000) such as /r/, /ɾ/, /ɹ/, /ʁ/ (total 72.85%) than sibilants (total 27,17%) in the imitation task, the nonce word imitation task resulted in a significantly higher rate (*p*=0,000) of sibilants such as /ʒ/, /ʃ/ and /s/ than rhotics.

We also conducted an acoustic analysis of the asssibilated rhotics produced by our Andalusian Spanish-speaking participants and compared them with those of the Ecuadorian Spanish speaker with respect to 'intensity', 'place', 'duration' and 'voicing' in a target-like manner (Figures 6-10).

Figure 2 shows that the distributions of duration for both Andalusian Spanish participants and the Ecuadorian Spanish speaker are similar - they share the same mean (177ms). However, the Andalusian Spanish-speaking participants' duration values ranged between 119ms to 283ms while they ranged between 90 ms to 390 ms for the Ecuadorian speaker. Moreover, moreover, variability was higher for the Spanish-speaking participants than the Ecuadorian speaker.

F2 values (Figure 7) were also practically the same with small differences in median and interqueartile ranges. The F2 medians were 2090.3 Hz and 2028.9 Hz for the Andalusian Spanish-speaking participants and the Ecuadorian Spanish speaker, respectfully. However, the means were different. The Andalusian Spanish-speaking participants' mean was 2075 Hz while for the Ecuadorian Spanish was 1997 Hz. F2 values were also generally produced as higher for the Andalusian Spanish-speaking participants (1236.1Hz to 2785.9369Hz) than the Ecuadorian Spanish speaker (1411.7Hz to 2729.5Hz).

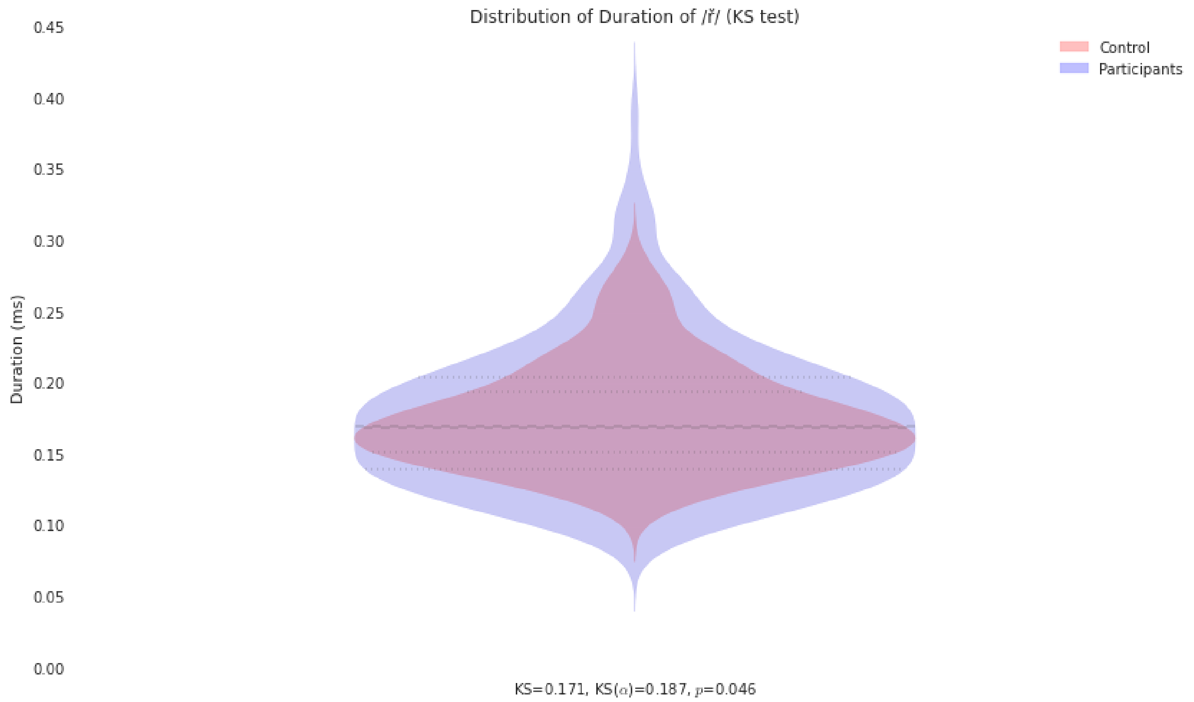


Figure 6. Duration of [ř] by Andalusian Spanish speakers and Ecuadorian Spanish speaker

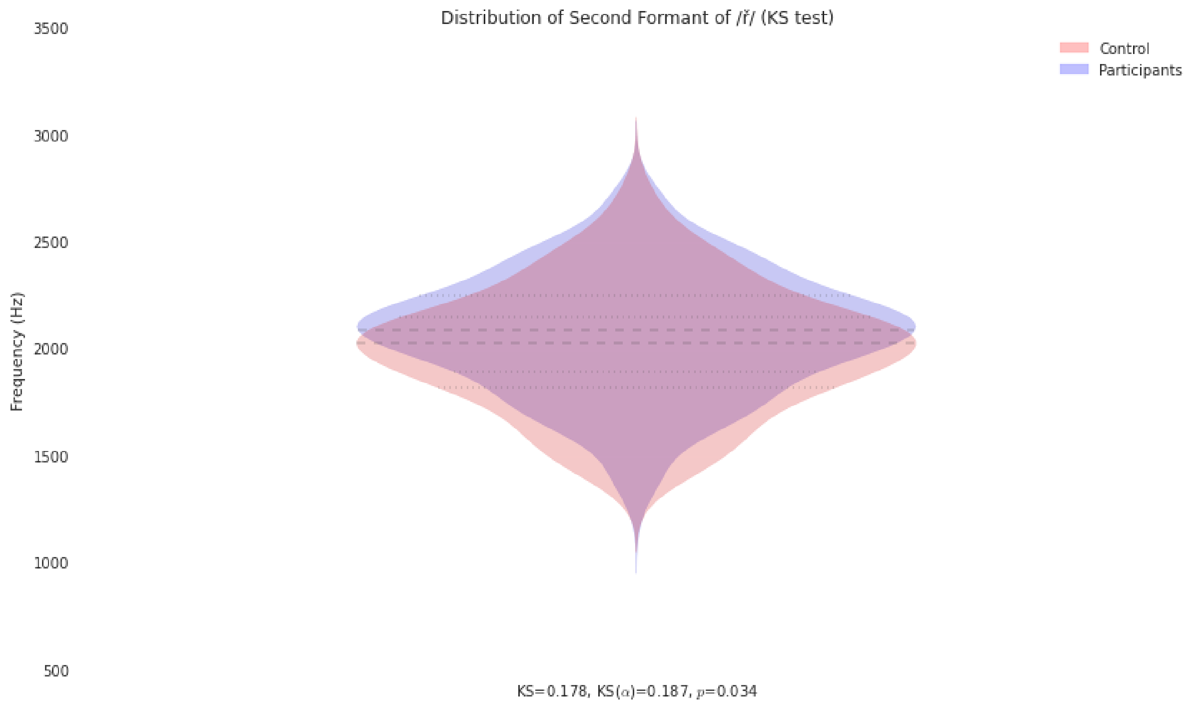


Figure 7. Distribution of F2 of [ř] by Andalusian Spanish speakers and Ecuadorian Spanish speaker.

Figure 8 shows the data of the distribution of CoG of the assibilated rhotic of both groups. The plot shows in this case that both groups, Ecuadorian Spanish and Spanish-speaking participants, are not drawn from the same distribution. Moreover, the means were 1025 Hz for the Andalusian Spanish-speaking participants and 616 Hz for the Ecuadorian speaker and were significantly different (*p=0.000*). Furthermore, the CoG ranged between 35.7 Hz and 3774.1 Hz for the Andalusian Spanish participants while it ranged between 182.6 Hz to 2192.6 Hz for the Ecuadorian speaker. These results suggest that Andalusian speakers do not produce the COG in a target-like manner.

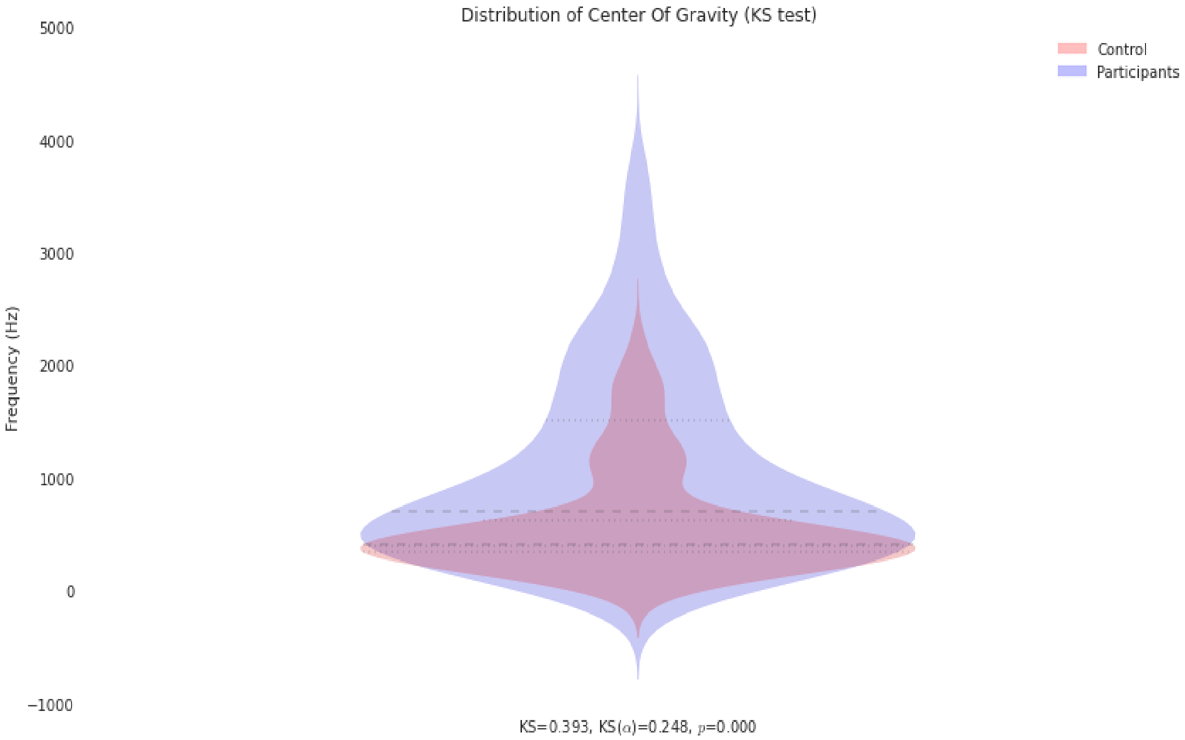


Figure 8. Distribution of CoG of [ř] by Andalusian Spanish speakers and Ecuadorian Spanish speaker

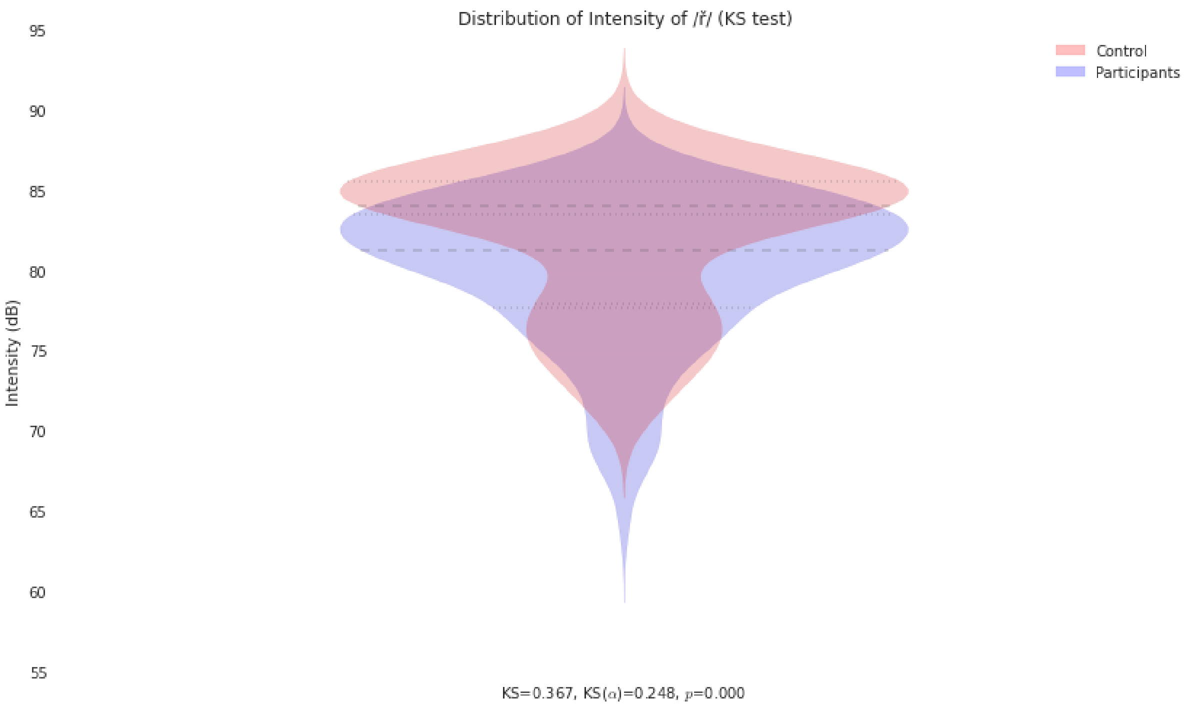


Figure 9. Intensity of [ř] by Andalusian Spanish speakers and Ecuadorian Spanish speaker

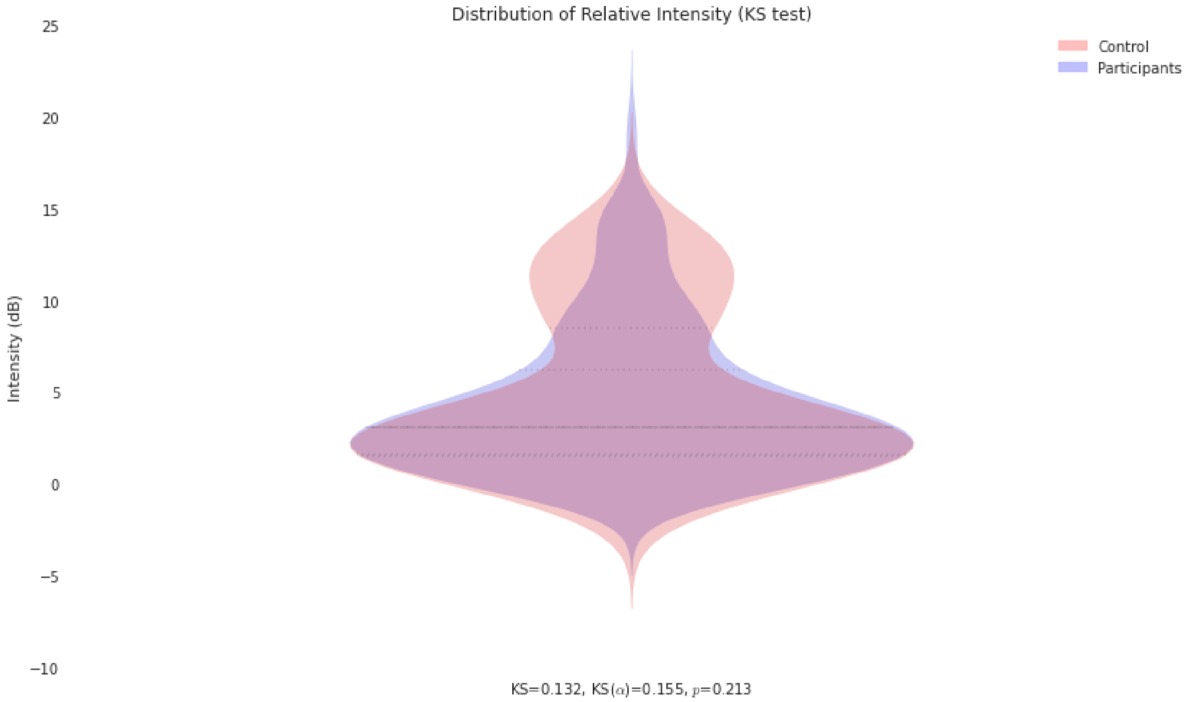


Figure 10. Relative intensity of [ř] by Andalusian Spanish speakers and Ecuadorian Spanish speaker

Figure 9 contains the results with respect to the intensity parameter. The distributions of both groups are different. An important aspect to highlight is that, although, it looks like there is a shift of intensity values of the Andalusian Spanish-speaking participants towards those of the Ecuadorian Spanish speaker, the difference between both distributions was significant (*p*=0.000). The Spanish-speaking participants produced the assibilated rhotic with less intensity than the Ecuadorian Spanish speaker. The intensity values ranged between 88dB and 87.0dB for the Andalusian Spanish speaker, while they ranged between 71.64 dB to 88.2 dB for the Ecuadorian Spanish speaker. The mean for the Andalusian Spanish-speaking participants was lower (80 dB) than its Ecuadorian Spanish counter-part (82dB) and the differences were significant (*p*=0.000).

Figure 10 reports the relative intensity of the assibilated rhotics by Andalusian Spanish speakers and the Ecuadorian Spanish speaker. The graph illustrates that the relative intensity of the assibilated rhotic by the Spanish-speaking participants falls within the range of 0.93dB and 19.74dB, while the Ecuadorian Spanish speakers' ranges between 1.13dB to 14.72dB. Both the Andalusian Spanish-speaking participants and the Ecuadorian Spanish speaker shared close values of intensity as their first quartile and median show (3.13dB by Spanish-speaking participants and 3.16dB by the Ecuadorian Spanish), and both groups produced the same amount of assibilated rhotic [ř] between 0dB and 5dB. However their behaviours differed because the Ecuadorian speaker produced more assibilated rhotics at higher values of intensity (more than 8dB) than the Andalusian Spanish-speaking participants. Although the KS test reports same distribution for the relative intensity, at *p*>0.2, we cannot trust that the Ecuadorian Spanish and the Spanish-speaking participants behave alike. Regarding to the mean, Spanish-speaking participants average was 4dB and 5dB for the Ecuadorian Spanish, and they were significantly different (*p*=0.000).

Discussion

We tested two hypotheses in this study. First, we predicted that if equivalence classification operates in the same way as in L2 phonetic and phonological acquisition, because Andalusian Spanish, similarly to English, includes both rhotics and sibilants, native Andalusian Spanish speakers' assibilated production patterns would be similar to those of the native English speakers reported in Rafat (2015). That is, although a relatively small percentage of assibilated rhotics would be attested in the D2 production data, assibilated rhotics would be categorized as 'similar' sounds and produced as other types of rhotics or sibilants, for the most part. Second, we predicted that knowledge of real words would make rhoticity more salient and result in a higher percentage production of assibilated rhotics in the real words imitation task in comparison with the nonce word imitation task. Furthermore, it was predicted that there would be a higher rate of sibilant production in the latter than the former task. Our results showed that all hypotheses in this study were confirmed. There were striking similarities between the patterns that emerged here for the production of assibilated rhotics for our native Andalusian Spanish speakers and those reported for the native English speakers in Rafat (2015). In the real word imitation task, assibilated rhotics were acquired only at the rate of 27.17%. The results in Rafat (2015) also showed that native English-speaking participants produced assibilated rhotics at a similar rate (23.13%). Moreover, as in Rafat (2015), the production patterns varied between the two tasks. In the real word imitation task, similarly to the audio-orthographic group in Rafat (2015), the participants for the most part produced L1-based rhotic sounds. However, in the nonce word imitation task, similarly to the audio-only group in Rafat (2015), the bulk of the participants' productions consisted of sibilants. The fact that the results of the real word imitation task echoed the results of the audio-orthographic group in Rafat (2015) and the results of the nonce word imitation task echoed the results of the audio-only group in Rafat (2015) leads us to conclude that knowledge of words can affect equivalence classification in both D2 and L2 acquisition. Rafat (2015) showed that assibilated rhotics exhibit various degrees of assibilation. Moreover, she proposed that when assibilated rhotics are highly assibilated, exposure to the orthographic cue <r> can make rhoticity, the less salient cue of assibilated rhotics, more salient, for the learners and lead to target-like productions. She also proposed that when rhotics are not heavily assibilated, exposure to <r> may result in L1-based transfer of the English rhotics or result L1 overriding the input. Here we explain the differences between the two tasks by proposing that knowledge of the target words, specifically the fact that these words are produced with a rhotic in Andalusian Spanish, makes the less salient feature of these rhotics (namely rhoticity) more salient, and results in either a target-like production or an D1-based rhotic transfer, such as the production of a trill or a tap. When participants do not have knowledge of the words, however, because there is nothing in L1 phonology to make participants notice the rhotic feature in the input and given that assibilation is a more salient feature than rhoticity for assibilated rhotics, native Andalusian participants are more likely to map these sounds on to sibilants in their D1.

Although the overall patterns of our D2 productions mirrored the production patterns of the L2 speakers in Rafat (2015), we must note that some differences were also noted. For example, whereas assibilated rhotics were mainly produced as a [ʃ] by the L2 auditory-only group in Rafat (2015), they were mainly produced as a [ʒ] in this study. We speculate that this might be because of the differences in the degree of voicing of assibilated rhotics of the Ecuadorian Spanish speaker in this study in comparison with the Mexican Spanish speaker in Rafat (2015). We will have to further explore this hypothesis in future work. We also noted four instances where assibilated rhotics were produced as [l] in the D2 production data. However, [l] was never attested in the production of the native L2 English-speaking participants' productions in Rafat (2015). We therefore attributed [l] productions to the fact that liquid neutralization is a characteristic of Andalusian Spanish (e.g., Ruiz-Peña, 2013). In all, we speculate that although equivalence classification may operate similarly in L2 and D2 learners, the D1 phonological processes may also exert an influence in D2 learners' productions. However, more data is needed before we can generalize this finding.

This study also conducted an acoustic analysis of the assibilated rhotics produced by the Ecuadorian Spanish speaker and our native Andalusian Spanish-speaking participants. According to the auditory transcription of the results, 27% of the target assibilated rhotics were realized as assibilated rhotics by our participants (they were thought to have both rhotic-like and sibilant-like qualities). A visual analysis of the spectrograms also showed that these assibilated rhotics exhibited a high degree of frication, suggesting that manner was produced in a target-like fashion in these realizations. However, an acoustic analysis of the other features associated with assibilated rhotics, showed that not all the acoustic parameters were produced in a target-like manner. Whereas duration and F2 were realized in a target-like manner, COG, intensity and relative intensity did not seem to be acquirable. Previously, manner has been said to be the most salient feature (Steriade, 1999) of rhotics (Ohala & Kawasaki, 1984). Manner was also the most acquirable feature for the French voiced dorsal fricative [ʁ] in Colantoni and Steele (2007). In addition, duration has been shown to be a cue that Spanish L2 learners rely on when identifying new L2 vowels (e.g., Bohn, 1995; Cebrian, 2006; Escudero, 2001). Escudero (2001) found that while Scotish-English speaking learners of Southern English had native-like perception of the Southern British English /i-ɪ/ contrast, Spanish-speaking learners used duration to identify these L2 vowels. Therefore, reliance on duration is a language-specific cue-reliance tendency in Spanish. With respect to our F2 results, we note that both F2 and COG correlate with place of articulation. Therefore, it is not clear why only the F2 values correlated with those of the Ecuadorian speaker's. In the future, we will need to also measure voicing. What is apparent, however, is that although 27% of the data created the precept of an assibilated rhotic in our transcribers, not all the parameters were produced in a target-like manner by our Andalusian Spanish-speaking participants.

Conclusions

In all, we believe that our study is important because it makes three new contributions to our understanding of D2 phonetic and phonological acquisition. First, we have shown a very robust similarity between the D2 and L2 production patterns that suggest equivalence classification operates similarly in both cases. Second, just like knowledge of orthography can modulate equivalence classification in native L2 participants, knowledge of words can modulate equivalence classification in native D2 learners. Moreover, based on the acoustic analysis of the assibilated rhotics produced by the Andalusian Spanish-speaking participants, although the Flege's SLM may predict the overall patterns of equivalence classification and hence D2 production patterns, it does not make adequate predictions about the relative difficulty of the phonetic features of the D2 sounds. Here, we have shown that while manner, duration and F2 were acquirable, other parameters such as COG, intensity and relative intensity were not. Moreover, we will need to investigate positional effects. Finally, there is some evidence in our data to suggest that D2 productions may also be additionally constrained by D1 phonological processes, although more data is needed to verify this.

We are also mindful of the fact that although one of the strengths of this study lies in the fact that it is a very controlled study that tells us how equivalence classification may operate in the very beginning stages of D2 acquisition, it is not a naturalistic study. In the future, we would like to extend our study to include a more naturalistic condition such as a conversation between the speakers of the two varieties of Spanish. Moreover, we hope to be able to also examine extra-linguistic factors such as, perception of D1 and D2 dialects on the dimensions of prestige, solidarity, social attractiveness and linguistic validity (Rindal, 2010), as well as the degree of contact with other dialects, and place of residence. In our study, our participants did not report any contact with Ecuadorian Spanish. Moreover, we know that although the Andalusian variety of Spanish is stigmatized in Spanish (e.g., Ruiz-Peña, 2013), Andalusian Spanish speakers are very proud of their variety of Spanish (e.g., Ruiz-Peña, 2013). In this study, we cannot really examine the effect of the social context, but we do believe that our participants, as a group, are not the most amenable to imitating another variety of Spanish. It would be interesting to compare assibilated rhotic production by Andalusian speakers with speakers of another variety of Spanish, who may relate differently to their D1 - where D1 may not be such a strong identity marker. We also think that the perceived degree of prestige that Ecuadorian Spanish enjoys in the Spanish-speaking will have to be further investigated as it may be another factor that may contribute to the low accurate production of these assibilated rhotics. Furthermore, our data is based on production and we will need to further validate our proposals regarding equivalence classification in D2 and L2 acquisition of assibilated rhotics by conducting a perception task.

References

Babel, M., (2009). *Phonetic and social selectivity in speech accommodation*. Doctoral dissertation, University of California, Berkeley.

Bohn, O.-S. (1995). *Cross language speech production in adults: First language transfer doesn't tell it all*. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279-304). Baltimore: York Press.

Blecua, B. (2001). *Las vibrantes del español: manifestaciones acústicas y procesos fonéticos*. PhD Thesis, Universidad Autónoma de Barcelona.

Carbonero, P. (2001). *Identidad lingüística y comportamientos discursivos. Sociolingüística Andaluza 12.* Universidad de Sevilla.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*, 372-387.

Escudero, P. (2001). The role of the input in the development of L1 and L2 sound contrasts: language specific cue weighting for vowels. *In Proceedings of the 25th annual Boston University conference on language development* (Vol. 1, pp. 250-261). Somerville, MA: Cascadilla Press.

Flege, J. (1995). *Second language speech learning: Theory, findings and problems*. In W. Strange (ed.*), Speech perception and linguistic perception: Theoretical and methodological issues* (pp. 233-277). York Press, Timonium.

Hualde, J. (2005). *The sounds of Spainsh*. Cambridge, UK: Cambridge University Press.

Jiménez, R. (1999). *El andaluz*. Arco/Libros, S.L. p.17-72

Lipski, J. M. (1994). *Latin American Spanish*. New York: Longman.

Ohala, J., & Kawasaki, H. (1984). Prosodic phonology and phonetics. *Phonology Yearbook, 1*, 113-127.

Rafat, Y. (2015). The interaction of acoustic and orthographic input in the L2 production of assibilated/fricative rhotics. *Applied Psycholinguistics*, *36*(1), 43-64.

Rindal, U. (2010). Constructing identity with L2: Pronunciation and attitudes among Norwegian learners of English. *Journal of Sociolinguistics*, *14*(2), 240-261.

Ruiz-Peña, E. (2013). *"Alma" o "arma", evidencia de la neutralización/l//r/en la variedad dialectal andaluza de Sevilla*", Master thesis, Western University.

Solé, M. (1998). Phonological Universals: Trilling, voicing and frication. *Berkeley Linguistics Society*, *24*, 427-442.

Solé, M. (2002). Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics*, 30, 655-688.

Steriade, D. (1999). *The phonology of perceptibility effects: The P-map and its consequences for constraint organization*. Unpublished manuscript, University of California, Los Angeles.

Appendix A: List of real words and fillers for the real word imitation task

Table 1. List of real words per position and stress

Position	Stressed	Unstressed
Word initial	remo risa ropa rusa ruta	rubí ramón rapé robé rosé
Medial - intervocalic	birra parra tarro porro burro	borré morral cerró carril barrí
Word final	poder calar sabor licor pulir	dólar sónar fúcar lémur césar

Table 2. List of fillers for the real word imitation task

Fillers			
glúten	pilló	fin	
lápiz	biblia	habla	
llanta	domé	folio	subí
llave	malla	mal	maná
gasto	gel	tos	mote
fajó	fuga	jefe	callé
pez	tan	flaco	tabla
lomo	boté	fumó	toldo
lobo	beca	plato	
mano	llamé	zafé	

Appendix B: List of nonce words and fillers for the nonce word imitation task

Table 3. List of nonce words per position and stress

Position	Stressed	Unstressed
	refo	rogú
	rube	refó
Word-inital	riga	raní
	renu	rupá
	raca	ricú
	firrá	porre
	nerró	hurri
Medial - intervocalic	murrí	lerra
	nurró	tarre
	carrí	lirra
	liper	júpir
	dafer	létar
Word final	padur	cásor
	zater	cáfor
	jalor	sígur

Table 4. List of fillers for the nonce word imitation task

Fillers			
zombón	fueya	zop	
loifu	mif	abce	
mul	moltre	blaspo	
astog	guybla	llejal	mofsú
jófa	luhom	gafu	bliapa
paxfi	feheje	onmex	tebó
fezá	temlla	omlan	gaox
jul	julmú	dolpa	sot
llopí	moan	nat	
fangué	naami	bizú	
		pizlo	

The discrimination of Spanish lexical stress contrasts by French-speaking listeners

Sandra Schwab[1], Joaquim Llisterri[2]
sandra.schwab@uzh.ch, joaquim.llisterri@uab.cat

[1] Universität Zürich and Université de Genève, [2]Universitat Autònoma de Barcelona

Abstract. The goal of the present research is to examine the role of the acoustic parameters involved in the discrimination of Spanish lexical stress contrasts by French-speaking listeners, and to validate the results of a previous study in which we used a stress identification task. The participants of the present experiment were ten French-speaking advanced learners of Spanish and ten French-speaking participants without knowledge of Spanish. They performed an AX discrimination task in which they heard pairs of Spanish trisyllabic words, and had to indicate whether the position of stress in the two stimuli was the same or different. The results support the idea that the perception of an accentual difference depends on the acoustic parameters involved in the manipulation applied to create a stress shift. More specifically, we found that the role of the acoustic parameters varies as a function of the accentual pattern and the competence in L2.

Keywords: lexical stress, stress 'deafness', prosodic transfer, L2 speech perception, French L1, Spanish L2

Introduction

It has been frequently noted that French learners of Spanish tend to place the stress on the final syllable of Spanish words (Gil, 2007; Rico, 2012), a fact that is explained as the manifestation of an accentual transfer, since French has been traditionally classified as a fixed-stress language, while Spanish is characterized as a free-stress language (Garde, 1968). In French, primary stress delimits sequences of words (stress groups or rhythmic groups) and appears at the end of such sequences, specifically in reading and in neutral speaking styles (Carton, 1974; Rossi, 1979; Vaissière, 1990). On the contrary, Spanish stress fulfils a distinctive role at the lexical level (Quilis, 1981, 1993), allowing for contrasts such as [ˈbaliðo] (*válido*, 'valid'), [baˈliðo] (*valido*, 'I validate') and [baliˈðo] (*validó*, 'he/she validated').

The acoustic phonetic realization of stress also differs in French and in Spanish. Although syllabic prominence is achieved through variations in fundamental frequency (f_0), intensity and duration in both languages, stress in French is realized with an increase in duration and, to a lesser extent, in f_0 (Léon & Martin, 2000; Léon, 2011); in Spanish, stress is usually the result of a combined increase of duration and f_0 values (Quilis, 1981).

Moreover, native speakers differ in the perceptual cues they use to detect accentual prominences. French listeners tend to privilege an increase in f_0 (Rigault, 1962), while changes in f_0 (Enríquez, Casado, & Santos, 1989) combined with changes in either duration or intensity appear to be necessary to identify the position of lexical stress in Spanish isolated words (Llisterri, Machuca, Mota, Riera, & Ríos, 2005).

These phonological and phonetic differences in the accentual systems might account for the difficulties experienced in production, but also in perception, by speakers of a fixed-stress language such as French when confronted to accentual contrasts in a free-stress language such as Spanish.

The role of the phonological categories of the first language (L1) as mediators in the perception of a second language (L2) was already acknowledged by the early European tradition of the Prague Linguistic Circle. The metaphors of 'phonological deafness' (*surdité phonologique*) (Polivanov, 1931) and of the 'phonological sieve' (*crible phonologique*) (Troubetzkoy, 1949) tried to capture the perceptual nature of the errors due to transfer from the L1 to an L2. Building on these ideas, the notion of 'accentual filter' (*crible accentuel*) has been introduced by several researchers as an explanation for transfer phenomena in the domain of stress (Billières, 1988; Borrell & Salsignac, 2002; Dolbec & Santi, 1995; Frost, 2010; Muñoz García, 2010; Salsignac, 1998).

In a series of studies on the perception of lexical stress by French speakers, Dupoux and his collaborators (Dupoux, Pallier, Sebastián Gallés, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián Gallés, 2001; Dupoux, Sebastián Gallés, Navarrete, & Peperkamp, 2008; Peperkamp & Dupoux, 2002) have put forward the hypothesis that a stress 'deafness' (a particular case of phonological 'deafness') might explain the difficulties exhibited by speakers of a language lacking contrastive stress when they are exposed to contrasts in accentual patterns. The results of their experiments indicated that when stimuli with phonetic variability were presented and a cognitively demanding task was used, French listeners, either monolingual or learners of Spanish, had difficulties in perceiving the position of stress which were not found in the native Spanish-speaking participants. This led the authors to conclude that "stress 'deafness' is better interpreted as a lasting processing problem resulting from the impossibility for French speakers to encode contrastive stress in their phonological representations" (Dupoux et al., 2008, p. 683).

Using a different approach, Mora, Courtois, and Cavé (1997) have shown that French listeners without knowledge of Spanish were able to correctly identify 87% of the stressed syllables in a sample of spontaneous speech in Spanish, although they did not necessarily rely on the same acoustic cues used by native Spanish listeners. Very similar levels of performance in a stress identification task (around 83%) have been reported by Muñoz García, Panissal, Billières, and Baqué (2009) for French speakers listening to isolated words and to words in a sentence context in Spanish; furthermore, participants with an advanced level of Spanish performed better than those with basic or intermediate knowledge of the language.

The results of all these studies suggest that the effects of the accentual filter might depend, among other factors, on the nature of the task performed by the participants and, in certain cases, on their level of proficiency in the L2. In order to shed some more light on the prosodic transfer that may occur in the perception of lexical stress, we have undertaken a series of experiments in which French listeners were exposed to accentual contrasts in Spanish.

The results of a first experiment showed that, when performing an identification (i.e. phonetic) task, French listeners were able to identify the position of lexical stress in approximately 70% of the cases, although the performance was influenced by the type of stress pattern; moreover, f_0 appeared as the most important parameter in the perception of the stress position and knowledge of Spanish influenced the sensitivity to the acoustic cues which signal the prominence of the stressed syllable (Schwab & Llisterri, 2010, 2011b).

In a second experiment, a shape-pseudoword matching task was adopted. We found that French-speaking listeners were able, after a short training, to encode and to retrieve the accentual information present in a small set of Spanish isolated pseudowords, although the responses to the acoustic manipulations performed on the stimuli lead us to hypothesize that the accentual representation acquired and stored by the French speakers was more rigid than the representation encoded by Spanish native speakers (Schwab & Llisterri, 2011a, 2012, 2014).

In the following sections, we will present the methodology and the results of a third experiment, in which a discrimination task has been used.

Method

Participants

Two groups of French speaking participants took part in the experiment: a group with advanced knowledge of Spanish and another one with no knowledge of the language. The advanced group was composed of 10 participants. They were between 21 and 36 years old and were all raised in a monolingual French speaking environment. They had been studying Spanish at the University of Neuchâtel (Switzerland) for 6-11 years. The group without knowledge of Spanish consisted of 10 students of the University of Neuchâtel. They were between 19 and 24 years old and were all raised in a French speaking monolingual environment. None of them reported good knowledge of Italian, which excludes the potential bias of knowing a free-stress Romance language.

Material

The corpus, taken from Llisterri et al. (2005), was composed of 4 triplets of trisyllabic words (CV.CV.CV) and 4 triplets of trisyllabic analogue pseudowords. All words and pseudowords could be proparoxytones (PP; e.g., [ˈbalið̞o], *válido*, ‘valid’), paroxytones (P; e.g., [baˈlið̞o], *valido*, ‘I validate’) or oxytones (O; e.g., [baliˈð̞o], *validó*, ‘he/she validated’).

The stimuli were divided into *Base* stimuli (i.e. without any manipulation) and *Manipulated* stimuli. For the creation of manipulated stimuli, we proceeded as follows: in proparoxytone words, f_0, amplitude and duration values for each vowel were replaced by the corresponding f_0, amplitude and duration values found in the equivalent paroxytone words (PP>P *Manipulated stimuli*); likewise, in paroxytone words, f_0, amplitude and duration values for each vowel were replaced by the corresponding f_0, amplitude and duration values found in the equivalent oxytone words (P>O *Manipulated stimuli*). In fact, manipulated stimuli resulted in a shift to the right of the accentual information, as can be observed in Figures 1 and 2.

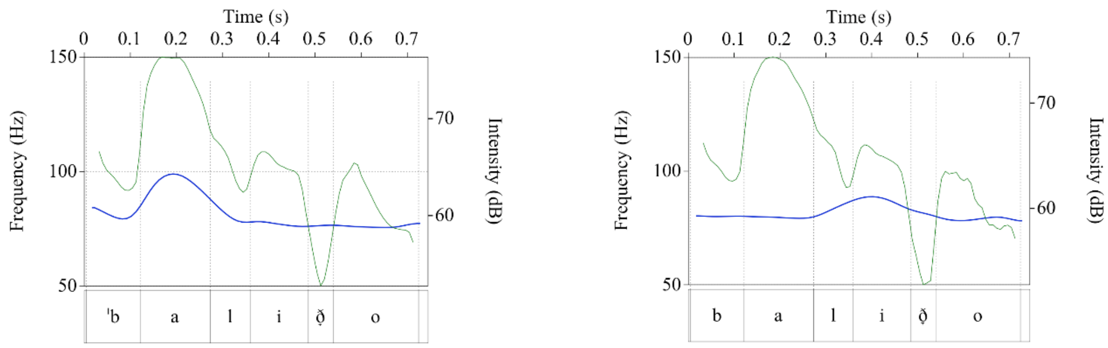


Figure 1. PP>P Manipulated stimulus: base stimulus *válido* (PP) on the left and the result of the manipulation of f_0 (in blue) using the values from *valido* (P) on the right.

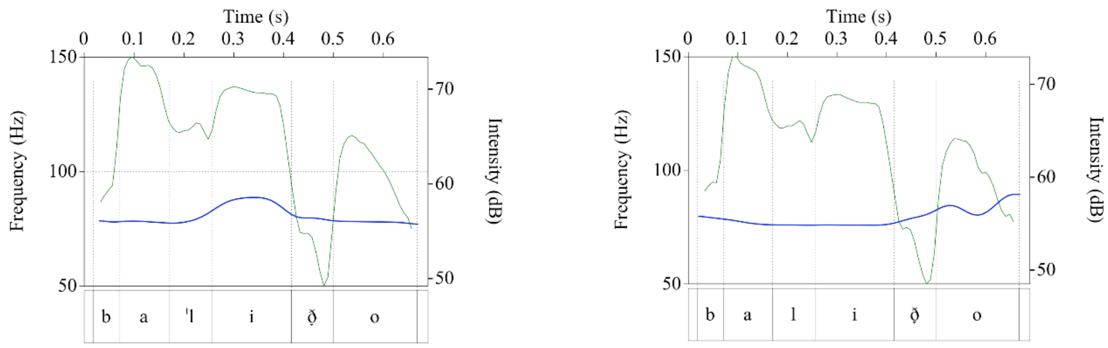


Figure 2. P>O Manipulated stimulus: base stimulus *valido* (P) on the left and the result of the manipulation of f_0 (in blue) using the values from *validó* (O) on the right.

The values were modified not only individually, but also simultaneously, obtaining the seven possible combinations of manipulated parameters: f_0, amplitude, duration, f_0+duration, f_0+amplitude, duration+amplitude, f_0+duration+amplitude. This strategy allows us to study the effects of each acoustic cue both in isolation and in combination with the others. All the manipulations were performed by resynthesis, using the PSOLA algorithm implemented in Praat (Boersma & Weenink, 2015).

During the test, the stimuli were presented in pairs in which a *Manipulated* stimulus was always presented with a *Base* stimulus. The *Base* stimulus appeared in half of the stimuli with the original stress pattern of the *Manipulated* stimulus (i.e. PP *Base* stimulus for PP>P *Manipulated* stimulus; P *Base* stimulus for P>O *Manipulated* stimulus) and, in the other half of the stimuli, with the intended shifted stress pattern of the *Manipulated* stimulus (i.e. P *Base* stimulus for PP>P *Manipulated* stimulus; O *Base* stimulus for P>O *Manipulated* stimulus). In total, 224 of different stimuli were used: 4 words and 4 pseudowords x 2 patterns x 7 manipulations x 2 pair members. Half the stimuli were presented in the *Base-Manipulated* order and the other half in the *Manipulated-Base* order.

Control pairs with identical stimuli were also included in the test. Among them, 24 were *Base-Base* pairs and 48 were *Manipulated-Manipulated* (4 words and 4 pseudowords x 3 manipulations x 2 patterns). In total, 296 trials were used in this experiment.

Procedure

Participants performed a stress AX discrimination task and were run individually. The experiment was run from a laptop using the DMDX software (Forster, 2012), which recorded the participants' responses. The participants listened to each trial (composed of a pair of stimuli) and had to indicate, as fast as possible, whether the position of the stress in the two members of the pair was "Identical" or "Different", by pressing the *Id* or *Diff* key on a keyboard. The two elements of the trial were separated by 500 ms. The participants had 2 seconds to answer and did not receive any feedback. The experiment began with a few training trials and lasted 20 minutes.

The 296 trials were divided into 4 blocks, each one containing 74 trials with the following conditions: 37 words and 37 pseudowords; 28 *Base-Manipulated* and 28 *Manipulated-Base* pairs; 6 *Base-Base* pairs (2 for each stress pattern: PP, P, O); 8 pairs for each of the 7 modifications; 12 control *Manipulated-Manipulated* pairs (6 PP>P and 6 P>O); 14 pairs for each accentual pattern (PP>P with P; PP with PP>P; P with P>O; P>O with O) The order (*Base-Manipulated* and *Manipulated-Base)* was counterbalanced across lexical status, manipulations, and stress patterns. Within each block, the trials were presented randomly, and the 4 blocks were also randomly distributed. Thus, each participant received a different presentation order.

Data analysis

First, the correct/incorrect responses to the control trials (i.e. identical pairs) were collected in order to ensure that the participants performed correctly the task. Then, we examined the Identical/Different (*Id*/*Diff*) responses of the test trials, composed of a *Manipulated* stimulus and a *Base* stimulus.

The two accentual patterns (PP>P and P>O) are hardly comparable, because stress is also associated with the prepausal status of the last syllable of the word in the P>O pattern. To that respect, Enríquez et al. (1989) noted that "para explicar la percepción acentual no sólo hay que tener en cuenta el parámetro que interviene, sino, además (y muy especialmente, en la Duración), el esquema acentual de la palabra . . . nos lleva a considerar una oposición entre segmentos interiores de palabra frente al segmento final de palabra, con comportamientos diferentes en cada caso" (p. 267). For that reason, we ran two separate analyses for PP>P and P>O stimuli in the case of pairs containing different stimuli.

Statistical analyses were carried out with the R software (Kuznetsova, Brockhoff, & Christensen, 2014; R Core Team, 2014). We ran the analyses on the Identical/Different responses using mixed-effects logistic regression models (Baayen, Davidson, & Bates, 2008). The dependent variable was the *Id*/*Diff* response. The predictors were the following: Competence in Spanish (*Advanced, No Knowledge*), Pair member (PP and P for PP>P stimuli; P and O for P>O stimuli), Lexical status (*Words, Pseudowords*) and Manipulation. The control variables were the presentation order of the pair (*Manipulated-Base, Base-Manipulated*) and the presentation blocks.

Participants and trials were entered as random variables. The significance of the main effects and interactions was assessed with likelihood ratio tests that compared the model with the main effect or interaction to a model without it. For clarity's sake, the results and figures are presented in percentages, although all statistical analyses have been performed on raw data (*Id/Diff* responses).

Considering, for example, the PP>P stimuli, an effect of pair member may be interpreted as follows according to the direction of the effect: 1) The manipulation triggers less “Different” (*Diff*) responses when the manipulated stimulus (PP>P) is paired with a PP stimulus than when paired with a P stimulus, meaning that the manipulation does not induce the perception of a stress shift. For example, the manipulated stimulus *valido* (PP>P) presents 10% of *Diff* responses when it is paired with the PP stimulus *válido* (PP>P paired with PP), whereas the same manipulated stimulus presents 90% of *Diff* responses when it is paired with the P stimulus *valido* (PP>P paired with P). 2) The manipulation triggers more *Diff* responses when the manipulated stimulus (PP>P) is paired with a PP stimulus than when paired with a P stimulus, meaning that the manipulation induces the perception of a stress shift. For example, the manipulated stimulus *valido* (PP>P) presents 90% of *Diff* responses when it is paired with the PP stimulus *válido*, whereas the same manipulated stimulus presents 10% of *Diff* responses when it is paired with the P stimulus *valido*. 3) The manipulation triggers the same number of *Diff* responses when the manipulated stimulus (PP>P) is paired with a PP stimulus than when paired with a P stimulus, meaning that the manipulation “does something, but not enough” for the stress shift to be clearly perceived. For example, the manipulated stimulus *valido* (PP>P) presents 60% of *Diff* responses when it is paired with the PP stimulus *válido*, and the same manipulated stimulus also presents 60% of *Diff* responses when it is paired with the P stimulus *valido*.

Results and discussion

Control trials

The participants' performance was between 95.24% and 100% of *Identical* responses for the trials composed of identical elements, which indicates that they performed the task properly.

PP>P Manipulated stimuli

As far as the *Id/Diff* responses are concerned, since the control variables (i.e., presentation order and blocks) showed no effect, they were removed from the model. The lexical status was also removed from the model, since it showed no effect and did not interact with other variables. Given the presence of the three-way interaction Competence x Pair member x Manipulation, we ran separate analysis for each manipulation, in order to determine whether the manipulation induces the perception of a stress shift (i.e., presence of an effect of pair member), and in order to examine the difference between the advanced participants and the participants with no knowledge of Spanish.

Manipulation of duration

Regarding the isolated manipulation of duration (see Figure 3), we observe an effect of Pair member, with more *Diff* responses when the manipulated stimulus was paired with P (90.23%) than when it was paired with PP (16.04%) ($\chi^2(1) = 23.01$, $p < .001$), which indicates that the manipulation of duration does not seem to induce the perception of a stress shift. Then, the results show an effect of Competence (Advanced = 50.38% and No Knowledge = 55.89%; $\chi^2(1) = 4.37$, $p < .05$), but no interaction Pair Member x Competence ($\chi^2(1) = 2.64$, *ns*).

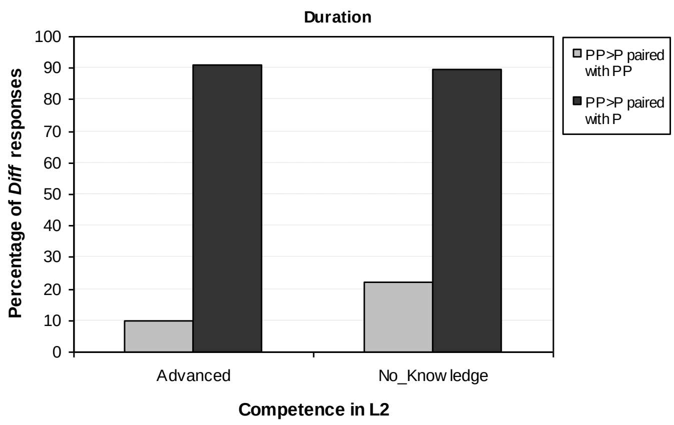


Figure 3. Percentage of *Different* responses as a function of the pair member (PP>P paired with PP, PP>P paired with P) and the competence in L2 (Advanced, No Knowledge) for the isolated manipulation of duration.

Manipulation of f_0

As far as the isolated manipulation of f_0 is concerned, we observe no effect of Pair member ($\chi^2(1) = 0.01$, *ns*), no effect of Competence ($\chi^2(1) = 0.19$, *ns*), and no interaction between both variables ($\chi^2(1) = 1.59$, *ns*). As can be seen in Figure 2, the manipulation of f_0 alone does not clearly induce the perception of a stress shift (67.66% of *Diff* responses for "PP>P paired with PP") and 61.62% for "PP>P paired with P"). Nevertheless, it "does something", although not sufficiently to clear-cut the perception between the PP and P stimuli.

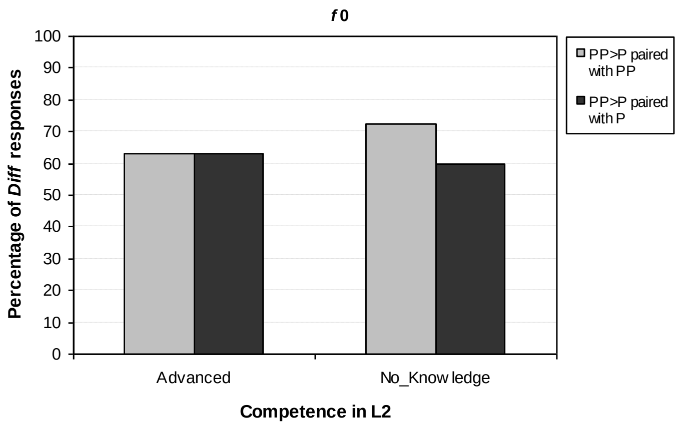


Figure 4. Percentage of *Different* responses as a function of the pair member (PP>P paired with PP, PP>P paired with P) and the competence in L2 (Advanced, No Knowledge) for the isolated manipulation of f_0.

Manipulation of intensity

With regard to the isolated manipulation of intensity (see Figure 5), a clear effect of the Pair member is observed, with more *Diff* responses for "PP>P paired with P" (95.28%) than for "PP>P paired with PP" (0.65%) ($\chi^2(1) = 41.81$, $p < .001$), which indicates that the manipulation of intensity alone does not induce the perception of a stress shift. Moreover, no effect of Competence ($\chi^2(1) = 1.29$, *ns*) and no interaction between both variables ($\chi^2(1) = 2.36$, *ns*) are noted.

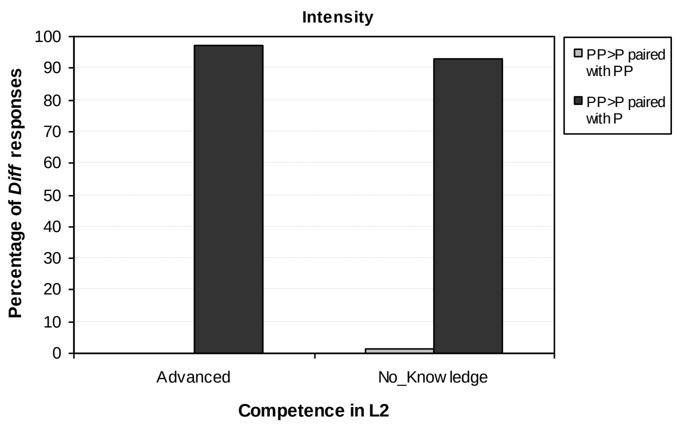


Figure 5. Percentage of *Different* responses as a function of the pair member (PP>P paired with PP, PP>P paired with P) and the competence in L2 (Advanced, No Knowledge) for the isolated manipulation of intensity.

Manipulation of duration and intensity

As for the combined manipulation of duration and intensity (see Figure 4), the results show an effect of the Pair member, with more *Diff* responses for "PP>P paired with P" (84.27%) than for "PP>P paired with PP" (17.21%) ($\chi^2(1) = 22.60$, $p < .001$). Thus, the manipulation of duration and intensity does not induce the perception of a stress shift. An effect of Competence is observed (Advanced = 48.76% and No Knowledge = 52.73%; $\chi^2(1) = 10.05$, $p < .01$), as well as an interaction between the Pair member and the Competence is present ($\chi^2(1) = 12.12$, $p < .001$): the participants with no knowledge of Spanish give more *Diff* responses than the advanced participants when the stimulus is paired with PP stimuli. In that sense, the former are more sensitive to the combined manipulation of duration and intensity than the latter.

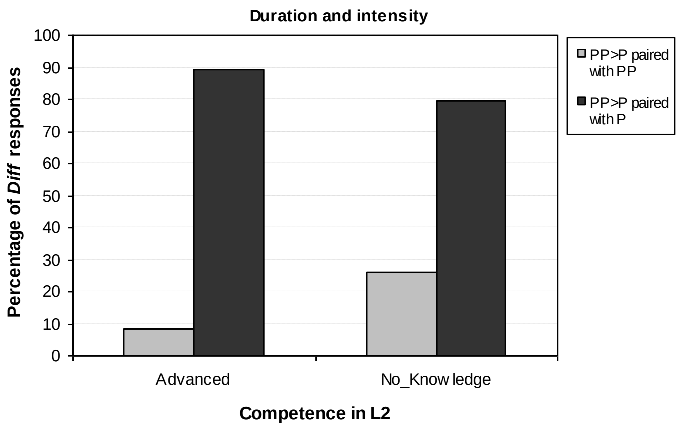


Figure 6. Percentage of *Different* responses as a function of the pair member (PP>P paired with PP, PP>P paired with P) and the competence in L2 (Advanced, No Knowledge) for the combined manipulation of duration and intensity.

Manipulation of f_0 and duration

As for the combined manipulation of f_0 and duration (see Figure 7), the results show an effect of Pair member, with more responses *Diff* for "PP>P paired with PP" (91.81%) than for "PP>P paired with P" (34.21%) ($\chi^2(1) = 22.21$, $p < .0001$). Therefore, the combined manipulation of f_0 and duration does induce the perception of a stress shift. No significant effect of Competence is observed ($\chi^2(1) = 0.97$, *ns*), although there are more *Diff* responses for the participants with no knowledge of Spanish (70.16%) than for the advanced participants (55.86%). Despite the smaller difference between "PP>P paired with PP" and "PP>P paired with P" in participants without knowledge than in advanced participants, no significant interaction is observed ($\chi^2(1) = 0.66$, *ns*).

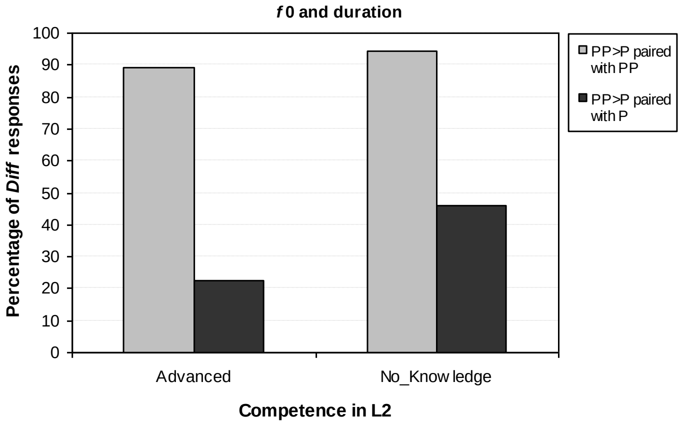


Figure 7. Percentage of *Different* responses as a function of the pair member (PP>P paired with PP, PP>P paired with P) and the competence in L2 (Advanced, No Knowledge) for the combined manipulation of f_0 and duration.

Manipulation of f_0 and intensity

Regarding the combined manipulation of f_0 and intensity, no effect of Pair member is observed ($\chi^2(1) = 1.00$, *ns*), in spite of the difference that can be noted in Figure 8 (77.87% of *Diff* response for the "PP>P paired with PP" and 59.87% for "PP>P paired with P"). Like in the case of the isolated manipulation of f_0, it seems, thus, that the combined manipulation of f_0, and intensity "does something", but not sufficiently to clear-cut the perception between the PP and P stimuli. Moreover, results show no effect of Competence ($\chi^2(1) = 1.21$, *ns*) and no interaction Pair Member x Competence ($\chi^2(1) = 0.24$, *ns*).

