



International Society of Experimental Linguistics

**ExLing 2021**

**Proceedings of 12<sup>th</sup> International Conference of  
Experimental Linguistics**

11-13 October 2021  
Athens, Greece

Edited by Antonis Botinis



National and Kapodistrian  
University of Athens





International Society of Experimental Linguistics

## **ExLing 2021**

Proceedings of 12<sup>th</sup> International Conference of

Experimental Linguistics

11-13 October 2021

Athens, Greece

Edited by Antonis Botinis

ExLing 2021  
Proceedings of 12<sup>th</sup> International Conference of Experimental Linguistics

Published by ExLing Society  
Electronic edition  
ExLing 2021  
Athens, Greece  
ISSN: 2529-1092  
ISBN: 978-618-84585-1-2  
DOI: 10.36505/ExLing-2021/12  
Copyright © 2020 ExLing Society

## Foreword

---

Welcome to the 12<sup>th</sup> International Conference of Experimental Linguistics ExLing 2021. Even this time, due to the COVID-19 pandemic uncertainty, we will restrict ourselves to a virtual conference, much like the ExLing 2020 conference last year.

Our Society is defined by a collective commitment to understanding language through the development of experimental methods in Linguistics. It is an established and international forum for generations of linguists, where new and established researchers, participate and discuss developments in linguistic research and diverse developments of experimental methodologies.

It began in 2006 in Athens with the first ExLing Conference, at the time an International Speech Communication Association (ISCA) ExLing Workshop; subsequently, it was hosted in places such as Paris, Saint Petersburg, Heraklion, Lisbon and since 2019 has been established as the annual International Conference of Experimental Linguistics.

ExLing 2021 has a repeated connection to the first one, as it takes place in Athens, but it goes virtually, exploring how technology would facilitate online language interaction of our members without losing the natural interpersonal communication that our Society has in high regard. To promote this interaction, this time, members of the ExLing Society have free access to the conference, so that the online presentation opens up to a wider audience.

This volume includes the proceedings of ExLing 2021. In addition to the main conference, ExLing 2021 hosts two special sessions encompassing diverse and energetic research domains that focus on experimental methods and current topics in Language Education and Language Pathology.

We would like to thank all ExLing 2021 participants and our keynote speakers, Edward Gibson, Robert J. Hartsuiker, Charles Hulme, Kathy Rastle, Martin Pichering, as well as colleagues from the International Advisory Committee and the Review Committee for their contribution to the successful outcome of the Conference.

Antonis Botinis  
ExLing Society

## Contents

---

<i>Rating nonnativeness in L1-Japanese L2-Arabic Speakers' Vowels</i> .....	1
Yahya Aldholmi, Sara Alotaibi, Malak Alrouqi .....	1
<i>Teen production and perception of dialectal Arabic speech rate</i> .....	1
Yahya Aldholmi, Asma Alqahtani, Rawan Aldhafyan .....	1
<i>An improved method for vowel space depiction</i> .....	1
Terese Anderson <sup>1</sup> , Grandon Goertz <sup>2</sup> .....	1
<i>The impact of L2 on L1 in students with learning disabilities</i> .....	1
Georgia Andreou, Maria Segklia .....	1
<i>SPEAKapp – Remote monitoring of language production to predict cognitive functioning</i> .....	1
Chiara Barattieri di San Pietro <sup>1</sup> , Valentina Simonetti <sup>1</sup> , Cristina Crocamo <sup>2</sup> , Maria Bulgheroni <sup>1</sup> .....	1
<i>Enhancing cross-linguistic influence through across-language priming</i> .....	1
Ioli Baroncini <sup>1</sup> , Jacopo Torregrossa <sup>2</sup> .....	1
<i>Inverse reaction time as an awareness measure in artificial grammar learning experiments</i> .....	1
Tsung-Ying Chen.....	1
<i>Sentence comprehension assessment in Russian</i> .....	1
Daria Chernova <sup>1</sup> , Artem Novozhilov <sup>1</sup> , Natalia Slioussar <sup>1,2</sup> .....	1
<i>Gradient acceptability between naïve and expert linguistic intuitions</i> .....	1
Jacee Cho <sup>1</sup> , Yafei Li <sup>2</sup> , Rebecca Shields <sup>2</sup> .....	1
<i>Acoustic and semantic processes during speech segmentation in French</i> .....	1
Mar Cordero-Rull <sup>1</sup> , Stéphane Pota <sup>2</sup> , Elsa Spinelli <sup>2</sup> , Fanny Meunier <sup>1</sup> .....	1
<i>Can locus equations model dialect-specific variation in coarticulation?</i> .....	1
Giovina Angela del Rosso .....	1
<i>Argument-adjunct asymmetry in long-distance wh-movement in Russian</i> .....	1
Yuliya Demina.....	1
<i>Grammatical-lexical pronoun dissociation in Moroccan Arabic agrammatism</i> .....	1
Loubna El Ouardi <sup>1</sup> , Samir Diouny <sup>2</sup> .....	1
<i>Constraints of lexical and grammatical aspect on event representations in Mandarin Chinese</i> .....	1
Shaohua Fang <sup>1</sup> , Fei Yuan <sup>2</sup> .....	1
<i>Uncovering variation in classifier assignment in Oceanic</i> .....	1
Michael Franjich <sup>1</sup> , Greville G. Corbett <sup>1</sup> , Alexandra Grandison <sup>2</sup> .....	1
<i>Professions and gender agreement in Russian</i> .....	1
Marina Frolova <sup>1</sup> , Natalia Slioussar <sup>1,2</sup> .....	1
<i>Applied Cognitive Linguistics and design of L2 figurative language material</i> .....	1
Ioannis Galantomos .....	1
<i>A multidimensional approach in teaching L2</i> .....	1
Matteo Greco .....	1
<i>The acquisition of English relative clauses by L1 Arabic and Korean speakers</i> .....	1
John Hitz.....	1

<i>The Universal Perceptual Model (UPM) of second language</i> .....	1
Georgios P. Georgiou.....	1
<i>An improved method of F0 determination</i> .....	1
Grandon Goertz <sup>1</sup> , Terese Anderson <sup>2</sup> .....	1
<i>Code-switching cost in word recognition</i> .....	1
Wanying Hu.....	1
<i>L2 idiom learning and L1-L2 similarity</i> .....	1
Ferdy Hubers, Catia Cucchiarini, Helmer Strik.....	1
<i>Syntactic priming may not lead to language change</i> .....	1
Irati Hurtado.....	1
<i>Pragmatic factors facilitate Condition C violations cross-linguistically</i> .....	1
Ioannis Iliopoulos, Claudia Felser .....	1
<i>Prosodic manifestations of conversational roles in two types of collaborative tasks</i> .....	1
Tatiana Kachkovskaia, Pavel Kholiavin, Anna Mamushina, Alla Menshikova .....	1
<i>L2 English request strategies in Cyprus setting</i> .....	1
Sviatlana Karpava .....	1
<i>Vocabulary skills in Down Syndrome and typical development</i> .....	1
Dimitra Katsarou <sup>1</sup> , Georgia Andreou <sup>2</sup> .....	1
<i>The role of foreign languages in the mobility journey of young Greeks</i> .....	1
Angeliki Kiapekaki.....	1
<i>Structural convergence in spoken English discourse</i> .....	1
Christina S. Kim, Gloria Chamorro .....	1
<i>Autistic traits in duration of utterance-final particles in Japanese</i> .....	1
Sachiko Kiyama <sup>1</sup> , Ge Song <sup>1</sup> , Kuniya Nasukawa <sup>2</sup> .....	1
<i>Exploring the role of verb frames in assessing semantic difficulty</i> .....	1
Meichun Liu, Zhuo Zhang .....	1
<i>The prosodic structure in French: Analysis with deep learning networks</i> .....	1
Philippe Martin .....	1
<i>Wug-testing phonetic prominence in Munster Irish</i> .....	1
Connor McCabe .....	1
<i>The production of Greek voiceless fricatives by young children</i> .....	1
Elena Nirgianaki <sup>1, 2</sup> , Vasiliki Kampaki <sup>2</sup> .....	1
<i>Processing of singular agreement controller in Turkish</i> .....	1
Ayşe Gül Özay Demircioğlu.....	1
<i>Noun plural inflection in German-speaking individuals with Down syndrome</i> .....	1
Martina Penke, Bernadette Wittecy .....	1
<i>Text presentation and information processing in Russian</i> .....	1
Tatiana Petrova .....	1
<i>Exploring inter-dialectal mutual intelligibility and SDA in Assamese</i> .....	1
Sabbah Qamri.....	1
<i>Language teaching and assessment in the context of World Englishes</i> .....	1
Aicha Rahal.....	1

<i>Validation of the Moroccan Arabic version of the Token Test</i> .....	1
Youssef Rami <sup>1</sup> , Samir Diouny <sup>2</sup> , Najib Kissani <sup>3</sup> , Assma Boudanga <sup>3</sup> .....	1
<i>Reverse transfer of metacognitive reading strategies of Moroccan trilingual learners</i> .....	1
Hassane Razkane <sup>1</sup> , Samir Diouny <sup>2</sup> .....	1
<i>Intonation of the Khakas fairytales</i> .....	1
Tatiana Ryzhikova, Ilya Plotnikov, Anton Shamrin.....	1
<i>Laryngeal activity in Barabian and the Surgut dialect of Khanty</i> .....	1
Tatiana Ryzhikova <sup>1</sup> , Ilya Plotnikov <sup>2</sup> , Nikolay Urtegeshev <sup>1</sup> .....	1
<i>HAT - A new corpus for experimental stylometric evaluation in Arabic</i> .....	1
Halim Sayoud, Siham Ouamour.....	1
<i>Affective norms for Chinese characters</i> .....	1
Thomas Schlatter.....	1
<i>Effects of agent position and orientation on perception and production</i> .....	1
Judith Schlenker <sup>1</sup> , Yulia Esaulova <sup>1</sup> , Barbara Zeyer <sup>1</sup> , Martina Penke <sup>1</sup> .....	1
<i>Effect of face mask and noise on word recognition by children and adults</i> .....	1
Anna Sfakianaki <sup>1</sup> , George P. Kafentzis <sup>1</sup> , Devora Kiagiadaki <sup>1</sup> , George Vlahavas <sup>2</sup> .....	1
<i>Global English metaphors on Greek press</i> .....	1
Elpida Sklika.....	1
<i>Extracting word-like units when two concurrent regularities collide: Electrophysiological evidence</i> .....	1
Ana Paula Soares, Alexandrina Lages, Helena Mendes Oliveira, Francisco-Javier Gutiérrez-Domínguez.....	1
<i>Factive verbs and island effects in Romanian</i> .....	1
Irina Stoica.....	1
<i>/t/ variation in Connemara English</i> .....	1
Kate Tallon.....	1
<i>Virtual rapport management: responses to positive online reviews</i> .....	1
Ly Wen Taw <sup>1,2</sup> , Shamala Paravasivam <sup>2</sup> , Alan Libert <sup>1</sup> , Christo Moskovsky <sup>1</sup> .....	1
<i>Using feature selection to evaluate pathological speech after training with a serious game</i> .....	1
Loes van Bommel <sup>1,2</sup> , Catia Cucchiari <sup>3</sup> , Helmer Strik <sup>3,4,5</sup> .....	1
<i>What is difficult in second language acquisition of syntax</i> .....	1
Jiaojiao Yao.....	1
<i>NP types of distractors and the processing of English cleft sentences</i> .....	1
Myung Hye Yoo, Satoshi Tomioka, Rebecca Tollan.....	1
<i>The Production of Emphatic Stress Test for Korean</i> .....	1
Jihyeon Yun <sup>1</sup> , Jee-Hye Chung <sup>2</sup> , Yeong-Wook Kim <sup>3</sup> , Il-Young Jung <sup>2,3</sup> .....	1
<i>Profiling the central auditory processing of speech sounds: a MMN multifeature study</i> .....	1
Maria Ioanna Zavogianni <sup>1,2</sup> , Ferenc Honbolygó <sup>1,3</sup> .....	1
<i>Thematic roles in dementia: the case of psychological verbs</i> .....	1
Eleni Zimianiti <sup>1</sup> , Maria Dimitrakopoulou <sup>2</sup> , Anastasios Tsangalidis <sup>2</sup> .....	1
<i>Oral and silent reading in adolescents: evidence from Russian</i> .....	1
Vladislav Zubov <sup>1</sup> , Tatiana Petrova <sup>2</sup> , Svetlana Alexeeva <sup>2</sup> .....	1







# Rating nonnativeness in L1-Japanese L2-Arabic Speakers' Vowels

Yahya Aldholmi, Sara Alotaibi, Malak Alrouqi

Department of Linguistics & Translation Studies, King Saud University, Saudi Arabia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This experiment investigates whether Arabic native speakers can detect nonnativeness (i.e., accentedness) in similar L1-Japanese L2-Arabic short and long vowels. We asked two native speakers of Arabic and two Japanese learners of Arabic as a foreign language to record 48 Arabic CVCVC vs. CVCV:C word pairs at a normal speech rate. We then isolated the second vowel from each pair and instructed 33 NSAs to rate, on a 1-7 scale (1= certainly native, 7= certainly nonnative), the nonnativeness of the target vowels. Results show that, irrespective of L1-L2 segment similarity, the raters were successfully able to detect accentedness in the nonnative stimuli with no overall difference in the nonnativeness ratings for short vs. long vowels.

Keywords: duration, accentedness detection, nonnativeness rating, Japanese, Arabic

## Introduction

Previous research assessing a listener's capacity to successfully detect accentedness/nonnativeness in L2 production (e.g., Flege 1984; Flege, Munro, & MacKay, 1995; Mackay, Flege, & Imai, 2006; Park, 2013; Almohareb, 2020; Zhi & Li, 2021) has utilized experimental materials ranging from entire passages to complete sentences, fragments, whole or partial words, syllables, segments, and even portions of segments. A handful of such studies have indicated duration as a critical factor contributing to the perceived degree of nonnativeness and provide evidence that longer materials, as common wisdom suggests, supply more information to listeners attempting to detect or rate nonnativeness. Stimulus duration has been examined collectively with other factors, for example level of proficiency, and previous results show that the detection of nonnativeness in a minimal utterance produced by highly proficient L2 learners is most reliable when using stimuli containing segments that are dissimilar to those in the learners' L1 (e.g., Park, 2013). Taking into consideration the role of both *stimulus duration* and *segment similarity* in rating nonnativeness, we examine whether native speakers of Arabic (NSAs) are able to detect nonnativeness in the production of three short and long vowels ([a-a:], [i-i:], and [u-u:]), each of which has a similar version in the counterpart language) by Japanese learners of L2 Arabic (JLAs). We also ask whether long vowels can provide more information than short ones to the NSAs making these judgments.

## Methodology

### Stimuli

In pursuit of the study objective, we carefully selected 48 Arabic CVCVC vs. CVCV:C minimal or nearly minimal pairs in which the target V2 was varied ([a-a:], [i-i:], and [u-u:]). The target vowels, which exist both in Arabic and Japanese, were always neighbored with obstruents, as in /nabat/ “grew” vs. /naba:t/ “plant”, in order to reduce the impact of consonantal information on nonnativeness ratings, although most of the consonants that appeared adjacent to the target vowels exist in both languages. We instructed two (one male and one female) NSAs and two (one male and one female) JLAs to produce the list at a normal speech rate, using the carrier phrase /qul-tu/ “I said”, similar to phrases such as /katab-tu (ʔajd<sup>s</sup>-an)/ “I wrote (as well)” used in previous studies (e.g., Tsukada, 2010). We then extracted the target vowels from each pair and placed each vowel between two beep tones of a lower amplitude.

### Procedure

Thirty-three NSAs (age  $M= 27.7$ ,  $SD= 4.7$ ) were instructed to rate, on a 1-7 scale (1= certainly native, 7= certainly nonnative), the (non)nativeness of each vowel. The experiment was self-paced and the participants were allowed to listen to each trial as many times as they wished. For familiarization purposes, each participant was provided with a few practice examples, which were excluded from the actual experiment, before they commenced the task.

## Results

The overarching results indicate that while nonnativeness ratings for long vs. short vowels in the nonnative stimuli were similar ( $Median= 4$  for both) in general, the raters were successfully able to detect nonnativeness in the nonnative stimuli overall as demonstrated by a difference in their ratings ( $Median= 4$  and  $2$  for vowels produced by JLAs and NSAs, respectively). These results are supported by the main effect outcome of a repeated measures regression model in which both *stimulus (non)nativeness* and *stimuli duration* were fit as predictor factors and with *nonnativeness ratings* as a predicted factor, Wald  $\chi^2(1) = 37.8$ ,  $p < .001$ , and by the interaction effects between the two predictor factors, Wald  $\chi^2(1) = 16.4$ ,  $p < .001$ .

As exhibited in Figure 1, although most of the ratings for each condition are relatively evenly distributed over the entire scale, there are two outstanding patterns. First, the participants rated approximately 52% of the native long vowels as *certainly native* and 18% of the vowels from the same condition as *native*; that is, roughly 70% of the native long vowels were rated as (*certainly*) *native*. Second, short vowels were responded to with the *uncertain* rate (28% for nonnative short vowels and 22% for native short vowels) more than the long vowels were (13% for nonnative long vowels and 8% for native long vowels).

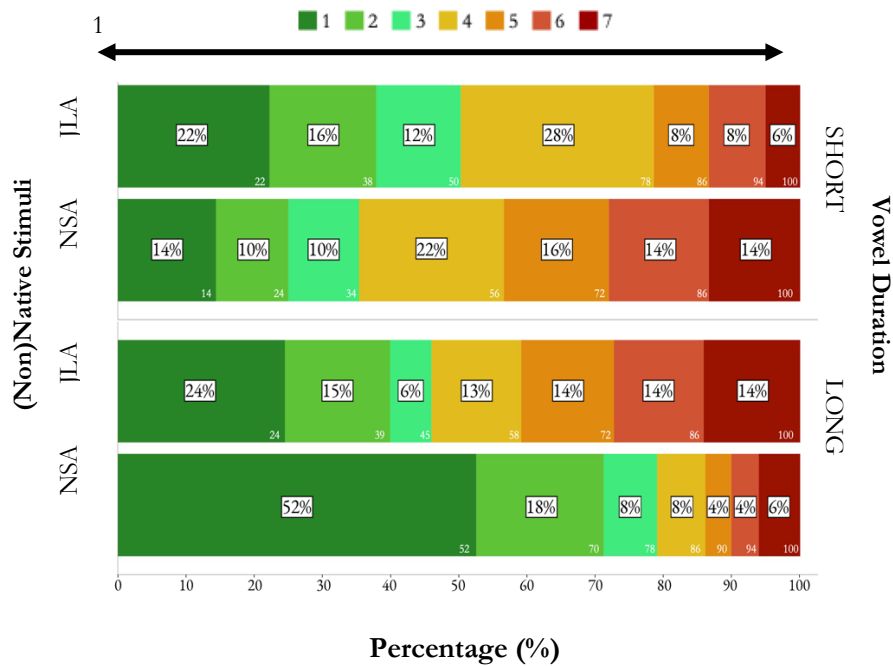


Figure 1. Percentages of (non)nativeness ratings for NSAs vs. JLAs' short vs. long vowels

## Discussion and conclusion

This study set out with two main related questions. The first probes whether NSAs can detect nonnativeness in three Arabic short vs. long vowels produced by JLAs. Although Japanese has contrastive duration and the vowels under investigation exist in the Japanese phonological repository, the current findings show that accentedness/nonnativeness can be still detected in L1-L2 similar segments. The second question asks whether long vowels facilitate nonnativeness detection more than short vowels. The findings reveal that both short and long vowels provide roughly the same magnitude of information in this respect.

Three pertinent points must be noted here. First, most previous studies that used a whole or partial segment to rate accentedness utilized consonants rather than vowels. For instance, Flege (1984) found that accentedness can be detected from a portion of a segment (30 ms of a /t/ burst). Harada (2006) instructed his Japanese raters to evaluate accentedness, focusing only on the bilinguals' singleton-geminate distinction and found out that the durational difference between the two categories was not sufficient to sound native-like.

The current findings add to the body of that research; vowels that exist in both L1 and L2 provide enough information for raters to detect nonnativeness.

Second, research on the durational short-to-long ratio in Arabic vs. Japanese vowels has shown both similarities and dissimilarities. For example, Tsukada (2009) states that the short-long ratio in Arabic and Japanese is not substantial (0.43 and 0.39 ms, respectively), while other studies (e.g., Tsurutani, Tsukada, & Ishihara, 2010) state that unlike Arabic vowels, Japanese short vowels are shorter than 50% of the duration of their long counterparts. Since accentedness was detected in both short and long vowels in the current study, it is likely that the learners were unable to maintain the durational ratio for both short and long vowels (i.e., duration was not Arabic-like in both types), which resulted in accentedness/nonnativeness in production and consequently in perception.

Third, despite careful attention to controlling the design of this study, there remain some limitations. In particular, this study relies on findings from previous research. A more careful study would recruit as many native and nonnative Arabic and Japanese talkers as possible, measure the durational ratio, and then use the talkers' productions as stimuli for the accentedness rating task. This would allow a closer inspection of the source of accentedness and yield more robust results.

## References

- Almohareb, S. 2020. The Effects of rater-specific characteristics on the rating of foreign accent (Doctoral dissertation, George Mason University).
- Flege, J.E. 1984. The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, 76(3), 692-707.
- Flege, J.E., Munro, M.J., MacKay, I.R. 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, 97(5), 3125-3134.
- Harada, T. 2006. The acquisition of single and geminate stops by English-speaking children in a Japanese immersion program. *Studies in Second Language Acquisition*, 28(4), 601-632.
- Mackay, I. R., Flege, J. E., Imai, S. 2006. Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics*, 27(2), 157-183.
- Park, H. 2013. Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, 41(2), 78-87.
- Tsukada, K. 2009. An acoustic comparison of vowel length contrasts in Arabic, Japanese and Thai: Durational and spectral data. *Int. J. Asian Lang. Process.*, 19(4), 127-138.
- Tsurutani, C., Tsukada, K., Ishihara, S. 2010. Comparison of native and non-native perception of L2 Japanese speech varying in prosodic characteristics. In *Proceedings of 2010 Australasian International Conference of Speech Science and Technology* (pp. 122-125).
- Zhi, N., Li, A. 2021. Acoustic salience in the evaluations of intelligibility and foreign accentedness of nonnative vowel production. *Lingua*, 256, 103069.

# Teen production and perception of dialectal Arabic speech rate

Yahya Aldholmi, Asma Alqahtani, Rawan Aldhafyan

Department of Linguistics and Translation Studies, King Saud University, Saudi Arabia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This study investigates the effect of native dialect on speech rate as produced and perceived by teenage speakers of three dialects of Arabic: Saudi, Syrian, and Egyptian. Fifteen participants reproduced ten Arabic sentences in their native dialect. The results of a production experiment show that, unlike Egyptian speakers, Saudi and Syrian speakers exhibit a high similarity in their speech rate. In a perception experiment, fifteen Saudi, Syrian, and Egyptian participants rated the tempo of their own dialect as well as the other two dialects on a 7-point scale. The findings reveal that the high similarity in the speech rate produced by Saudi and Syrian speakers does not entail a similarity in their perceived speech rates.

Keywords: dialectal Arabic, speech rate, production, perception, teens.

## Introduction

Previous research has explored the interrelationships between speech rate, age, and language/dialect across different dimensions. For instance, Robb et al. (2004) found that adult New Zealanders have higher speech rates than adult British and American speakers. This is at odds with the results of a later study, however, in which American children demonstrated faster speaking rates than New Zealander *children* (Robb and Gillon, 2007). Dialectal speech rate variations are not limited to dialects that exhibit substantial differences; rather, closely related dialects may also show notably incommensurate speech rates. For example, Jacewicz et al. (2009) found significantly higher articulation rates for Wisconsin speakers compared to North Carolina speakers.

When it comes to Arabic, speech rate in general has yet to receive adequate attention in the literature, but in particular there is a veritable lack of studies on Arabic speech rate as produced by *teenagers*. Therefore, the current two experiments supply *preliminary data* on speech rate as produced and perceived by teen speakers of three Arabic dialects. Specifically, we investigate the effect of *native dialect* on speech rate as produced and perceived by *teenage* speakers of Saudi, Syrian, and Egyptian Arabic, attempting to address two research questions: (1) To what extent is speech rate similar or dissimilar across the three distinct Arabic dialects under investigation? (2) How does a speaker of a given dialect's own speech rate impact her perceived speech rate of the other two dialects?

## Production experiment

### Methodology

The stimuli used in this experiment were ten seven-word Modern Standard Arabic (MSA) sentences ( $M$  of syll/sentence= 21.1,  $SD$ = 3.75). While most of the sentences were constructed by the experimenters, some were adopted from external sources. Fifteen female Egyptian, Syrian, and Saudi teen speakers participated in this experiment. Each participant was met individually and instructed to read each sentence and reproduce it in the participant's dialect at a normal speech rate while being recorded. The participants were encouraged to reproduce the same sentence multiple times whenever they had the urge to do so.

### Results

As shown in Table 1, the Egyptian participants were found to have a higher speech rate ( $M$ = 6.7 syll/s,  $SD$ = 0.95) than the Saudi and Syrian participants who themselves share a similar speech rate ( $M$ = 5.8 syll/s,  $SD$ = 0.73 and  $M$ = 5.7 syll/s,  $SD$ = 1.01, respectively). A repeated-measures ANOVA test revealed a statistically significant difference ( $F$  [2, 8]= 7.852,  $p$ = 0.01) attributed to the effect of Arabic dialects on speech rate. A post-hoc analysis test revealed statistically significant differences between Egyptian and Saudi dialects ( $p$ = 0.01), as well as between Syrian and Egyptian ( $p$ = 0.04). There was no statistically significant difference between Saudi and Syrian dialects ( $p$ = 0.6).

Table 1. Speech rate (syll/s) for Egyptian, Saudi, and Syrian dialects.

Dialect	Egyptian	Saudi	Syrian
Mean (SD)	6.7 (0.95)	5.8 (0.73)	5.7 (1.01)

## Perception experiment

In the first experiment we established a baseline for speech rate in the three dialects under investigation. In the second experiment, we sought to understand how participants perceive speech rate in their own dialects as well as in the other two dialects, following the methodology detailed below.

### Methodology

Ten sentences per dialect were selected from the 150 recordings supplied by the participants in the production experiment, making up a total of thirty trials. Among the thirty trials, each sentence appeared only once per dialect; that is, each sentence featured three dialectal versions. A different group of fifteen female Egyptian, Syrian, and Saudi teen speakers served as raters to judge the speech rate of the amplitude-normalized stimuli in the three dialects. It was necessary to match the age, gender, and education level of the participants in



the second experiment to those in the first in order to avoid any mismatch effects on speech rate perception. The raters were presented (online) with thirty randomly ordered auditory sentences and asked to rate the tempo of each using a 7-point rating scale (1= extremely slow, 4= normal, 7= extremely fast).

## Results

The Egyptian dialect tends to be perceived as faster than both Syrian and Saudi dialects, especially by the Saudi participants (*Median*= 6). The Egyptian dialect received a rating of fast in 78% of the responses and did not receive a rating of extremely slow at all across all stimuli and participants. The Saudi dialect is perceived as slightly slower than Egyptian and Syrian by the Egyptian and Syrian participants (*Median*= 3.5). None of the stimuli in the Saudi dialect were rated as extremely fast, and only 19% of the responses were rated as slightly fast, with most ratings clustered on the left side of the scale (from extremely slow to normal). The speech rate of the Syrian dialect was perceived as normal by speakers of all three dialects, including the Syrian participants themselves. One consistent observation across the three dialects is that the participants each perceive their own dialects as normal (*Median*= 4). A repeated-measures Generalized Linear Model indicated speech dialect as a statistically significant predictor of the perceived speech rate, Wald  $\chi^2(2)$ = 116.247,  $p$ = .0001. Participant dialect is also a statistically significant predictor of the perceived speech rate, Wald  $\chi^2(2)$ = 42.250,  $p$ =.0001. The two predictors show a statistically significant interaction effect, Wald  $\chi^2(4)$ = 35.093,  $p$ = .0001.

## Discussion and conclusion

The results of the production experiment show that Egyptian speakers are indeed faster speakers than are Saudi and Syrian speakers. The present results are consistent with previous findings regarding the effect of dialect on speech rate. For instance, Robb et al., 2004 offers evidence that New Zealand English speakers have faster speaking and articulation rates than do speakers of other English varieties. Leemann and Siebenhaar (2007) also examined speech rate in Swiss German dialects from two regions and found that Valais speakers demonstrate a higher speaking rate than that of Bernese speakers. Similarly, Leemann et al. (2014) found that speakers from Zurich have a higher speech rate than speakers from Bern. The speech rate recorded in this study is not comparable to the recorded speech rate for Jordanian speakers in Damhoureye et al. (2020) due to differences in units of measurement; Damhoureye and colleagues calculated speech rate using words while the present study measures in syllable per second. The speech rate of Egyptian in the present study is slightly higher than that previously recorded for Moroccan (average= 5.0 syll/s) in Vaane (1982) although it is still within the reported range (4.6-7.0) (Gósy, 1991) as the speech rate of Egyptian here is 6.7 syll/s. However, despite differences in participant gender and age, the speech rates for both Saudi (5.8

syll/s) and Syrian (5.7 syll/s) speakers are extremely similar to the average reported by Vaane (1982) for Moroccan (5.0 syll/s).

The findings of the perception experiment indicate that participant judgments of speech rate were affected by the dialect of the stimuli presented to them. Specifically, the Egyptian dialect, which was determined to be faster than the other two dialects in the production study, was perceived as a “fast” dialect. The current findings provide a preliminary investigation into the production and perception of speech rate for female Saudi, Syrian, and Egyptian female teens. These findings support further research on Arabic speech rate and are especially relevant to speech pathologists and dialectal Arabic teachers who work with Arabic-speaking youth. Future studies may include other dialectal Arabic varieties, increase the sample size, collect a larger amount of data, and control for syllable complexity (i.e., segment rate) in both methodology and an analysis of results, as this has been shown to be an important factor (e.g., Plug and Smith, 2021) in speech rate perception.

## References

- Damhoureye, M., Darawsheh, W., Qa'dan, W., Natour, Y. 2020. Preliminary Speech Rate Normative Data in Adult Jordanian Speakers. *Journal of Language Teaching and Research* 11, 204.
- Gósy, M. 1991. The perception of tempo, in Gósy, M. et al. (eds.), *Temporal factors in speech*, Budapest: Research Institute for Linguistics, 63-106, Hungarian Academy of Sciences.
- Jacewicz, E., Fox, R., O'Neill, C., Salmons, J. 2009. Articulation rate across dialect, age, and gender. *Language Variation and Change* 21, 233-256.
- Leemann, A., Kolly, M., Dellwo, V. 2014. Crowdsourcing regional variation in speaking rate through the iOS app ‘Dialäkt Äpp’. In *Proc. International Conference on Speech Prosody*, 217-221, Dublin.
- Leemann, A., Siebenhaar, B. 2007. Intonational and Temporal Features of Swiss German. In Trouvain, J., Barry, W. J. (eds), *Proc. 16<sup>th</sup> International Congress of Phonetic Sciences*, 957-960. Saarbrücken: Universität des Saarlandes.
- Plug, L., Smith, R. 2021. The role of segment rate in speech tempo perception by English listeners. *Journal of Phonetics*, 86, 101040.
- Robb, M., Gillon, G. 2007. Speech rates of New Zealand English- and American English-speaking children. *Advances in Speech Language Pathology* 9, 173-180.
- Robb, M., Maclagan, M., Chen, Y. 2004. Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics & Phonetics* 18, 1-15.
- Vaane, E. 1982. Subjective Estimation of Speech Rate. *Phonetica* 39, 136-149.

# An improved method for vowel space depiction

Terese Anderson<sup>1</sup>, Grandon Goertz<sup>2</sup>

<sup>1</sup>Chemistry and Chemical Biology, University of New Mexico

<sup>2</sup>Educational Linguistics, College of Education & Human Sciences, University of New Mexico

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

All speech is composed of combined individual sound waves that gives speech its particular qualities. Vowel space measurements are usually plotted in charts in which the second formant is plotted against the first formant. For this investigation, the formant values are taken from a computer program that employs a Fourier methodology to analyse the sound signal. This paper presents information about an improved method for vowel space depiction that was created by employing the mathematical integration of formant data. This model produces a more reliable mathematical model the vowel space. Each vowel formant can be represented by an algebraic equation.

Keywords: formant measurement, formant equation, vowel space, speech signal

## Introduction

Speech is a product of a vocal source which is the excitation signal, and a filter, or time-varying vocal tract. “The spectral envelope determines the relative magnitudes of the different harmonics, and it, in turn, is determined from the specific shape of the vocal tract during the phonation of that vowel” (Gold and Morgan 2000, p 25; Parker 1988).

A speech wave is a time evolution phenomenon that is modelled using partial differential equations that have a dependent variable that represents the wave value, an independent variable time, and one or more independent spatial variables (Elmore and Heald 1969; Pain 1993). Sound is the compilation of several simultaneous frequency waves that move through air. Sound waves interact as additive and subtractive sine wave components and form a unique sound, which is represented visually as a waveform in frequency space. *The essential idea is that speech is comprised of numerous sine waves, and each wave component has a distinct continuous mathematical function.*

The illustration in Figure 1, taken from page 765 of *The Handbook of Phonetic Sciences* (Ellis 2010), illustrates that a periodic function can be divided into its mathematical harmonic components. All sound waveforms can be mathematically represented by combinations of the mathematical equations that make up the waveform. Because speech is composed of multiple waves it is also possible to separate the sounds into formants.

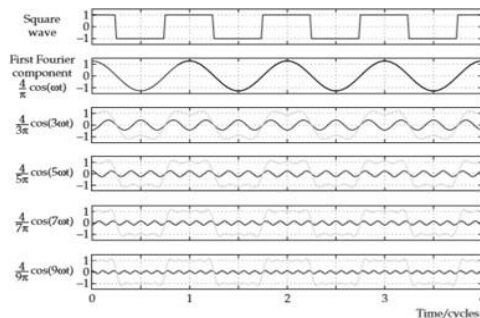


Figure 1. Each harmonic is an integer multiple of the fundamental frequency. Each of the five waves shown below the top square wave are components of the square wave.

### Speech as a mathematical phenomenon

A speech wave is modelled using partial differential equations, and these equations have an independent variable, time, and one or more independent spatial variables. The actual form of the wave is strongly dependent on the system's initial conditions, the boundary conditions, and system disturbances. Waves are described by their solutions to either linear or nonlinear partial differential equations (Elmore and Heald 1985).

There are numerous convolution methods that will separate speech into its component parts, the spectral envelope and excitation values. Once the speech is separated, both the spectral envelope and spectral fine structure can be efficiently parameterized. Methods of source-filter separation include linear prediction, cepstral analysis which produces a spectrum, and formant vocoding.

The vowel space values are traditionally produced by taking the linear average of the frequency values over time to produce a single value. These values are then plotted on a F1 vs F2 plot.

### Finding formant equations

Praat produces formant bands or regions, typically: F1, F2, F3, and F4. The Praat formant listing includes points of the wave equation, and not the equation, but the formant listing information is sufficient to determine the equation of the wave.

Separating the formant components of the speech wave allows the determination of a mathematical model equation which can be integrated and differentiated to produce the actual centroid of each function (see Kreyszig 1983, Stein 1973, and Thomas Calculus 2005). Laws of calculus state that the "average" or centroid for a continuous function is determined by the integral of that function. Equations were computed using MATLAB's Curve Fitting

Toolbox using the F1-F4 values from Praat. It was found that each sound formant can be effectively represented by an 8-term Gaussian equation. A total of 254 individual sounds were evaluated for comparison and to check the validity of the procedure. The general equation for a formant is the Gaussian Equation, below.

$$f(x) = A1 * e^{-(\frac{x-B1}{C1})^2} + A2 * e^{-(\frac{x-B2}{C2})^2} + A3 * e^{-(\frac{x-B3}{C3})^2} \\ + A4 * e^{-(\frac{x-B4}{C4})^2} \dots + A8 * e^{-(\frac{x-B8}{C8})^2}$$

This equation remains the same for all input formant data. These coefficient values (A1-A8, B1-B8, and C1-C8) are unique to each sound and formant and vary from small negative values to over 1000, and may include zero.

## Results

Integrating the equations of formant values consistently provides significantly better data fits than a method that uses a means procedure. Vowel spaces can be more accurately depicted. Integral values of the equations were compared to the traditional singular linear average (means) values of F1, F2, and F3 for the purposes of depicting vowel spaces.

Figure 2a (left) shows the F1 plot of a female speaker of Hmong producing a high front vowel. Linear modelling of the data does not result in very accurate information, and a  $R^2$  of 0.06. This F1 formant represents a modified wave function and so a wave equation can be fitted to the data, producing a much better fit as seen in Figure 2b (right). This is the same formant data as in Figure 2a but 2b has a  $R^2$  of 0.89.

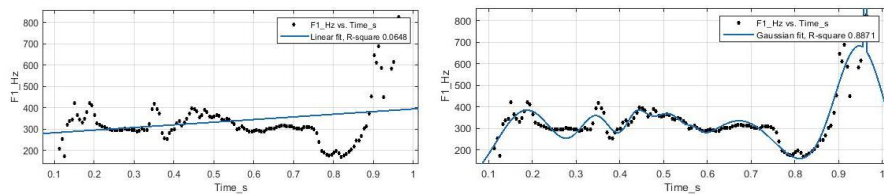


Figure 2a, 2b. 2a Left, The F1 of a high front vowel spoken by female speaker of Hmong, fitted with a linear equation (blue line). Right, the same vowel fitted with a Gaussian model that produces an equation that can be integrated (blue).

Integrating the Gaussian equation for each formant provides a more precise weighting of the information, which is a centroid value of the data points as seen in figure 3.

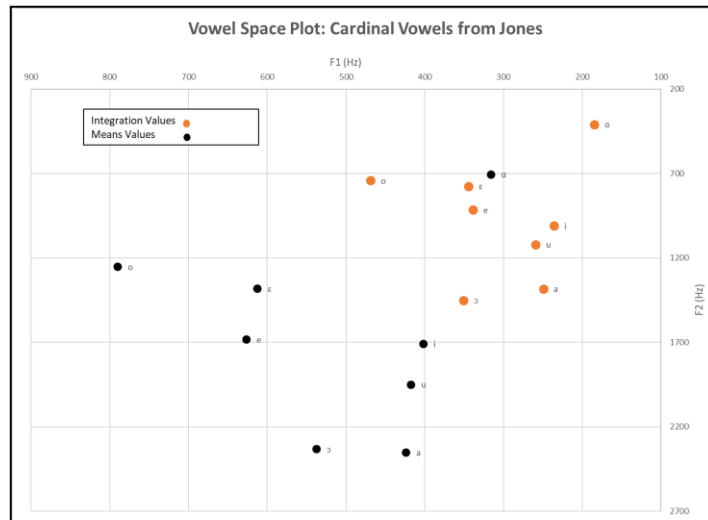


Figure 3. Vowel space plot of the cardinal vowels (Ladefoged, 2001). The red dots represent integrated vowel values, while the black dots depict average values of the formants. Vowels were not shown between slashes.

## Summary

This study delivered an effective model of formants which can produce integrated vowel space values that are more robust than the values created by means calculations.

## References

- Ellis, D. 2010. An Introduction to Signal Processing for Speech. In William Hardcastle, et al., *The Handbook of Phonetic Sciences*. New York: John Wiley & Sons.
- Elmore, W., Heald, M. 1985. *Physics of Waves*. New York: Dover.
- Gold, B., Morgan, N. 2000. *Speech and Audio Signal Processing*. NY: John Wiley & Sons.
- Kreyszig, E. 1983. *Advanced Engineering Mathematics*. Hoboken, NJ: John Wiley & Sons.
- Ladefoged, P. 2001. *A Course in Phonetics*. Harcourt College Publishers: Fort Worth, TX.
- Pain, H.J. 1993. *The Physics of Vibrations and Waves*. NY: John Wiley & Sons.
- Parker, S. 1988. *Acoustics Sourcebook*. NY: McGraw-Hill Book Company.
- Stein, Sh. 1973. *Calculus and Analytic Geometry*. New York: McGraw-Hill.
- Thomas, G., Weir, M., Haas, J., Giordano, F. 2005. *Thomas' Calculus, 11<sup>th</sup> Edition*. Boston: Addison-Wesley Longman, Inc.

# The impact of L2 on L1 in students with learning disabilities

Georgia Andreou, Maria Segkalia

<sup>1</sup>Department of Special Education, University of Thessaly, Greece

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Adequate knowledge of the first language (L1) has a positive impact on second language (L2) learning. Research has also confirmed a cross linguistic skills transfer from L2 to L1, according to the Cognitive Retroactive Transfer Hypothesis, which claims that an intervention in L2 linguistic skills can also benefit L1's corresponding skills even among students with learning disabilities. The present study examined the cross linguistic skills transfer of decoding and spelling skills from L2 (English) to L1 (Greek) among students with learning disabilities, after an intervention in L2. The results showed that transfer existed for decoding, but not for spelling, which seems to be language specific.

Keywords: CRT, cross-linguistic transfer, language

## Introduction

Students with L.D. and poor reading skills in their L1 will also show weaknesses in L2 because metalinguistic skills are common to all languages. However, children whose performance is high in L1 will perform well in L2 (Feder & Abu-Rabia, 2020).

Research has shown that there may be cross-linguistic transfer from L1 to L2 and its effect on L2 learning. However, only four researches study the transfer of these skills from L2 to L1 (Abu-Rabia & Bluestein-Danon, 2012; Feder & Abu-Rabia, 2020; Abu-Rabia & Shakkour, 2014; Abu-Rabia, Shakkour, & Siegel, 2013) based on the Cognitive-Retroactive Transfer Hypothesis (CRT), according to which, an intervention in L2 will improve the linguistic skills of both L2 and L1.

The aim of the present study is to examine further the transfer of decoding and spelling skills from English as an L2 to Greek as an L1 after an intervention program in L2, in students with L.D. and discuss them in terms of the CRT hypothesis. More specifically, our research hypotheses are:

- 1) It is expected that there will be an improvement of the participants in the decoding of L2 after intervention in L2.,
- 2) it is expected that there will be an improvement of the participants in the decoding of L1 after intervention in L2,
- 3) it is expected that there will be an improvement of the participants in the spelling of L2 after intervention in L2,
- 4) it is expected that there will be no improvement of the participants in the spelling of L1 after intervention in L2.

## Methodology

### Participants

35 Greek 8th grade students with Greek as an L1 and English as an L2 were the participants of the study. They were diagnosed with L.D. and they had all attended English classes since the 3rd grade only at school having the same level in L2 (beginners). The sample was divided into an experimental group (20 students) and a control group (15 students) with a combined average of 13.6 years of age.

### Procedure

Two weeks before the intervention both groups were tested in decoding and spelling in both languages. The same tests were administered two weeks after the intervention to all the students as well.

After the administration of the pre-tests the intervention in L2 began. All students of the experimental group participated in small group instruction sessions. The students had 45 minute lessons twice a week for a 5-month period.

## Results

In order to examine the statistical significance in L2 and L1 decoding and spelling skills before and after the intervention program, a nonparametric test (Wilcoxon test) for dependent samples was conducted.

In Table 1 the mean scores of the experimental group's L1 and L2 performance in each test for each linguistic skill before and after the intervention are presented.

Table 1. Mean scores of the experimental group in L1 and L2 decoding and spelling skills before and after intervention.

	Pre-intervention			Post-intervention		z-score	p value
	N	Mean	SD	Mean	SD		
L1 Decoding	20	94.65	3.13	102.20	1.88	-3.925	,000*
L2 Decoding	20	31.45	6.52	45.45	9.13	-3.924	,000*
L1 Spelling	20	35.15	5.35	35.50	5.26	,942	,346
L2 Spelling	20	24.20	4.92	31.50	4.81	-3,942	,000*

\*statistically significant difference ( $p < 0.05$ )

In Table 2 the mean scores of the control group's L1 and L2 performance in each test for each linguistic skill before and after the intervention are presented.



Table 2. Mean scores of the control group in L1 and L2 decoding and spelling skills before and after intervention.

L1 Decoding	Pre-intervention			Post-intervention			
	N	Mean	SD	Mean	SD	z-score	p value
	15	97.40	4.17	98.73	2.84	-1.569	,117
L2 Decoding	15	33.20	9.54	33.60	9.53	-1.038	,299
L1 Spelling	15	33.60	5.30	34.80	5.94	-2.412	,016
L2 Spelling	15	23.13	5.37	23.53	5.18	-.777	,437

## Discussion

The current study investigated the transfer of decoding and spelling skills from L2 to L1 after an intervention in L2 among students with LD. The findings of the present study indicate that the intervention program in L2 improved the performance of the participants in all English tests. Moreover, there was a similar improvement in L1 decoding but not in L1 spelling tests.

The first research hypothesis was verified. There was a statistically significant performance of the participants in the intervention group in the decoding skill in L2 after the intervention. Regarding the control group, no improvement in decoding skills was observed in the post-test since this group consisted of students with L.D. who did not receive any intervention.

The second research hypothesis was verified since there was an improvement in the decoding skill in L1 for the intervention group, but not for the control group. This improvement for the first group shows the cross-linguistic transfer of the skill from L2 to L1 confirming the CRT Hypothesis. It also seems that an intervention in L2 decoding can improve L1 decoding even in languages with different orthographic depth.

Furthermore, there was an improvement in the spelling skills of the intervention group in L2 after the intervention, and therefore verification of the third research hypothesis. In contrast, the control group did not show an improvement in its spelling skills during the final evaluation.

Moreover, it seems that interventions that focus on decoding and enhance phonological awareness, can also improve students' ability to encode words (Feder & Abu-Rabia, 2020). Also the intervention in both spelling and decoding skills in L2 improved the spelling skills of the intervention group students in this language.

Regarding the intervention group in the fourth research hypothesis, it seems that the spelling performance of L1 did not improve, because the spelling experiences from L2 were not transferred to L1. In contrast, the control group, showed a slight improvement in the spelling of L1 but not due to a transfer of the skill from L2 to L1, as the performance of the students in the control group in the spelling of L2 was constant in both measurements.

Moreover, the results for the intervention group are consistent with the findings of other studies that implemented a similar intervention program to the present study (Abu-Rabia & Bluestein-Danon, 2012; Abu-Rabia & Shakkour, 2014; Abu-Rabia, Shakkour, & Siegel, 2013).

## Conclusion

To conclude, it is understood that the systematic teaching of spelling and decoding skills in the foreign language is of high importance. It is also necessary to plan educational interventions in L1 and L2, so that both typical and non-typical students can benefit. Finally, the cooperation between the teachers of L1 and L2 language (and other languages, too), in schools, is considered important. In this way students with L.D. will have the opportunity to be successfully included in L1 and L2 classes.

## References

- Abu-Rabia, S., Bluestein-Danon, D. 2012. A study into the results of an intervention program of linguistic skills in English (L2) and its effect on Hebrew (L1) among poor readers: An examination of the cognitive-retroactive transfer (CRT) hypothesis. *Open Journal of Modern Linguistics*, 2(04), 131.
- Abu-Rabia, S., Shakkour, W. 2014. Cognitive retroactive transfer (CRT) of language skills among trilingual Arabic-Hebrew and English learners. *Open Journal of Modern Linguistics*, 4, 1-20.
- Abu-Rabia, S., Shakkour, W., Siegel, L. 2013. Cognitive Retroactive Transfer (CRT) of Language Skills among Bilingual Arabic-English Readers. *Bilingual Research Journal*, 36(1), 61-81.
- Feder, L., Abu-Rabia, S. 2020. An examination of differences in linguistic and meta-linguistic skills in English (FL) and Hebrew (L1): English intervention program for dyslexic, poor and normal readers. *The Journal of Educational Research*, 113(3), 226-243.

# **SPEAKapp – Remote monitoring of language production to predict cognitive functioning**

Chiara Barattieri di San Pietro<sup>1</sup>, Valentina Simonetti<sup>1</sup>, Cristina Crocamo<sup>2</sup>,  
Maria Bulgheroni<sup>1</sup>

<sup>1</sup>Ab.Acus s.r.l., Milan, Italy

<sup>2</sup>Department of Medicine and Surgery, University of Milano-Bicocca, Italy

<https://doi.org/10.36505/ExLing-2021/12>

## **Abstract**

Language production and comprehension can provide a useful perspective into an individual's mental health and cognitive abilities. SPEAKapp is a mobile application designed to deliver and analyze speech and language data for clinical and research purposes. It implements pre- and post-processing techniques based on Natural Language Processing (NLP) and Distributional Semantic Models (DSM) of language. The first functional prototype was tested for accuracy of data acquisition and elaboration, as well as for usability and acceptability in a pilot sample of fragile users. SPEAKapp showed good accuracy and replicability of results, and participants felt comfortable using the application. Further developments of the application are presented.

Keywords: language, NLP, mobile app, DSM, neuropsychology.

## **Introduction**

Speech and language assessment is a powerful yet unobtrusive tool that supports clinical assessment in various conditions implicating cognitive and affective symptoms. The current neuropsychological practice relies on standard paper-and-pencil batteries, requiring verbatim transcriptions and manual scoring. This is time-consuming and represents a barrier to frequent monitoring and remote assessment.

Technically, verbal responses are a stream of data suited to be processed with modern speech technologies. On the one hand, available automatic speech recognition software has reached remarkable accuracy. On the other, speech content can be modelled by relying on DSMs, which are automatic data-driven models of semantic representations that represent word meanings as numerical vectors in multi-dimensional spaces (Landauer & Dumais, 1997), and that can be used to simulate the structure of conceptual knowledge implied in the performance of semantic tasks (Mandera et al., 2017). However, to build upon these technologies and address actual clinical and research needs, it is pivotal to bridge the gap between end-users and clinicians on the one hand and developers and researchers on the other.

To do so, we developed "SPEAKapp", a system based on a mobile application to deliver and analyze speech and language data for clinical and research purposes. Precisely, SPEAKapp can: i) deliver standard neuropsychological tests and questionnaires; ii) collect audio responses; iii) perform speech-to-text transcription and noise removal, iv) score the results using digitalized deterministic approaches and semantic analysis; and v) store and manage output data. SPEAKapp frontend is implemented using the Flutter framework, whereas the backend is hosted on a cloud server. SPEAKapp relies on a commercial speech-to-text service (<https://cloud.google.com/speech-to-text>), as well as property algorithms to extract related meaningful semantic features implemented in Python. The system is GDPR compliant.

The main goal of the present study was to test: i) the reliability of the SPEAKapp system in terms of accuracy of automatic transcriptions; ii) the portability of the logic from a research environment to the final product environment, and iii) the usability and acceptability of the system in a pilot sample of potential clinical users.

## Methodology

A sample of 23 participants, Italian native speakers, among which 8 users of community mental health services, was recruited via personal referral. A categorical Verbal Fluency (VF) task was delivered through the app. A VF task is a standard neuropsychological test used to assess lexical retrieval: participants are asked to produce as many words as possible in a given semantic category (i.e., "animals") within a time limit (60 sec).

Automatic transcriptions were inspected for accuracy. Accuracy was calculated as  $TP + TN / (TP + TN + FP + FN)$ , where TP is the number of valid tokens correctly identified (True Positive), TN is the number of tokens correctly ignored as irrelevant (True Negative), FP is the number of tokens incorrectly transcribed (False Positive), and FN is the number of tokens not recorded although relevant (False Negative).

The results of the semantic analysis computed by the app were compared against the results of the original algorithms (Barattieri di San Pietro et al., 2020) in terms of the size of semantic clusters (the number of consecutive words produced that share similar properties) and the number of switches (i.e., the total number of transitions between these groups – Troyer et al., 1997) as computed from a word2vec (Mikolov et al., 2013) semantic space.

A System Usability Scale (SUS – Brooke, 1996) questionnaire was administered to the subgroup of participants who were users of community mental health services at the end of the data acquisition phase to test the acceptability and usability of the system.

## Results

The total number of spoken output (including sound fragments, repetitions, confabulations, conjunctions, false starts, and words) was manually identified ( $N = 782$ ). SPEAKapp recorded and transcribed  $N = 613$  tokens. Compared to the manual transcription for clinical purposes ( $N = 560$  words), SPEAKapp recorded and transcribed  $N = 68$  irrelevant tokens (0.12%), and failed to transcribe  $N = 15$  relevant tokens (0.03%). The resulting accuracy for research purposes of SPEAKapp was overall 89.71%.

On average, participants produced 23.13 words each (standard deviation  $SD = 8.34$ ). The mean number of switches was 9.61 ( $SD = 6.01$ ). The pooled mean of cluster size was  $M = 3.05$  words. The analysis was carried out twice, both with the original algorithms and the Python-embedded logic in SPEAKapp, and yielded identical results.

Analysis of the SUS questionnaire results revealed that participants felt comfortable using the application, whose functionalities were considered well integrated. Although not entirely autonomous in the use of the app, users felt that with a little initial training, they were able to learn what was needed to perform the tests. Overall results indicated that the users would feel comfortable using SPEAKapp again.

## Discussion

The present work aimed to test and validate the first functional prototype of SPEAKapp, a mobile application for language assessment that implements both standard deterministic approaches to test scoring as well as a set of novel indexes based on NLP and DSM. Results showed that the accuracy was adequate for clinical purposes and that data loss compared to manual transcription was nearly negligible. Results of the semantic analysis were in line with the expected, showing the successful translation to a commercial environment. Finally, the use of the application was considered practicable by participants.

To accommodate the needs of clinicians and researchers, the updated version of the app implements a comprehensive range of neuropsychological tests, such as ad-hoc versions of a prose recall test, a picture-naming test, a test with repetition of complex sentences, and an automatic series test. Tests are available in multiple languages. Given the growing interest in the analysis of speech acoustic analysis for clinical purposes, a PRAAT library (Boersma & Weenink, 2021) was implemented to extract fundamental frequency (F0), cycle-to-cycle perturbation measures of pitch (jitter), and amplitude (shimmer) as well as the harmonic-to-noise ratio (HNR). A designated web-based dashboard will enable clinicians to interact with the system for data entry and visualization. Collected data will feed a set of Machine Learning models on clinical outcomes based on changes over time of both standard and experimental indexes of language production to predict clinical visit annotations (target).

Beyond being a practical tool for data acquisition and transcription, the adoption of additional linguistic measures can measure subtle differences in language use. The integration of novel indexes based on verbal performance with standardized neuropsychological measures might lead to novel insights into mental health conditions, as well as to the identification of light and reliable indexes of cognitive functioning. In the long term, identifying a valid marker of treatment efficacy would support clinical research and innovation, facilitating the evaluation of new drugs' efficacy and the design of targeted approaches.

### Acknowledgements

SPEAKapp has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 857223.

### References

- Barattieri di San Pietro, C., de Girolamo, G. C., Luzzatti, C., Marelli, M. 2020. Mind your models! Distributional semantic models for the analysis of verbal fluency tasks in Schizophrenia Spectrum Disorders. The 26th Architectures and Mechanisms for Language Processing Conference, virtual congress.
- Boersma, P., Weenink, D. 2021. Praat: doing phonetics by computer [Computer program]. <http://www.praat.org/>
- Brooke, J. 1996. SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.
- Landauer, T. K., Dumais, S. T. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211.
- Mandera, P., Keuleers, E. & Brysbaert, M. 2017. Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, 92, 57-78.
- Mikolov, T., Chen, K., Corrado, G., Dean, J. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781
- Troyer, K., Moscovitch, M. Winocur, G., 1997. Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology*, 11(1), 138-46.

# Enhancing cross-linguistic influence through across-language priming

Ioli Baroncini<sup>1</sup>, Jacopo Torregrossa<sup>2</sup>

<sup>1</sup>Scuola Superiore di Dottorato, Università per Stranieri di Siena, Italy

<sup>2</sup>Institut für Romanische Sprachen und Literaturen, Goethe-Universität Frankfurt am Main, Germany

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

We investigate whether it is possible to prime a structure from a bilingual's one language (A) to the other (B), if this structure is grammatical in A, but not in B. Specifically, we focus on the use of the verb-subject-object word-order (VSO) in broad-focus sentences, which is allowed in Greek, but not in Italian. We use an across-language priming experiment to enhance the activation of VSO in Greek and observe whether this leads to the use of VSO in Italian. The results show an increase in the production of VSO in Italian following a Greek VSO prime. Overall, the study accounts for cross-linguistic effects in terms of the activation of a certain syntactic structure in a bilingual's processing system.

Keywords: cross-linguistic influence, structural priming, activation, ungrammatical structure, Greek-Italian bilinguals

## Introduction

Cross-linguistic influence in bilingual language production has been interpreted in terms of activation of a certain syntactic structure in a bilingual's processing system: A structure with a high degree of activation in one language exhibits the greatest tendency to appear in the other language (Serratrice 2016; Sharwood Smith and Truscott 2014). Structural priming experiments provide a privileged viewpoint to test the hypothesis that cross-linguistic influence from Language A to Language B is due to the activation of a certain structure in Language A. One can enhance activation of the target structure in Language A and observe whether this leads to an increased use of this structure in Language B. Across the literature, cross-linguistic priming experiments have mainly considered linguistic structures that are possible in a bilingual's both languages. For example, Hartsuiker et al. (2004) show that the use of passive sentences may be primed from Spanish to English. The aim of these studies is to understand whether the target structures have a common underlying representation across a bilingual's two languages (see also Vasilyeva et al. 2010). The aim of our study is to investigate whether structures that are grammatical in Language A but ungrammatical in Language B can be primed, too, in order to shed some new light on the cognitive mechanisms underlying cross-linguistic effects. We investigate the possibility to prime verb-subject-object structures (VSO) from

Greek to Italian. In Greek, both subject-verb-object (SVO) and VSO are possible word-orders in a broad-focus context, while in Italian VSO is not allowed (Roussou and Tsimpli 2006; Torregrossa et al. 2020).

## Methods

### Participants

36 Greek-Italian bilingual children (15 females, age range 7;5-11;10, M: 9;5) participated in the study. The children were all simultaneous or early sequential bilinguals (with an age of onset to Italian ranging between 0 and 3 years). At the moment of testing, they were attending an Italian immersion school in Athens (Greece), where Italian was the main medium of instruction. We refer to Baroncini & Torregrossa (in preparation) for further details concerning children's language exposure to Greek and Italian across different contexts over time.

### Materials and procedure

We designed an across-language priming experiment to investigate whether Greek-Italian bilingual children would produce an ungrammatical VSO sentence in Italian (target) after hearing a VSO sentence in Greek (prime).

The task was designed as a picture-description task in a bilingual mode: Children were asked to look at a picture appearing on the computer screen, listen to a sentence in Greek describing this picture (prime) and repeat it. Then a question in Italian appeared on the screen (*E adesso cosa succede qua?* "And what happens here?"), followed by a new picture. At this point, children had to describe the new picture in Italian. In order to keep the children more engaged in the task, all pictures were connected in a narrative.

As primes, we used VSO sentences and SVO sentences. Overall, children had to produce 40 target sentences in Italian, 20 preceded by a Greek SVO prime and 20 by a Greek VSO prime. Both primes and target sentences involved transitive verbs. We employed different lexical items in the prime and the target sentences. The verb to be used in the target sentence was provided in uninflected form in the picture, in order to prevent the results to be affected by children's lexical knowledge. We avoided the repetition of the subject and object constituents across the stimuli, in order to control for information structure.

## Results

First, we excluded the target sentences that were produced following an absent or incorrect repetition of the prime. This was done in order to make sure that the child listened to the prime properly and understood it correctly. Furthermore, we excluded sentences that were neither SVO nor VSO. The excluded sentences represent 27.29% of the total amount of targets.



Figure 1 shows the mean percentage of VSO sentences produced in Italian following a SVO and a VSO prime in Greek, respectively. After a SVO prime, the percentage of VSO sentences amounts to 4.47%. After a VSO prime, the percentage of VSO sentences increases by 3.23%, reaching 7.70%. Most of the produced target sentences were SVO, independently of whether the prime was SVO or VSO (95.53% and 92.30%, respectively).

We used R (R Core Team, 2012) and lme4 (Bates et al. 2012) to perform a Generalized Linear Mixed Model (GLMM) to model the probability of producing a VSO target as a function of the type of prime (SVO vs. VSO). We considered the production of SVO vs. VSO as the dependent variable, choosing the production of SVO as the reference level. As fixed effect, we used the type of prime (SVO vs. VSO). We specified random effects for participant and item. The GLMM-analysis reveals a significant effect of type of prime, indicating that participants tend to produce more VSO following a VSO prime, as reflected by the positive estimate ( $\beta = 3.33$ ,  $SE = 1.23$ ,  $z = 2.70$ ,  $p < .001$ ).

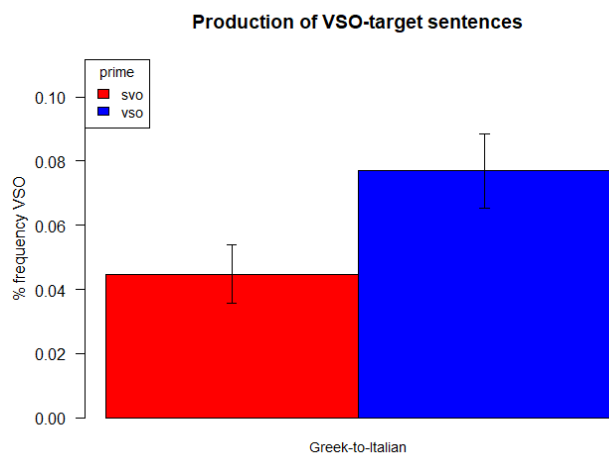


Figure 1. Mean percentage of VSO target sentences produced after a SVO- and a VSO-prime, respectively.

## Discussion

The results show that Greek-Italian bilingual children may produce VSO sentences in Italian, although VSO is not possible in broad-focus contexts in this language. Greek VSO primes enhance the production of VSO in Italian.

The enhanced production of VSO sentences after Greek VSO primes suggests that the activation of a structure in Language A (Greek) may motivate the use of the same structure in Language B, even if the structure is not available in Language B. This suggests that the activation of VSO in Greek motivates cross-linguistic influence from Greek to Italian in the use of VSO. Notably, some

instances of VSO are produced following a Greek SVO prime, too. We propose that this is due to the activation of the Greek language. In other words, the activation of the Greek language *per se* combined with the activation of VSO has a cumulative effect on the production of VSO in Italian, as observed in Figure 1. At this point, this is just a speculation. In Baroncini & Torregrossa (in preparation), the results of the present study are compared to the results of a within-language priming experiment involving the use of Italian only. This will allow us to understand whether VSO sentences are produced also if Greek is not activated.

Finally, it should be noticed that the production of VSO sentences is very limited compared to the production of SVO sentences. This may be due to children's sensitivity to the ungrammaticality of VSO in broad-focus contexts in Italian. However, it is not excluded that children prefer to use SVO in Greek, too. In Baroncini & Torregrossa (in preparation), we aim to solve this issue, by comparing the results of this study with a within-language experiment involving the use of Greek only.

## References

- Hartsuiker, R.J., Pickering, M.J., Veltkamp, E. 2004. Is syntax separate or shared between languages? Cross-linguistic syntactic priming in spanish-english bilinguals. *Psychological Science*, 15(6), 409–414.
- Roussou, A., Tsimpli, I. M. 2006. On Greek VSO again! *Journal of Linguistics*, 42(2), 317–354.
- Serratrice, L. 2016. Cross-linguistic influence, cross-linguistic priming and the nature of shared syntactic structures. *Linguistic Approaches to Bilingualism*, 6(6), 822–827.
- Sharwood Smith, M., Truscott, J. 2014. *The Multilingual mind: a modular processing perspective*. Cambridge: Cambridge University Press.
- Torregrossa, J., Andreou, M., Bongartz, C. 2020. Variation in the use and interpretation of null subjects: A view from Greek and Italian. *Glossa* 5(1): 95. 1-28.
- Vasilyeva, M., Waterfall, H., Gámez, P.B., Gómez, L.E., Bowers, E., Shimpi, P. 2010. Cross-linguistic syntactic priming in bilingual children. *Journal of Child Language*, 37(5), 1047–1064.

# Inverse reaction time as an awareness measure in artificial grammar learning experiments

Tsung-Ying Chen

Department of Foreign Languages and Literature, National Tsing Hua University,  
Taiwan

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Artificial grammar learning (AGL) experiments are frequently adopted to test implicit (unconscious) learning of linguistic generalizations by adult learners. Since explicit (conscious) learning seems inevitable for adult learners, awareness measures are necessary to assess if learners are conscious of acquired linguistic generalizations. Confidence level, for instance, would be uncorrelated or inversely correlated with learning performance if acquired linguistic knowledge is implicit. In this study, we examined if inverse reaction time (IRT) could be a fine-grained and objective confidence-based awareness measure in adult AGL experiments. With data from three phonological AGL experiments recruiting adult participants, we confirmed that confident responses are faster than unconfident ones and that slower (i.e., less confident) responses reflect the application of successfully acquired implicit knowledge.

Keywords: artificial grammar learning, inverse reaction time, awareness, confidence, phonology

## Introduction

In recent years, linguists have been using artificial grammar learning experiments (AGL; Reber 1967) to test various hypotheses regarding implicit (unconscious) and explicit (conscious) learning of linguistic generalizations. In typical AGL experiments, learners are exposed briefly to language input without being informed of hidden linguistic regularities and an upcoming test of their learning performance. Successful extension of hidden regularities to novel forms in the test by learners without awareness of the acquired generalizations would be viewed as evidence for implicit learning. Awareness measure is thus vital to the analysis of experimental results in AGL studies. In particular, for adult participants, artificial language learning is equivalent to L2 learning and an explicit learning process seems inevitable (e.g., Hulstijn 2005). One widely accepted awareness measure is the confidence level (CL) of learners' responses in test sessions. If learners' high response accuracy is uncorrelated or inversely correlated with their CL, the acquired knowledge is assumed as implicit (i.e., zero-correlation criterion; Dienes 2007). However, *subjective* CL does not always reflect adult learners' awareness of target knowledge as the learners, albeit with explicitly acquired target knowledge, could be too humble to report their certainty (e.g., Maie & DeKeyser 2020).

In this study, we thus propose to use inverse reaction time (IRT;  $-1000/\text{RT}$  in milliseconds) as a confidence-based awareness measure. IRT is more normally distributed than other response latency measures such as log-RT (Brysbaert & Stevens 2018) and comes with two primary advantages over subjective CL. First, IRT directly reflects learners' hesitation that arises from their CL, which is thus more *objective* than learners' own post-response CL ratings. Second, IRT is a gradient scale that can capture subtle changes in the effects of CL. In the rest of this squib, we will triangulate the relationship between IRT, CL, and implicit/explicit learning in adult AGL experiments with data from the author's three phonological AGL studies.

### Three phonological AGL studies

The three phonological AGL studies share a similar core experimental design, namely exposing *adult* learners to auditory training input and later testing them with novel items in an auditory acceptability judgment task. Comparisons were always made between two groups exposed either to training input supporting the learning of presumably implicit phonological knowledge or to input that did not. Learning performance was assessed in a two-alternative forced-choice acceptability judgment task in the three experiments. Core hypotheses in each study are detailed below.

In Chen (2020), the author tested if human learners have an inductive bias against a non-domain-final rising tone (R), a constraint that is phonetically natural and assumed to be implicit phonological knowledge. 53 participants were either exposed to disyllabic input without any non-final R (target group) or to disyllabic input without any non-final high tone (non-target group) – the latter tonal gap is phonologically arbitrary and perhaps unlearnable. Accordingly, only the target group was expected to demonstrate the implicit learning of the target constraint. In Chen (submitted), the learnability of the Obligatory Contour Principle (OCP) prohibiting adjacent level tones (OCP-Level; e.g., \*HL-LH, \*H-HL) was compared to that of OCP banning same adjacent tonal *unit* (OCP-Unit; e.g., \*HL-HL, \*LH-LH), and only widely attested OCP-Level was assumed to be learnable implicit phonological knowledge. 60 participants were exposed either to disyllabic input without adjacent tonal levels (target group) or to input without adjacent identical tonal units (non-target group), and only OCP-Level was expected to be acquired as an implicit generalization. Finally, in Chen (in prep.), the main research question is whether a skewed distribution of learning input in favour of shorter words could foster the implicit learning of local vowel harmony (VH) patterns. 60 adult learners were exposed to learning input with VH, but one group perceived more disyllabic items than trisyllabic ones (target group) and the other group listened to an equal number of disyllabic and trisyllabic items (non-target group). Only the target group was expected to converge on an implicit local VH grammar and extend the grammar to novel trisyllabic word forms.

## Data analysis

A total of 1,956 responses from the test sessions in Chen (submitted) and Chen (in prep.) were provided with binary CL ratings, which allowed us to establish the link between CL and IRTs. We regressed raw IRTs against Group (target vs. non-target), CL (yes vs. no), and their interaction in linear mixed-effects modelling, which indicated a significant main effect of CL ( $\beta = 0.173$ ,  $se = 0.023$ ,  $t = 7.67$ ,  $p < .001$ ). That is, IRTs were more negative (i.e., *faster*) for confident responses than for unconfident ones (Figure 1, left panel).

We then incorporated another 2,958 responses without CL ratings from the test sessions in Chen (2020) and coded binary response accuracy for all 4,914 responses for target and non-target groups depending on whether participants did correctly reject novel items violating learnable/unlearnable tonal constraints or extend local VH successfully to novel items. Our main prediction, according to the zero-correlation criterion, is that response accuracy does not correlate or is inversely correlated with IRTs only for the target groups that were expected to acquire implicit linguistic knowledge. To test the prediction, we regressed response accuracy against IRT ( $\xi$ -scored within participants), Group, and their interaction in logistic mixed-effects modelling, which indicates a significant IRT  $\times$  Group interaction ( $\beta = -0.106$ ,  $se = 0.044$ ,  $\xi = -2.401$ ,  $p = .016$ ; Figure 1, right panel). Crucially, for the target groups, a higher IRT predict a higher (and above-chance) response accuracy, whereas the chance-level accuracy regardless of IRT suggests no sign of learning for the non-target groups. The above findings are in line with our main prediction.

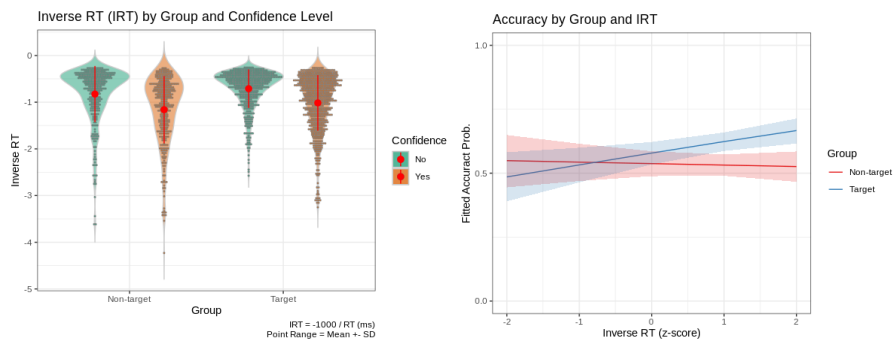


Figure 1. Descriptive raw IRT distributions by Group and Confidence Level (left), and response accuracy by Group and  $\xi$ -scored IRT (right).

## Discussion

With data from three adult phonological AGL experiments that were analyzed with IRT, we showed that (i) response latency is significantly shorter for *subjectively* confident responses, and (ii) slower responses were found to be more accurate only for the target groups expected to learn implicit phonological

knowledge. IRT should thus be viewed as a reliable and objective index of CL, which could in turn reflect the application of implicit/explicit knowledge in adult AGL studies. We speculate that the responses of adult learners in an acceptability judgment task in AGL studies are guided first by explicitly learned knowledge, which may or may not coincide with patterns hidden in the training input. The chance-level accuracy of more confident responses from the target groups (Figure 1, right panel), for instance, could be a case of applying an incorrect explicit generalization to their acceptability judgment. When an explicit generalization is not helpful for learners to make a definite decision, they hesitate and allow their intuition based on implicitly acquired target knowledge to lead the way. However, learners exposed to an unlearnable pattern or assigned to a condition that does not facilitate learning (e.g., the *non-target* groups) may simply fail to demonstrate both types of knowledge.

### Acknowledgements

The current study is funded by the Ministry of Science and Technology, Taiwan (108-2410-H-007-030-MY3) and has benefited from valuable discussions with James Myers.

### References

- Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, 1(1), 9.
- Chen, T.-Y. (2020). An inductive learning bias toward phonetically driven tonal phonotactics. *Language Acquisition*, 27(3), 331–361.
- Chen, T.-Y. (submitted). On the learnability of level-based and unit-based tonal OCP constraints: An artificial grammar learning study. Ms., National Tsing Hua University.
- Chen, T.-Y. (in prep.). A starting-small effect in the learning of opaque and transparent vowel harmony. Ms., National Tsing Hua University.
- Dienes, Z. (2007). Subjective measures of unconscious knowledge. *Progress in Brain Research*, 168, 49–64.
- Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, 27(2), 129–140.
- Maie, R., & Dekeyser, R. M. (2020). Conflicting evidence of explicit and implicit knowledge from objective and subjective measures. *Studies in Second Language Acquisition*, 42(2), 359–382.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6), 855–863.

# Sentence comprehension assessment in Russian

Daria Chernova<sup>1</sup>, Artem Novozhilov<sup>1</sup>, Natalia Slioussar<sup>1,2</sup>

<sup>1</sup>Institute for Cognitive Studies, St Petersburg University, Russia

<sup>2</sup>HSE University, Moscow, Russia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

We developed a Sentence Comprehension Test in Russian for syntactic competence assessment. It includes 60 unambiguous grammatically complex sentences of several types with comprehension questions aimed to test effective syntactic processing. The test does not show ceiling effects with adult native speakers. The results of the test correlate with verbal working memory span.

Keywords: syntactic complexity, sentence comprehension, Russian language

## Introduction

The ability to parse syntactically complex sentences efficiently is a crucial skill for text comprehension, and numerous studies demonstrated considerable by-subject variability in performance on syntactic processing tasks (Farmer et al. 2012). However, while tests on sentence processing are widely used with various neurologically impaired populations, tests for healthy adult native speakers are difficult to develop due to ceiling effects and have been created only for English so far (Acheson 2008; Dabrowska 2018). In the present paper, we present the Sentence Comprehension Test we developed for Russian.

## Materials

We created 60 unambiguous grammatically complex sentences of six types. These types are difficult to process according to previous experimental studies on Russian and other languages, see (1)-(6). All sentences were semantically reversible and unbiased. For every sentence, we created a question with a choice of two answers aimed to assess syntactic structure comprehension.

- (1) object relative clauses (see Price & Witzel 2017, Malyutina et al. 2018)

*Svidetel', kotorogo upomjanul v svoej rechi istec,*  
witness-NOM whom-ACC mentioned in his speech claimant-NOM  
*vskoobil so svoego mesta v zale suda.*  
jumped from his seat in room court

‘The witness that the claimant mentioned jumped up from his seat in the courtroom.’

Question: Who was mentioned?

Response options: A) the witness B) the claimant

- (2) spatial constructions (see Laurinavichyute et al. 2017)  
*Passazhir sprjatal v seryj jashbik kozhanyj chemodan.*  
passenger-NOM hid into grey box-ACC leather suitcase-ACC.  
‘The passenger hid the leather suitcase into the grey box.’  
Question: What was hidden where?  
Response options: A) the box in suitcase B) the suitcase in the box
- (3) temporal constructions (see Fedorova 2005)  
*Pered tem kak Tolja propylesit pol, Julja vyguljaet sobaku.*  
before Tolja vacuum-cleans floor Julia walks dog  
‘Before Tolja vacuum cleans the floor, Julia will walk the dog.’  
Question: What happens first?  
Response options: A) Tolja vacuum cleans the floor B) Julia walks the dog
- (4) sentences with high adjunct attachment in a complex noun phrase *Konvert*  
*peredali pomoshniku detektiva, sledivshemu*  
envelope-ACC gave assistant-DAT detective-GEN following-DAT  
*za podozrevaemym.*  
after suspect-INST  
‘The envelope was given to the assistant<sub>i</sub> of the detective<sub>j</sub>, following<sub>i</sub> the suspect.’  
Question: Who followed the suspect?  
Response options: A) the detective B) the assistant
- (5) sentences with low adjunct attachment in a complex noun phrase *Notarius napisal*  
*nasledniku millionera, zhivshego za granicej.*  
notary wrote heir-DAT millionaire-GEN living-GEN abroad  
‘The notary wrote to the heir<sub>i</sub> of the millionaire<sub>j</sub> living<sub>j</sub> abroad.’  
Question: Who lived abroad?  
Response options: A) the millionaire B) the heir
- (6) comparative constructions  
*Sherstjanaja jubka dlinnie shelkovej, no korоче l'njanoj.*  
woolen-NOM skirt-NOM longer silk-INST but shorter linen-INST  
‘The woolen skirt is longer than the silk one, but shorter than the linen one.’  
Question: Which skirt is longer?  
Response options: A) the silk one B) the linen one
- The test also included 40 filler sentences with a simpler syntactic structure (7).
- (7) *Na ploshhadke ja vstretil brata moego druga s bol'shoj sobakoj*  
at playground I met brother-ACC my friend-GEN with big dog  
‘At the playground I met my friend’s brother with a big dog.’  
Question: Who did he meet?  
Response options: A) his brother’s friend B) his friend’s brother



## Method

42 native speakers of Russian (29 female, age 19-32) took part in the experiment. The word-by-word self-paced reading methodology was used because otherwise the task would be too easy. The incorrect response was always a noun that was mentioned in the sentence, so the grammatical structure of the sentence had to be analyzed to give a correct answer. The order of the response options was random. Accuracy, word by word reading times, and time to give the answer were registered. We also measured participants' verbal working memory span using the Russian adaptation (Fedorova 2003) of the test by (Daneman&Carpenter 1980) as syntactic processing was shown to be affected by the characteristics of the working memory (Caplan & Waters 1999).

## Results and discussion

For the statistical analysis, we used logistic and linear mixed-effects regressions with random intercepts and slopes by participant and by item and Tukey's tests for post hoc comparisons. Test sentences were significantly more difficult to process than fillers, both in terms of correct answers (80.6% vs. 92.6% on average;  $\beta = 0.25$ ,  $SE = 0.04$ ,  $t = 6.01$ ,  $p < 0.01$ ) and reading and response times (711.4 ms vs. 595.6 ms,  $\beta = 0.25$ ,  $SE = 0.04$ ,  $t = 6.01$ ,  $p < 0.01$ ; 3484.2 ms vs. 3096.3 ms,  $\beta = 0.25$ ,  $SE = 0.04$ ,  $t = 6.01$ ,  $p < 0.01$ ). This proves the validity of the test.

Another important proof comes from the fact that we detected significant variation between participants. In target sentences, they made from 1 to 24 errors (98%-60% correct answers). We found a significant correlation between answer accuracy and working memory span test scores ( $r=0.59$ ,  $p<0.01$ ). The number of errors in filler sentences did not vary that much (from 0 to 8, which means 100%-80% correct answers, with  $\frac{3}{4}$  participants making no more than two errors).

There were also significant differences between most construction types. High/low attachment and comparative constructions had longer word-by-word reading times than other target sentence types ( $t > 6.74$ ,  $p<0.01$  for all pairwise comparisons). Spatial and comparative constructions had longer response times than other sentence types ( $t > 5.58$ ,  $p<0.01$  for all pairwise comparisons).

Finally, high and low attachment sentences triggered the largest number of incorrect responses (74.3% and 62.6%, respectively). Low attachment sentences were significantly different from all other types (except for high attachment,  $t > 3.78$ ,  $p<0.01$  for all pairwise comparisons), while for high attachment sentences, only some comparisons gave significant results. Importantly, every target type was significantly different from fillers, except for temporal one ( $t > 3.61$ ,  $p<0.01$  for all pairwise comparisons), which further proves the validity of the test.

Thus, sentence types that take more time to read and especially to answer are not the ones in which participants make more mistakes. This may point to two

different aspects of syntactic complexity. In some cases, arriving at any coherent interpretation is difficult; in the other cases, one arrives at some interpretation easily, but often this is not the correct one. The generalizations we made about different constructions may be useful for further processing studies. To conclude, our pilot study showed that the test we developed is far from trivial for the participants and does not show ceiling accuracy. After validation, it can be used in various studies as a tool to measure syntactic processing efficiency.

### Acknowledgements

The study is supported by RSF, project № 21-78-00064.

### References

- Acheson, D. J., et al. 2008. New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, 40, 278–289.
- Caplan, D., Waters G. 1999. Verbal Working Memory and Sentence Comprehension. *Behavioral and Brain Sciences*, 22, 1, 77-94.
- Dąbrowska, E. 2018. Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, 178, 222–235.
- Daneman, M., Carpenter P. 1980. Individual Differences in Working Memory and Reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 4, 450–466.
- Farmer, T.A., Misyak, J B., Christiansen, M.H. 2012. Individual differences in sentence processing. In M. J. Spivey, K. McRae, & M. F. Joannis (Eds.), *Cambridge handbooks in psychology. The Cambridge handbook of psycholinguistics* (p. 353–364). Cambridge University Press.
- Fedorova, O.V. 2003. Test po opredeleniju ob’ema operativnoj pamjati: istorija i sovremennoe sostojanie (in Russian, ‘Survey of the State of the Art in Verbal Span Test’). *Computational Linguistics and Intellectual Technologies*.
- Fedorova, O.V. 2005. Pered or posle: chto proshche. (in Russian, ‘Before or after: What is easier’). *Voprosy Jazykoznanija*, 2005, 6, 44-58.
- Laurinavichyute, A., Chrabaszcz, A., Farizova, N., Tolkacheva, V., Dragoy, O. 2017. Vlijanie sensomotornyh stereotipov na ponimanie prostranstvennyh konstrukcij: dannye dvizhenij glaz (in Russian, ‘The influence of sensorimotor stereotypes on the understanding of spatial constructions: eye-tracking evidence’). *Voprosy Jazykoznanija*, 3, 99-109.
- Malyutina, S., Laurinavichyute, A., Terekhina, M., Lapin, Y. 2018. No evidence for strategic nature of age-related slowing in sentence processing. *Psychology and Aging*, 33(7), 1045–1059.
- Price, I.K., Witzel J. 2017. Sources of Relative clause processing difficulty: Evidence from Russian. *Journal of Memory and Language*, 97, 208–244.

# Gradient acceptability between naïve and expert linguistic intuitions

Jacee Cho<sup>1</sup>, Yafei Li<sup>2</sup>, Rebecca Shields<sup>2</sup>

<sup>1</sup>Department of English and Language Sciences, University of Wisconsin-Madison, US

<sup>2</sup>Language Sciences, University of Wisconsin-Madison, US

## Abstract

The current study examines non-linguists' judgments about one structure which was reported in Li et al. (2012): interaction of adverbs in *wh*-movement which involves rare sentence types. Acceptability judgment data collected from 199 nonlinguists differ from the expert intuitions reported in Li et al. (2012). We argue that sentence acceptability judgments are necessarily influenced by extraneous factors; therefore, studying sentences that are hard to process or rare requires introspective judgments from linguists who can carefully contrast sentences under investigation by ignoring irrelevant information.

Keywords: gradient acceptability, introspective judgments

## Introduction

Some recent empirical studies have reported that nonlinguists' judgments may differ from those of linguists or those reported in the syntactic literature (Dąbrowska 2010; Gibson & Fedorenko 2013). In particular, Gibson and Fedorenko (2013) showed cases where linguists' intuitions reported in the literature are not shared by the general population. Gibson and Fedorenko (2013) proposed a number of possible reasons why linguists' judgments might be faulty and argued that quantitative data collected from a large number of naïve participants should be used to study syntax. Following this line of inquiry, the current study examines non-linguists' judgments about one structure which was reported in Li et al. (2012): interaction of adverbs in *wh*-movement (Section 2). Acceptability judgment data collected from 199 nonlinguists differ from the expert intuitions reported in Li et al. (2012). However, we do not think that these findings necessarily indicate that linguists' judgments reported in Li et al. (2012) are faulty or biased. Rather, we argue that linguists are better at ignoring irrelevant factors in judging grammaticality, in particular implausible or rare but grammatical sentences. Therefore, we argue that experts' introspective judgments are necessary to study certain types of sentences.

## Linguistic phenomenon

A well-known phenomenon in the study of syntax is Relativized Minimality (Rizzi 1990), which states roughly that a syntactic constituent cannot move across another constituent of the same type. This principle was claimed to affect adverb topicalization in a variety of languages in Rizzi (2001), based on the fact that the topicalization of an adverb is blocked by the presence of an intervening adverb, as in (1).

- (1) Topicalization: A' movement blocked by an intervening A' modifier
- a. Noisily, Sam (\*frequently) mows his lawn.
  - b. Noisily, Sam (\*apparently) mows his lawn.

Crucially, this principle is “Relativized” according to syntactic type, so that elements in argument (A) position block movement to another argument position, and elements in non-argument (A') position block movement to another non-argument position. Li et al. (2012) reported a more complex pattern of blocking with *wh*-movement of adverbs (1). Specifically, they noted that adverbs which are not themselves capable of undergoing *wh*-movement do NOT block *wh*-movement across them, despite being located in an A' position (1a). This pattern is unexpected assuming traditional Relativized Minimality, which would have predicted both (2a) and (2b) to be unacceptable.

- (2) Wh-movement: A' movement blocked by some intervening A' modifiers, but not others
- a. How noisily does Sam (\*frequently) mow his lawn?
  - b. How noisily does Sam (apparently) mow his lawn?

This study was based largely on the judgements of four native speakers of English, who were either PhD students or faculty members in linguistics.

## Methodology

We subsequently tried to replicate the empirical findings of this earlier study on a larger scale ( $n = 199$ ) using an acceptability judgment task with a 7-point Likert scale. To test the claim about the interaction between specific adverb combinations and *wh*-movement, we designed an experiment by manipulating two factors, sentence type (*wh*-question vs. declarative) and adverb combination (wh-wh vs. nonwh-wh), as illustrated in Table 1. The target condition is shaded.

Table 1. 2x2 factorial design.

Sentence type Adv combination	Declarative	<i>Wh</i> -question
Wh-wh	Lucas frequently sprayed the paint thickly.	*How thickly did Lucas frequently spray the paint?
Nonwh-wh	Lucas apparently sprayed the paint thickly.	How thickly did Lucas apparently spray the paint?

### Results

All scores including fillers were transformed into z-scores for statistical analysis. The z-score mean for each of the four conditions is presented in Figure 1.

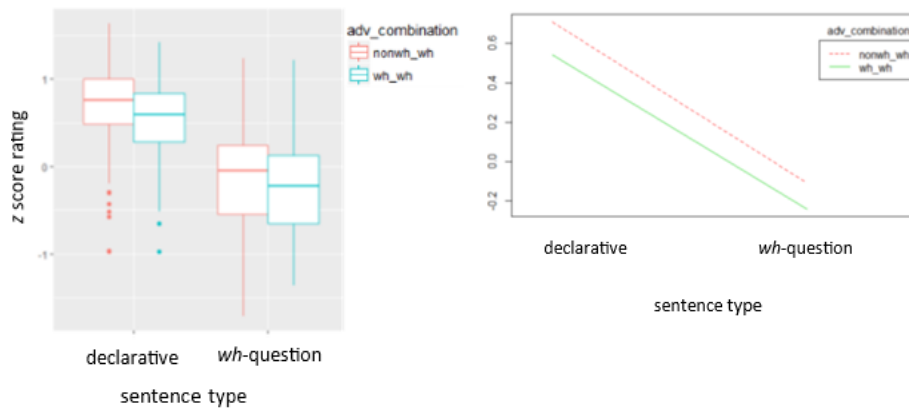


Figure 1. Mean z-scores for each condition. Figure 2. Interaction plot.

A linear mixed-effects regression analysis was conducted on z-transformed rating scores with Sentence type (*Wh*-question vs. Declarative) and Adv. Combination (Frequency-Manner (wh&wh) vs. Evidential-Manner (nonwh & wh)) as fixed factors. Subject and Item were included as random factors. Results indicated that there were significant main effects of Sentence type (Estimate=-.7, SE=.09, df=68, t=-8.59, p <.0001) and Adv. Combination (Estimate=-.2, SE=.07, df=207, t=-2.9, p <.003). Declarative sentences were rated higher than *wh*-questions and sentences containing evidential & manner adverbs were rated higher than sentences with frequency (wh-adverb) & manner (wh-adverb) adverbs. However, there was no significant interaction effect of these two factors, as shown in the interaction plot and the differences-in-differences (DD) score in Figure 2.