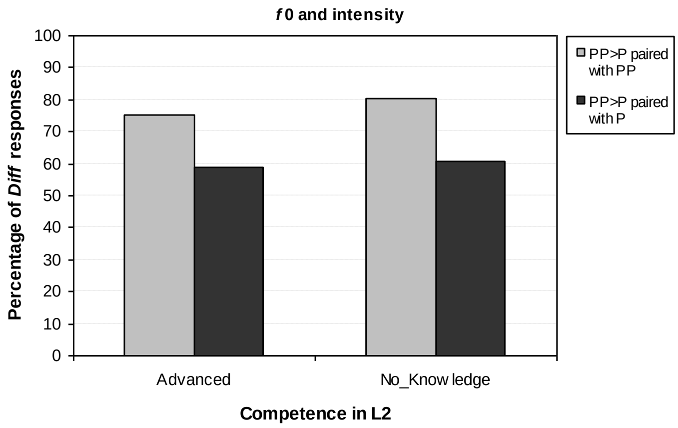


Figure 8. Percentage of *Different* responses as a function of the pair member (PP>P paired with PP, PP>P paired with P) and the competence in L2 (Advanced, No Knowledge) for the combined manipulation of f_0 and intensity.

Manipulation of f_0, duration and intensity

Finally, as for the combined manipulation of the three parameters (Figure 9), an effect of Pair member is observed ($\chi^2(1) = 31.53$, $p < .001$), with more responses *Diff* for "PP>P paired with PP" (95.80%) than for "PP>P paired with P" (32.72%). Therefore, as expected, this manipulation induces the perception of a stress shift. Moreover, no effect of Competence is noted ($\chi^2(1) = 0.01$, *ns*), although we observe more *Diff* responses for the participants without knowledge (70.44%) than for the advanced participants (58.09%). Moreover, the participants with no knowledge, in comparison with advanced participants, present a smaller difference between "PP>P paired with PP" and "PP>P paired with P" ($\chi^2(1) = 1.71$, $p < .01$).

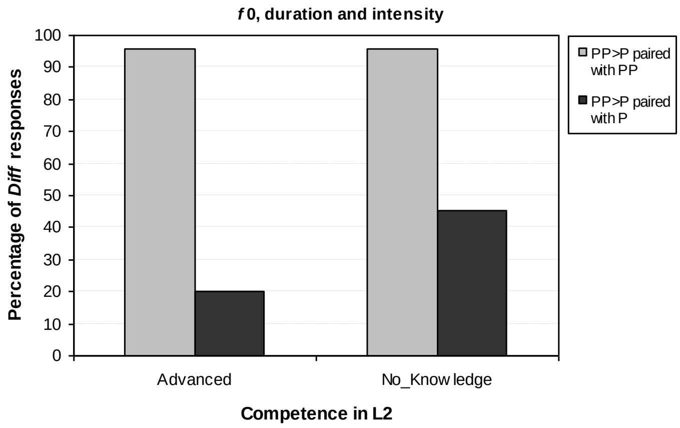


Figure 9. Percentage of *Different* responses as a function of the pair member (PP>P paired with PP, PP>P paired with P) and the competence in L2 (Advanced, No Knowledge) for the combined manipulation of f_0, duration and intensity.

Summary

In summary, the manipulation of duration and intensity, in isolation or in combination, does not trigger the perception of a stress shift in the case of PP>P stimuli. The manipulation of f_0, alone or with intensity, seems to "do something", but no sufficiently to clear-cut the perception of the stimulus as being different from the stimulus with the original or the shifted stress pattern. The role of the intensity seems minor, since it does not "help" f_0. On the other hand, the combined manipulation of f_0 and duration triggers the perception of the stress shift, with or without intensity. The differences between the advanced participants and the participants with no knowledge of Spanish are mainly observed when the manipulation involves duration. It seems that the participants with no knowledge are more sensitive to the manipulation of this parameter than the advanced participants.

P>O Manipulated stimuli

Given the presence of the three-way interaction Pair member x Manipulation x Competence, we ran separate analysis for each manipulation, in order to determine whether the manipulation induces the perception of a stress shift (i.e., presence of the effect of the Pair member), and in order to examine the difference between the advanced participants and the participants with no knowledge in Spanish. Since lexical status was not involved in the three-way interaction with competence, it was not included in further analyses. Regarding the control variables, whereas Block showed no effect and was removed from the analyses, the presentation order within the pair has a significant effect (i.e. more *Diff* responses for the *Base-Manipulated* than for *Manipulated-Base*) and was included in further analyses, although it will not be discussed in this paper.

Manipulation of duration

Regarding the isolated manipulation of duration (see Figure 10), we observe an effect of Pair member, with more *Diff* responses for "P>O paired with O" (79.49%) than for "P>O paired with P" (37.17%) ($\chi^2(1) = 6.71$, $p < .01$), which indicates that the manipulation of duration does not seem to induce the perception of a stress shift. Then, the results show no effect of Competence (($\chi^2(1) = 0.01$, *ns*) and no interaction Pair Member x Competence ($\chi^2(1) = 1.60$, *ns*).

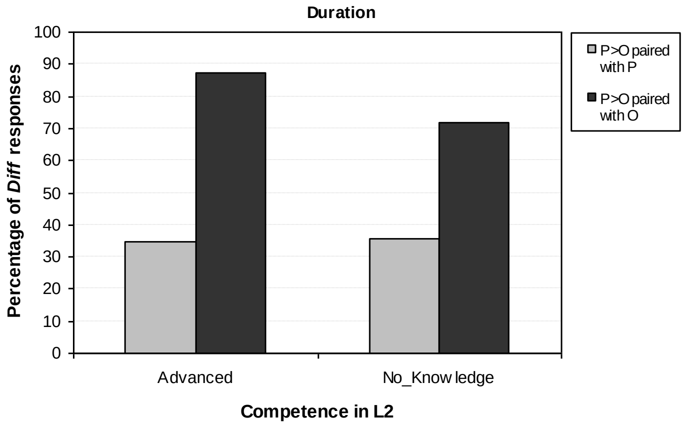


Figure 10. Percentage of *Different* responses as a function of the pair member (P>O paired with P, P>O paired with O) and the competence in L2 (Advanced, No Knowledge) for the isolated manipulation of duration.

Manipulation of f_0

As far as the isolated manipulation of f_0 is concerned (see Figure 11), we observe an effect of Pair member ($\chi^2(1) = 9.22$, $p < .01$) with more *Diff* responses for "P>O paired with O" (70.31%) than for "P>O paired with P" (39.63%). Moreover, no effect of Competence ($\chi^2(1) = 0.00$, *ns*) and no interaction between both variables ($\chi^2(1) = 2.76$, *ns*) were observed. These results indicate that the manipulation of f_0 alone does not trigger the perception of a stress shift.

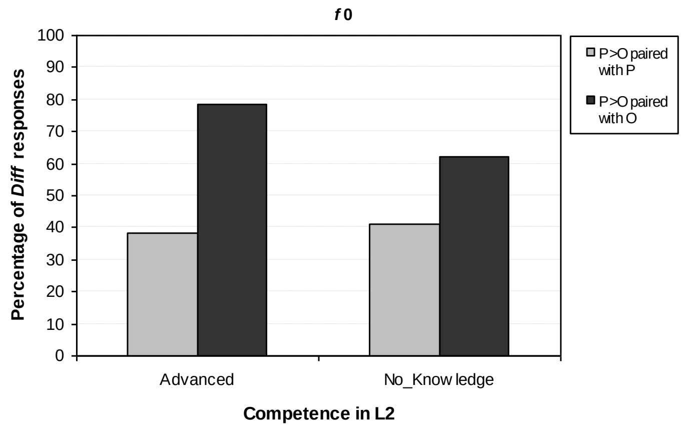


Figure 11. Percentage of *Different* responses as a function of the pair member (P>O paired with P, P>O paired with O) and the competence in L2 (Advanced, No Knowledge) for the isolated manipulation of f_0.

Manipulation of intensity

With regard to the isolated manipulation of intensity (see Figure 12), an effect of the Pair member is observed ($\chi^2(1) = 14.18$, $p < .001$), with more *Diff* responses for "P>O paired with O" (91.42%) than for "P>O paired with P" (22.50%), which indicates that the manipulation of intensity alone does not induce the perception of a stress shift. Moreover, no effect of Competence ($\chi^2(1) = 1.25$, *ns*) and no interaction between both variables ($\chi^2(1) = 0.25$, *ns*) are noted.

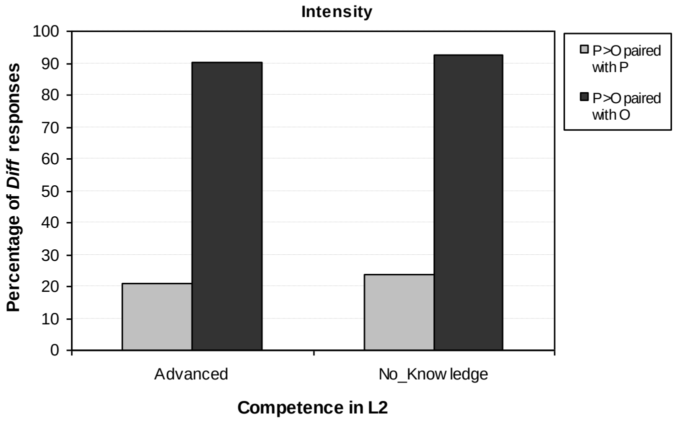


Figure 12. Percentage of *Different* responses as a function of the pair member (P>O paired with P, P>O paired with O) and the competence in L2 (Advanced, No Knowledge) for the isolated manipulation of intensity.

Manipulation of duration and intensity

As for the combined manipulation of duration and intensity (see Figure 13), the results show an effect of the Pair member, with more *Diff* responses for "P>O paired with O" (71.35%) than for "P>O paired with P" (36.98%) ($\chi^2(1) = 8.32$, $p < .01$). Thus, the manipulation of duration and intensity does not induce the perception of a stress shift. No effect of Competence is observed ($\chi^2(1) = 1.33$, *ns*), but a marginal interaction between the Pair member and the Competence (($\chi^2(1) = 3.17$, $p = .08$) has been found. The participants with no knowledge gave less *Diff* responses (44.6%) than the advanced participants (63.73%), especially when the manipulated stimulus was paired with an O stimulus. Participants without knowledge seem thus to be less sensitive to this manipulation than the advanced participants.

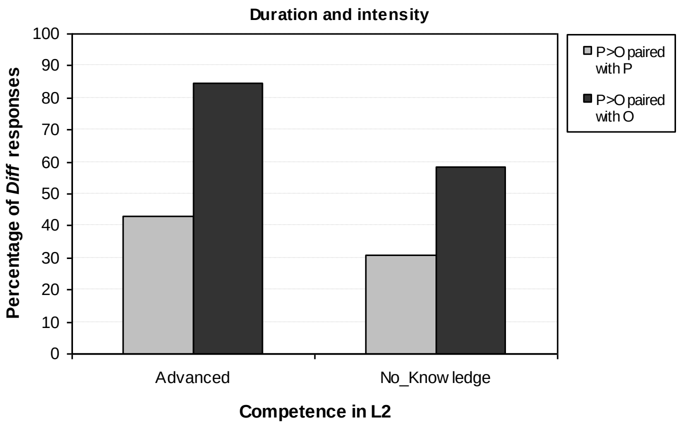


Figure 13. Percentage of *Different* responses as a function of the pair member (P>O paired with P, P>O paired with O) and the competence in L2 (Advanced, No Knowledge) for the combined manipulation of duration and intensity.

Manipulation of f_0 and duration

As far as the combined manipulation of f_0 and duration is concerned, the results show an effect of Pair member, with more *Diff* responses for "P>O paired with P" (64.91%) than for "P>O paired with O" (40.57%) ($\chi^2(1) = 6.35$, $p < .05$). Therefore, the combined manipulation of f_0 and duration induces the perception of a stress shift. An effect of Competence is observed ($\chi^2(1) = 6.09$, $p < .05$), as well as an interaction Pair Member x Competence ($\chi^2(1) = 7.60$, $p < .01$). As can be seen in Figure 14, the advanced participants present a greater difference between “P>O paired with P” and “P>O paired with O” stimuli than the participants with no knowledge.

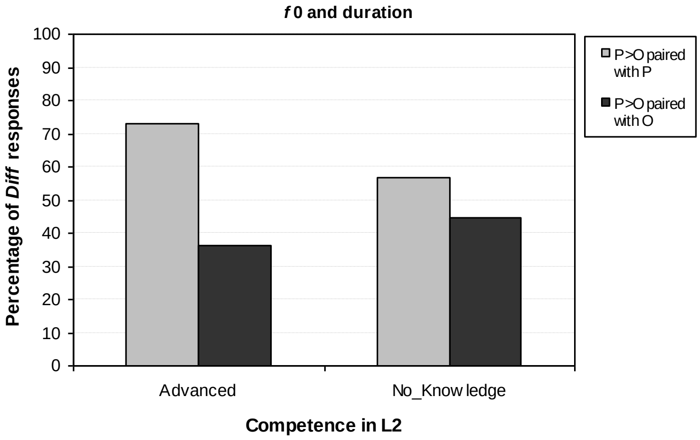


Figure 14. Percentage of *Different* responses as a function of the pair member (P>O paired with P, P>O paired with O) and the competence in L2 (Advanced, No Knowledge) for the combined manipulation of f_0 and duration.

Manipulation of f_0 and intensity

Regarding the combined manipulation of f_0 and intensity, no effect of Pair member ($\chi^2(1) = 0.84$, *ns*) and no effect of Competence ($\chi^2(1) = 0.32$, *ns*) are observed. An interaction between Pair Member and Competence is however present ($\chi^2(1) = 5.51$, $p < .05$). As can be seen in Figure 15, the Pair member effect goes in different direction in the advanced participants and in the participants with no knowledge. The former tend to perceive more differences when the manipulated stimulus is paired with the stimulus with the original pattern (“P>O paired with P”), while the participants without knowledge perceive more differences when the manipulated stimulus is paired with the stimulus with the shifted pattern (“P>O paired with O”).

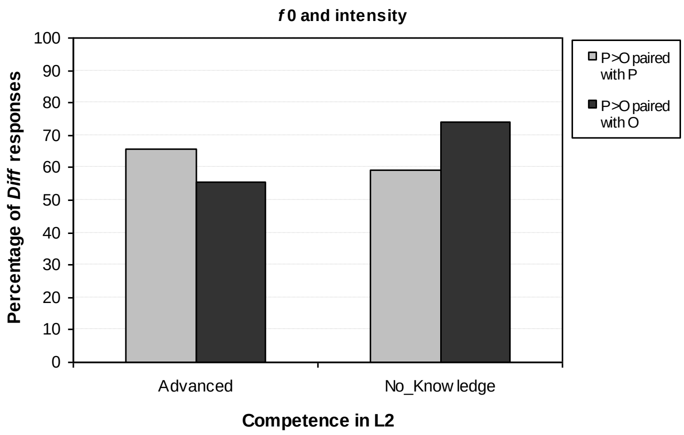


Figure 15. Percentage of *Different* responses as a function of the pair member (P>O paired with P, P>O paired with O) and the competence in L2 (Advanced, No Knowledge) for the combined manipulation of f_0 and intensity.

Manipulation of f_0, duration and intensity

Finally, as for the combined manipulation of the three parameters, an effect of Pair member is observed ($\chi^2(1) = 8.32$, $p < .001$), with more *Diff* responses for “P>O paired with P” (74.97%) than for

"P>O paired with O" (26.90%). Therefore, as expected, this manipulation induces the perception of a stress shift. An effect of Competence is noted ($\chi^2(1) = 1.33$, $p < .001$), as well as a marginal interaction between the Pair Member and the Competence ($\chi^2(1) = 3.17$, $p = .08$). As can be seen in Figure 14, the difference between "P>O paired with P" and "P>O paired with O" is smaller in the participants with no knowledge than in the advanced participants, which might suggest that the participants without knowledge are less sensitive to this manipulation than the advanced participants.

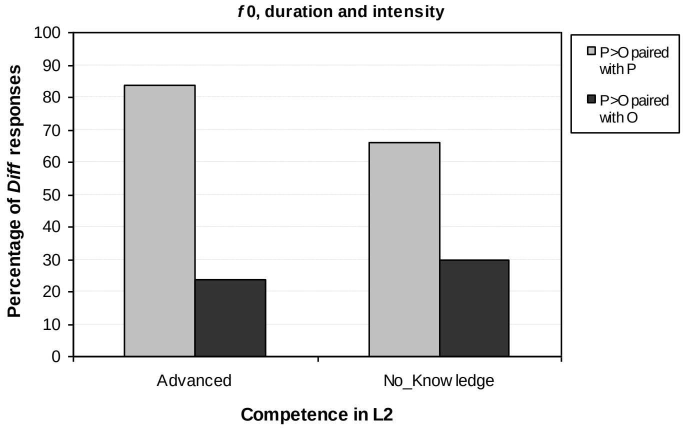


Figure 16. Percentage of *Different* responses as a function of the pair member (P>O paired with P, P>O paired with O) and the competence in L2 (Advanced, No Knowledge) for the combined manipulation of f_0, duration and intensity.

Summary

In summary, the combined manipulation of f_0 and duration, with or without intensity, clearly triggers the perception of a stress shift in P>O stimuli. The isolated manipulation of f_0, duration or intensity, as well as the combined manipulation of duration and intensity do not cause the perception of a stress shift. The combined manipulation of f_0 and intensity seems to "do something", but no sufficiently to clear-cut the perception of the stimulus as being different from the stimulus with the original or the shifted stress pattern.

Conclusion

In PP>P (e.g., *válido* manipulated using the values from *valido*) and in P>O (e.g., *valido* manipulated using the values from *validó*), f_0 seems to play the most important role in the perception of a stress shift, especially when combined with duration, whereas intensity plays a minor role. The main difference between the two accentual patterns resides in the isolated manipulation of f_0. While f_0 alone does not induce the perception of a stress shift in PP>P stimuli, it seems to "do something" in P>O stimuli, but not enough to clear-cut the perception of a stress shift. On the overall, the results from the discrimination test confirm the findings of a previous experiment in which an identification task was used (Schwab & Llisterri, 2010, 2011b).

The differences between the advanced participants and the participants without knowledge of Spanish mainly concern the role of duration, but they present an opposite trend in PP>P and P>O. Whereas it seems that the participants with no knowledge tend to be more sensitive to duration than advanced participants in PP>P stimuli, they are less sensitive in the case of P>O. This might be explained by the expectations that the participants with no knowledge might have from the French accentuation. As French stress is realized on the final syllable with an important lengthening (Léon, 2011), the participants without knowledge, not used to the phonetic realization of stress in Spanish, might have been less sensitive to duration in P>O than the advanced participants, because the lengthening of the final syllable in the Spanish stimuli was not as important as it would be in French.

To summarize, this investigation supports the idea that the perception of an accentual difference depends on the acoustic parameters used in the realization of the stress shift. More specifically, it has been shown that the role of the acoustic parameters varies as a function of the accentual patterns (PP>P and P>O) and the competence in L2. However, further work is needed to assess the effects of increasing the phonetic variability of the stimuli with the introduction of more voices and to explore the perception of lexical stress in words in context.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. doi:10.1016/j.jml.2007.12.005

Billières, M. (1988). Crible phonique, crible psychologique et intégration phonétique en langue seconde. *Travaux de Didactique du Français Langue Étrangère*, *19*, 5-30.

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer software]. Amsterdam: Department of Language and Literature, University of Amsterdam. Retrieved from http://www.praat.org

Borrell, A., & Salsignac, J. (2002). Importance de la prosodie en didactique des langues (application au FLE). In R. Renard (ed.), *Apprentissage d'une langue étrangère / seconde. 2. La phonétique verbo-tonale* (pp. 163-182). Paris: De Boeck Université.

Carton, F. (1974). *Introduction à la phonétique du français*. Paris: Bordas.

Dolbec, J., & Santi, S. (1995). Effets du filtre linguistique sur la perception de l'accent: Étude exploratoire. *TIPA, Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence*, *16*, 42-60.

Dupoux, E., Pallier, C., Sebastián Gallés, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, *36*(3), 406-421. doi:10.1006/jmla.1996.2500

Dupoux, E., Peperkamp, S., & Sebastián Gallés, N. (2001). A robust method to study stress "deafness." *The Journal of the Acoustical Society of America*, *110*(3), 1606-1618. doi:10.1121/1.1380437

Dupoux, E., Sebastián Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": the case of French learners of Spanish. *Cognition*, *106*(2), 682-706. doi:10.1016/j.cognition.2007.04.001

Enríquez, E., Casado, C., & Santos, A. de. (1989). La percepción del acento en español. *Lingüística Española Actual*, *11*, 241-269.

Forster, J. C. (2012). DMDX Updates Page [Computer software]. Tucson, AZ: Department of Psychology, University of Arizona. Retrieved from http://www.u.arizona.edu/~jforster/dmdx.htm

Frost, D. (2010). La surdité accentuelle: d'où vient-elle et comment la guérir? *Cahiers de l'APLIUT*, *29*(2), 24-43. doi:10.4000/apliut.684

Garde, P. (1968). *L'accent*. Paris: Presses Universitaires de France.

Gil, J. (2007). *Fonética para profesores de español: de la teoría a la práctica*. Madrid: Arco/Libros.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests in linear mixed effects models. R package version 2.0-20 [Computer software]. Retrieved from https://cran.r-project.org/web/packages/lmerTest/index.html

Léon, P. (2011). *Phonétisme et prononciations du français. Avec travaux pratiques d'application et corrigés* (6ème ed.). Paris: Armand Colin.

Léon, P., & Martin, P. (2000). Prosodie et technologie. In E. Guimbretière (ed.), *Apprendre, enseigner, acquérir: la prosodie au coeur du débat* (pp. 135-150). Rouen: Publications de l'Université de Rouen.

Llisterri, J., Machuca, M. J., Mota, C. de la, Riera, M., & Ríos, A. (2005). La percepción del acento léxico en español. In *Filología y lingüística. Estudios ofrecidos a Antonio Quilis* (Vol. 1, pp. 271-297). Madrid: Consejo Superior de Investigaciones Científicas - Universidad Nacional de Educación a Distancia - Universidad de Valladolid.

Mora, E., Courtois, F., & Cavé, C. (1997). Étude comparative de la perception par des sujets francophones et hispanophones de l'accent lexical en espagnol. *Revue PArole*, *1*, 75-86.

Muñoz García, M. (2010). *La perception et la production de l'accent lexical de l'espagnol par des francophones: aspects phonétiques et psycholinguistiques* (Thèse de doctorat). Université de Toulouse 2 Le Mirail - Universitat Autònoma de Barcelona.

Muñoz García, M., Panissal, N., Billières, M., & Baqué, L. (2009). Substance prosodique de l'accent espagnol: test expérimental en perception et en production d'une langue étrangère. In *Actes des 6èmes Journées d'Études Linguistiques* (pp. 15-20). Nantes, France. 18-19 Juin 2009.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress "deafness." In C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology 7* (pp. 203-240). Berlin: Mouton de Gruyter.

Polivanov, E. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague, 4: Réunion Phonologique Internationale Tenue à Prague (18-21/ XII 1930), 4,* 79-96.

Quilis, A. (1981). *Fonética acústica de la lengua española*. Madrid: Gredos.
Quilis, A. (1993). *Tratado de fonología y fonética españolas*. Madrid: Gredos.
R Core Team. (2014). R: A Language and Environment for Statistical Computing. Version 3.1.3. [Computer software] Vienna: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org
Rico, J. (2012). El acento y la sílaba en la clase de ELE. In J. Gil (ed.), *Aproximación a la enseñanza de la pronunciación en el aula de español* (pp. 75-92). Madrid: Edinumen.
Rigault, A. (1962). Rôle de la fréquence, de l'intensité et de la durée vocalique dans la perception de l'accent en français. In A. Sovijärvi & P. Aalto (eds.), *Proceedings of the 4th International Congress of Phonetic Sciences* (pp. 735-748). The Hague: Mouton.
Rossi, M. (1979). Le français, langue sans accent? *Studia Phonetica*, *15*, 13-51.
Salsignac, J. (1998). Perception de l'accent primaire de langues étrangères: Présentation d'une étude expérimentale. *La Linguistique*, *34*(1), 65-72. doi:10.2307/30249134
Schwab, S., & Llisterri, J. (2010). La perception de l'accent lexical espagnol par des apprenants francophones. In L. Baqué & M. Estrada (eds.), *La langue et l'être communiquant. Hommage à Julio Murillo* (pp. 311-328). Mons: Éditions du CIPA.
Schwab, S., & Llisterri, J. (2011a). Are French speakers able to learn to perceive lexical stress contrasts? In W.-S. Lee & E. Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1774-1777). Hong Kong, China. 17-21 August, 2011.
Schwab, S., & Llisterri, J. (2011b). The perception of Spanish lexical stress by French speakers: Stress identification and time cost. In K. Dziubalska-Kołaczyk, M. Wrembel & M. Kul (eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010* (Vol. 1, pp. 229-242). Frankfurt am Main: Peter Lang.
Schwab, S., & Llisterri, J. (2012). The role of acoustic correlates of stress in the perception of Spanish accentual contrasts by French speakers. In Q. Ma, H. Ding, & D. Hirst (eds.), *Proceedings of the 6th International Conference on Speech Prosody* (Vol. 1, pp. 350-353). Shanghai: Tongji University Press.
Schwab, S., & Llisterri, J. (2014). Does training make French speakers more able to identify lexical stress? *COPAL, Concordia Working Papers in Applied Linguistics*, *5*, 624-636.
Troubetzkoy, N. S. (1949). *Principes de phonologie*. (J. Cantineau, Trans.). Paris: Klincksieck. (Original work published 1939)
Vaissière, J. (1990). Rhythm, accentuation and final lengthening in French. In J. Sundberg (ed.), *Music, language, speech and brain* (pp. 108-121). New York: Macmillan.

Consonant harmony in children acquiring Farsi; typical vs. atypical phonological development

Froogh Shooshtaryzadeh[1], Pramod Pandey[2]
fshude112008@yahoo.com, pkspandey@yahoo.com

[1]Imam Khomeini International University, [2]Jawaharlal Nehru University

Abstract. This paper aims to examine place and manner harmony in children with typical and atypical phonological development who are acquiring Farsi and to compare the findings from this study with findings from similar studies on harmony in other languages. To collect data, 5 children with typical phonological development (ages: 2;8 to 4;0) and 5 children diagnosed with functional (non-organic) phonological disorder (ages: 4;5 to 5;9) were tested with a picture-naming task. During this, children should have produced 132 different names elicited by 132 pictures of items generally encountered in children's daily life, such as food, animals, and things. The data were complemented by a 15-30 minutes free recording of children's spontaneous speech. The primary examination of the data indicated some similarities and differences in harmony patterns in PD and TD children. Both groups showed a large number of manner-harmony instances and a small number of place-harmony instances. However, the two groups displayed differences in the types of place and manner harmony. Moreover, the comparison of the results of this study with results in similar studies on children acquiring other languages has demonstrated some significant differences. Contrary to findings in earlier studies that have indicated assimilation of coronals to dorsals in place harmony (Smith, 1973, Stoel-Gammon & Stemberger, 1994; Fikkert & Levelt, 2003; Gerlach, 2010), this study has found assimilation of dorsals to coronals and labials. Consideration of the results here within the Optimality Theory (OT) framework (Prince & Smolensky, 1993; McCarthy & Prince, 1994, 1995) shows that constraints, relating to harmony processes observed in other languages, are also present in Farsi; however, rankings differ in children acquiring Farsi. Furthermore, the findings of this study create doubts about the universality of PARSEDOR >> PARSECOR claimed by Goad (1997) and the cross-linguistic dominancy of dorsals over coronals in place harmony (Kiparsky, 1994). Eventually, considering our findings on manner harmony, in view of Wrights' approach (2001, 2004) to articulatory and perceptual characteristics of phonemic categories, has led to the conclusion that perceptual factors can also trigger harmony processes when articulatory limitations are lessened or removed. This study can lead to better understanding of phonological acquisition processes in Persian children and can shed light on the problems of children with phonological disorder, whch accordingly can help clinicians to come up with better intervention strategies.

Keywords: consonant harmony, typical phonological development, phonological disorder, Farsi

Introduction

Consonant harmony, or long-distance assimilation, is a process in which the articulatory characteristics of a consonant in one part of the word can affect the articulation of consonants in other parts of the word. The majority of studies on harmony in other languages has focused on place harmony and concluded that coronals are more likely to be the target of place harmony, while velars and labials are more likely to trigger harmony (Fikkert & Levelt, 2003; Fikkert, 2000; Gerlach, 2010; Pater & Werle, 2003; Smith, 1973; Stemberger & Bernhardt, 1997; Stoel-Gammon & Stemberger, 1994). There are also some studies on manner harmony (Dinnsen & Barlow, 1998; Dinnsen & O'Connor, 2001; Dinnsen, 1998; Vihman, 1978), which have discussed nasal and fricative harmony that targets glides and obstruent stops. Dinnsen (1998) has argued that when [continuant], [nasal], or [approximant] trigger the harmony, plosives and glides can be the targets of harmony. Moreover, Dinnsen and O'Connor (2001) claimed that various types of manner harmony indicate different limitations on what can serve as a target. This study aims to examine the above claims about place and manner harmony in the Typically Developing (TD) phonologies and in Phonological Disorder (PD) in children acquiring Farsi as their first language.

Method

Participants

The participants in this qualitative cross-sectional study are 5 children (3 girls and 2 boys) diagnosed with functional phonological disorder (PD) ranging in age from 4;6 years to 6;0 years, and five typically developing children (2 girls and 3 boys) ranging in age from 2;6 years to 4;0 years. The age difference is because a child is generally considered phonologically disordered if s/he remains unintelligible after 4 years old, a time when typically developing children are generally intelligible to strangers (e.g., Adams, Byers, Brown, & Edwards, 1997). Before this age, even if they are unintelligible, children are not classified as having a phonological disorder. To identify PD children from normal children, the candidates were examined by different specialists; a speech therapist checked the children for any speech problems, an audiometer checked their hearing normality, and a psychologist checked their cognitive abilities and mental health. Also, the children's medical profiles were considered and their parents filled out related questionnaires. There were also interviews with parents and teachers. The results of all these inquiries indicated that the PD children in the study are physically and mentally healthy and their speech problem is a result of a functional/nonorganic phonological disorder. All children come from middle-class families. They are primarily monolingual and speak standard Farsi (Tehrani accent) in most domains at home and in schools.

Speech assessment tools

The children-participants in this study were tested with a picture-naming task, which was devised based on the requirements of this research and features of the Farsi language. The task contained pictures of 132 familiar objects that have elicited the spontaneous production of 132 target words. The picture-naming task contained a good number of all types of consonants, i.e. 167 plosives, 14 affricates, 108 fricatives, 75 nasals, 73 liquids, and 16 glides. Except for phoneme /ʔ/ that is found only in word initial and medial positions in Farsi, all other consonants occurred in initial, word-medial and final word positions based on Farsi phonotactics. The test also included words with 1-6 syllables of all different types licensed in Farsi, i.e. CV, CVC, and CVCC. Finally, the test also included simple, complex, and compound words.

Data recording

To collect the data, each child was given the necessary instructions concerning the test in simple language. Later, each picture was presented to the child separately, he/she was asked to produce the name of the picture, and their productions were recorded. Sometimes, data were collected from a child during two-three sessions depending on his/her age and cooperation in answering the questions. The data were recorded by means of a solid sound recorder (Samsung Voice Recorder YP-VP1). The entire recording was done in a quiet place. In addition, there were 15-30 minutes of free recording for each child instigating motivation through play and reading stories.

Data processing

The recorded data was carefully transcribed by three judges using IPA. To ensure reliability, a consensus method was used to confirm the sound between two of the three judges. Then, the errors in children's productions were examined closely to determine the phonetic and phonemic inventories of each child and settle on the real cases of consonant harmony. To separate context free substitutions (quasi-harmonic error) from real harmony errors, all productions resulting from phonetic limitations were deleted in the list. Such cases are mainly observed in the PD group that has problems in the articulation of some fricatives and/or back consonants. Table 1 displays the phones absent from PD and TD children's phonetic inventories.

It should be noticed that the Farsi phonemic inventory includes 6 vowels, i.e. /e, æ, o, i, a, u/ and 23 consonants, i.e. /b p t d s z ʧ ʤ g k q ʔ r ʃ x ʒ v f h m n l j/. Moreover, Farsi always begins with a consonant and lacks initial consonant clusters.

Table 1. Phones ‘absent’ from the phonetic inventory of PD and TD children

PD group		TD group	
Se	/ʒ ʧ ʤ/	EI	/ʒ/
Ti	/s ʃ ʒ/	AI	/ʒ/
Ze	/s ʃ ʒ k x q/	Sa	—
Me	/ʃ ʒ/	Ma	/ʒ/
Hi	—	Ro	—
Hi	—	Ro	—

Results

The data collected through the picture-naming task from TD and PD children were analyzed and errors related to consonant harmony were examined. Two main types of harmony errors were identified in both TD and PD children, namely place and manner harmony. There are 138 potential contexts for consonant harmony per child. As shown in Table 2, the maximum number of harmony instances are allocated to manner harmony, i.e. 47 errors in the TD group and 68 errors in the PD group, while place harmony errors comprise 16 errors in the TD and 17 errors in the PD groups. Moreover, three types of manner harmony were recognized in the data, i.e. plosive, nasal, and fricative harmony. Plosive and nasal harmony was observed in both groups; however, fricative harmony was merely detected in the TD group. The TD children exhibited 35% plosive harmony, 28% nasal harmony, and 28% fricative harmony in their manner harmony errors produced by TD children, and the PD group illustrated 80% and 16% plosive and nasal harmony errors in their productions. Figure 1 shows manner and place harmony errors in PD and TD groups (from younger to older children).

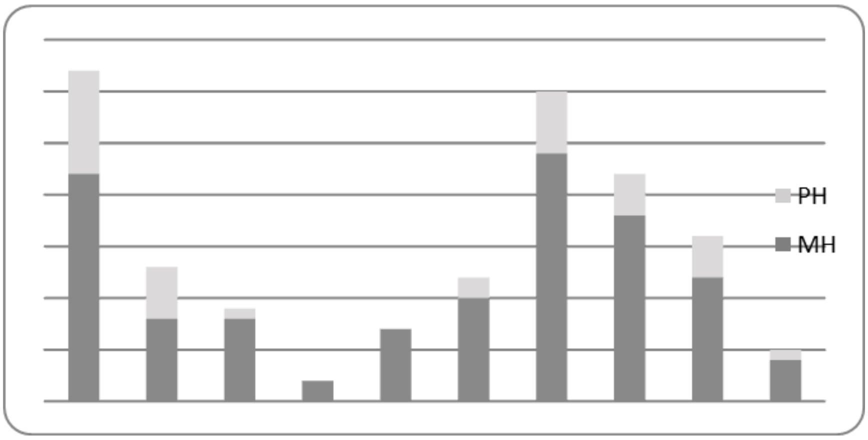


Figure 1: Manner harmony (MH) and place harmony (PH) errors in PD and TD groups

As the results have indicated for the TD children acquiring Farsi, dorsals generally harmonize to coronals or labials, and coronals harmonize to labials. For the PD children, dorsals generally harmonize to coronals, labials or the glottal stop/ʔ/, while coronals harmonize to labials. Also, a few instances of dorsal harmony are observed in a PD child. It should be noticed that those children that change dorsals to coronals and labials in harmony processes produce dorsals in other words and in different word positions.

Discussion

Place harmony

As mentioned in part 3, in TD children and most PD children dorsals harmonize to coronals or labials, and coronals harmonize to labials. Table 2 indicates some examples of place harmony errors produced by the TD children.

Table 2: Place harmony in El, Al and Sa (TD children)

	Target Word	Child Pronunciation	Gloss
El	*/guʃ/	[dus]	‘ear’
	*/guʃt/	[dust]	‘meat’
	*/gusfænd/	[dusfænd]	‘sheep’
	*/qarʧ/	[darʧ]	‘mushroom’
	*/qæza/	[dæza]	‘food’
	*/kuʧe/	[tuʧe]	‘lane’
	/mobl/	[momb]	‘sofa’
	/toxmomorq/	[momomoq]	‘egg’
	/ʔotobus/	[tobobus]	‘bus’
	/ʔæsb/	[bæsb]	‘horse’
Al	*/sæg/	[ʃæt]	‘dog’
	*/qaʃoq/	[gaʃod]	‘spoon’
	*/ mesvak/	[mestat]	‘toothbrush’
	*/hendune/	[ʃendune]	‘sofa’
	/qurbaqe/	[gowvave]	‘sheep’
Sa	/ketab/	[petap]	‘book’

As it is indicated in Table 2, most harmony errors in TD children (the examples identified with *) are related to dorsals which are harmonized to coronals, while both El and Al can produce the same dorsals in other words and contexts, as shown in Tables 3 and 4.

Table 3. Sounds targeted by harmony (g, q, k) produced correctly by El in other contexts

Target Word	Child Pronunciation	Gloss
/gol/	[gol]	‘flower’
/gav/	[gaf]	‘cow’
/ængur/	[nægo]	‘grape’
/xoʃgel/	[doʃgel]	‘pretty’
/ʧængal/	[nængal]	‘fork’
/gerje/	[gerje]	‘cry’
/tutfærængi/	[tutfærængi]	‘strawberr y’
/boʃqab/	[boʃqap]	‘plate’
/dæmaq/	[næmaq]	‘nose’
/qejʧi/	[qejʧi]	‘scissors’
/ʧaqu/	[daqu]	‘knife’
/kolah/	[kalah]	‘cap’
/kuh/	[kuh]	’mountain’

Table 4. The sounds targeted by harmony (g, q, k) produced correctly by Al in other contexts

Target Word	Child Pronunciation	Gloss
/ʧæng/	[ʧæng]	claw
/pælæng/	[pælæng]	tiger
/gævæzn/	[gævæs]	deer
/ ængoʃt/	[ængoʃ]	finger
/dæmaq/	[dæmaq]	nose
/dæsmalkaqæzi/	[dæsmalkaqæsi]	tissue
/ʤarubærqi/	[ʤarubæqi]	vacuum cleaner
/qofl/	[qofl]	lock
/badkonænk/	[badtonænk]	balloon
/toxmomorq/	[toxmok]	egg
/hævapejma/	[hæpejma]	airplane
/hæviʧ/	[hæviʧ]	carrot

Therefore, this study has illustrated that dorsals are targets while coronals and labials are triggers of place harmony in typically developing phonologies acquiring Farsi. This finding differs from those in other studies on harmony in typically developing children, mainly speaking English, where dorsals and labials are usually triggers, and coronals are targets of place harmony (Gerlach, 2010; Goad, 1997; Kiparsky, 1994; Pater & Werle, 2003; Pater, 2002; Smith, 1973). It seems that in Farsi, children prefer unmarked places (coronal and labial) to trigger harmony, and the more marked place (dorsal) to be the target of harmony.

However, place harmony observed in PD children is more complicated. As shown Table 5, the PD group exhibits fewer instances of place harmony, but with more variety. In this group, dorsals harmonize to coronals and labials, coronals harmonize to dorsals, and labials harmonize to coronals.

Table 5. Place harmony in PD children

	Target Word	Child Pronunciation	Gloss
Ti	/kælaq/	[dællaq]	crew
	/mesvak/	[pedtap]	toothbrush
	/xodkar/	[qoqkal]	pen
	/ʧængal/	[dæddal]	fork
	/bæstæni/	[dædtæni]	ice cream
	/dæsmal kaqæzi/	[dæbbal qaqædi]	tissue
	/sæg/	[gæk]	dog
Me	/zæbt/	[pæp]	recorder
	/maʃin/	[papin]	car
Ze	/ʤurab/	[tuwap]	socks
	/ʔænkæbut/	[ʔæbʔæput]	spider
	/tæxtexab/	[dæddebap]	bed
Se	/park/	[dark]	park
	/boʃqab/	[guʃga]	plate
Hi	/mesvak/	[mesbad]	toothbrush

It should be noted that /k, d, n, b/ that are targets of place harmony in Ti, and /p, /k/ that are targets of harmony in Se and Hi, can all be produced normally in other contexts by these children. Table 6 provides examples of the correct production of the sounds targeted in harmony by PD children in other contexts.

Table 6. Sounds targeted in harmony (/p, b, d, n, k/) produced correctly in other contexts

	Target Word	Child Pronunciation	Gloss
Ti	/dærja/	[dæra]	sea
	/kilid/	[kelid]	key
	/badkonæk/	[badtonæk]	baloon
	/ʔejnæk/	[ʔejnæk]	glass
	/kejk/	[gek]	cake
	/kif/	[kip]	bag
	/gorbe/	[gobe]	cat
	/moræbba/	[foʔæbba]	jam
	/pakkon/	[bakkun]	rubber
	/bærf/	[bæp]	snow
	/biskujit/	[biztuji]	biscuit
Se	/pelle/	[pelle]	stair
	/dæmpaji/	[dæmpaji]	slippery
	/lamp/	[lamp]	lamp
	/bæstæni/	[bædtæni]	ice cream
	/bærf/	[bærf]	snow
	/gorbe/	[ʔobe]	cat
	/tab/	[tab]	swing
Hi	/ʔænkæbut/	[ʔænkæbut]	spider
	/ʔejnæk/	[ʔejnæk]	glass
	/kilid/	[kilid]	key
	/badkonæk/	[badkonæk]	balloon

Moreover, some cases of coronal and labial harmony in PD children are not as straightforward as harmonies in TD children. For example, Ze cannot produce /x/ and, normally, in word medial and final positions produces it as coronal stops (d or t) while in the beginning of a word he produces it as with a glottal stop /ʔ/. However, in the word /tæxtexab/, though the first /x/ sound is produced as coronal stop (d) as usual, /x/ in the third syllable is harmonized with /b/ sound in syllable's coda and is labialized; therefore, [dæddebap] is produced for /tæxtexab/ instead of [dæddedap]. Similar cases of harmony have also occurred in other words produced by Ze, and also by Me, that shown in bold in Table 5. Me has problems in producing fricatives such as /z, ʃ/, and Ze has problems in producing /ʧ, k, r/. They both usually substitute these sounds with coronals (Table 7); however, in the presence of a labial, they are labialized (Table 5).

As it is observed in Table 5, PD children also exhibit instances of dorsal harmony as well as coronal and labial harmonies. The presence of dorsal harmony in PD children reminds of the presence of dorsal harmony in typically developing phonologies in other languages, such as Amahl's (Smith, 1973) and supports the claims of Optimality Theory (OT) (Prince & Smolensky, 1993; McCarthy & Prince, 1994, 1995). OT claims that constraints are universal; however, their ranking can be different in different languages. This approach can explain the cause of differences in the harmony patterns of Persian children with typical and atypical phonological development, and can also explain their differences and similarities with children acquiring other languages. Goad (1997) has explained Amahl's (Smith, 1973) consonant harmony patterns using OT by employing these constraints: ALIGNDORSAL, PARSEDORSAL, ALIGNCOR, and PARSECOR. Parse in the above constraints refers to a group of faithfulness constraints that needs the segments or features in the input to be parsed in the output. Therefore, these faithfulness constraints prefer candidates in which underlying elements have not been deleted. However, alignment represents a family of markedness constraints that requires a particular edge of a grammatical or prosodic category to match the particular edge of another grammatical or prosodic category (see McCarthy & Prince, 1993b for more details). To obtain the effect of harmony, Parse constraints should be ranked higher than alignment constraints for the same feature (see Goad, 1997 for discussion). To explain Amahl's harmony productions, Goad has suggested the following ranking for the above constraints (1997, p. 11):

PARSELAB, PARSEDOR >> ALIGNLAB, ALIGNDOR >> PARSECOR >> ALIGNCOR

Table 7: Sounds targeted in harmony (/z, ʃ/, /ʧ, k, r/) produced as coronal in other contexts

	Target Word	Child Pronunciation	Gloss
Me	/zænbur/	[tæbpur]	bee
	/mowz/	[xot]	banana
	/telvizijun/	[tevedun]	television
	/dæsmal kaqæzi/	[tædfan ?adædi]	tissue
	/medad ʃæmi/	[petad tæmi]	crayons
	/xærguʃ/	[?ætut]	rabbit
	/qaʃoq/	[?atud]	spoon
	/ʃotor/	[totoj]	camel
Ze	/radijo zæbt/	[dadijo dæhft]	recorder
	/?ænar/	[?ænaj]	pomegranate
	/zærrafe/	[dæ—jave]	giraffe
	/xorus/	[?ojut]	rooster
	/qejʧi/	[?etti]	scissors
	/ʧaqu/	[tato]	knife
	/moʧ/	[mut]	wrist
	/jæxʧal/	[tædtah]	refrigerator
	/park/	[pat]	park
	/pakkon/	[padton]	rubber
	/ʒakæt/	[dadæt]	jacket

Though the above constraint ranking can explain some PD children's harmony productions, another type of ranking is required to explain other PD children's and TD children's harmony productions. Tableaus 1, 2, and 3 illustrate the constraints and their ranking for a TD child, a PD child and Amahl (as in Goad, 1997), respectively. The sign ⇒ indicates the optimal output (i.e. the produced output). As displayed in the tableaus, the same constraints are present in all children's underlying grammar; however, the ranking can be different from one child (typical vs. atypical) to another or from one language (e.g., Farsi) to another language (e.g., English). The results of this study pose doubts about the universality of the ranking PARSEDOR >> PARSECOR claimed by Goad (1997) and the cross-linguistic dominance of dorsals over coronals in place harmony (Kiparsky, 1994). The similarity of the PD child's harmony pattern to Amahl's is also noticeable, indicating once more that all constraints are present and active in languages, even when their presence is hidden by the normal productions of speakers of a language.

Tableau 1. Place harmony in a TD child

Input: /guʃt/ 'meat'	ALIGNCOR	PARSEDOR
a. [gust]	*!	
b. ⇒ [dust]		*

Tableau 2. Place harmony in a PD child

Input: /xodkar/ 'pen'	ALIGNDOR	PARSECOR
⇒ [qoqkal]		*
[qodkal]	!*	

Tableau 3. Place harmony in Amahl (adapted from Goad, 1997)

Input: /stɔ:k/ 'stalk'	ALIGNDOR	PARSECOR
⇒ [gɔ:k]		*
[dɔ:k]	!*	

Manner harmony

As stated in section 3, there are three types of manner harmony, i.e. plosive, nasal, and fricative harmony. Plosive and nasal harmonies are observed in both groups. However, fricative harmony is observed only in the TD group. Tables 8 and 9 display some examples of manner harmony in the TD and PD groups receptively.

Table 8: Manner harmony in TD children

	Target Word	Child Pronunciation	Gloss
El	/kæfʃ/	[kæbʃ]	shoes
	/moræbba/	[roræbbas]	jam
	/mahi/	[vahi]	fish
	/lakpoʃt/	[dæfpos]	turtle
	/setare/	[dedare]	star
	/hævidʒ/	[hævis]	carrot
Al	/dændun/	[nændun]	tooth
	/mesvak/	[meStat]	tooth brush
	/gav/	[gab]	cow
	/ʒakæt/	[dʒadʒæt], [datæt]	jacket
	/ʧængal/	[ʧændʒal]	fork
	/dæhæn/	[næhæn]	mouth
	/saʔæt]	[ʃahæt]	o'clock
	/qurbaqe/	[govave]	frog
Sa	/ʃælvar/	[ʃælvah]	trousers
	/pærvane/	[pærbane]	butterfly
	/dæmaq/	[næmaq]	nose
	/toxmomorq/	[toxmomox]	egg
	/saʔæt/	[saxæt]	o'clock
	/qejʧi/	[qædʧi]	scissors
	/hævapejma/	[hæbapejma]	airplane
	/hæviʧ/	[hævis]	carrot'
Ro	/hævapeyma/	[hæmupeyma]	airplane
	/gusfænd/	[gusbænt]	sheep
	/pærvane/	[pærbane]	butterfly
	/kæfʃ/	[kæbʃ]	shoes
	/saʔæt/	[sahæt]	o'clock
	/gav/	[gap]	cow
	/dʒarubærqi/	[darubærqi]	vacuum cleaner

The manner harmony errors observed in the data from TD and PD children exhibit special characteristics that are worth mentioning. First, there are cases in which manner harmony happens between a substituted sound in the target word and the target of harmony, especially in PD group. In these cases, the target words are obtained in several steps. For example, Ti usually substitutes sounds like /t/ or /d/ for /s/, /ʃ/, /ʧ/ and this triggers manner harmony in some sounds. The following examples demonstrate the assumed processes in Ti:

	/mesvak/	**/doʧærxe/**
Substitution:	/s/→[d]: [medvak]	/ʧ/→[d]: [dodærxe]
Manner harmony:	/m/→[p]: [pedvak]	/x/→[k]: [dodærke]
Other processes:	[pedtak]	[todæxe]

Table 9. Manner harmonyin PD children

	Target Word	Child Pronunciation	Gloss
Sa	/ doʧærxe/	[dodæqen]	bicycle
	/ xærguʃ/	[?æ?uʃm]	rabbit
	/ qaʃoq/	[ʔaʃox]	spoon
	/ bæstæni/	[bæntæni]	ice cream
	/ zænbur/	[bænbuj]	bee
	/ sændæli/	[nændæli]	chair
	/sæbz/	[bæb]	green
Ti	/qejʧi/	[gedgin]	scissors
	/pærvane/	[bæmane]	butterfly
	/xodkar/	[qoqkal]	pen
	/radio zabt/	[dadijo dæ]	recorder
	/deræxt/	[dedævt]	tree
	/lakpoʃt/	[dabpot]	turtle
	/qofl/	[qufp]	lock
	/medad ʃæmi/	[bedad tæmi]	crayons
	/?ænar/	[?ænan]	pomegranate
	/van/	[man]	tub
	/livan/	[ʤiman]	glass
	/mesvak/	[pedtap]	toothbrush
	/maʃin/	[patin]	car
	/kuh/	[xuh]	mountain
	/doʧærxe/	[todæke]	bicycle
Ze	/medad ʃæmi/	[bedad tæmih]	crayons
	/dæmaq/	[dæbat]	nose
	/naxongir/	[tatodti]	nail sharpener
	/havapejma/	[?æbapejma]	airplane
	/ʃælvar/	[dæban]	trousers
	/pærvane/	[pæmane]	butterfly
	/radijo/	[dadijo]	radio
	/maʃin lebas ʃuji/	[vasin deban tuji]	washing machine
	/ʃirini/	[ʧinini]	sweet
	/lakpoʃt/	[dakbut]	turtle
	/toxmomorq/	[tovodod]	egg
	/?ænkæbut/	[?æb?æput]	spider
Me	/gævæzn/	[hævæt]	deer
	/toxmomorq/	[tovodod]	egg
	/ʃirini/	[ʧinini]	sweet
	/maʃin lebas ʃuji/	[papin tevan tuji]	washing machine
	/dæhæn/	[dæ?æn]	mouth
	/hævapejma/	[?ævavejma]	airplane
	/naxongir/	[tatoddi]	nail cutter
	/mesvak/	[beddap]	toothbrush
	/van/	[van]	tub
	/telefon/	[tetevun]	telephone
	/medad ʃæmi/	[petad tæmi]	crayons’
Hi	/mesvak/	[mesbad]	toothbrush
	/gav/	[gab]	cow
	/hævapejma/	[hæbapejma]	airplane
	/deræxt/	[dedæx]	tree

The abundance of such cases in PD children’s productions creates more complications in their manner harmony data relative to TD children. Furthermore, Dinnsen (1998), and Dinnsen and O’Connor (2001) have claimed that various types of manner harmony indicate different limitations on what can serve as a target. However, the children in this study have not indicated such limitations on what can serve as a target in this type of harmony. For example, in the nasal harmony errors of a TD child (EL, 2;6), liquids, stops, fricatives and affricates, all served as targets of nasal harmony (Table 10).

Table 10: Nasal harmony in El (TD child)

Target Word	Child Pronunciation	Gloss
/mobl/	[momb]	sofa
/sændæli/	[nændæli]	chair
/ʧængal/	[nængal]	fork
/dæmaq/	[næmaq]	nose
[xæmirdændun]	[xæminændun]	toothpaste
[badkonæk]	[nadtonæk]	ballon
[suzæn]	[nuzæn]	needle
[qænd]	[nænd]	cube sugar
[toxmomorq]	[momomoq]	egg

It should be reminded that except /ʒ/, El produces all Farsi sounds including the above harmonized sound i.e. /l, s, ʧ, d, b, t, q/ in all word positions in other contexts. Table 11 displays some examples of the correct production of the phonemes.

Table 11: Examples of correct production of /l, s, ʧ, d, b, t, q/

Target Word	Child Pronunciation	Gloss
/telefon/	[telefon]	telephone
/læb/	[læp]	lip
/fil/	[fil]	elephant
/boʃqab/	[boʃqap]	plate
/biskujit/	[biskovit]	biscuit
/setare/	[setare]	star
/qurbaqe/	[quqabe]	frog
/medad/	[mendad]	pencil
/saʔæt/	[sahæt]	o'clock
/ketab/	[ketab]	book
/dærja/	[dærja]	sea
/moʧ/	[moʧ]	wrist
/qejʧi/	[qejʧi]	scissors
/ʧi/	[ʧi]	what
/xorus/	[morus]	rooster

Third, as the results indicate, plosive and nasal harmony occurs in both TD and PD groups; however, fricative harmony only occurs in TD children. The presence of plosive and nasal harmony in the TD and PD groups and the absence of fricative harmony in the PD group are not surprising but justifiable through articulatory criteria. Nasals and plosives are less marked than fricatives on articulatory grounds because the production of fricatives requires more fine-grained coordination of articulators compared to that for plosives and nasals (Ladefoged, 2001). However, the occurrence of fricative harmony in TD children, in spite of being more marked on an articulatory basis, raises question about the motivation(s) behind it. To answer this question, we should consider the perceptual cues claimed by Wright (2001, 2004), who surveyed the perceptual features of approximants, fricatives, nasals and plosives. He considers two types of perceptual cues for segment identity, i.e. Internal Cues and Contextual Cues. The following hierarchies indicate the relative strength of internal and contextual cues in fricatives, nasals, and plosives:

Internal cues:
Fricatives > Nasals > Plosives
Contextual cues:
Plosives > Nasals > Fricatives

Regarding articulation, as explained above, fricatives are harder to produce than nasals. Nasals are also harder to produce than plosives because of their extra velum gestures (Samare, 1992; Ladefoged, 2001; Winters, 2002). Therefore, the articulatory hierarchy for the three manners of articulation ought to be as follows:

Articulation Hierarchy:
Plosive > Nasal > Fricative

Considering the articulatory ease and strong perceptual cues (i.e. contextual cues), plosives are predicted to be more likely to appear in children's early productions, because they are both easier to perceive and to produce. Nevertheless, though fricatives possess strong perceptual cues (i.e. internal cues), they are difficult to articulate. Nasals, though perceptually weaker than both plosives and fricatives, they are articulatorily easier than fricatives. Therefore, it is predicted that children with articulatory limitations have the tendency to use plosives and nasals as triggers of manner harmony, which can make the word easier for them to produce. This is the situation in PD children. However, in TD children who have fewer articulatory limitations relative to the PD children, the perceptual strength of fricatives dominates the articulatory strength of plosives and nasals in some harmony contexts. Thus, it seems that when there are fewer articulatory limitations in the production of speech sounds, perceptual factors can also motivate manner harmony. This conclusion implies that perceptual factors are able to stimulate phonological processes when articulatory limitations are lessened or removed.

Conclusion

This paper has analysed the results of a pilot study on consonant harmony in children with typical phonological development (TD) and children with functional phonological disorder (PD) acquiring Farsi. Examining the data relating to consonant harmony errors, this study concludes that triggers and targets of place harmony in children acquiring Farsi are partly different from those of children acquiring some other languages. Earlier studies on languages other than Farsi (e.g., English) have maintained that dorsals are triggers and coronals are targets of place harmony, while the present study illustrates that, in children acquiring Farsi, coronals are triggers and dorsals are targets of place harmony instead. Comparing the place harmonies in PD and TD children acquiring Farsi with the harmony errors of a child acquiring another language has illustrated that universal constraints, as Optimality Theory claims, are present in all languages, even when their presence is hidden by the normal productions of speakers of a language. Furthermore, the findings of this study do support the universality argument for PARSEDOR >> PARSECOR, as claimed by Goad (1997) nor the cross-linguistic dominance of dorsals over coronals in place harmony, as claimed by Kiparsky (1994). Eventually, employing Wrights' approach to articulatory and perceptual characteristics of phonemic categories (2001, 2004) in analyzing harmony patterns in TD and PD children has led to this conclusion: perceptual factors can also trigger harmony processes when articulatory limitations are lessened or removed. The findings of this study provide insights into the phonological development processes in children acquiring Farsi, and may help clinicians to improve their intervention strategies regarding PD children acquiring Farsi. Though this study has investigated harmony processes in Farsi to some extent, there are still other processes that have not been explored in Farsi yet and can form the substance of future research.

References

Adams, C., Byers Brown, B., & Edwards, M. (1997). *Developmental disorders of language*. London, UK: Whurr Publishers.

Dinnsen, D. A. (1998). On the organization and specification of manner features. *Journal of Linguistics*, *34*, 1-25.