## Discussion and conclusion

Gradient acceptability judgments between linguists and nonlinguists are largely due to the fact that they attend to different types of information in judging the acceptability/grammaticality of sentences (Francis, in press). Since sentence acceptability judgments are necessarily influenced by extraneous factors such as sentence complexity and frequency of occurrence “we cannot rely on non-expert judgements in the absence of theory-based reasoning and carefully-controlled stimuli” (Juzek et al., 2020, p.6). Therefore, it is necessary to have both quantitative data from non-linguists and judgments from linguists who can carefully contrast sentences under investigation by ignoring irrelevant factors.

## References

- Dąbrowska, E. 2010. Naïve v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review* 27, 1-23.
- Francis, E. (in press). *Gradient Acceptability and Linguistic Theory*. Oxford University Press.
- Gibson, E., Fedorenko, E. 2013. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28(1/2), 88-124.
- Juzek, T., Francis, E., Häußler, J. 2020. Syntax-semantics interferences: making a case for expert reasoning and multi-method approaches. *The Limits of Experimentation: Current Challenges in Experimental Linguistic Practice*.
- Li, Y., Shields, R., Lin, V. 2012. Adverb classes and the nature of minimality. *Natural Language & Linguistic Theory* 30(1), 217-260.
- Rizzi, L. 1990. *Relativized Minimality*. Cambridge, Mass.: MIT Press.
- Rizzi, L. 2001. Relativized minimality effects, in Mark Baltin and Chris Collins (eds.), *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell. 89-110.

# Acoustic and semantic processes during speech segmentation in French

Mar Cordero-Rull<sup>1</sup>, Stéphane Pota<sup>2</sup>, Elsa Spinelli<sup>2</sup>, Fanny Meunier<sup>1</sup>

<sup>1</sup>Université Côte d'Azur, CNRS, BCL, France

<sup>2</sup>Université Grenoble Alpes, CNRS, LPNC, France

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

We designed two experiments that tested the listeners' perceptual capacities during online segmentation of homophonic word boundaries while processing sentential information. In French, listeners often use variations in fine acoustic indices to detect word beginnings. We measured event-related potentials (ERPs) evoked by phonemically identical sequences, such as *l'affiche* ("the poster") and *la fiche* ("the sheet"), both [lafɛʃ], which were contained in either congruent or incongruent sentences. Results showed that although listeners can detect acoustic variations in homophonic sequences, these may not be salient enough when contextual information is also present. Shifting attention from sentence meaning (Task 1) to lexical information (Task 2), enhanced the listeners' perception of fine-grained acoustic details. Thus, top-down processes are likely to modulate speech perception and segmentation.

Keywords: speech segmentation, sentence processing, acoustic cues, ERP

## Introduction

To achieve successful comprehension, mappings are drawn from acoustic signal into lexical representations in memory (Davis, Rodd, 2011). Word boundaries can be detected by fine-grained acoustic details (Friederici, Wessels, 1993), which will depend on the listener's language. French listeners often use variations of acoustic information in content word beginnings to detect word boundaries. Several studies reveal that they can discriminate homophonic sequences such as *l'affiche-la fiche*, both [lafɛʃ] (Spinelli et al., 2007, 2010). Yet, both acoustic and contextual information are embedded during processing of ambiguous speech (Shoemaker, 2014).

In our study, we explored how French listeners use linguistic cues in processing ambiguous word boundaries within a specific context. To answer this, we presented homophonic sequences (HS) such as *l'affiche-la fiche* embedded in naturally produced sentences. Contextual information was presented under 3 different conditions: control, congruent, and incongruent. In Task 1, we used a semantical judgement task to detect whether French listeners can process fine-grained acoustic details from HS within a given context. In Task 2, we examined the cognitive cost of segmenting ambiguity during sentence processing. In a cross-modal identification task, attention was guided towards lexical instead of sentential information. To assess the listeners'

capacity to discriminate fine-grained acoustic details from HS, the N400 was considered as a relevant neurophysiological marker of interest. In ERP research, N400 is a negative component reflecting difficulties of lexical semantic integration for incongruent compared to congruent conditions, usually found between 200-600 ms after stimulus onset (Kutas, Federmeier, 2011). If fine-grained acoustic information is encoded during speech processing, listeners should be able to detect a semantic violation, yielding an N400 effect for incongruent context. If, however, HS' meaning is accessed through contextual cues, no N400 effect is expected since incongruent context will not be perceived as a semantic violation.

## Method

### Participants

Forty-nine native speakers of French took part in this study (Task 1: 21 participants, 11 females,  $21 \pm 2.42$  years old; Task 2: 28 participants, 22 females,  $22.2 \pm 4.3$  years old). All were right-handed and did not report any psychiatric, neurological, or hearing disorders.

### Stimuli

In both tasks, 144 sentences were presented: 48 control, 48 congruent, 48 incongruent. Control sentences consisted of grammatical sentences containing semantic violations of incoherent target words. For congruent and incongruent sentences, we used 24 pairs of HS in French (e.g., *l'ami-la mie*). They were included in sentences with a semantic constraint that either favored access to the meaning of the presented candidate (congruent) or in favor of the other candidate in the pair (incongruent). Sentences were recorded by a French female speaker, naïve to the study goals. To avoid acoustic differences, homophone candidates were excised from original sentences and introduced in either incongruent contexts (*cross-splicing*) or congruent contexts (*identity-splicing*). Sentence length ranged between 6 and 13 words ( $8.54 \pm 1.48$ ).

### Procedure

In Task 1, a written sentence appeared after a set of 6-13 spoken sentences, after which participants had to decide whether the written sentence was semantically related to the last sentence they have heard. In Task 2, a written word was presented after each auditory sentence and participants had to decide as quickly and as accurately as possible whether the written word was present in the auditory sentence. The selected written words were either present in the sentence (37.1% of the case), absent in the sentence (37.1%), or the target word homophone from the auditory sentence (25.8%).

EEG recordings were acquired using SynAmps2/RT amplifier (Compumedics), recorded from 64 active channels (Neuroscan Quik-Cap Neo Net), following the standard international 10/20 positioning system.



## **EEG pre-processing and analysis**

EEG data were pre-processed in MATLAB and analysed in R. A 0.1-48 Hz bandpass filter was applied. Epochs were extracted for a span of -200 to 1100 ms relative to article onset. Bad channels were interpolated, and blink and muscle artifacts were corrected. Reference was recomputed to average. Time windows of interest were selected based on global field power (Lehmann & Skrandies, 1980). Selected time-windows corresponded to 0-275, 275-550, 550-870, 870-1100 ms after critical word onset. Using the same time windows across tasks enabled comparisons. Signal amplitude was analysed through permutation tests carried separately for each time window.

## **Results in task 1**

Between 275-550 ms, a negative activity spread over frontocentral electrodes slightly right lateralized for control context, indicating the presence of an N400 component. It reached statistical significance only compared to congruent context. Compared to incongruent context, differences in the negative amplitude reached significance for central right electrodes (870-1100 ms).

## **Results in task 2**

An N400 component for control context was observed broadly (550-870 ms); then localized in frontocentral sides (870-1100 ms). For incongruent context, the N400 effect was localized in frontal scalp sides for both 550-870-ms and 870-1100-ms time windows. Significant differences were found between congruent and incongruent contexts over left frontal scalp sides.

## **Conclusions and general discussion**

In our paradigm, sentential information and HS were pointing to either the same (congruent condition) or different meaning interpretations (incongruent condition). Even though different acoustic indices were associated to HS, no differences were found in the processing of congruent and incongruent contexts (Task 1). Nevertheless, congruent and incongruent contexts yielded an N400 amplitude in Task 2, showing different amplitude sizes relative to their difficulty in integration. Although this effect can also be attributed to an effect of task (Kutas, Federmeier, 2011), results reflect the weighting of multiple segmentation strategies. This reveals the possible influence of a cue hierarchy based on individual saliency during speech processing (Rodd 2018; Mattys, 2003). Top-down processes involving attention may influence the perceptual processing of HS in continued speech. Differences in tasks 1 and 2 suggest that listeners may not use acoustic variations for lexical access when higher-level information (semantic) is salient; whereas fine-grained acoustic details may be used when attention is shifted towards lexical information. Thus, processing acoustic cues may not lead to reinterpretation of meaning when context is

present (Mattys et al., 2005), reflecting an enhancement of top-down contextual information over bottom-up perception.

## Acknowledgements

This work was supported by the French government through the UCA-Jedi project managed by the National Research Agency (ANR-15-IDEX-01) and by the CNRS Service and Research Unit Maison des Sciences de l'Homme et de la Société Sud-Est (MSHS Sud-Est, USR 3566) of the Université Côte d'Azur. We thank Ambre Denis-Noël for her contribution to the project.

## References

- Davis, M.H., Rodd, J. M. 2011. Brain structures underlying lexical processing of speech: Evidence from brain imaging. In Gaskell, G., Zwitserlood, P. (eds.) 2011, *Lexical representation: A multidisciplinary approach*, 197-230. Berlin: Mouton de Gruyter.
- Do Carmo-Blanco, N., et al. 2019. Processing of non-contrastive subphonemic features in French homophonous utterances: An MMN study. *J. Neurolinguistics*, 52.
- Friederici, A. D., Wessels, J. M. I. 1993. Phonotactic knowledge of word boundaries and its use in infant speech perception. *Percept. Psychophys.* 54, 287-295.
- Kutas, M., Federmeier, K. D. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.*, 62, 621–647.
- Lehmann., D., Skrandies, W. 1980. Reference-free identification of components of checkerboard evoked multichannel potentials fields. *Electroencephalogr. Clin Neurophysiol.* 48: 609–21.
- Mattys, S. L. 2003. Stress-based speech segmentation revisited. *Proceedings of Eurospeech: The 8th Annual Conference on Speech Communication and Technology*, 121–124. Geneva, Switzerland.
- Mattys, S. L., et al. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *J. Exp. Psychol. Gen.*, 134, 477-500.
- Rodd, J. M. 2018. Lexical Ambiguity. In S.-A. Rueschemeyer & M. G. Gaskell (eds.) 2018, *Oxford Handbook of Psycholinguistics*, 120-144. Oxford, UK: Oxford University Press.
- Spinelli, E., et al. 2010. An intonational cue to word segmentation in phonemically identical sequences. *Atten. Percept. Psychophys.* 72, 775-787.
- Spinelli, E., et al. 2007. Fine-grained access to targets and competitors in phonemically identical spoken sequences: the case of French elision. *Lang. Cogn. Process.*, 22:6, 828-859.

# Can locus equations model dialect-specific variation in coarticulation?

Giovina Angela del Rosso

Department of Fil.Co.Spe., Roma Tre-Sapienza University, Italy

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Cross-linguistic coarticulatory differences are experimentally challenging to disentangle. This paper aims to assess whether traditional locus equations constitute a suitable heuristic tool to highlight regional differences in coarticulation patterns. For this purpose, three highly similar dialects were chosen, i.e., Bari, Cagliari, and Palermo Italian. C-V effects were examined in CV and VC sequences (V=/a/; C=/p, t, k, f, s, m, n/). A two-step acoustic analysis was performed on 648 phonetically controlled utterances, read by 24 speakers, and sampled from the CLIPS corpus. Subsequently, traditional locus equations and alternative linear regressions were compared. The results show that locus equations do not adequately model variation of subtle dynamic patterns. Yet significant dialect-specific differences in coarticulation emerged fitting alternative linear regressions.

Keywords: acoustic phonetics; locus equations; coarticulation; dialect; Italian.

## Introduction

This paper aims to investigate whether locus equations represent a proper heuristic method to highlight significant cross-dialectal differences in the acoustics of coarticulation patterns among highly similar regional dialects of Italian. Coarticulation permeates speech models since its dynamics are bonded to the long-standing issues of variability and segmentation. Its patterns were held to be constrained by articulatory and aerodynamic demands, though they were later shown to be partly language-specific (Hardcastle & Hewlett 1999). Nonetheless, identifying language-specific differences is problematic, due to the flawed comparability of phonetic material across different languages. For this reason, three highly similar language varieties are examined in this study.

Experimentally, a classic method of indirectly assessing the degree of consonant-vowel (C-V) coarticulation is represented by first order locus equations (LEs, henceforth). They also capture articulatory-acoustic relationships, carrying information about places of articulation – see Perillo et al. (2015) and Bang (2017), for reviews. Indeed, LEs are regression lines fitting scatterplots of vowel formant frequencies. F2 values are measured at onset (F2<sub>ON</sub>, y-axis) and midpoint or steady state (F2<sub>MID</sub>, x-axis) across various vowel contexts. Their relation is formalized as  $F2_{ON} = F2_{MID} * k + c$ , where  $k$  is the slope of the regression line (i.e., the change in F2 during the transition) and  $c$  is the intercept between the regression line and y-axis (i.e., F2 at the beginning of

the transition). Steep regression lines indicate a high degree of coarticulation, whereas flatter slopes hint at lower degrees of coarticulation. C-V effects and the steepness of LE slopes decrease from labials to alveolars, while velars exhibit vowel-dependent split patterns (Hardcastle & Hewlett 1999).

## Material and method

Coarticulation patterns were compared across three highly similar regional dialects: Bari, Cagliari, and Palermo Italian. The extensive overlap of their vocabularies and phonological systems makes employing the same set of stimuli possible. The comparability is also enhanced in lexical neighbourhood density and according to the Output Constraints model (Hardcastle & Hewlett 1999).

Read speech material was extracted from the LF subcorpus of the CLIPS (corpus of Italian). Four female and four male speakers per variety were sampled. The words in the dataset meet the following criteria: they are content polysyllabic paroxytones; the Vs are stressed; the Cs are singleton; the phonetic environment was controlled; two or three repetitions per sequence were collected, depending on resources. Overall, 648 CV (/pa, ta, ka, fa, sa, ma, na/) and VC (/at, as, an/) utterances entered the dataset (27 words \* 24 speakers).

The dataset was manually segmented, annotated, and the trajectories of /a/ formant frequencies were semi-automatically analysed in Praat. The data were statistically processed in R. Specifically, the analysis was divided into two stages:

LEs were fitted using F2 frequencies extracted at two time points; F2<sub>ON</sub>/F2<sub>OFF</sub> (of CV and VC boundaries, respectively) and F2<sub>MID</sub>; language variety, manner, and place of articulation were implemented as additive effects;

an additional time point was analysed, i.e., 25% or 75% of the vowel duration of CV and VC sequences, respectively. F1 and F3 were measured at all three time points. These supplementary data were used to fit a series of alternative linear regressions, encompassing the same additive effects as in (I).

Finally, ANOVAs were performed to compare the models and identify the best solutions, which were then checked for overfitting; to measure accuracy and validate the results, the bootstrap resampling method was applied, through 200 random resamples with replacement.

## Results

The LEs plots reveal that the datapoints and regression lines are grouped by place of articulation (Figure 1). The intercepts of labials are lower than those of alveodentals, while velars have the highest intercepts. These results are in line with previous research findings relating LEs, degree of coarticulation, and place of articulation. However, if the variety variable is included, the variability sharply increases and prevents a straightforward detection of general patterns.

This is evidenced in the statistical analysis, summarised in Table 1. The ANOVAs identified M1 as the best LE-based model of stage I, accounting for

59% of the variation. However, M1 does not include language-specific differences since it relies solely on  $F2_{MID}$ , C place, and mode. The inclusion of the variety as an additive effect does significantly improve the predictive power.

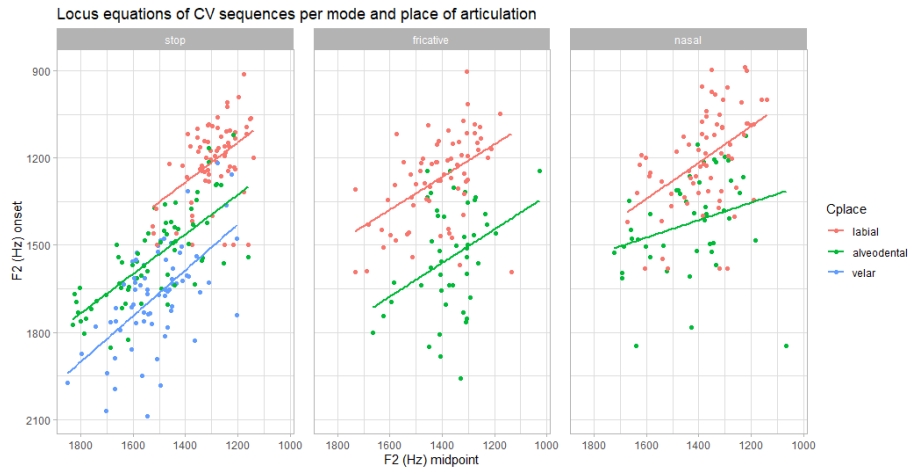


Figure 1. LEs fitting  $F2_{MID}$ - $F2_{ON}$  scatterplots, not divided per variety.

Table 1. Summary of the best linear models.

response		$F2_{ON}$ (CV)	$F2_{25\%}$ (CV)	$F1_{ON}$ (CV)	$F1_{OFF}$ (VC)
model		M1	M2	M3	M4
explanatory variable		$F2_{MID}$	$F2_{MID}$	$F1_{MID}$	$F1_{MID}$
		.54*** (.45, .62)	.82*** (.79, .86)	.44*** (.38, .49)	.69*** (.58, .81)
C place	labial	-222*** (-250, -194)	-80*** (-92, -69)	30*** (16, 44)	N/A
	velar	124*** (83, 165)	71*** (56, 87)	5 (-14, 24)	N/A
C mode	nasal	-82*** (-115, -48)			-74*** (-104, -45)
	stop	-23 (-55, 10)			-6 (-39, 27)
variety	CA		-19** (-31, -6)	22** (7, 38)	66*** (35, 97)
	PA		-15** (-28, -3)	24** (8, 39)	64*** (32, 95)
constant		761*** (633, 889)	282*** (228, 336)	301*** (258, 343)	77 (-15, 169)
$R^2$ (adj. $R^2$ )		.59 (.58)	.85 (.85)	.34 (.34)	.42 (.41)
residual SE		159	66	82	107
F-statistic		128 <sub>(5, 466)</sub> ***	502 <sub>(5, 466)</sub> ***	46 <sub>(5, 466)</sub> ***	27 <sub>(5, 186)</sub> ***

Nevertheless, adding the variety as an effect reveals statistically significant differences in all best models fitted in stage II. In particular, M2 employs  $F2_{MID}$  to predict  $F2_{25\%}$  and models 85% of the variation in the data, yielding the highest accuracy score among the regressions fitted in this study. M3 and M4 are especially noteworthy because they harness the first formant of CV and VC utterances, respectively – M3 uses  $F1_{MID}$  to predict  $F1_{ON}$  and M4 predicts  $F1_{OFF}$  through  $F1_{MID}$ . Finally, all models shown in Table 1 were validated after bootstrap resampling. No signs of overfitting were found; thus, no subsequent adjustments were made.

### Discussion and conclusion

This study confirms that LEs successfully model C-V effects across a variety of contexts, strongly depending on the place of articulation (Hardcastle & Hewlett 1999; Perillo et al. 2015; Bang 2017). Less consistent results were obtained as language-specific effects were evaluated in highly similar dialects.

In stage I, LEs failed to detect cross-dialectal differences, but they were later shown to be significant (stage II). Using more detailed data was crucial to avoid neglecting relevant sociolinguistic variation. Unlike Öhman (1966), this study shows that not only F2 but also F1 trajectories encode information that can be exploited to model regional differences in coarticulatory effects; yet no clear pattern emerged by relying on F3. The alternative regressions mimic the capability of LEs to capture variations between two time points. However, it is not clear whether this involves any articulatory meaning.

These findings support previous research indicating that relevant phonetic information is not exclusively conveyed by few «magic moments» (Carignan et al. 2020: 2). Conversely, relying only on well-known simple metrics runs the risk of oversimplification, especially affecting time-varying phenomena.

### References

- Bang, H.-Y. 2017. The acoustic counterpart to coarticulation resistance and aggressiveness in locus equation metrics and vowel dispersion. *Journal of the Acoustical Society of America* 141(4), EL345–EL350.
- Carignan, C., Hoole, P., Kunay, E., Pouplier, M., Joseph, A., Voit, D., Frahm, J., Harrington, J. 2020. Analyzing speech in both time and space. *Laboratory Phonology* 11(1: 2). 1–26.
- Hardcastle, W. J., Hewlett, N. (eds.) 1999. *Coarticulation: Theory, data and techniques*. New York, Cambridge University Press.
- Öhman, S. 1966. Coarticulation in VCV Utterances: Spectrographic Measurements. *The Journal of the Acoustical Society of America* 39(1). 151–168.
- Perillo, S., Bang, H.-Y., Clayards, M. 2015. Locus equation metrics as an index of coarticulation resistance: The effect of prosodic prominence. *Proceedings of Meetings on Acoustics* 25.

# Argument-adjunct asymmetry in long-distance wh-movement in Russian

Yuliya Demina

Institute of Linguistics, Russian Academy of Science, Russia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This paper investigates an argument-adjunct asymmetry in Russian in wh-questions. Such an asymmetry is described for many languages. In current relevant literature there are different approaches to wh-movement in Russian, but some questions are still unanswered, for example: is there a general argument-adjunct asymmetry in different types of questions? In present study, I conducted three grammaticality judgment experiments in an attempt to clarify the empirical picture regarding argument-adjunct asymmetry. According to the results, the island constraints hold in Russian for *chto*-clauses and indirect questions. *Chtbody*-clauses do not demonstrate island effects. However, it was also observed that there exists argument asymmetry which is valid for all types of indirect questions.

Keywords: argument-adjunct asymmetry, long-distance wh-movement, Russian, experimental study

## Argument-adjunct asymmetry

The purpose of this study is to find out how the asymmetry of arguments and adjuncts is manifested when they are extracted out of various types of interrogative structures in Russian, and also to consider the influence of interrogative structure type on the extraction's availability.

Previous works devoted to the problem of long-distance wh-movement in Russian, mainly describe subject-object asymmetry (Antonenko, 2010), or the asymmetry caused by the type of embedded clause complementizer. Hence, for example, some researchers (Khomitsevich 2008; Dyakonova 2009) claim that there is an asymmetry that occurs when the wh-phrases are extracted out of *chto*- ('what') and *chtbody*- ('why') embedded clauses. All researchers notice the impossibility of components extraction from an embedded clause with a complementizer *chto*. As for *kak* ('how') -clauses, Baylin in (Bailyn 2018) notes that their properties similar to *chto*-clauses. Extraction of wh-adjuncts and wh-subjects from it is unacceptable, whereas the extraction of wh-arguments is marginal. The marginality, apparently, is caused by the variability of the native speaker's judgments.

The asymmetry of extraction from the indirect question island is also well known (Boeckx 2012). An indirect question with an interrogative pronoun in Russian is also considered as a weak island, and its properties depend on the

type of extraction (Lyutikova 2009: 466). As for wh-movement, this type of question also shows an asymmetry: arguments can be extracted, but not adjuncts. A general indirect question in Russian is formed with *li* ('whether') particle, the presence of a *li* particle in question is mandatory.

Thus, based on previous works, I assumed that: a) long-distance wh-movement in Russian is restricted at the level of the embedded clause by complementizers; b) the extraction of wh-elements from indirect questions with interrogative pronoun is less acceptable than from indirect questions with *li* particle; c) d-linked constituents will be more acceptable than non-d-linked; d) the researchers describing this problem practically do not consider the asymmetry between arguments and adjuncts. Even if the asymmetry between the extraction of wh-phrases from *chto*- and *chtoby*-clauses is observable through introspective studies, then the potential asymmetry of arguments and adjuncts, as well as the combination of these factors, can only be established by formal quantitative methods.

## Experimental study

Each experiment included 'type of extracted elements' as an independent variable (wh-argument and wh-adjunct). Other independent variables included: in **Exp1** type of complementizer ('*chto*' or '*chtoby*'); **Exp2**: type of interrogative structure (indirect question with *li* particle and indirect question with interrogative pronoun) and D-linking (in term of (Pesetsky 1987)); **Exp3**: type of interrogative structure (indirect question with *li* particle, embedded clause with complementizer '*chto*' and embedded clause with complementizer '*kak*'). The stimulus material in each experiment was unified by a number of parameters: extracted wh-words, matrix predicate and length of sentence. Half of wh-arguments were animated (*kogo* 'who').

Participants had to judge the acceptability of the target sentence on a 7-point Likert scale (1 = unacceptable, 7 = acceptable). Every participant in total rated from 16, 32, and 24 target items, and the same quantity of fillers. Fillers divided on grammatical, ungrammatical and marginal ones. All stimuli and fillers were written materials. In case of Exp1, participants should read sentence word by word and then rate it, however, in Exp2 and 3 participants could see the whole sentence (target item or filler).

Exp1 involved 102 respondents, aged 15 to 70. Exp2 involved 136 people, 15 to 72 years old. Finally, Exp3 involved 83 people, from 23 to 72 years old. All participants were native speakers of Russian language, living in Russia. Responses of some participants were deleted due to too short time (approximately 35ms for each item response) or inadequate fillers rating.

Three experiments showed the following results. ANOVA analysis demonstrated that: in Exp1 the only significant factor was 'type of complementizer' ( $p$ -value  $< 2.2e-16$ ).



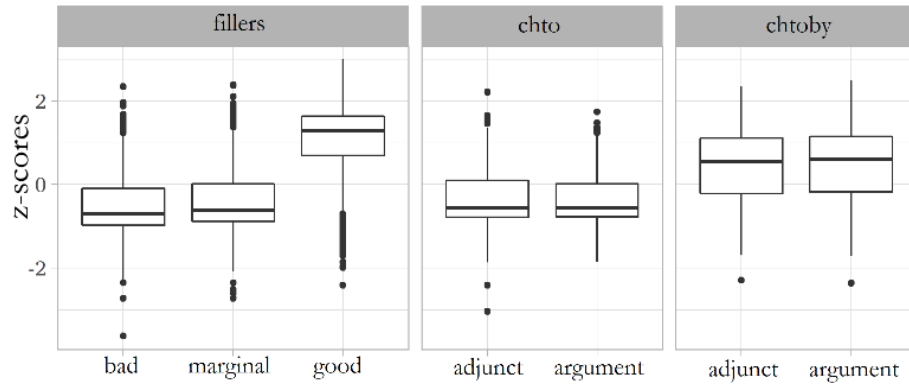


Figure 1. Experiment 1: z-scores of fillers and stimuli.

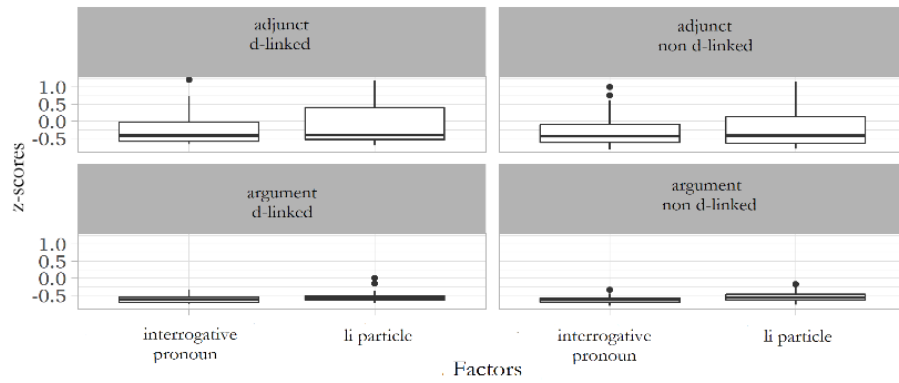


Figure 2. Experiment 2: z-scores of all stimulus.

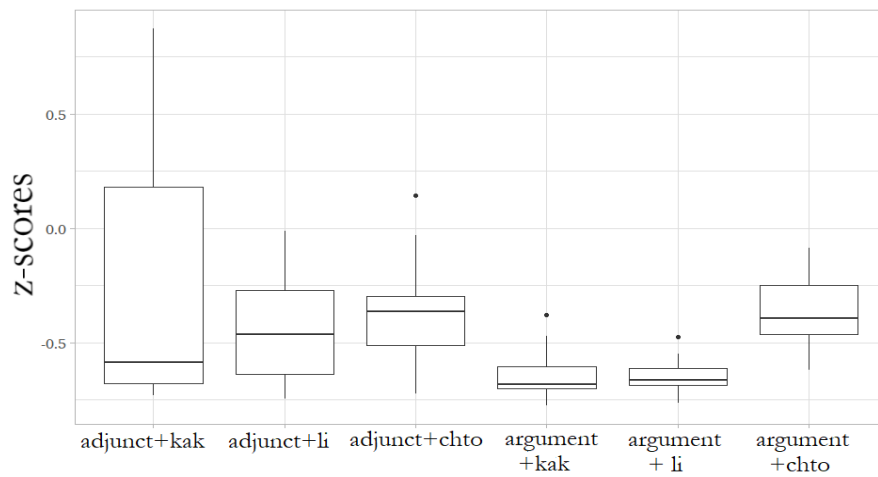


Figure 3. Experiment 3: z-scores of all stimuli.

Opposite, in Exp2 and Exp3 factor ‘type of extracted element’ was significant ( $p$ -value =  $2.04e-14$  in both cases). As for Exp2, stimuli containing d-linked constituents are slightly more acceptable (according to average z-scores and pairwise comparison data) than stimuli containing non-d-linked constituents, however ‘d-linking’ was not significant ( $p$ -value = 0.2158).

Exp1 and Exp3 indicate that the embedded clauses, headed by the complementizer ‘*cto*’, are strong islands, from which it is impossible to extract any elements. According to the results of Exp2, indirect questions in Russian also demonstrate islands constraint. Interrogatives with embedded clauses headed by complementizer ‘*ctoby*’ do not exhibit any island effects and allow extraction of any elements. Next, stimuli with extracted arguments from *ctoby*-clauses got higher scores than stimuli with adjuncts (Fig.1). In Exp2 and Exp3 all stimuli with adjunct got slightly higher scores than stimuli with arguments due to the possibility of late adjunction (Stepanov 2001).

Stimulus sentences with *kak*-clause and adjuncts showed the greatest variability in ratings (Fig.3). Structures with *kak*-clauses do not allow extracting wh-arguments, but extracting adjuncts is acceptable for a certain group of respondents. Probably, we can talk about several groups of respondents who process stimuli with adjuncts in different ways. Hence, among the participants of experiment 3 exist a group of participants, who can associate extracted adjunct with the matrix clause and rate such a stimulus as acceptable. The question of the status *kak*-clause is still open, since there is still not enough data to draw a conclusion about its properties.

## References

- Antonenko, A. 2010. Puzzles of Russian Subjunctives. U. Penn Working Papers in Linguistics, Volume 16.1.
- Bailyn, J. 2018. Scrambling the EGG. Handout 5: The Sensitive Feature People.
- Boeckx, C. 2012. Syntactic islands. – Cambridge University Press.
- Dyakonova, M. 2009. A phase-based approach to Russian free word order. LOT.
- Khomtseyevich, O. 2008. Sintaksicheskoe peredvizhenie iz pridatochnykh izyashitelnykh v russom yazyke. Vestnik Sankt-Peterburgskogo unversiteta. Seria 9. N 3, part 1.
- Liutikova, E. 2009. Otnositelnye predlozheniya s soyuznym slovom kotory: obtshaya charakteristika i svoistva peredvizheniya. Korpusnye issledovaniya po russkoy grammatike. Editors: Kiseleva K., Plungyan V., Rachilina E., Tatevosov S. Moscow: Probel, pp 436-511.
- Pesetsky, D. 1987. Wh-in-situ: Movement and unselective binding // The representation of (in)definiteness. 1987./ E. Reuland, A. Meulen (eds). Cambridge Massachusetts: MIT Press. P. 98–129.
- Stepanov, A. 2001. Late Adjunction and Minimalist Phrase Structure. Syntax, 4: 94-125

# Grammatical-lexical pronoun dissociation in Moroccan Arabic agrammatism

Loubna El Ouardi<sup>1</sup>, Samir Diouny<sup>2</sup>

<sup>1</sup>Applied Language & Culture Studies Lab, Chouaib Doukkali University, Morocco

<sup>2</sup>Clinical Neuroscience & Mental health Lab, Hassan II University, Morocco

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The ProGram theory assumes that pronouns can be classified into lexical and grammatical pronouns. Based on Boye & Harder's (2012) theory of the distinction between lexical and grammatical elements, it is predicted that grammatical pronouns are significantly more impaired than lexical pronouns in agrammatic aphasia. The objective of this study was to investigate whether Moroccan Arabic agrammatic subjects exhibit a dissociated processing ability between grammatical and lexical pronouns. Narrative speech was elicited from 5 agrammatic patients and 5 normal controls using the "Cookie Theft" picture description task from the Boston Diagnostic Aphasia Examination. Findings supported a grammatical-lexical dissociation in Moroccan Arabic agrammatism.

Keywords: Pronouns, agrammatism, Moroccan Arabic, ProGram theory

## Introduction

Agrammatism, often associated with Broca's aphasia, is characterized by omissions and substitutions of function words and grammatical morphemes. Agrammatic subjects are said to have trouble processing pronouns (Miceli & Mazzucchi, 1990). The current state of knowledge suggests that the pronoun deficit in agrammatism is not an across-the-board phenomenon, and that specific pronouns are more vulnerable to impairment than others (Avrutin, 2006).

Boye & Harder (2012) suggest a new conceptualization of the relation between the lexicon and the grammar, by arguing that whereas lexical items can stand alone and can convey the main point of a speech act, grammatical items cannot and are dependent on other items for their interpretation. Boye and Harder's (2012) account argues in disfavour of a general understanding of pronouns as closed-class items. It rather suggests a criterion by means of which a decision could be made as to whether a linguistic unit is grammatical or lexical. This is known as the focus test. The idea is that pronouns (or other structures) that cannot pass the focus test by means of clefting, focus particle or stressing are grammatical, whereas those pronouns that pass the test are classified as lexical. To this end, the present study examines the validity of the usage-based approach to grammatical status by drawing evidence from Moroccan Arabic (MA) agrammatism. Our goal is to use the focus test to first classify pronouns

as lexical or grammatical, and then investigate whether a dissociation exists between the production of the two pronoun categories in agrammatic speech. We predict that (1) patients with agrammatic aphasia (henceforth, PWAA) will produce fewer pronouns overall than non-brain-damaged participants (henceforth, NBDs), and that (2) grammatical pronouns will be more impaired than lexical pronouns in PWAA.

## Methods

### Participants

Data for the study were collected from 5 agrammatic patients (3 females, 2 males, mean age: 48, SD: 11.46; mean years of education: 6.2, SD = 1.9), and 5 non-brain-damaged participants (3 females, 2 males; mean age: 48.4, SD:10.73; mean years of education: 7; SD: 1.58). Months post-onset ranged from 8 to 13. PWAA received a diagnosis of Broca's aphasia with agrammatism based on the results of three measures: (1) the Short Version of the Moroccan Aphasia Test, (2) an action and object naming battery, and (3) a guided conversation from the Moroccan Arabic version of the Montreal-Toulouse Protocol of Aphasia Linguistic Examination (El Alaoui Fares et al., n.d). Patients met the classic criteria of agrammatic speech.

### Material and procedure

Narrative speech samples were collected from participants using the "Cookie Theft" picture from the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983). Speech samples were transcribed and analysed for the number of nouns and pronouns in addition to the percentage of grammatical and lexical pronouns produced.

### Classification of MA pronouns into grammatical or lexical pronouns

We used the focus test (clefting and focus particle) to classify 91 Moroccan Arabic pronouns as either grammatical or lexical. 47 pronouns were classified as grammatical, whereas 44 were classified as lexical.

## Results

The results indicating the performance of PWAA and NBDs are summarized in figures 1 and 2.

Between-group comparisons suggested that PWAA produced significantly fewer pronouns and nouns than NBDs (Mann Whitney U Test:  $Z=-2.611$ ,  $p=.009$  for pronouns;  $Z=-2.627$ ,  $p=.009$  for nouns). Comparisons also indicated that PWAA produced significantly fewer grammatical pronouns than NBDs ( $Z=-2.611$ ,  $p=.009$ ). The difference between the number of lexical pronouns produced by both groups did not reach statistical significance ( $Z=-.107$ ,  $p=.915$ ).

The NBD group descriptively produced more pronouns than nouns (mean pronouns=23.8; mean nouns=19.2), although this difference did not reach statistical significance (Wilcoxon Signed Ranks Test:  $Z=-1.214$ ,  $p=.225$ ). NBDs produced significantly more grammatical than lexical pronouns ( $Z=-2.023$ ,  $p=.043$ ). The PWAA group descriptively produced fewer pronouns than nouns (mean pronouns: 7; mean nouns: 12.4), although this difference did not reach statistical significance ( $Z=-1.461$ ,  $p=.144$ ). PWAA produced descriptively fewer grammatical than lexical pronouns (mean grammatical pronouns: 2.6; mean lexical pronouns: 4.4), although this difference did not reach statistical significance ( $Z=-1.089$ ,  $p=.276$ ).

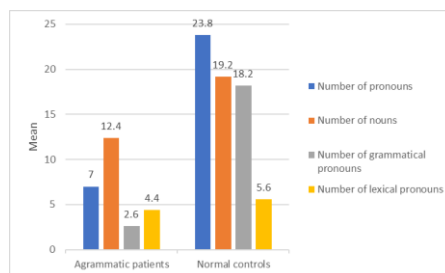


Figure 1. Mean number of nouns, pronouns, grammatical & lexical pronouns produced by agrammatic patients and normal controls in picture description.

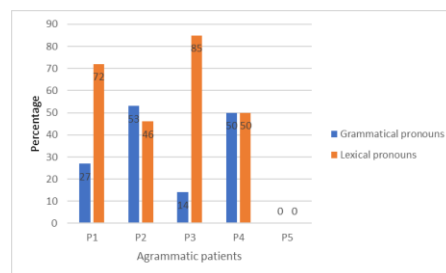


Figure 2. Percentage of grammatical & lexical pronouns produced by individual PWAA.

## Discussion

Our results confirmed our hypotheses. PWAA produced significantly fewer pronouns than NBDs, and produced fewer grammatical than lexical pronouns.

Regarding the grammatical-lexical distinction, the NBD group in this study produced significantly more grammatical than lexical pronouns. This aligns with the findings of Martinez-Ferreiro et al. (2018) who found that NBDs produced more grammatical pronouns than lexical pronouns. Following Martinez-Ferreiro et al. (2018), we advance that this reflects frequency effects optimizing the production of grammatical pronouns in the speech of normal participants. In addition, given the pro-drop nature of MA and the nature of the picture description task (focus on verb elicitation), a significant number of null subject pronouns (grammatical) were produced cliticizing to the main verb. Our results also indicated that PWAA had more difficulty producing grammatical (but not lexical) pronouns when compared to NBDs (Ishkhanyan et al., 2017).

Although the results of this study provided interesting support for the validity of the ProGram theory (Boye & Harder, 2012) in the sense that

grammatical pronouns were more vulnerable than lexical pronouns in the agrammatic group, it is important to note that the dissociation was not consistent across all the patients tested.

The study has both theoretical and clinical implications. A theoretical understanding of pronouns as belonging to the closed-class category of words is challenged on empirical grounds. Clinically, the current state of knowledge points to an additional symptom of agrammatic speech that can be used for linguistic diagnosis: reduction in the number of grammatical pronouns produced.

## References

- Avrutin, S. 2006. Weak syntax. In Y. Grodzinsky & K. Amunts (Eds.), *Broca's region* (pp. 49-62). Oxford University Press.
- Boye, K., Harder, P. 2012. A usage-based theory of grammatical status and grammaticalization. *Language*, 88, 1–44. doi:10.1353/lan.2012.0020
- El Alaoui Faris, M., Ettahiri, L., Fennich, R., Benbelaid, F. (n.d.). Montreal– Toulouse protocol of aphasia linguistic examination: MA version. Rabat: Faculty of Medicine.
- Ishkhanyan, B., Sahraoui, H., Harder, P., Mogensen, J., Boye, K. 2017. Grammatical and lexical pronoun dissociation in French speakers with agrammatic aphasia: A usage-based account and REF-based hypothesis. *Journal of Neurolinguistics*, 44, 1–16. doi:10.1016/j.jneuroling.2017.02.001
- Goodglass, H., Kaplan, E. 1983. *The assessment of aphasia and related disorders*. Philadelphia: Lea and Febiger.
- Martínez-Ferreiro, S., Ishkhanyan, B., Rosell-Clarí, V., Boye, K. 2018. Prepositions and pronouns in connected discourse of individuals with aphasia. *Clinical Linguistics & Phonetics*, 1-21. doi:10.1080/02699206.2018.1551935
- Miceli, G. Mazzucchi, A. 1990. Agrammatism in Italian: Two case studies. In L. Menn, L. Obler (Eds.), *Agrammatic aphasia: A cross-language narrative sourcebook* (pp. 717-816). John Benjamins.

# Constraints of lexical and grammatical aspect on event representations in Mandarin Chinese

Shaohua Fang<sup>1</sup>, Fei Yuan<sup>2</sup>

<sup>1</sup>Department of Linguistics, University of Pittsburgh, USA

<sup>2</sup>Faculty of Asian and Middle Eastern Studies, University of Cambridge, UK

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The current study addresses whether differential facilitation effects (dependent on lexical aspects) from perfectives and imperfectives observed in English and Cantonese would generalize to Mandarin Chinese by conducting a battery of offline and online experiments. The results indicated an across-the-board processing advantage of perfectives. Crucially, we found that the grammatical and lexical aspect interactively constrained the representation and processing of sentence processing in Mandarin. However, a robust facilitation emerged only for the perfective aspect when it co-occurred with accomplishment verbs but not for the imperfective aspect when it co-occurred with activity verbs.

Keywords: constraints, lexical aspect, grammatical aspect, facilitation, Mandarin

## Introduction

In Mandarin Chinese, aspect can be grammatically encoded. For example, the perfective aspect with *-le* and the imperfective aspect with *-zhe*. It can also be lexically encoded. For example, *du* ‘(to read)’ and *gu* ‘(to hire)’ are examples of activity verbs and accomplishment verbs. According to Peck, Lin and Sun (2013), accomplishment verbs can be mono-morphemic and resultative compounds. Verbs of both lexical aspects can be distinguished by the [ $\pm$ scalar] and the [ $\pm$ telic] features: Activity verbs are non-scalar and atelic; Accomplishment verbs are scalar and telic.

Language users use aspectual information to construct event representations during sentence comprehension. In a sentence-picture selection task, Madden and Zwaan (2003) found that the perfective aspect rather than the imperfective aspect facilitated English sentence comprehension. This effect is called *perfective facilitation* in Yap et al. (2009) where it was attested with Cantonese, and such an effect was shown to be modulated by lexical aspect in that perfective facilitation was evident with accomplishment verbs and imperfective facilitation culminated in activity verbs.

The present study adopts a cross-linguistic perspective examining whether the effects of grammatical and lexical aspects on the construction of event representations observed in English and Cantonese would generalize to Mandarin Chinese.

## Experiments

### Materials

A 2\*2 factorial design was adopted for the experiments, yielding a total of four conditions. The first independent variable was grammatical aspect (GA) which varied between perfective aspect (*-le*) and imperfective aspect (*-zhe*), and the second independent variable was lexical aspect (LA) with two levels, activity, accomplishment. Experimental items were sentences containing each of the four conditions in Table 1. Each participant read 36 sentences, consisting of 12 experimental sentences and 24 fillers, of which 10 sentences were catch trials. The experimental sentences were identical across each experiment.

Table 1. Lexical and grammatical aspect in Mandarin.