Dinnsen, D. A., & Barlow, J. A. (1998). On the characterization of a chain shift in normal and delayed phonological acquisition. *Journal of Child Language*, *25*, 61-94.

Dinnsen, D. A., & O'Connor, K. M. (2001). Typological predictions in developmental phonology. *Journal of Child Language*, *28*, 597-628.

Fikkert, P. (2000). Acquisition of phonology. In L. Cheng & R. Sybesma (eds.), The first glot international state–of–the–article book. The latest in linguistics. *Studies in Generative Grammar*, *48*, 221-250.

Fikkert, P., & Levelt, C. (2003). *Input, intake, and phonological development: The case of consonant harmony*. Ms, Leiden University and University of Nijmegen.

Gerlach, S. R. (2010). *The acquisition of consonant feature sequences: Harmony, metathesis and deletion patterns in phonological development*. PhD Dissertation. University of Minnesota.

Goad, H. (1997). Consonant harmony in child language: An optimality theoretic account. In S. J. Hannahs & M. Young-Scholten (eds.), *Focus on phonological acquisition* (pp. 113-142). Amsterdam: John Benjamins.

Kiparsky, P. (1994). Remarks on markedness. Paper presented at *TREND 2*.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages*. Blackwell Publishers Inc.

McCarthy, J., & Prince, A. (1994). The emergence of the unmarked: optimality in prosodic morphology. *Proceedings of the NELS*, *24*, 333-379.

McCarthy, J., & Prince, A. (1995). Faithfulness and reduplicative identity. In J. N. Beckman, L. W. Dickey & S. Urbanczyk (eds.), *Papers in optimality theory*, *18*, 249-384

Pater, J. (2002). Form and Substance in phonological development. In L. Mikkelsen & C. Potts (eds.), *WCCFL 21 Proceedings* (pp. 348-372). Somerville, MA: Cascadilla Press.

Pater, J., & Werle, A. (2003). Direction of assimilation in child consonant harmony. *Canadian Journal of Linguistics*, *48*(3), 385-408.

Prince, A. & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar*. Oxford, UK: Blackwell, reprinted 2002, 2004.

Samare, Y. (1992). *The phonetics of the Farsi language*. Tehran: Markaze Nashre Daneshgahi.

Smith, N. (1973). *The acquisition of phonology: A case study*. New York: Cambridge University Press.

Stemberger, J. P., & Bernhardt, B. H. (1997). Optimality theory. In M. Ball & R. Kent (eds.), *The new phonologies* (pp. 211-245). San Diego, CA: Singular Publishing Group.

Stoel-Gammon, C., & Stemberger, J. (1994). Consonant harmony and underspecification in child phonology. In M. Yavaş (ed.), *First and second language phonology* (pp. 63-80) San Diego: Singular Publishing Group.

Vihman, M. M. (1978). Consonant Harmony: Its scope and function in child Langugae. In J. H. Greenburg (ed.) *Universals of human language 2: Phonology* (pp. 281- 334). Stanford: Stanford University Press.

Winters, S. (2002). *Perceptual influences on place assimilation: a case study*. http:// www.ling.ohio-state.edu/~swinters/pdf/ Brigade.pdf.

Wright, R. A. (2001). Perceptual cues in contrast maintenance. In E. V. Hume & K. Johnson (eds.), *The role of speech perception in phonology* (pp. 251-277). San Diego: Academic Press.

Wright, R. A. (2004). A review of perceptual cues and cue robustness. In D. Steriade, R. Kirchner & B. Hayes (eds.), *Phonetically based phonology* (pp. 34-57). Cambridge, UK: Cambridge University Press.

No immersion, no instruction: Children's non-native vowel productions in a foreign language context

Ellen Simon[1], Ronaldo Lima Jr.[2], Ludovic De Cuypere[1]
ellen.simon@ugent.be, ronaldojr@letras.ufc.br, ludovic.decuypere@ugent.be

[1]Ghent University, [2]Federal University of Ceará

Abstract. This study aims to map the native Dutch and non-native English vowels of Belgian children who have not been immersed and have not received any school-based instruction in English, but who are exposed to it through the media. A fairly large and recent body of research addresses second language perception and production by early learners either through immersion in an L2-speaking community or through classroom-based instruction. However, there is also a vastly expanding number of children who live in a monolingual community and yet are exposed to English as a Foreign Language (EFL) from an early age through various media. This study addresses the question to what extent children acquire the English vowel system in such a context: is this type of exposure sufficient for them to create new phonetic vowel categories? Twenty-four Dutch-speaking children, aged between 9 and 12, participated in the study. They were all living in Belgium, and came from different dialect regions. None of them had received English instruction in school, but all of them reported having at least some sporadic contact with English, for instance, through television programmes or computer games. They all performed two Dutch picture-matching tasks, an English picture-naming task, and an English repetition task. The auditory stimuli were monosyllabic Dutch and English words containing each 12 Dutch and 11 English monophthongs. The vowel formants were analysed in Praat (Boersma & Weenink, 2011) by comparing the LPC (Linear Predictive Coding) analysis to the FFT (Fast Fourier Transform) spectrum. Lobanov-normalized vowel plots present the organization of these children's entire Dutch and English vowel spaces. The results focus on the English vowel contrasts /ɛ-æ/ and /ʊ-u/, as Dutch lacks these contrasts and has only one vowel in these areas of the vowel space (/ɛ / and /u/, respectively). The children produced a contrast between English /ɛ/ and /æ/ in the repetition task, but not in the picture-naming task. English /ɛ/, but not /æ/ was considerably different from the closest Dutch vowel /ɛ/. The children's English /ʊ/ and /u/ differed in terms of height (F1) and anteriority (F2), both in the repetition and the picture-naming task. The closest Dutch vowel, represented as /u/, did not differ from English /u/, and differed from /ʊ/ only in terms of height. The results suggest that 9-12-year-old Flemish children are at the beginning of creating new contrasts for non-native English vowels. This means that media-induced Second Language Acquisition should not be underestimated: even in contexts of L2 acquisition exclusively through media exposure children learn to produce contrasts between L2 vowels which do not exist in their L1.

Keywords: child second language phonology, vowels, production, acoustics, Dutch, English

Introduction and aims

This study aims to map the native (L1) and non-native (L2) vowels of children who have not yet received any school-based instruction in the L2, but who have been exposed to it in a non-immersion context. Studies on L2 phonological acquisition have typically focused on immersion contexts, often examining language acquisition by immigrants. In these contexts, once L2 acquisition starts, it is typically with intense exposure. The results of these studies (e.g., Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege, 2005; Gildersleeve-Neuman, Peña, Davis, & Kester, 2009; Darcy & Krüger, 2012) show that the L1 is generally still permeable in childhood, and that the children's L2 productions differ not only from those in their L1, but also from those of age-matched L1 children. Another set of studies on child L2 acquisition have focused on the effect of instruction on child L2 phonological acquisition, mostly examining the effect of age of onset of instruction on the attained proficiency level. The Barcelona Age Factor project (Muñoz, 2006), conducted longitudinally between 1996 and 2002, compared pupils for whom English instruction started at age 11 to pupils who started getting English instruction at age 8. Muñoz' conclusion of the project as a whole is that no group of learners performed even close to the native speakers that composed the control group. Late starters (age 11) performed better than early starters (age 8) at all phases of data collection, but the older learners' advantage decreased in the later collections. Conclusions of studies within the project focusing on perception and production (Fulana, 2006) and oral fluency (Álvarez, 2006; Mora, 2006) reached the same conclusion.

These studies suggest that, in contexts of maximal input, either through immersion or intensive instruction or training, children's L2 speech is influenced by their L1 and differs from that of age-matched L1 speakers. The question we address in this paper is what child L2 speech looks like in contexts of minimal input, i.e. in the absence of immersion or formal instruction. Such contexts are actually common: in many European countries, including Belgium, children are exposed to English through various media, such as computer games, television programmes and the radio before they get English classes in school.

In this study, we examine to what extent 9-12-year-old Dutch-speaking children living in Flanders have acquired the spectral quality of L2 English vowel sounds as the result of exposure to English through various media. Since children are exposed to multiple varieties of English (as is typical for English as a Foreign Language contexts, see Bohn & Bundgaard-Nielsen, 2007), the children's L2 vowels will not be compared to those of a control group of L1 speakers. Rather, we examine the internal organization of the children's L1 and L2 vowels spaces. In this paper, we will zoom in on two L2 vowel contrasts which do not occur in the L1, and address the following questions: (1) Do the children produce a contrast between the L2 English vowels in these pairs?, and (2) Do these productions differ from the closest L1 Dutch vowel?

Method

Participants

Twenty-four Dutch-speaking children, living in Flanders, Belgium, participated in Dutch and English production tasks. The mean age of the participants (9 girls, 15 boys) at the time of testing was 10;6 years (range: 9;10 to 12;2). Data were collected in three schools in different towns in Flanders, Ghent (n = 9), Erembodegem (n = 6) and Mol (n = 9), in order to examine potential effects of L1 regional variation. None of the children had received any formal L2 English instruction in school or made extended trips to English-speaking countries and no children reported having contact with native English speakers. However, all children in Belgium are exposed to English through the media and popular culture (music channels, English-spoken cartoon channels, computer games, English pop music, etc.), so that by the age of 9, they have a basic knowledge of English.

Tasks and procedure

All children performed a Dutch picture-matching task, an English picture-naming task and an English repetition task. In the Dutch picture-matching task, they were asked to match pictures while producing sentence of the form 'X belongs to Y', in which either X or Y was a target word (e.g., 'The *cheese* belongs to the mouse' - 'De *kaas* hoort bij de muis').

In the English repetition task, children saw pictures on a computer screen and heard the corresponding words over Bose™ headphones. They were instructed to repeat the words. Audio recordings, produced by a male and a female speaker of British English, were extracted from the online version of the Cambridge Advanced Learner's Dictionary (Upper intermediate – advanced) (Cambridge: Cambridge University Press, third edition, http://dictionary.cambridge.org). The English picture-naming task aimed at eliciting spontaneously produced words as opposed to repeated words. Children were shown six cards with four pictures on each and were asked to name the objects for which they knew the names in English.

Experimental set-up

The children were individually tested in a quiet room in their school, with no other person present besides the experimenter. All instructions were provided orally in Dutch. The recordings were made with a Sony clip microphone (ECMCS10), connected to a pocket-size Marantz Professional solid state recorder (PMD620). The recordings were made in Mono, with a sampling rate of 44.1 KHz. All tasks were performed in one session, and always in the order in which they are presented above.

Stimuli

All visual stimuli were black or coloured line drawings, taken from the web. The auditory stimuli were monosyllabic Dutch and English basic vocabulary words. Monosyllabic words with each of the 12 Dutch (/ɛ/, /ɪ/, /u/, /ɑ/, /i/, /ɔ/, /a/, /ʏ/, /o/, /ø/, /e/, /y/) and 11 English monophthongs (/ɛ/, /ɪ/, /ʊ/, /ɑ/, /i/, /ɔ/, /æ/, /ʊ/, /ɜ/, /ʌ/, /ɒ/) were selected, excluding schwa. Since the children's vocabulary in English was very limited, the consonantal context of the words could not be controlled for. All target words were high-frequency English words likely to be known by the majority of the children (mean log frequencies: picture-naming task: 9.954, SD 1.19; repetition task: 9.93, SD 1.27; frequencies from Balota, Yap, Cortese, Hutchison, Kessler, Loftis, Neely et al., 2007).

Analysis

The spectral analysis is based on measurements of the first and second formants. After the vowels were segmented in PRAAT (Boersma & Weenink, 2011), formant values predicted by LPC (Linear Predictive Coding) were manually checked against the FFT power spectrum (obtained by the calculation of the Fast Fourier Transform algorithm) of the central, most stable part of each vowel. This manual checking allowed for adjustments to be made in the ceiling frequency and/or the order of the LPC whenever necessary, which is essential when working with children, whose ceiling frequencies may vary considerably from one to another due to their still developing vocal tracts and typical high F0 values. A PRAAT script (Arantes, 2010) was used to visualize the LPC predictions against the FFT spectrum, and to change the parameters of analysis when necessary, and another script (Arantes, 2011) was used to later export all resulting F1 and F2 values to a spreadsheet. After extraction, F1 and F2 values were Lobanov normalized (Lobanov, 1971) and the output values were rescaled to Hertz, using the 'vowels' package (Kendall & Thomas, 2010) for R software (R Core Team, 2012). On the basis of visual inspection of the scatterplots (see Figures 1 and 3 in section 4), we identified 60 vowel productions with extreme values. After a close, manual examination of these 60 vowels, 49 observations were removed because background noise or extreme lengthening or whispering made the measurement unreliable. Thus, extreme values were deleted for technical reasons only, not because of their distance from the bivariate means. The normalized data were then used to create F1xF2 plots and to conduct joint multivariate tests.

In total, 793 Dutch and 1303 English vowels were retained in the analysis, leading to a total of 2096 vowels. For this paper, we focus on the analysis of two English vowel contrasts, which do not occur in Dutch. In these two pairs, Dutch has just one vowel in the area of the vowel space where English has two (see Table 1), and both pairs are hence predicted to be problematic for native speakers of Dutch.

Table 1. Three English vowel contrasts and the spectrally closest Dutch vowel

	English pairs	Closest Dutch vowel
1.	/E-Θ/ ('DRESS'-'TRAP')	/E/ ('MES')
2.	/Y-u/ ('FOOT'-'GOOSE')	/u/ ('HOEK')

Results

DRESS – TRAP vs. MES

Figure 1 presents a scatterplot of all productions of the English vowels /ɛ/ ('DRESS') and /æ/ ('TRAP') (left) as well as the closest Dutch vowel /ɛ/ ('MES') (right). All scatter plots are created with McCloy's (2015) PhonR package in R. The leftmost panel includes results of the picture-naming as well as repetition task. The rightmost panel includes only the English and Dutch picture-naming/matching task, since no repetition task was conducted in Dutch.

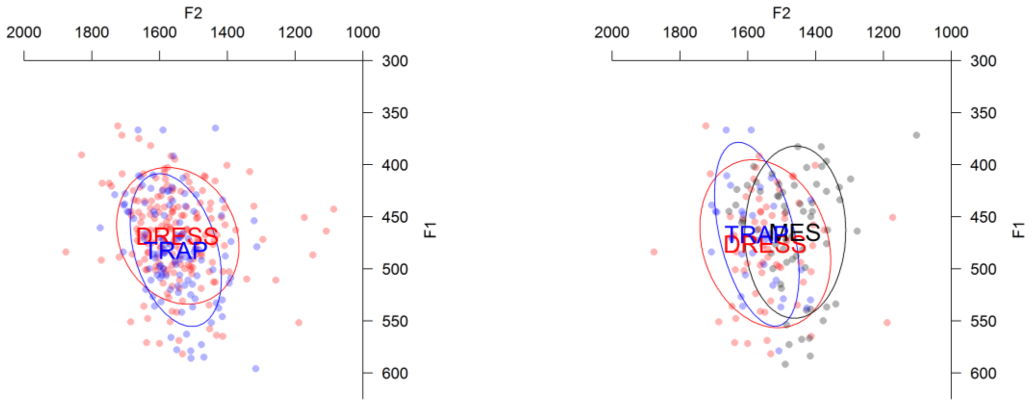


Figure 1. Scatterplot of English DRESS and TRAP (left) (spontaneous and repetition tasks), and comparison with Dutch MES (right) (spontaneous task only)

The scatter plots suggest a difference between DRESS and TRAP on F1 and a difference between MES and TRAP/DRESS on F2. The results of a joint multivariate test on the bivariate means for English DRESS and TRAP, controlling for TASK and REGION, show a significant effect of TARGET VOWEL in interaction with TASK (repetition vs. picture-naming/matching; Type II MANOVA: Hotelling-Lawley test, $P = 0.02$). (All statistical analyses were performed in R).

A post-hoc linear regression analysis on both formants separately indicates that TARGET VOWEL was significant in interaction with TASK for F1 ($P < 0.01$). The interaction plot in Figure 2 shows that F1 for TRAP is much higher than for DRESS in the repetition task, which is expected in English, but the reverse pattern can be observed in the picture-naming/matching task. While the 95% confidence intervals (the red bars) do not overlap in the repetition task, they do overlap in the spontaneous task, meaning that in the picture-naming task (referred to as the 'spontaneous' task) there is no evidence that a contrast is being made.

No difference between the target vowels was found for F2, which is in line with what the scatterplot in Figure 1 shows.

A multivariate comparison of DRESS and TRAP with the closest Dutch vowel, MES, again revealed a significant effect of TARGET VOWEL (Type II MANOVA test: Pillai test, $p < 0.001$). The post-hoc linear regression model showed that Dutch MES was significantly different from English DRESS in terms of F2 ($P < 0.001$)and F1 for the REGION Erembodegem. The difference with TRAP was not significant, neither in F1 nor F2.

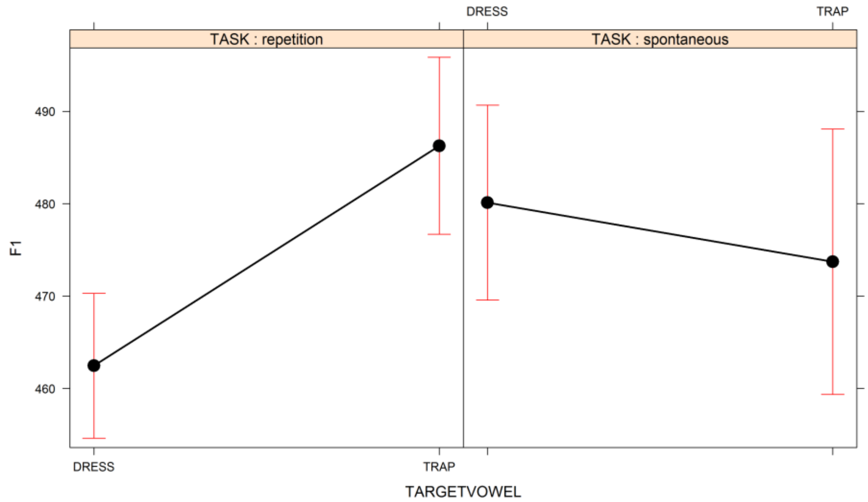


Figure 2. Interaction plot for TASK and TARGET VOWEL for F1

FOOT-GOOSE vs. HOEK

Figure 3 presents a scatterplot of all productions of the English vowels /Y/ ('FOOT) and /u/ ('GOOSE') (left) as well as the closest Dutch vowel /u/ ('HOEK') (right).

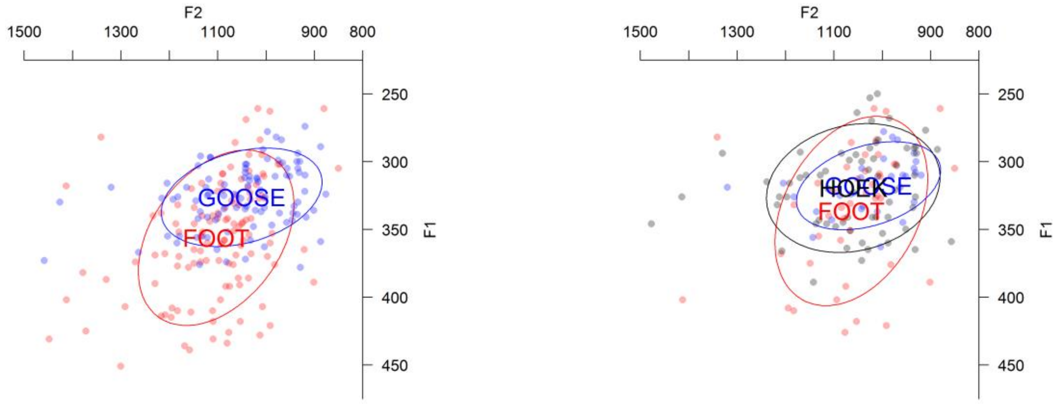


Figure 3. Scatterplot of English FOOT and GOOSE (left) (spontaneous and repetition task), and comparison with Dutch HOEK (right) (spontaneous task only)

As for the DRESS-TRAP contrast, a joint multivariate test on English FOOT and GOOSE productions revealed a highly significant effect of TARGET VOWEL, controlling for REGION and TASK (Type II MANOVA, Hotelling-Lawley test: $P < 0.001$). A post-hoc linear regression analysis confirmed that the two vowels differed significantly both in F1 and F2 ($P < 0.001$), again controlling for REGION and TASK.

A comparison with the closest Dutch vowel, HOEK, showed no evidence of a multivariate difference between the three vowels means (Type II MANOVA, Pillai test: $P = .054$). However, a post-hoc linear regression analysis revealed that Dutch HOEK was different from English FOOT in terms of F1 ($p = 0.02$), but not in terms of F2. No difference between HOEK and GOOSE was found in either F1 or F2.

Discussion and conclusions

This study addressed the question of whether Dutch-speaking children living in Flanders learn to create new categories for English vowels before they have received English instruction in school. In other words, is sheer exposure to English-spoken media sufficient for children to develop new L2 vowel categories, and to what extent do these vowel categories differ from the spectrally closest L1 Dutch vowels? For this paper, we zoomed in on two English vowel contrasts which do not occur in Dutch, namely /ɛ-æ/ and /ʊ-u/. Even though the DRESS-TRAP contrast is known to be difficult for native speakers of Dutch, both in perception (Broersma, 2005; Escudero, Simon, & Holgerer, 2012) and in production (Simon & D'Hulster, 2012), children produced these English vowels significantly different, both in terms of F1 and F2, but only in a repetition task. We found no evidence for a contrast between DRESS and TRAP in a picture-naming task, in which children had to retrieve their phonological representations of the L2 vowels. A comparison with the closest Dutch vowel, MES, conventionally represented by the phonetic symbol /ɛ/, showed that the children produced this Dutch vowel differently from English /ɛ/, both in terms of height and anteriority, but not different from English /æ/. With respect to the FOOT-GOOSE contrast, the results again showed that children produced a contrast between these vowels in terms of both height and anteriority, and this time they did so in both the repetition and the picture-naming tasks. The closest Dutch vowel, HOEK, represented as /u/, did not differ from English GOOSE, and differed from FOOT only in terms of height. In other words, even though the children's Dutch vowel is highly similar to both English vowels, the children managed to produce a contrast between these two L2 vowels.

To conclude, the results suggest that 9-12-year-old Flemish children are at the beginning of creating new contrasts for non-native English vowels. This means that media-induced Second Language Acquisition should not be underestimated: even in contexts of L2 acquisition exclusively through media exposure ('no immersion - no instruction'), children learn to produce contrasts between L2 vowels, which do not exist in their L1. The results are interesting in light of the relation between perception and production. A previous perception study with the same group of Flemish children (Simon, Sjerps, & Fikkert, 2012), based on mispronunciation detection tasks, showed that the children's perception of L2 English vowels was strongly influenced by their L1, but that the beginning of development of new categories could be detected. However, while the children are exposed to English-spoken media from an early age onwards, and get a considerable amount of L2 receptive input, they hardly ever produce the L2. Interviews with the child participants revealed that production of English was restricted to singing along with pop songs and the use of occasional English phrases with friends. Yet, despite this lack of productive practice, the children are at the beginning of creating new categories in their production, on the basis of their receptive input.

In addition, the results may have a pedagogical impact: children who are not immersed in the L2 and have not even had English classes in school yet, have an L2 vowel space which is different from their L1 vowel space, which is something that teachers in the first years of English language instruction in school may want to take into account when developing their teaching materials.

References

Álvarez, E. (2006). Rate and route of acquisition in EFL narrative development at different ages. In C. Munoz (ed.), *Age and the Rate of Foreign Language Learning* (pp. 127-155). Clevedon, Tonawanda NY & Ontario: Multilingual Matters.

Arantes, P. (2010). Formants.praat, v. 0.9 beta.

Arantes, P. (2011). Collect formants.praat, v. 0.11 alpha.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445-459.

Boersma, P., & Weenink, D. (2011). *Praat: Doing Phonetics by Computer.* [Computer Programme], http://www.praat.org.

Bohn, O. -S., & R. L. Bundgaard-Nielsen. (2007). Second language speech learning with diverse input. In T. Piske & M. Young-Scholten (eds.), *Input Matters in SLA* (pp. 207-218), Clevedon: Multilingual Matters.

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America*, *117*, 3890-3901.

Darcy, I., & F. Krüger. (2012). Vowel perception and production in Turkish children acquiring L2 German, *Journal of Phonetics*, *40*, 568-581.

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception, *Journal of Phonetics*, *40*, 280-288.

Fulana, N. (2006) The development of English (FL) perception and production skills: starting age and exposure effects. In C. Munoz (ed.), *Age and the Rate of Foreign Language Learning* (pp. 41-64). Clevedon, Tonawanda NY & Ontario: Multilingual Matters.

Gildersleeve-Neuman, C. E., Peña, E. D., Davis, B., & Kester, E. (2009). Effects of L1 during early acquisition of L2: Speech changes in Spanish at first English contact, *Bilingualism: Language and Cognition, 12 (2),* 259-272.

Kendall, T., & Thomas, E. R. (2010). Vowels: Vowel manipulation, normalization, and plotting in R. R package, v. 1.1, [Software available online: http://ncslaap.lib.ncsu.edu/tools/norm/].

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America*, *49,* 606-08.

McCloy, D. (2015). PhonR: tools for phoneticians and phonologists. R package version 1.0-3.

Mora, J. C. (2006). Age effects on oral fluency development. In C. Munoz (ed.), *Age and the Rate of Foreign Language Learning* (pp. 65-88). Clevedon, Tonawanda NY & Ontario: Multilingual Matters.

Munoz, C. (ed.) (2006). *Age and the Rate of Foreign Language Learning*. Clevedon, Tonawanda NY & Ontario: Multilingual Matters.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Online: http://www.R-project.org/.

Simon, E., Sjerps, M., & Fikkert, P. (2013). Phonological representations in children's native and non-native lexicon, *Bilingualism: Language and Cognition*, *17*(1), 3-21.

Simon, E., & D'Hulster, T. (2012). The effect of experience on the acquisition of a non-native vowel contrast, *Language Sciences*, *34,* 269-283.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children, *Journal of Phonetics*, *33*, 263-290.

Investigating the relationship between parental communicative behavior during shared book reading and infant volubility

Anna V. Sosa
anna.sosa@nau.edu

Northern Arizona University

Abstract. Accumulating evidence suggests that there is a strong, predictive relationship between prelinguistic vocalization (babble) and later language development. In general, "better" babblers become "better" talkers: that is, they have been shown to reach linguistic milestones sooner, to have faster rates of vocabulary acquisition, and to achieve superior language outcomes at later ages (summary in Stoel-Gammon, 1998). A number of child-internal factors have been identified that impact quantity and quality of babble and a few parent behaviors have been found to shape infant vocalization under controlled experimental conditions. However, the concurrent relationship between characteristics of naturally occurring child-directed speech and infant volubility has not been explored extensively. The current study seeks to answer the following question: which caregiver communicative behaviors are associated with increased infant volubility during naturally occurring play activities? Twenty parent-infant dyads (infants between 10 and 16 months) participated in 3 days of recording using the LENA Pro (Language Environment Analysis [LENA Foundation, Boulder, CO]) System. The LENA system includes a small digital recording device called a digital language processor that is worn by the child in a pocket in a vest. The processor can record up to 16 hours and is worn continuously by the child for at least 10 hours. The accompanying LENA software conducts automatic analyses of the recordings and generates estimates of a variety of different language measures, including amount of speech produced by adults in the child's environment and the number of vocalizations produced by the child. The data analyzed for the current study include one 15-minute play session per dyad with age-appropriate books. The play sessions occurred at the parent's convenience on one of the 3 days of recording and were not observed by researchers. Play sessions were orthographically transcribed and coded for the following parental communicative behaviors: 1) adult words per minute, 2) questions per minute, 3) directives per minute, 4) rejections/negations per minute, 5) engaging/excited expressions (e.g., sound effects, animal noises, gasps, claps, etc.), 6) rate of parental verbal responsiveness to infant vocalizations, and 7) parental imitation of infant vocalizations. Infant volubility was determined by calculating the number of speech-like vocalizations produced by the child per minute. Correlational analyses were conducted to investigate the relationship between parental communicative behaviors and infant volubility. The only parent behavior that was significantly correlated with infant volubility was parental verbal responsiveness; infants of parents who responded more consistently to their child's vocalizations babbled more during the play sessions. Results are consistent with previous work showing that parental responsiveness plays an important role in language development and has clinical implications for professionals working with families of young children who present with or who are at risk for delays in language acquisition.

Keywords: language input, prelinguistic vocalization, babble, parental responsiveness

Introduction

In his seminal paper in the field of child phonology, Jakobson (1968) referred to the prelinguistic period of vocal development (babble) as, "the purposeless egocentric soliloquy of the child" and as "biologically oriented 'tongue delirium'" (p. 24). Both expressions reflect a general opinion, which may have been widely held at the time, that babble is unrelated to language and that the study of language development should begin at the point when children begin producing true words. Over the past several decades, however, researchers in child phonology have investigated the phonetic similarities between babble and early words and have identified important relationships between prelinguistic vocalization and later language development. Stoel-Gammon (1998) summarizes some of the findings regarding the positive, predictive relationship between babble and later language development. In general, several studies have found correlations between babble "ability" and general speech and language skills, with "better" babblers, as determined by a variety of different measures, becoming "better" talkers: that is, the children who produced overall more as well as more complex babble reached linguistic milestones sooner, had faster rates of vocabulary acquisition, and achieved overall superior language outcomes at later ages.

Several reasons for the observed relationship between babble and later speech and language development have been discussed. Stoel-Gammon (1998) describes speech as having a skill component and that with more practice (i.e. more and better babble), comes greater control and precision of the movement. Furthermore, increased practice of sounds and syllables in babble may facilitate mapping of specific movement patterns to the resulting acoustic output. This is described by Stoel-Gammon (1998, 2011) as the auditory-articulatory 'feedback loop' and is necessary for the production of words. Additionally, babies who are better babblers may receive more responsive feedback from caregivers and may engage in more and longer vocal interactions with adults, both of which have been associated with better current and later language ability as well as with earlier attainment of major language milestones (Gros-Louis, West, Goldstein, & King, 2006; Tamis-LeMonda, Bornstein, & Baumwell, 2001; Zimmerman, Gilkerson, Richards, Christakis, Xu, Gray, & Yapanel, 2009).

Given the empirical evidence as well as the theoretical rationale for the association between prelinguistic vocalization and later language development, it follows that interventions focusing on increasing and expanding babble repertoire may be relevant for young children who are exhibiting or who are at risk for language delay. In order to implement this type of intervention, identification of specific parent behaviors that function to encourage babble is necessary. Several child-internal factors are known to affect quantity and quality of babble. These include language delay (Rescorla & Ratner, 1996), developmental disability (Paul, Fuerst, Ramsay, Chawarska, & Klin, 2011; Stoel-Gammon, 1997), and hearing status (Oller & Eilers, 1988). Some child-external or environmental factors have also been found that increase infant volubility under experimental conditions. These include contingent verbal and non-verbal parental responsiveness (Franklin, Warlaumont, Messinger, Bene, Nathani Iyer, Lee et al., 2014; Goldstein, King, & West, 2003; Goldstein & Schwade, 2008), "still face" paradigm (Hsu, Fogel, & Messinger, 2001), and parent imitation of child vocalizations (Dunst, Gorman, & Hamby, 2010). During natural interactions, parents use a variety of interactional and communicative behaviors to engage their infants. These behaviors include sound effects, songs, questions, directives, etc. Little is known, however, about the impact of these naturally occurring parental communicative behaviors on infant volubility. If behaviors that encourage infant vocalization can be identified, these behaviors may be used by parents and clinicians to increase infant volubility in children who are exhibiting signs of communication delay or who have risk factors for language delay.

The purpose of the current study was to explore the relationship between parent communicative behaviors and infant volubility during short parent-infant play sessions with age-appropriate books. The goal was to identify parent behaviors that are associated with increased vocalizations by the child. It was hypothesized that more overall parent talk, greater use of animated and engaging expressions and consistent parental responsiveness would be associated with increased infant volubility. While a number of studies have identified relationships between parent-infant communication and later language development (Gilkerson & Richards, 2009; Hart & Risley, 2003), few have investigated the relationship between naturally occurring parent communicative behaviors and concurrent infant volubility. Results of this study will add to that knowledge base.

Method

Participants

The data analyzed for the current study are taken from a larger study investigating the effect of type of toy used during play on quantity and quality of parent-infant communicative behavior (Sosa, in press). Data for the current study are from 20 parent-infant dyads recruited through posting of flyers in areas frequented by families with young children. The infant participants were between 10 and 16 months (*Mean* = 13.4) at the time of the study; 8 were male and 12 were female. The parents who participated in the study along with their children included 18 biological mothers and 2 biological fathers. American English was reported as the primary language used in the home. All parents had at least a high school degree and 15 parents had completed 4 or more years of post-secondary education. Racial/ethnic information was gathered at the time of enrollment. Of the 20 participating families, 18 self-reported as non-Hispanic white, 1 was Hispanic, and 1 was Native American.

Data collection

As part of the larger study, parent-infant dyads participated in 3 days of data collection taking place in their homes. Over the course of these 3 days, parents engaged in three 15-minute play sessions with their babies using three different toy sets, which were provided by the researchers. The toy sets included traditional toys (e.g., blocks and puzzles), electronic toys (e.g., baby laptop, electronic baby cell phone), and books (e.g., stiff board books with animal, color, and shape themes), all designed and marketed for children in this age range. Parents engaged in two 15-minute play sessions with each toy set. Data for the current investigation are taken from the first 15-minute play session with books only. The set of books consisted of five books; two books had a farm animal theme, two books had a shapes theme, and one book had a color theme. Parents were free to choose when during the day to play with the toys and were not directed to minimize natural distractions of the home environment. Therefore, there were sometimes other children, pets, or other adults present during the play session.

Play sessions were recorded using the LENA Pro (Language Environment Analysis [LENA Foundation, Boulder, CO]) System. The system includes a small digital recording device called a digital language processor that is placed in a pocket in a vest worn by the child. The processor records up to 16 hours of recorded sound and is worn continuously by the child for at least 10 hours. The accompanying LENA software conducts automatic analyses of the recordings and generates estimates of the amount of speech produced by adults in the child's environment, the number of child vocalizations, the number of adult-child conversational turns, and the amount of exposure to electronic noise (e.g., television). Parents were instructed to turn on the recording device when the baby woke up in the morning and to keep it running until the baby went to bed in the evening, resulting in recordings that were between 10 and 14 hours long. Because the recordings were done by the parents using the LENA Pro system, researchers were not present in the home during the play sessions. This was done in order to increase naturalness of parent-infant interaction and thereby increase ecological validity of study findings.

Parents also completed the *Lena Developmental Snapshot* before the first day of recording. This measure is a parent questionnaire that is completed together with the researcher and asks questions about expressive and receptive communication development. Based on parent responses, a standard score is generated, providing a quick estimate of overall communication development. The average standard score for the participating infants was 103.3 (*Mean* for the assessment = 100*; s.d.*= 15), and the range of scores for the participants was from 86 to 123, suggesting that all infants were exhibiting typical communication development, as per parent report.

Data coding

The audio recordings of the 15-minute play sessions were extracted from the longer recordings for coding and analysis. All parent utterances from the play sessions were orthographically transcribed by graduate student research assistants. These parent utterances were coded to generate seven different measures of parent communicative behavior. The measures are given, along with definitions and examples in Table 1.

Infant volubility (**VOL**) was determined by calculating the number of infant vocalizations produced per minute during the play session. An infant vocalization was defined as a speech-like utterance consisting of, at minimum, a voiced vowel. Cries, grunts, and vegetative noises were not coded as infant vocalizations.

Table 1. Description of measures of parent communicative behaviors that were analyzed

Measure of parent behavior	Definition	Example
Adult Words per minute (**AW**)	All words produced by the parent during the play session	n/a
Questions per minute (**QUEST**)	Questions produced by the parent during the play session	"Who says quack?" "Does this look like your little chick?"
Directives per minute (**DIR**)	Utterances produced by parents that were interpreted as an attempt to direct or redirect the child's attention or behavior	"Come here." "Can you make the sound of the snake."
Rejections/negations per minute (**REJ**)	Utterances produced by parent that included a rejection of the child's utterance or behavior	"That one's not the dog…" "No, nope, you cannot take my glasses."
Engaging/excited expressions per minute (**EXC**)	Parent productions of animal sounds, sound effects, gasps, claps, singing, baby games, interjections, nursery words, or baby's name	"Quack quack" "Oops" "Bang"
Verbal responsiveness (**RESP**)	Parent verbal response to child vocalization produced within 5 seconds of child utterance (calculated as a proportion by dividing the total number of parent responses per minute by the total number of child vocalizations per minute)	Child: (babbled utterance) Parent: "Oh yeah"
Imitation of child vocalizations per minute (**IMIT**)	Utterances produced immediately following a child vocalization that was an imitation of the child's babbled or linguistic utterance	*There were not examples of parent imitations of child vocalizations in the dataset*

Additional analysis was conducted using some of the measures derived from the LENA automatic analyses. The measures used included the adult word count (**AWC**) and the child vocalization (**CV**) percentile scores generated from all 3 days of recording. These measures reflect overall volubility of the child and the overall quantity of adult language heard by the child (with percentile scores generated from normative data).

Results

Inspection of the range and standard deviation showed that there was considerable variability between participants in terms of frequency of the parent communicative behaviors as well as infant volubility. Means, ranges, and standard deviations for each behavior are presented in Table 2. All values, with the exception of verbal responsiveness rate are presented in occurrences per minute.

Table 2. Mean, range, and standard deviation for parent communicative behaviors and infant volubility

Behavior	Mean	Range	Standard Deviation
AW	65.72	33.84 - 108.13	21.34
QUEST	5.08	2.52 - 9.59	1.9
DIR	2.07	.31 - 4.33	1.08
REJ	.11	0 - .67	.19
EXC	8.66	2.93 - 16.6	3.55
RESP	.55	.15 - 1.00	.21
VOL	4.02	.29 - 9.63	2.26

Parent verbal responsiveness is directly related to the frequency of child vocalizations; that is, there is only an opportunity for a response immediately after a child vocalization. In order to account for this, responsiveness is not reported in terms of absolute frequency of occurrence, but rather as the rate with which a parent responded verbally to a child vocalization (e.g., a responsiveness rate of .5 indicates that the parent responded verbally to the child's vocalizations 50% of the time).

In order to investigate the concurrent relationship between parent communicative behaviors and infant volubility during play with books, a series of correlational analyses were conducted. The correlations between six parent communicative behaviors (imitation of child vocalizations was excluded because it was not represented in the data) and infant volubility (i.e. child vocalizations per minute) is presented in Table 3. As evident in Table 3, there were no significant correlations between parent communicative behaviors and infant volubility.

Table 3. Correlations between parent communicative behaviors and infant volubility

	AW	QUEST	DIR	REJ	EXC	RESP
VOL	-.320	-.109	.092	.239	-.180	.069

Given that two of the infants presented as outliers in terms of their overall volubility rate (one infant vocalized only .29 times per minute and the other vocalized almost 10 times per minute), it was decided to run the correlational analyses again with these 2 outliers removed. Results of this analysis are presented in Table 4. This analysis resulted in one significant relationship; rate of parent verbal responsiveness was positively correlated with infant volubility. A scatter plot showing the relationship between rate of parent verbal responsiveness and infant volubility for the 18 parent-infant dyads is given in Figure 1.

Table 4. Correlations between parent communicative behaviors and infant volubility with outliers removed

	AW	QUEST	DIR	REJ	EXC	RESP
VOL	.032	.338	.033	.303	-.224	.554*

* Correlation is significant at the .05 level

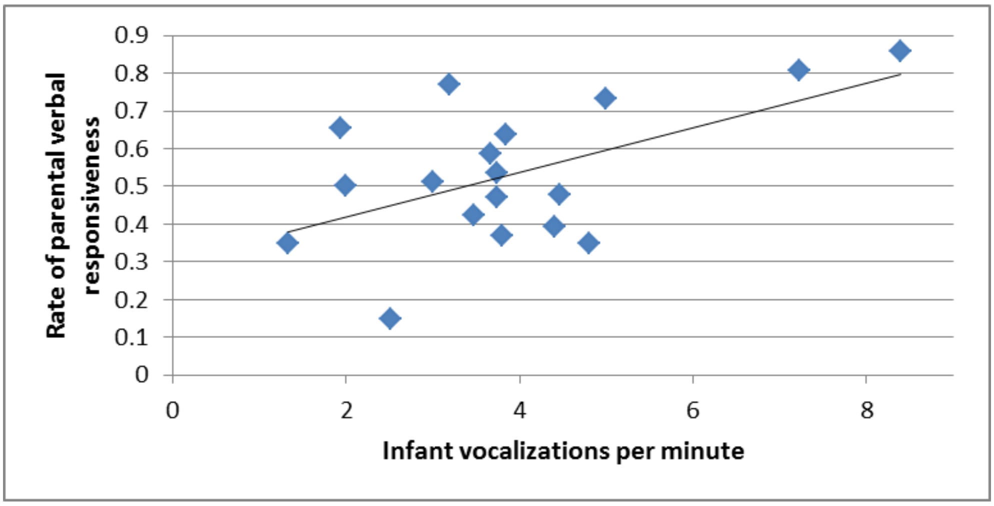


Figure 1. Scatter plot showing the relationship between verbal responsiveness and infant volubility (2 outliers removed)

Since only one measure of parent communicative behavior was correlated with concurrent infant volubility during the play sessions, the question arose as to whether infant volubility may be more closely related to a child's overall language environment, rather than to specific communicative behaviors present during a brief interaction. To explore this question, the automatic LENA measures generated from the 3 days of recording were analyzed to determine the relationship between general language environment, overall infant volubility during the 3 days, as well as infant volubility and adult words per minute during the play sessions. The measures of general language environment that were used in the analysis include Adult Word Count (AWC) and Child Vocalizations (CV) and are expressed as percentile scores. The correlation matrix showing the relationship between these measures is presented in Table 5.

Table 5. Correlations between adult and infant volubility during the play sessions (VOL and AW) and over the 3 days of recording (CV and AWC)

	CV	AW	AWC
VOL	.528*	-.320	.083
CV		-.261	-.072
AW			.550*

* Correlation is significant at the .05 level

Note. VOL: infant volubility during the play sessions; CV: overall infant volubility over 3 days of recording; AW: adult words per minute during the play sessions; AWC: overall adult words heard by child over 3 days of recording

Discussion

The purpose of the current study was to determine which, if any, parent communicative behaviors during a parent-infant play session were associated with infant volubility during the same session. Correlational analysis showed that the only parent behavior that was significantly correlated with infant volubility was verbal responsiveness. Children of parents with higher rates of verbal responsiveness vocalized more during the play session. There was no consistent relationship between quantity of language produced by the parent or any of the measures of communication style (e.g., use of questions, directives, engaging/excited expressions, etc.). Surprisingly, there was also no consistent relationship between overall amount of adult language heard by the child (as measured over 3 full days of recording) and infant volubility, either during the play session or over the 3 days of recording. There were, however, positive correlations between infant volubility during the play sessions and overall infant volubility as well as between adult words produced during the play sessions and overall counts of adult words produced. In other words, overall "talkative" babies babbled more during the play sessions and overall taciturn babies babbled less regardless of their parent's communicative behaviors, with the exception of verbal responsiveness. Similarly, overall talkative parents (i.e. those who produced more words heard by their child over 3 days of recording) were those who produced the most words during the play sessions.

The lack of a concurrent relationship between the number of words produced by parents and infant volubility is consistent with the results of Franklin, et al. (2014), who also found no relationship between parent and infant volubility during play sessions in a laboratory. Thus, while increased parent talk has consistently been associated with better language outcome for young children, this is the second study that has found that increased parent talk is not necessarily associated with concurrent infant volubility. It may be that the relationship between overall quantity of language input heard by infants and language development may only emerge over an extended period of time, not at a single measurement point.

Results of the current study are also consistent with previous work showing that contingent responsiveness by the parent (both verbal and non-verbal) shapes infant vocalization (Dunst et al., 2010; Goldstein & Schwade, 2008). A number of recent studies have concluded that in addition to just quantity of language input, quality of the communicative interaction (e.g., parental responsivity) is an important factor in language development (Zimmerman et al., 2009). Taken together, these results reinforce that parental contingent responsiveness is a strategy that should be encouraged in order to increase infant volubility and to support overall language development. Another parental behavior that has been found to increase and shape babble is imitation of the child's vocalization by the parent (Dunst et al., 2010). Surprisingly, imitation was not used by any of the parents during the plays sessions analyzed. While it's possible that the parents did imitate their children's' utterances during other activities and interactions, the complete lack of examples of imitations suggests that this is likely a relatively infrequent behavior. Thus, imitation as a strategy to increase infant volubility may need to be directly taught to and practiced by parents of children who are at risk for language delay.

A final, and important, consideration is that infant vocalization may shape parental communicative behavior as much as parent behavior shapes infant vocalization. That is, parents' interaction and communication style may change based on how much – or little – their child vocalizes. The direction of the change in parent communication, however, may vary depending on the parent and the specific behaviors of the child. For example, a parent of a child who naturally babbles very little may also reduce the amount of input they provide because they are not "pulled in" to communicative interactions by their child. On the other hand, a parent who observes that their child is not vocalizing very much may increase their overall language input as well as their use of engaging and animated expressions in an attempt to encourage more babbling. This bidirectional influence may explain why the hypothesized relationship between most of the parent communicative behaviors studied and infant volubility was not observed. Evidence for this possibility may exist in the observation that the mother of the baby who babbled the least produced the most adult words while the parent of the infant who vocalized the most produced fewer words per minute than all but 2 other parents.

While the results of the current study are in many ways consistent with previous work, it is important to consider limitations that may have impacted results. An important limitation of the present work is the relatively small sample size; caution should be used in generalizing results based on just 20 parent-infant dyads. Furthermore, data are based on a volunteer sample of relatively high-educated and ethnically homogenous participants and results based on a more diverse sample may have been different. Additionally, the activity of playing with books may have influenced the communicative interaction and different results may be found if interaction during different activities is analyzed. Future work should explore this possibility. Finally, any clinical implications of the current findings rest on the assumption that increasing infant volubility during the prelinguistic stage of development will have a direct, positive influence on language development and growth. To this author's knowledge, empirical evidence for this assumption is not available and remains to be explored in future work.

Conclusion

Results of the current study contribute to the evidence showing that parental contingent responsiveness plays an important role in language development, influencing both concurrent infant volubility as well as later language growth. The other parental communicative behaviors investigated, however, did not have an obvious impact on infant volubility. The results suggest that in working with families of children who are at risk for or who are already exhibiting communication delay, emphasis should be placed on increasing parental responsivity rather than on increasing overall quantity of parent talk. Additionally, a previously identified strategy for encouraging babble, imitation of child vocalizations, was not a strategy used by the parents during the play sessions analyzed. An important implication of this finding is that imitation is a strategy that may need to be explicitly taught to and practiced by parents in order to become established as a consistent part of their communicative repertoire.

References

Dunst, C. J., Gorman, E., & Hamby, D. W. (2010). Effects of adult verbal and vocal contingent responsiveness on increases in infant vocalizations. *Center for Early Literacy Learning, 3*(1).

Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Nathani Iyer, S., Lee, C., . . . Oller, D. K. (2014). Effects of parental interaction on infant vocalization rate, variability and vocal type. *Language Learning and Development*, *10*(3), 279-296.

Gilkerson, J., & Richards, J. (2009). The power of talk: Impact of adult talk, conversational turns, and TV during the critical 0-4 years of child development. *LENA Research Foundation,*

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(13), 8030-8035. doi:10.1073/pnas.1332441100 [doi]

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515-523. doi:10.1111/j.1467-9280.2008.02117.x [doi]

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509-516.

Hart, B., & Risley, T. (2003). The early catastrophe. *American Educator*, *27*(4), 6-9.

Hsu, H., Fogel, A., & Messinger, D. S. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: Factors associated with quantitative and qualitative differences. *Infant Behavior and Development*, *24*(1), 107-128.

Jakobson, R. (1968). *Child language: Aphasia and phonological universals* Walter de Gruyter.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, *59*, 441-449.

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry*, *52*(5), 588-598. doi:10.1111/j.1469-7610.2010.02332.x

Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech, Language, and Hearing Research*, *39*(1), 153-165.

Stoel-Gammon, C. (1997). Phonological development in down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews, 3*(4), 300-306.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children*. *Journal of Child Language, 38*(1), 1-34.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, *72*(3), 748-767.

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. Pediatrics, *124*(1), 342-349. doi:10.1542/peds.2008-2267.

Entropy as a measure of mixedupness in erroneous speech

Dimitrios Sotiropoulos[1], Elena Babatsouli[2]
dimsotirop@yahoo.com, ebabatsouli@ismbs.eu

[1]Technical University of Crete, [2]Institute of Monolingual and Bilingual Speech

Abstract. There are several types of grammatical and phonological errors that appear in normal and disordered speech during first language, second language, and bilingual development. In the literature, erroneous speech is evaluated by measuring these types of errors either individually (e.g., phoneme substitutions or deletions) or cumulatively (e.g., proportion of consonants correct (PCC), proportion of words correct (PWC), mean length of utterance (MLU) and its proportion (PMP) to the targeted MLU, phonological mean length of utterance (PMLU) and its proportion (PWP) to the targeted PMLU. These cumulative measures depend linearly on their component(s) and, consequently, their sensitivity to changes in their component(s) is constant. There are, however, instances in speech evaluation when a measure of higher sensitivity than that of linear measures would be advantageous in discriminating performance between different speech samples. For this reason, *Entropy* (E) is proposed as a measure for evaluating speech by computing the mixedupness of different types of errors and correctly produced (as targeted) grammatical or phonological parameter(s). Speech entropy is defined as $E = - \sum p_i \log_2 p_i$, where p_i is the frequency of the ith type of realization in proportion to the frequency of the targeted grammatical or phonological parameter(s) under examination, so that $\sum p_i = 1$. The sensitivity of entropy is compared analytically to that of linear measures for different values of their grammatical or phonological parameter-components. The analysis is complemented by computing the entropy in a bilingual child's English speech at the age of three years for different categories of word complexity. The analysis and application demonstrate the usefulness of the measure for evaluating speech samples and its advantage over linear measures for a considerable range of values of the grammatical or phonological parameters under consideration.

Keywords: entropy, measure, errors, speech, phonology, child, bilingual

Introduction

The quantification of measured speech errors has been of interest in the literature at least for the last ninety years. Nice (1925) introduced the average length of sentence (ALS), in terms of the number of words, to measure progress in child speech during development. Brown (1973) introduced a similar measure, the mean length of utterance (MLU), counting morphemes in the utterance, as a simple index of grammatical development. In language sample analysis (LSA) widely used by speech-language pathologists, the mean length of response (MLR), i.e. the average length of sentence (ALS), has found another name, the mean length of utterance in words (MLUw) in order to distinguish it from Brown's mean length of utterance in morphemes (MLUm). A comparison of these two measures was examined by Parker and Brorson (2005) for 40 language transcripts of 28 typically developing English speaking children between the ages of 3;0 and 3;10. The two measures were found to be very well correlated suggesting that MLUw may be used instead of MLUm, as the former is easier to compute. However, in all these measures that are concerned with grammatical and not phonological parameters, correctness of phonological segments is ignored.

Measurements of produced phonological segments, consonants and vowels, have also been discussed in the literature for a long time (see, for example, Ingram 1981). Shriberg, Austin, Lewis, McSweeney and Wilson (1997) proposed a refined measure of the proportion of consonants correct (PCC), computing the number of consonants correctly produced in context in proportion to the targeted consonants in the speech sample. With respect to whole-word correctness, Schmit, Howard, and Schmitt (1983) found that the measure of whole-word accuracy (WWA) favorably complements the proportion of consonants correct (PCC), based on data collected from children between the ages of 3 years and 3 three years and 6 months.

Measurements addressing phonological word complexity in child speech were proposed by Ingram and Ingram (2001) and Ingram (2002), the phonological mean length of utterance (PMLU), and its proportion to the targeted mean length of utterance, the proportion of word proximity (PWP). In computing the produced PMLU, correct consonants are counted twice as much as are vowels and substituted consonants. PMLU and PWP were defined as the arithmetic mean of their corresponding single word values in the utterance. Taelman, Durieux, and Gillis (2005) discussed how to use CLAN (MacWhinney, 2000) to compute PMLU from children's speech data. Several researchers have employed these measurements to evaluate speech performance not only in one language but also across two languages in bilingual child speech. Among these, Bunta, Fabiano-Smith, Goldstein, and Ingram (2009) compared 3-year old Spanish-English bilingual children to their monolingual peers to compute, among other quantities, PWP and the proportion of consonants correct, PCC. They found that while PWP and PCC differ in general, bilinguals only differ on PCC from their monolingual peers in Spanish. They further found that when comparing the Spanish and English of the bilingual participants, PCC was significantly different, but PWP was almost the same. Burrows and Goldstein (2010) compared PWP and PCC accuracy in Spanish-English bilinguals with speech sound disorders to age-matched monolingual peers. Macleod, Laukys, and Rvachew (2011) compared the change in PWP to that in PCC for two samples of twenty children each, both taken at the age of 18 months and 36 months. One of the samples involved monolingual English children while the other involved bilingual French-English children. For each sample, their results showed that the change in PWP was larger than that in PCC.

The proportion of phonological word proximity (PWP) was further examined by Babatsouli, Ingram, and Sotiropoulos (2014) not only per word but also *cumulatively* for all the words in a speech sample. They obtained an analytical expression for phonological word proximity (PWP) in terms of the proportion of consonants correct (PCC), the proportion of phonemes deleted minus added (PPD) and the proportion of vowels (PV) in the targeted speech sample. PWP is thus computed as the weighted average of single-word PWPs and not as their arithmetic mean as done in all previous studies. This way, the quantitative effects of PCC, PPD and PV on computed PWP are directly realizable in the whole speech sample.

In all of the above speech performance measurements, grammatical and phonological, the common feature is that they depend linearly on their components under consideration. This means that their sensitivity to changes in their components is constant, i.e. their slope is constant. Consequently, the effect of the component changes between different speech performances on the measures is known *a priory*. In practice, there are times that it will be advantageous to have a measure which is more sensitive than linear measures in order to enable better differentiation between speech performances.

For this reason, in the present paper, entropy is proposed as a measure of mixedupness in erroneous speech. The concept of entropy was introduced by Boltzmann in the 1870s as a measure proportional to the logarithm of the number of microstates that ideal gas particles could occupy. The definition of entropy used in the present study is the one proposed by Shannon (1948a, b) to measure the amount of information transmitted in a message, that is, $E = - \sum p_i \log_2 p_i$, where p_i is the probability that a particular piece of information of the message is transmitted. In our study here, p_i will be the frequency of the ith type of realization in proportion to the frequency of the targeted grammatical or phonological parameter(s) under consideration, so that $\sum p_i = 1$. In language, but not in the context of errors, entropy has been used to measure mixedupness of grammatical as well as phonological components, the latter initiated by Roman Jakobson, as discussed by Goldsmith (2000).

It will be shown that there is a considerable range of values of the grammatical or phonological parameter-components for which this measure is more sensitive than linear measures to changes of the components between different speech performances. This will be done first for correct-incorrect speech in terms of produced morphemes and consonants, comparing entropy to its linear counterparts, that is, the proportion of morphemes proximity (PMP) and the proportion of consonants correct (PCC).

Next, entropy will be defined for the different types of consonant realizations in a speech sample, correct, substituted, or deleted. The entropy defined so will be compared in terms of its sensitivity to a linear counterpart measure that we name proportion of consonants proximity (PCP), and is derived from the proportion of phonological word proximity (PWP) for the whole speech sample (Babatsouli et al., 2014) by setting the proportion of vowels (PV) equal to zero.

The analysis will be complemented by an application. A bilingual child's speech data in English at the age of three years will be utilized to compute both the child's entropy of correct and incorrect consonants and the entropy of consonants realizations (correct, substituted, deleted) for two categories of word complexity: monosyllabic words that only include singleton consonants and monosyllabic words with consonant clusters. The computed entropy will be compared to two linear counterpart measures, PCC and the proportion of consonants proximity (PCP), to see which measure differentiates better the child's speech performance across the two categories of word complexity.

The paper will end with conclusions followed by the list of references.

Entropy of morphemes or consonants correct/incorrect

The proposition here is to use entropy as a measure of mixedupness of different grammatical or phonological realizations in a speech sample that contains errors with respect to the targeted speech. The measure of entropy is defined as in Shannon (1948a, b). In this section, realizations will be limited to correct and incorrect without specifying incorrectness. Therefore, the entropy (E) of morphemes or consonants in a speech sample is defined as

$$E = - p \log_2 p - (1-p) \log_2(1-p) \quad (1)$$

where p is the proportion of produced morphemes to the targeted morphemes or the proportion of correctly produced consonants to the targeted consonants. The corresponding proportion of incorrect realizations is 1-p. It is noted that when the speech sample either contains no errors, p=1, or is full of errors, p=0, entropy attains its minimum value, zero, while when correct and incorrect realizations are of equal proportion, 0.5, entropy attains its maximum value, 1.

From the practical point of view, it is important to know how well entropies of different speech performances are differentiated depending on the values of p between the two performances. In other words, for speech evaluation purposes it is important that the measure used is *sensitive* to changes of p, i.e., the measure is capable of detecting relatively small changes of p. The sensitivity of entropy to very small changes of p can be seen by obtaining the entropy slope which, by differentiation of equation (1), is given as:

$$dE/dp = \log_2[(1-p)/p] \quad (2)$$

We can see that the entropy slope is very much dependent on the value of p. Near p being equal to 0 or 1, the entropy slope is very large, decreasing away from these values of p, diminishing as p approaches 0.5. Therefore, it is guaranteed that entropy will be a good measure for differentiating speech performances if the values of the proportions (p) of correctly produced morphemes or consonants to the targeted morphemes or consonants of the two performances are not both near 0.5 and are both either smaller than 0.5 or larger than 0.5. In cases where the p of one performance is smaller than 0.5 and the p of the other performance is larger than 0.5, entropy could be a good measure in terms of its sensitivity depending on the exact values of p. These remarks may be easier to see graphically in figure 1 where the entropy is plotted versus 1-p; the plot will be identical versus p since E has the same dependence on p as on 1-p.

Next, entropy is compared with known linear measures for morphemes and consonants to see which one of the measures is better to use when differentiating or evaluating speech performances. The entropy of morphemes cannot be compared directly with Brown's (1973) mean length of utterance (MLU), since MLU does not involve proportions of morphemes. However, the ratio of MLU to the targeted MLU, name it proportion of morphemes proximity (PMP), appropriate for use when comparing speech performances across languages, can be compared to the entropy (E) of morphemes defined by equation (1). PMP is plotted versus the proportion of incorrect morphemes (1-p) in Figure (1) together with E.

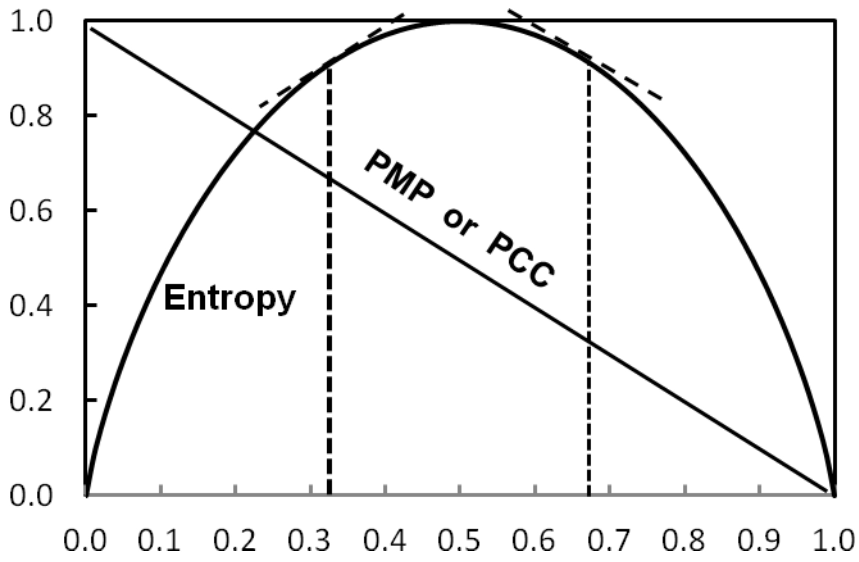


Figure 1. Entropy, PMP and PCC versus proportion (1-p) of incorrect morphemes or consonants

PMP has a constant slope equal to -1 (magnitude 1). Therefore, the sensitivity of PMP is uniform across values of p. Setting the magnitude of the entropy slope of equation (2) larger than 1, yields $p<1/3$ or $p>2/3$ as the values of p for which the entropy is guaranteed to be more sensitive than PMP to changes of p within this range of values. When $1/3<p<2/3$, PMP is guaranteed to be more sensitive than entropy to changes of p within this range of values. For changes of p crossing the values of the range boundaries, either measure can be more sensitive depending on the values of p across the two speech performances. At $p=1/3$ and $p=2/3$, the magnitudes of the entropy and PMP slopes are equal as shown in Figure 1.

The entropy of consonants can be compared directly to the linear measure PCC, the proportion of consonants correct (Shriberg et al., 1997). PCC has exactly the same dependence on the proportion of incorrect consonants (1-p) in the speech sample, as PMP has on incorrect morphemes. This is the reason that PCC is included in figure 1 as well. Therefore, the discussion above comparing the entropy sensitivity to the PMP sensitivity holds also true here in comparing the sensitivity of entropy with that of PCC and it will not be repeated.

Entropy of consonants realizations

Here the entropy of consonants realizations (correct, substituted, and deleted) will be defined and its sensitivity will be compared analytically to the sensitivity of a linear measure, the proportion of consonants proximity (PCP), for different speech performances.

Targeted speech consonants, TC, are realized either correctly as their targets, CC, or as substituted consonants, SC, or they are deleted, DC. Therefore, the following equation is true

$$TC = CC + SC + DC \qquad (3)$$

Dividing both sides of the equation by TC gives the following relationship between the proportions of realization types

$$SC/TC = 1 - CC/TC - DC/TC \qquad (4)$$

Now, the entropy of consonants realizations in a speech sample is defined as

$$E = - p_1 \log_2 p_1 - p_2 \log_2 p_2 - (1-p_1-p_2) \log_2(1-p_1-p_2) \qquad (5)$$

where p_1=CC/TC=PCC, p_2=DC/TC, and $1-p_1-p_2$= SC/TC. It is noted that when there are either no substitutions ($1-p_1-p_2=0$) or no deletions ($p_2=0$), or no correctly realized consonants ($p_1=0$), the entropy E of equation (5) attains its minima values (lower bound) according to the values of its non-zero components becoming identical to the entropy of equation (1) of the preceding section which was discussed. On the other extreme, the maxima values (upper bound) of entropy are attained when two of its three components are equal, with the maximum value of the upper bound of entropy resulting from the three components being equal. The upper and lower entropy bounds are plotted in Figure 2 against the proportion of consonants correct (p_1).

When both p_1 and p_2 change between two speech performances, the change in entropy is calculated as the difference between the two values of E of equation (5). When the p_1 and p_2 changes (dp_1 and dp_2) are very small, the change in entropy (dE) is given by

$$dE = dp_1 \log_2[(1-p_1-p_2)/p_1] + dp_2 \log_2[(1-p_1-p_2)/p_2] \qquad (6)$$

From this we can see that large entropy changes occur either near $p_1=0$ or near $p_2=0$ or near $1-p_1-p_2=0$. For either dp_1 or dp_2 being equal to zero, the entropy change becomes similar to that given by equation (2). When $p_1=(1-p_2)/2$, dp_1 does not affect the change in entropy and, similarly, when $p_2=(1-p_1)/2$, dp_2 does not affect the entropy change. In case $p_1=p_2=p$, the entropy change becomes

$$dE = (dp_1+dp_2) \log_2[(1-2p)/p] \qquad (7)$$

giving large entropy changes near p=0.5 (where there are no substitutions) and p=0 (all the realizations are substitutions), and small entropy changes near $p_1=p_2=1/3$ (the proportion of substitutions also equal to 1/3).

Now, the entropy change will be compared to the change of PCP, the linear measure of cumulative consonants speech performance in proportion to the targeted consonants. The proportion of consonants proximity (PCP) for a speech sample is defined as

$$PCP = (2CC + SC)/2TC \qquad (8)$$

in which correctly realized consonants are weighed twice as much as substituted consonants following Ingram and Ingram (2001) and Ingram (2002) who used these relative weights when computing the phonological length of utterance per word. Using equation (3) in equation (8) and rearranging, results in the following expression for PCP

$$PCP = (1 + PCC - PCD)/2 = (1 + p_1 - p_2)/2 \qquad (9)$$

This expression can be derived directly from the expression for the proportion of word proximity (PWP) for a speech sample given by Babatsouli et al. (2014) by setting the proportion of vowels (PV) equal to zero. For a given PCC, the lower PCP bound is given when deletions are most (no substitutions) and equal to 1-PCC, while the upper PCP bound is given when there are no deletions. Therefore, for a given PCC, PCP of equation (9) is bounded as

$$PCC \leq PCP \leq (1+PCC)/2 \qquad (10)$$

These PCP bounds are shown in Figure 2 together with the entropy bounds.

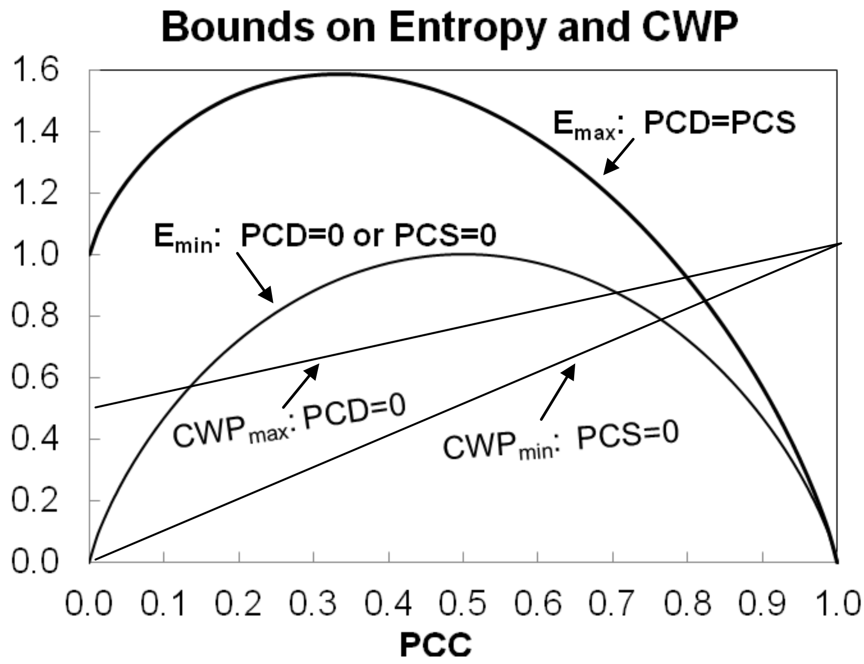


Figure 2. Maximum and minimum values of entropy and CWP versus PCC

Changes (ΔPCP) in PCP between speech performances are obtained from equation (9) as

$$\Delta PCP = (\Delta PCC - \Delta PCD)/2 = (\delta\ p_1 - \delta p_2)/2 \qquad (11)$$

This equation is valid for *arbitrary* changes in $\Delta PCC = \delta\ p_1$ and $\Delta PCD = \delta p_2$ between speech performances, whether they are small or large. When they are very small, $\delta p_1 = dp_1$ and $\delta p_2 = dp_2$.

Comparison between entropy changes and PCP changes can now be made in general either for very small changes in PCC and PCD by comparing equations (6) and (11), or for arbitrary changes in cases of special interest. Here, three cases are of special interest:

i) $\Delta PCC = \delta p_1 = 0$, and $\Delta PCD = \delta p_2$ is arbitrary (or vice versa). This case is applicable to differentiating speech performance across categories of word complexity: monosyllabic words, multisyllabic words, words without consonant clusters, and words with consonant clusters. Ingram (2015) reports that there are children with *speech sound disorders* (SSD) whose performance across these categories of word complexity is such that ΔPCC=0. For such cases, the magnitude of entropy changes in the performance across word categories is *guaranteed* to be *larger* than the magnitude of PCP changes when PCD values satisfy the following inequalities in respect of PCC values:

$$PCD < (1-PCC)(\sqrt{2}-1) \text{ or } PCD > (1-PCC)(2-\sqrt{2}) \qquad (12)$$

Therefore, for this range of PCC, PCD values, entropy will be a better measure to use than the linear PCP measure for differentiating performance across categories of word complexity in children with speech sound disorders. For example (12) yield that when PCC=0.5, entropy is more sensitive for PCD values outside the range (0.21, 0.29); and when PCC=0.2, entropy is more sensitive for PCD values outside the range (0.33, 0.47), in both cases with the upper limit of PCD values given by 1-PCC. When PCC and PCD values fall outside the range of values defined by inequalities (12), PCP will be a better measure to use than entropy for the same purpose. When PCC and PCD values cut across the limiting values defined by inequalities (12), the two measures need to be compared case by case as a conclusion cannot be reached in general *a priori*. Inequalities (12) were derived by setting the magnitude of dE of equation (6) larger than the magnitude of ΔPCP of equation (11) for $dp_1 = \delta p_1 = 0$.

i) ΔPCD= - ΔPCC. This may well happen when examining a child's performance across categories of word complexity with PCC being increased by the same amount as the decrease in the proportion of consonants deleted, meaning that the proportion of substituted consonants remains constant. The magnitude of entropy changes in the performance across word categories is *guaranteed* to be *larger* than the magnitude of PCP changes when PCC and PCD values fall in triangular regions defined by their vertices as:

$$\text{PCC}>\text{PCD}: (0,0), (2/3,1/3), (1,0) \text{ or } \text{PCC}<\text{PCD}: (0,0), (1/3,2/3), (0,1) \quad (13)$$

in which the first coordinate is the PCC value and the second coordinate is the PCD value. For this range of PCC, PCD values, entropy will be a better measure to use than the linear PCP measure for differentiating a child's speech performance across categories of word complexity. When PCC and PCD values fall outside the range of values defined by (13), PCP will be a better measure to use than entropy for the same purpose. When PCC and PCD values cut across the limiting values defined by the regions in (13), the two measures need to be compared case by case as a conclusion cannot be reached in general *a priori*. The range of values of (13) resulted from setting the magnitude of dE of equation (6) larger than the magnitude of ΔPCP of equation (11) with $dp_2 = - dp_1 = - \delta p_1$. This condition gave that the absolute value of $\log_2(p2/p1)$ must be larger than 1, from which the ranges of values of (13) were obtained.

iii) The third case of interest is motivated from the results of a study by Bunta et al. (2009), which concluded that when comparisons are made for the speech performance of bilingual English/Spanish children across the two languages, the phonological word proximity (PWP) did not vary while PCC values were significantly different. This indicates that PCP is almost invariable as well. Setting ΔPCP equal to zero in equation (11) gives ΔPCC=ΔPCD, meaning that while both correct consonants and deletions increase by the same amount, the proportion of substituted consonants to targeted consonants (1-PCC-PCD) decreases twice as much. In such a case, equation (6) yields that for all values of PCC and PCD (except for PCC=1 and PCD=0 or vice versa) the magnitude of the entropy slope which becomes

$$dE/dp=2\log_2[(1-p_1-p_2)/(p_1p_2)^{1/2}] \quad (14)$$

is larger than the magnitude of the PCP slope which is zero. Therefore, for such a case, entropy will be a far better measure to use than the linear measure of PCP, in order to differentiate speech performance.

An application

A Greek/English bilingual child’s performance is now evaluated on her English speech at the age of three years using both the entropy measure and the linear measures PCC, PCP, and PWP, the latter as defined by Babatsouli et al. (2014), cumulatively, for the whole speech sample, i.e.

$$\text{PWP} = \text{PCC}(1\text{-PV})/(2\text{-PV}) + \text{PPD}/(2\text{-PV}) \quad (15)$$

The measures will be compared to each other in terms of how well they differentiate speech performance across two categories of word complexity, monosyllabic words containing only singleton consonants and monosyllabic words containing at least one consonant cluster. The words used in computing the measures are all the English monosyllabic words produced by the child for a whole month at age 3;0, during digital recording of her speech. Details on how the data was acquired may be found in Babatsouli (2015) and it will not be repeated here as it is outside the scope of the present paper.

There were 47 monosyllabic words with singleton consonants produced by the child, which are: *back, beach, bed, bit, car, case, cat, cows, day, dog, door, fish, five, food, four, full, hair, have, head, here, hide, juice, kiss, leave, lick, look, loose, moon, more, mouse, nice, nose, now, pull, put, rain, red, right, sea, sit, table, wash, what, where, yes*.

Also, in the child's speech, there were 38 monosyllabic words with at least one consonant cluster, which are: *block, boots, box, bread, bridge, bring, brush, cold, called, clean, clear, clock, doesn’t,* *dream, farm, floor, found, grab, help, left, lost, milk, once, plain, seeds, slide, small, space, stopped, street, things, throw, toast, train, trash, trouble, try, washed*.

In a month's span, the child did not produce (realize) each targeted word in the same way, as she was in the midst of her phonological development, so that all of her word productions were included in calculating PCC and PCD; there were no vowel deletions. The results are:

singleton consonant words: PCC=0.64, PCD=0.0

consonant *clusters* words: PCC=0.51, PCD=0.19

Substituting these values in equations (1), (5), (9), and (15) yields values for the entropy of consonants correct/incorrect (E_1), the entropy of consonants realizations (E_2), the proportion of consonants proximity (PCP), and the proportion of phonological word proximity (PWP). The entropy of both consonant and vowel realizations (E_3) is also computed to be compared to PWP. For each word category, these values are:

singleton consonant words: E_1=0.94, PCC=0.64, E_2=0.94, PCP=0.82, E_3=1.53, PWP=0.87

consonant *clusters* words: E_1=1.00, PCC=0.51, E_2=1.47, PCP=0.66, E_3=1.91, PWP=0.69

From these values we conclude that: a) the proportion of consonants correct (PCC) differentiates the child's speech performance across the two word categories better than the entropy of consonants correct/incorrect (E_1). This is so because the two PCC values fall inside the range (1/3, 2/3) and in this range, as discussed in the first section above, it is guaranteed that PCC changes are larger than entropy changes, b) the entropy of consonants realizations (E_2) differentiates the child's performance across the two word categories better than the proportion of consonants proximity (PCP). This is not surprising in view of the discussion in the previous section above, since ΔPCC and -ΔPCD are almost the same (case ii) and the PCC, PCD values fall in the values defined by the left hand side of (13), c) the entropy of consonants and vowels realizations (E3) differentiates better the child's performance across the two word categories than the proportion of phonological word proximity (PWP).

It must be remarked that the application above is simply an example and in no way is intended to reach general conclusions on the usefulness of the applicability of the measures compared. Remarks comparing the measures in general were made in the analysis that was presented in the preceding sections.

Conclusion

Entropy was proposed as a measure of mixedupness in erroneous speech which is common in child phonological development. The entropy measure was defined and analyzed for consonants errors starting out without specifying their types and extending to consonant substitutions and deletions with respect to the targeted speech. The sensitivity of entropy, that is, its dependence on changes in its components, was compared to the sensitivity of linear measures in order to reach general conclusions on which measure is better to apply in evaluating speech performance and in differentiating performances between children as well as between word complexity categories of the same child. Specifically, the entropy of consonants correct/incorrect was compared analytically and in general to the linear measure of the proportion of consonants correct (PCC), and similar comparison was made between the entropy of consonant realizations to the linear measure of the proportion of consonants proximity (PCP). The analysis defined ranges in the values of the components (correct, incorrect, substitutions, deletions) in which either entropy or the linear measures are more sensitive and, thus, better in differentiating speech performance. While these ranges provide a guideline to practitioners on which measure to use for a large number of cases as far as measured speech components are concerned, they do not include all cases in general, as they are ranges of sufficient conditions, and therefore, we recommend that entropy is computed alongside the linear measures used to date, complementing speech evaluation for optimal results. The example used as an application and the possible scenarios of speech performance considered in the analysis clearly support this conclusion.

References

Babatsouli, E. (2015). Technologies for the study of speech: Review and an application [Special Issue on Language Disorders and ICT]. *Themes in Science and Technology Education*, *8*(1), 17-32.

Babatsouli, E., Ingram., D., & Sotiropoulos, D. (2014). Phonological word proximity in child speech development. *Chaotic Modeling and Simulation*, *4(3)*, 295-313.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Bunta, F., Fabiano-Smith, L., Goldstein, B. A., & Ingram, D. (2009). Phonological whole- word measures in three-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics and Phonetics*, *23*, 156-175.

Burrows, L., & Goldstein, B. A. (2010). Whole word measures in bilingual children with speech sound disorders. *Clinical Linguistics and Phonetics*, *24*, 357-368.

Goldsmith, J. (2000). On information theory, entropy, and phonology in the 20th century. *Folia Linguistica*, *XXXIV*(1-2), 85-100.

Ingram, D. (1981). *Procedures for the phonological analysis of children's language*. Baltimore, MD: University Park Press.

Ingram, D. (2002). The measurement of whole-word production. *Journal of Child Language*, *29*, 713-733.

Ingram, D. (2015). Whole-word measures: Using the pCC-PWP intersect to distinguish speech delay from speech disorder. In C. Bowen (ed.), *Children's speech sound disorders* (2nd ed., pp. 100-104). Oxford, UK: John Willey & Sons.

Ingram, D., & Ingram, K. (2001). A whole-word approach to phonological analysis and intervention. *Language, Speech and Hearing Services in Schools*, *32*, 271-283.

Macleod, A. A., Laukys, K., & Rvachew, S. (2011). The impact of bilingual language learning on whole-word complexity and segmental accuracy among children aged 18 and 36 months. *International Journal of Speech and Language Pathology*, *13*, 490-499.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Nice, M. M. (1925). Length of sentences as a criterion of a child's progress in speech. *Journal of Educational Psychology*, *16*, 370-379.

Parker, M. D., & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, *25*, 365-376.

Schmitt, L. S., Howard, B. H., & Schmitt, J. F. (1983). Conversational speech sampling in the assessment of articulation proficiency. *Language, Speech, and Hearing Services in Schools*, *14*, 210-214.

Shriberg, L., Austin, D., Lewis, B., McSweeney, J., & Wilson, D. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research*, *3*, 708-722.

Shannon, C. E. (1948a). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379-423.

Shannon, C. E. (1948b). A mathematical theory of communication. *Bell System Technical Journal*, *27* (4), 623-656.

Taelman, H., Durieux, G., & Gillis, S. (2005). Notes on Ingram's whole-word measures for phonological development. *Journal of Child Language*, *32*, 391-400.

Case studies of speech and language disorder in three children with autism spectrum disorder (ASD)

Roopa Suzana
roopasuzana@gmail.com

The English and Foreign Languages University

Abstract. Speech and language deficits occur in a number of developmental disorders like ASD, ADHD, cerebral palsy, Down's syndrome, etc. Inadequate language is a defining feature in Autism Spectrum Disorder. This paper examined and described the patterns of speech and language deficits in three children with ASD. The characteristic features of children with ASD, e.g., repetitive actions (hand flapping), self-destruction (head banging), temper tantrums and meltdowns, inadequate eye contact, sensitivity to touch and pain, etc., vary according to whether they are diagnosed as mild/borderline, moderate or severe. This paper examined the articulation and language disorders in three verbal children (aged twelve years), diagnosed to be 'mild' 'moderate' and 'severe' cases of autism. This study of the deviant patterns in their speech looked at certain aspects of speech and language, such as substitution of sounds, e.g., /r, w/ (with lip rounding). Substitution of the voiced alveolar non-lateral approximant with the voiced labio-velar semi-vowel, which is also an approximant to produce articulations, such as [θwi:], [dwes], [kwaɪ] and [gwi:n] instead of [θri:], [dres], [kraɪ] and [gri:n] respectively. Certain significant features were noted. It was possible to articulate /r/ when preceded by bilabial plosives, e.g., /pri:ʧ, braɪd, brʌðə/ and when /r/ occurs word initially, e.g., /red, rʌbə, raɪt, rəʊd/. The ability to articulate consonant clusters, addition of sounds, deletions, distortions/ substitutions and omissions of sounds, clenched teeth articulation, echolalia, scripting, incomplete sentences, sentence structure, use of the pronoun, comprehension/expression, fluency, voice quality and the ability to convey emotion through intonation were also studied. The study concluded with the differences and the commonalities that existed in their speech and language. It enabled the identification of sounds and sentence patterns that are abnormally produced, and concluded that deficits are greater in the child diagnosed as 'severe'. The findings of this study can help speech therapists design tailor-made speech therapies for such children.

Keywords: ASD, speech and language disorders, speech therapies

Introduction

Speech and language deficits occur in a number of developmental disorders like ASD, ADHD, Cerebral palsy, Down's syndrome, etc. Inadequate language for communication is a defining feature in Autism Spectrum Disorder. "One of the most striking features of the autistic child is impaired language: He communicates poorly or not at all, either by word or by gesture" (Churchill, 1978, p. 7). This paper examined and described the patterns of speech and language deficits in the English spoken by three verbal bilingual Indian/Greek children with ASD (aged twelve years), who were exposed to the English language from age of two years onwards. They were diagnosed to be 'mild' 'moderate' and 'severe' cases of autism. This study of the deviant patterns in their speech looked at certain aspects of speech and language such as: substitution of sounds, addition of sounds, deletions, distortions, articulation of consonant clusters, echolalia, scripting, incomplete sentences, sentence structure, use of the pronoun, comprehension/expression, fluency, phonation, voice quality and the ability to convey emotion through intonation. The aim of this study was also to relate the severity of autistic impairment with the severity of speech and language impairment of the child.

Autism and its diagnosis

Autism is a neurodevelopmental condition which affects the way in which information is processed in the brain. "Autism is a spectrum condition and as such takes many forms, from the non-verbal to the highly talkative for example, or from those who revel in sensory stimuli to those who find such encounters painful and distressing" (Murray, 2012, p. 1). Some major characteristic features exhibited by children with ASD as listed by the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition, May 2013) are given below:

A. Persistent deficits in social communication and social interaction across multiple contexts, as manifested by the following:
   1. Deficits in social-emotional reciprocity, ranging from failure of normal back-and-forth conversation to reduced sharing of interests and emotions, to failure to initiate or respond to social interactions.
   2. Deficits in nonverbal communicative behaviours used for social interaction, ranging from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expressions and nonverbal communication.
   3. Deficits in developing, maintaining, and understanding relationships, ranging from difficulties adjusting behaviour to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers.

B. Restricted, repetitive patterns of behaviour, interests, or activities, as manifested by at least two of the following:

   1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).

   2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns or verbal nonverbal behaviour (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat food every day).

   3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g, strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interest).

   4. Hyper- or hypo reactivity to sensory input or unusual interests in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement).

C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E. These disturbances are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay. Intellectual disability and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual disability, social communication should be below that expected for general developmental level. (Source: *www.autismspeaks.org/what-autism/diagnosis/dsm5-diagnostic -criteria*)

The above symptoms may vary in number and intensity based on which one is diagnosed as mild/borderline, moderate or a severe case of ASD. The diagnosis of autism is very complex. Simon Baron-Cohen says that it “is often carried out by a multidisciplinary team, typically taking two to three hours based on interview and observation” (Baron-Cohen, 2008, p. 37). The interview he notes will be conducted by “a child psychiatrist, clinical or educational psychologist, paediatrician and other health professionals” (Baron-Cohen, 2008, p. 38). It is generally the parents of the child and the information provided by them that aids diagnostic evaluation.

This paper does not deal with the criteria on which the diagnosis was made. For the purpose of this study, three children who were already clinically diagnosed as mild, moderate and severe cases of autism were selected. They were attending special schools that imparted special one-on-one education and training. The focus of this study has been on their speech and language for communication.

Aim and scope of this work

This work aimed to study the differences and the commonalities that existed in the speech and language of three verbal children of the same age (twelve years), diagnosed to be ‘mild’ ‘moderate’ and ‘severe’ cases of autism. It is hoped that this study will enable the identification of sounds and sentence patterns that are abnormally produced. The study will also look into all the aspects of speech and language deficits and conclude if we can relate the severity of autistic impairment with the severity of speech and language impairment of the child. The final findings of this study, it is hoped, may help speech therapists design tailor-made speech therapies for such children.

Selection of subjects

Three verbal children who were diagnosed as borderline, moderate and severe cases of autism were selected. The special schools to which they attended were visited and a conversation was made with them. Once the children felt a little comfortable, their speech was recorded. The three children were of the same age (twelve year olds born in 2003). One of them was a girl (a mild/borderline case) and the other two were boys (a moderate and severe case respectively). On the parents’ request, the names of these children will not be used in this paper. They were referred to as **Child I** (mild/borderline), **Child II** (moderate) and **Child III** (severe). Every other detail remained the same.

Case histories

It is essential for a study like this to have a detailed record of facts concerning the child, e.g., his/her parents, birth, milestones, the age at which the diagnosis was made, early intervention, treatment/therapies given etc. Given below is this information on the three children in in Table 1.

Table 1. Case histories

	Child I (mild/ borderline)	**Child II** (moderate)	**Child III** (severe)
Consanguineous marriage of parents	No	No	No
Milestones – Normal/ Delayed Head holding Crawling Sitting Standing Walking Speech development	Delayed 4-5 months 7 months 11 months 15 months 23 months At 10 months but regressed at 20 months later showed crests and troughs in development until 5yrs.	Delayed 5-6 months 12 months 14 months 18 months 24 months Started speaking at 4yrs	Delayed 14 months 20 months 24 months 28 months 36 months Started speaking after 10 years
Seizures	One episode (mild) at 6 years	Few episodes	Frequent episodes, first one 3 days after birth
Early Intervention	Yes	Yes	Yes
Hyperactivity	Yes	Yes	Yes
Diet	Yes – CFGF diet for 2-2.6 years	NO – CFGF diet	No – CFGF diet
Age of Diagnosis	24 months	20 months	20 months
Therapies Physiotherapy Speech therapy Occupationl atherapy	 Yes- 6 months Yes- 10 months Yes 12 months	 Yes- 10 months Yes- 24 months Yes 30 months	 Yes- 12months Yes- 24 months Yes 28 months

Group / Play therapy	Yes- 18 months	Yes- 24 months	Yes- 24 months
Education	Special school With integration (class 3) Niraj Public School, Hyderabad	Special school Spastic Society of India, Mumbai	Special school Spastic Society of India, Mumbai

Analysis of the data

The use of *phonemic transcription* to describe these speech patterns would assume a normal phonology. Since disordered speech has some degree of distortion an attempt has been made to describe these speech patterns with the help of *phonetic transcription* using the ExtIPA symbols.

At the very outset, the recorded data was subjected to auditory perception. After listening to it several times, initial impressions were noted. Later samples of speech were selected and an acoustic analysis was done using Praat (Boersma & Weenink, 2015). Acoustic analysis may well help the clinician to decide on the articulations produced by the patient (Ball, 1993, p. 184). This analysis proved to be very useful as all the additions, deletions and substitutions of sounds and the study of voicing/devoicing of certain sounds and the inconsistent amplitude used by these children could be substantiated by instrumental evidence.

CHILD I

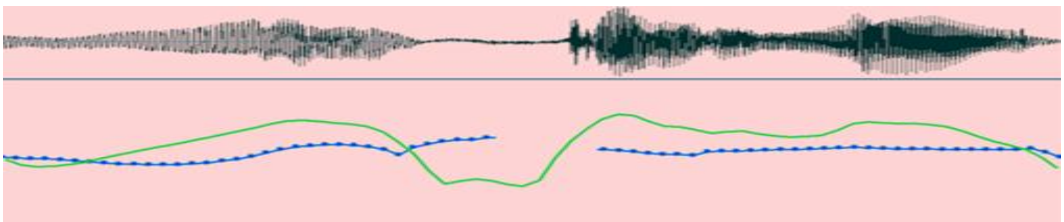
Diagnosed to be a case of mild/ borderline autism
Age: 12 years, Gender: girl, First symptom: no eye contact (at six weeks of age),
Regression of speech (at 20 months).

Substitution 1: /r/ → [w] (with lip rounding)

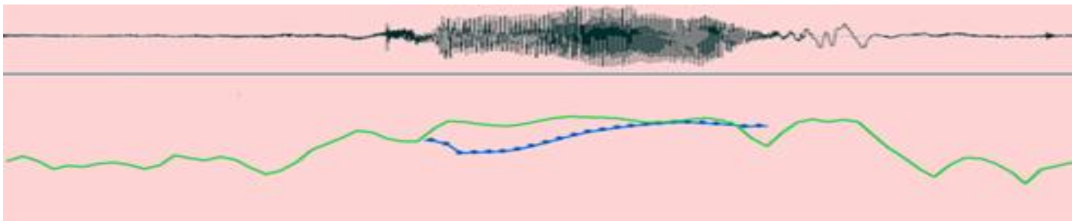
It was observed that there was a substitution of the voiced post alveolar frictionless continuant (non-lateral approximant) with the voiced labio-velar semi-vowel, which is also an approximant. Some examples of the same process observed in the child's speech are:

green → [gwi:]
three → [twi:]
dress → [dwes]
cry → [kwaɪ]
scratch → [skwæʧ]

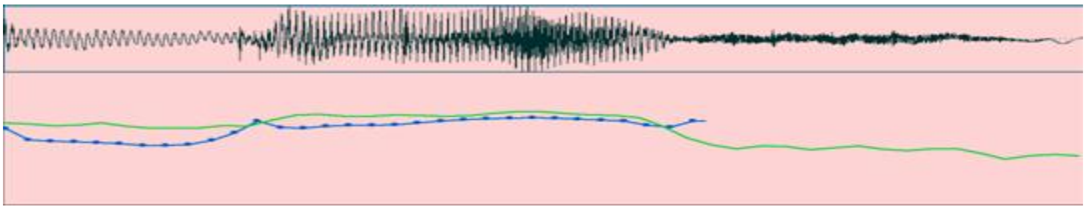
The speech waves of some of the words along with the audio files are given below:



green colour **[gwi:n kʌlə]**



***three* [twi:]**

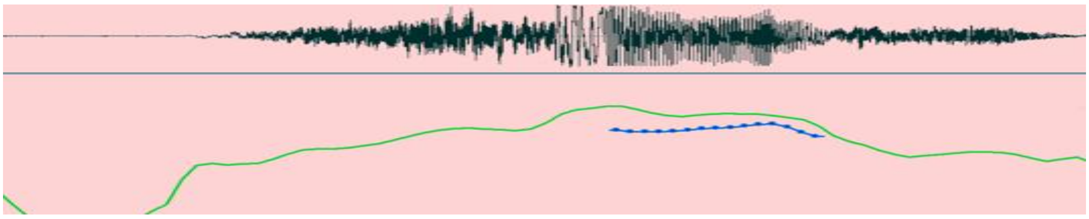


***dress* [dwes]**

However, it was also observed that this child could articulate /r/ when preceded by bilabial plosives as in *preach, bride, brother*, and when /r/ occurred word initially as in *red, rubber, right and rode* and word-medially as in *afraid* and *celebration*.

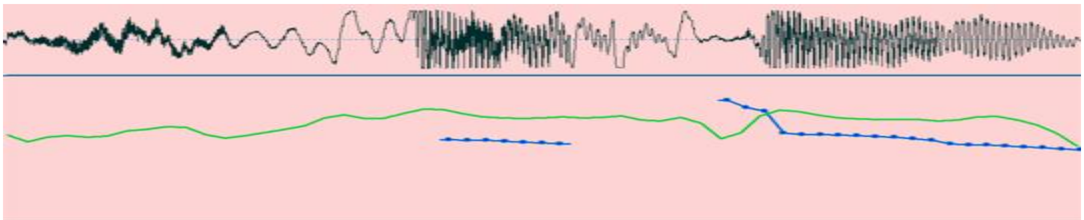
Substitution 2: /ʃ/ → [s]

ship → [sɪp]
shoes → [su:s]



***shoes* [su:s]**

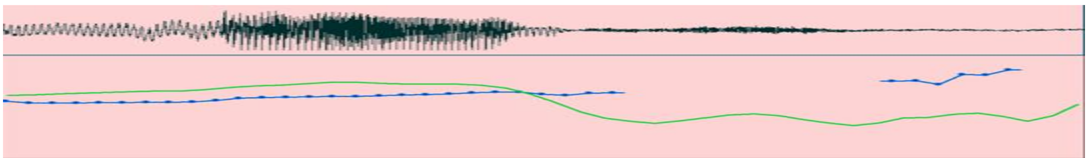
Substitution 3: /ŋ/ → [m]



***sweeping* [swi:pɪm]**

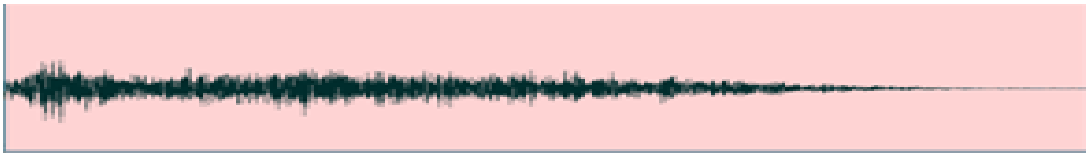
Some other words like *mopping* and *swimming* too were articulated as **[mɔ:pɪm]□ [swi:mɪm].**

Substitution 4: /z/ → [s]



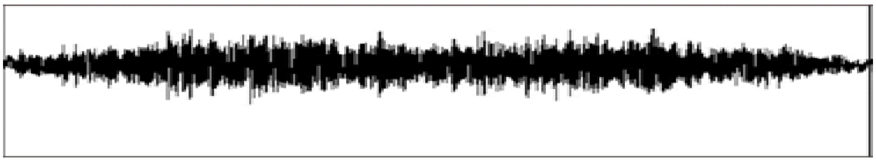
***nose* [no:s]**

The break in the pitch curve clearly shows that /z/ is realized as [s]. Given below is the zoomed wave of [s].

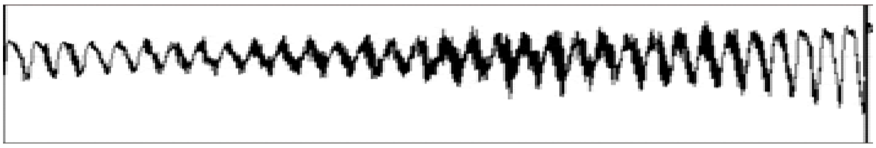


[s]

For a quick comparison, let us look at the speech waves of [s] and [z] given below.



Speech wave of [s]



Speech wave of [z]

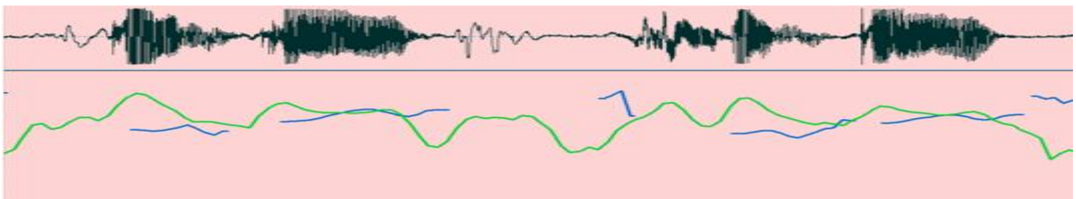
Therefore, it is evident that the child has substituted /z/ for [s].

In the four substitution patterns, 1, 2, 3 and 4, we notice that there is only change in the place of articulation while manner of articulation remains the same. (1. Approximants; 2. and 4. Fricatives; 3. Nasals).

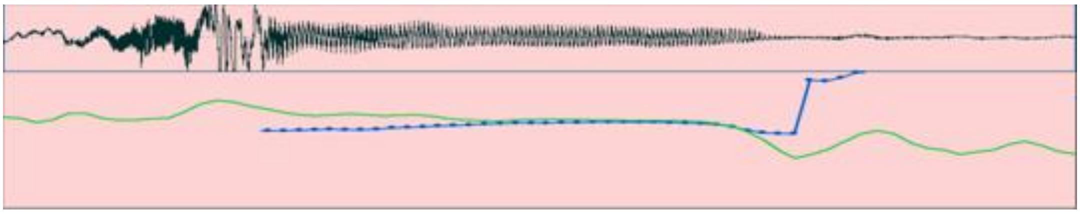
In Substitution 2, a palato-alveolar place of articulation changes to an alveolar place and in Substitution 3, we see that a velar place of articulation changes to a bilabial place. Therefore, we can say that fronting process is involved. There is no change in the voicing of the sounds. However, in the last substitution, we see that a voiced fricative is replaced by a voiceless fricative, the place of articulation being the same (alveolar).

Deletions/Elisions of sounds

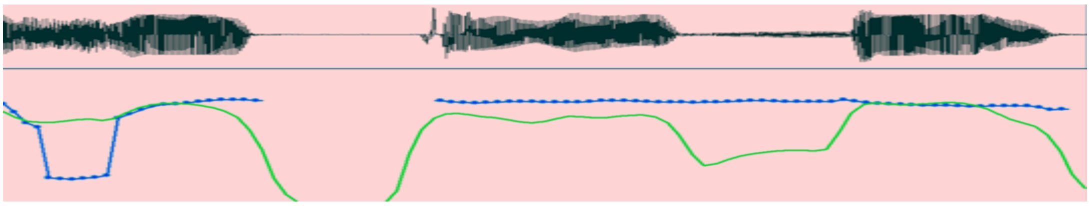
It was observed that in the speech of this child the sound that was almost always deleted was the voiced alveolar nasal, /n/.



***Thirteen, fourteen* [təti:] [fo:tɪ]**

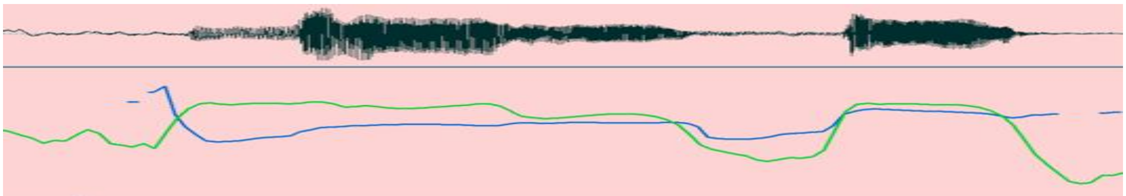


***June* [ʤu:]**



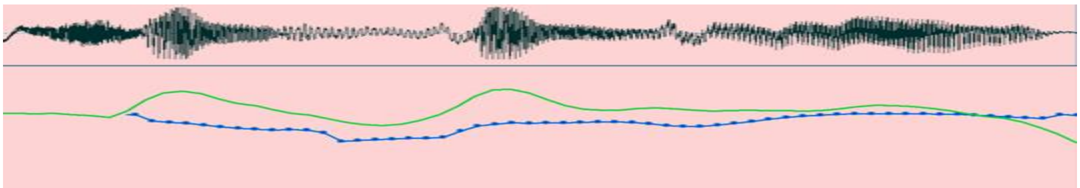
***open* in '*Open your mouth'* [o:p jɔ: maʊt]**

However, it was noticed that [n] was produced when it occurred in the initial position.



***November* [nəvembə]**

In the *January*, it was noticed that apart from /n/ deletion, there is also the deletion of the palatal /j/ and the diphthong /ʊə/. These three sounds were compensated by just one sound, a voiced bilabial plosive /b/. This process involves a change in the manner of articulation, as well. This is the only distortion of sounds noticed in the child's speech. The speech wave of *January* is given below.



***January* [ʤæbərɪ]**

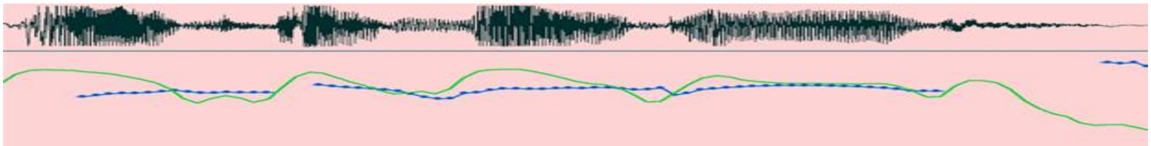
Word omission

This child did not largely omit words in her speech.

- The poem *Johny Johny Yes Papa* was recited without a single word being omitted in the entire poem.
- The *days of the week* were named and no day of the week was omitted.
- The *letters of the English alphabet a-z* were told without missing out on even one of them.
- The *numbers from 1-100* too were told without missing out any number.
- However, while listing out the months of the year, the child missed out the month 'May'.

Ability to articulate consonant clusters

This child was able to articulate consonant clusters. Sometimes, even though there was substitution of sounds, the child showed clear ability to produce consonant sounds in succession. The word *bottles* was produced with a final cluster as is evident in the speech wave below. The break in the pitch curve below the speech wave clearly shows the change of the final sound from /z/ to [s]. Nonetheless, the cluster was effectively produced.



***water bottles* [wɒtə bɒtls]**

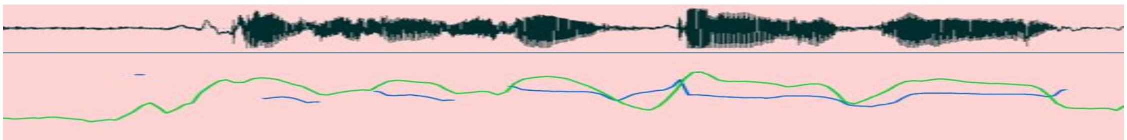
However, there was one instance that she could not produce a cluster and that cluster is /br/ in *brown*.

Echolalia and scripting

No echolalia was observed in the speech of this child. This child showed occasional scripting but the interesting part is that there was no irrelevant speech. For example, she would recite a part of the rhyme '*what a hog to swallow a dog'* whenever she saw a dog on the road!

Complete sentences/ Sentence structure

The child's speech was at a 4-5 word sentence level. The sentences were complete. Some of the sentences are: *This is a tiger*, *This is a nose*, *This is a stomach***,** etc. The child showed knowledge of the singular and plural. Given below is the speech wave of one of the sentences. The child was able to use the SVO order in the sentences.



This is a tiger

It was also observed that she produced the plural of *tooth* as *teeths*. However, there was evidence of almost immediate self-correction and the plural form '*teeth'* was used.

Use of the personal and possessive pronoun

The child has made fair use of the personal pronouns *I* and the possessive pronoun *my*. When asked questions, like '*How many eyes do you have*' or '*When is your birthday?* She answered in full sentences using personal/possessive pronouns like *I* and *My*. The responses were: *'I have two eyes*' and **'***My birthday is on 22nd September'.*

Possessive case

Sentences like: '*This is mommy book*' were observed where there was no use of the possessive case. There were four other sentences without the use of the possessive case. When the child was asked the names of her parents and sister, her responses were as given below:

'My father name is ....' *'My mother name is…'*
'My sister name is…' *'My school name is. .'*

Comprehension/Expression/Pragmatics

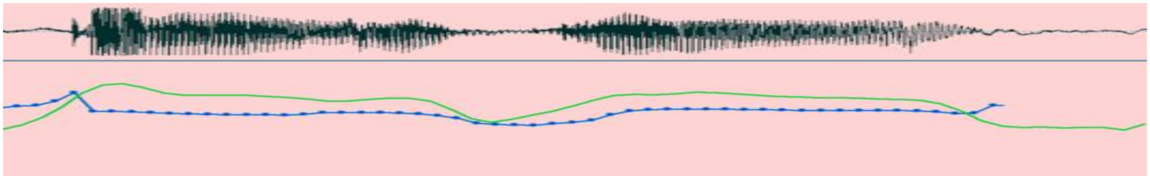
The child's comprehension was found to be fair. She understood and was able to carry forward three step instructions. Her understanding of language and its usefulness was also fair. For example when asked what things were put in the fridge, she answered relevantly unlike the answers given by Child II. The answers given were '*water, vegetables milk, horlicks bournvita*. She could also answer questions like '*Do you like your school?*' and '*Can you name your friends in school?*' relevantly and named three of her friends in school correctly.

Fluency

This child spoke sentences of 4-5 words fluently but she could not initiate a conversation. She answered almost all questions asked. She could answer almost all yes-no questions and questions beginning with *What*, *Who* and *Where* but not questions beginning with *When* and *Why*. This child can also ask for things she needs.

Phonation

Many speech disorders exhibit an inability to manipulate the voicing contrast of some of the sounds in the target language. There was just one instance where the voiceless alveolar plosive sound /t/ was realized as a voiced sound. Given below is the speech wave of the word *tiger* where /t/ is realized as a voiced sound. This is evident from the presence of the pitch curve during the articulation of the /t/ in [taɪgə]. However, the word did not sound like [daɪgə].



***tiger* [taɪgə]**

Voice quality and intonation

The child's voice quality was neither completely normal nor hoarse. This child had no control over the amplitude of her voice. She sounded unnecessarily too loud on some occasions and too soft on some other occasions. There was no evidence of the use of intonation patterns in the child's own speech but she did distinguish between a strict or an angry tone (falling pitch) and a polite tone (a rising pitch) in others voices.

CHILD II.

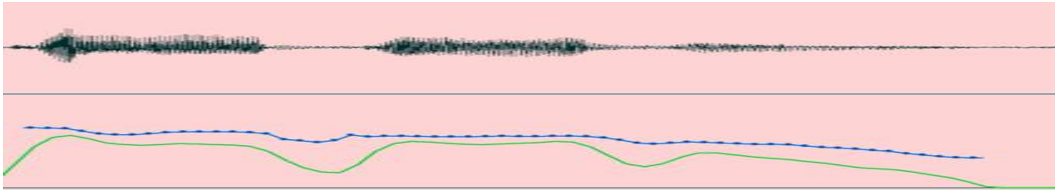
Diagnosed to be a case of moderate autism
Age: 12 years, Gender: boy, First symptom: delayed speech development

Substitutions

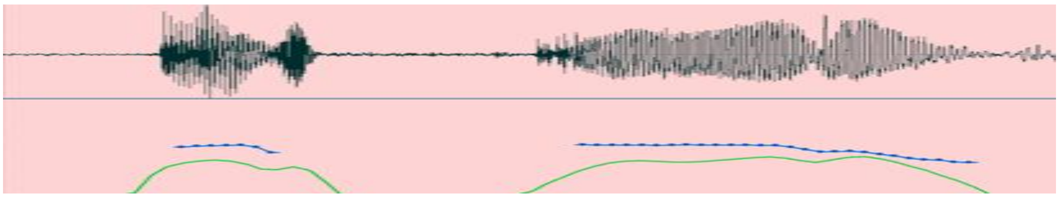
1. *Target sound* /r/ ⟶ *Output sound* / l /
2. *Target sounds* /t/ /s/ /k/ ⟶ *Output sound* /θ/

The words *aeroplane*, *tongue*, *Saturday*, *ice-cream* and *nose were* pronounced as [elople:n], [θʌŋg], [θa:təde:], [a:sθi:m], [no:s], respectively.

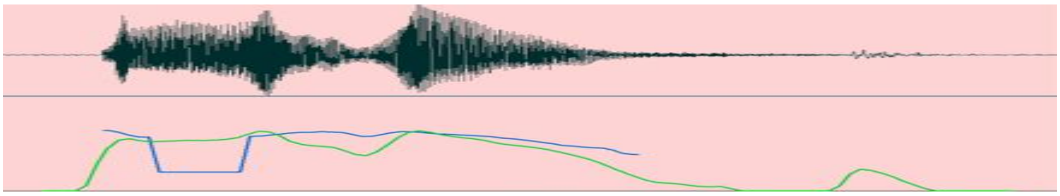
We see a general fronting process adopted by this child. The post-alveolar frictionless continuant has been replaced by the alveolar lateral. The alveolar and velar sounds have been substituted by the dental sound. The speech waves of some of the words are given below.



Saturday **[θa:təde:]**

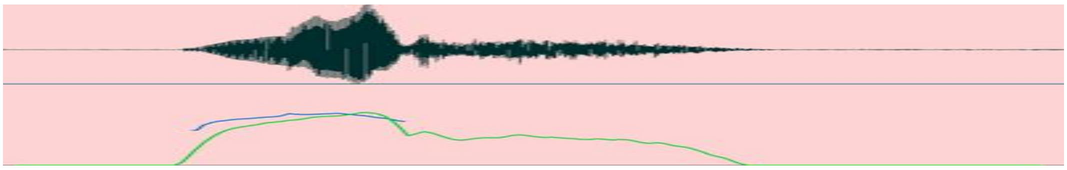


ice-cream [a:sθi:m],



tiger **[θaɪgə]**

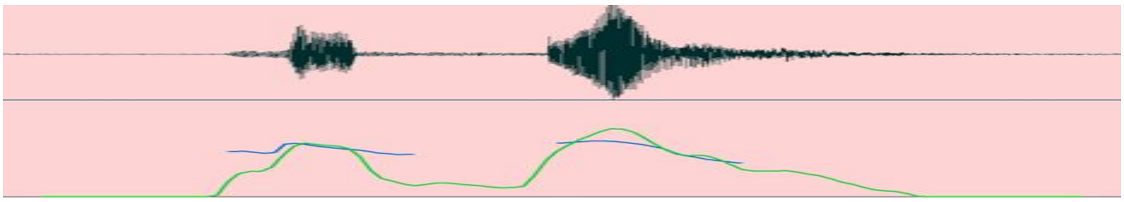
In the word *nose* however we see that the voiced alveolar fricative /z/ was substituted by its voiceless counterpart [s].



nose **[no:s]**

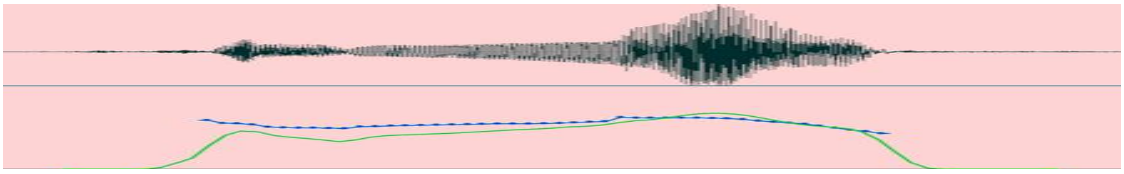
Deletions/Elisions of sounds

In the word *birthday*, it was observed that the fricative sound [θ] was deleted and the word was uttered as [bəde:]. Partial final devoicing was observed in /d/ of the second syllable [de:]. Although this voiced plosive does not occur at a word final position, it is evident in the speech wave given below. The break in the pitch curve shows partial devoicing. The stress is on the second syllable.



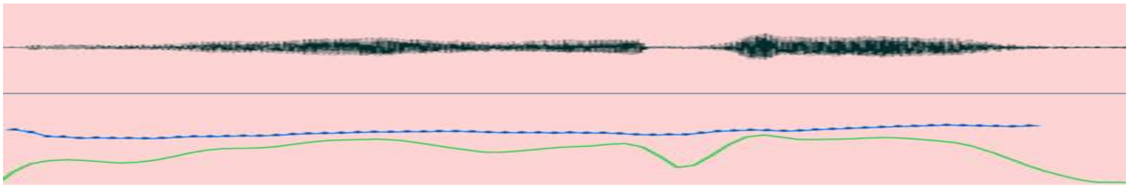
birthday **[bəde:]**

The following speech wave is that of the word *stomach* was pronounced as [sʌma:]. The /t/ after /s/ and the final sound /k/ were dropped.



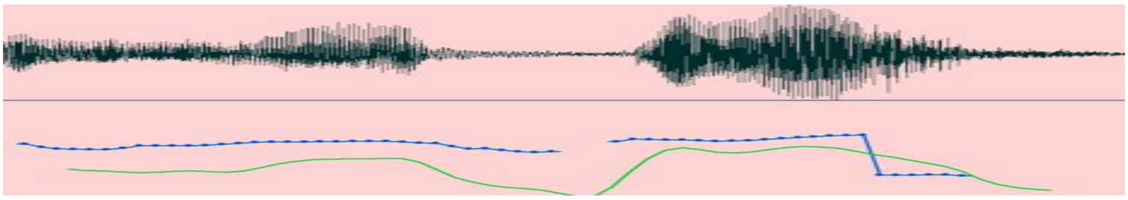
***stomach* [sʌma:]**

In the word *Wednesday*, the initial labio-velar semivowel /w/ was pronounced as [ʊ] where there was neither rounding of the lips nor was there a labio-dental articulation. The disyllabic word was divided into three syllables and uttered as [ʊe-nə-de]. The fricative sound /z/ was dropped and the final diphthong /eɪ/ was a long pure vowel [e:] This word was also articulated with very low strength of articulation.



***Wednesday* [ʊe-nə-de]**

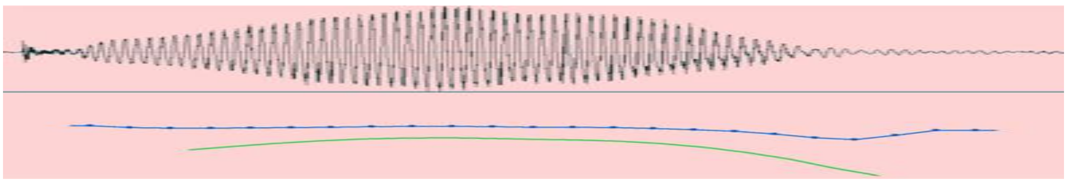
In the word *Thursday*, it was observed that the fricative sound /z/ was dropped and the final diphthong /eɪ/ was a long pure vowel [e:] Partial final devoicing was observed in /d/ of the second syllable, although the voiced plosive does not occur at a word ending position.



***Thursday* [tɜ:de:]**

In the words *thirteen, fourteen, seventeen, eighteen, nineteen and watermelon*, the final /n/ sound was deleted.