Grammatical Aspect	Lexical Aspect	
	Accomplishment	Activity
Perfective <i>-le</i>	mai- <i>le</i> (to have bought)	chi- <i>le</i> (to have eaten)
Imperfective <i>-zhe</i>	mai- <i>zhe</i> (buying)	chi- <i>zhe</i> (eating)

### Participants and procedure

Eighty-two native Mandarin speakers participated in a self-paced reading (SPR) task, and an untimed acceptability judgment task (AJT). Another group of 65 participants completed a speeded acceptability judgment task (SAJT). Experiments were distributed via Ixweb Farm (Drummond, 2013). In the SPR, each sentence was presented word-by-word in a non-cumulative fashion, followed by a comprehension question. In the AJT, participants rated each sentence using a 7-point Likert Scale. In the SAJT, sentences were presented one word at a time at a rate of 300 ms for participants to respond with yes/no within 3s.

### Data treatment and analysis

For AJT, no participant was rejected for statistical analysis since each of them consistently rated the unambiguously acceptable catch trials above the unacceptable ones. Raw ratings were then converted into z-scores. For SPR, before statistical modelling, data were cleaned and trimmed such that two participants who scored below 80% on the comprehension questions were removed, and the RTs above 2000 ms were further discarded. RTs were subsequently log transformed. Linear mixed-effects models and logistic mixed-effects models were used for the SPR/AJT data and the SAJT data respectively. Independent variables were sum coded; participant and experimental items were treated as the random effects.



## Results

Figure 1 visualizes the descriptive results for AJT. There was a main effect of grammatical aspect, indicating that perfective sentences were rated reliably higher than imperfective sentences ( $b=-0.325$ ,  $p<.0001$ ). GA significantly interacted with LA. Noticeably, sentences containing accomplishment verbs were more acceptable with perfective aspect than with imperfective aspect ( $b=-0.921$ ,  $p<.0001$ ). Imperfective sentences were less acceptable when they contained accomplishment verbs than when they contained activity verbs ( $b=-0.724$ ,  $p<.0001$ ). Figure 2 visualizes the descriptive results for SAJT. A main effect of grammatical aspect revealed that perfective sentences were more likely to be accepted than imperfective sentences ( $b=-1.170$ ,  $p=.0069$ ). GA and LA significantly interacted such that imperfective sentences with activity verbs ( $b=-3.380$ ,  $p<.0001$ ) were more likely to be accepted than those with accomplishment verbs. Moreover, perfective sentences were more likely to be accepted than imperfective sentences when they contained accomplishment verbs ( $b=-3.407$ ,  $p<.0001$ ). For the SPR, RTs of two regions were analysed – the critical region where GA locates, and the spillover region where numeral-classifier locates. A reliable interaction between both aspects was observed at the critical region, indicating that this region was read faster for perfective sentences with accomplishment verbs than for sentences with activity verbs ( $b=-0.064$ ,  $p=.0392$ ). A main effect of GA was obtained at the spillover region, reflecting an overall slowdown for the imperfective sentences relative to the perfective sentences ( $b=0.030$ ,  $p=.0916$ ). Moreover, we obtained a reliable interaction between both aspects at the spillover region, which was driven by the reduced RTs on the perfective sentences relative to the imperfective sentences when they contained accomplishment verbs ( $b=0.074$ ,  $p=.0174$ ).

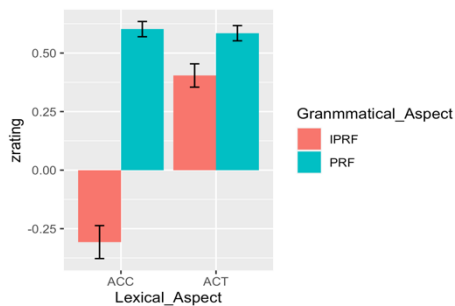


Figure 1. Mean z-ratings.

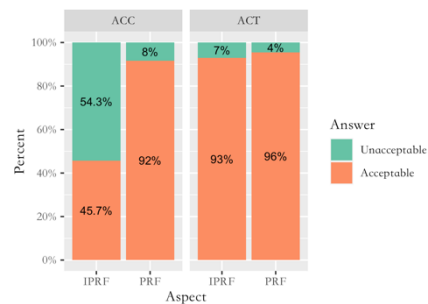


Figure 2. Speeded acceptability judgments.

## Discussion and conclusion

The results show an advantage of perfective aspect over imperfective aspect across each measure, suggesting that the perfective aspect relative to imperfective aspect could facilitate processing to a larger extent in Mandarin. This could be interpreted by the fact that the imperfective aspect with a focus on the internal structure of an event usually yields a longer retention rate relative to the perfective aspect, thus leading to a slower cognitive processing. (Madden & Zwaan, 2003; Yap et al., 2009).

Crucially, it was clear from the results that GA and LA interacted to facilitate sentence processing across the board. However, different from Yap et al. (2009), we only observed the perfective facilitation with accomplishment verbs, but not the imperfective facilitation with activity verbs in Mandarin. Specifically, in the AJT, the fact that sentences with Accomplishments + *le* were rated higher than those with Accomplishments + *zhe* demonstrated a facilitation of Accomplishments + *le*. In the SAJT, the facilitation of Accomplishments + *le* was also robust, as reflected in a higher likelihood of accepting sentences with Accomplishments + *le* relative to those with Accomplishment + *zhe*. Similar facilitative patterns emerged during moment-by-moment processing in that the spillover region was processed faster in the condition of Accomplishments + *le* relative to Accomplishments + *zhe*.

We argued that the interaction between LA and GA was driven by semantic compatibility: Accomplishment verbs being telic are semantically compatible with the perfective aspect (*-le*) which usually denotes a completed event, whereas Activity verbs being atelic are semantically compatible with the imperfective aspect (*-zhe*) which usually denotes an ongoing event.

## Acknowledgments

We would like to thank Dr. Foong Ha Yap, Dr. Carol J. Madden, and Dr. Alan Juffs for their helpful comments and discussion regarding this project.

## References

- Drummond, A. 2013. Ibx farm. Online server: <http://spellout.net/ibexfarm>, 831.
- Madden, C. J., Zwaan, R. A. 2003. How does verb aspect constrain event representations? *Memory & cognition*, 31(5), 663-672.
- Peck, J., Lin, J., Sun, C. 2013. Aspectual classification of Mandarin Chinese verbs: A perspective of scale structure. *Language and Linguistics*, 14(4), 663.
- Yap, F. H., Chu, P. C. K., Yiu, E. S. M., Wong, S. F., Kwan, S. W. M., Matthews, S., ... Shirai, Y. 2009. Aspectual asymmetries in the mental representation of events: Role of lexical and grammatical aspect. *Memory & cognition*, 37(5), 587-595.
- Zacks, J. M., Tversky, B. 2001. Event structure in perception and conception. *Psychological bulletin*, 127(1), 3.

# Uncovering variation in classifier assignment in Oceanic

Michael Franjeh<sup>1</sup>, Greville G. Corbett<sup>1</sup>, Alexandra Grandison<sup>2</sup>

<sup>1</sup>Surrey Morphology Group, University of Surrey, UK

<sup>2</sup>School of Psychology, University of Surrey, UK

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

We discuss the results of a video vignettes experiment that uncovers the variation of noun-classifier assignment in the possessive classifier system of six Oceanic languages. The results show that languages vary in their noun-classifier assignment, with some languages displaying relatively fixed assignment, similar to a grammatical gender system.

Keywords: classifiers, gender, grammaticalisation, Oceanic, video vignettes

## Motivation

In typical Oceanic possessive classifier systems, a noun can occur with different classifiers, depending on how the possessed item is used by the possessor (Lichtenberk, 1983). For example, *oei* ‘water’ in Vatlongos (Vanuatu) occurs with either the DRINK (1a) or the GENERAL classifier (1b). The ability of a noun to occur with different classifiers is termed overlap.

1a. *mu*            *oei*            **man**  
3sg.drink    water            drink.cl.3sg  
‘he drank his water’

b. *mikas*        *vatin*        *ni*        *oei*            **nan**  
3SG.wash    head.3SG    PREP    water        GENERAL.CL.3SG  
‘he washed his head with his water’

In marked contrast, North Ambrym’s (Vanuatu) cognate for water – *we* – occurs only with the DRINK classifier, not the GENERAL classifier, whether drinking is involved (2a), or not (2b):

2a. *mamnu*        **man**            *we*  
3SG.drink    DRINK.CL.3SG    water  
‘he drank his water’

b. *mwebagwo*        *boton*        *ne*        **man**            *we*  
3SG.wash        head.3SG    PREP    DRINK.CL.3SG    water  
‘he washed his head with his water’

We argue that North Ambrym’s innovative system shows some properties of a gender system: a noun occurs with a particular classifier regardless of contextual interactions. What is more, example 1b from Vatlongos shows a match between the general semantics of the verb ‘wash’ and the GENERAL classifier, whereas in North Ambrym there is a mismatch between the general semantics

of the verb ‘wash’ and the semantics of the DRINK classifier (2b). We expect more gender-like systems to display more mismatches between the verb and the classifier. We ask whether gender systems can indeed emerge from classifiers in this way.

## Methodology

We designed a suite of experiments to compare possessive classifier systems in six representative Oceanic languages: Merei, Lewo, Vatlongos, North Ambrym (Vanuatu), Nêlêmwa and Iaii (New Caledonia). Each has a different inventory of classifiers, from a simple two-way distinction (Merei) up to a more complex inventory of 23 (Iaii). Table 1 shows the languages, classifiers and participants.

Table 1. The sample languages, number of classifiers and number of participants.

Language	Classifiers	Participants
Merei	2	21
Lewo	3	23
Vatlongos	4	23
North Ambrym	5	23
Nêlêmwa	19	11
Iaii	23	16

Table 2. The nouns tested, typicality of interaction, context and classifier.

Noun	Typical interaction	Atypical interaction
green coconut	drink (DRINK) / eat (FOOD)	sit (general)
dry coconut	eat (FOOD) / drink (DRINK)	sit (general)
paper	draw on (GENERAL)	eat (FOOD)
eggs	eat (FOOD) / sell (GENERAL)	drink (DRINK)
mango	eat (FOOD) / drink (DRINK)	
coffee	drink (DRINK)	eat (FOOD)
milo	drink (DRINK)	eat (FOOD)
coconut oil	rub on skin (GENERAL)	eat (FOOD) / drink (DRINK)
washing liquid	wash with (GENERAL)	drink (DRINK)

We focus on the methodology and results from a video vignette experiment. Participants of the different languages watched the same 24 video vignettes; these were designed to investigate the use of classifiers when the items depicted in the vignettes are used in different interactional contexts. Participants were asked to describe in one simple sentence what the actor was doing with their possessions, thus evoking a possessive classifier. Since the classifier inventories vary across the sample languages, we tested their three main classifier categories – GENERAL, DRINK and FOOD. In Table 2 we give the

noun for the item depicted, the typical and atypical interactions along with their expected classifiers.

This method allowed us to investigate three key criteria (i) the amount of overlap between noun and classifier, (ii) the semantic (mis)match between verb and classifier, and (iii) the typicality of contextual interaction.

## Results

ANOVAs were calculated using the *rstatix* package (Kassambara, 2020) in R (R Core Team, 2020). There was a significant main effect between languages for the overlap scores,  $F(5, 111) = 7.55$ ,  $p < .001$ ,  $\eta_p^2 = .25$ . Pairwise comparisons comparing individual languages indicated that North Ambrym and Merei both had significantly lower overlap scores than Lewo, Vatlongos, Nêlêmwa and Iai (largest  $p < .02$ ).

There was a significant main effect between languages for the semantic mismatch scores between verb and classifier,  $F(5, 111) = 9.301$ ,  $p < .001$ ,  $\eta_p^2 = .3$ . Pairwise comparisons comparing individual languages found that North Ambrym had significantly more semantic mismatches between the verb and classifier than all other languages (largest  $p = .005$ ). Furthermore, Iai had significantly fewer semantic mismatches than all other languages bar Merei (largest  $p = .018$ ). Additionally, Vatlongos had significantly more semantic mismatches than Merei ( $p = .043$ ).

When accounting for the difference between typicality of contextual interaction, there was a main effect between languages for semantic mismatches between verb and classifier for typical interactions,  $F(5, 111) = 5.863$ ,  $p < .001$ ,  $\eta_p^2 = .209$ . Pairwise comparisons revealed that North Ambrym had significantly more semantic mismatches than all other languages (largest  $p = .02$ ). Furthermore, Lewo had significantly more mismatches than Vatlongos ( $p = .031$ ) and Iai ( $p = .025$ ). There was also a main effect between languages for semantic mismatches between verb and classifier for atypical interactions,  $F(5, 111) = 6.213$ ,  $p < .001$ ,  $\eta_p^2 = .219$ . Pairwise comparisons revealed that North Ambrym had significantly more semantic mismatches than all other languages bar Nêlêmwa and Vatlongos (largest  $p = .001$ ). Furthermore, Iai made significantly fewer mismatches than Vatlongos and Nêlêmwa (largest  $p = .018$ ). Additionally, both Lewo and Merei made significantly fewer mismatches than Vatlongos (largest  $p = .022$ ).

A significant negative correlation was found between the overlap score and semantic mismatch score ( $r = -.468$ ,  $p < .001$ ) showing that participants who used more overlapping constructions between a noun and a classifier displayed less semantic mismatches between verb and classifier.

## Discussion

The video vignette experiment has demonstrated different ways to compare classifier systems, by comparing systems based on the degree of noun-classifier overlap, typicality of interaction and semantic (mis)match between verb and classifier. The results reveal that some innovative systems do indeed function more like a gender system, in having a more fixed assignment system.

In terms of noun-classifier assignment, North Ambrym shows the most fixed assignment, and so its system is the one which comes closest to a gender system; at the other extreme, Vatlongos displays the most overlap, and has the most classifier-like assignment system. Furthermore, North Ambrym also has the largest amount of semantic mismatch between verb and classifier, in both typical and atypical interactions.

All six languages display higher amounts of semantic mismatch between verb and classifier for atypical interactions, revealing that the periphery of the classifier systems function similarly. Speakers use a more gender-like fixed assignment system with atypical interactions than with typical interactions. The core typical interactions display lower amounts of mismatch, showing that languages display more overlap between noun and classifier in these contexts.

The significant negative correlation between overlap score and semantic mismatch reveals that when speakers use a system that is more classifier-like and displays more overlap between noun and classifier, they are more likely to have fewer semantic mismatches between the verb and classifier.

Finally, the results for Merei display a more gender-like system. However, this is partly due to Merei only having two classifiers – a GENERAL and a CONSUMABLE classifier. The CONSUMABLE classifier includes nouns that refer to both drink and food entities. As the vignettes were designed to investigate three main classifiers – FOOD, DRINK and GENERAL, there is more likely to be less overlap in Merei, due to the combination of two semantic categories.

## Acknowledgements

We gratefully acknowledge the ESRC (grant: ES/R00837X/1: Optimal categorisation: the origin and nature of gender from a psycholinguistic perspective). We thank our 117 participants in Vanuatu and New Caledonia, in particular, Willie Salong (North Ambrym); Elder Simeon Ben (Vatlongos); Korau Melio (Lewo); Melkio Wulmele, Adam and Hester Pike (Merei); Willion Phadom (Nêlêmwa); Wejê Bae (Iaai); and Anne-Laure Dotte (Nêlêmwa and Iaai), for their help in conducting the experiments.

## References

- Kassambara, A. 2020. Rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.6.0.
- Lichtenberk, F. 1983. Relational Classifiers. *Lingua* 60(2-3):147–176.
- R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

# Professions and gender agreement in Russian

Marina Frolova<sup>1</sup>, Natalia Slioussar<sup>1,2</sup>

<sup>1</sup>Department of Liberal Arts and Sciences, St Petersburg University, Russia

<sup>2</sup>School of Linguistics, HSE University, Russia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

In Russian, most nouns denoting professions are historically masculine, but can now be used as common gender: with both masculine and feminine agreement. At the same time, some of these nouns have paired feminine nouns (e.g. *žhurnal'ist* 'journalist<sub>M/(F)</sub>' – *žhurnal'istka* 'journalist<sub>F</sub>'). We investigated for the first time how the availability of such pairs influences the processing of common gender nouns with feminine agreement. We found that online (in a self-paced reading task), this factor does not play a significant role, while stereotypicality (which professions are perceived as stereotypically male or female) does. Offline (in an acceptability judgment task), the situation is the opposite.

Keywords: grammatical gender, profession, stereotype, processing, Russian.

## Introduction

Russian language has three grammatical genders: masculine (M), feminine (F) and neuter (N). The gender of the noun cannot be unambiguously determined from its inflectional affixes (although there are some clear tendencies), but becomes evident from agreeing adjectives, participles and verbs. Like in many other languages, most nouns denoting professions are masculine in Russian.

How to call a female director or a female author in Russian? Two routes are available. Firstly, a corresponding grammatically feminine noun can be formed (e.g. *žhurnal'ist* 'journalist<sub>M</sub>' – *žhurnal'istka* 'journalist<sub>F</sub>'). Russian uses a variety of suffixes to form such nouns, and many masculine nouns denoting professions do not have an established feminine counterpart at all. Secondly, Russian has so-called common gender, and an originally masculine noun can be used with masculine and feminine agreement (e.g. *nash* / *nasba vrach* 'our<sub>M</sub> / our<sub>F</sub> doctor'). This route is available for any masculine noun denoting profession, even when a paired feminine noun exists.

Many studies focus on processing of gender agreement with nouns denoting professions in different languages (e.g. Carreiras et al., 1996; Gygax et al., 2008), including Russian (Garnham, Yakovlev, 2015; Magomedova, Slioussar, 2021; Slioussar, Generalova, 2018). But it was never investigated how the availability of paired feminine nouns influences the processing of common gender (historically masculine) nouns with feminine agreement. This question is interesting both for gender linguistics and for theoretical morphology. We address it in the present study in two experiments focusing on online processing and offline judgment.

## Self-paced reading experiment

### Method

63 native speakers of Russian participated in the study. We constructed 24 stimulus sentences like (1). The subject was always a common gender noun. Half of the nouns had a paired feminine noun, the other half did not (we avoided nouns with colloquial pairs). In both groups, half of the nouns denoted professions that Russian speakers perceive as stereotypically male, while the other half denoted stereotypically female professions (based on the study by Garnham and Yakovlev (2015)). The sentences appeared in two conditions, with masculine and feminine agreement, as (1) shows, and were divided into two experimental lists together with 50 grammatical fillers.

- (1) V reklame kosmetolog blestjashche prezentoval/prezentovala novyj krem.  
in advertisement cosmetologist brilliantly presented<sub>M/F</sub> new cream  
'In an advertisement, the cosmetologist brilliantly presented a new cream?'

The word-by-word self-paced reading methodology was used. To ensure that participants are reading properly, one third of the sentences was followed by questions with a choice of two answers. No participants were excluded based on low accuracy, and below, only RTs are discussed.

### Results and discussion

Average RTs per region in different conditions are shown on Figures 1 and 2 (*p* stands for nouns with feminine pairs, *un* for unpaired nouns, *f/m* for feminine or masculine agreement on the predicate).

Mixed-effects regressions with random intercepts and slopes by participant and by item were used for the statistical analysis. The fixed factors were pairedness, stereotypicality and verb gender. For the first two factors, we were interested not in their main effects (sentences about different professions contained different lexical items), but in their interactions with the verb gender factor. We analysed data in different regions and found significant results only in the regions 5 and 6.

In region 5, the verb gender, stereotypicality and their interaction were significant ( $\beta = 51.19$ ,  $SE = 13.63$ ,  $t = 3.76$ ,  $p < 0.01$ ;  $\beta = 59.02$ ,  $SE = 13.59$ ,  $t = 4.34$ ,  $p < 0.01$ ;  $\beta = -45.65$ ,  $SE = 19.58$ ,  $t = -2.33$ ,  $p = 0.03$ ). In region 6, the verb gender and its interaction with stereotypicality were significant ( $\beta = 49.01$ ,  $SE = 14.45$ ,  $t = 3.39$ ,  $p < 0.01$ ;  $\beta = -44.19$ ,  $SE = 21.17$ ,  $t = -2.09$ ,  $p = 0.05$ ). In other words, sentences with feminine agreement are processed more slowly than the ones with masculine agreement (i.e. it is still a less expected option), but this delay is less pronounced with stereotypically feminine professions. We can conclude that in online processing, the existence of a paired feminine noun does not play a significant role, while stereotypicality does.



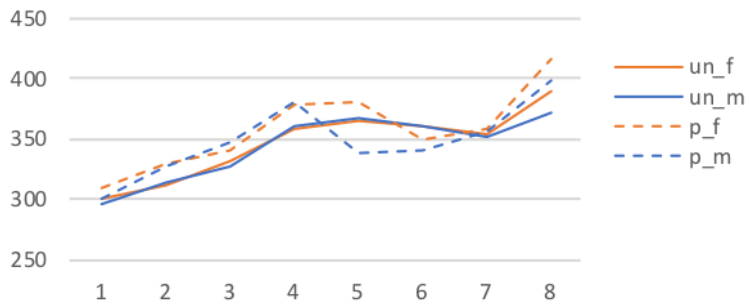


Figure 1. Average word-by-word RTs (in ms) with nouns denoting stereotypically female professions.

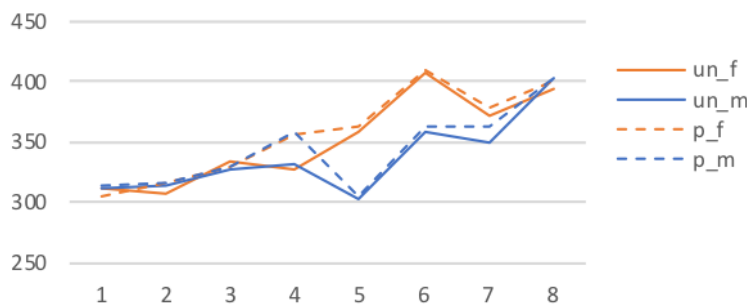


Figure 1. Average word-by-word RTs (in ms) with nouns denoting stereotypically male professions.

## Acceptability judgment experiment

### Method

40 native speakers of Russian participated in the study. We constructed 24 stimulus sentences like (2). Like in the first experiment, the subject was always a common gender noun, and we had an equal number of paired and unpaired nouns denoting stereotypically male and female professions. The proper names made it clear that target sentences described women. We also had 24 analogous filler sentences with male proper names. Participants were asked to judge the acceptability of the sentences on a 1 to 5 scale.

- (2) Psikholog Lidia Mikhajlovna chutkaja.  
 psychologist L.M. sensitive<sub>F</sub>  
 ‘Psychologist Lidia Mikhajlovna is sensitive.’

## Results and discussion

The average ratings in the paired conditions were 3.8 for stereotypically female professions and 3.7 for stereotypically male ones, while in the unpaired conditions, they were 2.8 and 3.2, respectively. We used mixed-effects ordinal regressions with random intercepts and slopes by participant and by item. The fixed factors were pairedness and stereotypicality. The results were the opposite to the online experiment: pairedness was significant ( $\beta = 2.73$ ,  $SE = 0.42$ ,  $t = -7.54$ ;  $p < 0.01$ ), while stereotypicality was not, although numerically, there is a tendency for stereotypically female professions to receive higher ratings.

We assume that the pairedness factor reaches significance only offline because it requires a complex metalinguistic judgment (figuring out that there is another, better way to express the same meaning). We also hypothesize that stereotypicality is significant only online because it influences predictability: feminine agreement is used more often with stereotypically female professions and is more readily expected after nouns denoting such professions. Judging how natural a sentence sounds, readers do not take stereotypes into account. This is a non-trivial result because stereotypes are very persistent and, if we go beyond its strictly scientific implications, a very encouraging finding.

## Acknowledgements

The study was supported by the grant 75288744 from St Petersburg University.

## References

- Carreiras M. et al. 1996. The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology A*, 49, 639-663.
- Garnham A., Yakovlev Y. 2015. The interaction of morphological and stereotypical gender information in Russian. *Frontiers in Psychology*, 6, art. 1720, 1-12.
- Gygax P. et al. 2008. Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, 23, 464-485.
- Magomedova, V., Slioussar, N. 2021. Gender and case in Russian nouns denoting professions and social roles. *Computational Linguistics and Intellectual Technologies*, 20, 483-491.
- Slioussar N., Generalova A. 2018. Grammaticheskie xarakteristiki i gendernye stereotipy pri obrabotke soglasovanija po rodu v russkom jazyke (in Russian, Grammatical characteristics and gender stereotypes in the processing of gender agreement in Russian). *The 8th International Conference on Cognitive Science: Abstracts*. Moscow: Institute of psychology RAS. Pp. 933–935.

# Applied Cognitive Linguistics and design of L2 figurative language material

Ioannis Galantomos

Department of Language and Intercultural Studies, University of Thessaly, Greece

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The goal of this paper is to put down the practical issues related to the design of a cognitive linguistic-based material for teaching L2 figurative language. Figurative language, such as metaphors, idioms and metonymies, occurs effortlessly in various modes of speech and it is closely related to L2 communicative competence. Within Applied Cognitive Linguistics, figurative language is a complex phenomenon and has given rise to notions, such as motivation, mental organization and mental imagery. These notions have been proven to foster L2 figurative language long-term retention and lexical precision. Given that Cognitive Linguistics is a cognitively demanding approach, certain actions should be undertaken in order to design learning material and make L2 figurative language instruction feasible.

Keywords: L2, figurative language, Applied Cognitive Linguistics, teaching material, CEFR

## Introduction

Lakoff and Johnson's seminal *Metaphors we live by* (1980/2003) stood as the basis for the emergence of Cognitive Linguistics (/Semantics) and made explicit that figurative language is the norm in everyday language practice. Figurative language serves key functions, such as description, explanation, clarification, summation, agenda management, humour and evaluation (Semino, 2008).

From the cognitive linguistic perspective, figurative language is a multifaceted phenomenon; it involves language, the human conceptual system, socio-cultural features and neural and bodily activity (Kövesces, 2005). Applying this radical view of figurative language to second language (henceforth L2) instruction, has been proven to be beneficial for L2 learners in terms of long-term retention and lexical precision (e.g. Hoang & Boers, 2018). These promising results are based on the notions of *motivation*, *mental organization* and *mental imagery* (Boers, 2011).

In particular, motivation is central to human cognition and explains how and why a particular meaning of a figurative expression has arisen (Lakoff, 1987). From this perspective, motivation makes figurative language more memorable (Boers, 2018). Categorization refers to the inevitable and unconscious ability of forming categories based on perceived similarity (Taylor, 2003). Maldonado (2008) claims that categories are very useful for the peripheral rules of the target language that often pose more difficulties to L2 learners. Finally, it is suggested

that learners tend to form strong conventional images in order to describe certain figurative language units, such as idioms (Gibbs, 1994).

## **Applications**

### **Theoretical considerations**

Cognitive Linguistics is a cognitively demanding approach (Gutiérrez Pérez, 2017). For this approach to be beneficial, the advantages should become straightforward to L2 learners from the very beginning (Boers & Lindstromberg, 2006).

Given that not all concepts are figurative (Danesi, 2008), a cognitive linguistic-driven instruction should be seen as a supplementary technique and not as the sole means for L2 vocabulary instruction (Boers, 1999).

Lastly, L2 learners should be exposed to figurative language from early on, that is from CEFR (=Common European Framework of Reference for language, Council of Europe, 2001)-based A2 proficiency level (Littlemore, Krennmayr, Turner, & Turner, 2014).

### **Practical issues**

With reference to more practical issues that are expected to rise when designing a cognitive linguistic-driven teaching material, it is suggested that figurative language appears in context (Peleg, Giora, & Fein, 2004).

Second, the selected texts should be authentic in order to be interesting and increase learners' motivation and degree of engagement (Peacock, 1997).

Third, learners should be told that figurative language is ubiquitous in ordinary discourse (Lakoff & Johnson, 1980/2003). MacLennan (1994) claims that L2 figurative instruction will be fruitful only if learners are explicitly told that figurative language is an integral aspect of everyday communication and it cannot be ignored.

Fourth, L2 learners tend to connect images to figurative language (Gibbs, 1994). Thus, visual and graphic tools will result in organization and clarity, trigger prior knowledge and provide opportunities for interaction with key content at a more complex level (Malette, 2020).

Fifth, Piquer-Píriz (2011) argues that L2 learners need to be familiarized with the core senses of polysemous words which are present and of everyday use in any classroom. If a learner knows the basic/core meaning of an L2 word and is familiar with strategies, such as metaphor and metonymy, then s/he will be able to understand, use, and produce the semantic extensions (/figurative meanings) of these words.

Sixth, Sökmen (1997) points out that an organized vocabulary is better learnt than random lists. Boers (2000) has shown that the lexical organization of figurative language under metaphoric themes (=conceptual metaphors) raised learners' metaphor awareness and facilitated retrieval and long-term retention.

Seventh, figurative language organization under conceptual metaphors entails teaching L2 figurative language based on the scheme “A is B” (Danesi, 1995).

Eighth, Danesi & Grieve (2010) argue that familiarization with figurative language can be developed through a *conceptual syllabus*. In a conceptual syllabus, units will be organized around conceptual domains, such as *love, time, weather* and *ideas* along with grammatical and communicative information regarding their functions and frequency in ordinary language use. Alternatively, units can be planned around salient or less highly productive concepts of the target language (Danesi & Grieve, 2010).

Lastly, research has shown that raising L2 learners’ awareness of the origin of figurative language can contribute to their long-term retention and eventually to better vocabulary acquisition (Boers, 2001; Boers, Eyckmans, & Stengers, 2007). Hence, it will be beneficial for L2 learners to become familiar with patterns of cross-linguistic and cross-cultural differences and similarities (Boers & Demecheleer, 2001).

## References

- Boers, F. 1999. Learning Vocabulary Through Metaphoric Awareness. *Études et Travaux* 3, 53-65.
- Boers, F. 2000. Metaphor awareness and vocabulary retention. *Applied Linguistics* 21(4), 35-43
- Boers, F. 2001. Remembering figurative idioms by hypothesising about their origin. *Prospect* 16, 35-43.
- Boers, F. 2011. Cognitive Semantic ways of teaching figurative phrases. *Review of Cognitive Linguistics* 9(1), 227-261.
- Boers, F. 2018. Exploring Linguistic Motivation. In Lontas, J.I. (ed.) 2018, *The TESOL Encyclopedia of English Language Teaching*, vol. 5, pp. 1-6. Hoboken, NJ, Wiley. DOI: 10.1002/9781118784235.eelt0940.
- Boers, F., Demecheleer, M. 2001. Measuring the impact of cross-cultural differences on learners' comprehension of imageable idioms. *ELT Journal* 55(3), 255-262.
- Boers, F., Lindstromberg, S. 2006. Cognitive linguistic applications in second or foreign language instruction: rationale, proposals, and evaluation. In Kristiansen, G., Achard, M., Dirven, R., Ibáñez, Fr. (eds.) 2006, *Cognitive Linguistics: Current Applications and Future Perspectives*, pp. 305-355. Berlin & New York, Mouton de Gruyter.
- Boers, F., Eyckmans, J., & Stengers, H. 2007. Presenting figurative idioms with a touch of etymology: more than mere mnemonics? *Language Teaching Research* 11(1), 43-62.
- Council of Europe 2001. *Common European Framework of Reference for Languages: Learning, teaching assessment*. Cambridge, CUP.
- Danesi, M. 1995. Learning and teaching languages: the role of “conceptual fluency”. *International Journal of Applied Linguistics* 5(1), 3-20.
- Danesi, M. 2008. Conceptual errors in second-language learning. In De Knop, S., De Rycker, T. (eds.) 2008, *Cognitive Approaches to Pedagogical Grammar*, pp. 231-256. Berlin & New York, Mouton de Gruyter.
- Danesi, M., Grieve, K. 2010. Conceptual Error Theory and the Teaching of Italian. *ITALICA* 87(1), 1-20.

- Gibbs, R.W. Jr. 1994. *The Poetics of Mind. Figurative Thought, Language, and Understanding*. Cambridge, CUP.
- Gutiérrez Pérez, R.G. 2017. Teaching Conceptual Metaphors to EFL Learners in the European Space of Higher Education. *European Journal of Applied Linguistics* 5(1), 87-114.
- Hoang, H., Boers, F. 2018. Gauging the association of EFL learners' writing proficiency and their use of metaphorical language. *System* 74, 1-8.
- Kövecses, Z. 2005. *Metaphor in Culture. Universality and Variation*. Cambridge, CUP.
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago, The University of Chicago Press.
- Lakoff, G., Johnson, M. 1980/2003. *Metaphors We Live By*. Chicago & London, The University of Chicago Press.
- Littlemore, J., Krennmayr, T., Turner, J., Turner, S. 2014. An Investigation into Metaphor Use at Different Levels of Second Language Writing. *Applied Linguistics* 35, 117-144.
- MacLennan, C. 1994. Metaphors and prototypes in the learning and teaching of grammar and vocabulary. *IRAL* 32(1), 97-110.
- Maldonado, R. 2008. Spanish middle syntax: A usage-based proposal for grammar teaching. In De Knop, S., De Rycker, T. (eds.) 2008, *Cognitive Approaches to Pedagogical Grammar*, pp. 155-196. Berlin & New York, Mouton de Gruyter.
- Malette, B. 2020. Concept Circles: A Graphic Organizer that Fosters Deeper Understanding of Critical Vocabulary and Concepts. In Liontas, J. (ed.) 2020, *TESOL Encyclopedia of English Language Teaching*, pp. 1-9. Hoboken, NJ, Wiley. DOI: 10.1002/9781118784235.eelt0996.
- Peacock, M. 1997. The effect of authentic materials on the motivation of EFL learners. *ELT Journal* 51(2), 144-156.
- Peleg, O., Giora, R., & Fein, O. 2004. Contextual Strength: the Whens and Hows of Context Effects. In Noveck, I.A., Sperber, D. (eds.) 2004, *Experimental Pragmatics*, pp. 172-186. Basingstoke, UK, Palgrave Macmillan.
- Piquer Píriz, A.M. 2011. Motivated word meanings and vocabulary learning. The polysemy of hand in the English for Young Learners classroom. *Metaphor and the Social World* 1(2), 154-173.
- Semino, E. 2008. *Metaphor in Discourse*. Cambridge, CUP.
- Sökmen, A.J. 1997. Current trends in teaching second language vocabulary. In Schmitt, N., McCarthy, M. (eds.) 1997, *Vocabulary: Description, Acquisition and Pedagogy*, pp. 237-257. Cambridge, CUP.
- Taylor, J.R. 2003. *Linguistic Categorization*. Oxford, OUP.

# A multidimensional approach in teaching L2

Matteo Greco

Scuola Universitaria Superiore IUSS di Pavia, Italy

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This paper aims to discuss some good techniques gaining better results in second language acquisition during L2 classes. The core idea is that the “whole” person is involved in the learning activity and, therefore, the teaching experience has to deal with a multidimensional approach: it should adopt creativity tasks, such as painting and drawing, as well as practical ones, such as cooking and reciting; it should adopt unmoving activity, such as singing, as well as moving activity, such as jumping and dancing; etc. Each technique will be discussed on a theoretical-driven background.

Keywords: L2 teaching; language acquisition; Practical-Guide

## A Theoretical Framework

Teachers are more and more demanding of successful techniques in order to improve their ability and efficiency. It is well known from Lenneberg's (1967) pioneering studies on the Critical Period Hypothesis, that L2 learners need different input and approaches depending on their age. For example, it has been argued that learners from 0 to 6 years old are able to learn every natural language without any effort and just from direct contact; learners from 6 years old to the end of puberty gradually lose their natural ability and, therefore, they increasingly need to be overtly educated. Learners after puberty completely lose the ability to learn a language from direct contact alone, requiring deliberate attention and extensive comprehensible input in the target language. Although the Critical Period Hypothesis has been challenged (for a critical discussion, see Muñoz, Singleton, 2011), there is an international consensus on the fact that there is a continuous linear decline in the capacity to learn a second language: from the completely naturalistic way of newborns to the completely structured way of adults. Clearly, teachers who want to perform courses in the best way should adopt different techniques depending on which phase the students belong to. For example, while the correct acquisition of the phonological components of the L2 is natural during the first phase (Tahta, Wood, Lowenthal, 1981), in the last one it is extremely difficult, contrasting with the easier acquisition of the other grammatical components, such as syntactic, semantic and pragmatic ones (Isik, 2000).

Teaching second languages requires many tools and abilities (Greco, 2021) and very often it is not easy to make decisions on how to arrange L2 courses. Theoretical literature is rich in suggestions, but it often lacks practical indications (Kanno and Stuart, 2011). On the other hand, practical books

present many activities without a theoretical explanation for them. This study wants to melt together these points of view, by discussing some practical indications, which are based on a theoretical scientific background, and it will focus on the youngest learners, i.e. those from 0 to 6 years old.

### **Activities and strategies in L2 Teaching**

One of the first helping tools in language acquisition projects is the evaluation of the similarities between the students' native language with the L2 ("cross-linguistic influence in second language acquisition"; see Odlin, 2012). For example, a well-established principle in this field is that people use the knowledge of their L1 to acquire a new language in all the grammar domains (phonetics, lexicon, etc.). An immediate implementation on this point is that it will be easier for students to start from the similarity between L1 and L2 (Molina et al. 2013). The bottom line of this strategy is to start from the sound and the words that, eventually, are identical in the two languages.

#### **Passive listening**

Passive listening is one of the most stable principles in the teaching of a second language since the 1960s / 1970s. A way to perform it is to conduct the class in the L2 language, either completely or not. Other strategies can be adopted, such as listening to a story at the end of a class once a week for just 10 minutes. During the story time, students are engaged in a twofold way: they are very attentive to the tale, focusing on it with passion, and the teacher has the chance to interrupt the reading activity to question and ask clarifications. This is particularly appreciated and efficient in order to increase the vocabulary knowledge, confirming what Elley (1989) proposed: stories repeated three times as well as working on some specific words considerably increases the vocabulary learning. Other activities very useful for passive listening are the applied projects in L2 (such as, cooking and art classes). We know from the literature (Long, 2006) that when students learn a foreign language and, at the same time, some subject-specific content in the same language they gain higher results than a language-only approach.

#### **Active Speaking**

Speaking in an L2 is undoubtedly the goal of every L2 course. Many activities can be implemented to reach this goal, such as pronunciation tasks, repetition tasks, and singing activities. Very often, languages display sounds that are either different or absent in the learners' mother tongue language. For example, Italian and Standard English differ in the /r/ pronunciation: Italian /r/ is a voiced alveolar trill consonant; English /r/ is a voiced alveolar approximant consonant. This issue is particularly challenging for L2 learners. An easy way to help students is to pay deliberate attention to pronunciation (Trofimovich, Gatbonton, 2006), such as looking for some acoustic analogies with such sounds. In this case, some animal sounds, like frogs, are perfect for learning



how to trill the tongue in order to make a perfect Italian /r/. This also meets a principle for which it seems to be easier to learn a new sound in a word or pseudo-words that «has no previous associations for a learner, than to learn it as part of a known word» (Nation, 2011: 449).

Coming to the repetition tasks, it can be arranged in multiple ways. Students have to greet in the second language and they have to answer some polite questions – such as, ‘how do you feel today?’, ‘what day is it?’, etc.– at the beginning of every class. They can also review the content of a previous class at the beginning and at the end of each class at least 3 consecutive times. This assured deep memorization of both the lexicon and the small sentences. Moreover, the story times and the applied projects further give the opportunities to review the acquired knowledge. Finally, students can also sing a L2-song (the same song for a whole month) every class. Songs may refer to some lexical-phonological arguments planned for that month. According to the literature (see, among others, Ludke, 2016), using songs is very useful for better and faster lexicon acquisition, gaining long-term benefits. This mix of activities should allow students to perform sentences, however simple, in a one-year program. It is very important that students learn small full sentences which may guide them in building the more complex one. More specifically, lexical chunks representing pre-patterned grammatical structures may be replicated by changing the lexical vocabulary, reducing the communicative stress and improving the knowledge of the L2 grammatical features.

### **Some strategies**

One of the key features of a good L2 project is the emotional engagement of the students: the whole person is involved in the learning activity and, therefore, the teaching experience has to deal with a multidimensional and multiple-sensorial approach. For instance, according to Lazaraton (2004), the interaction between gesture and speech during a vocabulary explanation shows that «classroom L2 learners receive considerable input in a nonverbal form that may modify and make verbal input (more) comprehensible» (p. 111).

Other strategies can involve some dancing and gym activities, very useful for both keeping the attention up and learning the vocabulary referred to the parts of the body, and having fun at the same time. Painting and drawing are very useful in learning many things, such as food names. Showing pictures of animals when students are asked to shout out their names is particularly appreciated. Watching videos on some handicrafts work very well in the review activity: videos are stopped by the teacher any time the teacher wants to ask some words or phrases referring to something in the given video frame (movie-talk strategy). Finally, an important strategy is to arrange the class according to an “8-minute rule”: every 8 minutes the ongoing activity has to be stopped and changed due to the limited attention span of young students, mixing the multidimensional and multiple-sensorial activities seen above. For example, the first 8 minutes can be dedicated to reviewing (repetition and song); then, 8

minutes can be dedicated to crafts and games; and, finally, the last 8 minutes can be dedicated to a movie-talk activity.

### Concluding Remarks

The present work provided some good techniques and strategies L2 teachers may adopt in order to obtain better and stronger involvement from the student in the early scholar stage (from kindergarten to second grade). The full involvement of the students is the key feature of the learning activity as well as the theoretical-guided multidimensional approach performed by the teacher. Activities are based on listening and speaking strategies, which range from creativity tasks, such as painting and drawing, to practical ones, such as cooking and reciting; from unmoving activity, such as singing, to moving activity, such as jumping and dancing; from memory tasks, such as repetition by heart, to “out the blue” tasks, such as shouting new words out; from traditional support materials, such as paper, to informatic ones, such as videos; etc. Future research will arguably give quantitative information on all the techniques discussed here.

### References

- Greco, M. 2021. A brief practical-guide for L2-teachers: K-2 American children learn Italian. *Italiano LinguaDue*, 13 (1), 141–156.
- Isik, A. 2000. The role of input in second language acquisition. *Review of Applied Linguistics* 129, 1, 225-274.
- Kanno, Y., Stuart, C. 2011. Learning to Become a Second Language Teacher: Identities-in-Practice. *The Modern Language Journal*, 95, 236-252.
- Lazaraton, A. 2004. Gesture and speech in the vocabulary explanations of one ESL Teacher: a microanalytic inquiry. *Language Learning* 54 (1), 79-117.
- Lenneberg, E. H. 1967. *The Biological Foundations of Language*. New York, John Wiley and Sons.
- Long, M. H. 2006. Recasts: the story so far. In Long, M. H. (ed.), *Problems in SLA*, 75-116, Mahwah (NJ), Erlbaum.
- Ludke, K. M. 2016. Singing and arts activities in support of foreign language learning: an exploratory study. *Innovation in Language Learning and Teaching* 12, 371- 386.
- Molina, G. T., Cañado, M. L. P., Agulló, G. L. 2005. Current Approaches and Teaching Methods. Bilingual Programmes. In McLaren, N., Madrid, D., Bueno, A. (eds.), *TEFL in Secondary Education*, Editorial Univesidad de Granada, Granada.
- Nation, I.S.P. 2013. *Learning Vocabulary in Another Language*. CUP
- Muñoz C., Singleton, D. 2011. A critical review of age-related research on L2 ultimate attainment. *Language Teaching* 44, 1, 1-35.
- Odlin, T. 2012. Crosslinguistic Influence in Second Language Acquisition. *The Encyclopedia of Applied Linguistics*, 1562- 1568.
- Tahta S., Wood M., Lowenthal, K. 1981. Age changes in the ability to replicate foreign pronunciation and intonation. *Language and Speech* 24, 4, 363-372.
- Trofimovich, P., Gatbonton, E. 2006. Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal* 90 (4), 519-535.

# The acquisition of English relative clauses by L1 Arabic and Korean speakers

John Hitz

Department of English, University of Central Oklahoma, US

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This study examined the interaction of language transfer and language processing on the second language acquisition of English relative clauses (RCs) by native speakers of Arabic and Korean at intermediate- and advanced-level English proficiency. Experimental participants completed self-paced reading and sentence judgment tasks on-line. Results show that L1 transfer has a powerful impact on intermediate-level learners; however, transfer effects fade at advanced levels, when L2 RC processing constraints are more influential. Taken together, the results of the sentence judgment task and the reading time results indicate that L1 influence and language processing interact in the L2 acquisition of English RCs, and that the mystery of L1 transfer may be solved by analysing its interaction with other SLA factors.