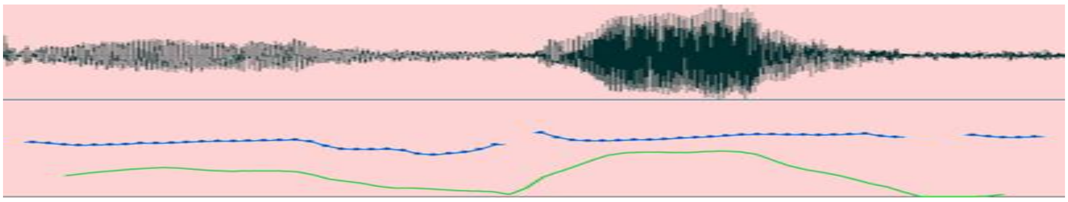
In the words *teeth* and *table* the word final sounds /θ/ and /l/ were dropped. Given below is the speech wave of *teeth.*



***teeth* [ti:]**

Ability to articulate consonant clusters

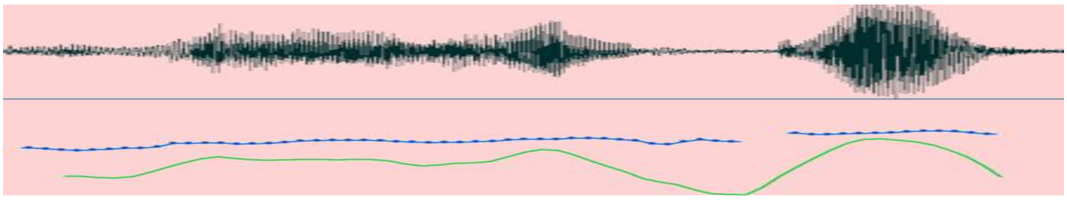
It was observed that the child found it difficult to articulate *consonant clusters*. Here, we observe the child's inability to articulate consonant clusters: /st/, / tj/, /nz/. Given below is the speech wave of the word *Tuesday*.



***Tuesday* [tu:zde:]**

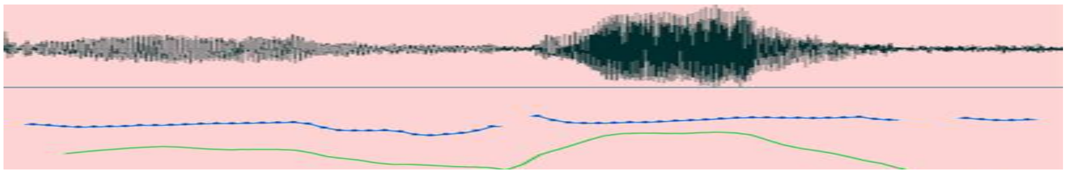
Phonation

This child's speech showed an inability to manipulate the voicing contrast to a large extent. There was evidence of *voicing* of voiceless sounds and *devoicing* of voiced sounds The speech wave given below is that of the word *Friday*. It has been observed that the voiceless fricative /f/ is realized as a voiced consonant. And there is also final partial devoicing of /d/.

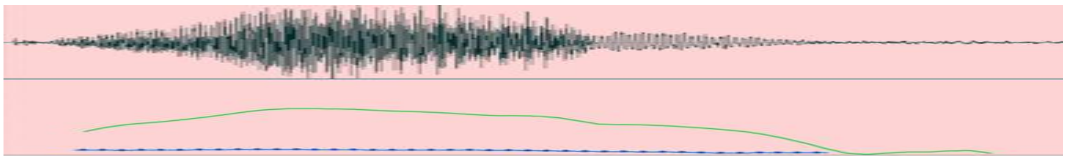


***Friday* [fraıde]**

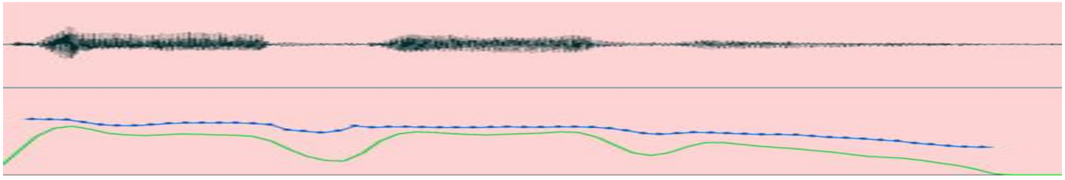
Tuesday was pronounced without the palatal /j/. The voiceless alveolar plosive /t/ is realized with vibration of the vocal cords which is evident from the pitch curve. Initial partial devoicing of /d/ is also seen.



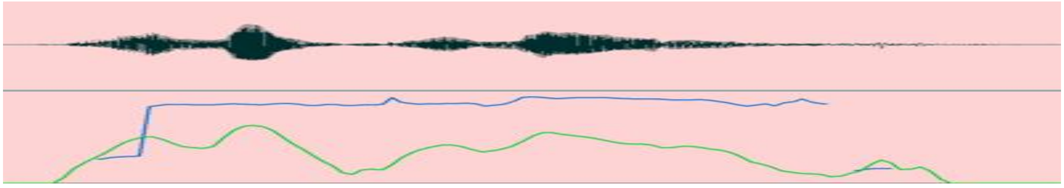
***Tuesday* [tu:zde:]**



***train* [tre:n]**



Saturday **[θa:təde:]**



aeroplane **[elo:ple:n]**

Note the continuous pitch curve without a break in the above speech waves.

Echolalia and Scripting

Echolalia has been observed to be present but only to a little extent. Scripting has not been observed.

Omissions of words

When this child was asked to count numbers, he successfully did it from 1-10. When prompted, he continued to count from 11 to 20. However, he ommitted numbers 15, 16 and 20. There were also missing letters in the rendition of the English alphabet [*ABCD..F…I..KLM..OPQR…W..YZ*].

There was also omission of words as the child recited the poem '*Johnny Johnny*'. The missing words are put in brackets.

Johnny Johnny, yes papa
Eating …..(sugar)……..*no papa*
Telling lie………………(no papa)
(open your mouth)…uttered something which sounded like (*Opera*)

Complete sentences/ sentence structure/ use of the pronoun.

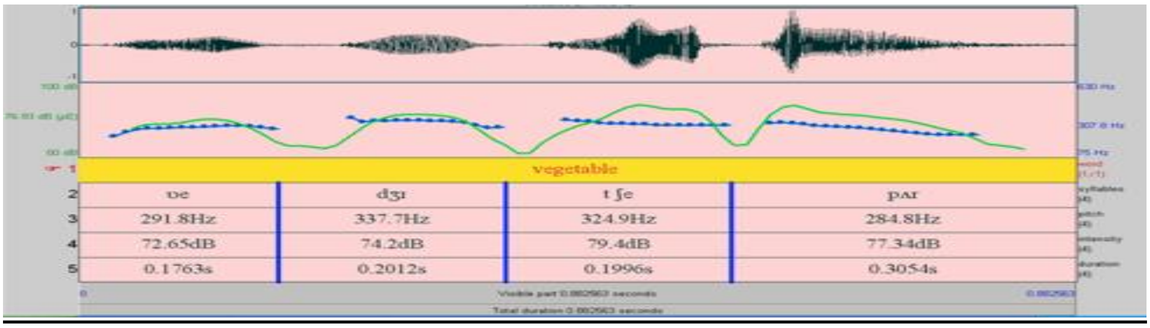
The child did not speak complete sentences to analyze the sentence structures. The child had no use of pronouns. Not even pronouns *you*/*your* because every time he needed to be asked a question, his name had to be used. For example, 'what does 'NAME OF CII' like?' or 'Which class is 'NAME OF CII' studying in?

Comprehension/Expression/Pragmatics

This child's comprehension was just enough to perform a single step instruction. When asked a question like '*what do you put in the fridge?*' some of the words listed by him were: '*bread, banana, watermelon, onion, bus*', some relevant and others completely irrelevant. Also, he also could not answer questions like '*Do you like your school?*' and '*Can you name some friends in your school?*' which CI could answer relevantly.

Fluency

The child was still not at the sentence level of communication. It was observed that even longer words were not spoken in a single stretch of articulation. A word like *vegetable* with four syllables was broken down to its syllables and uttered with distortion as [ʊe.dʒɪ.ʧe.pʌr]. Evidence of this is given below. The speech wave of *vegetable* given below shows the transcription of the syllables as uttered by the child.

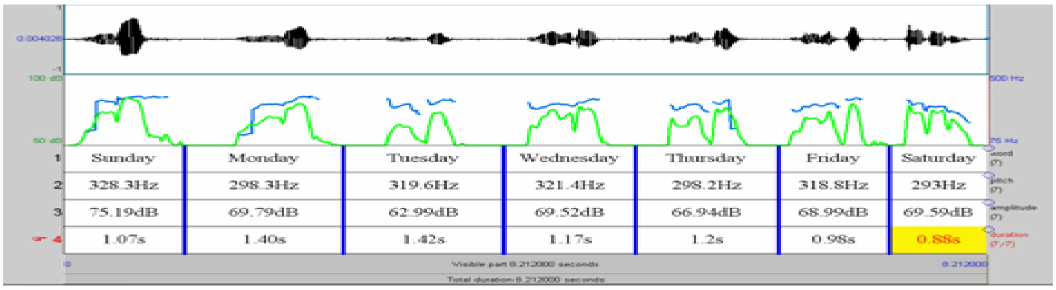


vegetable **[ʊe.dʒɪ.ʧe.pʌr]**

Tiers were drawn and the values of pitch, intensity and duration were noted. It was also observed that the stress (in the most prominent syllable which is the one with the highest value of pitch and intensity) was on the third syllable instead of the first.

Voice quality and intonation

The child's voice quality was not normal. It sounded hoarse and generally showed inconsistency in loudness. Evidence of *days of the week* is shown in the following speech wave:



Inconsistent amplitude in *days of the week*

Tiers were added to the speech wave and were numbered: 1, 2, 3 and 4. These tiers showed the **word** in the first tier and the values of **pitch, amplitude** and the **duration** in the 2nd, 3rd and 4th tiers, respectively. We notice here that *Tuesday* showed the least amplitude at 62.99dB. However, it must be noted that the child did not miss out any day.

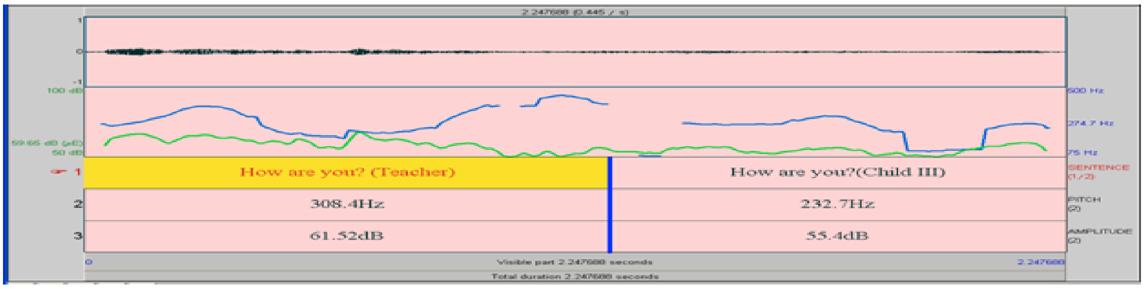
CHILD III

Child III - Diagnosed to be a case of severe autism
Age: 12 years, Gender: boy, Age of onset: 1.5 year, First symptom: seizure

This child who was diagnosed as a severe case of autism showed almost no language development. His teacher was asked to encourage him to speak because he was familiar with her. Regrettably, not much data could be gathered for analysis of his speech. The teacher had to ask him the question '*what is your name?*' repeatedly but there was no response. Then she tried to prompt him by starting the answer for him: '*My*……' '*My*…… ' '*My name is* …….'. Only after 4-5 repetitions did the child repeat what the teacher said.

The child repeated the question '*how are you?*' after the teacher asked him the question 3-4 times. Finally, the child merely repeated the answer uttered by the teacher, '*I am fine!*'. When asked to count numbers, the child did not respond. In the hope of getting him to count, the teacher started counting and repeated the number *one* about eight times. Yet, the child was unable to take off from there. Each number was repeated 3-4 times and the child would repeat the same number in a very low voice. When it came to number *four* the teacher listened to the child's low voice and said, '*loudly 4…*' and the child simply repeated '*loudly 4….*'

Therefore, we can conclude that this child has echolalia to a large extent. Even those utterances he repeated were of a very low amplitude. The following speech waves of both the teacher's and the child's voices for '*How are you?'* show that the average values of pitch and amplitude of the child's utterance is very low when compared to the teacher's voice.



Echolalia with very low amplitude

This child, a case of severe autism showed extremely limited speech and lacked adequate speech and language for communication.

Final Findings

The number of phonological substitutions/deletions and other distortions at the level of sound, sentence structure and expressions were considered as a rough and ready measure of severity of speech and language disorders. It is clear in Table 2 that speech and language disorder is most severe in Child III. Child III showed very little development in his speech and language which resulted in very little data to analyze. Child I showed the least disorder severity. In fact, her speech and language development was the best as compared to the other two children.

Table 2. A comparative picture

Speech/Language disorder	CHILD I	CHILD II	CHILD III
Substitution	/r/, /ʃ/, /ŋ/ and /z/ were substituted	/r/, /t/, /s/, /z/ and /k/ were substituted	n/a
Deletions/elisions sounds	/n/ was deleted in the word final position	/θ/, /t/, /k/, /z/, /n/ and /l/ were deleted in the medial and final positions	n/a
Omissions of words	Not observed	Yes, to a large extent	n/a
Distortions	One instance	A few instances	
Ability to articulate consonant clusters	Yes	No	
Echolalia	No	Small extent	Large extent
Scripting	One instance	Did not observe	n/a
Complete sentences	Yes (4-5 words)	No	No
Sentence structure	Correct	No sentences	No sentences

Ability to use the pronoun	Yes	No	No
Comprehension/Expression	Fairly present	Very limited	Absent
Fluency	Fair level	Not present	n/a
Phonation	One instance of absence of voicing contrast	Many instances of absence of voicing contrast	n/a
Voice quality	Fair	Fair	Bad
Intonation	Has the ability to distinguish between angry and polite tones	Does not have the ability to distinguish between angry and polite tones	n/a
Voice quality	Fair	Fair	Bad
Intonation	Has the ability to distinguish between angry and polite tones	Does not have the ability to distinguish between angry and polite tones	n/a

It is noted that most of the disorders at the sound level are *phonological* rather than *phonetic* (target sounds were replaced by sounds that are part of the phonemic inventory of the same language, i.e. English. All these children were exposed to the English language and had also been studying it in their schools since the age of 2;5.

Indeed it is not just the number of substitutions/deletions and other distortions that were the criteria for deciding on severity level but also the kinds of distortions/substitutions that each child employed. Some distortions resulted in perceptual confusion but some others in which the clinician was unable to decipher the sound were considered more severe. For example, if the child uses [ɬ] (voiceless alveolar lateral fricative) and [ɮ] (voiced alveolar lateral fricative) in place of /s/ and /z/ there is considerable difficulty in perceiving the sounds and they also sound markedly unnatural. However, in this study no atypical sounds were produced by the children.

Conclusion

This study demonstrated different levels of speech and language disorder in these three children. The findings of this study could help speech clinicians to train children with extensive and appropriate practice. Results could assist in establishing correct place of articulation, thereby helping the child perceive contrast. The clinician may or may not be able to recognize the distortions of certain sounds on the basis of mere auditory perception, as these children also exhibited a rapid rate of speech. A quick analysis of this kind would help clinicians take off therapy in the right path from the beginning without wasting much time. This study reiterates the fact already ascertained: that severity of autistic impairment relates to severity of speech and language impairment in children.

References

Ball, M. J. (1993). *Phonetics for Speech Pathology*. London, UK: Whurr Publishers Ltd.

Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer* [computer program]. http://www. praat.org/

Churchill, D. W. (1978). *Language of autistic children*. Washington D.C: V.H. Winston & Sons.

Eisenson, J. (1986). *Language and speech disorders in children*. New York, NY: Peramon Press.

Murdoch, B. E. (1990). *Acquired speech and language disorders*. London, UK: Chapman & Hall.

Murray, S. (2012). *Autism*. London, UK: Taylor& Francis.

Renfrew, C. E. (1972). *Speech disorders in children*. Oxford, UK: Peramon Press.

The impact of the critical period on reading and pronunciation in L2

Urszula Swoboda-Rydz[1], Marcin Chlebus[2]
urydz22@gmail.com, mchlebus@wne.uw.edu.pl

[1]Polish Language Department ‘Polonicum’, University of Warsaw
[2]Faculty of Economic Sciences, University of Warsaw

Abstract. The critical period (CP) hypothesis applies to first and second language acquisition (SLA). The decline of second language (L2) learning abilities does not occur suddenly, but late learners usually do not reach the uniformed success of first language (L1) acquisition. The strongest predictor of the final outcome of L2 remains the age of acquisition (AoA), which correlates negatively with morphosyntax and phonology, but not with semantics and learning new vocabulary. Reading in English (L2) is a necessary skill for many professionals, who often feel discouraged to keep up with L2, particularly if they started learning it at a later age. Therefore, the aim was to check differences in L2 speed of reading and pronunciation between early and late learners. *Method*. The study involved students (43), teachers (13) and doctors of medicine (38), whose L1 was Polish. In the phonological part, participants read aloud short sentences and 10 chosen phonemes were assessed. In the reading part, the relative speed of reading in English (RSRE) was measured. A survey included some personality features and linguistic behaviour of the participants such as listening, speaking and writing in English. In a multivariate analysis of the impact of determinants on the RSRE, panel data estimators, such as Random Effects (RE) and Generalized Least Square estimator (GLS) were considered. They take into account heteroscedasticity and cross-correlation of error term in the study group. *Results and discussion*. An RSRE equal to 1 means reading in English is as fast as in Polish. AoA had an impact on the expected RSRE, which increased by 0.279, if the start of L2 learning was between 7-14; but by 0.726, when AoA $\geq$ 15, which means that these people read relatively more slowly in L2 than early learners. Being a teacher, studying for extended exam in English and a language contact index combining speaking, listening and reading, decreased expected RSRE by 0.346, 0.318 and 0.666, respectively. Declared amount of time spent reading English and the number of hours of L2 education did not affect the expected RSRE. *Conclusion*. The pronunciation results are inconclusive and cannot be explained by AoA. AoA has a significant influence on the reading speed in L2, which approximates the speed of reading in L1, if a person started learning L2 before the age of 7 and has adopted a linguistic behaviour.

Keywords: critical period, panel data analysis, phonology, reading, second language acquisition

Introduction

A critical period originates from observations of behavioural patterns and physiological experiments with animals. It implies that a given ability may be acquired only if appropriate environmental stimuli are present within a particular time window during an ontogenic development.

The term “critical period” (CP), coined by Lenneberg (1967), suggests a sharp decline in plasticity after an end-point, following which learning is impossible or difficult (Johnson & Newport, 1989; Singleton & Ryan, 2004). The term “sensitive period” is used to imply a gradual decline in learning abilities (Newport, 2006; DeKeyser, 2000).

Neurobiological experiments have indicated that the CP for phonetic perception in a first language (L1) begins prior to 1 year and syntax develops between 1.5 and 3 years of age (Kuhl, 2010). After puberty, acquisition of the first language is strongly deficient (Newport, 2006).

When the critical period hypothesis applies to second language acquisition (SLA), there is no consensus whether there is a cut-off point or a gradual decline (Birdsong, 2005; Zhu, 2011). A longitudinal survey of immigrants to Australia failed to find a pattern of discontinuation, typical of CP (Chiswick, Lee, & Miller, 2004), so did a 1990 US Census of about 25,000 Chinese and 39,000

Spanish immigrants (Bialystok & Hakuta, 1999; Hakuta, Bialystok, & Wiley, 2001), and a 2000 US Census (Chiswick & Miller, 2008).

Although it is possible for adults to learn a foreign language, late learners are characterised by lack of a success typical of L1 acquisition and poor final attainment (Johnson & Newport, 1989). Only early learners in the age group 3-7 years were able to achieve native results compared to late learners >15 years of age.

The strongest predictor of the final outcome in L2 is believed to be age of acquisition (AoA), which refers to the age at which exposure to L2 starts. The term has been used mostly in the situation of immigrants, while age of first exposure (AoE) describes a situation when a learner starts L2 education at school, visits a foreign country for the first time or starts contact with L2 relatives. In the paper, only the term AoA will be used to refer to the time of first contact with a foreign language which is cultivated afterwards.

AoA correlates negatively with morphosyntax and phonology, but no such correlation has been found between age effects and semantics and learning new vocabulary (DeKeyser & Larson-Hall, 2005; Birdsong, 2006). Some authors suggested the notion of a "multiple critical period" (Long, 1990; Seliger, 1978; Knudsen, 2004) arguing that there are different CPs for phonology or morphosyntax and some linguistic structures in L2 (Lee & Schachter, 1997; Weber-Fox & Neville, 1996).

A decline in second language (L2) learning capacities points to maturation processes, which take place between birth and puberty (Lenneberg, 1967; Long, 1990; Penfield & Roberts, 1959; Pinker, 1995; Pulvermüller & Schumann, 1994; Scovel, 1988). Physioneurological evidence of cortical maturation evidence involves:

- lateralization
- completion of myelinisation which results in reduced plasticity and difficulty in learning,
- metabolic turnover, which is the highest rate in the first decade and reaches a low and stable level around puberty
- synaptogenesis, which peaks around 2-4 years and stabilizes between 10 and 15 years of age (Uylings, 2006)
- hormonal changes around puberty other physiological changes after 30 such as neuritic plaques, neurofibrillary tangles and other degenerative features,
- the reduction of dopamine D2 receptors, which starts at 20 and lasts through the lifespan and results in decreased execution of functions, verbal fluency and perceptual speed (Birdsong, 2006).

The final acceptance or refusal of the CP depends on its definition; however, the age effect remains a fact. Children learn differently from adults and are able to take advantage of implicit learning, while adults require explicit learning and depend more on declarative memory (Ullman, 2001; DeKeyser, 2005). There is also more variation among late learners than early ones (Fillmore, 1979). The decline of L2 learning abilities does not occur suddenly, but becomes a gradual process after 6-7 years of age (DeKeyser, 2000) and learners cannot achieve nativelike achievements after that age (DeKeyser, 2000). AoA also affects cerebral representation of L2 and language processing (Paradis, 2004).

The point of discussion is which age or segment of AoA constitutes a maturational milestone which impacts SLA. The age of three limits phonological abilities (Flege, 1981) and shows different cortical involvement during sentence processing (Mueller, 2006). AoA impacts mostly L2 phonology (Flege, Yeni-Komshian, & Liu, 1999). Phonological acquisition above the age of five or six results in foreign accent (DeKeyser, 2000). The age of six years is a time of formation of a dense neuronal network, which may decrease or increase in the lifespan due to intensive learning (Uylings, 2006).

Many authors suggested that the critical period should be up to 7 years of age. When L2 is acquired after that age, it does not overlap with dominant L1 areas, it is less lateralized and the degree of proficiency decreases, reaching adult levels by the end of adolescence (Pinker, 1995; Birdsong, 2005). However, it was found in adopted children after the age of 7 and 8 that Korean (L1) was replaced by French (L2), and their brains did not show any stimulation when exposed to the first language (Pallier, Dehaena, Poline, LeBihan, Argenti, Dupoux, & Mehler, 2003).

Some researchers point to the onset of puberty around 11-12 years of age, while others indicate that of 15 or 16+ considering that neural maturation continues and mielinisation finishes in the third decade in humans (Weber-Fox & Neville, 1996; Singleton & Ryan, 2004). Semantic processing has not been found affected by AoA (Mueller, 2006).

Ultimate linguistic attainment seems to be determined by intensive use and neurological responses to linguistic input. In a few studies, late learners were able to present nativelike performance in a new language (Ioup, Boustagui, El Tigi, & Moselle, 1994). Brain processing in L2 depends on the effect of age and proficiency. In comprehension tasks, L2 proficiency shows a stronger impact than AoA on cortical involvement of L1 and L2 (Dehaena, Dupoux, Mehler, Cohen, Paulescu, Perani, van de Moortele, Léhericy, & LeBihan,1997; Perani, Paulescu, Sebastian, Dupoux, Dehaena, Bettinardi, Cappa, Fazio, & Mehler, 1998; Chee, Caplan, Soon, Sriram, Tan, Thiel, & Weeks, 1999). These observations are congruent with the “convergence hypothesis” (Green, 2003) which postulates that L2 processing is similar to L1 when L2 proficiency increases.

Aim

In Poland, English as a second language is taught at school, but it is sporadically continued after graduation from university. Reading in English is a necessary skill for many professionals and it is an independent activity from speaking. Adults often feel discouraged to keep up with foreign language education, particularly if they started learning L2 at a later age. Given the fact that semantic processing is less age dependent (Dekeyser, 2005; Mueller, 2006), it seems worthwhile to check if there are any differences in the speed of silent reading/comprehension in L2 and in pronunciation between late and early learners. Additionally, the influence of other factors on the relative speed of reading in English was analysed.

Method

The study involved three groups of participants: first year medical students at Warsaw Medical University, doctors of medicine of different specialties from Warsaw, and university teachers of English at Warsaw Medical University, whose mother tongue is Polish (see Table1). All participants were informed of the purpose of the research and gave an oral consent to it.

The examination included a pronunciation part, a reading part, and completion of a survey. It was conducted individually in a quiet room. The research lasted from March until July 2015.

Table 1. Distributions of the participants among analysed groups

Group of participants	No. of participants	Percentage %
1st year medical students	43	45.74
Medical doctors	38	40.43
Academic teachers	13	13.83

Group characteristics

The students attended three university groups randomly chosen out of sixteen. At Warsaw Medical University, students are appointed to groups during recruitment process, mainly based on alphabetical order. The exclusive criteria were: L1 different from Polish and the age >22 years, which could mean a break in English education longer than 1 year. Doctors were invited to participate in the study if they had achieved a high grade in an English exam, part of their specialisation exam, which was assessed by the researcher. All English teachers from Warsaw Medical University were included, except for four of them who refused to participate.

Students (aged 19-22, mean 19.8) started to learn English (L2) mainly at 6-7; doctors (aged 28-59, mean 43.4) and teachers (aged 34-65, mean 53.2) started learning the L2 at different ages.

The three groups shared tertiary level of education, medical interests and good socio-economic status, with the exception of students who were at the beginning of their university career.

Phonological part

Students and doctors were asked to read aloud ten sentences containing ten phonemes that are absent in the Polish language. The production of each phoneme was assessed as correct or incorrect by the researcher after reading each sentence. This task was not performed for university teachers of English, as they were assumed to pronounce the sounds correctly.

Reading part

In the reading part, the relative speed of reading in English (RSRE) was measured. The participants silently read sentences that appeared one by one on a computer screen and decided whether the sentence in Polish and English made sense. The last word in each sentence was underlined and was purposefully changed by the researcher every eight sentences (anomalous). The participants were instructed not to try to look for the right word if they felt the sentence made no sense, not to judge the content of the sentence, or personalise it. They were also instructed to read as fast as they reasonably could in order to judge if the sentence made sense or not.

A dedicated computer programme measured absolute and relative reading times. For this purpose, the sentences were “combined” into 33 pairs of a length ± 3%. The RSRE was a fraction of the time of reading an English sentence to the time of reading a Polish sentence. An RSRE equal to 1 means that the speed of reading is as fast in English as in Polish. Higher values of the RSRE mean that reading time in English is longer and its speed is lower. Calculating the relation between reading time in English and Polish was important, as it allowed minimisation of individual differences due to eye problems, reading habits, reflectivity, impulsiveness, and age.

At the beginning, all participants underwent a mock test consisting of three Polish and English sentences, so that they could get acquainted with the computer program and practise how to signal their decision. The right key (→) was used when the participant thought that the sentence made sense, while the left one (←) was used when the participant believed that the sentence did not make sense. The sentences were written in a Times New Roman font size 12. The washout time between sentences was five seconds that was counted on the screen.

Subsequently, the proper test started with thirty-three sentences in the Polish language appearing one by one on the screen. When the Polish part was completed, a short survey with the researcher was conducted in English. Then the remaining thirty-three English sentences appeared on the screen and decisions were taken respectively. Finally, each participant could view the obtained result.

Examples of paired sentences from the research are given below. Pair 12 was anomalous, which means that the last word was replaced and the original word is given in brackets. In pair 13, the Polish sentence needed more time to be assessed quickly as it was a difficult metaphor, while in pair 15, the English sentence was difficult due to the word *diligence*, which was rarely used by doctors and students (see Figure 3).

12. Alkohol nie daje odpowiedzi, ale pozwala zapomnieć, jakie było **badanie**. (pytanie)
12. Never let your work drive you. Master it and keep it in complete **darkness.** (control)
13. Jeśli nie możesz zmienić swoich przyjaciół... Zmień przyjaciół.
13. People rarely succeed unless they have fun in what they are doing.
15. Naturze człowieka leży rozsądne myślenie i nielogiczne działanie.
15. What we hope to do with ease, we must learn first to do with diligence.

Metaphors

The research aimed to measure time of reading. For this purpose, quotations in the two languages, mostly containing metaphors, were chosen as minimal meaningful texts. According to Lackoff: "metaphor is understanding and experiencing one kind of thing in terms of another", which was supposed to be a challenge for readers. The metaphors did not contain proper names or cultural references in order to avoid intercultural misunderstandings.

Survey

The survey was conducted in English and was used to help set the mind into English mode (Grosjean, 2008). The participants were inquired, among others, about (see also Table 2):

- when they started learning English
- how many years they had learned English and what the frequency of English lessons was
- how much time they spend on average reading in Polish and in English in a week
- whether they like "small risk" in life
- students only were asked what kind of school leaving exam (*matura*) in English they took - basic or extended
- doctors only were asked whether their participate actively in English-speaking conferences and meetings, or whether they have spent at least 6 months in an English-speaking country in the last 2 years, whether they talk regularly in English with a person in their close family or at work at least once a week, and whether their listen to spoken English for more than one hour per week

The total amount of English classes was calculated, assuming a fixed number of weeks at school, school breaks, and weeks when classes are usually missed, such as the last week of a school year. The final number constituted a model number of English classes because it did not consider differences between schools, or sick leaves.

Data

For the 94 participants in the three groups, the RSRE of 33 pairs of sentences and other characteristics specific to particular pairs were calculated. As a result, 3102 observations were taken into consideration while modelling. The list of collected characteristics is presented in Table 2. The structure of the data is panel-like, where participants are units and successive pairs of sentences are treated as measurements of the phenomenon. The panel is balanced, since for each unit results for all 33 pairs of sentences are available.

The information obtained made it possible to carry out a multivariate analysis of the impact of individual determinants on RSRE.

Table 2. Distributions of the participants in analysed groups

Characteristics	Description	Usage
Characteristics collected during the experiment		
NO	Sentences serial number	ALL
ID	Participant ID	ALL
GROUP	Group ID (student, medical doctor, teacher)	ALL
CO_SENT_PL	Content of sentences in the Polish language	ALL
TIME_R_PL	Time required to make a decision in case of sentences in Polish	ALL
CH_PL	Participant's choice whether the sentence in Polish made sense	ALL
INFOSENSE_PL	Information whether the sentence in Polish made sense	ALL
MISTAKE_PL	Information whether the participant made a mistake while assessing the	ALL

	sense of the Polish sentence	
CO_SENT_ENG	Content of sentences in the English language	ALL
TIME_R_ENG	Time required to make a decision in the case of sentences in English	ALL
CH_ENG	Participant's choice whether the sentence in English made sense	ALL
INFOSENSE_ENG	Information whether the sentence in English made sense	ALL
MISTAKE_ENG	Information whether the participant made a mistake while assessing the sense of an English sentence	ALL
RSRE	The relative speed of reading in English (RSRE)	ALL
Study variables (concerning all participants)		
SEX (M=1 F=2)	Participant sex	ALL
AGE	Participant age	ALL
START ENG	Year of age when the participant started learning English	ALL
LEFT_HAND	Information whether the participant is left-handed or right-handed	ALL
Study variables (concerning students)		
ENGEX (B=1, E=2)	Information whether the participant took an extended or basic secondary school-leaving exam (*matura*) in English	STUDENTS
CITY/TOWN	Size of the city/town from which the participant comes	STUDENTS
HR_E	Estimated number of hours in the English language in an elementary school	STUDENTS
HR_G	Estimated number of hours in the English language in a junior high school (*gimnazjum*)	STUDENTS
HR_S	Estimated number of hours in the English language in a secondary school	STUDENTS
HR_TOT	Estimated total number of hours in the English language in school	STUDENTS
READ_WK	Number of hours which the student spends reading in English per week	STUDENTS
LISTEN	Number of hours which the student spends listening to English speech	STUDENTS
ARTICLES	Number of articles which the student reads in English per week	STUDENTS
BOOKS	Number of books which the student reads in English per week	STUDENTS
CERTIFICATE	Information whether the participant has a certificate in English	STUDENTS
Study variables (concerning doctors)		
LISTEN	Information whether the doctor listens to English	DOCTORS
CONFER	Information whether the doctor actively participates in conferences in English	DOCTORS
CONTACT	Information whether the doctor speaks English regularly	DOCTORS
SUM_ENG	Combined index informing about the degree of English use in everyday life	DOCTORS
TIME_READ	Information on how much time per week the doctor reads in English	DOCTORS
RISK	Information whether the doctor likes little risk	DOCTORS

Econometric framework

During the econometric analysis, a joint analysis was conducted for all participants and two additional analyses exclusively for students and for medical doctors. Performing additional analyses stems from the fact that additional information gathered from students and doctors cannot be included in the general model due to dissimilarities between groups.

Because data are in a panel form, panel data estimators were taken into consideration. The general formula of the estimated model is described as follows:

$$y_{it} = \alpha + x_{it}\beta + u_i + \varepsilon_{it} \qquad (1)$$

where y_{it} is the RSRE for a particular participant and a particular pair of sentences t, x_{it} is the matrix of characteristics that explain the level of RSRE; α and β are the coefficients to be estimated; u_i is the individual effect and ε_{it} is the error term.

In order to estimate coefficients in the Equation 1, different methods may be used. In this study, the Breusch and Pagan LM test for random effects, the F test for individual effects, the fixed-effects estimator and the Hausman specification test were used to decide on the best estimator among the Fixed-effects, Random-effects and POLS estimators (for details see Baltagi, 2013). Then, as an extension, autocorrelation within panels (by the serial correlation in the idiosyncratic errors of a linear panel-data test described in Wooldridge, 2010), heteroscedasticity across panels (by the LR test) and cross-sectional correlation (as in the Frees, Friedman, and Pesaran tests described in Sarafidis & De Hoyos, 2006) were examined. Based on the properties of the error term, an appropriate Generalized Least Square estimator was used (Greene, 2012).

The econometric framework described above was used in all three analyses.

Results

Phonetic production

The phonetic test was conducted among students who started learning English earlier, and doctors who were late learners, except for four of them.

Fifty percent of phonemes were significantly better pronounced by students (see Table 3). They included /iː/, /ð/, /ɜr/, dark /l/ and /ə/. Only /ɪ/ was pronounced better by doctors though the difference was not statistically significant.

Table 3. Percentage of participants with correct pronunciation of 10 chosen phonemes

Percentage of participants with correct pronunciation of the sounds (IPA)			
Phoneme	**% of students**	**% of doctors**	**p**
/iː/	90	71	0.0007
/ n+iː/	76	71	0.423
/ɪ/	33	39	0.377
alveolar /t/	43	34	0.191
/ð/	98	79	0.00003
/ŋ/	98	95	0.248
/æ/	86	76	0.071
/ɜr/	93	84	0.046
/ə/	50	26	0.0005
dark/ l/	86	63	0.0002

The best pronunciation in both groups involved /ð/, /ŋ/ and / ɜr/. These sounds are practised at school and they do not pose a particular problem to Polish speakers if given enough attention. The words in which the phonemes occurred, *weather*, *building*, *bird*, are frequent and learnt well. The worst pronunciation in both groups involved /ɪ/, /ə/, and /t/. Differentiating between /ɪ/ and /iː/ and the correct production of the schwa, /ə/, was problematic. Even though articulation of these sounds is not particularly difficult, practice is often neglected by learners, whose oral communication is impeded by lack of interest, and by teachers who become discouraged by the lack of learners' enthusiasm in mastering their pronunciation. The sound /n+iː/ should be pronounced as palatal /ŋ/ + /iː/, and not alveolopalatal /ń/ + /iː/, which is a typical feature in the Polish language. /t/ as an alveolar must be noticed by learners or explicitly pointed out by teachers to achieve correct pronunciation. However, even if alveolar /t/ is mispronounced as dental, it does not lead to misunderstandings.

Preliminary statistical results

Following presentation of phonological results, results on reading are discussed. Before results of the multivariate analysis of the impact of individual determinants on RSRE are considered, preliminary univariate analysis results are analysed.

Initially, it is worth considering the average value of RSRE depending on the AoA of participants. In the upper chart of Figure 1 that presents the average values of the RSRE for all participants, an upward tendency of the ratio might be noticed, which may be interpreted as follows: the later one started learning English, the stronger the tendency of the relative speed of reading in English. This tendency can be seen among students and doctors.

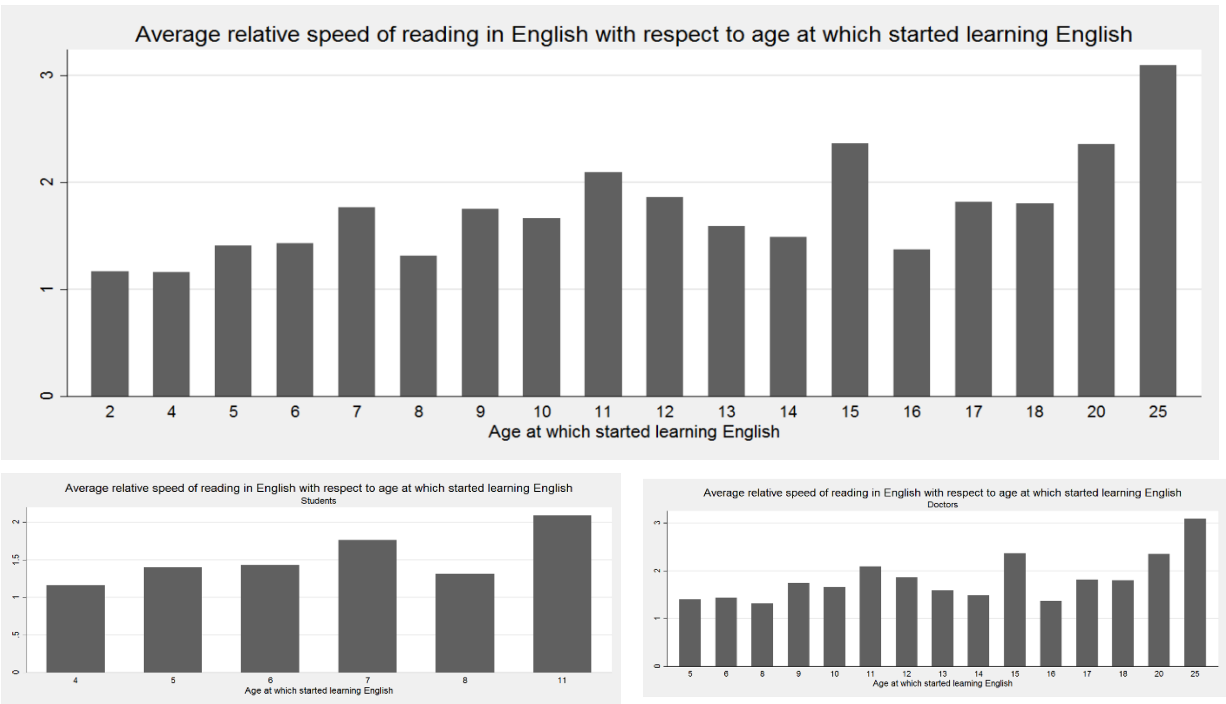


Figure 1. The average values of the RSRE for all participants, students and medical doctors with respect to age at which they started learning English

Another issue to be considered is information in histograms of the RSRE values (see Figure 2). The histogram for all participants shows that the RSRE distribution is mostly concentrated between 1 and 2. The average value of the RSRE is equal to 1.58 with a standard deviation of 0.98. Among the students, the distribution of the RSRE is similar in all groups, although a slight shift towards higher values may be noticed. For doctors, an interesting fact is that the RSRE <1 was not observed, which was true for both students and teachers.

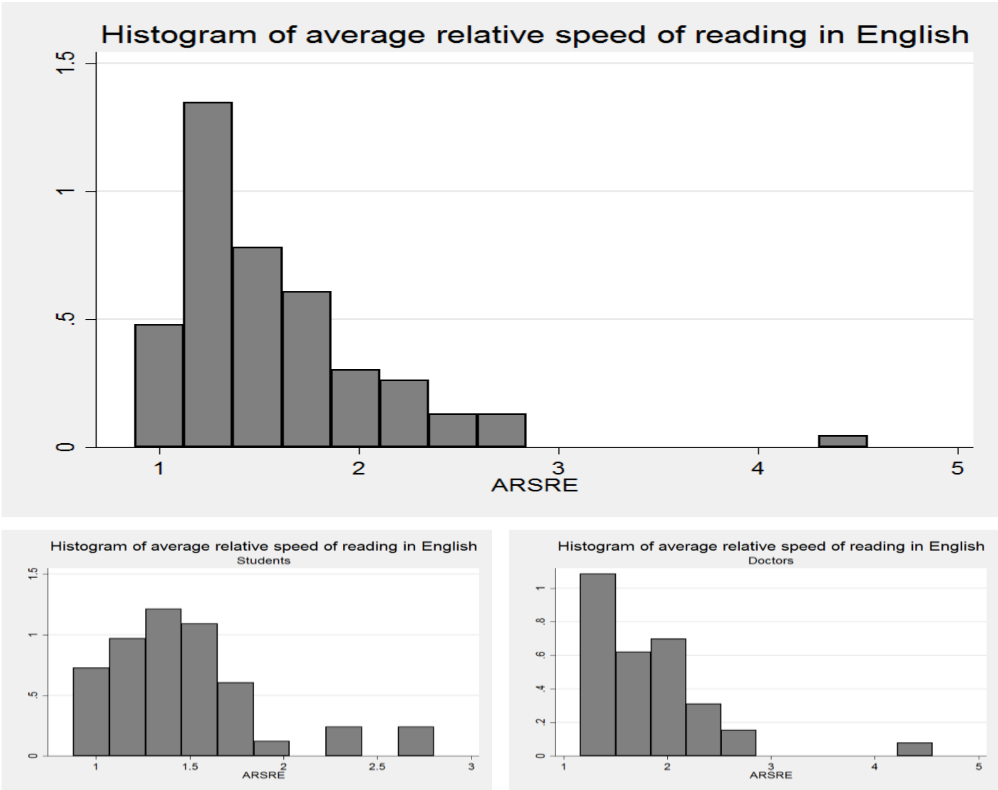


Figure 2. Histograms of the RSRE for all participants, students and medical doctors

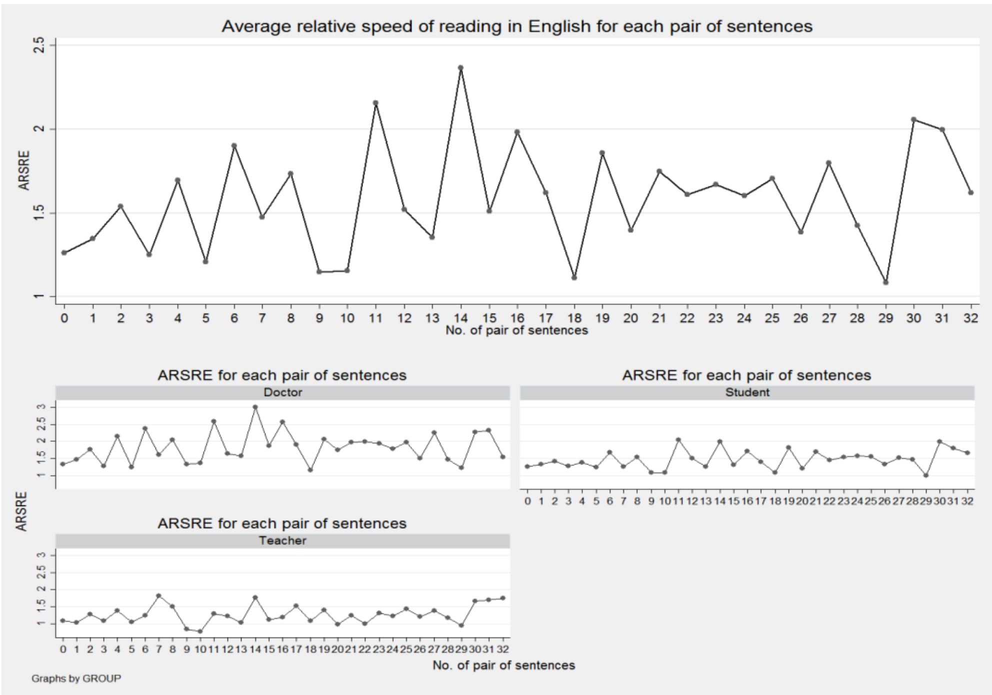


Figure 3. The average values of the RSRE for all participants, students and medical doctors with respect to pairs of the sentences

The last element worth considering in the preliminary study is average values of the RSRE for individual sentences (see Figure 3). It is noticed that there is no strong upward tendency in the average values of the RSRE based on the length of sentences. It is also noticed that the highest volatility in the average values of the RSRE is among doctors, which could mean that doctors have the biggest problems in reading when a sentence in English is relatively difficult. It has to be noted, however, that these conclusions have been drawn from the univariate analysis; the multivariate analysis shows that the cause does not lie in the group of doctors' as such, but that there are other factors correlated with the fact of being a doctor.

The model based on all participants

Following discussion of the results of univariate analysis, results of the multivariate analysis are presented. In studying determinants of the RSRE, based on the available data for all participants, a set of nine variables was prepared (see Table 4) that potentially could affect the value of the indicator.

According to the results, the Random Effects (RE) estimator proved to be the best among basic panel data estimators. Taking into account the fact that autocorrelation in error terms was not present but heteroscedasticity and cross-correlation were identified, an appropriate GLS estimator was used that takes into account heteroscedasticity and cross-correlation of error terms. In this case, the GLS estimator was used for these characteristics which turned out to be relevant in the RE model. Additionally, the extended model was analysed (GLS Ext) for the GLS estimator, in which all characteristics were taken into account and the process of insignificant characteristics elimination was performed independently from the RE model. The results for the RE and GLS models are shown in Table 4.

Table 4. Results for the model based on the all participants

	GLS	RE	GLS Ext
Teacher	-0.346***	-0.349*	-0.373***
	(0.042)	(0.140)	(0.044)
Has sense in ENG	-0.114***	-0.082*	-0.085**
	(0.028)	(0.035)	(0.033)
Mistake in PL	-0.130***	-0.103*	-0.088**
	(0.025)	(0.049)	(0.033)
Mistake in ENG	0.346***	0.334***	0.351***
	(0.033)	(0.051)	(0.036)
Learnig start 7-14	0.279***	0.311**	0.297***
	(0.036)	(0.101)	(0.037)
Learnig start 15+	0.726***	0.798***	0.739***
	(0.048)	(0.169)	(0.073)
Sentence Length	0.201***	0.216***	0.230***
	(0.038)	(0.032)	(0.038)
Gender			0.125***
			(0.026)
Left			0.152***
			(0.034)
Constant	1.369***	1.314***	1.225***
	(0.045)	(0.080)	(0.046)
N	3102	3102	3102

* p<0.05, ** p<0.01, *** p<0.001

The occurrence of the cross-correlation in the error term means that some of the RSRE values unexplained by the model (sentences, in particular) tend to be similar for most participants. It should be related to the existence of easier and tougher sentences in English irrespective of sentence length, for which participants needed either less or more time to comprehend compared to adequate Polish sentences.

Results obtained in the model based on the available data for all participants may be sum up as follows. Firstly, according to the results obtained, AoA between the ages 7-14 increased the expected RSRE by 0.279; but when AoA ≥15, the expected RSRE increased by 0.726, which means that people with AoA after the age of 7 years read relatively more slowly in L2 than people with AoA before the age of 7 years; so, reading speed decreases when AoA is late. These findings may be evidence that AoA has a significant influence on the RSRE.

Secondly, there is no difference in the expected RSRE between students and doctors. The only difference is found in teachers: the expected RSRE decreased by 0.346. The difference in teachers is expected as L2 teachers should be able to read relatively faster in L2 than other groups. Results obtained for students and doctors are interesting, because they show that these two groups have similar abilities in reading in L2, even though at first glance differences between the two groups are expected. This finding becomes even sounder supported by the irrelevance of variables that may distinguish the groups of students and doctors (such as age).

Additionally, for sentences numbered 11 and onwards, the expected RSRE increased by 0.201, which means that sentence length has an influence on the expected RSRE, but the relation is of a rather jumpy character than linear. Furthermore, it may be seen that both a mistake in an English sentence and lack of mistake in a Polish sentence (increased the expected RSRE by 0.346) and a mistake in Polish and lack of mistake in English (decreased the expected RSRE by 0.130) have an influence on the expected RSRE. Also, the participant decision that sentence in L2 “makes sense” has an influence on the expected RSRE (decreased by 0.114).

The model based on students

For students, except for variables included in the model for all groups, additional variables were added (see Table 7), and analysed separately.

Choice of an adequate model was analogous to that for a model for all groups of participants. Also, in this case, an attempt to extend the model for the GLS estimator was made, but the results obtained were counterintuitive (due to a model overfitting issue) and because of that they are omitted in Table 5.

In the student group, similar results were obtained as in the model based on all participants:

1. AoA at ages 7-14 increased the expected RSRE (by 0.617) compared to AoA before 7 years of age. There were no students who started to learn English after the age of 15.
2. A mistake in an English sentence and lack of a mistake in a Polish sentence increased the expected RSRE; a mistake in a Polish sentence and lack of a mistake in an English sentence decreased the expected RSRE.
3. Also, sentence length increased the expected RSRE (by 0.084), however, here the optimal threshold is sentence number 15.

Even though main findings stay valid, some minor differences are found among the variables also used in the model for all participants. Among students, only decision on whether the sentence “makes sense” or “does not make sense” had no significance, but only if the sentence really made sense or not (anomalous). Additionally, in this model, being a female increased the expected RSRE, which means that females read relatively more slowly in L2 than males.

The only relevant additional information gathered among students with impact on the RSRE is choosing an extended secondary school-leaving exam (*matura*) in English, which lowered the expected RSRE by 0.318. None of the remaining variables were found to be significant - neither the number of hours of education in English, nor the amount of time spent reading in English, nor the size of home town.

Table 5. Results for the model based on data for students

	GLS	RE
Sentences have sense	-0.068**	-0.109*
	(0.022)	(0.045)
Mistake in PL	-0.052***	-0.089
	(0.016)	(0.059)
Mistake in ENG	0.299***	0.247***
	(0.015)	(0.061)
Gender	0.312***	0.303**
	(0.016)	(0.093)
Learnig start 7-14	0.617***	0.630***
	(0.035)	(0.144)
Learnig start 15+	0.000	0.000
	(.)	(.)
Ext. maturity exam	-0.318***	-0.328***
	(0.012)	(0.092)
Sentence Length	0.084***	0.065
	(0.018)	(0.038)
Constant	1.383***	1.454***
	(0.032)	(0.101)
N	1419	1419

* p<0.05, ** p<0.01, *** p<0.001

Table 6. Results for the model based on data for doctors

	GLS	RE	GLS extended
Mistake in ENG	0.452***	0.505***	0.400***
	(0.017)	(0.092)	(0.019)
Learnig start 7-14	0.280***	0.281	0.400***
	(0.022)	(0.434)	(0.035)
Learnig start 15+	0.708***	0.720	0.712***
	(0.030)	(0.456)	(0.050)
Sentence Length	0.302***	0.286***	0.236***
	(0.042)	(0.061)	(0.027)
Risk	-0.392***	-0.401	-0.340***
	(0.022)	(0.239)	(0.033)
Has sense in ENG			-0.083***
			(0.016)
English score = 1			-0.352***
			(0.019)
English score = 2			-0.549***
			(0.029)
English score = 3			-0.666***
			(0.067)
Constant	1.499***	1.526***	1.752***
	(0.049)	(0.430)	(0.065)
N	1254	1254	1254

* p<0.05, ** p<0.01, *** p<0.001

The model based on doctors

For doctors only, except for the variables included in the model based on all groups, additional variables were added (see Table 2), and analysed separately.

Choice of an adequate model was analogous to that of the model for all groups of participants.

In the doctor group**,** similar results were obtained again as in the model for all participants and students. Again, AoA at ages 7-14 increased the expected RSRE (by 0.280), and AoA >15 years of age increased the expected RSRE even more (by 0.708). Sentence length increased the expected RSRE by 0.302 with the sentence threshold on number 11 and further. Additionally, in the GLS extended model like in all groups, the participant's choice that the sentence "made sense" in English decreased the expected RSRE.

Among doctors (in comparison to all participants and students), some minor changes are also noted, specifically a mistake in a Polish sentence and lack of a mistake in an English sentence do not influence the expected RSRE.

Additionally, two factors among doctors had an influence on the RSRE. Firstly, being a risk lover decreased the expected RSRE. It may be due to the fact that risk lovers take decisions faster because they accept the risk of making a mistake. Secondly in the GLS extended model, combined language contact index had an influence on the expected RSRE: the higher the value of the index, the lower the expected RSRE.

Discussion

In the phonological part, the phonemes are divided into "easy" i.e. manageable to demonstrate and produce, and more "difficult" i.e. demonstrating and requiring more attention. Thus, phonemes /ŋ/, /ð/, /ɜr/, /i:/, /æ/, and dark /l/ (the last one appears in Polish) were produced well in both groups, with better results among students in percentage: 98/95, 98/79, 93/84, 90/71, 86/76, 86/63, respectively. Difficult phonemes, such as /n+i:/, /ə/, /ɪ/, and alveolar /t/showed poor results in percentage 76/71, 50/26, 33/39, 43/34, respectively. Successes and failures in pronunciation can be explained by other factors than age effect. The better phonological outcome among students may result from more opportunity to speak at school, while doctors were only sporadically exposed to spoken English as for them the main linguistic source was reading. Doctors seemed to pay less attention to pronunciation during communication. Anyway, AoA was in all, except one participant, >3 years, which is relatively late for phonological development.

The main conclusion from the obtained results in models analysing the RSRE is that AoA had an impact on the RSRE: the later the beginning of acquisition the slower the RSRE. This conclusion is confirmed in all groups, all participants, students and doctors.

Among biological factors, being left-handed and being a woman prolonged the RSRE. There were four left-handed participants and their number may be insufficient for drawing conclusions. The fact of men being faster is difficult to explain at this stage.

The fact that the RSRE was sensitive to mistakes in English (increased) and in Polish (decreased) is a result of the fact that a mistake prolonged thinking time usually and changed the value of the fraction of RSRE respectively.

Some peaks in Figure 3 of anomalous sentences, numbers 5, 11, 14, 19, 27, and 30 show a need for more time to read and process. Troughs may result from a pair of sentences with an easy metaphor in English and difficult in Polish. The most characteristic is little volatility among teachers and less volatility among student compared to doctors.

The threshold sentence among doctors was 11, after which the RSRE was longer, while among students it was 15, which may be a combination of at least three factors: earlier AoA, being younger and having had more practice in reading texts.

The amount of hours of formal education did not improve the RSRE; reading may not be done during lessons or teachers may not make students read at home. Declared reading time did not influence the RSRE, probably reflecting participants' wishful thinking.

In terms of the RSRE, a value of 1 means that the speed of reading in L2 is like in L1 and it would be found in well-balanced bilinguals. In our study on the basis of the model for all participants, it can be noticed (Table 4) that only teachers, who started learning English before the age of 7 reached the expected RSRE (ceteris paribus) close to 1. It suggests than a division between early and late learners lies before and after the age of 7. Based on the same model, both students and doctors achieved an RSRE oscillating around 1.3, if they started learning English (L2) before the age of 7 (ceteris paribus). Having analysed the students' (Table 5) and doctors' results (Table 7), it can be noted that it was possible for students to achieve the expected RSRE close to 1 but only if they started learning English before the age of 7, and chose to pass the extended school-leaving exam in English, *matura*. Also, doctors who scored 3 in the combined language contact index had the expected RSRE close to 1 (1.086, ceteris paribus), only in the GLS extended model. Being a risk lover, decreased the RSRE among doctors, but this predisposition to take risk does not indicate linguistic abilities, only a faster deciding process and acceptance of making a mistake.

What improved the RSRE: a) among students was the choice of *matura*, which necessitated practising all linguistic skills, and b) among doctors was the contact index combining speaking, listening and reading. Looking for additional contact with L2 may be called a linguistic lifestyle, which means that subjects continued the use of L2 despite the termination of education or lack of necessity to do so. This linguistic lifestyle seems to be important in monolingual countries like Poland.

Finally, it is not possible to draw conclusions concerning AoA and L2 pronunciation because the results can be explained by education-derived factors and not only by age affect. The results show that in order to improve the speed of reading in L2 to a value similar to the speed of reading in L1, it is necessary not only to use English intensely but also to start learning English before the age of 7. Both AoA and adopted linguistic behaviour have an impact on the RSRE. The reading speed in L2 approximated the speed of reading in L1 only if L2 education started before the age of 7 and the participant chose a particular linguistic behaviour, being a teacher, studying for an extended exam in English as a student, or using L2 on a regular basis as a doctor.

References

Baltagi, B. H. (2013). *Econometric Analysis of Panel Data* (5th ed.). Chichester, UK: Wiley.

Bialystok , E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (ed.), *Second language acquisition and the critical period hypothesis* (pp. 161-181). Mahway, NJ: Erlbaum.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. M. B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp.109-127). New York: Oxford University Press.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. In M. Gullberg & P. Indefrey (eds.), *The cognitive neuroscience of second language acquisition* (pp. 9-58), Malden, MA: Blackwell.

Chee, M., Caplan, D., Soon, C., Sriram, N., Tan, E., Thiel, T., & Weeks, B. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron*, *23*, 127-137.

Chiswick B. R., & Miller, P. W. (2008). A test of the critical period hypothesis for language learning. *Journal of Multilingual and Multicultural Development*, *29*(1), 16-29.

Chiswick, B. R., Lee, Y. L., & Miller, P.W. (2004). Immigrants' language skills: The Australian experience in a longitudinal study. *International Migration Review, 38*(2), 611-654.

Dehaena, S., Dupoux, E., Mehler, J., Cohen, L., Paulescu, E., Perani, D., van de Moortele, P. F., Léhericy, S., & LeBihan, D. (1997). Anatomical variability in the cortical representation of first and second languages. Neuroreport, *8*, 3809-3815.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499-533.

DeKeyser, R. M, & Larson-Hall, J. (2005). What does the critical period really mean? In J. F.Kroll, A. M. B. Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp.88-108). New York, NY: Oxford University Press.

Fillmore, L. W.(1979). Individual differences in second language acquisition. In C. Fillmore, D. Kempler, & W. Wang (eds.), *Individual differences in language ability and language behaviour* (pp. 203-228). New York: Academic Press.

Flege, J. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL* Quarterly, *15*, 443-455.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory and Language*, *41*, 78-104.

Green, D. W. (2003). The neural basis of the lexicon and the grammar in L2 acquisition: The convergence hypothesis. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (eds.), *The interface between syntax and the lexicon in second language acquisition* (pp. 197-218). Amsterdam: Benjamins.

Greene, W. H. (2012). *Econometric analysis* (7^{th} ed.). Upper Saddle River, NJ: Prentice Hall.

Grosjean, F. (2008). *Studying bilinguals.* Oxford, UK: Oxford University Press.

Hakuta, K., Bialystok, E., & Wiley, E. (2001). Critical evidence: A test of the critical period hypothesis for second-language acquisition. *Psychological Science*, *14*, 31-38.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition, 16*, 73-98.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of the maturational state on the age acquisition of English as a second language. *Cognitive Psychology*, *21*, 60-99.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behaviour. *Journal of Cognitive Neuroscience*,*16*(8), 1412-1425.

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, *67*(5), 713-27.

Lee, D., & Schachter, J. (1997). Sensitive period effects in binding theory. *Language Acquisition*, *6*(4), 333-362.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York, NY: Wiley.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*, 251-285.

Mueller, J. L. (2006). L2 in a nutshell: The investigation of second language processing in the miniature language model. In M. Gullberg & P. Indefrey (eds.), *The cognitive neuroscience of second language acquisition* (pp. 235-270). Malden, MA: Blackwell.

Newport, E. L. (2006). Language development, critical period. In *Encyclopedia of cognitive science* (pp. 737-740. Retrieved on October 11, 2015 Online http://www.bcs.rochester.edu/people/newport/newport-ecs-a0506.pdf.

Pallier, C., Dehaena, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex*, *13*, 155-161.

Paradis, M. (2004). *A neurolinguistics theory of bilingualism.* Amsterdam/Philadelphia: Benjamins.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms.* Princetown, NJ: Princetown University Press.

Perani, D., Paulescu, E., Sebastian, N., Dupoux, E., Dehaena, S., Bettinardi, V., Cappa, S. F., Fazio, F., & Mehler, J. (1998). The bilingual brain: Proficiency and the age of acquisition of the second language. *Brain*, *121*, 1841-1852.

Pinker, S. (1995). *The language instinct*. New York, NY: Harper Collins.

Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language* Learning, *44*, 681-734.

Sarafidis, V., & De Hoyos, R. E. (2006). Testing for cross-sectional dependence in panel-data models. *The Stata Journal*, *6*(4), 482-496.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newsbury House.

Seliger, H. W. (1978). Implications of a multiple critical periods hypothesis for second language learning. In W. C. Ritchie (ed.), *Second language acquisition research: Issues and implications* (pp. 11-19). New York, NY: Academic Press.

Singleton, D., & Ryan, L. (2004). *Language acquisition: The age factor* (2^{nd} ed.). Clevendon, UK: Multilingual Matters.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in the first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*, 105-122.

Uylings, H. B. M. (2006). Development of the human cortex and the concept of the ‘critical’ and ‘sensitive periods’. In M. Gullberg & P. Indefrey (eds.), *The cognitive neuroscience of second language acquisition* (pp. 59-90), Malden, MA: Blackwell.

Weber-Fox, C., & Neville, G. H. (1996). Maturational constraints on functional specializations for language processing: ERP evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, *8*, 231-256.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT Press.

Zhu, W. (2011). The critical period of L2 acquisition studies: Implications for researchers in Chinese EFL context. *Journal of Language Teaching and Research*, *2*(6), 1217-1226.

Investigating early language development in a bilectal context

Loukia Taxitari[1,2,3], Maria Kambanaros[2,3], Kleanthes K. Grohmann[1,3]
taxitari@ucy.ac.cy, maria.kambanaros@cut.ac.cy, kleanthi@ucy.ac.cy

[1]University of Cyprus, [2]Cyprus University of Technology, [3]Cyprus Acquisition Team

Abstract. The study of language development has focused on monolingual, and more recently bilingual development, but a much under-studied situation exists for children who grow up exposed to two dialects of the same language. One such case can be found in the bilectal linguistic community of Cyprus, where two varieties of the same language, Cypriot Greek and Standard Modern Greek, co-exist and shape language development. This study presents the Cypriot Greek adaptation of the MacArthur Bates Communicative Development Inventory (CDI) along with data from five age groups, toddlers between 18 and 30 months of age. The preliminary data already show a clear pattern of increase in vocabulary production across ages, as expected, and a semantic profile of children which agrees with models of lexical development from other languages. The CDI has the potential to become a valuable tool for research and clinicians on the island, and the Greek-speaking world in general, but also to provide researchers with a thorough understanding of very early language development in the bilectal community of Cyprus.

Keywords: bilectalism, CDI, diglossia, language acquisition, lexical development, toddlers

Introduction

The investigation of language development over the past decades has mainly focused on monolingual development (e.g., Golinkoff, Hirsh-Pasek, & Hollich, 1999; Clark, 2004), although bilingual development has gained ground over the last 20 years, with researchers studying different aspects of it, from lexical and phonological development to the effects of bilingualism on cognitive function (e.g., Pearson & Fernández, 1994; Werker & Byers-Heinlein, 2008). A grey area between the two extremes, monolingualism and bilingualism, has recently received much-needed attention (Grohmann, 2014; Grohmann & Kambanaros, to appear): discretely *bilectal* populations (Rowe & Grohmann, 2013), that is, speakers in linguistic communities traditionally characterised as diglossia, where more than one variety of the same language co-exist. One such case is Cyprus, where the local dialect, Cypriot Greek (CG), co-exists with the standard variety, Standard Modern Greek (SMG); but this approach can arguably be extended to countries in which distinct dialects co-exist with a higher standard, such as Germany, Great Britain, Italy, Norway, or Switzerland.

Recent research on the development of Greek in Cyprus suggests that CG-speaking children acquire morphosyntax differently from their monolingual SMG-speaking peers in mainland Greece (Grohmann, 2011; Grohmann & Leivada, 2012; Kambanaros, Grohmann, Michaelides, & Theodorou, 2012). Also Durrant, Delle Luche, Cattani, and Floccia (2014) compared the phonological representations of familiar words between mono- and bidialectal 18-month-olds in British English and found that only monodialectal children could detect phonological mispronunciations of words, suggesting that multidialectalism may impact the degree of specificity of one's phonological representations in early infancy. The question then is whether children who are exposed to more than one language variety grow up as monolinguals or bilinguals, or whether there could be a third, intermediate, option between monolingualism and bilingualism with its own special characteristics (for discussion, see Kambanaros, Grohmann, Michaelides, & Theodorou, 2014; Grohmann & Kambanaros, to appear).

Recently, Taxitari, Kambanaros, and Grohmann (2015) used the CG adaptation of the MacArthur-Bates Communicative Development Inventory (CDI) to look at 2-3 year olds' lexical development, through the study of translation equivalent (TE) pairs in a first pilot study with the tool. TE pairs refer to words with a different lexical form in two varieties with the same meaning. CG-speaking children were reported to produce many such TE pairs, that is, both a CG and an SMG word for the same concept; this behaviour is suggested to arise in contrast to mutual exclusivity, as the reluctance to attach two labels to the same concept, evidenced in both monolingual and bilingual children from around 2 years of age (Markman & Hutchinson, 1984; Au & Glusman, 1990; Markman, Wasow, & Hansen, 2003). CDI data from English- and French-speaking children, however, show that bilingual children actually make use of multiple labels for a single concept from very early in life, exhibiting a lack of, or overriding, mutual exclusivity from as young as 13 months of age (De Houwer, Bornstein, & De Coster, 2006). Similarly, CG-speaking children use words from both varieties, CG and SMG, to refer to the same concept, departing from monolingual children's behaviour that are rather reluctant to attach two labels to the same concept.

The CDI is not limited to TE pairs, however. It has been widely used to describe children's language abilities at different ages, month by month: the number of words understood and produced, most popular words or semantic categories, word use, grammatical development, and more (Fenson, Dale, Reznik, Bates, Thal, & Pethick, 1994; Fenson, Bates, Dale, Goodman, Reznick, & Thal, 2000; Fenson, Marchman, Thal, Dale, Reznik, & Bates, 2007; Jørgensen, Dale, Bleses, & Fenson, 2010). Also, percentiles of collected samples can be produced and new data compared to available norms in order to help identify children at risk for language and communication difficulties. Data for various adaptations, both monolingual and bilingual, are now available online and can be used for comparisons between languages (see the websites of CLEX at http://www.cdi-clex.org and the Wordbank at http://wordbank.stanford.edu).

In the current study, we present data from the CG-CDI in children between 18 and 30 months, in 3-month intervals. The CG-CDI has been adapted for Greek-speaking Cyprus and data from parents of young children are currently being collected in an effort to better understand very early language development on the island. The aims of the study were two-fold: to study early lexical development, including the creation of a lexical-semantic profile of bilectal children's language development, and to investigate specific aspects of bilectal children's early language development which could give us clues to the question of how this group of children is best described linguistically. For the latter, we focus on TE pairs, which provide information on how flexibly concepts and words (acoustic forms) are treated by children in this bilectal population.

Method

Participants

Parents of children in five age groups (18, 21, 24, 27, and 30 months) participated in this study. Table 1 shows the mean age and standard deviation, as well as the gender distribution in each of the five groups. All children were recruited for the LexiKyp project (CG-CDI) through online outreach (Facebook, Cyprus Acquisition Team lab website, LexiKyp project website), other advertisements in the form of leaflets (nurseries, children's clinics, playgrounds), and recruiting events around Cyprus. Some parents were approached directly by the research team and others volunteered by contacting the project administrator themselves or signing up through an online registration system.

Table 1. Participants' information for five age groups

Age group in months	Number (girls/boys)	Mean age in months (standard deviation)
18	43 (22/21)	17.97 (.45)
21	27 (14/13)	20.81 (.26)
24	24 (9/15)	24.03 (.12)
27	27 (16/11)	27.04 (.31)
30	36 (14/22)	30.19 (.68)

Along with the CG-CDI, parents were asked to answer a number of demographic questions which might relate to and affect language development, modelled after the Language and Background Development Questionnaire (Paradis, Genesee, & Crago, 2011; Paradis, Emmerzael, & Duncan, 2010). These questions targeted information about different aspects of the child's development (premature birth, birth order in the family, frequent ear infections) and language environment (exposure to languages other than Greek, having a housemaid at home from a different country, or one of the parents not being Greek Cypriot), as well as parents' educational level and any history of language problems in the family (see section *Demographic Questions* below for more).

All children who participated were exposed only to (Cypriot) Greek from birth, and on a daily basis. None of the children was systematically exposed to any other language; children were excluded if the parents reported that the child was exposed to another language more than 10 hours per week. All children were full-term (less than 6 weeks premature) and had no history of hearing problems or ear infections. The questionnaire was completed only by mothers.

CG and SMG

Over the past decades there has been considerable discussion in the literature regarding an exact definition of the linguistic situation in Greek-speaking Cyprus. Recently, Rowe & Grohmann (2013) suggested that the country is currently transitioning through a state of diglossia and Tsiplakou (2014) argues for a partial convergence of a Cypriot *koiné* to Standard Modern Greek through innovative, structurally mixed forms together with systematic language alternation in the form of code-switching, code-mixing, and register shifting. This Cypriot *koiné* is the variety used in urban centres on the island, retaining many of the characteristics of CG but also leaving behind many of the features of CG geographical sub-varieties and replacing them with more standard-like features. For the purposes of this paper, by CG we will refer to the CG *koiné*.

Differences between CG and SMG can be traced at all levels of linguistic analysis:

- Concerning phonology, CG and SMG mainly differ in terms of certain consonants (germination and no voiced stops in CG), which make the *koiné* sound distinctly different from SMG.
- CG and SMG differ in several aspects of their inflectional morphology; however, within the *koiné*, CG and SMG often become mixed up with features from either being used with structures from the other variety.
- In terms of the lexicon, CG and SMG share a large proportion of their vocabulary, with certain lexical tokens existing only in one or the other variety, and others having different meaning across varieties.
- CG and SMG share most of Modern Greek syntax, but there are also certain CG-specific structures such as enclisis in indicative declaratives, *wh*-question formation, or the syntactic expression of focus.

At all linguistic levels, there are similarities and differences between the two varieties, with some levels more closely related than others. Phonology and syntax seem to remain quite distinct in the two varieties, while morphological features tend to be more mixed in the *koiné*. There is still considerable debate in the literature whether CG and SMG form part of a continuum or not, and the question which arises is when exactly during language development these different features are acquired and when they become separated (or even merged).

Although CG and SMG differ across all levels of linguistic analysis, for the purposes of this paper, we focus on the lexicons of the two varieties which are largely common.

CG-CDI: Words and Sentences

The CG adaptation of the MacArthur Bates Communicative Development Inventory *Words and Sentences* (Fenson et al., 1994) was used in this study. The *CDI: Words & Sentences* consists of two sections: Part I: *Words Children Use* and *Word Use*, and Part II: *Sentences & Grammar* and *Word Combinations*.

A number of demographic questions were given to parents before the questionnaire, which were based on an adaptation for Greek Cypriot parents of the *Developmental and Language Background Questionnaire* (Paradis et al., 2010; Paradis et al., 2011; Taxitari et al., 2015).

The first part of the CDI focuses on words and their use. The first section consists of a long list of words and parents are asked to mark if their child produces the items on the list. The questionnaire includes a total of 819 words, divided in 24 categories: Sounds (18), Animals (56), Toys (24), Food and Drink (84), Vehicles (18), Home Objects (63), Furniture and Rooms (40), Clothes (34), Outside Things (34), People (39), Body Parts (30), Games and Routines (45), Verbs (96), Descriptive Words (50), Places to Visit (24), Quantitatives and Articles (21), Pronouns (27), Propositions and Words for Place (30), Colours and Shapes (14), Numbers (21), Modal and Auxiliary Verbs (18), Connectives (9), Words for Time (12), and Question Words (12). Although words are presented in isolation, some context is provided to parents as words are divided in different semantic categories. 'Grammatical' categories also exist, such as modal and auxiliary verbs or question words.

The CG-CDI is the adaptation of the CDI in CG containing both SMG and CG forms. As far as the lexicons of the two varieties are concerned, differences between CG and SMG might be found both lexically and phonologically. So there are three ways a concept might behave across the two varieties:

- a concept might be lexically the same, for example, the words for *hand or mouth*, where the word could further be phonologically different (SMG [ˈçeɾi] and CG [ˈʃeɾi] for *hand*) or identical ([ˈstomɐ] in both varieties for *mouth*);
- a single concept might be lexically different in CG and SMG, for example, the word for *head* ([cefɐˈli] in SMG and [cʰːelle] in CG);
- a concept could exist in only one of the two varieties, for example, [tʰːoɾos] in CG is equivalent to *bath towel*, which does not exist as a single word in SMG where instead the word for *towel* in general is used, [peˈt͡setɐ].

In the CG-CDI, we list as separate entries only items which differ lexically. For this, we include both concepts with different words in the two varieties and concepts which can be found in only one variety. Words which differed phonologically in the two varieties were entered in the CG-CDI as a single entry, for example the above [ˈçeɾi] and [ˈʃeɾi] for *hand*, and where possible the two different pronunciations were provided. Parents were asked to mark which pronunciation their child uses.

The CG-CDI contains a total of 819 words (108 words found only in CG, 23 are Non-Language Specific Words, and 688 words found in SMG and CG) for 728 concepts. Fewer concepts exist than words because a single concept can correspond to both a CG and an SMG word, as described above. There are thus 91 such TE pairs, with words from both varieties that correspond to a single meaning. The number of words included in this adaptation is remarkably higher than other monolingual versions, such as the American English CDI. The number is also fairly lower than bilingual CDIs, which include two different lists of words, one from each language. The CG adaptation includes words from both varieties in one long list. The reason is that CG (as well as SMG) is a variety of Greek, and CG and SMG share a big portion of their lexicon. This results in very few CG-only words in the CG-CDI, and a large number of shared words.

The second section of Part I includes five questions on the child's use of words, and parents need to answer them on a 3-point scale ("never" - "sometimes" - "often") depending on how often their child uses the word in the particular way. These questions relate to the use of words in the absence of the actual object or event: if the child uses the word to refer to the past or the future, if the child talks about absent objects, if she will bring an object when someone asks for it, and if the child will talk about someone's object in the absence of the person to which the object belongs.

Demographic Questions

For the purposes of the current study a shorter version of the *Developmental and Language Background Questionnaire* was created, which is based on the ALEQ and ALDeQ questionnaires originally developed by Paradis et al. (2010, 2011) and subsequently modified in COST Action

IS0804 (Tuller, 2015). The questionnaire had been translated into Greek for a previous CG-CDI study, in order to control for the different factors which could affect children’s lexical development (Taxitari et al., 2015). The LexiKyp project version included the following sections:

1. general information about the child (name, birth date, gender, order of birth in family)
2. the child’s health history (frequent ear infections or other illnesses)
3. exposure to other languages (if and how much the child is exposed to another language, is there a housemaid from another country in the household, does one of the parents come from another country)
4. the parents’ educational level
5. any history of language difficulties/impairments in family

Procedure

The contact details of all volunteers in the study were registered in the LexiKyp database, which stored contact details and birth dates of children at different ages. Parents (exclusively mothers) were contacted when their child reached the right age for the study in one of five age groups: 18, 21, 24, 27, and 30 months. They were reminded about the project and the procedure, and asked if they would still like to take part. Parents who agreed to participate received an email which contained instructions on how to reach the online version of the CG-CDI on the SurveyMonkey website, along with a password for entering the study and a unique participant code for each parent. The first page of the online questionnaire gave the parent all the necessary information about the CG-CDI and asked for the parent’s consent to proceed. Each parent participated in only one age group; if, however, parents did not complete the questionnaire after the first contact, they were contacted again at a different age, unless they explicitly asked otherwise.

Parents were asked to complete the CG-CDI at their own time and place, but preferably when they would be uninterrupted. If they needed to stop and continue later, they could save their responses, sign out, and continue the completion at a later time. They were asked not to talk to other people or to the child herself while completing the questionnaire, and solely rely on their own knowledge about their child’s language and communicative skills.

In Part I, the vocabulary checklist of the CG-CDI, parents were instructed to mark the field if their child produces a word, or leave it unmarked otherwise. They were also informed that they would find some words in CG in the word list and that they would sometimes find a word for an object in both CG and SMG. They were instructed to mark the version their child uses or mark both if the child uses both. They were also instructed to accept different pronunciations of the word from the child, as long as the word is systematically used by the child to refer to the concept in question.

Scoring

For every item in the CG-CDI vocabulary checklist that the parent reported their child produced, a single point was given; fields left unmarked received no points. Extra words that the parent added were not considered. Words in the two varieties which correspond to the same concept were marked as TEs; for example, SMG [pɐsxɐˈlit͡sɐ] and CG [pɐpɐˈɾunɐ] for *ladybird*. There are 91 such pairs in the CG-CDI; each word received one point to yield a total vocabulary score for each child. In order to calculate a conceptual vocabulary score, all TEs received one point, irrespective of whether the child produced only the CG word, only the SMG word, or both. Following De Houwer et al.’s (2006) terminology, CG and SMG words which make up a TE pair are called *members* of that pair. So, when a child produces only the CG or only the SMG member, she is said to produce a *singlet*; when, on the other hand, the child produces both members of the pairs, she is said to produce a *doublet*.

Measures

In order to test children’s productive vocabulary in the bilectal CG-CDI, two measures were calculated, *total vocabulary score* (the total number of words the child can say, coming from both

SMG and CG) and *total conceptual vocabulary* (by subtracting the number of doublets a child says from her total vocabulary score). Also, the number of TE pairs produced was measured as well as the number of singlets and doublets produced in these pairs. Additionally, a *total CG score* and a *total SMG score* were calculated from the TE pairs. Total scores were also calculated for the following grammatical categories:

- Nouns: animals, food & drink, vehicles, toys, house objects, outside objects, body parts, places to visit, clothes, furniture & rooms, people, numbers, colours & shapes
- Verbs: verbs, modal & auxiliary verbs
- Function words: pronouns, quantitatives & articles, questions, prepositions & words for place, connectives & particles
- Adjectives: descriptive words
- Adverbs: words for time
- Other: sounds, games & routines

Analysis

The main statistical analysis employed was an Analysis of Variance (ANOVA) comparing *production* (total, doublets, singlets, SMG, CG, grammatical categories) across age groups and gender. Pearson *r* correlations were also run between the different grammatical category scores and the total production scores in the CDI.

Results

Vocabulary Production

A univariate ANOVA with total vocabulary score as a dependent variable, and age group and gender (male vs. female) as fixed factors, revealed significant main effects of age group, $F_{(4,156)} = 33.98$, $p < .001$, $\eta^2 = 1$, and gender, $F_{(1,156)} = 5.12$, $p < .05$; $\eta^2 = .61$, but no interaction between age and gender, $F_{(4,156)} = .79$, $p = .53$. Figure 1 shows the increase in vocabulary production across ages and separately for each gender; table 2 shows the mean vocabulary score for each age group separately, collapsed for gender.

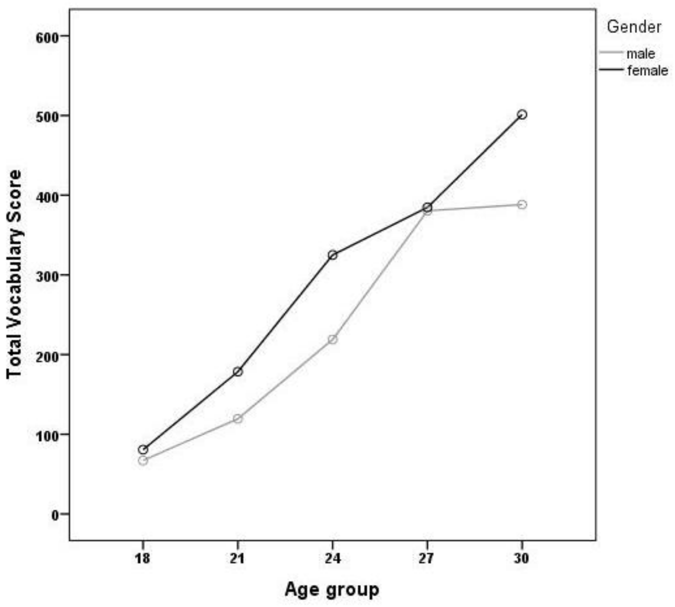


Figure 1. Increase in total vocabulary score across ages, separately for boys and girls (significant increase in total vocabulary by age, $F_{(4,156)} = 33.98$, $p < .001$, $\eta^2 = 1$)

Table 2. Mean vocabulary score for the five age groups, collapsed for gender

Age group in months	Number (girls/boys)	Mean vocabulary score (standard deviation)
18	43 (22/21)	74.05 (81.23)
21	27 (14/13)	150.04 (139.64)
24	24 (9/15)	258.67 (158.71)
27	27 (16/11)	382.96 (194.29)
30	36 (14/22)	432.17 (207.22)

In order to further investigate the main effect of gender, independent-samples *t*-tests were run for each age group independently comparing word production in boys and girls. No significant differences were found between the groups in any of the five ages (*p*s > .05).

Conceptual Vocabulary

As with the total vocabulary score, a univariate ANOVA with total conceptual vocabulary score as a dependent variable, and age group and gender (male vs. female) as fixed factors, was run. As the previous analysis, it revealed significant main effects of age group, $F_{(4,156)} = 34.73$, $p < .001$, $\eta^2 = 1$, and gender, $F_{(1,156)} = 5.09$, $p < .05$; $\eta^2 = .61$, but no interaction between the two, $F_{(4,156)} = .79$, $p = .56$. Figure 2 shows the increase in conceptual vocabulary across ages and separately for each gender.

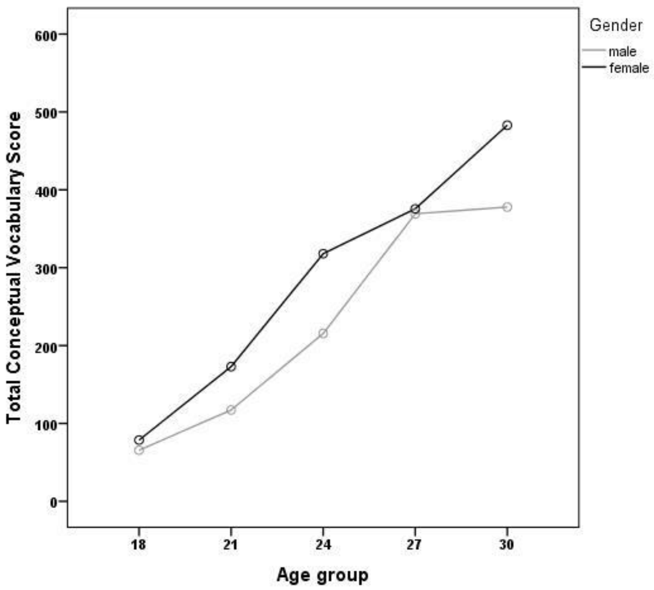


Figure 2. Increase in total conceptual vocabulary score across ages, separately for boys and girls (significant increase in total conceptual vocabulary by age, $F_{(4,156)} = 34.73$, $p < .001$, $\eta^2 = 1$)

Grammatical Class

Univariate ANOVAs were run separately for each of the five grammatical classes, with percentage of the total vocabulary as a dependent variable and age as a fixed factor. A significant increase in the percentage of the children's total vocabulary was shown for Nouns, $F_{(4,156)} = 5.61$, $p < .001$, $\eta^2 = .98$, Verbs, $F_{(4,156)} = 18.82$, $p < .001$, $\eta^2 = 1$, Adjectives, $F_{(4,156)} = 19.89$, $p < .001$, $\eta^2 = 1$, and Adverbs, $F_{(4,156)} = 13.84$, $p < .001$, $\eta^2 = 1$. A significant decrease was found for Other Words, $F_{(4,156)} = 17.14$, $p < .001$, $\eta^2 = 1$, and no change in the percentage as a fraction of the total for Function Words, $F_{(4,156)} = 1.29$, $p = .26$. Figure 3 presents the grammatical classes as a fraction of the total vocabulary across ages in a pie-chart plot.

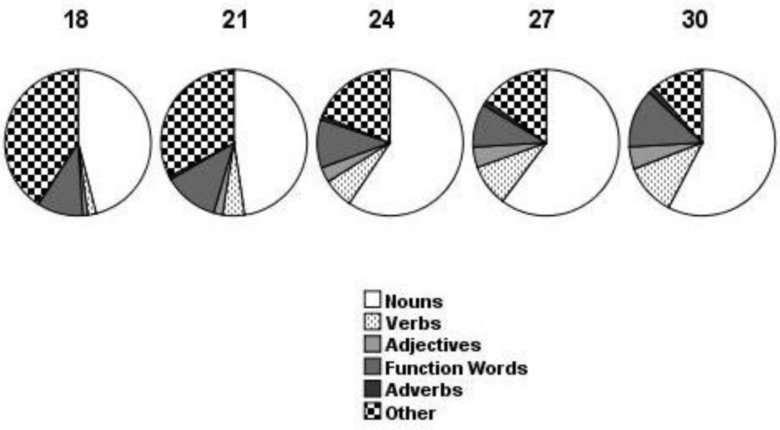


Figure 3. Percentage of each grammatical class as a fraction of children's total vocabulary

However, because of the high variability in children's profiles and total vocabulary scores at these early stages of lexical development, a second analysis was run without a division in age groups but taking into account only the children's vocabulary scores. These were correlated with the percentage of each grammatical class as a fraction of the total vocabulary in Pearson *r* correlations. As with the ANOVA, total vocabulary score correlated positively with Nouns, $r_{(157)} = .33$, $p < .01$, Verbs, $r_{(157)} = .84$, $p < .01$, Adjectives, $r_{(157)} = .77$, $p < .01$, and Adverbs, $r_{(157)} = .64$, $p < .01$, and negatively with Other Words, $r_{(157)} = -.69$, $p < .01$. There was no correlation with Function Words, $r_{(157)} = .09$, $p = .29$.

Translation Equivalent Pairs

A univariate ANOVA with number of TE pairs produced as a dependent variable, and age group and member type (single vs. doublet) as fixed factors, showed significant main effects of age, $F_{(4,314)} = 39.28$, $p < .001$, $\eta^2 = 1$, and member type, $F_{(1,314)} = 171.17$, $p < .001$, $\eta^2 = 1$, as well as an interaction between age group and member type, $F_{(4,314)} = 9.65$, $p < .001$, $\eta^2 = 1$.

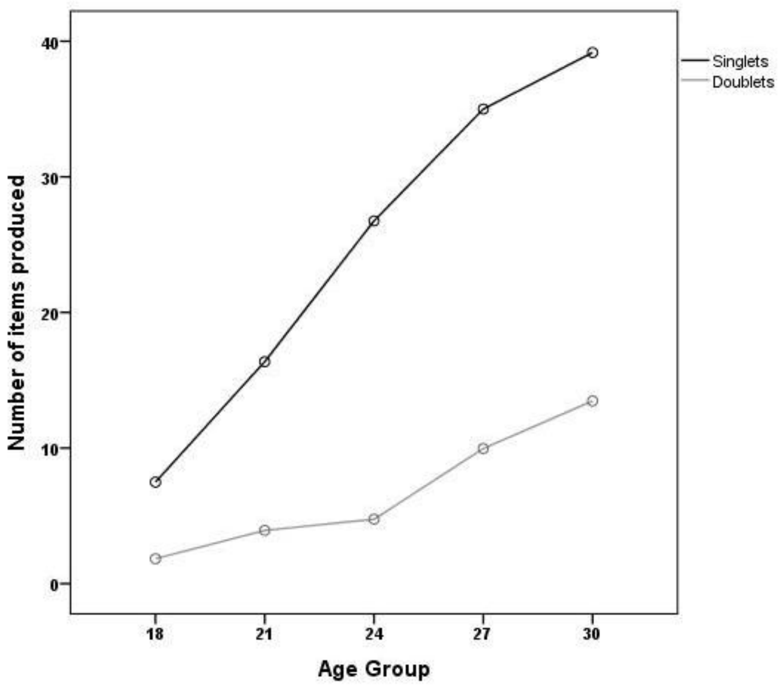


Figure 4. Translation Equivalent Pairs produced across ages, shown as singlets and doublets separately

Figure 4 shows the number of singlets and doublets produced across ages. Further one-way ANOVAs for doublets and singlets separately showed a significant increase in production for both across ages ($p < .001$).

A second analysis of the TE pairs focused on whether the words produced in those pairs came from CG or SMG. A univariate ANOVA with number of TE pairs produced as a dependent variable, and age group and variety (CG and SMG), revealed a main effect of age, $F_{(4,314)} = 38.09$, $p < .001$, $\eta^2 = 1$, but no effect of variety, $F_{(4,314)} = 2.1$, $p = .15$, or an interaction between the two, $F_{(4,314)} = .88$, $p = .48$. Figure 5 shows the production of CG and SMG words as part of TE pairs across ages.

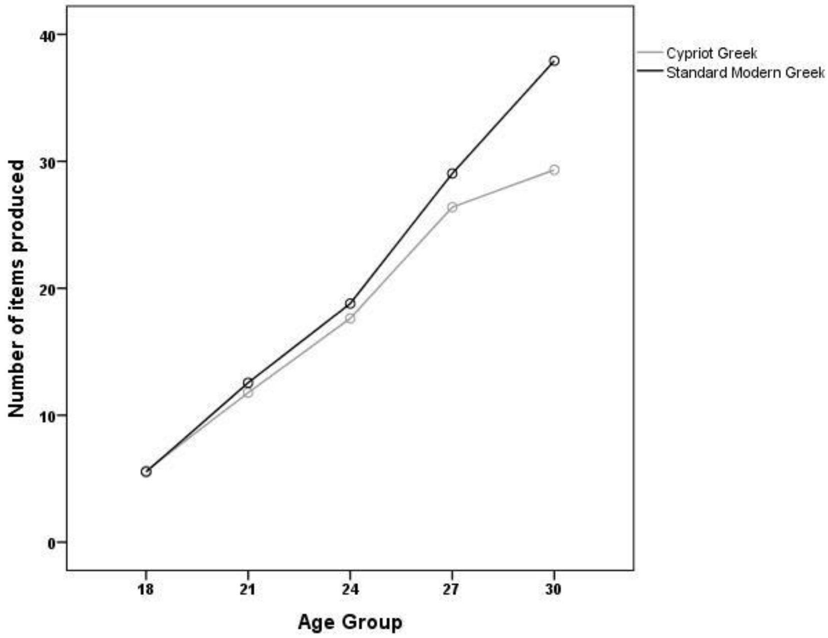


Figure 5. Translation equivalent members produced from each variety across ages

Discussion

The first aim of this study was the first investigation of language development in children who grow up as bilectal speakers in the diglossic community of Greek-speaking Cyprus. The children studied fell into five age groups spanning from 18 to 30 months of age. The collected data showed a clear increase in word production across these ages, similar to other CDI adaptations and to what can be expected from the word learning literature (see the CLEX website for data from several languages). This suggests that the CDI, which has been adapted for many languages (and cultures), is proving a suitable and powerful tool for the study of early language development in Cyprus as well.

An overall difference in word production between boys and girls was found, with girls producing more words. However, this difference disappeared when each age group was tested individually. Gender differences are not unexpected, although they are not found in all languages. Studies with the American English CDI report differences between boys and girls that place girls on average about one month ahead of boys, although these differences account for less than 2% of the variation found within and across ages, and they are mainly limited to production (Fenson et al., 2000). The fact that no differences are found for comprehension suggests that gender differences might actually be an artefact of the cultural environment in which the child is being brought up in.

A division of children’s productive vocabulary into grammatical categories also showed a nice progression from low variability within children’s early lexicons to high variability as children become more advanced word learners. The categories of words from which their lexicons are composed are in agreement with Caselli, Casadio & Bates’ (1999) four-stage model of lexical development: in Stage 1 lexicons are composed of *routines and word games*, which corresponds to our Other Words category. Stage 2 involves *reference*, and occurs between 50 and 200 words when lexicons are mainly composed of nominals, just like the lexicons of 18-month-olds in the CG-CDI study are mainly composed by Nouns (as well as Other Words). Stage 3 involves *predication*, and begins to develop after children have accumulated vocabularies of 100-200 words, similar to the increase in Verb and Adjective production found in this study. Finally, Stage 4 involves *grammatical function words*, and occurs after children have accumulated vocabularies of more than 400 words. This is also evident in our data in the absence of a notable increase in target words between 18 and 30 months of age. This is possibly due to the fact that these young children’s lexicons have not grown large enough yet to exhibit an increase in Function Words; however, function words are present from very early on, though they are thought to be memorised routines rather than actual grammatical markers (Caselli et al., 1999).

A well-known fact in the study of lexical development (and language development in general) is the high variance of children within and across ages (e.g., Fenson et al., 1994; 2000). Our five age groups also exhibited high variance. For this reason, an additional correlational analysis was run which did not include any pre-division of children into age groups, but instead compared the size of their lexicons (total vocabulary score) and the percentages of the different grammatical categories. This showed the same pattern as the analysis of the five age groups, suggesting that, besides the high variance, age groupings can still be legitimate in the analysis of lexical development. However, our groupings included children with a 3-month age difference and this could have allowed for the comparable results between the two analyses; groupings with smaller intervals (of the scale of one month) might not be equally informative, and instead a correlation analysis which takes into account the size of the lexicon might be more appropriate.

A final analysis involved children’s conceptual vocabulary and TE pairs. When TE pairs were taken away from the children’s total vocabulary, their conceptual vocabulary showed the same pattern as their total word production. An increase in concept production was noted with increased age, and girls overall produced more concepts than boys. The TE pairs analysis showed the simultaneous production of both singlets and doublets in every age group, and both of them increased with age. Children in Greek-speaking Cyprus from as young as 18 months of age can produce one or two words for a single concept coming from either variety of the language, CG and SMG. As they grow older, they learn more singlets (i.e. more concepts), but they also learn more doublets (i.e. two words for the same concept). This is in agreement with previous findings for bilectal children who acquire CG (Taxitari et al., 2015), but also with bilingual children who comprehend and produce words from two languages for a single concept (De Houwer et al., 2006). This flexibility of bilingual children to use more than one label for a single concept is taken as evidence against mutual exclusivity, a bias which guides monolingual children’s language development. Bilectal children are shown here to exhibit similar behaviour to bilingual children, which might suggest that bilingual and bilectal children could be closer on the monolingualism-bilingualism continuum than previously thought.

Conclusions

The LexiKyp project is the first large-scale investigation of language development in the bilectal community of Cyprus. Here we present data from five age groups, from 18 to 30 months of age. It is a first effort to study lexical development on the island and produce a semantic profile of these children. At the same time we aim to extend this profile to include grammatical development as well, and to study the relationship between vocabulary and grammar in a morphologically rich language, such as Greek. The CG-CDI is expected to become a valuable tool for researchers and clinicians on the island and the Greek-speaking world in general.

Additionally, we aim to provide answers to the question of where these children are on an assumed monolingualism-bilingualism continuum, applying the idea of comparative bilingualism (Grohmann, 2014) to much younger children; some initial evidence from the use of TE pairs suggests that there could be many similarities between bilectal and bilingual children, which could extend beyond the vocabulary to other aspects of language acquisition and cognitive development.

References

Au, T. K., & Glusman, M. (1990). The principle of mutual exclusivity in word learning: To honor or not to honor? *Child Development*, *61*(5), 1474-1490.

Caselli, C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language*, *26*(1), 69-111.

Clark, E. V. (2004). How language acquisition builds on cognitive development. *Trends in Cognitive Sciences*, *8*(10), 472-478.

De Houwer, A., Bornstein, M. H., & De Coster, S. (2006). Early understanding of two words for the same thing: A CDI study of lexical comprehension in infant bilinguals. *International Journal of Bilingualism*, *10*(3), 331-347.

Durrant, S., Delle Luche, C., Cattani, A., & Floccia, C. (2014). Monodialectal and multidialectal infants' representation of familiar words. *Journal of Child Language*, doi:10.1017/S03050000914000063.

Fenson, L., Bates, E., Dale, P. S., Goodman, J., Reznick, J. S., & Thal, D. (2000). Measuring variability in early child language: don't shoot the messenger. *Child Development*, *71*(2), 323-328.

Fenson, L., Dale, P. S., Reznik, S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), 1-185.

Fenson, L., Marchman, V.A., Thal, D.J., Dale, P.S., Reznik, J.S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes Publishing.

Golinkoff, R. M., Hirsh-Pasek, K., & Hollich, G. (1999). Emerging cues for word learning. In B. MacWhinney (ed.), *The emergence of language* (pp. 305-330). Hillsdale, NJ: Lawrence Erlbaum Associates.

Grohmann, K. K. (2011). Some directions for the systematic investigation of the acquisition of Cypriot Greek: A new perspective on production abilities from object clitic placement. In E. Rinke & T. Kupisch (eds.), *The development of grammar* (pp. 179-203). Amsterdam: John Benjamins.

Grohmann, K. K. (2014). Towards comparative bilingualism. *Linguistic Approaches to Bilingualism*, *4*, 336-341.

Grohmann, K. K., & Kambanaros, M. (to appear). The gradience of multilingualism in language development: Positioning bilectalism within comparative bilingualism. *Frontiers in Psychology: Language Sciences*.

Grohmann, K., & Leivada, E. (2012). Interface ingredients of dialect design: Bi-*x*, socio-syntax of development, and the grammar of Cypriot Greek. In A. M. Di Sciullo (ed.), *Towards a biolinguistic understanding of grammar: Essays on interfaces* (pp. 239-262). Amsterdam: John Benjamins.

Jørgensen, R. N., Dale, P. S., Bleses, D., & Fenson, L. (2010). CLEX: A cross-linguistic lexical norms database. *Journal of Child Language*, *37*(2), 419-428.

Kambanaros, M., Grohmann, K. K., Michaelides, M., & Theodorou, E. (2012). Comparing multilingual children with SLI to their bilectal peers: Evidence from object and action picture naming. *International Journal of Multilingualism*, *10*, 1-22.

Kambanaros, M., Grohmann, K. K., Michaelides, M., & Theodorou, E. (2014). On the nature of verb–noun dissociations in bilectal SLI: A psycholinguistic perspective from Greek. *Bilingualism: Language and Cognition*, *17*(1), 169-188.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1-27.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241-275.

Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, *43*(6), 474-497.

Paradis, J., Genesee, F., & Crago, M. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning* (2nd ed.). Baltimore, MD: Brookes Publishing.

Pearson, B. Z., & Fernández, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning*, *44*, 617-653.

Rowe, C., & Grohmann, K. K. (2013). Discrete bilectalism: Towards co-overt prestige and diglossic shift in Cyprus. *International Journal of the Sociology of Language*, *142*, 119-142.

Taxitari, L., Kambanaros, M., & Grohmann, K. K. (2015). A Cypriot Greek adaptation of the CDI: Early production of translation equivalents in a bi-(dia)lectal context. *Journal of Greek Linguistics*, *15*, 122-145.

Tsiplakou, S. (2014). How mixed is a 'mixed' system? The case of the Cypriot Greek *koiné*. *Linguistic Variation*, *14*, 161-178.

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 301-330). Bristol: Multilingual Matters.

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in Cognitive Sciences*, *12*(4), 144-151.

Rhythmic contrast between Swedish and Albanian as an explanation for L2-speech?

Mechtild Tronnier[1], Elisabeth Zetterholm[2]
mechtild.tronnier@ling.lu.se, elisabeth.zetterholm@isd.su.se

[1]Centre for Languages and Literature, Lund University
[2]Department of Language Education, Stockholm University

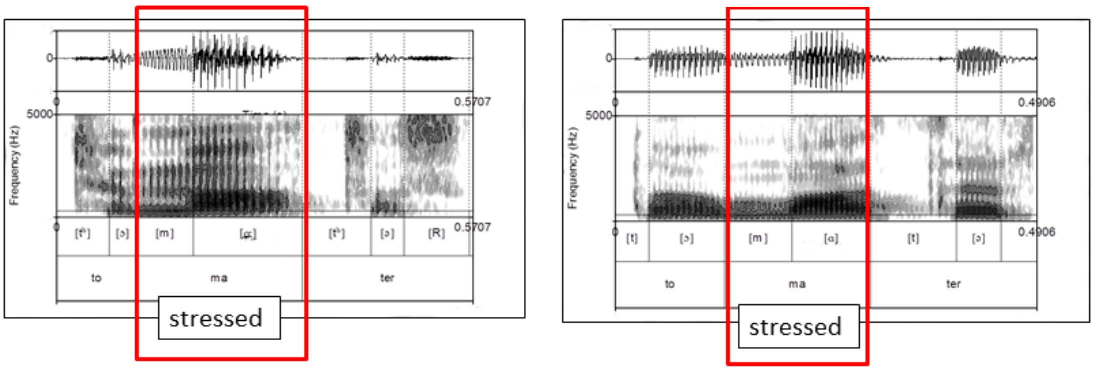
Abstract. Based on observations of the rhythmic structure of L2-speech produced by L1-speakers of Albanian – which suggest the occurrence of transfer – a study is presented here that compares durational aspects between the two languages. In order to do this, speech read by Swedish and Albanian L1-speakers was recorded and investigated, and normalized durational factors were analysed. The results, however, do not support the assumption that there is variation in the rhythmic structure between the two languages. According to the results, transfer cannot explain previous observations.

Keywords: language rhythm, prominence, transfer, Albanian, Swedish

Introduction

The acquisition of rhythm and the contrast in prominence in a second language is challenging for learners. Not only does placement of stress on the appropriate syllable in the word has to be learned, but also features which are used to express such a contrast, and the extent to which they are used, have to be acquired, as well. Such features are determined by variation in sound intensity, segment and/or syllable length, the presence of tonal accents and the degree of articulatory precision. Furthermore, the level of prominence of a sequence of syllables in an utterance is not always binary, but may be primary, secondary or tertiary.

When studying the accented L2-speech of Swedish produced by L1-speakers of Albanian, it was not always clear which syllable in a word carried the highest level of stress (Tronnier & Zetterholm, 2013). Therefore, one of the foreign accent features, is not simply due to incorrect stress placement, but requires further explanations. Observations from auditory analysis were that the reduction of the vowel – which is usually required in unstressed syllables in Swedish – was not carried out sufficiently.



Figures 1 and 2. Examples of the word *tomater* [tʰɔˈmɑːtəʁ] "tomatoes", pronounced by a speaker with L1-Swedish (left) and an L1-speaker of Albanian producing L2-Swedish (right). The illustration shows differences in length of both the stressed syllable (in the red frame) and the unstressed syllables produced by the two speakers.

Visual inspection of the speech wave gave the impression that vowels seemed to be of a similar length and at almost equal distance from each other, whether or not these were part of an anticipated stressed or unstressed syllable. An illustration of the difference between L1-speech and L2-speech of Swedish concerning that aspect is given in Figures 1 and 2.

To acquire insight into whether these observations are based on factors related to the rhythmic character of Albanian – i.e. the L2-learners' first language – and therefore a matter of transfer or on artefacts which emerged due to the experimental set up, a comparative study of the rhythmic structure of Albanian and Swedish, produced by L1-speakers of both languages, was carried out and the results are presented below.

The initial intention was to also analyse L2-Swedish produced by the same L1-speakers of Albanian. But due to poor performances during the recording sessions when reading the Swedish version of "The Northwind and the Sun" by most of the L2-speakers (i.e. extensive speech and reading errors, pauses, interruptions, hesitations and re-takes), that plan had to be abandoned.

Background

The nature of rhythm is based on the impression that a sequence of sounds, but also of other events recurs in a regular way. Such cyclic repetitions co-occur with the fact that subunits within the repeated parts are grouped together and that there is a clear division between the regularly repeated parts (cf. Bruce, 2012). With regard to sound perception, it has been reported from psychological experiments that listeners tend to impose a rhythmic structure onto clearly monotonous sound sequences (ibid.). Rhythm is therefore a basic human phenomenon.

Spoken language is also subject to rhythmic structuring and differences in rhythm are components which give the impression that different languages do not sound alike. The acquisition of the rhythmic structure in a foreign language is thus a matter that has to be taken into consideration in L2-instruction. It has been shown, however, that when presented with a non-lexical item - the CV-syllable "sa" - which preserved the original durational patterns of vowels and consonants from natural speech in synthesised speech samples, listeners could identify individual languages and distinguish between languages perceptually (Rasmus & Mehler, 1999).

In a historic description of language rhythm, languages were categorised as either stress-timed or syllable timed (Pike, 1945; Abercrombie, 1967). In that account, it was proposed that certain units appear in equal temporal intervals. Such isochrony was assumed to apply for the interval from one stressed syllable to the next in a stress-timed language and for each syllable interval for syllable-timed languages. The concept of isochrony has also been extended for the mora in a third type of languages, i.e. mora-timed languages. An example of a stress-timed language is English, of a syllable-timed language French and of a mora-timed language Japanese. However, the concept of isochrony has been questioned, and measurements of the intervals in focus have not given substantial support for its existence (e.g., Dauer, 1983). In addition, experiments with synthesized speech, in which isochrony has been strictly maintained, have shown that listeners experienced the rhythmic structure as being clearly unnatural (Bruce, 2012).

Instead, it has been proposed, that languages are more or less stress-based (Dauer 1983). In that way, different languages are found at different places along a scale of stress-basedness, depending on how central stress is in that particular language. Within that framework, rather than isochrony, three factors a) higher ratio of average syllable length between stressed and unstressed syllables, b) phonotactic complexity and c) vowel reduction in the unstressed syllable are considered to be representative for a language to fulfil the criteria to be placed at the higher end of the scale of stress-basedness. At that end of the scale, those languages have been placed, which traditionally were classified as stress-timed.

There is, however, evidence that some languages, like Polish, show mixed structures, e.g., phonotactically complex syllables co-occur with unstressed syllables that lack vowel reduction. Contrast in prominence can thus be manifested in a temporal domain, i.e. variation in duration, and in a quality domain, variation in articulatory excursion (c.f. Barry & Andreeva, 2001).

Interaction between different prosodic aspects can also have an influence on rhythmic structure, and languages with lexically flexible stress are mainly found among those which had previously been classified as stress-timed. In addition, it should not be forgotten that phonological aspects of quantity can have an influence on the rhythmic character of a language. Along these lines, the interaction of prominence at different levels should be taken into account as well. The occurrence of lexical stress (primary, secondary), phrasal prominence, prominence based on the information structure within an utterance, voluntary focus and the degree and way in which adverse elements are neutralised contribute to the nature of rhythm of each language. Such interactions are relevant factors which one should bear in mind when comparing the rhythmic structures of two or more languages.

When comparing the rhythmic structure between languages or dialects, the question about which units are those affected and are subject to a perceived difference still remains. Thus, a difference can be related to variation in duration and/or articulatory excursion concerning syllables as a whole, codas, vowels only or consonants, and consonant clusters.

Rhythmic typology: Swedish

Swedish has been assumed to be a syllable-timed language, similar to other Germanic languages (c.f. Engstrand, 2004; Nishihara & van de Wejer, 2012). Experimental studies, however, have shown that isochrony cannot be confirmed either for interstress intervals or for syllables in general (Eriksson, 1991; Strangert, 1985). Other criteria proposed to better typify the rhythmic character of languages with an alternative classification, i.e. the notion of syllable-based languages (Dauer, 1983) are, however, fulfilled to a large extent. Along these lines, Swedish comprises stressed syllables which, on average, have longer durations in the flow of speech than unstressed syllables. Furthermore, in unstressed syllables, vowel quality is more neutralised than in stressed syllables. Reduction phenomena also occur in coda consonants in unstressed syllables. And finally: Swedish has a complex phonotactic structure, where the consonant clusters in mono-morphemic codas can contain three consonants (e.g., *hemsk* “terrible”) and poly-morphemic codas can contain even more consonants (e.g., *skälmskt* “slyly”). It should, however, be pointed out, that in the latter case, consonant articulation is somewhat reduced in the flow of speech and not all consonants are pronounced completely.

On a lexical level, Swedish prosody incorporates aspects of quantity. The stressed syllable is always heavy and the unstressed syllable is light (Bruce, 2012). In addition, there is a structural diversity in the stressed syllable, which is also distinctive. This diversity manifests itself on the one hand by the occurrence of a long vowel followed by no consonant (*bo* [bu:] “live, reside (infinitive form)”) or a short and single consonant (*bok* [bu:k] “book”), and on the other hand by the occurrence of a short vowel followed by a long consonant (*fall* [fal:] “fall”) or a consonant cluster (*falsk* [falsk] “false”). The contrast in syllable weight between stressed and unstressed syllables already provides a basis for Swedish to belong to those types of languages with a rhythmic structure which were classified as being strongly stress-based.

It should also be pointed out that Swedish has primary plus secondary stress on different elements in a compound. Syllables carrying secondary stress are subject to quantity rules just like syllables carrying primary stress, and they are therefore heavy. The difference in stress realisation lies in the absence of a tonal accent on the secondary stress, which is a salient feature of the primary stress.

In running speech, the lexical stress in function words becomes neutralised, which contributes to the rhythm of Swedish speech. Furthermore, phrasal accents, information structure and focus lead to the stronger prominence of certain syllables, which would otherwise not always be considered as being stressed.

Rhythmic typology: Albanian

Literature on the prosodic system of Albanian is not extensive. It is known that there is at least one stressed syllable in a word. However, it is not clear if this is distinctive and flexible, as in Swedish, or fixed. According to Garlén (1988), the example of the minimal pair 'bari “the grass” and ba'ri “(shep-)herds” indicates the occurrence of distinctive stress, whereas according to Lloshi (1999), stress is mainly fixed on the final syllable, which leads to a trochaic rhythm.

A distinction in vowel quantity is found in some Albanian dialects, like Gheg, which is spoken in Kosovo and northern Albania. This is pointed out by Granser and Moosmüller (2003) in their investigation of vowel quality variation in stressed syllables. To what extent quantity distinction is restricted to stressed syllables is unclear, but it will be assumed that this is the case.

According to a survey on potential pronunciation problems for learners of English with Albanian as their L1 (Alimemaj, 2014), Albanian has length and weight on the last two syllables which co-occur with the stressed syllables. In addition, it is reported that Albanian is different from English in that every syllable is almost equal in length (ibid.). This remark is valuable for the current study, as this feature was observed to occur in L2-Swedish produced by L1-speakers of Albanian. However, no references to experimental studies are given in that survey (ibid.), which confirms the apparently impressionistic assessment.

Matter of contention

The focus of this contribution is a search for an explanation onto why L1 Albanian L2-learners of Swedish distribute prominence unlike L1-speakers of Swedish, rather than a search for a typological account for any of these two languages to be more or less stress-based. Details of rhythm-typology and stress-typology are presented above to demonstrate the complexity of the involved factors when approaching the current question.

The present study

Methodological approach

As a first approach, measurements of comparable units between the different languages were carried out in this study. As stress is very much connected to the vowel in a syllable and because it is easier to single out vowel onsets than syllable boundaries in connected speech, the procedure chosen was to measure the length of units from vowel onset to next vowel onset, thus obtaining the length of "quasi-syllables". For reasons of quantity features in the individual languages, this procedure was also preferred to measurements of e.g., vowel and/or consonant length. An example of segmentation can be seen in Figure 3.

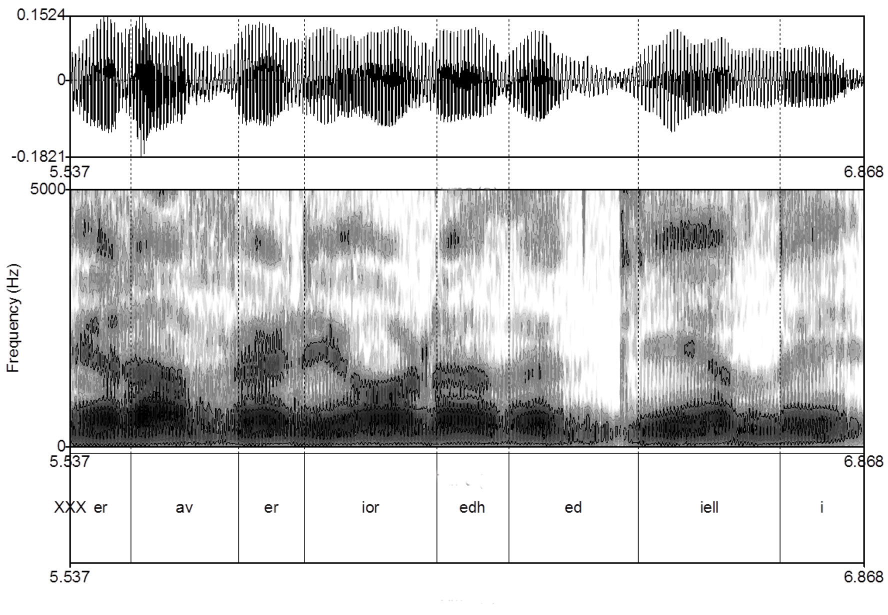


Figure 3. An example of the segmentation

Speakers, speech material and recordings

The speech material used for the present analysis was produced by seven L1-speakers of Albanian and seven L1-speakers of Swedish. All the Albanian speakers currently live in the South of Sweden and have been living there for a different period of time, between three months and 20 years. They all originate from Kosovo and they speak the Gheg variant of Albanian. The age of the Albanian speakers ranges from 25 to 54 years. The L1-speakers of Swedish also live in the South of Sweden, where they have been brought up. Therefore they also speak a variant of Southern Swedish, more precisely, they speak a variant of the Scanian dialect. Their age ranges from 23 and 49 years.