Keywords: language processing, language transfer

## Introduction

Some SLA studies of RC acquisition have argued that L2 learners tend to accept and produce resumptive pronouns (RPs) in the L2 if RPs are acceptable in the L1 (Gass, 1979). Thus, they claim that transfer effects result when L2 learners judge stimuli such as the following:

- 1(a). The relatives<sub>i</sub> [who<sub>i</sub> we visited \_\_\_\_\_<sub>i</sub> last night] enjoyed the evening.  
(gapped condition)
- 1(b). \*The relatives<sub>i</sub> [who<sub>i</sub> we visited *them*<sub>i</sub> last night] enjoyed the evening.  
(resumptive condition)

The RC in sentence 1(a) is described as a “gapped” RC because the parser must link the filler, *who*, with the gap after *visited*. In 1(b), *them* is the resumptive pronoun.

Other SLA studies (e.g., Tezel 1998) have argued that L2 learners’ acceptance and production of RPs in RCs in the L2 is an effect of language processing. As the distance between the filler and the gap increases, the parser is strained because it must hold the filler in short term memory while it searches for the gap. The parser must also use surrounding RC context to infer the gap location; however, in RCs with RPs, the difficulty involved in searching for an empty category is removed as the empty category is overtly expressed. Therefore, RCs with RPs are easier to process.

I argue that both L1 transfer and language processing constraints affect the second language acquisition of RCs, and designed an experiment to tease them apart.

## Methodology

Twenty-one L1 speakers of Arabic, 50 L1 Korean speakers, and 16 L1 English speakers participated in the study. The L1 Arabic and Korean participants were subdivided into intermediate- and advanced-level L2 English proficiency groups based on their scores on the grammar and reading sections of the Michigan Test of English Language Proficiency.

Experimental items consisted of English subject (SRC), direct object (DORC), and oblique RCs (ORC), counterbalanced in gapped and resumptive conditions as in sentences 1(a) and 1(b). Participants reviewed a total of 48 experimental items and 56 fillers. All experimental items were presented online and were randomized by the E-Prime application. Participants read sentences presented in cumulative fashion at their own pace. Each key press revealed a word of an item and would remain on the computer screen until all words in the sentence appeared with subsequent key presses. After all words were read, the sentence disappeared from the screen, and then participants were prompted to rate items on a 4 point scale in which 1=certainly correct, 2=possibly correct, 3=possibly incorrect, and 4=certainly incorrect.

## Predictions

1. L1 influence should cause the L1 Arabic speakers to accept English RCs with RPs more frequently than L1 Korean speakers at equivalent proficiency levels because RPs are acceptable in Arabic direct object and oblique RCs, whereas Korean bans RPs in these RC types.
2. The non-native speakers should accept items in the RP condition more frequently than native speakers because RPs can facilitate RC processing.
3. The non-native speakers should take longer than the native speakers to read the experimental items.
4. The native speakers of Arabic should read RPs more quickly than the native speakers of Korean and English. Native speakers of Arabic will find RPs in English RCs quite normal because they are acceptable in Arabic, whereas L1 Korean and English speakers will pause in surprise on encountering them because they rarely appear in Korean and English RCs.

## Results

Transfer effects were evident in comparisons within L1 groups of RCs in the gapped and resumptive conditions. The Arabic speakers preferred gapped

SRCs to SRCs with RPs, as they do in Arabic ( $t=-7.62, p<.0001$ ). All of the L1 Arabic speakers at the intermediate level spoke Saudi Arabian Arabic, a dialect that allows RPs and gaps in DORCs. They may have transferred their L1 grammar to their judgments of DORCs in the gapped and resumptive conditions; as predicted, the Arabic speakers' average ratings of DORCs in the gapped and RP conditions was close to 2 on the rating scale ("possibly correct") and there was no statistically significant difference between them ( $t=-.33, p=1.00$ ). The Arabic-speaking participants were expected to prefer English ORCs with RPs to gapped ORCs because gapped ORCs are banned in Arabic, a prediction that was confirmed ( $t=4.01, p=.035$ ).

Most of the sentence judgments made by the intermediate-level Korean speakers are in line with L1 transfer hypotheses as well. Korean RCs license gaps and ban RPs in all the RC types used in this study, a distribution that seems to have influenced the participants' judgments of the SRCs and DORCs, but not the ORCs. The Korean participants clearly preferred gapped SRCs to SRCs with RPs ( $t=-7.71, p<.0001$ ) and preferred gapped DORCs to DORCs with RPs ( $t=-5.20, p=.0002$ ), but seemed to perceive gapped ORCs and ORCs with RPs as equally unacceptable ( $t=.85, p=1.00$ ), perhaps because the complexity of these RCs caused excessive strain on the parser.

For advanced-level non-native speakers, the sentence judgment task shows only one possible transfer effect. The Arabic speakers' ratings of gapped and resumptive ORCs did not contrast significantly ( $t=-3.21, p=.37$ ), perhaps because some of them may have transferred their L1 preference for RPs in ORCs just enough to ensure that there was no significant difference. The L1 Korean speakers, on the other hand, preferred gapped ORCs to ORCs in the RP condition at statistically significant levels ( $t=-12.24, p<.0001$ ), a result consistent with the distribution of gaps and RPs in Korean ORCs.

The sentence judgment task also showed that the non-native speakers at both English proficiency levels accepted RPs in DORCs and ORCs more frequently than the native speakers of English at statistically significant levels, a result indicating that RPs alleviate the strain placed on the parser during sentence processing.

The reading time results provide evidence of language transfer. The L1 Korean speakers took longer to read experimental items with RPs as compared with gapped items (Korean intermediate comparison:  $t=6.73, p<.0001$ ; Korean advanced comparison:  $t=7.89, p<.0001$ ). When this comparison was run on Arabic speakers' reading times, however, no statistically significant differences emerged, consistent with experimental predictions (Arabic intermediate comparison ( $t=-3.08, p=.08$ ; Arabic advanced comparison:  $t=-3.15, p=.0532$ ). The Korean speakers, perceiving RPs as odd or unexpected constituents due to L1 influence, slowed down to read them, whereas the Arabic speakers did not decrease their reading speed to the same degree because RPs are common in their L1.

Reading time data of individual words demonstrate that non-native speakers may process RCs more slowly than native speakers, but not in a manner that is qualitatively different. For example, all participants read the verb and the remaining words in the RC predicate more slowly than any individual words preceding them, a result of the extra time required by the parser to assign theta roles on encountering the verb and to find the gap afterwards. The L2 learners, however, needed more time to read individual words in the RC predicate than native speakers, and therefore may have expended comparatively more effort to assign theta roles and find gaps. Therefore, it seems that L2 learners can perform the same RC processing operations as native speakers (i.e. theta role assignment, finding gaps), but with less speed and efficiency. This finding contradicts Marinis et al. (2005), who claim that L2 learners cannot develop complex RC representations with fillers and gaps.

## Discussion

Filipović and Hawkins (2013) claim that L1 transfer is likely to occur when it enhances L2 processing, and when it will not cause communicative disruptions. These claims are supported by the experimental results, which indicate that RPs facilitate RC processing, a phenomenon that may have also increased the likelihood of L1 transfer in the case of the L1 Arabic speakers. Additionally, RPs in English RCs do not lead to breakdowns in communication, as demonstrated by the occasions when native speakers of English accept and produce them. In summary, contrary to Tezel (1998), transfer and processing interact in the SLA of RCs. SLA may therefore benefit from an approach that integrates phenomena such as transfer and processing, and that views SLA theories as complementary descriptions of a complex reality rather than as clashing viewpoints, an integrative paradigm Filipović and Hawkins (2013) describe as Complex Adaptive Systems Theory (CASP).

## References

- Filipović L., and Hawkins, J.A. (2013). Multiple factors in second language acquisition: The CASP model. *Linguistics* 51: 145–176.
- Gass, S. (1979). Language transfer and universal grammatical relations. *Language Learning* 29(2): 327-344.
- Marinis, T., Roberts, L., Felser, C., and Clahsen, H. (2005). Gaps in second language processing. *Studies in Second Language Acquisition* 27: 53-78.
- Tezel, Z. (1998). The acquisition of restrictive relative clause configurations in English as a second language. Ph. D. Dissertation, Indiana University of Pennsylvania.

# The Universal Perceptual Model (UPM) of second language

Georgios P. Georgiou

Department of Languages and Literature, University of Nicosia, Cyprus

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Several speech models have been developed to examine second language (L2) speech patterns, considering that the acquisition of L2 sounds is often challenging for most learners. The Universal Perceptual Model (UPM) is a newly introduced model which provides predictions about the discrimination of L2 phone contrasts. In this paper, the central tenets and current evidence about the model's predictions are briefly discussed, while some revisions are also proposed. UPM predicted with success the discrimination accuracy of nonnative phone contrasts, indicating that it can be a useful L2 speech model. Future research should investigate further the predictability of the model.

Keywords: Universal Perceptual Model, second language, speech perception, assumptions

## Introduction

Several speech acquisition models have been formed to predict the difficulties of learners in perceiving and producing the L2 phones. A newly established speech model, the *Universal Perceptual Model* (UPM) was firstly introduced in Georgiou (2021b). This model has been developed to inform and update the current theory of L2 acquisition and to provide more precise predictions about the ability of learners to discriminate challenging L2 segmental contrasts.

UPM is based on the *Functional Reorganization Hypothesis* (FRH) (Werker, 1995) which argues that although the discrimination of nonnative contrasts declines in adults, there is not a complete 'loss' of sensitivity as a result of L1 experience; instead, a functional reorganization occurs (Werker & Pegg, 1992). Thus, UPM supports that all possible human speech sounds have a mental correlate in the human brain from birth. In contrast, nonnative sounds are initially *inactive* and *disoriented* phonetic units which are activated upon L2 learning onset, and are oriented toward native productions as L2 experience increases. So, at the initial stage, the L2 productions do not match those of native speakers since the robust L1 phonological system works as a filter for L2 sounds (Trubestkoy, 1939)

UPM supports that the phonological space is filled with phonetic categories and there are not any mechanisms that lead to the formation of further phonetic categories as mental representations for each speech sound attested in every human language are available from birth prior to language experience.

The attainment of a native-like pronunciation is not impossible, according to UPM, as speech sounds are ‘universal units’ and can be activated and oriented toward L1 productions at any time of human life if some preconditions are met. Universal units are perceptual in nature, constraining the perception of phonetic categories extracted from the speech signal.

### UPM predictions

UPM introduced three types of perceptual similarity of a single L2 phone to one or more L1 phones. These types include degrees of overlap. Degrees of overlap were firstly used by Faris et al. (2016) for uncategorized L2 phones (i.e., those that failed to reach a predefined categorization threshold). UPM denies the use of any categorization thresholds to form its predictions since different thresholds might affect the categorization type. According to their degree of overlap (see Figure 1), L2 phones might be:

- (1) *non-overlapping*: identified in a different set of above-chance L1 categories,
- (2) *partially overlapping*: to have at least one shared above-chance category
- (3) *completely overlapping*: both nonnative phones are identified within the same above chance L1 category or set of L1 categories.

Above-chance categories are those selected more often than chance. Percentages are used to find the chance score; for example, if the script responses include 5 L1 categories, the chance score is 20%. *One-sample t-tests* are then employed to test whether the classification percentages of the L2 sounds in each L1 category are significantly different or not from the chance score; if  $p < .05$ , classification is more often than chance. Non-overlapping contrasts would be the easiest to distinguish followed by partially overlapping and completely overlapping contrasts. However, completely overlapping contrasts might be distinguished in the same manner as partially overlapping contrasts if the goodness-of-fit ratings (GOF) of the two classified phones differ from each other ( $p < .05$ ).

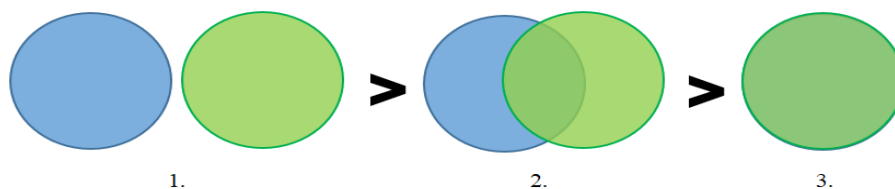


Figure 1: Overlapping degrees of UPM.



## Revised predictions

The predictions of UPM need some clarifications and extensions. For instance, the discrimination of two L2 phones without above-chance responses might be poor to excellent depending on the phonetic similarity of the two L2 phones. This similarity can be estimated by comparing the basic articulatory characteristics of the phones (e.g., place and manner of articulation and voicing for consonants, and height, backness, and lip rounding for vowels). In contrast, the discrimination of one L2 phone that includes at least one above-chance response and another L2 phone without below chance responses might be very good-to-excellent since learners perceive some similarity to one phone and one or more L1 phones and no similarity to the other L1 phone (Faris et al., 2018). Also, above-chance categories should be considered only i. those of which the classification percentage differs significantly from the chance score and ii. those of which the classification percentage is equal or more than the chance score. So, for example, as it can be seen in Table 1, although the 4% classification of the Italian [ɛ] in the Cypriot Greek [i] differed significantly from the 20% chance score ( $p < .05$ ), this response is not above-chance since it is below the chance score. In partially overlapping contrasts, the discrimination might depend on how close the one L2 phone is to the common L1 response of the other L2 phone. For example, if one L2 phone is above chance with a 100% classification score in an x L1 phone, and one other L2 phone is above chance with 70% in an x L1 phone and 30% in a y L1 phone, these phones will be discriminated less accurately than if one L2 phone is above chance with 100% classification score in an x L1 phone, and one other L2 phone is above chance with 60% in an x L1 phone and 40% in a y L1 phone.

## Current evidence

Fifteen Cypriot Greek learners of Italian participated in two psychoacoustic tasks in which they were asked to classify L2 Italian vowels (in [bV] context) in their L1 phonological system, and to discriminate pairs of Italian vowels ([i—e], [e—ɛ], [ɔ—o], [o—u]). Another 10 Italian speakers participated as the control group (see Georgiou, 2021b).

The results of the classification task are shown in Table 1. The discrimination task showed that [i—e] and [e—ɛ] (partially overlapping contrasts) had 67% and 76% correct responses respectively, [ɔ—o] (non-overlapping) had 81% correct responses, and [o—u] (completely overlapping) had 52% correct responses. The analysis has been conducted in R with the use of mixed-effect models. It was found that [o—u] significantly differed from all vowel contrasts ([ɔ—o]:  $\beta = -29.07$ ,  $SE = 4.02$ ,  $t = -7.234$ ,  $p = < .0001$ ; [i—e]:  $\beta = 15.60$ ,  $SE = 4.02$ ,  $t = 3.883$ ,  $p = .0054$ ; [e—ɛ]:  $\beta = 24.00$ ,  $SE = 4.02$ ,  $t = 5.973$ ,  $p = < .0001$ ). The [i—e] contrast significantly differed from [ɔ—o] ( $\beta = -13.47$ ,  $SE = 4.02$ ,  $t = -3.352$ ,  $p = .0270$ ).

Table 1: Classification of the Italian vowels in the Cypriot Greek phonetic categories (%). The parentheses show the GOF (Georgiou, 2021b).

Italian vowels	Cypriot Greek vowels				
	[i]	[e]	[a]	[o]	[u]
[i]	<b>100</b> (4.27)				
[e]	<b>47</b> (3.14)	<b>53</b> (4.18)			
[ɛ]	4 (3.05)	<b>96</b> (4.52)			
[a]			<b>100</b> (4.68)		
[o]				19 (3.19)	<b>81</b> (4.1)
[ɔ]				<b>97</b> (4.32)	3 (3.9)
[u]				9 (4.23)	<b>91</b> (4.21)

Bold represents above-chance responses  $> 20\%$ ,  $p < .05$ .

## Conclusions

UPM successfully predicted the discriminability of the vowel contrasts, yielding that the non-overlapping contrast had the best discrimination followed by the partially overlapping and the completely overlapping contrasts. Future empirical research should aim at testing the predictions of the model in (a) a context where the L2 is widely spoken, (b) speakers with an L1 other than Greek and/or with a different L2, (c) consonantal contrasts, (d) diphthongs, (e) varying contexts (e.g., target phone in initial, final position), and (f) learners with advanced proficiency in the L2.

## References

- Faris, M. M., Best, C. T., Tyler, M. D. 2016. An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *Journal of the Acoustical Society of America* 139, 1, EL1-EL5.
- Faris, M. M., Best, C. T., Tyler, M. D. 2018. Discrimination of uncategorized non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics* 70, 1-19.
- Georgiou, G. P. 2021b. Toward a new model for speech perception: The Universal Perceptual Model (UPM) of Second Language. *Cognitive Processing* 22(2), 277-289.
- Trubetzkoy, N. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht. Translated 1969 by Christine Baltaxe as *Principles of Phonology*. Berkeley and Los Angeles, University of California Press.
- Werker, J. F. 1995. Exploring developmental changes in cross-language speech perception. In Gleitman, L.R., Liberman, M. (eds.), *An invitation to cognitive science. Language: An invitation to cognitive science*, 87–106. The MIT Press.
- Werker, J. F., Pegg, J. E. 1992. Infant speech perception and phonological acquisition. In Ferguson, L. M. C., Stoel-Gammon, C. (eds.), *Phonological Development: Models, Research, and Implications*. York, Parkton, MD.

# An improved method of F0 determination

Grandon Goertz<sup>1</sup>, Terese Anderson<sup>2</sup>

<sup>1</sup>Educational Linguistics, University of New Mexico

<sup>2</sup>Chemistry and Chemical Biology, University of New Mexico

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This research investigated the structure of the speech Fundamental Frequency (F0). The principal objective of this study was to create an improved method for analyzing, understanding, and plotting the F0 values of spoken words. The F0 is created by a variety of models that produce the F0 values for plotting. These F0 values are a represented average of the lower range frequency of sound.

Keywords: fundamental frequency, sub-harmonic modelling, Chebyshev transform

## Introduction

F0 is often understood to be a single, slightly varying waveform resulting from vocal fold vibrations in the throat, which would be consistent with the understanding of a hollow tube model of the throat that is often used to explain the phenomenon.

## F0 determination technology

The F0 of speech is an estimated value and is not measured. Tsangas et al. (2014) and Staudacher, et al. (2016) report that there are at least 10 pitch estimation algorithms for use by linguists. F0 estimation algorithms include Nearly Defect-Free (NDF), TEMPO, SWIPE, SHRP, AAC, and YIN. (Pertusa and Iñesta 2012). Roark (2006) mentioned that there are more than 70 algorithms to estimate F0, which illustrates the difficulty in determining fundamental frequencies. Verde, De Pietro and Sannino (2018) describe a personalized methodological F0 determination that provides better than 77% accuracy, 72% sensitivity and 81% specificity.

Speech contains higher formant bands that are integer multiples of a common low frequency, the fundamental frequency. These harmonic sounds evoke a pitch corresponding to their F0 (Micheyl, Ryan, & Oxenham 2012). The fundamental frequency is a source for the harmonics and must be estimated, inferred, or modelled from the harmonic values (Pertusa and Iñesta 2012; Tsanas, Zañartu, Little & McSharry 2012). Staudacher et al. (2016) mention that, “Many pitch detection algorithms (PDAs) analyze a speech signal by partitioning it into segments and calculating the respective fundamental frequencies (short-term analysis). The length of the segments (frames) limits the minimum frequency or the maximum period to be determined.”

### **Physiology and F0 modeling**

Tsanas, Zañartu, Little & McSharry (2012) mention that “The estimation of the fundamental frequency ( $f_0$ ) is a critical problem in the acoustic characterization of speech signals.” They observe that without genuine ground truth information it is impossible to validate many of the F0 determination algorithms. Tsanas et al. (2012) also summarize that the F0 varies in time, the F0 may change between vocal cycles, the sub-harmonics of the actual F0 appear frequently, and the vocal tract resonances affect the vocal folds resulting in harmonics which may be multiples of the actual F0. For their research Tsanas et al. (2012) generated 92 sustained vowel /**a**/ sounds from physiological examinations of humans and computed the ground truth F0 time series.

Two main variables that affect the range of vocal fold vibration frequency are vocal fold elongation and tissue fiber stress, but other factors come into play. Tsanas et al. (2014) notes that the F0 can change if the F0 is affected by vocal tract resonances that affect the vocal folds in the form of feedback resulting in additional sub-harmonics. These facts indicate that the F0 could contain additional component waveforms. These fractional F0 sub-harmonics appear in the spectrum and modify the waveform values of the F0. Titze et al. (2016) note that the average F0 is predominantly determined by vocal fold length, and other factors include the freedom of movement of the laryngeal muscles that control elongation and collagen density or nonlinearity in tissue fiber tension. The vocal fold tissues consist of three main layers, epithelium, non-muscular lamina propria, and muscle. When the fibers of one layer are under tension, “the layer can be considered a ‘thick string vibrating in a viscous soup.’” The string modes of vibration dominate over the gel modes of vibration and so the F0 is largely determined by the fiber component, but the combined properties of the gel and the fibers add to the total range of normal frequencies. Mannell (2007) notes that waves can interfere with themselves, that the supra-glottal vocal tract may attenuate some sound waves, and that waves can be reflected due to radiation impedance when the vocal tract opens into larger space.

Voice production depends on more than one underlying morphologic parameter including laryngeal framework mechanics, the depth of the vocal fold tissue layers, vocal fold boundary geometry and tissue fiber stress resulting in the F0 that is regulated through two distinct mechanisms (Titze et al. 2016).

### **Procedure**

The CHEBYprime computer program written in Matlab was used to evaluate the vowel F0s and provide data on all the formants. Two hundred ten vowel sounds produced by male and female speakers were evaluated. CHEBYprime uses Chebfun routines (Trefethen 2000; Driscoll, Hale, & Trefethen 2014), a Chebyshev Transform (Boyd 2001), and Singular Value Decomposition procedure (Gold and Morgan 2001) to produce verifiable distinct formants,

consistent with harmonic principles. The F0 was determined by the pitch determination algorithm, SHRP (Sun 2002), which is based on subharmonic to harmonic ratio.

## Results

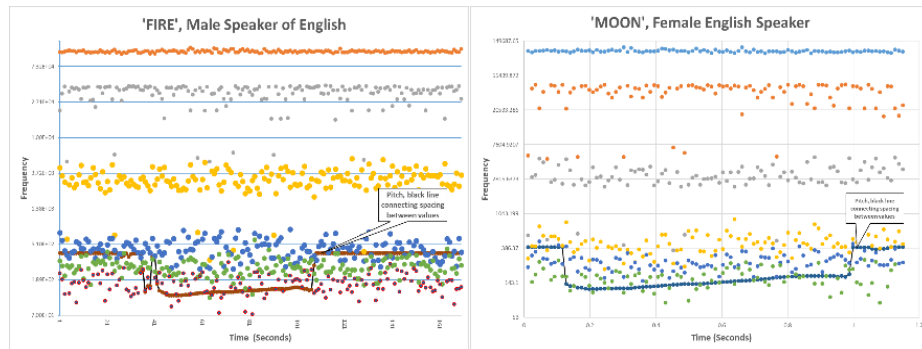


Figure 1a left, 1b right. Formant plots showing clear formant regions as individual points and the F0 linear plot occupying space in the lower formant region. Formants are close to each other, but do not overlap.

The Chebyshev algorithm was able to distinguish vowel formants into harmonic bands of formants, with the uppermost band being out of the hearing range. A Chebyshev transform is suited for speech data analysis because can precisely measure the modulation irregularities of speech data, and does not manipulate output data into the imaginary domain.

## Conclusion

F0 was found to occupy the same frequency space as some of the lower formant bands. Data plots indicate that F0 values are a combination of the lowest of the formant frequency bands, as evidenced by formants occurring in the same frequency (Hertz) region as the F0. This complex F0 structure may arise from vocal folds, vocal fold inconsistencies, echo, and irregularities in the throat.

## References

- Boyd, J. 2001. Chebyshev and Fourier Spectral Methods. Mineola, NY: Dover.
- Gold, B., Nelson M. 2000. Speech and Audio Signal Processing. NY: John Wiley & Sons, Inc.
- Mannell, R. 2007. Standing Waves and Resonance. Macquarie University. [www.ling.mq.edu.au/speech/acoustics/frequency](http://www.ling.mq.edu.au/speech/acoustics/frequency).
- Micheyl, C., Ryan, C., Oxenham, A. 2012. Further evidence that fundamental-frequency difference limens measure pitch discrimination. *Journal of the Acoustical Society of America*. 131(5) May 2012.
- Pertusa, A., Iñesta, J. 2012. Efficient methods for joint estimation of multiple fundamental frequencies in music signals. *EURASIP Journal on Advances in Signal Processing* 2012. 2012:27.
- Roark, R. 2006. Frequency and Voice: Perspectives in the Time Domain. *Journal of Voice*, Volume 20, Issue 3. Pp 325-354
- Staudacher, M., Steixner, V., Griessner, A., Zierhofer, C. 2016. Fast fundamental frequency determination via adaptive autocorrelation. *EURASIP Journal on Audio, Speech, and music Processing*. 2016:17. DOI 10.1186/s13636-016-0095-8.
- Sun, X. 2002. The Determination, Analysis, and Synthesis of fundamental Frequency. Dissertation: Northwestern University.
- Titze I., Riede, T., Mau, T. 2016. Predicting Achievable Fundamental Frequency Ranges in Vocalization Across Species. *PLoS Computational Biology* 12 (6): e1004907. doi:10.1371/journal.pcbi.1004907
- Trefethen, L. 2000. Spectral methods in Matlab. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Driscoll, T., Hale, N., Trefethen, L. (eds.). (2014). Chebfun Guide.
- Tsanas, A., Zañartu, M., Little, M., McSharry, P. 2012. Robust fundamental frequency estimation in sustained vowels using ensembles. *IEEE Transactions on Audio Speech Language Process*.
- Verde, L., De Pietro, G., Sannino, G. 2018. A methodology for voice classification based on the personalized fundamental frequency estimation. *Biomedical Signal Processing and Control*. Vol. 42(April), pp 134-144.

# Code-switching cost in word recognition

Wanying Hu

School of Chinese as a Second Language, Peking University, China

## Abstract

There is still no consensus on the source of code-switching costs. This study focuses on whether the CS cost comes from language processing, especially the word recognition procedure. 37 Chinese-English bilinguals have taken lexicon decision tasks. The results show that orthography does not affect the switching costs, which means there may be no cost in the early stage of mental lexicon processing. The choice of linguistic nodes (L1/L2) affects the switching costs, which means that the code cost may come from the later stage of mental lexicon processing. Our results support that Chinese Hanzi, pinyin, and English store in the same mental lexicon.

Keywords: code-switching, cost, word recognition, orthography, language node

## Introduction

Code-switching refers to the alternating use of two languages in a single utterance, a sentence, or other language components. It is one of the salient features of bilingualism. The significant cost (the CS cost means switching languages relative to staying in the same language) and asymmetrical costs (the CS cost between the first language to the second language and L2 to L1 are asymmetrical) are two common phenomena in code-switching.

The source of code-switching costs is one of the core issues in code-switching from a psycholinguistic perspective, but there is still no consensus. Some studies believe that code-switching cost is from language processing, especially from the mental lexicon (Grainger & Beauvillain, 1987). The other view is that the CS cost comes from task switching or factors other than language (Thomas & Allport, 2000). The Bilingual Interactive Activation Plus Model (BIA+, Dijkstra & van Heuven, 2002), combining the two viewpoints, believes that both mental lexicon and tasks could be the sources of code-switching costs.

This study focuses on whether the CS cost comes from language processing, especially the word recognition procedure. To study the cost source in word recognition, we choose orthography (Chinese dual scripts, Hanzi and Pinyin<sup>1</sup>) and language nodes (L1 or L2) recognition process to test the CS costs. The former represents the early stage of mental lexicon processing, while the latter represents the late stage.

## Method

### Participants and materials

37 Chinese adult learners of English have participated in this experiment. Their English is fluent with a TOEFL score over 80 or equivalent.

180 simple Chinese words from HSK levels (standard Chinese level test with 6 levels in total) 1 to 4 will be used, including 90 nouns and 90 verbs. The words are translated into Pinyin and English directly. The lists are divided into three groups: Hanzi group, Pinyin group, and English group (see Table 1). We confirm that each group includes 30 nouns and 30 verbs, and the levels of words are arranged randomly. We balanced the frequency and concreteness of the words. We also balanced the length and the number of syllables between the English and Pinyin groups.

Table 1. examples of stimulus.

	Hanzi	Pinyin	English
Noun	衣服	fēngǐng	apple
Verb	毕业	zhǔnbèi	invite

All the materials compose three groups. The first group is the Pinyin-Hanzi group with 60 words in Pinyin and 60 words in Hanzi. The order of stimulus is fixed: Hanzi-Hanzi-Pinyin-Pinyin-Pinyin-Hanzi (the data of the underlined parts will collect) to make the repetition trials coming from a pure repetition trial and exclude delaying influences of switching trials (Mosca & de Bot 2017). The second group is the Hanzi-English group, and the third group the English-Pinyin group. Each participant has a different order, sequenced in a Latin Square design. And they finish all the groups twice, 6 blocks in total.

### Procedure

This experiment was conducted with OpenSesame. Participants' task is to decide whether a word is a noun or a verb. Before the online section, they read all the words in a paper version to review whether they are nouns or verbs. Then they practice reading by switching between Hanzi and Pinyin.

During the experimental session, participants judge whether a word is a verb or a noun by pressing a button. The reaction time and error rate are recorded. After the online test, the proficiencies and vocabulary of participants' English and Chinese are tested.

## Results and discussion

### Orthography influence on code-switching

As it is difficult to measure the proficiency differences between Hanzi and Pinyin, we controlled pre-word proficiency and selected the Chinese (the



second words in English-Hanzi/ English-Pinyin groups) to investigate the influence of orthography on the cost (the general results show in Figure 1).

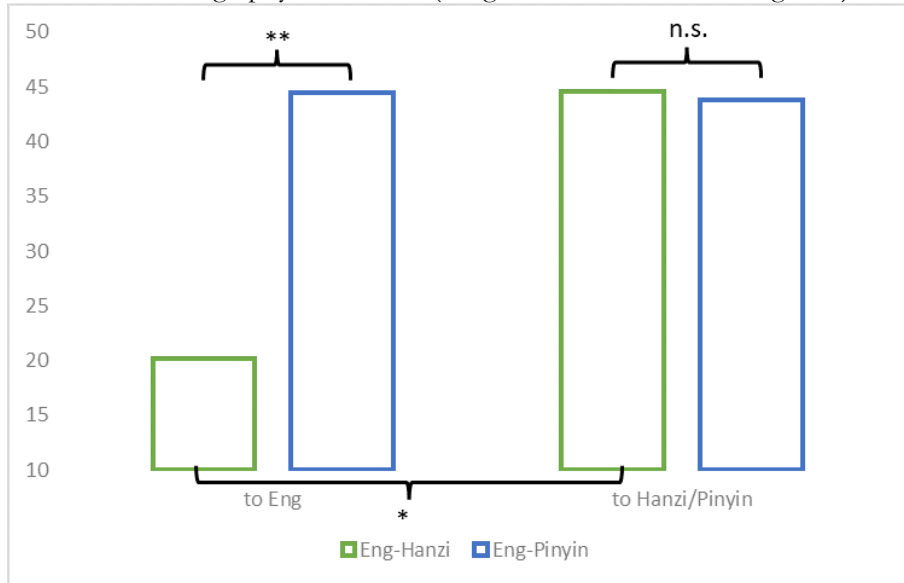


Figure 1. Switching costs among English, Hanzi and Pinyin.

By subject, the results of repeated measure ANOVA shows that the main effect of code-switching is significant,  $F(1,36)=19.361, p<0.05, \eta^2=0.350$ , and the statistical power  $\beta=0.990$ . The main effect of different orthographies to English is significant,  $F(1,36)=248.848, p<0.05, \eta^2=0.874, \beta=1$ ; The interaction is not significant,  $(1,36)=0.002, p>0.05, \eta^2=0, \beta=0.050$ . The results from analysis by item and the mixed-effects model by R are consistent with the former analysis. The paired-samples T-test shows there is no difference between the cost from English to Hanzi ( $M=44.61\text{ms}, SD=68.63$ ) and the cost from English to Pinyin ( $M=44.43\text{ms}, SD=41.02$ ),  $t(36)=0.043, p>0.05$ .

The orthographic differences between Hanzi and Pinyin do not affect the switching. The result implies that the early orthographic stage of word recognition is not the place where produces cost. Therefore, the results show that the orthographies of Chinese, Pinyin, and English are in the same mental lexicon. Their orthographies activate simultaneously during word processing.

### Language nodes effects on code-switching

For this question, We analyze the switching between English and Hanzi, as it is hard to measure the proficiency of Pinyin. We used language proficiency as covariance and establish a mixed effect model analysis by R. The results show that the interaction between language nodes and code-switching is significant ( $t=2.956, =0.018$ ), which means that language nodes (L1/L2) significantly affected the generation of code-switching costs, consistent with the prediction

of the BIA+ model (Dijkstra & van Heuven, 2002). This model assumes that the language node activation in the late stage of word recognition is a source of the cost. Code-switching involves the process of breaking through the recognition threshold of a single language. This process requires cognitive consumption. Our results indicate that when language changes, the language judgment process of bilinguals produces costs.

## Conclusion

In this study, the sources of code-switching costs were investigated during the recognition process of orthography (Hanzi/Pinyin dual system) and language nodes. The results show that: (1) There is no code-switching cost in the early orthographic processing of word recognition. At this stage, the orthographies of both languages are activated simultaneously; (2) Switching costs are generated at the late stage of word recognition, especially in the decision of language node. The results partly support the BIA+ model and indicate that Chinese Hanzi, Pinyin, and English are stored in the same mental lexicon, and the mechanism of Pinyin, Hanzi, and English are consistent.

## Notes

1. Chinese has a dual-script system: the logographic script Hanzi and phonological script Pinyin. Each Chinese word can be written in the logographic script (Hanzi) and Pinyin. These two scripts have a one-to-one corresponding relationship. Hanzi is used commonly in daily life while Pinyin is a useful tool for children and second language learners.

## References

- Dijkstra, T., van Heuven, W.J.B. 2002. The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism-Language and Cognition*, 5(3), 175-197.
- Grainger, J., Beauvillain, C. 1987. Language Blocking and Lexical Access in Bilinguals. *The Quarterly Journal of Experimental Psychology-Human Experimental Psychology*, 39(2), 295-319.
- Mosca, M., de Bot, K. 2017. Bilingual Language Switching: Production vs. Recognition. *Frontiers in Psychology*, 8. doi: 10.3389/fpsyg.2017.00934
- Thomas, M.S.C., Allport, A. 2000. Language Switching Costs in Bilingual Visual Word Recognition. *Journal of Memory and Language*, 43(1), 44-66.

# L2 idiom learning and L1-L2 similarity

Ferdy Hubers, Catia Cucchiarini, Helmer Strik

Centre for Language Studies, Radboud University, the Netherlands

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Learning idiomatic language positively impacts L2 proficiency, but idiomatic expressions like *take the bait* pose difficulties to second language (L2) learners, possibly because only limited classroom time can be devoted to practicing idioms. Alternative methods providing practice through Computer Assisted Language Learning (CALL) systems have gained importance. A CALL study on German learners acquiring Dutch idioms suggested that learning L2 idioms may be affected by L1 and L2 similarity. To further investigate L2 idiom learning in relation to L1-L2 similarity, a study was conducted with L1 Arabic learners of Dutch L2. The results show that while CALL-based practice enhances L2 idiom knowledge, the degree of learning is affected by L1-L2 similarity.

Keywords: idioms, computer assisted language learning, L1-L2 similarity.

## Introduction

Idiomatic expressions like *spin a yarn* pose difficulties to second language (L2) learners, while research shows that they are eager to learn idioms, that this positively impacts L2 proficiency and that idioms are essential L2 vocabulary, (Cieślicka 2006). Classroom activities can devote limited time to practicing idioms, so Computer Assisted Language Learning (CALL) systems (Türker 2016) have been proposed to practice idiomatic expressions. In a CALL study on German learners acquiring Dutch idioms (Cucchiarini, Hubers & Strik 2020) cross-language overlap (the degree to which L2 idioms correspond to L1 idioms) and idiom transparency (the degree of correspondence between the literal and the figurative meaning of an expression) had an impact on learning idioms.

Considering that idiomatic expressions are rooted in the linguistic and cultural background (Boers et al. 2004), L2 idiom learning might be affected by the specific L1-L2 combination, and, in particular, the degree of L1-L2 similarity (i.e., the linguistic distance between the L1 and the L2). Linguistic distance measures have indeed been shown to be impressive predictors of L2 proficiency scores between Indo-European L1s (Schepens, van der Slik & van Hout, 2016).

To investigate how cross-language overlap and transparency impact L2 idiom learning in relation to L1-L2 similarity, a study was conducted with L1 Arabic learners of Dutch L2 as in Cucchiarini et al. (2020) with German L2 learners. Since Arabic is less similar to Dutch than German, the current study

would allow comparisons between different degrees of L1-L2 similarity. We address the following research questions:

1. To what extent does CALL-based practice contribute to L2 idiom knowledge in Arabic learners of Dutch?
2. To what extent is L2 idiom knowledge related to L1-L2 similarity and idiom properties such as transparency and cross-language overlap?

## Method

A group of 14 Arabic L2 learners of Dutch (mean age 23, SD = 3.4; intermediate proficiency level, mean LexTale score 57.2, SD = 5.7) participated in this study. Their performance was compared to that of 42 comparable German L2 learners of Dutch studied in Cucchiarini et al. (2020). A pre-test post-test within subject-design was adopted. From our native benchmark database (Hubers et al. 2018, 2019) 60 idiomatic expressions were selected based on idiom properties and idiom knowledge scores, 30 expressions for intensive practice (12 presentations during practice) and 30 expressions for limited practice (2 presentations during practice). Cross-language overlap was determined by a Dutch-Arabic bilingual as in Cucchiarini et al. (2020), using the following categories: (1) Dutch idiom does not exist in Arabic (NE), (2) Dutch idiom does exist in Arabic, but in completely different words (DW), (3) Dutch idiom does exist in Arabic and has *n* content words in common (nW), and (4) Dutch idiom has a word-to-word correspondent in Arabic (AW). The subjects participated in four consecutive experimental CALL sessions in which they completed four exercise types on Dutch idioms and received instantaneous, automatic feedback from the CALL system. Through pre- and post-tests the participants' idiom knowledge (multiple-choice questions) and vocabulary knowledge (LexTale, Lemhöfer & Broersma 2012) were tested, see Cucchiarini et al. (2020) for details of the methodology.

## Results

The idiom knowledge of the Arabic learners seems to improve after CALL-based training, but not as much as that of the German learners from Cucchiarini et al. (2020) (see Figure 1). To address our research questions and to statistically test the pattern found in Figure 1, we performed a logistic mixed effects regression analysis (Table 1). The performance on the multiple-choice questions (correct/incorrect) formed the dependent variable in the analysis. Five-point scalar Transparency scores were converted to a binary variable (Opaque and Transparent).

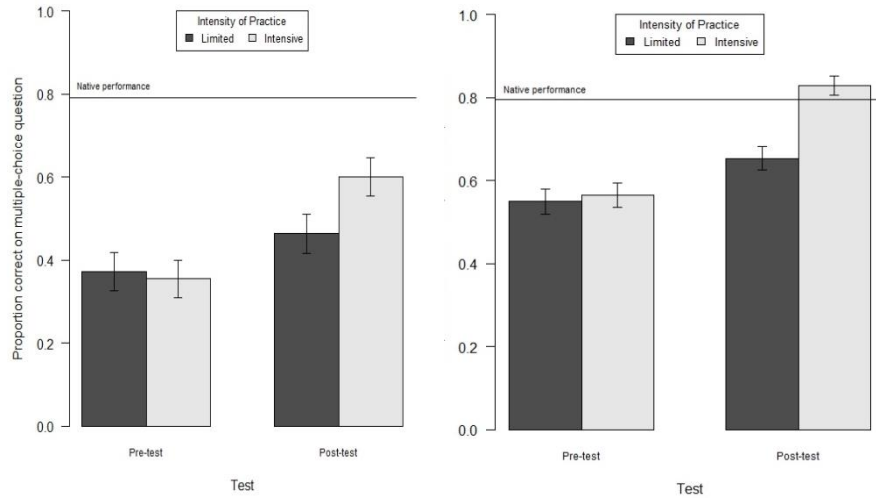


Figure 1. Mean proportion correct and SEs (pre- and post-test) for idioms with limited and intensive practice for Arabic (left panel) and German (right panel) L2 learners. Horizontal line shows mean native performance from a previous study.

Table 1. Final logistic mixed effects regression model.

<b>Fixed effects</b>		<b>Beta</b>	<b>Std. Error</b>	<b>z value</b>
(Intercept)		-4.3690	1.6585	-2.634 **
Pre-test vs. Post-test		-0.4898	0.1532	-3.197 **
Intensive vs. Limited practice		0.7071	0.1533	4.612 ***
Opaque vs. Transparent		0.4043	0.2059	1.964 *
Cross-language overlap DW vs. NE		0.3846	0.2597	1.481
Cross-language overlap nW vs. NE		0.0646	0.3976	0.163
Cross-language overlap AW vs. NE		-0.3291	0.6659	-0.494
LexTale score		0.0674	0.0286	2.354 *
Pre-test x Intensive		-0.7301	0.2184	-3.343 ***
<b>Random effects</b>		<b>Variance</b>	<b>Std. Deviation</b>	
Idiom	Intercept	0.2547	0.5047	
Participant	Intercept	0.3325	0.5767	

Fixed effects in the final model were: (1) Test (reference category: Post-test), (2) Intensity of Practice (reference category: Limited), (3) Transparency (reference category: Opaque), (4) Cross-Language overlap (reference category: NE), (6) LexTale score, and (7) Test x Intensity of Practice. Idioms and Participants (both random intercepts only) were added to the model as random effects. Other interactions did not significantly improve the model fit and were excluded from the analysis. A significant interaction effect emerged between

Test and Intensity of Practice ( $p < .001$ ). In the pre-test no significant difference was found between the idioms with respect to Intensity of Practice (relevelled version of the model:  $\beta = -0.23$ ,  $SE = 0.16$ ,  $p = .883$ ), while in the post-test the Arabic L2 learners performed better on idioms that received intensive practice than on idioms that received limited practice ( $p < .001$ ). Both at pre- and post-test the participants performed significantly better on transparent idioms than on opaque idioms ( $p < .05$ ). While vocabulary knowledge positively affected their performance ( $p < .05$ ), cross-language overlap did not.

## Discussion and conclusions

CALL practice had a significant, positive impact on idiom learning. Cross-language overlap significantly impacted idiom learning in the German speakers (see Cucchiari et al. 2020), but not in the Arabic speakers. For transparency an effect was observed only at post-test for the German learners (Cucchiari et al. 2020), while for the Arabic learners the effect was already visible at pre-test. These results lead us to conclude that while CALL-based practice enhances L2 idiom knowledge, the degree of learning is affected by L1-L2 similarity.

## Acknowledgements

We are grateful to our colleagues in the NWO ISLA project (360-70-510), which is financially supported by the Netherlands Organisation for Scientific Research (NWO).

## References

- Cieślicka, A.B. 2006. Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research* 22(2), 115-144.
- Cucchiari, C., Hubers, F., & Strik, H. 2020. Learning L2 idioms in a CALL environment: The role of practice intensity, modality, and idiom properties. *Computer Assisted Language Learning*.
- Hubers, F., Cucchiari, C., Strik, H., & Dijkstra, T. 2019. Normative Data of Dutch Idiomatic Expressions: Subjective Judgments You Can Bank on. *Frontiers in Psychology*, 10, 1–15. <https://doi.org/10.3389/fpsyg.2019.01075>
- Hubers, F., van Ginkel, W., Cucchiari, C., Strik, H., & Dijkstra, T. 2018. Normative data on Dutch idiomatic expressions: Native speakers. *DANS [Dataset]*. <https://doi.org/https://doi.org/10.17026/dans-zjx-hnsk>
- Lemhöfer, K., & Broersma, M. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Schepens, J. J., van der Slik, F., & van Hout, R. 2016. L1 and L2 Distance Effects in Learning L3 Dutch. *Language Learning*, 66(1).
- Türker, E. 2016. Idiom acquisition by second language learners: the influence of cross-linguistic similarity and context. *The Language Learning Journal*, 1–12.