The material consisted of recordings of read speech of the story “The Northwind and the Sun”, produced in Swedish by the Swedish L1-speakers and produced in Albanian by the L1-speakers of Albanian. The speakers were asked to read the story twice, and the second version was used for further analysis.

The recordings were made on various occasions and in varied settings and locations. The recordings of the Albanian speakers were all carried out in a quiet school classroom with a Roland Digital Audio Recorder R-05 and a directed lavalier microphone (Shure). The recordings with the Swedish speakers on the other hand, were carried out in studio-like booths, furnished with damped walls at Lund University, using the same recording equipment as in the classroom settings.

Data Analysis

The recordings were manually segmented in PRAAT by inserting marks/boundaries at vowel onsets. In that way, the duration of a segment from vowel onset to the successive vowel onset - a quasi-syllable - could be calculated.

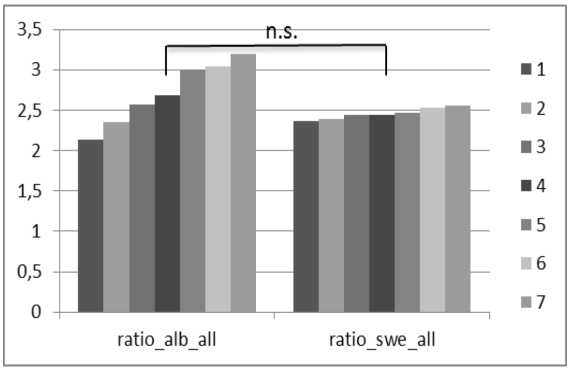
For each speaker, the average ($\bar{x}$) and standard deviation (sd) of the duration of those quasi-syllables was calculated. The measure of the standard variation shows the average alteration of syllable length and is more important for the current analysis than the values of the average. The larger the standard deviation, the larger the contrast in syllable weight and vv. In addition, for each speaker, the standard deviation to average ratio was calculated ($\bar{x}$/sd) to normalise the data for differences in speech tempo. A low value of the ratio represents a larger variation in duration of the quasi-syllables.

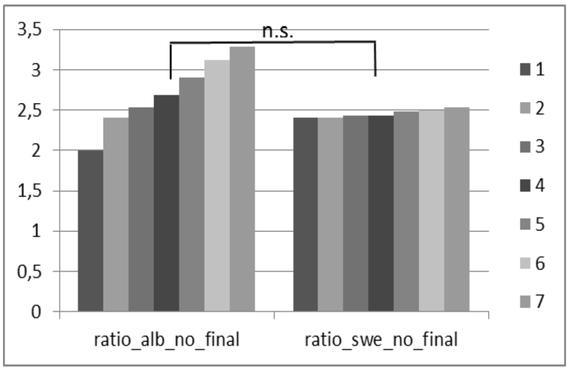
As phrase-final lengthening is a known trait, its influence on durational variation was also tested. Therefore, statistic tests between the data from the two languages were carried out for both: a) the data set including the phrase final quasi-syllables, and b) the data set excluding the phrase final quasi-syllables. In that way, T-tests for independent samples were carried out for the values of the ratios between the two languages.

Results

The range of variation in syllable duration between the two languages is not significant ($p > 0.07$ for the data including phrase final syllables and $p > 0.1$ for the data excluding phrase final syllables). Thus, syllable length does not vary to a larger extent for the speakers of Swedish than for the speakers of Albanian. The results are depicted in Figures 4 and 5, where L1-speakers of Albanian are presented in the left cluster and L1-speakers of Swedish in the right cluster.

In addition, Figure 4 shows the distribution of the average ratios for the both groups of L1-speakers when values for phrase-final syllables are also included in the statistical analysis, thus ignoring the effect of phrase-final lengthening. Figure 5, on the other hand, shows the same distribution, but excludes the values for phrase-final syllables. For both languages, it can be confirmed that the data including the values of the phrase final syllables show no significant difference from the data excluding those values ($p > 0.9$ for Albanian and $p > 0.9$ for Swedish). From both figures, it can be seen that regardless of the inclusion or exclusion of the phrase-final syllable into the statistic calculation, that there is more conformity for the ratios obtained from the Swedish speakers than for the Albanian speakers. Hence, speaker variation is larger for the Albanian speakers.





Figures 4 and 5. Average ratios for each speaker grouped by L1 (left block for Albanian and right block for Swedish in each figure). Figure 4 (left) includes phrase final quasi-syllables and Figure 5 (right) excludes phrase final quasi-syllables.

Discussion

The results of the present study do not support the assumption that neutralisation of syllable weight between stressed and unstressed in L2-Swedish produced by L1-speakers of Albanian, as found in an earlier study (Tronnier & Zetterholm, 2013), is based on transfer of rhythmic formation from L1. Based on the analysis of the range of length variation of quasi-syllables, the data and the analysis presented here do not show a significant difference in the rhythmic organisation between L1-Swedish and L1-Albanian speech. The results obtained here point to a similarity in length variation between the two languages, in that none of the languages presents us with larger variation in length of the quasi-syllables than the other. Larger variation could be interpreted to suggest that there is a clear difference in prominence between stressed and unstressed syllables. Lack of a clear length contrast as a feature in Albanian was pointed out by Alimemaj (2014) as a potential obstacle for L1-Albanian speakers learning English. Such dissimilarity does not seem to apply when comparing Albanian and Swedish, according to the analysis above.

One interesting aspect which the obtained data reveals (cf. Figures 4 and 5) is that the Swedish speakers show much more conformity in the ratio-values that represent the range of variation. The Albanian speakers show a more spread picture, where speaker 1, with a low value for the ratio, presents us with a large variation in duration, even larger than any Swedish speaker. The ratio obtained from speaker 7 in the Albanian group, however, tends to correspond to the expected outcome, based on previous observations. There is no explanation that can be given, other than external factors such as a different degree of comfort for the various speakers during the recording session.

In this study, however, only aspects of the length of rather large chunks of speech (the quasi-syllable) were analysed. This method had been chosen to overcome issues concerning quantity factors and questions of segmentation. Alternative duration measurements might represent a better way to find an explanation why a lack of rhythmic contrast in L2-Swedish produced by L1-speakers of Albanian was previously observed. In this sense, more detailed measurements of vowels and consonantal parts present in speech (*%V*, *ΔC*, *varcoC*, etc., cf. Dellwo, 2009) might give a better insight into the way in which Albanian differs from Swedish in its rhythmic structure, and if that could account for the rhythmic structure of L2-Swedish produced by L1-Albanian speakers.

Moreover, a closer investigation of qualitative factors (Barry & Andreeva, 2001) could give another insight into differences in rhythm between the two languages. Originally, however, alterations in durational factors in L2-speech were observed rather than differences in the use of e.g., articulatory reduction.

Conclusion

The results obtained in this study do not provide an explanation as to why L2-speakers of Swedish with Albanian as their L1 seem to vary the duration between stressed and unstressed syllables so little. As it was shown above, the durational variation between quasi-syllables in both languages is not as dissimilar as expected. On the basis of this study, transfer between L1-Albanian to L2-Swedish L2 as assumed in earlier observations cannot be accounted for. It must therefore be concluded that previous observations that syllable length varied less for Albanian L2-speakers of Swedish than for L1-Swedish speakers are based on behavioural grounds. For example, the production of L2-speech in a reading task might have led the L2-speakers to strongly focus on pronouncing the new text clearly and, therefore, produce fairly unnatural speech. Another explanation for the obtained results may be found perhaps in the unsatisfactory methodology used here, i.e. comparing the normalised duration of what was called "quasi-syllables". Other methods of analysis may thus be more suitable for this type of investigation, and will be considered in a follow-up study.

Acknowledgement

This study was made possible by using the facilities available at the Humanities Laboratory at Lund University and with the support of the staff there.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh, UK: Edinburgh University Press.

Alimemaj, Z. M. (2014) English phonological problems encountered by Albanian learners. *European Scientific Journal*, *10*, 1857-7431.

Barry, W. J., & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, *31*(1), 51-66.

Bruce, G. (2012). *Allmän och svensk prosodi*. Lund: Studentlitteratur.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51-62.

Dellwo, V. (2009). Choosing the right rate normalization method for measurements of speech rhythm. In S. Schmid, M. Schwarzenbach, & D. Studer-Joho (eds), *La dimensione temporale del parlato* (pp. 13-32). Torriana: EDK Editore.

Engstrand, O. (2004). *Fonetikens Grunder*. Lund: Studentlitteratur.

Eriksson, A. (1991). *Aspects of Swedish Speech Rhythm*. Gothenburg Monographs in Linguistics 9. Gothenburg University: Department of Linguistics.

Garlén, C. (1988). *Svenskans fonologi*. Lund: Studentlitteratur.

Granser, T., & Moosmüller, S. (2001): The schwa in Albanian, In *Proceedings of the 7th International Conference on Speech Communication and Technology* (pp. 317-320). Aalborg.

Lloshi, X. (1999). Albanian. In U. Hinrichs (ed.), *Handbuch der Südost-Europa-Linguistik* (pp. 277-299). Wiesbaden: Harrassowitz Verlag.

Nishihara, T., & van de Weijer, J. M. (2012). On syllable-timed rhythm and stress-timed rhythm in world Englishes: Revisited. *Bulletin of Miyagi University of Education*, *46*, 155-163.

Pike, K. (1945). *The intonation of American English*. Ann Arbor, MI: The University of Michigan Press.

Rasmus, F., & Mehler, J. (1999). Language identification based on suprasegmental cues: A study based on resynthesis. *Journal of the Acoustical Society of America*, *105*(1), 512-521.

Strangert, E. (1985). *Swedish speech rhythm in a cross-language perspective*. Almqvist & Wiksell International, Stockholm, Sweden.

Tronnier, M., & Zetterholm, E. (2013). Observed pronunciation features in Swedish L2 produced by L1-speakers of Albanian. *Studies in Language and Culture*, *21*, 85-88.

The effect of age of onset on long-term attainment of English (as L2) pronunciation in instructional settings in Spain

Katherine Elisa Velilla García, Claus-Peter Neumann
elisavelilla@gmail.com, cpneuman@unizar.es

University of Zaragoza

Abstract. Beginning in the late 1990s, the starting age for foreign language (FL) learning in Spain was progressively moved to the beginning of primary education and to preschool (Morales Gálvez, Arrimadas Gómez, Ramírez Rueda, López Gayarre, & Ocaña Villuendas, 2000), assuming that the earlier you start a foreign language, the better you will acquire it. The theoretical foundation that seems to support this assumption is the Critical Period Hypothesis (Lenneberg, 1967), which posits the existence of a threshold age after which starting to learn a language will never lead to full competence. A less categorical version of this hypothesis conceives of sensitive periods for different aspects of language (Long, 2013, Meisel 2013) with gradual offsets rather than abrupt discontinuities. Both versions have remained controversial (see, for example, White & Genesee, 1996; Birdsong, 2005). There is some agreement that a critical or sensitive period exists on the phonological level, probably caused by a change of perception, which becomes increasingly categorical while the structure of the L1 phonemic system is acquired (Brown, 2000; Ioup, 2008). In instructional settings in Spain, some findings appear to qualify this view. A large-scale research (the Barcelona Age Factor project) suggests long-term phonological advantages for very young starters on the level of perception but not for pronunciation (Fullana, 2006). However, the youngest starting age in the project was 8 while some researchers assume the age span during which perception changes to be between 5 and 7 (e.g., Flege, 1995). It would, thus, be interesting to observe if the earlier starting age in the current Spanish school system might have any significant effect on long-term attainment of English pronunciation. To do so, we recorded the speech of 20 adult Spanish speakers, 10 of whom had started to learn English at preschool while the rest had done so at the age of 8 or later, all other variables being equal. The speech was analysed for accuracy. The early starters achieved an average rate of 1,8 errors per 100 words (while the later starters achieved 7,2, which is 4 times higher). These results suggested that starting to learn English at preschool has a significant effect on long-term attainment of pronunciation in a Spanish instructional context. However, when we replicated the study with 20 new subjects, the results were much less conclusive, yielding no statistically significant difference.

Keywords: sensitive periods, long-term attainment, pronunciation

Introduction

In recent decades, Spanish children have started to learn a foreign language at progressively earlier ages. In order to address the perceived low English level of Spanish speakers, the moment when a foreign language is first taught to children was moved from the fifth grade of Primary School to the third grade in 1991 and eventually to the first grade and to the preschool stage, at first experimentally in various communities in the late 1990s (Morales Gálvez et al., 2000, p. 88-91), and then officially in all Spain in 2006. The common sense assumption behind these changes is, of course, that the earlier you start to learn a foreign language, the better you will acquire it.

The theoretical foundation that seems to support this assumption is the Critical Period Hypothesis (Lenneberg, 1967), which posits a threshold age after which starting to learn a language will never lead to full competence. A less categorical version of this hypothesis conceives of sensitive periods or phases for different aspects of language (Long, 2013; Meisel, 2013) without clear-cut threshold ages and with gradual offsets rather than abrupt discontinuities. Neither of the two versions has ever been conclusively confirmed and both have, in fact, remained controversial. Critics point out that the fossilization of a second language learner's interlanguage (Selinker, 1972) might be due to other factors than the learner's starting age (e.g., de Bot, Lowie, &Verspoor, 2006, p. 66-67), and that the fact that even Selinker allows for a small percentage (5%) of adult learners who do achieve a native-like level calls the hypothesis into question. Indeed, there are empirical studies (e.g., White & Genesee, 1996; Birdsong, 2005) that do report native-like L2 levels for late starters.

There is some agreement, however, that a critical or rather sensitive period might exist on the phonological level. Authors like Flege (1995) and Wode (1994) have argued that between the ages of five and seven the child establishes discrete L1 phonetic categories, which is why the child's aural perception gradually changes from a continuous mode, in which all phonemes are perceived as they really sound, to a mode of categorical perception, in which a phoneme of an L2 is assimilated to a similar-sounding phoneme in the L1 (see also Brown, 2000; Ioup, 2008).

Nevertheless, these authors' observations were made in natural second-language environments. If we turn to instructional settings, some findings in Spain appear to qualify their view. A large-scale research, the Barcelona Age Factor project, calls long-term phonological advantages for very young starters into question (Fullana, 2006). However, when the BAF project was carried out, English instruction in Spain started in the third form of primary education, so that the youngest starting age in the project was 8 while the above-mentioned age range, during which perception changes, is supposed to be between 5 and 7 (e.g., Flege, 1995). Since the starting age of English in the Spanish educational system was moved to the first year of primary education (and even to the preschool period) in the late 1990s, nowadays there are many young adults in Spain who started to learn English at the age of 6 or even earlier, i.e. before the end of the age range postulated by Flege and others. It would, thus, be interesting to observe if this earlier starting age in the current Spanish school system might have any significant effect on long-term attainment of English pronunciation.

Objectives

In this study, we wanted to verify whether an early starting age (7 years or younger) of learning English as a foreign language in Spanish instructional settings has a significant effect on long-term attainment of English pronunciation.

Methodology

Our subjects consisted of 20 adult Spanish speakers (in the age range between 20 to 47) all of whom had learned English in an instructional setting. None of them had been brought up bilingually, nor did any of them have close relatives who were native speakers of English. None of them had spent any significant time in any English-speaking country. They were divided into two groups of 10 subjects each: group A and group B. Group A had started to learn English at the age of 8 or later, while Group B had done so at an earlier age. Since we wanted to isolate the aspect of pronunciation, we decided to apply a discrete point test, providing all subjects with the following text (consisting of 180 words), chosen because of its syntactical simplicity, which would not put any cognitive demands on the subjects. Lexically the text consists of very basic high-frequency items, whose pronunciation any English learner beyond the beginner lever should be familiar with:

> *Hi! I'm Jennifer! I am 9 years old. I live in Houston, Texas with my mother, father, and two brothers. I like going to school but I hate doing homework and taking exams. At school, I study English, Spanish, Science, Social Studies and Mathematics. I love going to school and seeing my friends and teachers every day. I also like to play baseball after school. I don't have any sisters but my best friend, Olga, is just like my sister. We tell each other everything. We also study and watch TV together. When I grow up I'm going to be a nurse and take care of sick people.*

Each subject had to read the text aloud. Their speech was recorded and analysed for phonetical accuracy, classifying errors into interlingual and intralingual errors. The former represent errors of interference, a process in which an L1 structure is mistakenly transferred to the L2; the latter refer to errors that arise out of the very process of acquiring the L2 (Ellis, 1994, p. 59-60). Intralingual errors, in turn, were subdivided into errors of overgeneralization, i.e. errors that result from using an acquired rule in a context where it does not apply (e.g., forming the past tense on '-ed' with irregular verbs) and errors of simplification, in which the speaker reduces the correct forms to simpler ones so as to facilitate communication (frequently through omission of morphemes or segments, e.g., the omission of the third-person singular '-s').

The results of this initial research were striking: group B achieved 1,8 errors per 100 words while group A achieved 7,2, which is 4 times higher, suggesting a significant effect of the starting age on ultimate attainment. However, we suspected a bias in the formation of the groups. We found out that several of the subjects in group A had actually learnt English as a second foreign language, their first foreign language being French, something we had not expected. In those cases, since the subjects' eventual university entrance exams would have included French and not English, the instruction they had received in English was rudimentary at best and exclusively based on grammar and vocabulary, to the detriment of the receptive and productive skills. Therefore, they cannot be compared to subjects whose first foreign language was English and who received a more balanced instruction in that language at school. For these reasons, we decided to replicate the study with 20 new subjects, this time making sure that all the subjects in both groups had had English as a first foreign language at school. Since this second study has a much higher degree of validity (both groups being equal except for the starting age), the data reflected in this paper correspond to the second study.

Results

Group A

In this group we find many mistakes repeated by several subjects. However, most of them are idiosyncratic errors, this means errors committed by one subject which were not committed by others or were only repeated by one or two individuals. One of the most common mistakes among English speaking students is the inclusion of what is known as the epenthetic vowel: the insertion of a vowel in a word to make its pronunciation easier. With Spanish learners of English, this frequently affects words beginning with /s/ followed by a consonant. Thus, we find this pronunciation error committed by several of the subjects.

The word *school* whose phonetic transcription is /skul/, is found eleven times wrongly pronounced as [eskul]. This is because individuals have inserted the vowel /e/ at the beginning of the word in order to facilitate its pronunciation. The reason for that is that in Castilian the phoneme /s/ never forms any consonant clusters at the beginning of a word, so there was a negative transfer, the subjects uttering the words that begin with "s + consonant" like similar words in their language, such as *escuela*, *estudiar* and *español*. We find the same type of error in the words *study* (pronounced [estʌdi]) and *studies* (pronounced [estʌdiz]), as well as *Spanish* (pronounced [espænɪʃ]). Since these errors result from a negative transfer, they are classified as interlingual errors.

Another common mistake, which mainly occurs in the Spanish speakers born in Spain, is the pronunciation of the English /h/ (a glottal fricative) as the Castilian phoneme /x/, a velar fricative represented in writing by the letter 'j'. This error commonly occurs when we find the phoneme /h/ at the beginning of words, as in our data with the words: *hate* /heɪt/, *have* /hæv/, and *hi* /haɪ/, which some subjects have pronounced [xeɪt], [xæv] and [xaɪ] respectively, clearly cases of interlingual errors.

Another common mistake among Spanish speakers is the pronunciation of /r/. The English phoneme /ɹ/ is a postalveolar approximant, while the Spanish /r/ is an alveolar tap or trill, which some of our subjects applied to the pronunciation of the word *nurse* /nɜɹs/, reflected in the table above by doubling the /r/ [nɜrrs].

Similar to the above-mentioned errors and, therefore also classified as interlingual errors, are the cases in which participants pronounced some words of the text as if they were Spanish words as happened with the word *years*, which was mistakenly pronounced, on several occasions, as [jeʌrs] or [jɪʌrz], and *study*, pronounced [estudi/] instead of [stʌdi] (also containing the above-mentioned epenthetic vowel).

Less frequent cases were [doɪŋ] for *doing*, [ʌlso] for *also*, [nurs] for *nurse*, [haɪt] for *hate*, [broðərz] for *brothers* and [lov] for *love*.

Table 1. Group A errors

Type of errors		**Errors**
Interlingual errors		[espænɪʃ] (x1), [jeʌrs] (x3), [jɪʌrz](x1), [tɛsəs] (x1), [broðərz] (x1), [xeɪt] (x1), [hʌɪt] (x1), [eskul] (x11), [estʌdiz] (x3), [estʌdi] /(x3), [stʊdi] (x1), [nurs] (x1), [nɜrrs] (x1), [doɪŋ](x1), [esæmz] (x1), [lov] (x1), [ʌlso](x1), [xaɪ] (x1), [xæv] (x1), [gʊɛn] (x1)
Intralingual errors	Simplification	[tiʧər] (x1), [sɪstər] (x1), [sɪst] (x1), [ʤʌs] (x1), [ʤʌt] (x1), [ma] (x2), [wɑt] (x1), [ɪ] (x1), [doʊn] (x1), [zæmz] (x1)
	Overgeneralization	[gru] (x1)
Unclassifiable		[ʃpænɪs] (x1), [het] (x1), [nʌrs] (x1), [siʤɪŋ] (x1), [ɛɪx] (x1), [wɪtʃ] (x1)

A special case within the interlingual errors occurs with the words *Texas* /tɛksəs/ and *exams* /ɪgzæmz/, in which the letter "x" is pronounced /s/ by some students. This type of error is common among learners of English in Spain since in some Spanish regions, the letter "x" is pronounced /s/, and this pronunciation is transferred to the pronunciation of "x" in English words.

Another particular case of interlingual error is the pronunciation of *when* as [gʊɛn]. In Spanish the bilabial approximant /w/ does not exist, and Spanish speakers frequently pronounce it as a lenis bilabial plosive, [b]. Colloquially, however, many Spanish speakers exchange /b/ for [g] in *bueno*, pronouncing it [gʊɛno]. It is this colloquial (and very frequent) exchange that is transferred to the pronunciation of "when". Spanish speakers generally pronounce *would* and *wood* as [gʊd].

As to the intralingual errors made by the participants, most of them were errors of simplification, more precisely of elision: the subjects omitted the pronunciation of one or more phonemes in a few words: *sisters* [sɪstər] and [sɪst], *teachers* [tiʧɜr], *just* [ʤʌs] and [ʤʌt], *watch* [wat], *exams* [zæmz], *is* [ɪ], *my* [ma] and *don't* [doʊn].

The only error of overgeneralization made by this group was the pronunciation of *grow* /groʊ/ pronounced as [gru], applying the acquired form of the irregular past tense to the verb in general.

Other errors made by the subjects are mistakes that we have labelled as "non-classifiable", which means that they do not fit within the above-mentioned categories. Among these non-classifiable errors we can find *nurse* /nɜrs/ as [nʌrs], *with* /wið/ as [witʃ], *hate* /heɪt/ as [het], *seeing* /siɪŋ/ as [siʤɪŋ] and *each* /iʧ/ as [ɛɪx]. All these cases represent idiosyncratic errors that cannot be pinned down either to interference from L1 nor to overgeneralization or simplification but rather seem to correspond to individual misreadings of the words involved.

Group B

As in group A, a very common interlingual error in this group was the insertion of the epenthetic vowel before words beginning with /s/ followed by another consonant as happened with *school*, *studies*, *study* and *Spanish*, where many individuals inserted the vowel /e/ at the beginning.

Other interlingual errors were made with the words *doing*, *when*, *also*, *hi* and *English*, all of which were pronounced as if they were Spanish words by some subjects, so that the negative transfer affected not only a single phoneme, but the whole word. As in group A, we find *when* pronounced as [gʊɛn], which has already explained above. On the other hand, we find the word *everything* pronounced as [ɛvriθɪŋx], which is due to the fact that in Spanish the written letter “g” is in certain contexts realized as a velar fricative, a pronunciation that Spanish speakers frequently transfer to English words ending with “-g” or “-ng”.

Table 2. Group B’s errors

Type of errors		**Errors**
Interlingual errors		[espænɪʃ] (x2), [doɪn] (x1), [eskul] (x4), [gʊɛn] (x1), [eŋglɪʃ] (x1), [estʌdi] (x1), [ɛvriθɪŋx] (x1), [ʌlso] (x1), [xaɪ] (x1), [estʌdiz] (x1)
Intralingual errors	Simplification	[wat] (x1), [ʤʌt] (x2), [doʊn] (x2), [ɪ] (x2), [saɪən] (x1)
	Overgeneralization	[laɪv] (x1), [tɔkɪŋ] (x1), [tɪl] (x3), [heɪv] (x1)
Unclassifiable		[bʌd] (x1), [xʌv] (x1), [het] (x1), [lɪv] (x1), [bet] (x1), [aɪm] (x1)

Within intralingual errors of simplification, we found the elision of phoneme /z/ in *is*; the phoneme /s/ in the words *just* and *science*, pronounced [saɪən[; the phoneme /t/ in *don’t* and the phoneme /ʧ/ in *watch*.

We found four intralingual errors of overgeneralization. In the case of the word *tell* (found three times erroneously pronounced as [til]), the subject has successfully acquired the pronunciation of the letter “e” as /i/ and applies it to contexts (“e” + final consonant) in which it should be pronounced /ɛ/. Two more cases of overgeneralization occur in the words *live*, pronounced as [laɪv] (erroneously applying the aquired pronunciation rule “i” + consonant + “e” = /aɪ/) and *have* as [heɪv/ (overgeneralizing the rule “a” + consonant + “e” = /eɪ/). A complex case of overgeneralization is represented by the realization of *taking* as [tɔkɪŋ]. Here the speaker applies the spelling of the irregular past tense form (*took*), wrongly pronouncing it as /ɔ/, to a regular form of the verb

Among the Non-classifiable errors we find the words *but* [bʌd] and [bet]), *hate* [het] and [xʌv], *love* [lɪv] and, finally, *I* [aɪm].

Table 3. Comparison: Group A vs. Group B

	Interlingual errors	**Intralingual errors**	**Unclassifiable errors**	**Total number of errors**
Group A	37	12	6	55
Group B	15	14	6	35

Conclusion

In the second study, the early starters achieved an average rate of 1,9 errors per 100 words while the later starters achieved 3,1. In spite of this clear difference, a t-test yields a p-value of 0,16, which suggests that the difference is not statistically significant. One can observe that the biggest difference between the two groups can be found in the category of interlingual errors, but even if the p-value for this category is slightly better (at 0,13), the difference is still not statistically significant. Our second study can therefore not confirm conclusively that starting to learn English at preschool has a significant effect on long-term attainment of pronunciation in a Spanish instructional context (nor on the types of errors produced).

References

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & M. B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 109-126). Oxford, UK: Oxford University Press.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (ed.), *Second language acquisition and linguistic theory* (pp. 4-63). Oxford, UK: Blackwell.

de Bot, K., Lowie, W., & Verspoor, M. (2006). *Second language acquisition: An advanced resource book*. New York, NY: Routledge.

Ellis, R. (1994). *The study of second language acquisition*. Oxford, UK: Oxford University Press.

Flege, J. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore, MA: York Press.

Fullana, N. (2006). The development of English (FL) perception and production skills: Starting age and exposure effects. In C. Muñoz (ed.), *Age and the rate of foreign language learning* (pp. 41-64). Clevedon, Avon: Multilingual Matters.

Ioup, G. (2008). Exploring the role of age in the acquisition of a second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition* (pp. 41-62). Amsterdam: John Benjamins.

Lenneberg, E. (1967). *Biological foundations of language*. New York, NY: Wiley.

Long, M. (2013). Maturational constraints on child and adult SLA. In G. Granena & M. Long (eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 3-41). Amsterdam: John Benjamins.

Meisel, J. M. (2013). Sensitive phases in successive language acquisition: The critical period hypothesis revisited. In C. Boeckx & K. K. Grohmann (eds.), *The Cambridge handbook of biolinguistics* (pp. 69-85) Cambridge, UK: Cambridge University Press.

Morales Gálvez, C., Arrimadas Gómez, I., Ramírez Rueda, E., López Gayarre, A., & Ocaña Villuendas, L. (2000). *La enseñanza de lenguas extranjeras en España*. Madrid: Ministerio de Educación, Cultura y Deporte.

Selinker, L. (1972). *Rediscovering interlanguage*. New York, NY: Longman.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, *12*, 233-265.

Wode, H. (1994). Nature, nurture, and age in language acquisition: The case of speech perception. *Studies in Second Language Acquisition*, *16*, 325-345.

Russian-English intonation contact: Pragmatic consequences of formal similarities

Nina B. Volskaya
volni@phonetics.pu.ru

Saint-Petersburg State University

Abstract. The paper describes formal similarities in the tonal patterns of Russian and English rising-falling intonation and states the differences in perception, application and functional load in the two languages with consequences on L1-L2 contact. It focuses on recent changes in the realization of the Rise-Fall in the speech of young native Russians and regards perceptual consequences for representatives of the older generation of Russian speakers. It presents the results of a perceptual study on this intonation and reports cases of miscommunication between the two generations of Russian speakers, projecting possible situations of miscommunication between Russian speakers of English as L2 and English learners of Russian .

Keywords: intonation contact**,** the rise-fall, Russian intonation, English intonation, L1-L2 prosodic and pragmatic interference

Introduction

Rising intonation is thought to be a universal feature of interrogation and non-finality (Hirst & Di Cristo, 1994). At the same time, its concrete patterns may be language specific. Rising-falling intonation has always been considered to be the most common pattern for general questions and non-finality in Russian. Importantly, this phonetically complex contour involving a rise on the tonic syllable followed by a fall in the post-tonic part is associated by native Russians with rising intonation. It is the most frequent intonation in spontaneous and read Russian speech. The shape of the tone as such is not associated with emotional meaning though, just as any contour, it can have variants used in expressive speech, e.g., a widened pitch range and increased rising/falling intervals suggesting some sort of emotion, the most common of which is surprise.

Observations over intonation patterns in Russian and English indicate more formal similarities than expected. First of all, there is the Rise-Fall in the English inventory of complex tones (O'Connor& Arnold, 1973), and there is rising-falling realization of the phonetically simple High Drop. Without going deep into detail in the discussion about the phonological status of the Rise-Fall in English, let us concentrate on perceptual and pragmatic aspects: first, both tones (simple and complex) are falling tones to any English ear; second, the Rise-Fall is not a neutral intonation: it is associated with a whole set of emotions, from great surprise and challenge to very unfriendly or a haughty attitude (O'Connor& Arnold, 1973).

Observations over the language behavior of Russian speakers of English and foreign speakers of Russian allow us to admit that this intonation pattern is often misinterpreted phonetically, phonologically and, consequently, pragmatically. “Intonation and prosodic features ... are part and parcel of pragmatic aspects of language, exerting a subtle, yet decisive, influence on the way in which native speakers perceive and interpret the linguistic behaviour of the non-native speaker. An intonational "error" is probably more serious than a segmental one because segmental mistakes do not relate to the pragmatic aspects of speech as directly as do suprasegmental mistakes” (Toivanen, 2001). In this paper we shall try to have a closer look at some of them.

The Rise-Fall in Russian

Russian is known for its specific questioning (a yes-no question) and non-final intonation: a rise on the tonic syllable followed by a steep fall on the post tonic part, if any (Figure 1).

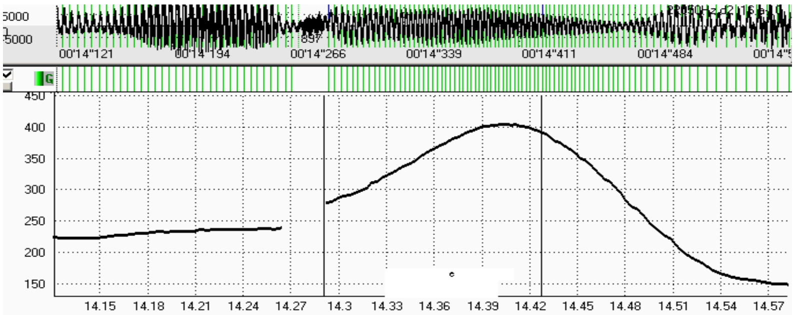


Figure 1. Typical intonation pattern for Russian general questions: F_0 peak on the tonic vowel

In neutral questions and non-final intonation units, the F_0 peak should coincide with the tonic vowel (Brysgunova, 1980). In non-final units, there is a somewhat smaller F0 excursion on the tonic syllable. The falling tone in the post-tonic part is normally rather abrupt. In questions, it reaches the speaker's lowest pitch; and in non-final units it often drops to medium pitch. This type of rising-falling intonation is the most common for general questions and non-finality and it is the most frequent intonation pattern in spontaneous and read Russian speech. Phonologically and perceptually, this phonetically complex contour is associated with *rising* intonation by native Russian speakers. The shape of the tone as such is not associated with any emotional meaning. As any contour, however, it may have variants in expressive speech (e.g., widened pitch range and increased rising/falling intervals) suggesting some sort of emotion, the most common of which is surprise.

In recent studies of intonation variation in Russian spontaneous and read speech, we came across realizations of the rise-fall characterized by shifting of the F0 peak further to the right, so that it is either late in the vowel (or the tonic syllable), or outside the syllable altogether (Volskaya, 2008). They were commonly used by young Russian speakers (Figure 2).

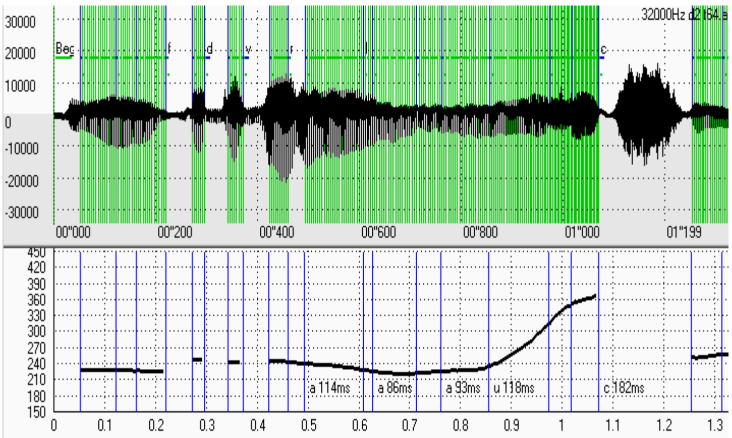


Figure 2. Late F0 peak on the post-tonic syllable of the word "ulitse" in a non-final unit (female speaker)

These observations (Volskaya, 2008) were confirmed by a follow-up study devoted to the phonetic realizations (i.e. late F0 timing) of the non-final and question intonation in Russian. The study, supervised by the author of this article, showed that these patterns have become particularly common in young Russian speakers (Demidchik, 2009).

Right-shifting of the F0 peak in the rise-fall in questions was mentioned by Brysgunova (1984) as a possible means for providing special emphasis on the question, and adding to it a note of astonishment, criticism, challenge, etc. Thus questions accompanied by a late F0 peak placement are by no means neutral requests for information. The rise-fall with a displaced F0 peak in questions as well as non-final units should be regarded differently by representatives of older and younger generations of Russian speakers. Here, we face a mismatch between young speakers' intended neutral request for information and its misinterpretation by the representatives of the older generation, having their own views on how neutral questions ought to sound like: in this situation the real intentions of a young speaker may be misinterpreted, and a communication conflict may result.

Experiment design and results

To confirm or reject this proposition an auditory experiment was carried out under the supervision of the author using research material specially designed for the purpose (Demidchik, 2009). That is, 90 general questions and non-final intonation units characterized by a late F0 peak placement were selected from interviews where five female speakers aged 20-22 were recorded. They were presented to three groups of listeners: school children aged 12-16, students aged 20-24 and a group of listeners aged 50-60 respectively. The subjects were to answer whether the question they hear is a neutral one, or tick the word indicating a particular emotion if they perceive it in the intonation unit they hear. Results are presented in the Table 1.

Table 1. Distribution of the types of response in the three groups of listeners (%)

Type of response	Subjects, responses (%)		
	School-children	Students	Grown-Ups
Neutral	55	55	33
Emotional	34	39	67
No answer	11	6	0

Although utterances produced with a rise-fall nucleus having a late F0 peak should have been associated with a pronounced emotional reaction, as it was stated by Brysgunova (1984), the data presented in Table 1 suggest, that for two groups of young listeners the effect of late F0 timing is much smaller, than for the third group of subjects ("grown-ups", potentially, their parents or grandparents), who associated it with a particular set of unintended emotions: impressed, surprised, happy, on the one hand (these attitudes were observed in the group of young listeners as well) , and challenging, reproachful, antagonistic, haughty, boasting etc. on the other (attitudes not in the list of young subjects' responses).

We may conclude that misunderstanding which (already) exists between two generations (generation gap) may be intensified by differences in the mental prosodic lexicon of young Russian speakers and representatives of older generation. There is a clear mismatch between:

- intention (neutral request for information)
- realization (a displaced F0 peak), and
- perception and interpretation (unintentional and often unpleasant for the listener emotional coloring of the message).

This seems to be an interesting and special case of the intra-language interference.

The Rise-Fall in English

As stated above, authors are not unanimous in defining the phonological status of the rise-fall: some consider it as a phonetic variant of the High Drop (see Crystal 1976 for discussion). At the same time, in traditional models of English intonation (Kingdon, 1957; O'Connor& Arnold, 1973; Cruttenden, 1986), it is widely acknowledged that the rise-fall signals strong emotions, either negative or positive, when the speaker is impressed, either favorably or unfavorably. It is claimed to convey various attitudinal meanings from obviousness to exasperation. According to O'Connor and Arnold (1973), who include the rise-fall in the inventory of English complex falling tones (the Jackknife), it may sound challenging, antagonistic, authoritative, haughty. Others report such attitudes as "teachingly reproachful" (Schubiger, 1958), asserting, signaling great annoyance or satisfaction (Crystal, 1976; Gunter, 1972). The most common general label seems to be "great surprise".

At the same time, in certain varieties of English, rising-falling intonation is widely used in *neutral* discourse, as it is not associated with any of the listed meanings: "In fact the Welsh employ the rise-fall … in circumstances where it would not be used in Southern England" (Jones, 1967). D. Jones (1967) among others, also mentions "a displaced F0 peak as a means for providing extra emphasis".

Discussion

As follows from the description, the English rise-fall is very similar in its phonetic realization to the Russian rising-falling intonation with a displaced F0 peak described above. It seems to suggest a similar set of attitudes or emotional meanings. The problem is that young Russian speakers, unaware of the set of attitudes associated with it in their native language may use it in their English speech, as well. Moreover, for Russian speakers the rising-falling intonation is phonologically and perceptually rising: when they ask a question using this intonation they mean nothing but a neutral request for an answer. What can we expect from this situation in L1-L2 contact? For an English speaker this intonation is phonologically falling. Heard in a question, it can be interpreted as "insistence on agreement", "seeking confirmation" (Volskaya, 1985), coupled with the attitudes normally associated with the Rise-Fall.

Of course, there is no reason why students learning English should not pronounce general questions with falling intonation. In teaching practice of English intonation at the Department of Phonetics, we no longer adopt the standard assumption in intonation literature that there exists for certain types of sentences only one common neutral tonal pattern: for yes-no questions a (low) rising tone is postulated. Students are well aware of the fact that questions of this type are more often than not pronounced with falling intonation. Moreover, they learn from literature (e.g., O'Connor & Arnold, 1973) and from experience (a two-year course on English pronunciation, including intonation based on O'Connor and Arnold's (1973) system of Tone Groups), that "any syntactic pattern can be spoken on any intonation pattern" (Brown, 1975). "Any" here means any from the set which native English speakers use automatically and which L2-English speakers ought to learn for use in appropriate situations.

Conclusion

Comparing rising-falling intonation patterns in Russian and English, we discovered a very curious situation: differences in the functional load and interpretation of the rise-fall in various parts of England may lead to a intra-language inter-dialectal prosodic interference. In Russian, differences in phonetic realization of the rise-fall have resulted in the communicative conflict between younger and older generations: children and their parents seem to speak different languages.

Formal tonal similarity of the rise-fall intonation contour in L1 and L2 may lead speakers to believe erroneously in its functional "sameness" as well, and thus in the possibility of a positive transfer. In that case, he would have to face the consequences. As far as L2 intonation is concerned, since the phonetic realization of the rise-fall in Russian and English displays similarity, a Russian learner of English may not find it difficult to produce the English rising-falling intonation pattern, but he is not aware of the effect this intonation may have on the listener. On the other hand, native English speakers may also fail to interpret the message and intentions of their partner correctly. This mismatch may lead to misinterpretation of the speaker's or listener's intentions and, as a result, of his or her personality.

Rising-falling intonation is observed in many other languages. In French, for example, intonational contours inventory has a rising-falling tone which presents very interesting semantic properties. It has been called “intonation d’implication” by Delattre (1966) suggesting that the contour conveys an implicit meaning. Beside this, the “implication” contour conveys various attitudinal meanings from obviousness to irritation, and is also used to mark contrastive focus. In English the contour which is used to convey implication as well as contrast is the Fall-Rise. By and large, intonation has been largely ignored in second language acquisition studies; many books on L2 learning include no or little reference to intonation or prosody; the subject seems to be too complicated, and empirical studies on prosodic interference are very few. At the same time, empirical research on native and target language intonation may shed light on the processes and consequences of prosodic interference and yield important social, sociolinguistic information about the prosody of speech varieties within a language and across languages.

References

Brysgunova, E. A. (1980). Intonation. In N. Shvedova (ed.), *Russian Grammar II* (pp. 96-122). Moscow: Nauka.

Brysgunova, E. A. (1984). *Emotsionaljno-stilisticheskije razlichija russkoj zvuchashchej rechi.* Moscow: MGU.

Brown, G. (1975). Phonological theory and language teaching. In J. O. B Allen & S. P. Corder (eds), *Techniques in Applied Linguistics* (pp. 98-121). London , UK: Cambridge University Press.

Cruttenden, A. (1986). Intonation. Cambridge, UK: Cambridge University Press.

Crystal, D. (1976). *Prosodic systems and intonation in English.*Cambridge, UK: Cambridge University Press.

Delattre, P. (1966). Les dix intonatons de base du Français. Online: http://mathilde.dargnat.free.fr/ INTONALE/ article-Delattre1966.pdf

Demidchik, L. (2009). *Foneticheslaja realizatsija i funktsionaljnaja nagruzka voskhodjasche-niskhodjashchei intonatsii v rechi russkoj molodezhi. SPb:SPbGU* (Diploma paper) (in Russian).

Gunter, R. (1972). Intonation and its relevance. In D. Bolinger (ed.), *Intonation* (pp. 194-215). Harmondsworth, UK: Penguin.

Hirst, D., & Di Cristo, A. (eds.) (1994). *Intonation: A survey of twenty languages*. Cambridge, UK: Cambridge University Press.

Jones, D. (1967). *The pronunciation of English*. Cambridge, UK: Cambridge University Press.

Kingdon, R. (1957). *The groundwork of English intonation.* London, UK: Longman.

O'Connor, J. D., & Arnold, G. F. (1973). *Intonation of colloquial English* (2nd ed.). London, UK: Longman.

Schubiger, M. (1958). *English intonation: Its form and function*. Tubingen: Niemeyer.

Toivanen, J. (2001). *Perspectives on intonation: English, Finnish and English spoken by Finns. Forum Linguisticum,* Band 37. Frankfurt: Peter Lang.

Volskaya, N. B. (1985) *Relevantnye priznaki intonatsionnoy interferentsii.* PhD thesis. Leningrad: Leningrad State University (in Russian)

Volskaya, N. B. (2008). Sootnoshenije intonatsionnogo tipa i variantov. *Materialy XXXVII Mezhdunarodnoj filologicheskoj konferenctii Formaljnye metody analiza russkoj rechi.* 7. Saint-Petersburg: SPbGU (in Russian)

Voice onset time in heritage speakers and second-language speakers of German

Joost van de Weijer[1], Tanja Kupisch[2]
vdweijer@ling.lu.se, tanja.kupisch@uni-konstanz.de

[1]Lund University, [2]University of Konstanz and UiT The Arctic University of Norway

Abstract. In this study, we examine possible effects of childhood country, residence country and age of onset, on pronunciation in heritage speakers and second-language speakers. For this purpose, we compare voice onset time (VOT) realizations in German across three speaker groups: a group of monolingual German speakers, a group of early bilingual (2L1) speakers who also spoke French, and a group of French native speakers who learned German as their second language (L2). The 2L1bilingual speakers had grown up either in Germany or France. The L2 speakers either lived in France, or had moved to Germany at the time of data collection. All participants were highly proficient in German, even though the L2 speakers scored lower on a test of grammatical and vocabulary knowledge than the early bilingual speakers. The VOT measurements were longest in the monolingual speakers. The measurements in the other two groups, while shorter than those in the monolingual group, fell within the range of what has previously been reported about German VOT. Remarkably, the L2 speakers did not differ significantly from the 2L1s, nor did we find significant effects in terms of childhood and residence country. We therefore conclude that neither an early age of onset nor a stay in the country where the heritage or L2 is spoken are necessary conditions for successful realization of VOT.

Keywords: voice onset time, heritage speakers, L2 speakers, French, German

Introduction

Heritage speakers are typically defined as speakers who grow up hearing and speaking a minority language in a naturalistic setting at home and who later become more proficient in the societally-dominant language. It has been claimed that proficiency in the heritage language does not always develop at age appropriate levels and it is often not mastered at a “native-like level” eventually (e.g., Benmanoun, Polinsky, & Montrul, 2013), although there is some controversy on this issue (e.g., Pascual y Cabo & Rothman, 2012; Kupisch, 2013; Putnam & Sanchez, 2013 for different views). There is indeed a substantial amount of studies showing that in some domains of grammar, heritage speakers do not attain monolingual-like end states during adulthood in spite of their early exposure to the language (see Benmanoun, et al., 2013 for an overview). Typically, the explanations are sought in the quantity of input, i.e. insufficient or decreasing exposure during pre-school and early school years, or input quality, i.e. exposure to attrited speech or speech by second language (henceforth L2) speakers.

Comparatively little research has been done in the domain of heritage speakers’ phonology. It is generally assumed that this domain is relatively well preserved (e.g., Rothman, 2009, Benmamoun, Montrul, & Polinsky, 2013), and existing studies indicate that heritage speakers have an advantage over L2 speakers (Au, Knightly, Jun, & Oh, 2002; Oh, Jun, Knightly, & Au, 2003; Chang, Yao, Haynes, & Rhodes, 2011; Kupisch, Barton, Hailer, Kostogryz, Lein, Stangen, & van de Weijer, 2014). At the same time, they rarely come to be perceived as native speakers in real life situations, or when their global accent is judged (e.g., Kupisch, Barton, Hailer, Kostogryz, Lein, Stangen, & van de Weijer, 2014). So far, the extant research has failed to determine the sources of non-native like attainment, possibly owing to methodology. First, the samples in heritage speaker studies are more often than not heterogeneous with respect to (i) age of onset (AoO) in the majority language and (ii) whether or not the heritage speakers were born in the heritage country or in the new host country. This means that prior linguistic knowledge when starting to acquire the majority language and length of residence in the new host country (implying a change in quantity and quality of input sources for the heritage language) might individually or jointly influence the speaker’s attainment in the heritage language.

The present study is concerned with these issues. We will compare Voice Onset Time (VOT) realization in German, produced by a group of monolingual (L1) German speakers, a group of early bilingual (2L1) French-German speakers and a group of French native speakers who spoke German as a late L2. Thus, the latter two groups differ in the age at which they started to learn German. The 2L1 speakers did this from birth, whereas the L2 speakers started at a later age. Furthermore, half of the 2L1 speakers grew up in France (thus being heritage speakers of German) whereas the other half had lived in Germany for most of their lives (thus being heritage speakers of French). Comparison between these two groups allows us to evaluate potential effects of input quantity during childhood, as the language that these two groups presumably heard most during their childhood differed. The L2 German speakers (L1 French) all spent their childhood in France, but half of them had moved to Germany at the time they were recorded. This permits us to evaluate the effect of the language environment at a later age. The data collection is based on spontaneous speech. All participants were proficient speakers of German, capable of holding a fluent conversation for half an hour on any kind of topic. As we will explain in more detail below, the VOT of German voiceless stop consonants is distinctly longer than that of their French counterparts. It therefore may be used as an indication of how well the speakers pronounced German. Specifically, long VOT is expected to sound more German-like than short VOT.

VOT in German compared to French

VOT differentiates the language-specific realizations of voiced (/b, d, g/) and voiceless (/p, t, k/) plosives. It refers to the interval between the release of the stop and the onset of voicing (Lisker & Abramson, 1964, p. 389). According to Lisker and Abramson (1964), there exist three different types of VOT: (i) voicing lead (voicing starts before the release), (ii) short voicing lag (voicing begins with the release or shortly after it), (iii) long voicing lag (voicing starts late after the release). Many of the world's languages distinguish two categories of stops, voiced and voiceless. In French, (i) voicing lead with negative VOTs characterizes voiced stops, and (ii) short voicing lag (with VOT values as short as 30 ms) characterizes voiceless stops. In German, voiced stops are produced with (ii) a short voicing lag, while voiceless stops are produced with (iii) a long lag. Thus, German voiceless stops have longer VOTs than French ones. In this study, we focus on the VOT of /k/. For this consonant, VOT ranges of 30-49 ms and 37-67 ms have been reported for French and German respectively (see Lein, Kupisch, & van de Weijer, forthcoming for an overview).

Findings on VOT differ substantially due to several factors that have a joint impact on VOT production. First, many languages display a hierarchy of shorter to longer VOTs ranging from /p/ over /t/ to /k/. VOT can further be influenced by syllable stress, speech rate, word length and the quality of the following vowel. Relatedly, stops in isolated words are said to have longer VOT than those in spoken sentences and spontaneous speech. Finally, there may be regional variation, as has been reported for German by Braun (1996, p. 25). Despite variation, VOT is traditionally considered as the categorical unit par excellence that characterizes voice: a plosive is voiceless if it falls into a certain VOT range, but if it crosses the relevant threshold, it is automatically perceived as voiced. Eimas, Siqueland, Jusczyk, and Vigorito (1971) showed that even newborns at one month of age can discriminate voiced from voiceless stop consonants on this basis. Since the German VOT of voiceless stops is noticeably longer than that of the French voiceless stops, we can predict that a French influence on German will result in relatively short VOTs compared to monolingual German speakers (i.e. resulting in short lag, similar to the VOT of German voiced stop /g/). Note, finally, that aspirated plosives, the focus of our analysis, are produced with a long lag and are considered to be more marked than their short lag counterparts. This means that they are typically acquired later and potentially more vulnerable to language influence or even attrition. In the present study, we address the following research questions:

(i) Do early simultaneous bilinguals differ from monolingual L1 speakers with respect to their VOT in German despite having the same AoO?

(ii) How do heritage speakers of German compare to bilingual speakers with the same language combination but speaking German as their majority language?

(iii) Do simultaneous bilinguals produce more monolingual-like VOTs than late L2 speakers, thus having an advantage through their early exposure to the language?

(iv) Do L2 speakers who live in the L2 country have an advantage in VOT production over L2 speakers who live in the L1 country?

Method

Participants

The speaker sample consisted of seven German L1 speakers, 14 2L1 speakers who acquired both German and French simultaneously from birth, and 14 L2 speakers whose native language was French and who started to learn German at the age of 11 or later, i.e. long after what is typically considered a "sensitive" period for pronunciation. The 2L1 speakers all came from families in which one of the parents spoke German and the other one spoke French with them, following the "one person - one language" strategy. Half of them grew up in France, and had moved to Germany during adulthood. On average, they had lived in Germany for 10 years (range 0.5-20.0 years) at the time they were recorded. The other half grew up and lived in Germany. The L2 speakers all had French as their first language. They reported that they had started to learn German when they were on average 15.4 years old (range 11-25 years). Half of them had moved to Germany during adulthood, and had been living in Germany for 1.58 years (range 2 weeks - 9 years) at the time of the recording. The 2L1 and the L2 speakers all completed a cloze test as an assessment of their general proficiency in German. This test consists of 45 items and focuses on grammatical and vocabulary knowledge.

Table 1 provides information about the speakers' country of origin, country of residence, age, and proficiency. As shown by the numbers in the table, the speakers were of approximately equal age range. In terms of their scores on the cloze test, there were differences, with the 2L1 groups obtaining higher scores than the L2 groups, and, within these two groups, the speakers from or residing in Germany obtaining higher scores than those from or residing in France.

Table 1. Speaker overview

Group	***n***	**childhood country**	**residence country**	**age mean (range)**	**cloze test mean (range)**
Monolinguals	7	Germany	Germany	35.0 (24-70)	---
2L1s from France	7	France	Germany	34.4 (24-40)	33.6 (13-41)
2L1s from Germany	7	Germany	Germany	29.3 (20-42)	41.0 (39-44)
L2s in France	7	France	France	30.4 (21-45)	20.7 (11-26)
L2s in Germany	7	France	Germany	33.6 (22-49)	23.5 (11-37)

Speech elicitation and material selection

The 2L1 and L2 speakers presented above had been recorded earlier, as part of the HABLA corpus (http://www1.uni-hamburg.de/exmaralda/files/e11-korpus/public/index.html). The data were collected during interviews with native speakers, thus representing naturalistic speech. Most L1 speakers had been recorded earlier as well (Lein et al., forthcoming). Unlike the 2L1 and the L2 speakers, the L1 speakers were shown a set of images of objects that started with a stop consonant, and were instructed to tell a narrative using the images. These data were collected for the purpose of eliciting a high number of nouns starting with a voiceless plosive and can thus be considered semi-spontaneous.

A total of 1045 words with /k/ in initial position followed by a vowel were extracted from the recordings for analysis. The target words were all monosyllabic (e.g., *kommst* 'come', *kurz* 'short') or disyllabic with stress on the first syllable (e.g., *Kiste* ‘box’, *Kajak* ‘kajak’). In German, this implies that the consonant is aspirated. Function words as well as content words were included. The place of articulation of the vowel following /k/ was classified as high (e.g., /i, u/) or low (e.g., /a, o/). Table 2 provides an overview of the material. As a result of the different recording situations, there was some imbalance in the material. A comparatively large part of the sample was produced by the L1 speakers, relatively more function words were selected from the L2 speakers, and the words produced by the 2L1 speakers were relatively more often only one syllable long. VOT was measured as the time interval between the release of the consonant and the onset of vocal-fold vibration. The boundaries of this interval were determined by visual inspection of the waveform and spectrogram within the speech editor Praat (Boersma &Weenink, 2013).

Table 2. Material overview (proportions within parentheses)

Group	tokens	function words	high vowel	disyllabic
Monolinguals	344	42 (0.12)	112 (0.33)	244 (0.71)
2L1s from France	141	32 (0.23)	33 (0.20)	36 (0.26)
2L1s from Germany	164	41 (0.25)	41 (0.29)	26 (0.16)
L2s in France	166	70 (0.42)	40 (0.24)	77 (0.46)
L2s in Germany	230	88 (0.32)	72 (0.31)	110 (0.48)

Results

Figure 1 shows VOT boxplots for each of the five groups. The overall longest VOT values were within the L1 group (mean VOT 76 ms). Within the two 2L1 groups, the speakers who grew up in Germany produced somewhat longer VOT (mean VOT 58 ms) than those who grew up in France (mean VOT 51 ms). The L2 speakers who had moved to Germany as adults produced longer VOT (mean VOT 56 ms) than those who lived in France (mean VOT 48 ms).

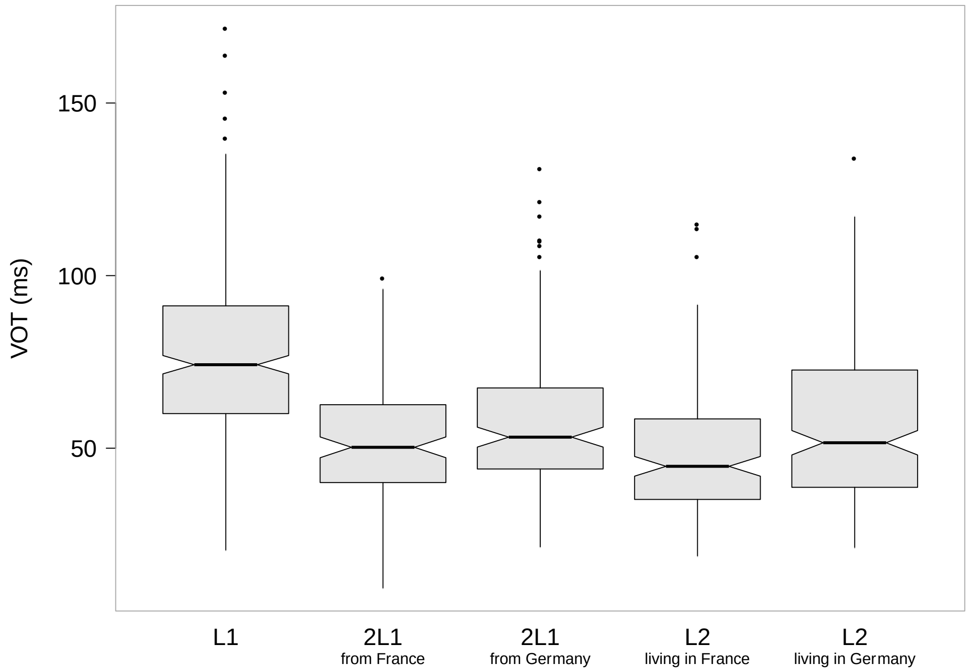


Figure 1. VOT across speaker groups

The differences between the groups were compared in a mixed-effects regression analysis. Four contrasts (planned comparisons) were formulated that together stood for the overall group effect and were of primary interest for the study, i.e. they were chosen to answer the four research questions described above. The two bilingual groups were contrasted against one another (contrast 1), and so were the two L2 speaker groups (contrast 2). Additionally, the L2 speakers were compared with the 2L1 speakers (contrast 3), and, finally, the L1 speakers were compared with all the four other groups (contrast 4). Contrasts 1 and 2, respectively, relate to the effects of the childhood country and of residence country. Contrast 3 relates to the effect of AoO. Finally, contrast 4 relates to an overall effect of being a monolingual native speaker of German or not. We also included possible effects of word length in syllables, vowel height and word type (function or content word), in order to control for the imbalance in the dataset described above.

Table 3. Regression output

	Estimate	**SE**	**df**	***t***	***p***
Intercept	50.545	2.695	155.5	18.756	0.000
2L1 from France versus 2L1 from Germany (contrast 1)	-8.899	5.558	31.5	-1.601	0.119
L2 in France versus L2 in Germany (contrast 2)	-5.971	5.483	29.8	-1.089	0.285
2L1 versus L2 (contrast 3)	3.949	3.938	31.6	1.003	0.324
L1 versus the rest (contrast 4)	-20.708	4.239	27.4	-4.885	0.000
vowelheight (low or high)	14.120	1.489	1023.8	9.482	0.000
word length in syllables (1 or 2)	2.475	1.377	1030.6	1.798	0.073
word type (content or function word)	-1.237	1.621	1032.9	-0.763	0.446

The results of the analysis are presented in Table 3. VOT in the L1 group was approximately 20 ms longer than that in the other four groups together. This was a significant difference. The estimated difference between the two bilingual groups was almost 9 ms but this difference was not significant. The difference between the two L2 speaker groups was almost 6 ms but this difference was not significant either. The difference between the L2 speakers and the bilinguals was almost 4 ms and non-significant. Finally, there was a rather large and significant effect of vowel height, in that high vowels elicited approximately 14 ms longer VOT than low vowels. The effect of word length was positive and marginally significant, and the effect of word type was small and not significant.

Discussion

Our first research question was whether L1 and 2L1 speakers produce different VOT ranges, although they all started acquiring German from birth. The average VOT of the stop consonants produced by the 2L1 speakers was significantly shorter (55 ms) than that of the L1 speakers (77 ms). At first sight, this seems to suggest that the 2L1 speakers did not receive sufficient German input during their childhood, since one of their parents did not speak German to them, and (for the 2L1 group from France) because German was not the language spoken in their childhood country. We note, however, that the average value of 55 ms is higher than the highest value produced for French VOT in the literature (cf. Lein et al, forthcoming), and that the value of 77 ms exceeds the highest value for German reported in the literature. For these reasons we do not believe that the observed values in the 2L1 speakers were unnaturally short, that is, being not “German-like”. Rather, we believe that the observed values for the L1 speakers are exceptionally high and should be seen as representative of carefully pronounced speech rather than of spontaneous speech.

Our second research question concerned the potential effect of the language spoken in the childhood country. The 2L1 speakers who grew up in Germany had somewhat longer VOTs than those who grew up in France, i.e. the heritage speakers of German. The estimated difference was not statistically significant, and we therefore cannot safely conclude that the language spoken in the childhood country affects the pronunciation of the heritage language. However, we see at least two reasons to be cautious in this conclusion. First of all, the estimated difference was approximately 9 ms which, even if not significant, is substantial. Second, the 2L1 speakers who had grown up in France, all had moved to Germany at the time they were recorded. It might very well be the case that these speakers' pronunciation would have been more French-like had they still lived in France. For these reasons we do not want to exclude the possibility that the majority language of their childhood environment affects pronunciation, and we think that this issue deserves further exploration in a future study, if possible with a larger speaker sample.

Our third research question was whether 2L1 speakers have more German-like VOT than L2 speakers, supposedly because they have been exposed to German from an earlier age. The estimated average VOT produced by the L2 speakers was not significantly shorter than that produced by the 2L1 speakers. In fact, the difference in average VOT between the two groups was only minimal. Assuming that the 2L1 speakers produced native-like VOT, this finding is remarkable since L2 speakers are widely reported to lag behind L1 speakers in pronunciation. As mentioned in the introduction, the L2 speakers who participated in this study were generally very proficient in German, and yet they scored well below the 2L1 speakers on the cloze test. We can conclude from this that in spite of their late AoO, L2 speakers can learn to produce VOT as close to monolingual-like ranges as those of 2L1 speakers at a comparable level of proficiency. We think this finding deserves further discussion. Differences between L2 and L1 speakers are typically explained with reference to AoO. However, if AoO is a crucial factor in the acquisition of phonology, then we expect a difference between L2 and 2L1 speakers, contrary to our results. L2 speakers in this study performed on a par with the 2L1 speakers, even in spite of the fact relatively low scores on the cloze test. So, if anything, the L2 speakers should be disadvantaged, but they are not. An alternative explanation, potentially invalidating our findings, is that we coincidentally selected speakers exceptionally good in pronunciation but poor in lexicon or morphology, as evidenced by their results on the cloze test. We consider this an unlikely scenario. Crucially, our results suggest that 2L1s are influenced by their second native language, as little or as much, as L2 speakers are influenced by their first language. Explanations that hinge on AoO alone can therefore not sufficiently explain the mechanisms at play in 2L1 and L2 acquisition. Differently put, both types of learners are capable of producing German-like VOTs, i.e. VOTs that differ noticeably from those typical of monolingual L1 French speakers.

The final research question was whether L2 speakers who lived in Germany produced more German-like VOT than those who lived in France. The L2 speakers who lived in Germany produced longer VOT than those who lived in France, but this difference was not statistically significant. We therefore cannot conclude that moving to the country where the second language is spoken necessarily has a positive effect on the pronunciation of the second language, but we see our results as an indication that it might. Alternative explanations are also possible. Speakers who move to the L2 country may be more motivated to learn the second language and therefore have better pronunciation than those who stay in their home country. Note also that the difference between the two L2 groups (contrast 2) was somewhat smaller than that between the 2L1 groups (contrast 1), though not reaching significance in either case. This may suggest that exposure to and use of the majority language, i.e. the language heard and spoken relatively often, plays a comparatively more important role during childhood than during adulthood.

Conclusion

We observed that the pronunciation of L2 speakers may be as good as that of 2L1 speakers in spite of the fact that the L2 speakers started to learn the language at a much later age. We acknowledge of course, that VOT is only one aspect of pronunciation among many others, but it is one that suits itself particularly well for the language pair that was studied here. Furthermore, we saw indications that the language spoken in the country where a speaker lives, either during childhood or during adulthood, may have an influence on the speaker's pronunciation, suggesting that speakers are able to adapt to the ambient language not only as children but also as adults. While the results show that AoO is not a crucial factor, the question about other potential factors, e.g., the exact length of residence in the country where the target language is spoken and current language use, remains open at the moment and will be addressed in the future.

Acknowledgments

We wish to thank Miriam Geiss and Luana D'Agosto for comments and their assistance in the data analysis.

References

Au, T. K., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, *13*, 238-243.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, *39*(3/4), 129-181.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program]. Version 5.3.52. Retrieved from www.praat.org/.

Braun, A. (1996). Zur regionalen Distribution von VOT im Deutschen. In A. Braun (ed.), *Untersuchungen zu Stimme und Sprache/Papers on Speech and Voice* (pp. 19-32). Stuttgart, Germany: Steiner.

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *Journal of the Acoustical Society of America*, *129*, 3964-3980.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303-306.

Kupisch, T. (2013). A new term for a better distinction? A view from the higher end of the proficiency scale. *Theoretical Linguistics*, *39* (3-4), 203-214.

Kupisch, T., Barton, D., Hailer, K., Kostogryz, E., Lein, T., Stangen, I., & van de Weijer, J. (2014). Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal*, *11* (2), 123-150.

Lein, T., Kupisch, T., & van de Weijer, J. (forthcoming). Voice onset time and global foreign accent in German-French simultaneous bilinguals during adulthood. *International Journal of Bilingualism*.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.

Oh, J., Jun, S., Knightly, L., & Au, T. (2003). Holding on to childhood language memory. *Cognition*, *86*, 53-64.

Pascual y Cabo, D., & Rothman, J. (2012). The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics*, *33*(4), 1-7.

Putnam, M., & Sanchez, L. (2013). What's so incomplete about incomplete acquisition? A prolegomenon to modeling heritage language grammars. *Linguistic Approaches to Bilingualism*, *3*, 478-508.

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, *13*(2), 155-163.

Segmental difficulties in French learners of German

Jane Wottawa[1], Martine Adda-Decker[1], Frédéric Isel[2]
jane.wottawa@univ-paris3.fr, martine.adda-decker@univ-paris3.fr, fisel@u-paris10.fr

[1]LPP, UMR 7018 CNRS - U. Paris 3/Sorbonne Nouvelle
[2]MoDyCo, UMR 7114 CNRS - Université Paris Ouest Nanterre La Défense

Abstract. The French Learners Audio Corpus of German Speech (FLACGS) was recorded to study the quantity and nature of French learners' pronunciation difficulties in German on a segmental level across three different tasks: word repetition, reading and picture description. The corpus was transcribed manually. The orthographic transcription was automatically aligned with the MAUS-web service. Among others, the data suggests that French learners of German have difficulties with vowel quantity contrasts as well as presence of /h/ onset on a segmental level. Duration is a valid cue to investigate vowel quantity as well as /h/ onset production. The French learners performed well across the tasks on vowel quantity distinction, except for the contrast /a/-/a:/ in the reading task. French learners of German produce identical durations for /a/-/a:/ that neither match the usual short vowel or the long vowel duration. French learners of German might only have one /a/-sound they can produce without any auditory input. That could explain why the duration for /a/-/a:/ is not clearly associated to the short-long vowel duration pattern. The quantity contrast between /a/-/a:/ was well performed by the French learners of German in the repetition task. That result suggests that the omitted contrast in the reading task is not due to erroneous perception. Regarding the /h/ onset, /h/ onset production decreases with higher production complexity in the task. At least three out of four possible /h/ onsets are produced as /h/ onset by French learners. The others are widely replaced by empty onsets, about 15% of the uttered words with /h/ onset, except for the reading task. In reading, French learners prefer a glottal stop to an empty onset. This result could be explained by decoding efforts. In reading and picture describing, French learners of German tend to produce longer /h/ onsets than German natives. The French learners may aim to be unambiguous by insisting on the first segment of the word. Across the tasks, French learners of German behave native-like for the vowel quantity contrast and the /h/ onset in the repetition task. This result suggests that French learners of German perceive vowel quantity and /h/ onsets well. The speech corpus is not a resource that allows us by itself to conclude whether the participants have achieved contrastive perception of the vowel quantity contrast or the /h/ onset, however.

Keywords: second language learning, speech corpus, segmental difficulties

Introduction

The pronunciation of a foreign language (L2) is conditioned by the phonological system of the mother tongue (L1). Mastering the phonological system of the L2 improves the communication with native speakers. Flege (1995) highlights that L2 speech production can be erroneous, and that the production skills of L2 learners do not only depend on perception skills in the L2.

In the literature, there are very few studies investigating French natives learning English or German. In 2014, Shoemaker investigated syllable boundaries perception in French learners of English. The author shows that in perception, French learners of English are more sensitive to the presence of glottal stops than to aspiration and that, by consequence, glottal stops are a more salient cue to syllable boundaries than aspiration. With respect to the intelligibility of L2 productions, native German listeners underwent a perception test of German vowels produced by French native speakers (Zimmerer & Trouvain, 2015a, b). The results showed that French learners' short vowels are perceived less well than their long vowels in minimal pairs by German native speakers. Another study by the same authors (2015) focussed on /h/ onset production of French learners in German and German native speakers in read speech. The researchers found that German native speakers have voiced and unvoiced /h/. French learners of German globally produce more unvoiced /h/ that the native speakers but they tend to produce only small amounts of empty onsets. Glottal stops, on the other hand, are more frequent especially for learners rated as beginners.

Studying vowel quantity contrast and /h/ onsets in German speech of French learners is part of a larger project in the framework of which we want to investigate whether detailed knowledge, awareness and practice of segmental and supra-segmental differences between the L1 and the L2 of a speaker help to improve his L2 pronunciation. To this purpose, a speech corpus was recorded to identify segmental and supra-segmental challenges for French learners of German in German speech.

Corpus definition, collection and content

Participants

For the *French Learners Audio Corpus of German Speech* (FLACGS), all participants were recruited in Paris, France. Participation was on a voluntary basis. In return, they received an incentive USB key and pronunciation feedback. The recordings took approximately 45 minutes per participant.

French learners of German (FG)

20 FG (10 women and 10 men) were recorded in Paris (France). The women were aged between 20 and 30 years, the men between 24 and 32. All FG as well as their parents had only French as a first language (L1). They auto-evaluated their German skills based on the Common European Framework of Reference for Languages (CEFRL). In both gender groups, all levels were represented: A1/A2 up to C2. Moreover, all participants learned English at school.

German native speakers (GG)

20 GG (10 women and 10 men) were recorded in Paris (France) except for two of them who were recorded in Germany. The women were aged between 22 and 47 years, the men between 30 and 45. All GG as well as their parents had only German as L1. Except for one female and one male participant who had no knowledge of the French language, all GG were highly proficient in French (B1/B2 up to C2+ according to the CEFRL). They all learned English at school. The great majority lived in France for several years.

Even if the GG were born and raised in different regions of Germany, all spoke in standard German for the recordings.

Tasks

The participants performed three tasks of increasing production complexity:

1. <u>Repetition task</u>
   Participants heard small sentences over headphones they repeated immediately.
2. <u>Reading task</u>
   Participants read aloud the short stories *Nordwind und Sonne* and *Die Buttergeschichte*.
3. <u>Picture description</u>
   The picture description task was the only task without linguistic input (Figure 1).

The repetition task aimed to investigate how FG produce long and short vowel contrasts, consonants and consonant clusters that are unusual or different in the French language as well as lexical stress in different word positions.

Carrier sentences (*Er sagt ... klar und deutlich* and *Ich sage ... klar und deutlich*) including 55 distinct words in central accented position were recorded by a female native German speaker. The participants listened to all the spoken utterances in a randomized order over headphones and repeated them.

Figure 1. Picture description task

The material of the repetition task was composed of words with lexical stress in different positions (word-initial syllable, word-final syllable, penultimate syllable and ante-penultimate syllable), minimal pairs with long and short vowels (e.g., *Hüte* /hy:tə/ and *Hütte* /hʏtə/) minimal pairs with voiced/unvoiced distinction in plosive (e.g., *glauben* /glaʊ̯bən/ and *klauben* /klaʊ̯bən/). We also included words that are difficult to pronounce because of their phonotactics (for French natives, challenging consonants, clusters and glottal stops between vowels in adjacent syllables) e.g., *Schächtelchen* /ˈʃɛçtəlçən/ and *erobernde* /ɛɐʔˈɔbɐndɛ/. The aim of the repetition task was to check FGs' reproduction skills of utterances in L2.

The participants had to read two short stories *Nordwind und Sonne* and *Die Buttergeschichte* frequently used in phonetic studies (e.g., in the Kiel corpus, Kohler, 1996). Conflicting orthographic conventions between L1 and L2 are possible sources of pronunciation difficulties. For instance the graphic <z> is pronounced /z/ in French but /ts/ in German. Another example is the graphic sequence <au> which is pronounced as the diphthong /aʊ̯/ in German, but /o/ in French. The aim of the reading task is twofold:

(i) check overall FG pronunciation difficulties, when reading;
(ii) focus on difficulties which may arise due to conflicting orthographic conventions between German and French.

Table 1. Summary of the FLACGS-corpus

Name	*French Learners Audio Corpus of German Speech* (FLACGS)
Language	German
Speakers	40 speakers (20 male and 20 female) • 20 L1 German • 20 L1 French, L2 German (Level of competence: A2-C2)
Volume	ca. 30 h of speech (ca. 15h & ca. 15h) (35 250 words)
Content	• repeated (55 words) • read (347 words) sem • semi-spontaneous speech (49 - 239 words)
Transcription	manual using the German orthography
Alignment	MAUS-webservice (automatic) and manual checking

Both languages, French and German, use the Latin alphabet. But the letters and letter combinations do not necessarily code the same sounds e.g., *Mantel* is produced as [ˈmantəl] by GG. FG are more likely to pronounce [mãˈtɛl] as the letter combination <an> corresponds to a nasal vowel in written French. The reading task also allows us to compare prosodic patterns in different places of the utterance, e.g., to compare how word stress is realized in the beginning, the middle and at the end of an utterance with respect to prosody.

The description task aims to collect semi-spontaneous speech. All participants described the same picture. We concentrated our analysis on uttered words like *Haus*, *Mädchen*, *Junge* and *Sonne*. The image description is the only task where the participants did not have any linguistic support (prior written sentences, spoken utterances) to help them with their speech production. Before the participants started the image description, we made sure they knew the names of the items and actions represented in the picture by asking them to name all items.

Methods

Transcription and alignment

First, manual checking and potential corrections of the orthographical text of both the repeated and read material was carried out, if necessary. A manual orthographical transcription of the spontaneous speech part (description task) was realized. Transcriptions included spontaneous speech specific events, such as hesitations, disfluencies and false starts.

Second, the MAUS-webservice (Munich AUtomatic Segmentation - web service, Schiel, (1999); Kisler, Schiel, and Sloetjes (2012); https://clarin.phonetik.uni-muenchen.de/BASWebServices /#/services) performed the alignment of the speech signal with its transcription. The MAUS aligner generates a TextGrid-file that can be opened with Praat (Boersma & Weenink, 2001). Transcription checking is almost real-time whereas the manual transcription of the spontaneous speech took about 3 minutes for 1 minute of recordings. MAUS needs these orthographic transcriptions to segment the speech signal.

Third the TextGrid files generated by MAUS were checked. They comprise three tiers corresponding to three segmentation or annotation levels: the orthographic word, the canonical pronunciation of the word (phonemic level) and the aligned phones (phonetic level). The automatic alignment of each sound-file was checked manually for boundaries and labelling. Phone boundaries of targeted words were manually corrected if necessary. We also checked some pronunciations, for instance when MAUS had to perform a grapheme-to-phoneme conversion for words that were not included in its dictionary. Performing those adjustments took about 2min for 1min of automatically aligned speech.

Acoustic analysis

Acoustic parameters were measured using Praat scripts. The first four formants, energy, intensity and voicing rate were extracted from the sound files for each phoneme. The TextGrid file provides information about segmental durations. The short and long vowel contrast as well as the presence of /h/ onsets only required information on segment duration that could be extracted from the TextGrids.

Results

Short and long vowels

German natives produce the phonologically short and long vowels in minimal pairs by acoustic duration (and vocalic timbre, i.e. formants differences). In French, this duration contrast is absent. We want to investigate whether GG speakers make duration distinctions for all vowel pairs and whether duration distinction is better for some vowel pairs than others. We are also interested in knowing whether FG speakers are able to make duration distinctions and what the impact of the different tasks would be.

Duration contrast of vowels across tasks

In Figure 2, mean vowel duration in milliseconds for the short-long vowel pairs /ɪ/-/i:/, /ʏ/-/y:/, /a/-/a:/ and /ɔ/-/o:/ for both participants groups are plotted.

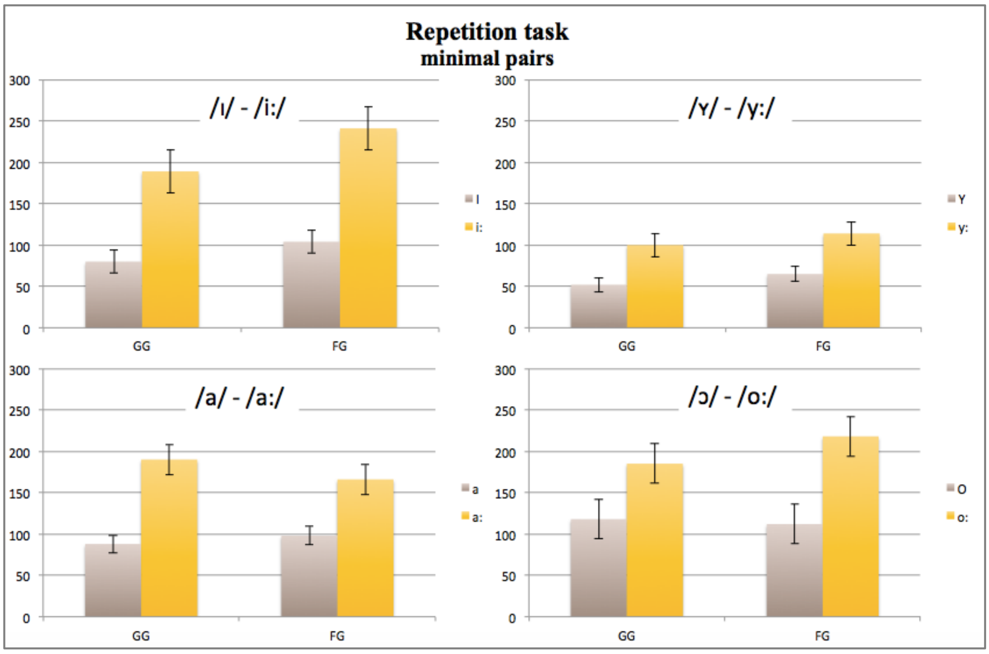


Figure 2. Vowel duration in stressed word positions in milliseconds, repetition task

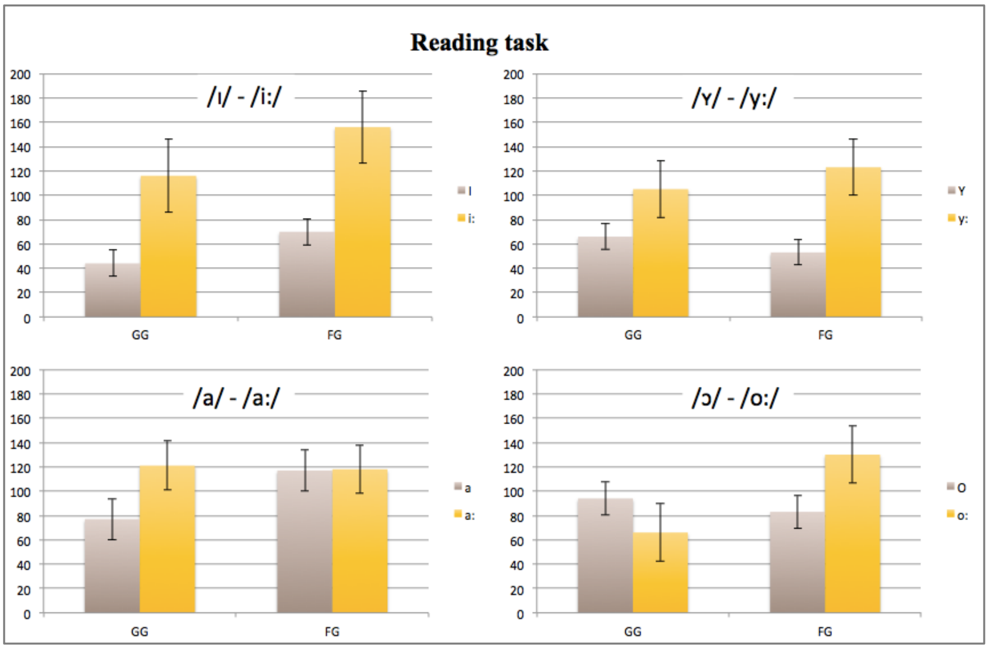


Figure 3. Vowel duration in stressed word positions in milliseconds, reading task

In the repetition task, the FG are quite successful in imitating GG speakers' duration oppositions. FG generally produce vowels with longer durations compared to the GG vowels. The central open long vowel /a:/ and the half-open back vowel /ɔ/ are exceptions to this. In both participants groups, GG and FG, statistically significant duration differences are made for all short-long vowel pairs that occurred during the repetition task.

In the reading task (e.g., Figure 3), statistically significant differences are made by the FG for all long-short vowel pairs except the /a/-/aː/ contrast. The /ɔ/-/oː/ contrast regarding the duration pattern is better performed by the FG during reading.

/h/ onset

In German, /h/ onsets exist frequently, and minimal pairs between /h/ onsets and /ʔ/ onsets also exist: *Haus* /haʊ̯s/ *versus aus* /ʔaʊ̯s/. The French language does not have a phonological /h/. That is why they tend to omit /h/-onsets in foreign languages they might learn.

We predict that FG will replace /h/ onsets with empty onsets or with /ʔ/ onsets. We also think that /h/ onsets produced by the FG should have a shorter duration than those produced by GG. And finally, global /h/ onset production should decrease as the production tasks gets more complex.

/h/-onset across tasks

Table 2 recapitulates all possible /h/ onsets and their realization by FG. First, we observe that /h/ onset production decreases with increasing complexity of the production task. Still, a surprisingly high number of /h/ onsets are actually produced by FG: at least three out of four for the most complex task, the picture description.

Furthermore, Table 2 confirms that /h/ onsets are more likely to be replaced by empty onsets than glottal stops except for the reading task where FG produced one out of five /h/ onsets as a glottal stop.

	Repetition task	**Reading task**	**Picture description**
[h] onset	85%	78%	75%
[ʔ] onset	1%	20%	9%
Empty onset	14%	2%	16%
Tokens	**77**	**104**	**71**

Table 2. /h/ onset realisations ([h], [ʔ], empty) across the three different tasks in FG speakers

Figures 4, 5 and 6 present the produced /h/ onsets by the FG and their duration in milliseconds. For the repetition task again, FG behave like GG when they produce the /h/ onset. There is no statistical difference of /h/ onset duration between GG and FG (e.g., Figure 4).

In the reading task, represented in Figure 5, FG produce significant longer /h/ onsets for most vowel contexts except in the right context of rounded vowels.

Regarding the picture description, /h/ onsets produced by FG are globally longer than those produced by GG. The word *Hunden* is an exception to that trend. *Hunden* presents a complex morphology: stem+plural+dative. FG who use such complex words are very proficient in German. They are more likely to adjust their production to the native model.

Discussion

Short and long vowels

For the short and long vowel opposition, GG and FG make a duration opposition in both repetition and reading task. The picture description was not included in the analysis because not enough words with long and short vowels in stressed positions were produced. This result shows that FG are sensitive to duration variations in vowels, even if this contrast has no phonological value in French.

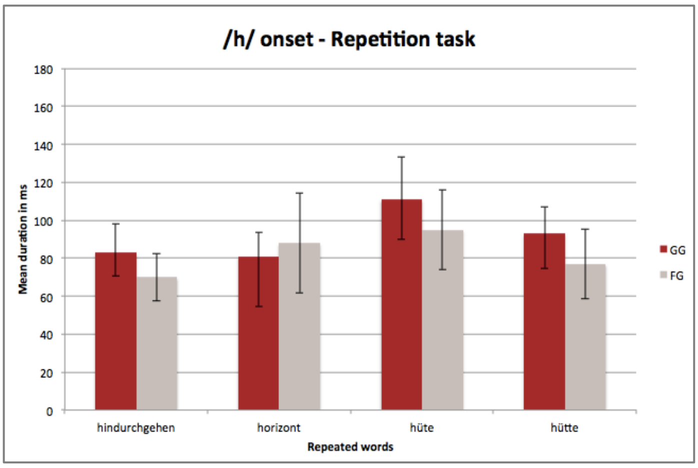


Figure 4. Duration (in ms) of /h/ onset, repetition task

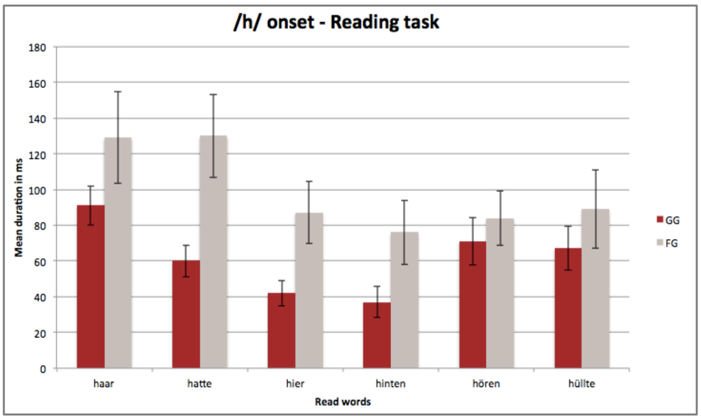


Figure 5. Duration (in ms) of /h/ onset, reading task

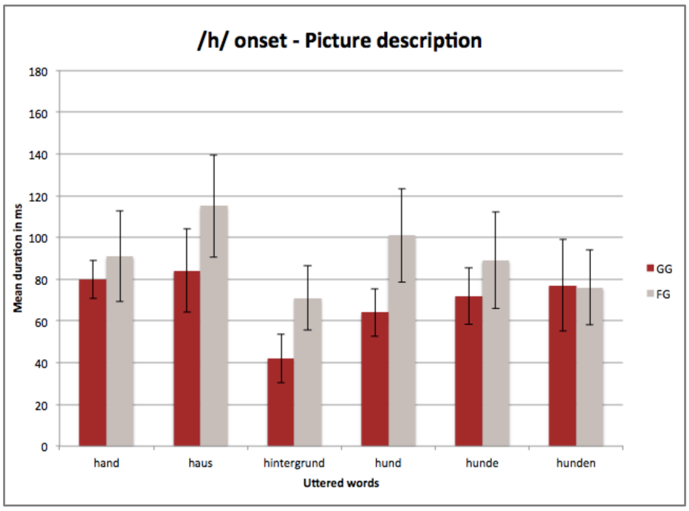


Figure 6. Duration (in ms) of /h/ onset, picture description

In the repetition task, FG behave native-like in contrasting minimal pairs. This result suggests that FG can perceive vowel duration and can repeat the pattern in their oral production. However, being able to reproduce vowel quantity patterns does not automatically mean that FG have contrastive perception of vowel quantity. Especially for minimal pair production, a great number of participants thought they produced the same word twice.

Regarding the reading task, FG realized vowel quantity surprisingly well. That could be due to orthographic cues. In German orthography short vowels are often followed by a double consonant e.g., *so<u>ll</u>te* and long vowels are often followed by a graphic <h> (Dehnungs-h) e.g., *frü<u>h</u>*. On one hand, the /ɔ/ - /o:/ contrast regarding the duration pattern is better performed by the FG than the GG. This vowel contrast does exist in French, which could explain the better performance of the FG. It is also possible that FG overgeneralized the short-long vowel pattern, as lax vowels tend to be shorter than tense vowels. Vowel duration can be influenced by the words' sentence position. This criterion could explain why the production of GGs' short vowel /ɔ/ is longer than their production of the long vowel /o:/. On the other hand, FG do not produce any difference in vowel quantity for the vowel pair /a/-/a:/. Both vowels match the mean duration of GGs' /a:/. Compared to the other three vowel pairs, in the reading task, FG produced the /a/-/a:/ pair shorter than the other long vowels but longer than the other short vowels. French learners of German might only have one /a/-sound they can produce without any auditory input. That could explain why the duration for /a/-/a:/ is not clearly associated to the short-long vowel duration pattern.

/h/ onset

At least three out of four /h/ onsets are produced by FG. If /h/ onsets are not produced, they are mostly replaced by empty onsets except in reading. In reading, almost all unrealized /h/ onsets are replaced by glottal stops rather than empty onsets. The glottal stop can be due to cognitive efforts towards orthographic decoding. First, the graphic representation <h> could trigger an onset production instead of leaving it empty. Second, as the /h/ phoneme does not exist in French. FG speakers may put a lot of effort to produce this glottal fricative. The production effort, if not successful, could result in a glottal stop. Both explanations relate to the conflicting orthographic conventions with respect to <h>, which is known to be pronounced /h/ in German but not in French.

Empty onsets that are produced in the repetition task and during the picture description concern about 15% of the uttered words with an expected /h/ onset. This result suggests that the production of empty onsets instead of /h/ onsets is not linked to orthography. Both, the repetition task and the picture description furnished no written input at all.

When producing /h/ onsets, FG speakers tend to emphasize their durations as compared to those produced by GG in both the reading and the picture description tasks.

Conclusions

A corpus of non-native German speech by French natives was recorded. Participants had to perform three different tasks of increasing production complexity. Our results show that segmental difficulties are task-related.

FG are able to produce duration contrasts that are not phonological in French in the repetition and the reading task, except for the /a/-/a:/ pair in reading. The picture description was not taken into account for the vowel quantity contrast because of its limited number of tokens with short and long vowels. FGs' ability of vowel production does not allow us to conclude about their contrastive vowel perception. To investigate whether FG contrastively distinguish between short and long vowels a perception test has to be performed.

Regarding /h/ onsets, a surprisingly high number of /h/ onsets is produced as actual [h] onsets across all three tasks. Except in the reading task, /h/ onsets are more likely to be replaced by empty onsets than by glottal stops. FG tend to exaggerate /h/ onset durations. This information indicates that FG are well aware of the difficulties they have in producing /h/ onsets.

Across the tasks, French learners of German behave native-like for the vowel quantity contrast and the /h/ onset in the repetition task. This result suggests that French learners of German perceive vowel quantity and /h/ onsets well. The *French Learners Audio Corpus of German Speech* is not a resource that by itself allows us to conclude whether the participants have achieved contrastive perception of the vowel quantity contrast or the /h/ onset.

Acknowledgments

This work was made possible through a Sorbonne Nouvelle University PhD funding to the first author. It was also supported by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

References

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer [computer program]. Retrieved from www.praat.org/.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Kisler, T., Schiel, F., Sloetjes, H. (2012). Signal processing via web services: the use case WebMAUS. In *Proceedings Digital Humanities* 2012 (pp. 30-34). Hamburg, Germany.

Kohler, K. J. (1996). Labelled data bank of spoken standard German: the Kiel corpus of read/spontaneous speech. In *Proceedings of Fourth International Conference on Spoken Language ICSLP 96* (Vol. 3, pp. 1938-1941). IEEE.

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In *Proceedings of the ICPhS* (pp. 607-610).

Shoemaker, E. (2014). The exploitation of subphonemic acoustic detail in L2 speech segmentation. *Studies in Second Language Acquisition, 36*(04), 709-731.

Zimmerer, F., & Trouvain, J. (2015a). Perception of French speakers' German vowels. In *Proceedings of the Sixteenth Annual Conference of the International Speech Communication Association*.

Zimmerer, F., & Trouvain, J. (2015b). Productions of /h/ in German: French vs. German speakers. In *Proceedings of the Sixteenth Annual Conference of the International Speech Communication Association*.

Phonetic and phonological acquisition in Persian speaking children

Talieh Zarifian[1], Yahya Modarresi[2], Laya Gholami Tehrani[1], Mehdi Dastjerdi Kazemi[3]
t.zarifian@uswr.ac.ir, modarresi@ihcs.ac.ir, lgtehrani@uswr.ac.ir, dastjerdi@riec.ac.ir

[1]University of Social Welfare and Rehabilitation Sciences, [2] Institute of Humanities and Cultural Studies, [3]Institute for Exceptional children, Research Institute of Education

Abstract. _Purpose:_ This study aimed to answer 3 main questions in Persian speaking children: (1) What are the ages for normative acquisition and mastery for each phoneme? (2) What is the percentage of consonant, vowel and phoneme correct? (3)What phonological processes can be seen? _Methods_:The samples were gathered from 387 children aged between 3-6 years old using a 27 singleword picture-naming articulation test for the consonant acquisition study and 54 single word picture naming phonological test for the phonology study. _Results_: Findings revealed that all participants acquired all 23consonants and six vowels and two diphthongs by age 3;0 based on the 75% criterion in two positions (syllable initial and syllable final) and mastered all Persian phonemes by age 3;6, except /s/,/z/, /ʒ/, and /r/ at the 90% criterion in two positions (syllable initial and syllable final). /ʒ/ and /r/ were mastered by age 3;11; /s/ and /z/ were mastered by age 4;6. By age 6;00, children produced 94.57% of consonants correct, the percent of vowel correct was 99.8% of vowels correct and the percent of phoneme correct was 96.3% of phonemes correct. By age 3;00, syllable deletion, consonant and vowel assimilation had disappeared. Between ages 3-4, there was a major decline in the following processes: gliding, affrication, deaffrication, prevocalic voicing, vowel substitution, metathesis, stopping. Between ages 4-5, the following processes were declining: final consonant deletion and fronting. _Practical implications_: The following phonological processes were attributed to atypical production because they weren't found in any group of children at more than 10%: backing, initial consonant deletion, insertion, sound preference, gemination, degemination, nasalization, denasalization, and deletion of more than two syllables. These findings seem to provide useful information for speech-language pathologists for assessing Persian speaking children and designing treatment objectives in Persian.

Keywords: acquisition, Persian, phonological process, percent correct, consonant, vowel

Introduction

Nowadays there is real interest in knowing about phonetic and phonological development in languages other than English. The focus of this study is on Persian. The Persian language (also known as Farsi) is a member of the Western Iranian branch of the Indo-Iranian family within the Indo-European language family (Keshavarz & Ingram, 2002). It is the official language of Iran, Afghanistan and Tajikistan, and it is also widely spoken in some other countries, such as India, Bahrain, and among immigrants in Europe, US, and the Pacific countries. There are various accents of the Persian language spoken in Iran and in other Persian speaking countries.

The phonetic and phonological system in Persian

The present study focused on the Persian language spoken widely in Tehran. The sound system of the Persian language (sometimes known as formal Persian) is discussed briefly in the following section.

Persian syllable structure

There are three syllable structure patterns (i.e. cv, cvc, cvcc). Persian syllables can't be initiated with vowels, and when a word starts with a vowel, it includes a glottal /ʔ/ as the syllable onset (e.g., 'ʔɑb' 'water'). Persian syllable structure only permits word-final clusters, while initial clusters occur in loan words. Tables 1 and 2 present the syllable structure and cluster patterns in Persian.

Table 1. Persian syllable structures

Persian syllable	example
cv	/mu/ 'hair'
cvc	/tup/ 'ball'
cvcc	/kæfʃ/ 'shoe'

Table 2. Cluster patterns in Persian

cluster pattern	example
stop-stop	/ʤoGd/ 'owl'
fricative-fricative	/ kæfʃ/ 'shoe'
fricative-stop	/ʔæsb/ 'horse'
stop-fricative	/sæbz/ 'green'
l clusters(c-l or l-c):	
l+ b, t, d, k, G, ʔ	/Gælb/ 'heart'
l + f, v, s, z, x, h	/tælx/ 'bitter'
l + m	/zolm/ 'injustice'
r clusters(c-r or r-c):	
r + b, t, d, k, g, G, ʔ	/morG/ 'hen'
r + x, s, ʃ, f, z, v	/færʃ/ 'carpet'
r + ʧ, ʤ	/Gɑrʧ/ 'mushroom'
r + m, n	/gærm/ 'warm'
J clusters(j- c):	
j+ d, t, b, k, ʔ	/kejk/ 'cake'
j + s, z, ʃ, f	/ʔejʃ/ 'luxury'
j + m, n	/bejn/ 'between'
j + l, r	/sejl/ 'flood'
m clusters(m-c or c-m):	
m + t, d, ʔ, G, n, p	/lɑmp/ 'light'
m + s, z, ʃ	/læms/ 'to touch'
m + l, r	/hæml/ 'to carry'
n clusters:	
n + b/ d/ g/ ʔ	/bænd/ 'strap'
n + ʧ /ʤ	/konʤ/ 'corner'

Persian sound system

The Persian language consists of 23 consonants, six vowels and two diphthongs(Bijankhan, 2006; Hall, 2007; Keshavarz & Ingram, 2002; Samarah, 1977; Yarmohammadi, 1965). The Persian phonemes inventory includes all English stops and affricates, plus /G/ and /ʔ/. In formal Persian, final /ʔ/ is usually deleted. Comparing with English, Persian does not have the velarized nasal /ŋ/. While there are seven fricatives in English and Persian (i.e. /f/, /s/, /ʃ/, /h/, /v/, /z/, and/ʒ/), there are some differences in the fricative system of the two languages with English having two dental-fricatives /θ/ and /ð/, and Persian having the uvular fricative /x/. Additionally, both languages have similar liquids and glides on a phonemic level (i.e. /j/, /w/, /r/, /l/). However, the phoneme /r/ in the two languages is not phonetically the same. The Persian /r/ is a trill rather than a liquid approximant as in English

(Keshavarz & Ingram, 2002). Tables 3 and 4 present consonants, vowels and diphthongs in official Persian, respectively.

Table 3. The Persian consonantal system in IPA

Place of Articulation	**bi-labial**	**labio-dental**	**dental-alveolar**	**alveolar**	**alveolo-palatal**	**palatal**	**velar**	**uvular**	**glottal**
Manner of Articulation									
Stop	p b		t d				k g	G	ʔ
Fricative		f v		s z	ʃ ʒ			χ	h
Affricate					ʧ ʤ				
Glide						j			
Lateral				l					
Trill				r					
Nasal	m			n					

Table 4. The Persian vowel system

Place of Articulation	Vowels		Diphtongs	
	front	back		
high	i	u	ei	oʊ
mid	e	o		
low	æ	ɑ		

Development of phonological processes/error patterns

Although the occurance of phonological processes/error patterns described in three longitudinal case studies(Fahim, 1995; Meshkatoddini, 1989; Nourbakhsh, 2002) and five observational cross-sectional studies(Damirchi, 2010; Derakhsande, 1997; Ghassisin, 2006; Reza Pour, Tahbaz, & Mehri, 1999; Shirazi, Mehdipour-Shahrivar, Mehri, & Rahgozar, 2010), there is little information about determining the age, or age range, at which the various error patterns appeared or disappeared in the speech of normally developing children. Table 4 presents the most common error patterns in the speech of Persian speaking children based on information in the afore-mentioned studies.

Table 5. Most common phonological processes/error patterns in Persian speaking children

Derakhshandeh, (1997)	**Reza Pour et al., (1999)**	**Ghassisin (2006)**	**Shirazi, et al., (2010)**	**Damirchi (2010)**
N:56 (male/female)	**N**:100 (male/female)	**N**:60 (male/female)	**N**:128 (male/female)	**N**:96 (male/female)
Age ranges: (3;9-3;11 & 4;9-4;11)	**Age range**: (2;0-3;6)	**Age range**: (2;0-3;11)	**Age range**: (2;0-3;11)	**Age range**: (2;0-5;11)
Most common error patterns				
gliding, cluster reduction, fronting, stopping, fricating, final consonant deletion, syllable deletion, voicing, lateralization, affrication, assimilation	gliding, cluster reduction, fronting, final consonant deletion, initial & medial consonant deletion, voicing, metathesis, assimilation	gliding, cluster reduction, fronting, final consonant deletion, initial & medial consonant deletion, voicing, metahtesis, assimilation	gliding, cluster reduction, fronting, final consonant deletion, initial & medial consonant deletion, voicing, metahtesis, assimilation	gliding, cluster reduction, deaffrication, final consonant deletion, initial consonant deletion, stopping, backing assimilation

The results of these studies are useful but not comprehensive. The way that the reseachers determined error patterns is controversial. The criteria for determining errors on single error occurrence can warrant the existence of a particular error pattern, but it is questionable and it needs to make a distinction between one instance of an error, which may take place by chance or occur due to developmental fluctuation, and the frequent occurrence of an error type that represents a certain tendency in a child's speech (Dodd, Holm, Hua, & Crosbie, 2003).

The current study

This study aims to answer three main questions with regard to Persian speaking children's developing phonologies: (1) What are the ages for normative acquisition and mastery for each phoneme? (2) What is the percentage of consonant, vowel and phoneme correct? (3)What phonological processes can be seen in Persian speaking children between the ages 3-6?

Method

Subjects

The samples were gathered from 387 children aged between 3-6 years old, using a 27 singleword picture-naming articulation test for the consonant acquisition study and 54 single word picture naming phonological test for the phonology study.

A total of 387 children (191 boys and 196 girls, aged 36-72 months, M(SD):53.7($\pm$10.1), attending 12 nurseries and kindergartens in Tehran were recruited after obtaining their parents/guardians consents following ethics approval from the Medical Ethics Committee for the University of Social Welfare and Rehabilitation Sciences.

Only monolingual Persian speaking children with no background on speech therapy were included. Children were selected through a simple convenience sampling. The exclusion criteria were structural deficits (e.g., cleft palate), permanent hearing loss, Persian as a second language at home, autism spectrum disorder and dysarthria. These were determined by the child's medical record history, a clinical examination by an experienced Speech Language Pathologist and reports from parents. Participants were tested in a quiet place for the duration of the articulation test, attempt imitation, and tolerate cuing (Dodd et al., 2003; Holm, Crosbie, & Dodd, 2007). Table 6 reports demographic data on the participants.

Table6: Age (in months) and gender of the participating children

Age group (month)	number	Mean age (month/day)	SD	percent
35-42	60	39/3	1.89	15.50
43-48	82	45/6	1.51	21.18
49-54	60	51/3	1.64	15.50
55-60	68	57/3	1.67	17.57
61-66	61	63/5	1.73	15.76
67-72	56	69/3	2.18	14.47
Total	387	53/7	10.09	100

Note: SD= standard deviation

Materials

A 27 singleword picture-naming articulation test for the consonant acquisition study and 54 single word picture naming phonological test for the phonology study (Zarifian, Tehrani, Modaresi, Dastjerdi- Kazemi, & Salavati, 2014b) were used to assess the children’s speech abilities.

Data analysis

A broad phonetic transcription was made online after production of any words. Further, all testing procedure was audio-video recorded. Audio-video recordings were made through an assessment procedure to allow the revision of transcription and transcription reliability measurements. The Gold Wave software was used for detailed analysis and refining the audio recordings. The examiners reviewed each transcription with reference to its audio-video recording to ensure the accuracy of online transcriptions.

All utterances were audiotaped and immediately transcribed by the researcher using the International Phonetic Alphabet (IPA) revised 2005. The utterances were transcribed again later from audio-recordings to check the original transcription, and a second rater (a native Persian phonetician) additionally transcribed 13.4% of the data. The inter-rater-reliability agreement was at 96.5%.

The metrical was analyzed to provide normative data on the acquisition of phonetic and phonemic inventories and phonological process use in Persian-speaking children. The criteria set for each sub-analysis are described in the following section.

Results

Phonetic inventory

Results revealed that all participants acquired all 23 consonants, six vowels and two diphthongs by age 3;0 based on the 75% criterion in two positions (syllable initial and syllable final), and mastered all Persian phonemes by age 3;6, except /s/, /z/, /ʒ/, and /r/ with at the 90% criterion in two positions (syllable initial and syllable final). /ʒ/ and /r/ were mastered by age 3;11 while /s/ and /z/ were mastered by age 4;6. Table 7 presents this information. The age of acquisition and mastery were calculated based on the Amayreh and Dyson method (Amayreh & Dyson, 1998).

Table 7. Phonetic acquisition according to75% and 90% criteria

Age group	age	75%	90%
1	3;0- 3;5	p, b, t, d, ʔ, m, n, f, v, ʃ, h, x ʧ, ʤ, l, j, k, g, G, ʒ, r, s, z	p, b, t, d, ʔ, m, n, f, v, ʃ, h, x, ʧ, ʤ, l, j, k, g, G
2	3;6- 3;11		ʒ, r
3	4;0- 4;5		s, z
4	4;6-4;11		
5	5;0- 5;5		
6	5;6- 5;11		

Vowels

The investigation of the vowel system indicated that all participants had acquired and mastered vowels before age 3;0.

The percentage of consonant, vowel and phoneme correct

Three quantitative measures were used to calculate the percentage of consonant, vowel and phoneme correct (PCC, PVC and PPC, respectively). For calculating the percentage consonants correct (PCC), the percentage of consonants pronounced correctly was divided by the total number of consonants elicited in the Phonological Assessment (Zarifian, Tehrani, Modaresi, Dastjerdi -Kazemi, & Salavati, 2013). For the percentage vowels correct (PVC), the percentage of vowels pronounced correctly was divided by the total number of vowels elicited in the Phonological Assessment (Zarifian, Tehrani, Modaresi, Dastjerdi- Kazemi, & Salavati, 2014a, 2014b) and finally for the percentage phonemes correct (PPC) (Zarifian et al., 2013), the percentage of phonemes pronounced correctly were divided by the total number of phonemes elicited in the Phonological Assessment. By the age of 6, children produced: 94.57% of consonants correct, 99.8% of vowels correct, and 96.3% of phonemes correct.

Developmental phonological processes

Phonological error patterns are defined as consitent differences between child and adult realisations of the target words; they are categorized at two levels: syllable error patterns and substitution error patterns. There is a general tendency that error patterns affect a group of sounds (Bankson & Bernthal, 1998; Dodd et al., 2003; Grunwell, 1987; Ingram, 1981). The criteria for assigning an error pattern as age appropriate was that more than 10% of children in an age group had to exhibit the error pattern at least twice for gliding, affrication, deaffrication, prevocalic voicing, lateralization, backing, nasalization, denasalization, germination, degemination, thrilling, vowel substitution, methatesis, initial consonant deletion, consonant/vowel assimilation/ harmony, weak syllable deletion, insertion, consonant, and four times for stopping, fronting, final devoicing, cluster reduction, final consonant deletion. Table 8 lists the chronology of phonological processes in the Persian speaking children.

By the age of 3, syllable deletion, consonant and vowel assimilation had disappeared. Between ages 3-4, there was a major decline in the following processes: gliding, affrication, deaffrication, prevocalic voicing, vowel substitution, metathesis, stopping. Between ages 4-5, the following processes were declining: final consonant deletion and fronting. The following phonological processes were attributed to atypical production because they weren’t found in any group of children at more than 10%: backing, initial consonant deletion, insertion, sound preference, gemination, degemination, nasalization, denasalization, and deletion of more than two syllables.

Uncommon processes are listed and explained below: backing, final consonant voicing, thrilling, germination, degemination, sound preference, nasalization, denasalization, insertion, deletion more than one syllable, initial consonant deletion, voicing/devoicing.

Table 8. Developmental phonological processes in Persian-speaking children

process		3;0-3;5	3;6-3;11	4;0-4;5	4;6-4;11	5;0-5;5	5;6-5;11
Substitution process	gliding						
	affrication						
	deaffrication						
	stopping						
	Prevocalic voicing						
	fronting						
	final devoicing						
Syllable error patterns	cluster reduction						
	Final consonant deletion						
	metathesis						
	consonant harmony						
	vowel harmony						

Discussion

The phonetic and phonemic acquisition and the percentage of phoneme accuracy were studied in the speech sample of 387 Iranian Persian-speaking children, aged between between 3;0 and 5;11 years. The results showed that as age increased, the phonetic and phonological skills developed. In this study, two aspects of speech development were considered: the age of acquisition of sounds (phonetic acquisition) and the ages that error patterns became evident and disappeared (phonemic acquisition). Analyzing the gathered data showed that phonological skills would develop with age. Children's speech becomes more accurate as they get older. They articulate more sounds correctly and use fewer error patterns. Analyzing performance in six monthly age bands revealed a gradual progression of speech accuracy. Significant differences were identified between groups of children aged 3;0-3;11 years, 4;0-5;5 years, and 5;6-5;11 years on the percentage of consonant correct (PCC). Differences were found between three age groups of children aged 3;0-3;5 years, 3;5-5;5 years and 5;5-5;11 on the percentage of vowel correct (PVC). Based on the percentage of phoneme correct (PPC) we again have three age groups: 3;0-4;5 years, 4;5-5;5 years and 5;5- 5;11, that were produced correctly. Accuracy increased with age. The acquisition of vowels is assumed to be complete by age three, therefore it is not assessed explicitly in this study. The sequence of sound acquisition reported in this study was consistent with English-speaking studies: /p, b, t, d, ʔ, m, n, f, v, ʃ, h, x,ʧ, ʤ, l, j, k, g, G/ were among the first sounds acquired, while /s, z,/ were the last sounds acquired and mastered. The age of acquisition for sounds was similar to Dodd et al. (2003) with the exception of /ʃ, ʤ, r, l/. They used a 90% accuracy criterion (the child had to produce the sound accurately at least 90% of the time) but it is unclear what proportion of children in an age band had to have 90% accuracy for an age of acquisition to be assigned to a sound. The current study implemented a phonetic approach. The researchers included a sound in a child's inventory if it was produced spontaneously or or in imitation.

When children are first exposed to a word they may imitate it correctly (e.g., Guri 'teapot') once the word is a lexical item, they may then go on to use a system-level sound substitution (e.g., Guri). The word is pronounced as [guli] by a five year old child. Error patterns decreased with age. Ninety percent of the assessed children over five years of age had error-free speech. Voicing had resolved by 3;0 years; stopping by 3;6; weak syllable deletion and fronting by 4;0 years. Deaffrication and cluster reduction were resolved by 5;5 years. Liquid gliding persisted up to six years. The results of this study are consistent with Dodd et al.(2003), who reported that the majority of error patterns resolved rapidly between 2;5 and 4;0 years. The results supported this hypothesis. No gender differences were found between age groups.

Conclusion

These findings will provide useful information for speech-language pathologists assessing Persian speaking children and designing treatment objectives in Persian.

References

Amayreh, M. M., & Dyson, A. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research*, *41*(3), 642-653.

Bankson, N., & Bernthal, J. (1998). Factors related to phonologic disorders. *Articulation and phonological disorders*, 172ą232.

Bijankhan, M. (2006). *Phonology: Optimal Theory*. Tehran: Samt.

Damirchi, Z. (2010). *The Study of Phonological Processes in 2-6 year old Farsi-speaking children.* (MA), Iran University of Medical Science, Tehran.

Derakhsande, F. (1997). *The sound system of Persian speaking children.* (MA), Iran University of Medical Science, Tehran.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of British English-speaking children. *Clinical Linguistics & Phonetics*, *17*(8), 617-643.

Fahim, M. (1995). *Marahele Roshd-e Zaban-e Farsi ya "Ektesab-e Zaban-e Madari.* Paper presented at the *The Third Linguistic Conference*. Tehran, Iran.

Ghassisin, L. (2006). *Study of phonological Processes in 2-4 year old Isfahani speaking children.* (MA), Iran University of Medical Sciences, Tehran.

Grunwell, P. (1987). *Clinical phonology*. Baltimore: Williams & Wilkins.

Hall, M. (2007). Phonological characteristics of Farsi speakers of English and l1 Australian English speakers' perceptions of proficiency. *Unpublished MA thesis. University of Curtin, Croatia*.

Holm, A., Crosbie, S., & Dodd, B. (2007). Differentiating normal variability from inconsistency in children's speech: normative data. *International Journal of Language & Communication Disorders*, *42*(4), 467-486.

Ingram, D. (1981). *Procedures for the phonological analysis of children's language*. Baltimore: University Park Press.

Keshavarz, M. H., & Ingram, D. (2002). The early phonological development of a Farsi-English bilingual child. *International Journal of Bilingualism*, *6*(3), 255-269.

Meshkatoddini, M. (1989). Roshde sedaha va nezame avayi-e zaban dar goftar-e koudak. Speech and The first Grammar in Persian speaking children. *Journal of the Faculty of Literature and Humanity* Sciences, *5*(2), 29-49.

Nourbakhsh, M. (2002). *First language acquisition (Persian child) from 36 to 50 months of age syntactic and morphological study.* (MA), Islamic Azad University, Tehran.

Reza Pour, M., Tahbaz, S., & Mehri, A. (1999). *Phonological Processing in Children with 24-42 months in Tehran.* (BA Unpublished thesis), University of Welfare and Rehabilitation Sciences, Tehran.

Samarah, Y. A. (1977). *The arrangement of segmental phonemes in Farsi*. Faculty of Letters, University of Tehran, Tehran, Iran.

Shirazi , T., Mehdipour -Shahrivar, N., Mehri, A., & Rahgozar, M. (2010). Study of phonological processes of 2-4 year old Farsi-speaking children. *Journal of Rehabilitation*, *12*, 17-22.

Yarmohammadi, L. (1965). *A Contrastive Study of Modern English and Modern Persian*. Indiana University.

Zarifian, T., Tehrani, L., Modaresi, Y., Dastjerdi -Kazemi, M., & Salavati, M. (2013). Percentage of correct consonants in Persian speaking children and its psychometric features *Journal of Exceptional Children*, *4*, 45-53.

Zarifian, T., Tehrani, L., Modaresi, Y., Dastjerdi- Kazemi, M., & Salavati, M. (2014a). The percentage of vowel correct scale in Persian speaking children. *Iranian Rehabilitation Journal*, *12*(19), 5-8.

Zarifian, T., Tehrani, L., Modaresi, Y., Dastjerdi- Kazemi, M., & Salavati, M. (2014b). The Persian version of phonological test of diagnostic evaluation articulation and phonology for Persian speaking children and investigating its validity and reliability. *Journal of Audiology*, *23*(4), 10-20.