# Syntactic priming may not lead to language change

Irati Hurtado

Department of Spanish and Portuguese, University of Illinois, US

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This study examines the hypothesis that syntactic priming can lead to language change among bilingual speakers living in language contact areas. A group of second-generation immigrants (Spanish-English bilinguals) living in the United States completed one within-language and one cross-linguistic priming experiment designed to prime participants with overt subject pronouns in Spanish. Each experiment consisted of a pre-test, a treatment, and a post-test task. Participants showed a significant priming effect in both treatment tasks, but the effect did not extend to any of the post-tests. Results suggest that syntactic priming might not be driving language change at the individual level, contrary to what previous studies have proposed.

Keywords: syntactic priming, language change, subject pronoun expression

## Introduction

The mechanisms that drive language change are still unknown, although some potential factors have been proposed in the literature. In this regard, Jäger and Rosenbach (2008) claimed that syntactic priming, the psycholinguistic mechanism whereby a speaker repeats the syntactic structure of a previous utterance, could be one of the mechanisms of language change at the individual level. Some studies have tested this hypothesis by examining bilingual speakers living in language contact areas. A good example of this are the studies conducted by Torres-Cacoullos and Travis (2011, 2016) with Spanish-English bilinguals from New Mexico. These studies found that there was a significant priming effect during spontaneous conversations, both in monolingual and in bilingual (code-switching) mode. Furthermore, the studies showed that within-language priming yielded a stronger effect than cross-linguistic priming.

However, in order to claim that syntactic priming leads to language change, it is important to examine whether speakers continue to use the target structure outside of the priming task itself or in a different conversation (Chang 2008, Pickering and Garrod 2019). If the priming effect does not extend beyond the task or conversation, it would constitute evidence that priming cannot be one of the triggers of language change at the individual level. Despite this evidence being crucial to accept the hypothesis that syntactic priming can lead to language change, no study up to date has included any type of post-test in its experimental design.

## Method

The present study seeks to investigate the hypothesis that syntactic priming can lead to language change. To do this, two experiments were designed (one within-language and the other cross-linguistic) consisting of a pre-test, a treatment, and a post-test each.

The target structure chosen for the study was overt subject pronoun expression in Spanish. Overt subject pronouns are grammatical but pragmatically redundant in Spanish, so speakers usually omit them. On the contrary, English requires overt subject pronouns. Therefore, this structure is ideal to test whether exposure to overt subject pronouns in Spanish (within-language priming) or overt subject pronouns in English (cross-linguistic priming) affects participants' rates of overt pronoun usage in Spanish.

## Participants

A group of 33 Spanish-English bilingual speakers completed the study. These were all second-generation immigrants raised and currently living in the U.S. Midwest.

## Materials

A total of 160 short stories (2-3 sentences long each) were created for the priming experiments. Out of those 160 stories, 120 were in Spanish and 40 were in English. The stories described daily situations in the 2<sup>nd</sup> person singular form. However, none of the stories contained any overt subject pronouns. Thus, all subjects were either lexical or were omitted (this latter only applied to Spanish). Stories were then divided into 8 lists of 20 items each.

Four of those lists (containing only stories in Spanish) were used in the pre-test and post-test tasks of the experiments. Since there were two pre-test and two post-test tasks total, the 4 lists were counterbalanced across participants so that they all went over each of the lists once. Additionally, every story in these lists was complemented by a set of 3-4 keywords.

The remaining 4 lists (2 containing stories in English and 2 in Spanish) were used for the treatment tasks. Here, the stories in English were in fact translations of the stories in Spanish. Thus, there were only two unique lists of stories in this set of four. The lists in Spanish were used in the treatment task of the within-language experiment and the lists in English were used in the treatment task of the cross-linguistic experiment. These lists were also counterbalanced so that participants went over each unique list once. Furthermore, because these stories were used in the treatment tasks, each story was complemented by a prime sentence in the form of a prompt containing an overt subject pronoun (either 'tú' in Spanish or 'you' in English, depending on the experiment).



## Design and procedure

Participants completed the study individually in two experimental sessions. In the first session, participants completed the within-language experiment (pre-test, treatment, and post-test). In the second session, participants completed the cross-linguistic experiment (pre-test, treatment, and post-test). All participants were given a short 10-minute break between tasks and there were at least 2 weeks in between experimental sessions.

During the priming experiments, participants were exposed to the stories one by one on a computer screen and were asked to produce a sentence to continue each of the stories by using either the keywords provided (pre-test and post-test tasks) or the information in the prompt (treatment tasks). All responses had to be produced out-loud in Spanish and were recorded in a computer using Audacity (i.e., a popular audio software).

## Results

All sentences produced in the priming experiments were transcribed and entered into a spreadsheet where additional codes were included to facilitate the analysis process (i.e., experiment, task, participant, item, and response).

Descriptive results showed that participants produced a higher rate of overt subject pronouns in the treatment tasks than in the pre-test and post-test tasks (Figure 1). Furthermore, the rate of overt subject pronouns was numerically higher in the within-language priming experiment than in the cross-linguistic priming experiment. Lastly, overt pronoun production in the post-test tasks did not differ much from that of the pre-test tasks.

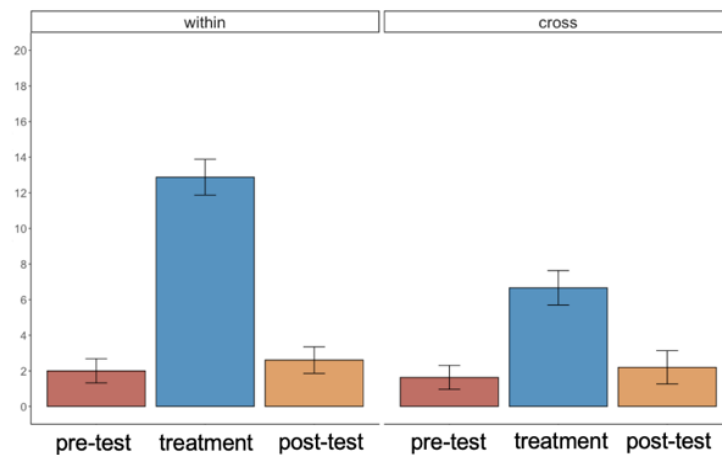


Figure 1. Mean use and *SE* of overt subject pronouns by experiment and task.

In order to confirm whether these observed differences were statistically significant, a mixed-effects binomial logistic regression model was ran using the *lme4* R package (Bates et al. 2015) with response as the binomial dependent variable and experiment and task as independent variables. Participant and item were entered in the model as random effects. The model confirmed that there was a significant priming effect in both treatment tasks ( $\beta = 4.04$ ,  $z = 19.75$ ,  $p < .001$ ) and that the effect was significantly stronger in the within-language experiment than in the cross-linguistic experiment ( $\beta = 1.75$ ,  $z = 11.93$ ,  $p < .001$ ). Moreover, rates of overt pronoun production from the post-test tasks were not significantly different from those of the pre-test tasks ( $\beta = -0.53$ ,  $z = -2.11$ ,  $p = .087$ ).

### General discussion

The purpose of this study was to assess the hypothesis that syntactic priming leads to language change. Two priming experiments including post-tests were designed (one within-language one cross-linguistic) and a group of Spanish-English bilinguals was tested. It was hypothesized that if the priming effect did not extend to the post-test tasks, priming could not be one of the mechanisms driving language change at the individual level. Results showed that this was indeed the case, as there was a significant priming effect in both treatment tasks which did not extend to any of the post-test tasks. Results demonstrated that syntactic priming is a short-lived phenomenon unlikely to impact future language production, at least in the case of the construction tested in these experiments.

### References

- Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67, 1-48.
- Chang, F. 2008. Implicit learning as a mechanism of language change. *Theoretical Linguistics*, 34, 115-122.
- Jäger, G., Rosenbach, A. 2008. Priming and unidirectional language change. *Theoretical Linguistics justified*, 34, 85-113.
- Pickering, M., Garrod, S. 2019. Priming and language change. In Hundt, M., Mollin, S., Pfenninger, S. (eds) 2019, *The changing English language: Psycholinguistic perspectives*, 173-190. UK, Cambridge University Press.
- Torres-Cacoulllos, R., Travis, C. 2011. Testing convergence via code-switching: Priming and the structure of variable subject expression. *International Journal of Bilingualism*, 15, 241-267.
- Torres-Cacoulllos, R., Travis, C. 2016. Two Languages, one effect: Structural priming in spontaneous code-switching. *Bilingualism: Language and Cognition*, 19, 733-753.

# Pragmatic factors facilitate Condition C violations cross-linguistically

Ioannis Iliopoulos, Claudia Felser

Potsdam Research Institute for Multilingualism (PRIM), University of Potsdam,  
Germany

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Binding Condition C prohibits coreference between a referential expression and a pronoun that c-commands it. Several non-structural factors have been argued to increase the acceptability of Condition C violating coreference interpretations, however. We report the results from a questionnaire study investigating the influence of two pragmatic factors - aspectual backgrounding and the temporal ordering of events - on the acceptability of Condition C violations in English and German. Our results show that coreference interpretations are pragmatically facilitated in both languages, lending support to pragmatic approaches to Condition C. Our results suggest, furthermore, that the relative strength of pragmatic or discourse-level factors may differ even across typologically closely related languages.

Keywords: Condition C, aspectual backgrounding, temporal ordering

## Background

Binding Condition C was originally proposed as a syntactic constraint which holds that a referential (R-)expression such as a proper name must be free (Chomsky, 1981). Coreference between an R-expression and a pronoun or another R-expression that c-commands it is prohibited, as illustrated in (1).

However, exceptions to Condition C, with an R-expression and a c-commanding pronoun being able to refer to the same entity or individual as in (2) below, have frequently been noted in the theoretical linguistic literature (see e.g. Grodzinsky & Reinhart, 1993).

(1) \*He<sub>i</sub>/\*John<sub>i</sub> said that John<sub>i</sub> would win. (Chomsky 1981: 193, ex. 25)

Whilst coreference in (2) is uncontroversial, the degree to which other Condition C violations are considered acceptable seems to vary among speakers and to be influenced by multiple factors beyond syntax. Most of the available evidence is merely anecdotal, however.

(2) He<sub>i</sub> is [Colonel Weisskopf]<sub>i</sub>. (Grodzinsky & Reinhart 1993: 78, ex. 19a)

Harris and Bates (2002, exp.1) carried out a coreference judgement task to examine how clausal backgrounding would affect the acceptability of Condition C violating coreference readings in English. Either progressive aspect as in (3b) or the past perfect tense were used as a way of backgrounding the main clause.

- (3) a ?He<sub>i</sub> threatened to leave when Billy<sub>i</sub> noticed that the computer had died.  
b ?He<sub>i</sub> was threatening to leave when Billy<sub>i</sub> noticed that the computer had died.

Their results showed that coreference in Condition C violating contexts was permitted about 60% of the time in the non-backgrounded condition (3a), and that the acceptance rate rose significantly (to above 70%) when the main clause was backgrounded as in (3b). Harris and Bates concluded that aspectual backgrounding has a similar function as clausal subordination in facilitating coreference readings. Acceptable condition C violations have also been observed in experimental studies on German (Drummer & Felser, 2018; Patterson & Felser, 2019), but the role of pragmatic factors was not examined here.

Several theoretical approaches have tried to account for Condition C violability. While earlier pragmatic approaches (e.g. Bolinger, 1979; Huang, 2000; Levinson, 1991) had difficulty accounting for the role of c-command in coreference computation, Schlenker (2005) provides a theoretical framework that combines syntax and pragmatics-driven approaches. He claims that a referential NP can be used to refer to a pre-introduced referent only when it affects the denotation of the description or provides a pragmatic effect (e.g. by using epithets). These two criteria are evaluated every time a new sentence is processed. Another factor to be considered is the prominence of the entities involved. According to Schlenker (2005: 403), one way of rendering an entity more prominent is to denote it by an expression in a c-commanding position.

From the perspective of pragmatic approaches to Condition C, variation in the acceptability of Condition C violations can be explained by assuming that speakers may perceive the need for using a referential NP differently when it comes to referring to a previously mentioned entity. To our knowledge, the question of whether pragmatic factors facilitate Condition C violations in a similar way across languages has never been examined experimentally.

## Methodology

We carried out a questionnaire study to gauge the extent to which aspectual backgrounding (as in 4 and 5) and the temporal ordering of events (6) affect the likelihood of a c-commanding cataphoric pronoun being interpreted as coreferential with an R-expression (such as *Jamie* in (4)) in English and German.

- (4) He {crossed/was crossing} the road when Jamie raised his hand to wave goodbye.
- (5) She {climbed/had already climbed} up the tree when Cindy saw a giant squirrel.
- (6) She met some colleagues {before/after} Emma had lunch with her brother.

61 native German (mean age: 24.8, SD: 7.5) and 61 native English speakers (mean age: 38.1 years, SD: 13.9) completed a web-based antecedent evaluation questionnaire in their native language. The design of the English and German questionnaires was parallel. Ten experimental item pairs were created as in (4) and (5), with either progressive aspect or past perfect used as backgrounding devices. Note that, unlike in English, progressive aspect is not grammaticalized in German but can be signalled explicitly by a prepositional 'pseudoprogressive' construction (e.g. Er war am Arbeiten, lit. 'He was at the work<sub>KINF</sub>'). Ten further experimental stimulus pairs were created which contained temporal conjunctions (6). While the conjunction *before* signals that the two events described appear in chronological order, the conjunction *after* signals the reverse order. For each stimulus sentence, participants were asked to indicate whether coreference between a pronoun and a named character was possible.

## Results

For the statistical analysis we ran three generalised linear mixed-effects models in R, one for every condition pair: 1. *progressive/non-progressive*, 2. *past perfect/non past perfect*, 3. *before/after*, using the *glmer()*-function of the *lme4*-package. All models contained GROUP (ENG, GER) and CONDITION as fixed effects in interaction, with subject and item as random effects. CONDITION was significant for the first two condition pairs ( $p < .01$ ), and GROUP for the third one ( $p < .01$ ). Their interaction was significant for the first and the third model ( $p < .01$ ). Post-hoc tests showed that in both participant groups, both kinds of aspectual backgrounding led to a significant increase in coreference interpretations, confirming and extending previous findings for English (Harris & Bates, 2002). However, the type of temporal conjunction did not reliably affect German speakers' responses, whilst English speakers allowed for significantly more coreference interpretations for *before* than for *after*. In addition, Table 1 shows that English speakers were more reluctant than German speakers to violate Condition C across all conditions (overall coreference acceptance: 38% for English vs. 67% for German).

Table 1. Mean proportions of accepted coreference interpretation per language group and condition (ENG = English, GER = German).

	Condition	Language	Coreference
+backgrounding	+past perfect	ENG	0.62
		GER	0.75
	+progressive	ENG	0.62
		GER	0.75
-backgrounding	-past perfect	ENG	0.23
		GER	0.40
	-progressive	ENG	0.29
		GER	0.59
temporal connectives	before	ENG	0.31
		GER	0.79
	after	ENG	0.19
		GER	0.74

## Conclusion

Our results confirm that clausal backgrounding can facilitate Condition C violation, making backwards anaphora easier. Aspectual backgrounding seems to promote coreference in Condition C contexts cross-linguistically, while the temporal ordering of events affected coreference interpretation only in English. Taken together, our results support pragmatic approaches to Condition C but indicate that pragmatic constraints on coreference may vary in strength across different languages even if these are closely related typologically.

## References

- Bolinger, D. 1979. Pronouns in discourse. In Givón, T. (eds.), *Syntax and semantics* 12, *Discourse and syntax*, 313-342. New York, Academic Press.
- Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht, Foris.
- Drummer, J.-D., & Felser, C. 2018. Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language* 101, 97-113.
- Harris, C. L., Bates, E. A. 2002. Clausal backgrounding and pronominal reference: A functionalist approach to c-command. *Lang Cogn Process* 17, 237-269.
- Huang, Y. 2000. *Anaphora. A cross-linguistic study*. Oxford: Oxford University Press.
- Levinson, S. 1991. Pragmatic reduction of the Binding Conditions revisited. *Journal of Linguistics* 27, 107-61.
- Patterson, C., & Felser, C. 2019. Delayed application of binding condition C during cataphoric pronoun resolution. *Journal of Psycholinguistic Research* 48, 453-475.
- Schlenker, P. 2005b. Minimize restrictors! (Notes on definite descriptions, condition C and epithets). In Maier, E., Bary C., Huitink J. (ed.), *Proc. of Sinn und Bedeutung*, 385-416, Nijmegen, The Netherlands.

# Prosodic manifestations of conversational roles in two types of collaborative tasks

Tatiana Kachkovskaia, Pavel Kholiavin, Anna Mamushina, Alla Menshikova

Department of Phonetics, St Petersburg University, Russia  
<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The acoustic features of the speaker's voice are liable to change due to a number of factors, such as success of communication, social distance between the interlocutors etc., and conversational role. The paper aims to find out how pre-defined roles of leader and follower are manifested in the speakers' prosodic features depending on the task they perform: map task vs. card-matching game. In general, information giver tends to speak louder and with higher maximum F0; temporal changes between roles are speaker-dependent. The two types of tasks differ in the prosodic manifestations of conversational roles: in map task the roles of giver and follower differ much more consistently; in card games speakers rarely reveal prosodically well-defined roles.

Keywords: speech prosody, dialogue, role-switching, map task, speaker variability.

## Introduction

Much of the research of dialogue speech uses collaborative tasks to initiate active involvement of both interlocutors into the conversation. Among the most popular tasks are various types of card games, and the popular map task, where one speaker needs to describe a route drawn on a map to his/her interlocutor. It is often the case with collaborative tasks that speakers play either the role of information giver (the one who describes something, interviews the other speaker etc.), or the role of information follower (the one who listens to a description, is interviewed etc.). It has been shown that conversation roles are indeed a source of intra-speaker variability. It has been shown that information givers have longer turn duration [Forsyth et al. 2007] and more speech disfluencies [Bull, Aylett, 1998]; pause duration at turn transitions is greater when the turn is passed from the giver to the follower [Branigan et al. 1999]; between roles there are differences in speech rate [Karpiński et al. 2014]. Including the parameter of role in the analysis enables the researcher to observe curious interactions between role and gender, to trace in more detail the strategies of speech entrainment.

In this research we are estimating the influence of conversational roles on the speakers' prosodic features. As an example of a highly role-dependent collaborative task we take the classical map task. For comparison, we apply the same procedure to a card game which is presumably less role-dependent.

## Material

This research is based on the corpus of Russian dialogue speech SibLing [Kachkovskaia et al. 2020]. The basic set of speakers consists of 10 pairs of same-gender siblings aged 23 to 40, and each of them was recorded in five dialogue settings: with the other sibling, a close friend of the same gender and similar age, a stranger of the same gender and similar age, a stranger of the opposite gender and similar age, and a stranger of the same gender, greater age and higher job position (“boss”). In total, the corpus contains 90 dialogues.

The card matching game was based on searching for similarities in two decks of ten whimsical pictures. The speakers took turns to describe their picture; thus, they were supposed to swap roles several times. In map task, the interlocutors were asked to guide each other through a set of schematic maps, changing the roles 4 times. The maps were not identical—some of the objects were swapped or replaced, for longer discussion. The recorded dialogues were segmented into subdialogues, each containing a discussion of a single route or card. In map tasks, speakers usually had clearly defined roles. In card games, roles could be distinguished well only at the beginning; then, the discussion became more like a free conversation and, as a result, more role-neutral. This is why for the card matching task we only analysed the first two card descriptions.

## Method

For each turn in a recording, the following features were extracted: mean and maximum  $F_0$ , mean and maximum loudness, speech tempo. We removed all the turns shorter than 1.2 s from the analysis, to avoid the noise in the data that could be brought by short or mistakenly detected turns.  $F_0$  was computed using REAPER software (<https://github.com/google/REAPER>); for loudness we implemented the formula described in GeMAPS [Eyben et al. 2016]. Speech tempo was computed in syllables per second based on orthographic transcriptions. Each hesitation was counted as 3 syllables.

A series of Welch's two-tailed independent t-tests was run to analyze the significance of differences in feature values between different roles.  $H_0$  was stated as follows: there is no significant difference between the ways a person speaks as leader and as follower. For each speaker in each dialogue and for each task, t-tests were run to test this hypothesis for each of the prosodic features.

## Results

Figure 1 illustrates the differences between role switching in card game and map task for maximum  $F_0$  and speech tempo. The data are grouped with respect to speaker's gender and the interlocutor's relationship to the speaker.

Statistical analysis has shown that there are much more dialogues with significant difference between the roles within the map task than within the card game. Maximum  $F_0$  is typically higher in leaders than in followers; the



difference is statistically significant in 77 % cases in map tasks and only in 13 % cases in card games. Mean  $F_0$  is a weaker cue which hardly differs across roles at all. This is a reasonable finding, as speakers can easily change their maximum  $F_0$  but not the minimum; as a result, mean  $F_0$  has smaller variation. Speech tempo differs across roles more rarely than maximum  $F_0$ : 36 % cases in map task and 8 % cases in card games. But more importantly, some participants spoke faster as leaders (in map tasks, 26 %), while some—slower (in map tasks, 10 %). Loudness is generally higher in leaders than in followers. Maximum loudness shows significant differences across roles in 99 % cases in map tasks and 30 % cases in card games; mean loudness—in 91 % and 28 %, respectively.

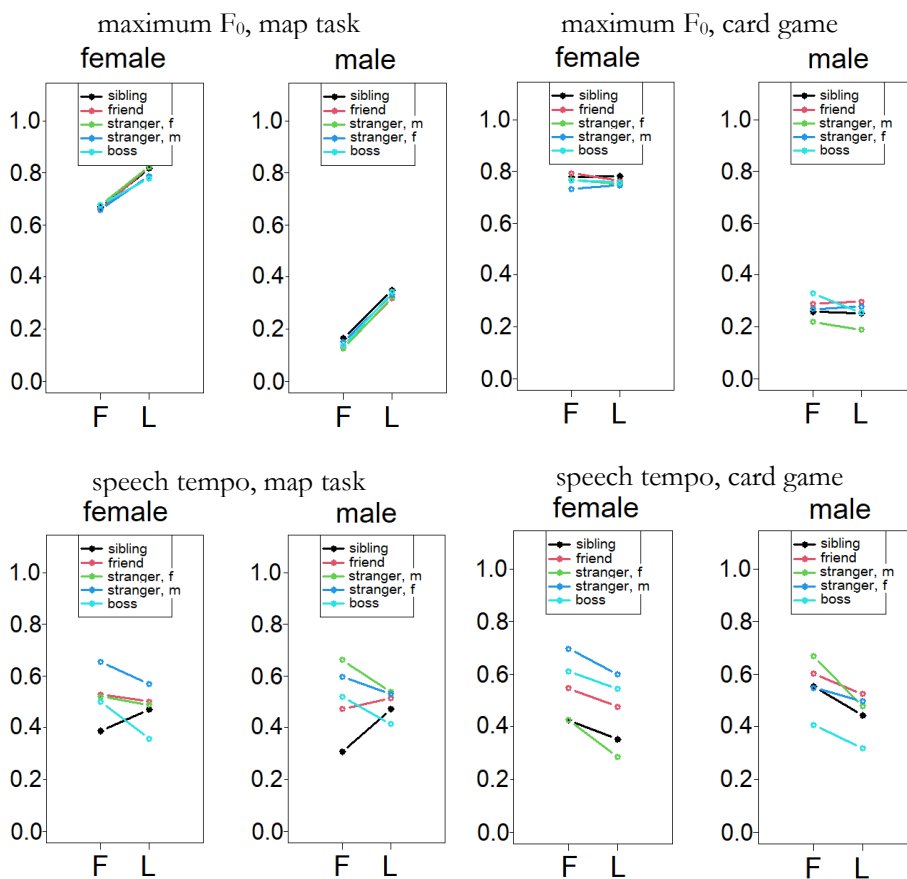


Figure 1. Relative differences in prosodic features across conversational roles (follower, F, vs. leader, L) in map tasks and card games. Various colours refer to types of relationship between the speaker and his/her interlocutor.

## Discussion

When the collaborative task requires well-defined conversation roles, the interlocutors indeed speak differently according to whether they need to behave as leaders or followers. The card-matching game initiates a different type of speakers' behaviour: while in the map task the interlocutors play the “externally imposed” roles, in the card game they seem to behave more naturally. In the latter case there might still be some hierarchy, but this is probably due to the actual speaker's roles. In fact, determining the speakers' roles and hierarchy in a natural dialogue is not an easy task by itself, as relying on various types of questionnaires only provides subjective data. Knowing about the changes in acoustic features across roles could help the researchers solve this problem.

Speakers' behaviour in terms of expressing conversational roles might depend on the relationship between the interlocutors. This can be seen in Figure 1 for tempo in map task, where dialogues between siblings show opposite tendencies compared with other dialogues. A thorough analysis of all possible factors will be the next step of this research.

## Acknowledgements

This work is supported by the Russian Science Foundation, grant 19-78-10046 “Phonetic manifestations of communication accommodation in dialogue”.

## References

- Branigan, H., Lickley, R., McKelvie, D. 1999. Non-linguistic influences on rates of disfluency in spontaneous speech. Proc. 14<sup>th</sup> International Congress of Phonetic Sciences, San Francisco, USA.
- Bull, M., Aylett, M. 1998. An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. Proc. 5<sup>th</sup> International Conference on Spoken Language Processing, 1175–1178, Sydney, Australia.
- Eyben, F., Scherer, K. R., Schuller, B., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., Truong, K. P. 2016. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing* 7 (2), 190–202.
- Forsyth, R., Clarke, D., Lam, P. W. Y. 2007. Timelines, transcription and turn-taking: A corpus-driven approach to simultaneous speech. Proc. 4<sup>th</sup> Corpus Linguistics conference, University of Birmingham, UK.
- Kachkovskaia, T., Chukaeva, T., Evdokimova, V., Kholiavin, P., Kocharov, D., Kriakina, N., Mamushina, A., Menshikova, A., Zimina, S. 2020. SibLing corpus of Russian dialogue speech designed for research on speech entrainment. Proc. 12<sup>th</sup> Language Resources and Evaluation Conference, 6556–6561, Marseille, France.
- Karpiński, M., Klessa, K., Czoska, A. 2014. Local and global convergence in the temporal domain in Polish task-oriented dialogue. Proc. 7<sup>th</sup> International Conference on Speech Prosody, 743–747, Dublin, Ireland.

# L2 English request strategies in Cyprus setting

Sviatlana Karpava

Department of English Studies, University of Cyprus

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This study investigated pragmatic competence and awareness of L2 learners of English. The focus was on their request strategies in L2 English, level of (in)directness, request perspective, internal and external modification and the task effect. The participants of the study were 80 Cypriot Greek undergraduate students. The pragmatic tests: discourse completion task, multiple-choice questionnaire and a role-play were used in order to assess L2 learners' pragmatic competence in oral discourse. The tasks had four conditions regarding power and social distance variables: [+social distance/status equals], [-social distance/status equals], [+social distance/hearer dominance] and [-social distance/hearer dominance]. Overall, it was found that social distance, hearer dominance, power, familiarity and imposition affect the degree of (in)directness, request perspective, internal and external modification.

Key words: Requests, (in)directness, perspective, internal, external modification

## Introduction

Request speech acts and mitigating strategies in L1 and L2 have been widely and thoroughly studied across languages and cultures revealing cross-linguistic differences (Economidou-Kogetsidis, 2013). Appropriate use of the request form in a social context depends on such factors as L1, culture, age and gender of the interlocutors, social distance and power relations as well as the degree of imposition. Native speakers acquire pragmatic competence subconsciously through exposure to native language from birth, which is not the case with L2 learners who need to put a conscious effort in order to learn pragmatic rules in L2 (Webman Shafran, 2019). This study aimed to investigate (in)directness of L2 English request strategies used by L1 Cypriot Greek (CG) undergraduate students, request perspective as well as internal and external modification and whether they are affected by the type of the task: oral vs. written along with power and social distance variables. The focus was on their pragmatic behaviour and interactional practices in different social settings and possible transfer from L1 CG into L2 English (Ogiermann & Bella, 2020; Taguchi & Roever, 2017).

## Study

The participants of the study were 80 Greek Cypriot (CG) first year undergraduate students (17-25 years old, 44 male and 36 female, normal speech

and hearing), who were learners of L2 English at a private, English-speaking university in Cyprus. Their L2 English proficiency was from low intermediate to advanced (5-9 IELTS scores, Mean 6.5). The pragmatic tests based on the Speech Act Theory as a theoretical framework. Discourse completion task (DCT) and multiple-choice questionnaire) were implemented in the research. In addition, an oral role-play task was used in order to assess L2 learners' pragmatic competence in oral discourse. The tasks had four conditions regarding power and distance variables: [+social distance, status equals], [-social distance, status equals], [+social distance, hearer dominance] and [-social distance, hearer dominance], see examples (1)-(4) of the four conditions of the DCT:

- (1) This is your first day at university. During your class you realize that you have forgotten your pen at home. You don't know the student who sits next to you but you want to borrow a pen from him/her. What would you say? [+social distance/status equals];
- (2) It is a lunch time. You are at the cafeteria with your friends. It seems that your mobile phone does not work but you urgently need to call home. You want to use the mobile phone of your friend. What would you say? [-social distance/status equals];
- (3) Next week you are going to take part in the university sports competition. You want to ask your professor to extend the deadline for the assignment submission. What would you say? [-social distance/hearer dominance];
- (4) You are in an unknown city and you need to go to the city centre. You see a policeman and want to ask him for directions. What would you say? [+social distance/hearer dominance]. The linguistic (socio-economic) background questionnaires were used as well.

## Results and discussion

The results showed that the participants had a tendency to use conventionally indirect strategy more than other strategies in all three tasks, which supports the idea of centrality of conventionalization in politeness, which is in line with the frequency-based or habit-based account of politeness and conventionalization that is related to inherent evaluation of the expression, the context and "meta-knowledge about not what expressions mean but how often they mean that." (Terkourafi, 2015: 17), see Table 1.

The participants had an overall strong preference for the hearer perspective due to a possible L1 transfer from CG in terms of pragmatic knowledge and preferred perspective, see Table 2. The findings of our study give a further evidence in support of the general pattern of the underuse of lexical modifiers by L2 English learners in comparison to native English speakers (Economidou-

Kogetsidis, 2008, 2009), see Tables 3 and 4 regarding internal and external modification.

Table 1. Degree of (in)directness: Total scores in the three tasks.

Task	D	CI	Non-ID	NAR
MCQ (640)	85	446	77	32
	13.29%	69.64%	12.10%	4.97%
DCT (640)	51	545	38	6
	7.96%	85.18%	5.93%	0.93%
ORP (640)	52	452	61	75
	8.18%	70.68%	9.48%	11.66%
Chi-square DCT/ORP	$\chi^2=12.293$ , df=1, p=.714	$\chi^2=39.337$ , df=1, p=.589	$\chi^2=5.533$ , df=1, p=.938	$\chi^2=2.734$ , df=1, p=.950
Chi-square DCT/MCQ	$\chi^2=15.721$ , df=1, p=.473	$\chi^2=45.236$ , df=1, p=.037	$\chi^2=22.858$ , df=1, p=.118	$\chi^2=4.288$ , df=1, p=.368
Chi-square MCQ/ORP	$\chi^2=13.104$ , df=1, p=.665	$\chi^2=38.715$ , df=1, p=.306	$\chi^2=11.884$ , df=1, p=.455	$\chi^2=9.139$ , df=1, p=.331

\*MCQ=Multiple Choice Questionnaire; DCT=Discourse completion task; ORP=Oral Role Play; D=Direct; CI=Conventionally Indirect; Non-CI=Non-Conventionally Indirect; NAR=No Act of Request.

Table 2. Request perspective: Written vs. oral task.

Request Perspective	Hearer	Speaker	Inclusive	Impersonal
DCT (634)	347	226	12	49
	54.73%	35.64%	1.89%	7.74%
ORP (565)	345	121	5	94
	61%	21.33%	0.86%	16.81%
$\chi^2$	72.703	36.017	.995	9.828
df	1	1	1	1
p	.066	.142	.319	.875

\*DCT=Discourse completion task; ORP=Oral Role Play.

Table 3. Internal modification: Written vs. oral task.

IM	ZM	MP	CD/O	D	U	H	S	C	A
DCT (634)	346	203	59	12	5	0	5	2	2
	54.57	32.01	9.35	1.89	0.78	0	0.78	0.31	0.31

ORP (565)	332	188	29	3	7	4	1	1	0
	58.83	33.18	5.17	0.45	1.29	0.64	0.22	0.22	0
$\chi^2$	41.74	44.16	13.39	.195	.631	N/A	–	–	–
df	1	1	1	1	1	N/A	–	–	–
p	.759	.669	.572	.907	.427	N/A	–	–	–

\*IM=internal modification; ZM=zero marking; MP=marker ‘please’; CD/O=consultative devices/openers; D=downtoners; U=understaters; H=hedges; S=subjectivizers; C=cajolers; A=appealers/ \*DCT=Discourse completion task; ORP=Oral Role Play.

Table 4. External modification: Written vs. oral tasks.

EM	ZM	G	D	P	G/P	PR/MP	I/M	A
DCT (634)	176	289	2	30	7	3	10	117
	27.76	45.58	0.34	4.73	1.1	0.47	1.57	18.45
ORP (565)	358	114	1	0	3	6	3	80
	63.36	20.28	0.21	0	0.43	1.07	0.43	14.22
$\chi^2$	41.56	45.98	.018	–	.153	.195	.284	27.45
df	1	1	1	–	1	1	1	1
p	.925	.556	.894	–	.695	.658	.867	.493

\*EM=external modification; ZM=zero marking; G=grounders; D=disarmers; P=preparators; G/P=getting a precommitment; PR/MP=promise of a reward/make a promise; I/M=imposition minimizer; A=apology.

## References

- Economidou-Kogetsidis, M. 2008. Internal and external mitigation in interlanguage request production: The case of Greek learners of English. *Journal of Politeness Research: Language, Behaviour, Culture* 4(1), 111-138.
- Economidou-Kogetsidis, M. 2009. Interlanguage request modification: The use of lexical/phrasal downgraders and mitigating supportive moves. *Multilingua* 28(1), 79-112.
- Economidou-Kogetsidis, M. 2013. Strategies, modification and perspective in native speakers' requests: A comparison of WDCT and naturally-occurring requests. *Journal of Pragmatics* 53(1), 21-38.
- Ogiermann, E., Bella, S. 2020. An interlanguage study of request perspective: Evidence from German, Greek, Polish and Russian. *Contrastive Pragmatic*, 1-30.
- Taguchi, N., Roever, C. 2017. *Second Language Pragmatics*. Oxford University Press, Oxford.
- Terkourafi, M. 2015. Conventionalization: A new agenda for im/politeness research *Journal of Pragmatics* 86, 11-18.
- Webman Shafran, R. 2019. Level of directness and the use of please in requests in English by native speakers of Arabic and Hebrew. *Journal of Pragmatics* 148, 1-11.

# Vocabulary skills in Down Syndrome and typical development

Dimitra Katsarou<sup>1</sup>, Georgia Andreou<sup>2</sup>

<sup>1</sup>School of Early Childhood Education, Aristotle University of Thessaloniki, Greece

<sup>2</sup>Department of Special Needs Education, University of Thessaly, Greece

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Down Syndrome is the most common developmental disorder characterized by mild to moderate intellectual disability. Several studies have reported poor language and prosodic skills, phonological problems due to their deficits caused by intellectual disabilities. This research aims to study over time the expressive use of language in 10 children with Down Syndrome and in 10 children with typical development matched on their chronological age and nonverbal intelligence. We used t tests analyses and the results showed statistically significant development of the oral vocabulary skills in the children with typical development, but not in the children with Down Syndrome, with the typically developing children showing faster and better development in the oral vocabulary skills than the children with Down Syndrome.

Keywords: Down Syndrome, toddlers, expressive language, longitudinal study, Greek language

## Introduction

Down syndrome is the most common developmental disorder characterized by mild to moderate intellectual disability. Several studies have reported poor language and prosodic skills, phonological problems due to their deficits caused by intellectual disabilities, mouth and tongue abnormalities and dental issues (eg Abbeduto, Warren & Conners, 2007; Levy & Eilam 2013). Older children with Down Syndrome show expressive vocabulary deficits relative to nonverbal cognitive levels. Expressive vocabulary levels on standardized tests and during language samples for children and adolescents with Down Syndrome were delayed to comparison groups of younger typically developing children at similar mental age or to chronological age matched groups. These deficits extend across the domain of vocabulary and phonology production (eg Griego et al 2015).

Few longitudinal linguistic studies have been conducted so far, mainly in the past decades, due to many mental and body health issues these children face. Most of these studies found that vocabulary skills in children with Down syndrome appear to lag behind their cognitive development (eg Zampini & D'Odorico 2013). Given the paucity in Greek language this research aims to study over time the oral vocabulary skills in 10 children with Down Syndrome

and 10 children with typical development matched on their chronological age and nonverbal intelligence.

## Methodology

Our hypotheses are a) children with typical development would present better oral vocabulary skills over time than children with Down Syndrome and b) children with typical development would increase their oral vocabulary skills over time more than children with Down Syndrome. For the purpose of this study 10 children with Down Syndrome and 10 children with typical development were tested. All children were matched on their chronological age and on nonverbal intelligence with a standardized test (Tzouriadou, Anagnostopoulou, Toutountzi, & Psoinou, 2008- Detroit Test of Learning Aptitude). Children with Down Syndrome had their mental intelligence measured by WISC III, three months before the study by a certified psychologist, and all of them had moderate intellectual disability. Moreover, children with Down Syndrome were tested by karyotype test, had a typical trisomy 21 and all of them were born and raised in Greece and Greek was their native language. Neither of the children had any previous hearing or any other clinical problems that could affect the results. Parents gave their consent and all necessary ethics were taken into consideration.

Children were firstly tested at the age of 4 to 6 years and again at the age of 12 to 14 years with a standardized Language Test, and more specifically with the subscale of oral vocabulary (Tzouriadou, Singolitou, Anagnostopoulou, & Vakola, 2008- L-a-t-o). The first oral vocabulary task, suitable for ages 4 to 6 years old, consists of two parts. The first part has 14 items and measures the child's ability to find the correct described word. For example, the researcher says /thelo na mu pis mia leksi pu arhizi apo "pa" ke simeni to ptino pu kani "papapa"/ (I want you to tell me a word that starts with "pa" and means a bird that sounds like "papapa") the child has to say /papja/ (duck). Each correct answer is awarded with one point and each wrong with 0 points. Maximum possible score for this part is 14 points.

The second part consists of 15 items and measures the child's ability to describe well known words. For example, the researcher says as an example /ti ine skilos- ine ena zoo, yavyizi, fora luri/ (What is a dog? An animal, barks, has a leash). And then asks the child another well known word and the child has to give an accepted description that matches this word. Each correct answer gives one point and each wrong 0 points. Maximum possible score for this part is 15 points. Maximum score for the whole task is 29 points.

The second oral vocabulary task, suitable for ages 12 to 14 years old, has 13 items and measures the child's ability to find the correct described word with similar methodology as the previous task. Maximum possible score for this part is 13 points. The second part consists of 13 items and measures the child's ability to describe well known words with similar methodology as the previous



task. Maximum possible score for this part is 13 points. Maximum score for the whole task is 26 points.

At the end of all meetings, data were collected and analysed by SPSS version 20.

## Results

Descriptive statistics indicators for the total of the correct answers for all participants in the two measurements are presented below. The highest mean of correct answers is presented in the oral vocabulary task for the age group of 12 to 14 years old and the lowest mean of correct answers in the oral vocabulary task of children aged 4 to 6 years. Low values for Skewness and Kurtosis were observed in all parts, indicating normal distribution (Table 1).

Table 1. Descriptive Statistics of correct answers in both groups.

Test	N	Mean Age	SD age	Min	Max	Skewness	Kurtosis
Oral Vocabulary 1	20	5.5	0.9	1	22	0.7	0.9
Oral Vocabulary 2	20	13.5	0.9	3	23	0.6	0.8

The 10 toddlers with typical development ( $M = 480$ ,  $SD = 34.5$ ) compared to 10 toddlers with Down Syndrome ( $M = 425$ ,  $SD = 31$ ) in their oral vocabulary skills demonstrated significantly better performance,  $t(18) = 2.1$ ,  $p < .05$ . After eight years the 10 children with typical development ( $M = 536$ ,  $SD = 2.5$ ) compared to the same 10 same children with Down Syndrome ( $M = 137$ ,  $SD = 2.9$ ) again demonstrated better oral vocabulary skills,  $t(18) = 4.9$ ,  $p < .05$ . Additionally, two paired-samples t-tests were conducted to compare performance in oral vocabulary skills between two different time periods. The first one when children were 4 to 6 years old and the second when children were 12 to 14 years old. There was a significant difference in the scores for children with typical development in the first measurement ( $M=9.4$ ,  $SD=1.14$ ) and in the second measurement ( $M=17.4$ ,  $SD=0.9$ );  $t(9)=-8.00$ ,  $p < 0.005$ . As for children with Down Syndrome no significant differences were observed between the first measurement ( $M=1$ ,  $SD=1.89$ ) and the second measurement ( $M=4.3$ ,  $SD=1.67$ ),  $t(9)=-3.3$ ,  $p > 0.005$ .

## Discussion

These results suggest that children with typical development performed better in both time periods than children with Down Syndrome. Moreover, our findings show that oral vocabulary skills increase as children with typical

development grow older, whereas oral vocabulary skills of children with Down Syndrome increase but in a slower rate than their typical developing counterparts. Our findings come in line with previous research conducted in this field that show that children with typical development exceed in their performance over children with Down Syndrome (eg Zampini & D'Odorico 2013). Since this is the first time a longitudinal study in greek language is performed, we can assume that deficits in oral vocabulary in children with Down Syndrome may be caused, in addition with the above mention difficulties, by the peculiarity of Greek language in morphology and in complexity of consonants and vowels, although further studies need to be done in this fruitful field. However, our findings must be treated with caution since the sample is small and should not be generalized to the total of children with Down Syndrome.

## References

- Abbeduto, L., Warren, S., Conners F. 2007. Language development in Down Syndrome: From the prelinguistic period to the acquisition of literacy. *Mental Retardation and Developmental Disabilities Research* 13, 247-261.
- Grieco, J., Pulsifer, M., Seligson, K., Skotko, B., Schwartz, A. 2015. Down Syndrome: cognitive and behavioral functioning across the lifespan. *American Journal of Human Genetics* 169, 135-149.
- Tzouriadou, M., Anagnostopoulou, E., Toutountzi, E., Psoinou, M. 2008. *Detroit Test of Learning Aptitude*. Thessaloniki, Aristotle University of Thessaloniki and Ministry of Education.
- Levy, Y., Eilam, A. 2013. Pathways to language: A naturalistic study of children with Williams Syndrome and children with Down Syndrome. *Journal of Child Language* 40, 106-138.
- Tzouriadou, M., Singolitou, E., Anagnostopoulou, E., Vakola, I. 2008. Psychometric criterion of Language Adequacy (L-a-T-o). Thessaloniki, Aristotle University of Thessaloniki and Ministry of Education.
- Zampini, L., D'Odorico, L. 2013. Vocabulary development in children with Down Syndrome: Longitudinal and cross sectional data. *Journal of Intellectual and Developmental Disability* 38, 310-317.

# The role of foreign languages in the mobility journey of young Greeks

Angeliki Kiapekaki

Groupe d'Études Orientales, Slaves et Néo-helléniques, University of Strasbourg, France

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The role of foreign languages at the participation of young Greeks at the European mobility programs is the main subject of this communication. The target audience is former participants of the mobility programs (Erasmus , Lifelong Learning Program, Erasmus +). This research is characterized primarily as a qualitative study, with certain features of a quantitative study. According to our results, foreign languages have versatile functions for the participation in a mobility program: they can be simultaneously a condition and a reason, at the same time as they are an essential effect of mobility. Also, mastery of foreign languages is mandatory in order to be able to participate in mobility programs. Likewise, participants decide to take part at a mobility program and to move to another country in order to practice at least one foreign language or learn a new one. The analysis of the corpus reveals that the choice to learn a foreign language is determined by different factors and based on various motivations. Finally, language policy can also play an important role and have an impact on the linguistic repertoire of individuals. Policy choices can thus create the framework for learning a foreign language and determine the host country of the mobility.

Keywords: mobility program, language acquisition, language policy

## Introduction

The present communication focuses on the role of foreign languages at the participation of young Greeks at the European mobility programs.

European mobility programs were the target subject in many studies. Teichler's work (2015) for example has shown that the motivations for organized mobility are mainly linguistic, academic, professional, socio-cultural and personal. The four major advantages of temporary study abroad are academic, cultural, linguistic and professional advantages. Teichler also found out that students who have benefited from an international mobility experience can have visible and less visible effects. We aim to study the essential role of foreign languages in the participation at the European mobility programs.

## Methods and materials

In order to investigate the importance of foreign language at the European mobility program, we opted for an interdisciplinary research. The topics and the methods relate to social sciences, linguistics, and educational sciences. The

two axes of our reflection are the mastery of foreign languages, on the one hand, and the European mobility programs, on the other.

The target audience is made up of former participants in the following mobility programs: Erasmus (1987- 2013), Lifelong Learning Programs (2007-2013), Erasmus + (2014-2020) and European Voluntary Service (EVS) (1996-2013, 2014-2020). These programs are an expression of the language policy applied at European level and they offered the ideal field to study multiple topics.

The methodology used for this research is a mixed method. As Hesse-Biber (2010) says it is “a research design that uses both quantitative and qualitative data to answer a particular question or set of questions”. In this case, quantitative and qualitative methodology is used in order to collect and analyse data.

As for the data collection, our corpus is constructed with 164 responses to a questionnaire, 54 Success stories (published on the official IKY website) and 12 semi-structured interviews. As for the analysis, we applied descriptive statistics and content analysis.

## Results

The nature of the European mobility programs demands a minimum language proficiency. It was not so strange to see that the young Greek participants have a multilingual language profile (Figure 1) but also the omnipresence of the English language. During the mobility and mostly in an academic setting, English is seen as *Lingua Franca* between speakers of different languages, by choice or by compulsion (JENKINS, 2011).

The foreign languages have versatile functions ; they can be simultaneously a condition and a reason, at the same time as they are an essential effect of mobility. Also, fluency in foreign languages is compulsory in order to be able to participate in the mobility. Likewise, participants move to another country in order to practice at least one foreign language or learn a new one. Language skills play a constitutive role in participation in mobility programs as they are the condition, the reason, the purpose, the motive and the result of the mobility.

As for the language acquisition, policy choices at different levels and in different areas can thus create the framework for learning a foreign language and determine the host country of the mobility. Our data analysis showed that 4 different policies can determine the language acquisition of a foreign language.

Finally, we understand that foreign language acquisition and mobility are correlated ; the Greek participants started with the learning procedure of one or two languages. The desire to mobility lead the to participate in an European mobility program: Mobility allows participants to acquire the ability to integrate multicultural environments, to coexist and work with people from different

cultures, practice foreign language, have an European or international experience and finally improve a foreign language or learn a new one.

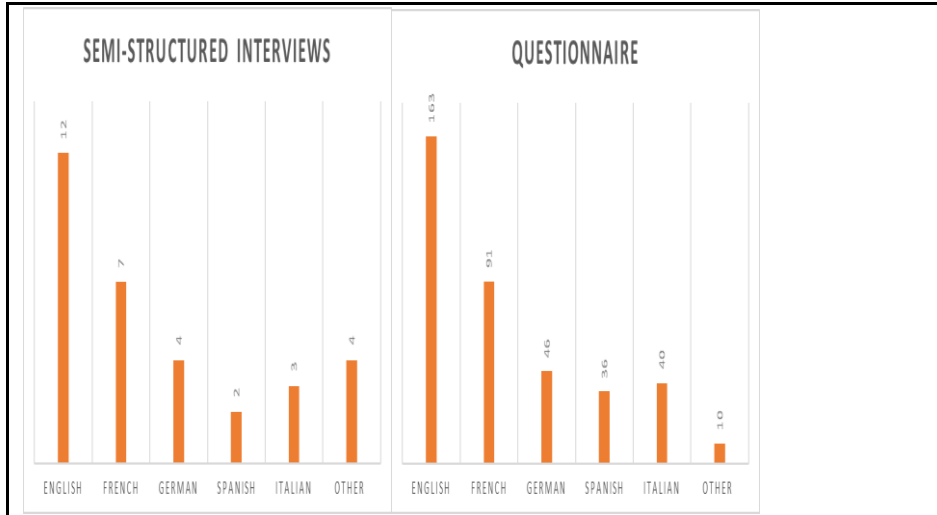


Figure 1. The linguistic skills of the informants.

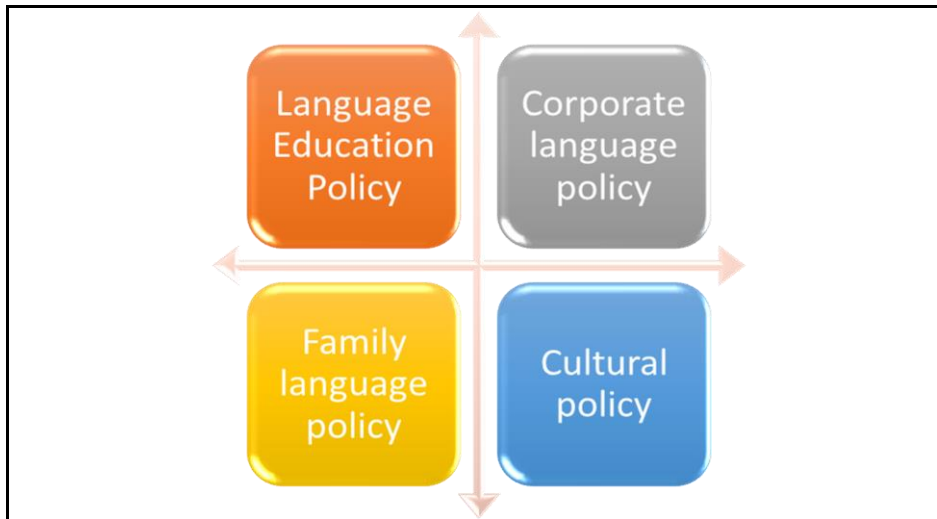


Figure 2. The policies that influence language acquisition.

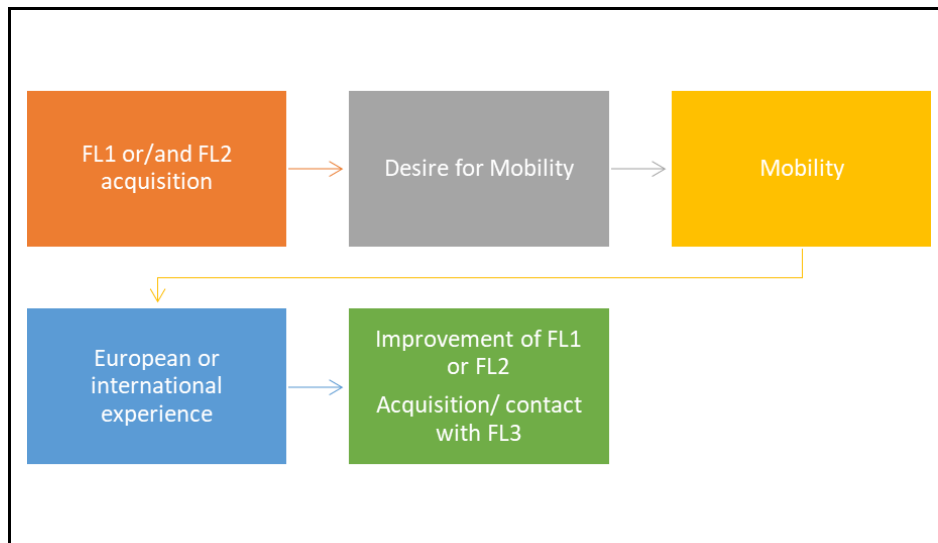


Figure 3. Language acquisition via the European mobility programs.

## Conclusions

To conclude, we understand that mobility and plurilingualism have an “action-reaction” type relationship. Language skills are the main strength of mobility. Whether mobility concerns students or employees, it is based on the language skills of individuals. Language policy determines the linguistic repertoire of participants. And, the choice of mobility is closely linked to the linguistic repertoire of the individual, which determines the host country of the mobility.

## References

- Hesse-Biber, S. 2010. *Mixed methods research: merging theory with practice*. New York : Guilford Press.
- Jenkins, J. 2013. *English as a Lingua Franca in the International University: The Politics of Academic English Language Policy*. London : Routledge.
- Teichler, U. 2015. The impact of temporary study abroad. In *Social interaction, identity and language learning during residence abroad*. Amsterdam : European Second Language Association.

# Structural convergence in spoken English discourse

Christina S. Kim, Gloria Chamorro

Department of Modern Languages and Linguistics, University of Kent, UK

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This study extends the logic of prior studies showing phonetic convergence between interlocutors to the structural domain. We ask whether listeners' adaptation of the syntactic forms they produce depends on how they perceive their interlocutor on measures of interpersonal similarity. Structural priming was used to assess the extent in which interlocutor characteristics influence structural convergence in dialogues between native speakers of different varieties of British English (Lancashire and South-East). Our findings suggest that structural priming is mediated by a speaker's perception of their similarity to their interlocutor, as assessed based on sociolinguistic cues.

Keywords: structural priming, dialogue, native English speakers, sentence production

## Introduction

Multiple lines of research show that interlocutors readily adapt to each other in dialogue. For example, phonetic convergence is observed in interactive conversation and word-shadowing (Goldinger 1997). Such adaptation can be modulated by listeners' perceptions of interlocutor characteristics, like voice, gender, nativeness (Babel & McGuire 2015; Kim et al. 2011). Similarly, lexical alignment reflects speakers' perception of interlocutors' in-/out-community status (Tobar-Henriquez et al. 2021). Structural priming – the tendency of a speaker to produce a syntactic alternant that they have used previously – can also be socially mediated by, for example, social desirability or accent typicality (Fraundorf & Jaeger 2016; Hwang & Chun 2017; Kim & Chamorro 2021). This study extends this research by asking whether any structural priming observed between native British English speakers depends on how one interlocutor perceives the other on interpersonal similarity measures.

## Methodology

We created a computer-based picture-matching game to elicit descriptions of ditransitive events. The task was a dialogue version of the picture-description task used in classical structural priming studies (Bock, 1986). Participants (n=29) saw a series of pictures on a screen. They were told to determine whether pictures matched those of the other “player”, who was a confederate, used only double object (DO) forms, and conversed with the participant over headsets from another room.

A verbal-guise task (Cooper 1975; Zahn & Hopper 1985) assessed participant's impressions about the two confederates before the experiment. Participants first heard a recording of a speaker from South-East England and provided judgments about the speaker's attributes (e.g. attractive, trustworthy) on a 1-7 scale. The process was repeated for a Lancashire speaker. This yielded 15 scores (one per attribute) for each speaker, for each participant, creating a vector of scores for each attribute. We used principle components analysis to reduce original attributes to nine (PC1-9), which explained 95% of the score variance. Participants also marked their hometown and where they thought the confederates were from on a map (Map distance). Participants were assigned randomly to one of the confederates to play the game with.

A logistic mixed-effects regression model predicting DO responses was used to analyse the responses; predictors included: lexical bias (verb's bias toward the prepositional dative (PD) form, from a norming study), trial, verb distribution (whether only alternating verbs were used, or included non-alternating verbs), PCs 1-9, map distance, lexical overlap (whether the verb in the current trial matched the verb on the previous trial), and two-way and three-way interactions. Model comparison was used to remove terms that did not contribute to model fit; the maximal random effects structure supported by the data was used.

## Results

There was a strong main effect of lexical bias ( $\beta=-5.22$ ,  $SE=0.80$ ,  $p<0.0001$ ), with DO forms less likely to be produced the more strongly PD-biased the verb was. The lexical bias effect weakened over trials ( $\beta=0.046$ ,  $SE=0.018$ ,  $p<0.001$ ), suggesting that with sufficient exposure, even strongly PD-biased verbs became more likely to be used in a DO sentence (for  $Trial=\mu_{Trial}-1SD$ :  $\beta=-6.14$ ,  $SE=0.88$ ,  $p<0.0001$ ; for  $Trial=\mu_{Trial}+1SD$ :  $\beta=-4.21$ ,  $SE=0.86$ ,  $p<0.0001$ ). There was also a PD-bias:Verb distribution interaction ( $\beta=-1.39$ ,  $SE=0.56$ ,  $p<0.05$ ), with alternating-only lists showing a greater negative effect of strong PD-bias relative to full distribution lists (for full distribution:  $\beta=-3.84$ ,  $SE=0.62$ ,  $p<0.0001$ ; for alternating only:  $\beta=-6.61$ ,  $SE=1.23$ ,  $p<0.0001$ ). This suggests that prior lexical knowledge can be relied on more heavily for alternating-only lists, where it is never violated; by contrast, for full-distribution lists, half of the confederate's sentences would go against the participant's prior lexical knowledge. This dependence of PD-bias on Verb distribution weakened over the course of the experiment ( $\beta=0.038$ ,  $SE=0.018$ ,  $p<0.05$ ).

The effect of perceived similarity with the speaker was shown most directly by the main effect of Map distance ( $\beta=-0.085$ ,  $SE=0.034$ ,  $p<0.05$ ), with participants producing fewer DO structures as their estimate of the map distance between their hometown and the speaker's increased. In addition, PCs 2 ( $\beta=0.19$ ,  $SE=0.12$ ,  $p<0.05$ ), 5 ( $\beta=0.29$ ,  $SE=0.13$ ,  $p<0.05$ ), 8 ( $\beta=-0.65$ ,



SE=0.16,  $p<0.0001$ ) and 9 ( $\beta=0.71$ , SE=0.25,  $p<0.001$ ), which encoded participants' perceptions of speaker-specific attributes, influenced how likely participants were to produce DO sentences.

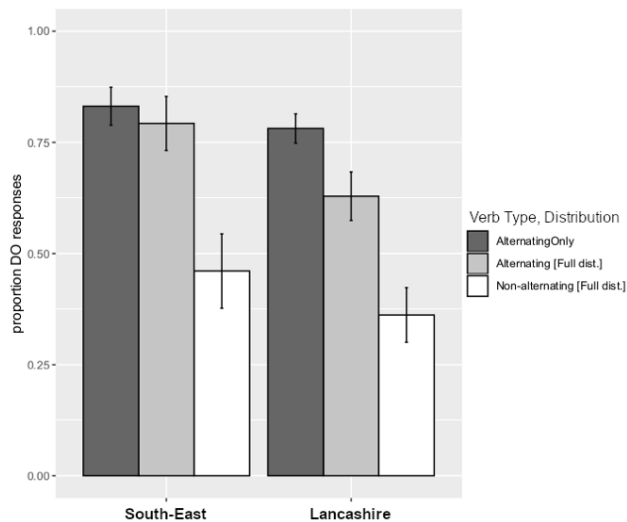


Figure 1. Mean proportions of DO productions by Speaker type.

Finally, there was a strong negative main effect of Lexical overlap ( $\beta=-0.18$ , SE=0.032,  $p<0.0001$ ), with fewer DO forms associated with lexical repetition trials overall. This is in the opposite direction of the general lexical boost pattern reported in prior studies. However, the marginal interaction of Lexical overlap, Map distance, and Trial ( $\beta=0.00088$ , SE=0.00052,  $p<0.1$ ) points at a possible explanation: while no-overlap trials show a Map distance effect that remains relatively stable throughout the trials, overlap trials show a marked decrease in the Map distance effect over trials. In other words, participants started out being reluctant to use the same form for adjacent trials with the same verb (some of which would have required producing an anomalous sentence), and this reluctance was more pronounced when the interlocutor was perceived to be from further away from the participant. This dependence of Lexical overlap on Map distance weakens as trials progress.

## Discussion

This study was aimed at investigating whether speakers' perceptions of their proximity with their interlocutor influenced structural convergence. The main effects of principle components 2, 5, 8 and 9 suggest that judgments made by the participants of individual personality attributes based on the speaker's voice influenced how likely they were to produce DO sentences.

A main effect of Map distance was also revealed. This effect was found across the board, with participants producing fewer DO structures when their perceived distance between their hometown and the speaker's increased. This does not necessarily suggest that speakers "trusted" the competence of their interlocutors more when they perceived them to be similar (small Map distance) to them, and less when they perceived them to be dissimilar (larger Map distance); that is, instead of a trade-off (i.e. relying on own prior knowledge v. interlocutor's competence, which would have appeared as a PD-bias:Map distance interaction), participants just showed less convergence overall with more distant interlocutors. Similarly, Map distance did not interact with Verb distribution. Such an interaction might have been observed, for instance, if participants were seeking an explanation for the atypical distribution of sentence forms in the full-distribution lists (i.e. those containing alternating and non-alternating verbs), and attributed the grammatical anomalies to the speaker's distance (and therefore dissimilarity) from themselves. However, in our data, the Map distance effect was no stronger for full distribution than alternating-only lists.

With regards to the interactions involving Verb distribution, the stronger PD-bias effect for alternating-only compared to full-distribution lists suggests that prior lexical knowledge is used more when it is more reliable in context (as it is for alternating-only lists). In addition, a reliable lexical boost was observed for lists with only alternating verbs, but not for lists featuring the full distribution, suggesting that the lexical boost relies to some extent on the output form being well-formed.

Overall, the findings reveal that speakers' reliance on prior lexical knowledge shifts with a number of contextual factors, including the overall distribution of well-formedness in the discourse context, the perception of similarity/distance with the interlocutor, and perceived non-linguistic attributes of the interlocutor. We suggest that structural convergence can be modulated by interlocutor-specific sociolinguistic and social information.

## References

- Fraundorf, S.H., Jaeger, T.F. 2016. Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language* 91, 28-58.
- Hwang, H., Chun, E. 2018. Influence of social perception and social monitoring on structural priming. *Cognitive Science* 42, 303-313.
- Kim, C.S., Chamorro, G. 2021. Nativeness, social distance and structural convergence in dialogue. *Language, Cognition and Neuroscience*.
- Tobar-Henriquez, A., Rabagliati, H., Branigan, H.P. 2021. Speakers extrapolate community-level knowledge from individual linguistic encounters. *Cognition* 210.

# Autistic traits in duration of utterance-final particles in Japanese

Sachiko Kiyama<sup>1</sup>, Ge Song<sup>1</sup>, Kuniya Nasukawa<sup>2</sup>

<sup>1</sup>Department of Linguistics, Tohoku University, Japan

<sup>2</sup>Department of English, Tohoku Gakuin University, Japan

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This study investigated the effects of individual autistic traits on productive prosody of Japanese utterance-final particles (UFPs) as social markers expressing the speaker's moods, based on previous studies reporting atypical patterns in speech prosody by people with autism spectrum disorders (ASDs). Analysis of the prosodic features of UFPs obtained from a role-play task by typically developed native Japanese speakers revealed that individual speakers' autistic traits significantly affected the duration of the UFPs, but intensity and pitch did not show any significant effects. Speakers with higher autistic traits tended to utter UFPs in shorter times, in comparison with those with lower autistic traits. This study provides evidence that the atypical prosody associated with autistic traits may be reflected in the duration of mood morphemes.

Keywords: Japanese, right periphery, utterance-final particle, prosody, autistic traits

## Introduction

Prosody plays a key role in communicating a speaker's emotional state as well as linguistic meanings. It has long been claimed that individuals with autism spectrum disorder (ASD), which is characterized by difficulties with social communication and restricted and repetitive behavior, display atypical prosodic features in their use of pitch, duration and intensity (Asperger, 1944). However, some recent studies have rejected this long-held claim, suggesting instead that individuals with ASD have no specific difficulties in productive prosody (McCann & Peppé, 2003). This lack of consensus could stem from the heterogeneity that exists within the ASD population, but it could also reflect shortcomings in the way earlier experiments were conducted, where results were often based on various factors (e.g. grammatical, semantic, emotional) which were not specified (Dahlgren et al., 2018).

We explore one possible solution to control these confounding effects by focusing on utterance-final particles (UFPs) in East- and South-East Asian languages, because UFPs are bound morphemes with no substantial meanings, and speakers can freely attach UFPs to an utterance to indicate their attitude or mood (Yap et al., 2014). To elucidate the effects of individual autistic traits on prosody, this study measured the prosodic features of the high-frequency Japanese UFPs *-ne* (other-oriented UFP) and *-yo* (self-oriented UFP), with

reference to the Autism-Spectrum Quotient (AQ) for typically developed (TD) young adult native Japanese speakers. It was hypothesized that individual AQ has a stronger association with the prosody of other-oriented *-ne* than with self-oriented *-yo*, since autistic traits are concerned with the other’s mental states.

## Methods

We obtained data (Table 1) on the prosodic features of Japanese UFPs from 27 native Japanese speakers (average age  $21.1 \pm 2.7$  years, 13 males) by utilizing a role-play task where each participant uttered 80 sentences with either the UFP *-ne* or *-yo* in hypothetical dialogs in a soundproof chamber. Linear mixed effects (LME) models were constructed for the features of duration, intensity, and pitch. Specifically, the index of pitch was examined in terms of mean, standard deviation (SD), and difference between maximum and minimum, to track the intonation patterns. In sum, dependent variables of the modelling were mean duration, intensity, and pitch, and pitch SD, and difference. For each of the five dependent variables, the LME modeling included fixed effects of UFP type (*-ne/-yo*) as a categorical variable and AQ (out of 50) as a continuous variable, and random effects of participant and item.

## Results and discussion

We found a significant correlation between the participants’ AQ (i.e., a higher score indicates a higher autistic tendency) and the duration of UFPs, whereas correlations of AQ with other indices of intensity and pitch were not significant (Figure 1). The LME modeling (Table 2) revealed that TD young adults with a higher autistic tendency uttered the UFPs with a shorter duration, in comparison with those with a lower autistic tendency. This effect was stronger for other-oriented *-ne* than for self-oriented *-yo* (Figure 2).

The result is consistent with a previous electrophysiological study of Japanese UFP comprehension (Kiyama et al., 2018), which reported that young TD adults with higher autistic traits exhibited a particular hypersensitivity to the UFP *-ne* as a marker of social distance with others. By focusing on the prosody of UFPs, this study provides evidence that AQ-associated atypical prosody may be reflected in the duration of mood morphemes.

Table 1. Means (SDs) of prosodic features of Japanese UFPs *-ne* and *-yo*.

UFP Type	Duration (ms)		Intensity (db)		Pitch (Hz)	
	Men	Women	Men	Women	Men	Women
<i>-ne</i>	286.8	331.2	70.6	71.0	135.0	196.2
	(78.4)	(67.9)	(4.3)	(5.8)	(29.7)	(30.7)
<i>-yo</i>	227.9	237.3	66.5	69.6	134.5	189.6
	(63.9)	(41.9)	(5.3)	(5.4)	(27.4)	(28.3)

Notes: UFP = utterance-final particle, Men ( $n = 13$ ), Women ( $n = 14$ ).

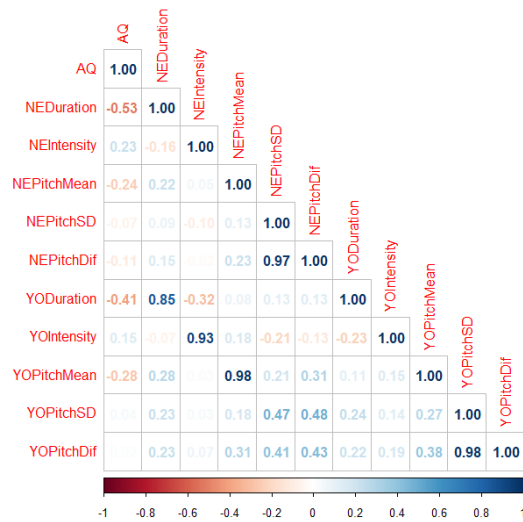


Figure 1. Correlation matrix among prosodic features of Japanese UFPs *-ne* and *-yo*, in addition to individual speaker’s AQ ( $N = 27$ ).  
 Notes: UFP = utterance-final particle, AQ = Autism-Spectrum Quotient.

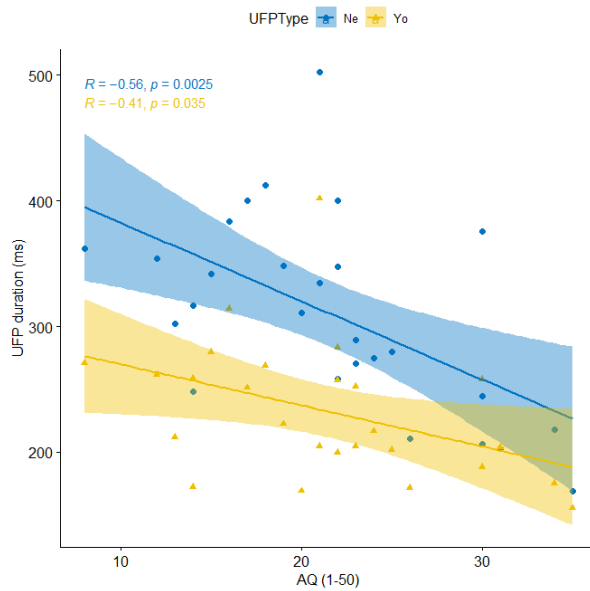


Figure 2. Plot of individual native speakers ( $N = 27$ ) in terms of AQ (x-axis) and duration (y-axis) of Japanese UFPs *-ne* (blue) and *-yo* (yellow).  
 Notes: UFP = utterance-final particle, AQ = Autism-Spectrum Quotient. Shades represent 95% confidence intervals.

Table 2. Fixed effects of UFP type (*-ne* and *-yo*) and AQ on UFP prosody.

Contrast	$\beta$	95% CI [LL, UP]	t	df	p
<b>Duration</b>					
(Intercept)	0.419	[0.167, 0.671]	3.261	28.5	0.003
UFP type	-0.786	[-0.935, -0.638]	-10.360	34.0	< 0.001
AQ	-0.432	[-0.671, -0.193]	-3.547	23.9	0.002
UFP type x AQ	0.173	[0.036, 0.309]	2.480	26.7	0.020
<b>Intensity</b>					
(Intercept)	0.187	[-0.118, 0.493]	1.201	64.7	0.234
UFP type	-0.318	[-0.548, -0.088]	-2.709	80.3	0.008
AQ	0.131	[-0.103, 0.364]	1.096	6.4	0.312
UFP type x AQ	-0.024	[-0.121, 0.074]	-0.473	26.4	0.640
<b>Pitch Mean</b>					
(Intercept)	-0.040	[-0.373, 0.293]	-0.236	15.8	0.816
UFP type	-0.021	[-0.138, 0.095]	-0.361	62.2	0.719
AQ	0.048	[-0.200, 0.296]	0.382	7.6	0.713
UFP type x AQ	0.013	[-0.067, 0.093]	0.323	27.2	0.749
<b>Pitch SD</b>					
(Intercept)	-0.274	[-0.355, -0.194]	-6.686	62.4	< 0.001
UFP type	0.464	[0.277, 0.652]	4.847	29.4	< 0.001
AQ	-0.071	[-0.207, 0.065]	-1.020	10.7	0.330
UFP type x AQ	0.067	[-0.114, 0.248]	0.728	26.7	0.473
<b>Pitch Difference</b>					
(Intercept)	-0.237	[-0.320, -0.153]	-5.541	61.2	< 0.001
UFP type	0.402	[0.190, 0.614]	3.715	26.8	0.001
AQ	-0.096	[-0.234, 0.042]	-1.368	9.4	0.203
UFP type x AQ	0.061	[-0.138, 0.260]	0.599	25.3	0.555

*Notes:* UFP = utterance-final particle, AQ = Autism-Spectrum Quotient,  $\beta$  = standardized partial regression coefficients, CI = confidence interval, LL = lower limit, UL = upper limit,  $N = 27$ .

## Acknowledgements

This study was supported by JSPS KAKENHI Grant Number 19H00532.

## References

- Asperger, H. 1944. Die Autistische Psychopathen im Kindesalter. *Archiv für Psychiatrie und Nervenkrankheiten*, 117, 76–136.
- McCann, J. Peppé, S. 2003. Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders*, 38, 325-350.
- Kiyama, S., Verdonschot, R., Xiong, K., Tamaoka, K. 2018. Individual mentalizing ability boosts flexibility toward a linguistic marker of social distance: An ERP investigation. *Journal of Neurolinguistics*, 47, 1-15.

# Exploring the role of verb frames in assessing semantic difficulty

Meichun Liu, Zhuo Zhang

Department of Linguistics and Translation, City University of Hong Kong, China

<https://doi.org/10.36505/ExLing-2021/>

## Abstract

This paper explores the role of verb frames in semantic difficulty. We built a textbook corpus based on textbooks in primary schools, and the grades of the textbooks were treated as semantic difficulty levels. The difficulty levels were further mapped onto the performance of verb frames. The study found that the verb-frame diversity is highly correlated with semantic difficulty level with statistical significance and that the distribution of verb frames may correspond to the skill theory (Fischer & Corrigan, 1981) in cognitive language development. The skill theory could be used to group verb frames by tiers and roughly predict their distributional trends. Thus, verb frames could be helpful as a semantic reference for textbook planning and readability assessment.

Keywords: verb frames; semantic difficulty; textbooks; readability; skill theory

## Introduction

In readability assessment, lexical and syntactic cues are widely used, but semantic cues are quite limited. The newly emergent frame-based features look promising, though they have not yet been incorporated into a working model. Lee, Liu & Cai (2020) first proposed that frame-based features can potentially be utilized in Chinese readability assessment. However, the study just focused on ten frequent verbs across three frames and used less than 1,600 manually annotated sentences to find the potential indicators.

This paper aims to find out how verb frames contribute to semantic difficulty through a quantitative corpus-based study. Two hypotheses are proposed.

H1: There is a strong correlation between verb-frame diversity and difficulty level with statistical significance.

H2: The distributions of verb frames mostly vary in line with the three tiers—sensory-motor, representational, and abstract—proposed in the skill theory of cognitive development (Fischer, 1980; Fischer & Corrigan, 1981).

## Literature review

Speakers understand the meaning of a word by knowing its background frame (Fillmore and Atkins, 1992:76-77). The Archi-frames indicate broad semantic domains distinguished by self-contained conceptual schema (Fillmore, 1982).

Language learning is based on actual usage, and language complexity comes from the interaction of cognition and use (Bybee, 2010). As some verbs are learned first in the L1 acquisition of verb-centered constructions (Jing-Schmidt, 2019: 13-31), complex texts may contain more diverse verb frames (H1). This skill theory regards cognitive development as "the construction of hierarchically ordered collections of specific skills," which can theoretically be divided into three tiers: sensory-motor tier, representational tier, and abstract tiers with gradually increased complexity (Fischer, 1980; Fischer & Corrigan, 1981). Therefore, frames from the lower tier should decrease while higher-tier frames should increase as the text becomes more advanced (H2).

## Methodology

We propose the verb-centered frame-based model of semantic complexity to map verbs with semantic difficulty and use this model to study how verb frames may influence semantic difficulty. This part mainly introduces the materials and two experimental designs corresponding to the research questions.

### Materials

A corpus with about a million characters was built based on nine sets of Chinese textbooks at the primary school level, and the grade levels of the textbooks were used as semantic difficulty levels. Then, verb frames were annotated automatically based on Mandarin VerbNet (<http://mega.lt.cityu.edu.hk/~yufechen/#/>), a Chinese-approximate semantic database with annotated verb frames.

### Experimental designs

This section introduces the designs of two experiments.

Experiment#1 We randomly selected 1,000 sentences from textbooks at each grade level, divided them into 50 groups of 20 sentences each, and calculated the average number of the different verb frames (verb-frame diversity) in each group to minimize the effect of sentence length. We then repeated the random selection ten times and calculated the Spearman correlation between the diversity of verb frames and the grade level of the text.

Experiment#2 The proportion of each frame at each difficulty level was calculated, and frames were categorized into the three tiers guided by the skill theory. Linear regression models were employed to explore the frame distribution trends towards difficulty level regarding skill-based frame tiers.

## Results and discussion

The results show a strong correlation between the diversity of verb frames and the difficulty level,  $r = 0.89$  ( $> 0.8$ ),  $p < 0.001$  ( $< 0.05$ ), with verb frames becoming more diverse as difficulty increases. Furthermore, the distribution of



the frames largely corresponds with the skill theory of cognitive language development (Fischer & Corrigan, 1981), and the distributions of most dominant frames by difficulty fit the regression models with high coefficients of determination. The tier-specific trends of Archi-frame distributions are displayed in Table 1. As the semantic difficulty level increases, the proportions of Tier 1 frames (COMMUNICATION, SELF MOTION & PERCEPTION) decrease, whereas Tier 3 Frame COGNITION increases.

Table 1. Analysis of dominant Archi-frames distribution by three tiers.

Archi-frame	Typical verbs	Tier	Distribution trend
COMMUNICATION	说 ( <i>shuō</i> , 'say') 谈 ( <i>tán</i> , 'talk')	1: Sensory-motor	Downward
SELF MOTION	来 ( <i>lái</i> , 'come') 走 ( <i>zǒu</i> , 'walk')	1: Sensory-motor	Downward
PERCEPTION	听 ( <i>tīng</i> , 'hear') 触 ( <i>chù</i> , 'touch')	1: Sensory-motor	Downward
EXISTENCE	有 ( <i>yǒu</i> , 'have') 具有 ( <i>jùyǒu</i> , 'possess')	2: Representational	No obvious trend (Slightly upward)
TRANSFERENCE	给 ( <i>gěi</i> , 'give') 学习 ( <i>xuéxí</i> , 'learn')	2: Representational	No obvious trend (Slightly upward)
COGNITION	相信 ( <i>xiāngxìn</i> , 'believe') 决定 ( <i>juéding</i> , 'decide')	3: Abstract	Upward

The findings agree with the empirical knowledge of language usage that the more abstract meaning is considered to be more difficult. We further explored other less dominant frames and found similar trends in tiers. Also, even in the same tier, the frames displaying upward trends are usually considered more difficult and easier to be used in a metaphorical way. For example, CAUSED-POSITION and CAUSED-MOTION all belong to Tier 2, but the former displays a downward trend, so the verbs within the frames are generally less difficult than those within the latter frame. For example, CAUSED-POSITION verbs such as 摆放 (*bǎifàng*, 'place'), 搁置 (*gēzhì*, 'put aside'), 镶嵌 (*xiāngqiàn*, 'embed') care more about the action of putting something somewhere whereas CAUSED-MOTION verbs such as 背负 (*bèifù*, 'bear'), 搬动 (*bāndòng*, 'move') and 提携 (*tíxié*, 'support') are grounded in the physical action of carrying something somewhere, which can be easily used in a metaphorical way like the examples below.

1. 背负 压力 / *bèifù yālì* / bear pressure / to bear pressure
2. 用 科学的 力量 搬动 这座 大山  
use scientific power move this mountain  
*yòng kēxuéde lìliàng bāndòng zhèzuò dàshān*  
to move this mountain with the power of science
3. 互相 提携 / *hùxiāng tíxié* / mutual support / to support each other

## Conclusion

As a pioneering quantitative investigation on Chinese verb-frame behaviors by difficulty, the study shows that the diversity of verb frames is highly correlated with semantic complexity with statistical significance. The skill theory could be applied to categorize verb frames by tiers, compare the general difficulty of verb frames, and even roughly predict the distributional trends of several dominant frames. Thus, verb-frame diversity and distribution can potentially serve as a semantic reference for textbook planning and readability assessment.

## Acknowledgments

We gratefully acknowledge the grant support (Project#9360163) of the Hong Kong Institute of Data Science, City University of Hong Kong.

## References

- Lee, J.S., Liu, M., Cai, T. 2020. Using verb frames for text difficulty assessment. Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet, 56–62.
- Fillmore, C.J., Atkins, B. T. S. 1992. Toward a frame-based lexicon: The semantics of RISK and its neighbors. In Lehrer, A., & Kittay, E. (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 75–102). Hillsdale: Erlbaum.
- Fillmore, C.J. 1982. Frame semantics. In the Linguistic Society of Korea (Ed.) *Linguistics in the morning calm* (pp. 111–137). Seoul: Hanshin Publishing Co.
- Bybee, J. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87, 477–531.
- Fischer, K.W., Corrigan, R. 1981. A skill approach to language development. In R. E. Stark (Ed.) *Language behavior in infancy and early childhood* (pp. 245–273). Elsevier: North Holland.
- Jing-Schmidt Z. 2019. Corpus and Computational Methods for Usage-Based Chinese Language Learning: Toward a Professional Multilingualism. In: Lu X., Chen B. (eds) *Computational and Corpus Approaches to Chinese Language Learning*. Chinese Language Learning Sciences. Springer, Singapore. [https://doi-org.ezproxy.cityu.edu.hk/10.1007/978-981-13-3570-9\\_2](https://doi-org.ezproxy.cityu.edu.hk/10.1007/978-981-13-3570-9_2)

# The prosodic structure in French: Analysis with deep learning networks

Philippe Martin

LLF, UFRL, Université de Paris, France

## Abstract

Experiments using supervised deep learning algorithms were executed on a large set of sentences selected from the SIWIS (Yamagishi et al., 2016) corpus in order to investigate properties of the prosodic structure in French. Based on a phonological model assuming dependency relations between accent phrases (Martin, 2018), the training corpus was annotated with a modified ToBI notation system, encoding prosodic events located on stressed syllables (vowels), i.e. pitch accents, and syntactic groups final syllables (vowels), i.e. boundary tones, which in French occur on the same accent phrases final syllable.

Keywords: Deep learning, French, ToBI, prosodic structure.

## Introduction

Various deep learning models are well suited to classify, process and generate time series such as speech prosodic events, and may appear at first very attractive to obtain pertinent phonetic and phonological information in this field. However, the large number of parameters (typically over 1,000,000) obtained from deep learning training make extraction of meaningful phonetic and phonological information almost impossible. Therefore, instead of using raw acoustic data for training (e.g., WaveNet), feature engineering derived from specific knowledge in the domain could be used not only to speed up the learning iteration process, but also to evaluate the pertinence of the descriptive features extracted from raw acoustic data from the performance of the selected deep learning model.

## Data

In order to obtain some prosodically meaningful results, sentences from a training corpus were annotated with as melodic contours as well as with a modified ToBI notation. These annotations pertain to prosodic events located on stressed syllables vowels, i.e. pitch accents, and syntactic groups final syllables vowels, i.e. boundary tones. Pitch accents and boundary tones in French occur on the same accent phrases (i.e. stress group) final vowel. Merged prosodic events are classified as rising or falling, above or below the glissando threshold, i.e. perceived as a melodic change or as a static tone. Sentence final and falling prosodic events before pauses constitute special categories.

The glissando value gives an indication pertaining to the actual perception of melodic changes. It is evaluated by the formula  $(st2-st1) / (t2-t1)$ , with  $st1$  and  $st2$  being the stressed vowel starting and ending F0 frequency value in semitones at times  $t1$  and  $t2$ . The glissando threshold, which approximately determines the limit for the perception of a change in pitch, lies in the  $0,16 / (t2-t1)^2$  and  $0,32 / (t2-t1)^2$  range (Rossi, 1971).

Rising and falling melodic changes above the glissando threshold are labelled respectively as  $H^*H^-$  and  $L^*L^-$  in the proposed modified F-TOBI system, and *Cris* and *Cfal* using contour notations. Rising or falling melodic changes below the glissando threshold are labelled  $H^*/L^*$  or *Cneu*. Sentence conclusive terminal prosodic events are annotated as  $L^*L\%$  or *Cdec* (declarative case) and  $H^*H\%$  or *Cint* (interrogative case). The falling contour before pause is labelled  $L^*\#$  or *Cfap*, Sentence post final declarative and interrogative prosodic events in a rheme-theme configuration use the  $H^*/L^*$  or  $C0n$  for the declarative case, and  $H^*/H\%$  for the interrogative configuration. These categories have been shown to adequately indicate the dependency relations between stress groups, which in turn define the sentence prosodic structure.

#### Prosodic annotation: F-ToBI and melodic contours









$H^*/L^*$ / <i>Cneu</i> neutralized, under the glissando threshold	
$L^*L^-$ / <i>Cfal</i> falling, above the glissando threshold	
$H^*H^-$ / <i>Cris</i> rising, above the glissando threshold	
$L^*\#$ / <i>Cfap</i> falling, before a pause > 250 ms	
$L^*L\%$ / <i>Cdec</i> final conclusive declarative	
$H^*/L^*$ / $C0n$ post final declarative	
$H^*H\%$ / <i>Cint</i> final conclusive interrogative	
$H^*H\%$ / <i>Cin</i> post final interrogative	

Figure 1. Table of prosodic events annotation symbols in both the modified F-ToBI and contour systems

### Training and testing

The French SIWIS corpus (Yamagishi et al., 2016) contains 4477 relatively short sentences with various syntactic structures, read by 31 different speakers. Automatically selected stressed syllables candidates were validated perceptually.

The processing steps implemented in C++ and Python (using Keras and TensorFlow libraries) are as follows:

- Resampling of the 4477 recordings from 44100 Hz to 22050 Hz to better accommodate the automatic segmentation algorithm.
- Automatic segmentation into words and phones with API annotation, using a forced alignment algorithm comparing corpus sentences with TTS generated speech.
- Generation of the fundamental frequency curve for each sentence (spectral based algorithm).
- Data encoding with fundamental frequency values at the beginning and end of annotated stressed vowels, as well as narrow band spectrograms limited to 1300 Hz. These data implicitly include vowel and word duration, syllabic rate, intensity variation, etc.

## Experimental results

For indirectly evaluate the efficiency of the selected descriptive features, two deep learning training and testing experiments were conducted, using two types of input data related to stressed vowels:

1. Fundamental frequency curve F0, using the spectral comb method
2. Partial narrow band Fourier spectrum, limited to 1300 Hz.

3300 samples constitute the training set, encoded as images of 111 by 25 pixels. The training model has two hidden layers, with respectively 200 and 150 nodes. The Rectified Linear (i.e.  $x = \max(0, x)$ ) activation function was used for both. The output had 8 classes of prosodic events, as defined above (the two post-finals were excluded). The training algorithm used the Adam optimization algorithm, and operates on batches of 32 samples and an embedding of 100.

The training results pertaining to the two types of input data are:

For the fundamental frequency values obtained by the spectral comb method:

```
1/46 [.....] - ETA: 0s - loss: 0.4056 - accuracy: 0.8504
46/46 [=====] - 0s 650us/step - loss: 0.4760 - accuracy: 0.8070
<tensorflow.python.keras.callbacks.History object at 0x000001DE67350AC0>
```

For the narrow band Fourier spectrum, limited to the 0-1300 Hz range:

```
1/26 [>.....] - ETA: 0s - loss: 0.0107 - accuracy: 1.0000
14/26 [=====>.....] - ETA: 0s - loss: 0.0244 - accuracy: 0.9039
26/26 [=====>.....] - 0s 4ms/step - loss: 0.0287 - accuracy: 0.9927
<tensorflow.python.keras.callbacks.History object at 0x000001DE5DB823A0>
```

The final accuracy, respectively 0.8070 and 0.9927, reflects the efficiency of feature selection to encode the data. A low accuracy indicates that some prosodic events belonging to different classes could not be separated, i.e. by using F0 alone, so that narrow band spectrograms appear more appropriate than direct fundamental frequency values for training. This could be explained by occasional errors made in F0 detection, whereas there is no implied calculation of F0 by using a narrow band spectrogram.

Two hundred prosodic events not part of the training set was processed for testing. It reveals that most errors, i.e. wrong classification of prosodic events, were linked to the glissando threshold used in annotation of the training set. For instance, events categorized as H\*H- (or Cris) could be occasionally identified as H\*/L\* (or Cneu) and conversely. Indeed, the glissando threshold is a parametric approximation leading to erroneous classification when its value is close to the threshold between perceived melodic change and static tone perception of melodic events.

## Conclusions

Machine learning is a powerful tool to obtain an efficient discrimination of classes of a very large number of objects. However, this classification power does not deliver any comprehension pertaining to the differences between objects belonging to different classes, whereas our intuitive knowledge allows us to establish these classes and to classify any new object in one of these classes.

We may therefore consider using deep learning processes indirectly to get some insight pertaining to linguistic knowledge, as “ machine knowledge” remains hard to interpret.

## References

- Martin, Ph. 2018. Intonation, structure prosodique et ondes cérébrales, London: ISTE, 322 p.
- Rossi, M. 1971. Le seuil de glissando ou seuil de perception des variations tonales pour la parole. *Phonetica*. n° 23, 1-33.
- Yamagishi, J. et al. 2016. The SIWIS French Speech Synthesis Database, University of Edinburgh. School of Informatics. The Centre for Speech Technology Research, <https://doi.org/10.7488/ds/1705>.

# Wug-testing phonetic prominence in Munster Irish

Connor McCabe

Phonetics and Speech Laboratory, Trinity College, Dublin

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

8 speakers of Munster Irish were presented with a series of disyllabic nonwords and directed to read them aloud in a carrier phrase. Each nonword corresponded to a different pairing of syllable weights (e.g. light-heavy, heavy-light), said to determine lexical stress placement in the variety. A binomial logistic regression examined phonetic measures of prominence as predictors of syllable location, and mixed-effect multiple linear regressions evaluated weight-pairings as predictors of cross-syllable change in the same measures. Results suggest a great deal of inter- and intraspeaker variation, and no clear role of weight in determining assignment of prominence. This is relevant for work on the complex stress system typically attributed to Munster Irish, and for critical examinations of stress description beyond Irish.

Keywords: Phonetics, prominence, weight, Bayesian, Irish

## Introduction

Munster Irish (MI) is said to exhibit a complex system of weight-sensitive lexical stress, diverging from the initial stress of other Irish varieties (Ó Sé 1989; Doherty 1991). A ternary weight hierarchy – VV/V: > [ax] > V – is used in a stress domain of the first three syllables of a word. Initial stress obtains when this domain contains only light syllables, and otherwise the rightmost heavy syllable in the domain is stressed. Stress is non-contrastive. There are numerous lexical exceptions, and the system's origins and precise details are disputed.

Formal accounts of MI stress typically use impressionistic, non-L1 descriptions by early-20<sup>th</sup>-century dialectologists as input (e.g., O'Rahilly 1932). Recent, ongoing phonetic investigation of conservative recordings from 1928 has questioned the accuracy of these complex descriptions (McCabe 2021).

A 'wug' (nonword) task was devised to investigate whether present-day MI speakers productively use syllable weight to assign stress to unknown words.

## Methods and materials

The syllables /b<sup>v</sup>a/, /b<sup>v</sup>ax/, and /b<sup>v</sup>ɑ:/ were combined to create 36 di- and trisyllabic nonwords. The present discussion is restricted to disyllables.

Targets were presented in the carrier phrase *Cad a dúirt an \_\_? Dúirt an \_\_ Tá.* ('What did the \_\_ say? The \_\_ said 'Yes.'). with the second instance used for measurement. The first instance allowed for brief familiarisation with the

target. Participants were told to treat the nonwords as foreign governmental titles, akin to the untranslated use of Irish titles like *Taoiseach* (‘chieftain’/‘prime minister’) in English-language media, to facilitate relatively natural reading. Participants were directed to emphasise the answer given (i.e. *Tá* ‘Is’/‘Yes’), in order to draw phrasal focus, which has the potential to mask or distort the implementation of lexical stress (de Lacy 2014), away from the nonword.

8 L1-MI participants were recruited online. Elicitation sessions took place over Zoom due to public health restrictions, with participants being coached to record themselves locally using Audacity. Materials were presented in PowerPoint via Zoom, with stimuli in pseudorandomised orders over 5 trials, for a total of 360 tokens (8 participants x 9 disyllabic targets x 5 trials).

Segmentally flawed tokens (e.g. /b<sup>v</sup>axa/ or /b<sup>v</sup>ab<sup>v</sup>ab<sup>v</sup>a/ for target /b<sup>v</sup>ab<sup>v</sup>a/) were discarded. Participant error rates ranged from 0-32.5%, leading to a total of 322 usable tokens.

Table 1. Disyllabic nonwords representing all 9 permutations of the 3 traditionally described weight-categories.

Weight	Light (V)	/ax/	Heavy (V:)
Light (V)	baba /b <sup>v</sup> ab <sup>v</sup> a/	babach /b <sup>v</sup> ab <sup>v</sup> ax/	babá /b <sup>v</sup> ab <sup>v</sup> ɑ:/
/ax/	bacha /b <sup>v</sup> axa/	bachach /b <sup>v</sup> axax/	bacha /b <sup>v</sup> axa/
Heavy (V:)	bába /b <sup>v</sup> ɑ:b <sup>v</sup> a/	bábach /b <sup>v</sup> ɑ:b <sup>v</sup> ax/	bábá /b <sup>v</sup> ɑ:b <sup>v</sup> ɑ:/

## Analysis and results

Praat was used to label individual words and syllables, and to automatically extract the following syllable-level measures: duration, mean intensity, minimum, maximum, and mean F0, mean F1, and mean F2. Minimum and maximum F0 were used to calculate the range of F0 on a given syllable. Measurements were z-scored at the level of the individual speaker to facilitate interspeaker comparison. Statistical analysis explicitly avoided reference to prescribed location of phonological ‘stress’, instead focussing on the relationship between phonetic prominence(s) and syllable position.

First, a binomial logistic regression was fitted using Bayesian methods. Speaker-normalised mean intensity, F0 range, mean F1, mean F2, and the interaction between mean intensity and F0 range were used as predictors of the log-likelihood that a given syllable was the final syllable of a disyllabic target. In other words, this asked whether an exaggeration in one or more phonetic exponents of prominence favoured a particular syllable position, and indeed whether different exponents (e.g. pitch and intensity) behaved uniformly. F0 range was selected over a measure of F0 height to focus exclusively on pitch *excursion*, as both rising and falling pitch accents are attested in Irish; high F0 does not necessarily indicate stress. F1 and F2 were included as indicators of vowel reduction in line with previous work on lexical prominence in the variety (Blum 2018). Duration was not used due to its robust contrastive status in the



language. Random slopes by speaker, target, and repetition were included for all predictors. These incorporate the potential for wide variability between speaker strategies, between targets of different weight structure, and for change as the nonwords became more familiar over the course of the task. Weakly informative, normally-distributed priors were used for a model comprising 4 chains of 5000 iterations with a warmup of 1000 iterations.

Speaker normalised mean intensity and mean F2 emerged as the best predictors of a final-syllable classification, both with negative slopes. Increased intensity and decreased vowel backness appear to disfavour second syllables, albeit with very wide 95% credible intervals. All other credible intervals substantially overlap with 0, meaning that a null effect should not be ruled out.

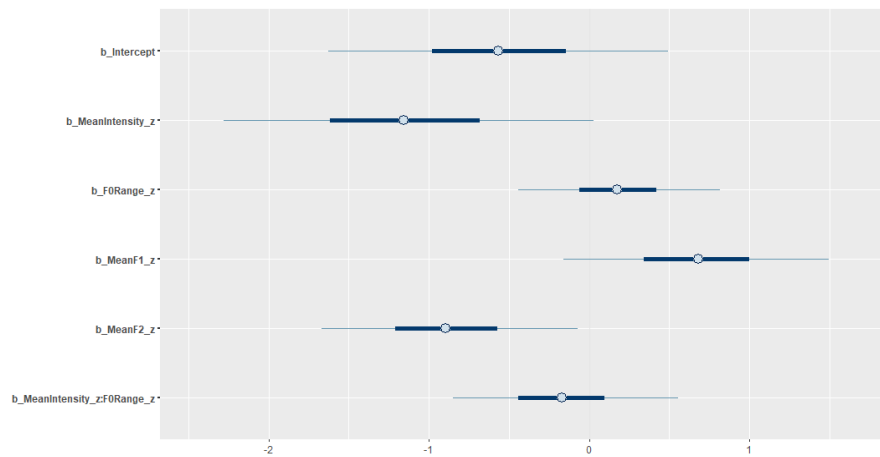


Figure 1. Joint posterior distribution of population-level effects for the binomial logistic regression, showing 95% credible intervals for change in log-likelihood of a final-syllable classification per increase of one standard deviation in the parameter(s) in question.

Additionally, 4 Bayesian mixed-effect multiple linear regressions were fitted using target identity as a predictor of cross-syllable change in speaker-normalised mean intensity, F0 range, mean F1, and mean F2, again with weakly informative, normally distributed priors and random slopes fitted by speaker and repetition. This looked for evidence of consistent directionality of change in phonetic prominence(s) in different weight pairings. For example, light-heavy *babá* might be expected to predict positive cross-syllable change in intensity and/or pitch range if speakers consistently stress the heavier final syllable, and vice versa for heavy-light *bába*.

None of said mixed-effect linear regressions returned evidence of predictive value for any of the parameters used. 95% credible intervals for the effect of all target identities on slope of cross-syllable change all heavily overlap with 0.

Individual distributions are not included due to space constraints. These results fail to provide any obvious support for a productive role of syllable weight in the implementation of straightforward phonetic correlates of lexical stress.

## Discussion and conclusion

Native speakers of MI were asked to produce disyllabic nonwords corresponding to pairings of the three different weight categories traditionally said to determine the location of lexical stress in the variety. Results of a binomial logistic regression broadly suggest that, even with different weight-pairings taken into account, increased mean intensity and decreased vowel backness both weakly favour initial syllables, with no measured parameter appearing to favour final syllables. This may suggest a weak preference for initial stress in unfamiliar words, regardless of weight structure. Results of a series of mixed-effect multiple linear regressions do not indicate predictive utility for different weight-pairings with regard to cross-syllable change in measures of intensity, pitch range, and vowel quality.

These results are preliminary, and subject to ongoing refinement as part of the author's PhD. The sample size presented is small, although Bayesian methods allow for greater inferential flexibility in this regard than frequentist statistics. However, the near total lack of evidence for speakers' use of weight-based criteria for assigning prominence to unfamiliar nonwords is of interest.

It is impossible to collect new, controlled data from the era in which impressions of MI's purportedly weight-based stress system were initially recorded. The present data are able to divorce questions of stress and prominence from complex etymological concerns, and focus directly on speakers' default strategies for assigning prominence to unfamiliar words. A weight-based account of stress in historical or present-day MI cannot yet be ruled out. Nevertheless, the above results suggest, at the very least, that the existence and productivity of such a system should not be taken for granted.

## References

- Blum, E. 2018. Allophony-driven stress in Munster Irish. Manuscript, Rutgers University.
- Bürkner, P. 2018. Advanced Bayesian Multivel Modeling with the R packing brms. *The R Journal* 10(1), 395-411.
- Doherty, C. 1991. Munster Irish stress. In Proceedings of the 10<sup>th</sup> West Coast Conference on Formal Linguistics, 115-126.
- de Lacy, P. 2014. Evaluating evidence for stress systems. In H. van der Hulst (ed.) *Word Stress: Theoretical and Typological Issues*. Cambridge: CUP, 149-193.
- McCabe, C. 2021. Revisiting assumptions stress and 'stress-shift' in Munster Irish. Poster. *Phonetics & Phonology in Europe IV*, 21-23 June 2021.
- O'Rahilly, T.F. 1932. *Irish Dialects Past & Present*. Dublin: Brown & Nolan.
- Ó Sé, D. 1989. Contributions to the study of word stress in Irish. *Ériu* 40, 147-178.

# The production of Greek voiceless fricatives by young children

Elina Nirgianaki<sup>1,2</sup>, Vasiliki Kampaki<sup>2</sup>

<sup>1</sup>Department of Speech & Language Therapy, University of Peloponnese, Greece

<sup>2</sup>Department of Speech & Language Therapy, Metropolitan College, Greece

## Abstract

The present study aims to identify age-related developmental patterns for the Greek language, by analyzing children's and adult's fricative productions. Frication duration, as well as first and second spectral moments (mean, variance) were measured in the first stressed syllable of real disyllable words produced by young children and adult female speakers. A significant effect of age was revealed for fricative duration, with children presenting significantly higher values than adults for all fricatives. The effect of age was also significant for the first and second spectral moments, but not for all fricatives, while the effects of place of articulation and post-fricative vowel on the examined parameters suggest similar intra-category differences for the different age groups.

Keywords: children, Greek, voiceless fricatives, duration, spectral mean, variance

## Introduction

Several studies that have employed acoustic analysis in order to investigate the acquisition and development of speech in children indicate age-related differences in the development of children's obstruent productions (e.g. Nittrouer 1995; Nissen & Fox 2005; Li 2012). Differences in the acoustic and spectral characteristics of speech sounds produced by children and adults may indicate that children's articulatory gestures are not as precisely specified as those of adult speakers (e.g., Nittrouer 1995; Nissen & Fox 2005; Li 2012). Of particular interest has been the development of the voiceless sibilant contrasts, which are also important since these sounds are often among the ones presenting articulatory disorders (Bernthal et al., 2018). Moreover, exposure to a specific language seems to play an important role on the way children develop speech, suggesting different developmental patterns even for speech sounds that are common across different languages (Li 2012).

With regards to the production of Greek fricatives, Nirgianaki (2014) has investigated fricatives produced by Greek speaking adults, finding significant differences among the different places of articulation, such as temporal, spectral and amplitudinal. However, the acoustic studies on the description of Greek sounds produced by children are rare and, focus on a limited subset of fricatives either singleton or in clusters (e.g., Syrika et al. 2011).

In this line of research, the present study aims to investigate temporal and spectral characteristics of Greek voiceless fricatives, and particularly, duration,

first and second spectral moments (M1-mean, M2-variance), as produced by young children compared to those produced by adults.

### Method

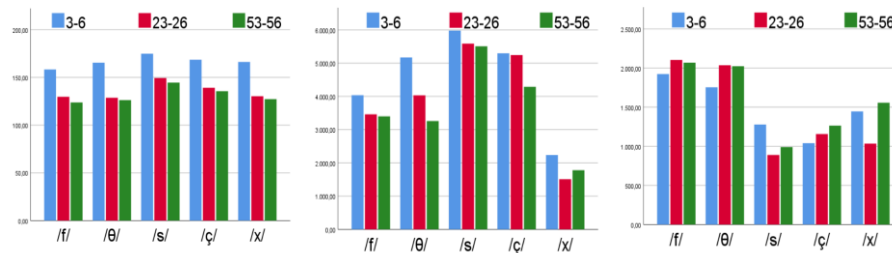
The Greek voiceless fricatives /f θ s ç x/ were recorded in real two-syllable words of the form 'CVCV within the carrier phrase /i i'kona 'ðixni 'ena/'mia \_\_\_ / (“The image shows a \_\_\_”). Each fricative was in initial position, followed by all five Greek vowels (/a e i o u/). Words beginning with /x/ were only followed by the vowels /a o u/, since the allophone /ç/ appears before the front vowels /e i/. Each participant was prompted an image of the target word and the written phrase “The image shows a ...” (next to the image), and then listened to the phrase. Nine female speakers, 3 in each of the age groups 3-6, 23-26 and 53-56 years old, produced the experiment’s material three times at a normal tempo in a quiet room. A consent form was signed by the parents. Duration, as well as first and second spectral moments (M1-mean, M2-variance) -in the middle of each fricative- were measured in Praat software (Boersma & Weenink 2019). SPSS (v26) was used for data statistical analysis.

### Results

A 3-way Anova (age group X place X vowel) with fricative duration as the dependent variable revealed a main effect of the age group ( $F(2)=143.302$ ,  $p<0.001$ ). Post hoc tests indicated significantly higher values for the children’s group than both adult groups. Subsequent tests for each fricative separately, revealed significant differences between children and adults for all fricatives. A main effect was also revealed for place of articulation ( $F(4)=14.460$ ,  $p<0.001$ ), with the alveolar fricative /s/ being significantly the longest, followed by /ç/, /x/, /θ/ and /f/. The following vowel had also a significant effect on frication duration ( $F(4)=13.258$ ,  $p<0.001$ ) (fricatives before i u > e o > a).

Regarding spectral mean, a similar 3-way Anova revealed a main effect of age ( $F(2)=14.459$ ,  $p<0.001$ ), with significantly higher values for children than the adult groups. Subsequent tests for each fricative revealed significant differences between children and adults for the fricatives /θ/, /s/, and /x/. A main effect was also revealed for place of articulation ( $F(4)=67.979$ ,  $p<0.001$ ), with /s/ having the highest value followed by /ç/, /θ/, /f/, /x/ and all differences being significant except for the one between /θ/ and /f/.

In the 3-way Anova with spectral variance as the dependent variable, a main effect of the age was revealed ( $F(2)=3.260$ ,  $p=0.039$ ), with the age groups 3-6 and 53-56 having significantly higher values than the group 23-26. Subsequent tests for each fricative revealed significant differences between the children and the adult groups for the fricatives /θ/ and /s/. A main effect was also revealed for place of articulation ( $F(4)=109.874$ ,  $p<0.001$ ). /s/ had the lowest values followed by /ç/, /x/, /θ/, /f/. Significant differences were revealed between /s/ and /f/, /θ/, /x/, and between /ç/ and /θ/, /x/.



Figures 1-3. Duration (ms), spectral mean and variance (Hz) of Greek voiceless fricatives as produced by children 3-6 and adults 23-26 and 53-56 years old.

According to 2-way ANOVAs within each age group, duration did not differentiate fricatives in children group, though it significantly differentiated /s/ from /f/, /θ/, and /x/ in both adult groups. Spectral mean significantly differed among all fricatives (except for /f/-/θ/) for the adults, while for children it significantly differentiated /f/ from all other fricatives, /θ/ from /s/, /x/, /s/ from /ç/, /x/, and /ç/ from /x/. Variance differentiated the front fricatives /f/ and /θ/ from all other fricatives in all age groups (except for /θ/ from /x/ in children). Hierarchical structure of fricatives as well as with respect to each vowel were similar for the different age groups for all variables.

Table 1. Mean duration (ms), M1 and M2 (Hz) values per fricative for the three age-groups (3-6, 23-26, 53-56 years old).

Fricative	Duration			M1			M2		
	3-6	23-26	53-56	3-6	23-26	53-56	3-6	23-26	53-56
f	158	130	124	4034	3461	3399	1924	2104	2068
θ	165	130	126	5171	4030	3259	1754	2036	2024
s	175	149	145	5978	5586	5506	1278	890	990
ç	169	139	136	5295	5244	4289	1040	1156	1265
x	166	130	127	2235	1508	1780	1446	1034	1557

## Discussion and conclusions

Overall, the results indicate significant temporal differences in fricative production between young children and adults as well as differences in spectral mean and variance.

In terms of duration, fricatives produced by children were significantly longer than those produced by adults. However, according to Nissen and Fox's (2005) research on English fricatives, although English speaking adults present shorter durations than children in /f/ and /θ/ productions, this changes for the alveolar and palato-alveolar fricatives, with children exhibiting shorter durations. Moreover, duration values of Greek speaking children were not found to differentiate fricative place of articulation, which is not the case

neither for Greek adult speakers (Nirgianaki 2014) nor for English speaking adults and children (Jongman et al. 2000; Nissen & Fox, 2005).

Fricatives produced by children also exhibited higher spectral mean values than those produced by adults; however, this has not been observed for the respective English fricative categories (Nissen & Fox 2005). Children's variance values were higher for /θ/ and /s/, revealing a more diffused spectrum for these fricative productions compared to the ones of adult speakers.

By confirming the existence of significant differences between children and adults in Greek fricative productions, the present study provides support to the argument that the immaturity of children's anatomical structures and motor control of vocal organs could be responsible for them (Vorperian et al. 2009). Furthermore, the accordance between the intra-category acoustic patterns of children and adult productions, as well as differences between the present results and previous ones reported for English similar sounds (Nissen & Fox 2005), support further the claim that children 'are attuned to language-specific relevant acoustic dimensions in mastering speech production' (Li 2012).

## References

- Bernthal, J.E., Bankson, N.W., Flipsen, P. 2018. Φωνολογιές και Αρθρωτικές Διαταραχές. Επιμ. Παπαθανασίου, Η. εκδόσεις Κωνσταντάρας.
- Boersma, P., Weenink, D. 2019. Praat: doing phonetics by computer [Computer program]. Version 6.0.50.
- Jongman, A., Wayland, R. & Wong, S. 2000. Acoustic characteristics of English fricatives. *J Acoust Soc Am*, vol.108, n.3, pp.1252–1263.
- Li, F. 2012. Language-specific developmental differences in speech production: A cross-language acoustic study. *Child Development*, vol.83, n.4, pp.1303–1315.
- Nirgianaki, E. 2014. Acoustic characteristics of Greek fricatives. *J Acoust Soc Am*, vol.135, n.5, pp. 2964 – 2976.
- Nissen, S. L. & Fox, R. A. 2005. Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *J Acoust Soc Am*, vol.118, n.4, pp.2570–2578.
- Nittrouer, S., 1995. Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *J Acoust Soc Am*, vol.97, n.1, pp.520-530.
- Syrika, A., Nicolaidis, K., Edwards, J. & Beckman, M.E. (2011). Acquisition of initial /s/-stop and stop-/s/ sequences in Greek. *Lang. Speech* 54(3), 361-386.
- Vorperian, H. K. & Kent, R. D., 2008. Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. *J Speech, Lang Hear Res*, December, vol.50, n.6, pp.1510–1545.

# Processing of singular agreement controller in Turkish

Ayşe Gül Özay Demircioğlu

English Language Education Department, TED University, Turkey

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The present study investigates whether the properties of the singular agreement controller in Turkish comitative construction ease the processing of the comitative construction and the singular agreement controller checks the agreement type on the verb (singular vs. plural). The data collected from 134 native Turkish speakers indicated that only one of the singular agreement controllers: the second person *sen* 'you' (both singular and plural agreement) was processed slower than the others, and there was no significant main effect of the agreement type (plural vs. singular).

Keywords: comitative construction, sentence processing, Turkish

## Introduction

### Comitative constructions in Turkish

Comitative is a broad term utilized to express the functions of modality, and accompaniment, and instrumentality (Haspelmath, 2004; McNally, 1993; Stassen, 2000). Almost all languages own their way of conveying the comitative meaning. Every language employs a different marker to convey this meaning, such as *mit* 'with' in German, *ile* 'with' in Turkish, or *with* in English.

Comitative construction in Turkish is expressed with a free post-position *ile* 'with'. This post-position is also formed with a bound allomorph *(y)la/le* as it is directly attached to the word. Turkish comitative construction can convey the relations of modality as in (1a), instrumentality in (1b) and coordination in (1c) (Göksel & Kerslake, 2004; Kornfilt, 1997).

- (1) a) Basri maç-ı heyecan-la bekli-yor-Ø.  
Basri.nom match-ACC excitement-with wait-PRSNT-3SG  
'Basri is waiting for the match with excitement.'
- b) Basri ekmeğ-i bıçak-la kes-ti-Ø.  
Basri.nom bread-ACC knife-INSTR cut-PAST-3SG  
'Basri cut the bread with a knife.'
- c) Basri ile ben restoran-a git-ti-m.  
Basri.nom with I restaurant-DAT go-PAST-1SG  
'I went to the restaurant with Basri.'

As shown in (1c), *ile* ‘with’ may refer to the human companion and appears between two nouns ‘Basri’ and ‘I’. However, the verb agrees with the second noun, first-person singular. This indicates that the structure is expressed as a prepositional phrase and the second DP checks the agreement. I referred to this DP as **singular agreement controller** as it checks the agreement on the verb and causes the singular agreement.

The interesting thing is that Turkish also allows the sentences as in (2). In other words, the comitative construction functions as comitative coordination, and no effect of singular agreement controller is observed. This structure can be interpreted as symmetrical comitative construction as both participants play equal roles. Nevertheless, the singular and plural reading of the same structure may create ambiguity until the verb is seen in the sentence.

- (2) Basri ile ben restoran-a git-ti-k.  
Basri.nom with I restaurant-dat go-PAST-1PL  
‘Basri and I went to the restaurant.’

### Research questions

- 1) Do the properties of the singular agreement controller (1st, 2nd and 3rd person) in the comitative construction ease the processing of the comitative construction?
- 2) Do the properties of the singular agreement controller in the comitative construction make one of the two readings more likely (singular vs. plural)?

### Expectations

As there is no previous literature on this topic, my null hypothesis is to find no effect of the properties of singular agreement controller on the interpretation comitative construction in Turkish. This study is significant because there is no previous literature investigating the comitative construction in Turkish from the experimental point of view.

### Methodology

#### Participants

Data were collected from 134 native Turkish speakers (97 females and 36 males). They had no problem with their visions and no problem with their language or reading. Their ages were between 18 and 22 ( $M=19.15$ ,  $SD=0.88$ ).

#### Instruments

The experimental items consisted of eight regions starting with a comitative construction with a singular agreement controller followed by an adverbial of place, the verb, and an adverbial of time. As the independent variables, the agreement on the verb and the singular agreement controller were changed



across the lists. The three types of singular agreement controllers (1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> person) were distributed across four lists with eight experimental items mixed with 12 filler items. A sample experimental item with the first-person singular agreement controller was shown in Table 1.

Table 1. Distribution of Regions.

R1	R2	R3	R4	R5	R6	R7	R8
Senle	ben	resim	sergisi-nde	gez-(i)yor.du-k	uzun	zaman	önce.
with you	I	exhibition-LOC	visit-prog.past-1pl	long	time	ago	
‘You and I were visiting the painting exhibition a long time ago.’							

Regions 6 and 7, spill-over regions, are critical regions. The verb region was not chosen as the critical region because its length differs since Turkish is an agglutinative language and the length of the affixes differs across the cases. Data were gathered through a self-paced reading task via a non-cumulative moving-window paradigm. Ibex Farm Program was used to collect data (Drummond, 2016). Ibex Farm is an easy way access via a link on any browser. The experiment started with five practice items, and they experimented.

### Data Analysis

The participants’ accuracy rates for the comprehension questions were checked, and those with less than 70% accuracy were removed. To eliminate the extreme values, the reading times lower than 100 ms (0.05% of all data) and higher than 3000 ms (0.2% of all data) were removed. Secondly, data were transformed via Lg10 transformation. The reaction times above 2.5 standard deviations were removed (4% of all data). The removed values were replaced by the mean of the region of this specific item. In the analysis, the dependent variable was the reaction times while the independent variable was the singular agreement controller (1<sup>st</sup>, 2<sup>nd</sup> or 3<sup>rd</sup>) and agreement on the verb (singular and plural). One-way Repeated-Measures ANOVA was conducted with a within-subject variable, the singular agreement controller (1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup>), and a between-subject factor, the agreement type on the verb (plural and singular).

### Results

One-way Repeated-Measures ANOVA (item analysis) indicated no significant main effect of the agreement type [ $F_2(1, 9)=2.1, p=.177$ ], but there was a significant main effect of the singular agreement controller [ $F_2(1, 9)=6.6, p=.017$ ], in addition to two-way interaction between agreement type and the singular agreement controller [ $F_2(2, 9)=5.4, p=.028$ ] in Region 6. Post hoc comparisons using the Bonferroni correction demonstrated a significant difference between 2<sup>nd</sup> and 1<sup>st</sup> (*sen* ‘you’ and *ben* ‘I’) and between 2<sup>nd</sup> and 3<sup>rd</sup> (*sen* ‘you’ and *o* ‘he/she’) ( $p<.05$ ). The second-person singular agreement controller was the hardest to process. Nevertheless, the same analysis on Region 7

showed no main effect of agreement [ $F_2(1, 9) = .19, p = .677$ ], no main effect of singular agreement controller [ $F_2(2, 9) = 1.2, p = .338$ ] or no interaction between two [ $F_2(2, 9) = 0.66, p = .542$ ].

## Discussion and conclusion

To summarize my results, when the singular agreement controller is investigated, only one of the singular agreement controllers: the second person *sen* 'you' was processed slower than the others. There was no significant main effect of the agreement type (plural vs. singular).

While investigating why the second person was hardest to process, the person hierarchy can be visited. Regarding person hierarchy, in the traditional account, the first person is followed by the second and the third (1st > 2nd > 3rd) (Filimonova, 2002). Therefore, person hierarchy cannot explain my results since it predicts that the processing of the comitative construction will be easier when the singular agreement controller is *sen* 'you' or *ben* 'I' compared to *o* 'he/she', contrary to fact. I do not know the motivation for this processing difficulty, but one part considers both semantic and morphological factors. As the 3rd person is not marked on the verb in Turkish, it may lessen its processing load, and as to the 1st person, as Filimonova (2002) suggested, it is higher in the hierarchy. Therefore, it may not be surprising to see its easiness of processing. For further research, it would be worth investigating the processing of the second-person *sen* 'you' in the regular coordination to see whether this processing difficulty in comitative construction is because of the nature of the comitative.

## References

- Drummond, A. 2016. Ibex farm. <https://github.com/addrummond/ibex>
- Filimonova, E. 2002. Person hierarchy and its implications: the case of Aymara. In Crevels, M., van de Kerke, S., Meira S. & van der Voort, H. (eds.) 2002, Current studies on South American languages [Indigenous Languages of Latin America, 3], 99-213. Leiden, Research School of Asian, African, and Amerindian Studies (CNWS).
- Göksel, A., & Kerslake, C. 2004. Turkish: A comprehensive grammar. Routledge.
- Haspelmath, M. 2004. Coordinating constructions: An overview. *Typological Studies in Language* 58, 3-40. <https://doi.org/10.1075/tsl.58.03has>
- Kornfilt, J. 1997. Turkish. Routledge.
- McNally, L. 1993. Comitative coordination: A case study in group formation. *Natural Language & Linguistic Theory* 11(2), 347-379. <https://doi.org/10.1007/BF00992917>
- Stassen, L. 2000. AND-languages and WITH-languages. *Linguistic Typology* 4(1), 1-54. <https://doi.org/10.1515/lity.2000.4.1.1>

# Noun plural inflection in German-speaking individuals with Down syndrome

Martina Penke, Bernadette Wittecy

Department of Special Education and Rehabilitation, University of Cologne, Germany

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The aim of the study was to investigate noun plural inflection in German speaking individuals with Down syndrome (IDS) in comparison to a control group of typically developing (TD) children. 40 noun plurals with different German inflectional endings were elicited. Accuracy scores as well as error types were analysed. Group comparisons indicated that noun plural inflection is affected in the participants with DS. In contrast to the TD children the dominant error type of the IDS were omissions, i. e. unmarked forms. This finding suggests that the observed deficit is not restricted to inflectional morphology per se, but also involves a violation of a prosodic constraint operating on the output of German noun plural inflection.

Keywords: Down syndrome, morphology, plural inflection, developmental language disorder

## Introduction

Down syndrome, caused by a trisomy of chromosome 21, is typically associated with delays and deficits in language acquisition. It has been argued that inflectional morphology is particularly affected in individuals with Down syndrome (henceforth IDS) (Chapman et al. 1998). Detailed analyses of inflectional deficits are, however, still relatively sparse. Moreover, studies have come to divergent findings with respect to inflectional deficits in IDS: while some have reported such deficits (e.g. Eadie et al. 2002, Penke 2018), others have found inflectional morphology to be unimpaired (e.g. Christodoulou & Wexler 2016, Ring & Clahsen 2005). We will contribute to this research by presenting data on German noun plural inflection.

German noun plurals can be marked by the inflectional endings *-s*, *-e*, *-er*, *-n*, *-en*, or can remain unmarked. Native German plural nouns are subject to a prosodic constraint that requires the plural form to end in a reduced syllable, i.e. an unstressed syllable with Schwa or a syllabic sonorant (e.g. *Bär* - *Bären* 'bear(s)', *Tisch* - *Tische* 'table(s)', *Kind* - *Kinder* 'child(ren)') (Neef 1998).

## Method

### Participants

31 monolingual German-speaking children and adolescents with DS (12 female), aged 4;07 to 19;02 years (*M* 14;05), participated in the study. Two of

them had a mild hearing loss, for the remaining IDS no permanent hearing loss had been diagnosed. Nonverbal mental age of the participants with DS was assessed using the SON-R 2.5-7 (Tellegen et al. 2007). It ranged from 2;11 to 6;05 years ( $M$  4;05). 26 monolingual typically-developing (TD) children, matched in chronological age to the MA of the IDS (3;04-5;07 years,  $M$  4;05) served as a control group.

### Procedure

40 noun plurals (eight plurals each for the endings *-s*, *-er*, *-e*, *-n* and *-en*) were elicited from each participant in a previously randomized order. All items were matched for lemma and plural-form frequency according to the CELEX database (Baayen et al. 1993). To tap into the productive abilities of the participants all tested items were of relatively low frequency.

Participants saw a picture with one object which they were asked to name, and were then presented with a picture showing three of these objects to elicit a plural form. They were tested individually after a short familiarization with the task. Test items were presented in the same previously randomized order. Participants' reactions were transcribed and included in the analysis if the target or a related noun was produced intelligibly.

### Data analysis

For each participant the percentage of correctly produced plural forms was calculated. In addition, the produced errors were analysed and categorized as either omission errors, i.e. production of the singular form without an inflectional ending (e.g. *Bär* instead of *Bären*), or substitution errors with an incorrect ending (e.g. *Bäre* instead of *Bären*). Group comparisons were performed using Welch's *t*-test and repeated measures ANOVA.

### Results

The participants with DS produced on average 38.2 analysable nouns out of the 40 elicited forms (95.5%). In the control group the mean number of analysable nouns was 39.8 (99.5%). The mean accuracy score of the group of IDS was 32.8% ( $SD$  24.7%, range 0%–85.0%), whereas the mean accuracy score of the TD group was 62% ( $SD$  13.6%, range 35.0%–82.5%), a significant difference ( $Welch(48.125) = 5.65$ ,  $p < .001$ ,  $d = 1.432$ ). Accuracy scores did not correlate with chronological or mental age in both participant groups ( $p > .1$  each).

The results of the error analysis are displayed in Fig. 1. They show that most of the errors in the group with DS were omission errors. In the group of TD children, on the other hand, substitution errors constituted the dominant error type. The repeated measures ANOVA with GROUP as between-subjects factor and ERROR TYPE as within-subjects factor yielded a significant GROUP\*ERROR TYPE interaction ( $F(1,55) = 12.76$ ,  $p = .001$ ,  $\eta_p^2 = .188$ ), confirming that the distribution of errors differed in the two participant groups.

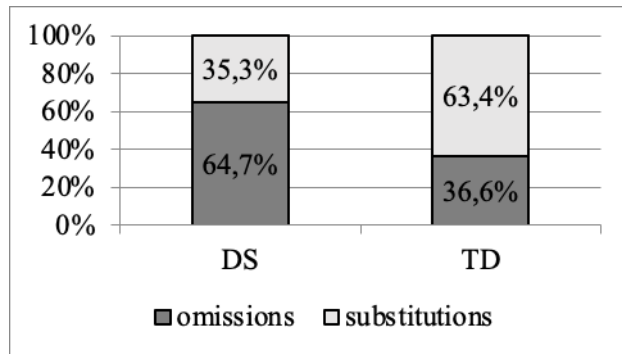


Figure 1. Mean percentage of omission and substitution errors.

A prosodic constraint requires all native German noun plurals (i.e. plurals other than *-s*-inflected) to end in a reduced syllable. For nouns requiring a plural form on *-e*, *-er*, or *-en* an omission of the plural ending (e.g. *3 Bär*, *3 Tisch*, *3 Kind*) results in a violation of this constraint. Error analysis yielded that a substantial proportion of the incorrectly produced plural forms for these nouns were left unmarked by the participants with DS ( $M$  61.5%) and were, thus, violating the prosodic constraint on plural forms. This proportion was significantly lower in the group of TD children (39.2%) ( $Welch(54.898) = 2.64, p = .011, d = .69$ ). The high proportion of produced plural forms that do not adhere to the prosodic constraint on German plural nouns suggests that this prosodic constraint is not fully operative in the participants with DS. Interestingly, however, the huge majority of the unmarked forms produced by the IDS were accompanied by a quantifier (e.g. the numeral *three* or the quantifier *many*) (86.4%), suggesting that the concept of plural was expressed by the quantifier instead of the unavailable inflected plural form.

## Discussion

Our results provide evidence that noun plural inflection is affected in German children and adolescents with DS. The significant difference in performance to a group of TD children matched in chronological age to the nonverbal mental age of the participants with DS indicates an inflectional deficit that cannot be accounted for by the cognitive limitations of IDS. These findings support previous research that found inflectional morphology to be impaired in IDS (e.g. Eadie et al. 2002; Penke 2018).

Moreover, the findings on incorrect, unmarked plural forms indicate that the observed deficit is not restricted to inflectional morphology per se, but encompasses the prosodic constraint that operates on the output of German noun plural inflection and requires these forms to end in a reduced syllable.

Despite these deficits, the observation that most unmarked nouns were produced with a preceding quantifier suggests that our participants with DS have already grasped the concept of plurality and were expressing it via a quantifier when they did not succeed in marking plurality by the inflectional ending (see Clark & Nikitina 2009).

### Acknowledgements

We would like to thank all participants and their families as well as everyone who was involved in data collection and analysis. This research was supported by the German Science Foundation (DFG, grant PE 683/3-1).

### References

- Baayen, H., Piepenbrock, R., van Rijn, H. 1993. The CELEX lexical database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Chapman, R.S., Seung, H.K., Schwartz, S.E., Kay-Raining Bird, E. 1998. Language skills of children and adolescents with Down syndrome: II. Production deficits. *Journal of Speech, Language, and Hearing Research*, 41 (4), 861–873.
- Clark, E.V., Nikitina, T.N. 2009. One vs. more than one: Antecedents to plurality in early language acquisition. *Linguistics*, 47(1), 103–139.
- Christodoulou, C., Wexler, K. 2016. The morphosyntactic development of case in Down syndrome. *Lingua*, 184, 25–52.
- Eadie, P.A., Fey, M.E., Douglas, J.M., Parsons, C.L. 2002. Profiles of grammatical morphology and sentence imitation in children with specific language impairment and Down syndrome. *Journal of Speech, Language, and Hearing Research*, 45(4), 720–732.
- Neef, M. 1998. The reduced syllable plural in German. In Fabri, R., Ortmann, A., Parodi, T. (eds.), *Models of inflection*, 244–265, Tübingen: Niemeyer.
- Penke, M. 2018. Verbal Agreement Inflection in German Children with Down Syndrome. *Journal of Speech, Language, and Hearing Research*, 61, 2217–2234.
- Ring, M., Clahsen, H. 2005. Morphosyntax in Down's syndrome: Is the extended optional infinitive hypothesis an option? *Stem Spraak en Taalpathologie*, 13(1), 3–13.
- Tellegen, P., Laros, J., Petermann, F. 2007. SON-R 2,5-7. Non-verbaler Intelligenztest. Göttingen, Hogrefe.

# Text presentation and information processing in Russian

Tatiana Petrova

Institute for Cognitive Studies, St Petersburg University, Russia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This study is based on the Cognitive Theory of Multimedia Learning (Mayer 2009). We investigated how readers process and comprehend information when reading and listening texts of different formats: infographics, audiotext, infographics combined with audiotext, written format. Biographies of four Russian writers were used as the material. All the stimuli were of the same size and the same level of readability. In a four-group design experiment, 32 foreign students and 32 native speakers of Russian examined four texts in four different formats, answered the factual and analytical after-the-text questions, gave the keywords and estimated the difficulty of each text. The overall results made it possible to build a hierarchy of best-perceived formats for both groups of participants.

Keywords: text processing, multimodal text, text format, and comprehension

## Introduction

Multimodal representation of information is getting more and more popular nowadays in various spheres of life. Polycode texts have become an integral part of modern communication. Studies of infographics (i.e. drawings, graphs, diagrams, etc. accompanied by short captions), sketch noting (visual notes consisting of hand-written text and visual elements), advertising texts, multimedia courses integrating verbal and non-verbal means, oral and written presentation of material, and allowing to present complex information in the optimal way, gain ground. Polycode text is a text composed of attributes belonging to different semiotic systems, namely of two non-homogeneous parts: verbal (lingual/spoken) and non-verbal (belonging to other than natural language semiotic systems) (Paivio, 2006). Mayer's (2009) Cognitive Theory of Multimedia Learning suggests that effective engagement with the material presented in the form of polymodal text occurs due to the need to switch attention between text and image, oral signal and text, and establish the connection between these elements. This leads to integration of the new information into the existing cognitive system, as well as the ability to use the acquired information in the future.

Numerous experimental studies of reading mechanisms have been carried out in recent years. However, the text level as an integral unit remains understudied, at least on Russian language material. In particular, of essential

interest is how the text format is related to the text comprehension quality. What type of signal presentation (verbal text or written text) would be most effective in the e-learning format for certain readers/listeners; whether the text presentation of information in a non-verbal form promotes or, conversely, interferes with perception and understanding, and vice versa, how concomitant verbalization affects the understanding and memorization of figurative information and its emotional assessment, remains unclear.

## Hypothesis

Reading, processing, and comprehension in texts of different formats are different: multimodal text is easier to process and comprehend than written or oral text of the same content.

## Experimental design

### Material

Four infographics containing interesting facts from the biographies of four Russian writers were used as the material. Infographics were extracted from Russian newspapers (<http://aif.ru>) and then converted them into regular text format. All the texts and infographics were of the same length and the same level of readability (checked via <http://readability.io/>). The infographics contained the same proportion of verbal and non-verbal elements (<https://cloud.mail.ru/public/HVsi/ujuEL4zFZ>). Afterwards, all the texts were read by the same dictator (female, 23 years old, Russian) and audio-recorded. Thus, we got the 16 stimuli to check how participants integrate text-figure information when processing and understanding the same text in four different formats: infographics (graphic visual representation of information), audiotext, infographics combined with audiotext, plain written format.

Table 1. Main characteristics of the experimental texts.

Text	Readability level	N of signs	N of words	N of sentences
Text 1 'Pushkin'	7,71	202	1520	16
Text 2 'Chekhov'	7,52	198	1446	21
Text 3 'Gorky'	7,62	196	1398	15
Text 4 'Yesenin'	7,71	193	1405	20

### Participants

32 students learning Russian as a foreign language (experimental group, female=21, Mage=23, SD=4, B1-B2 level of Russian language defined by the Common European Framework of Reference for Languages) and 32 native speakers of Russian (control group, female=27, Mage=23, SD=8) participated in the experiment.



## Procedure

In a four-group experimental design, the participants examined four texts in four different formats. Afterwards, they answered six factual and six analytical questions after each text, identified the keywords, and estimated the subjective difficulty of each text using five scales (from  $-2$  to  $+2$ ). Text comprehension was also controlled by the cloze-test technique. Every participant examined each format of the text and each biography only once. Each text was presented on the computer screen for 5 min. For the oral format there was an opportunity to listen the audiotext twice. All the stimuli were presented in randomized order. The experiment was carried out through the platform <http://coreapp.ai>. It lasted around 50 minutes for each foreign participant, and around 25 minutes for Russian native speakers.

## Results

We used JASP for calculating the statistics (version 0.14.1.0) and the Mann-Whitney U test to compare the results we obtained. We found the factor of format of presentation to be significant for all the experimental texts. We revealed that audiotext is the most challenging text format for both groups of participants. Written text is the easiest and better-perceived format for foreign students. No significant differences were found in processing the text presented in infographics, written form and combined (audio + written) modalities in a group of Russian native speakers (all  $p$ s  $> 0.05$ ). The link to the data is the following: <https://cloud.mail.ru/public/HfFZ/Uknu6NFF5>. The overall results made it possible to build a hierarchy of best-perceived formats for the foreigners: written text — infographics — multimodal text (infographics + audiotext) — audiotext.

## Discussion

In the present study, we examined what format of text presentation contributes to more successful information processing and understanding. We found out that audiotext is the most difficult format of presentation. Surprisingly, we revealed that written text is the easiest text format for those who learn Russian as a foreign language. Infographics appeared to be more difficult to process and worse to use for retrieving the essential information. Presumably, these results are due to the difficulties in parallel cognitive processing of graphic and verbal elements (Blinova, Shcherbakova 2019). In the process of perceiving a polycode text, such as infographics, the information contained in it undergoes double decoding: when the image concept is extracted, it is "superimposed" on the verbal text concept, and interaction of the two concepts leads to creation of a single general concept (rendering) of a polycode text, which confirms the idea, that information perceived through different channels, including verbal and iconic, is integrated and processed by a person in a single universal subjective

code of thinking (Zhinkin 1982; Paivio 2006; Fernández-Fontecha et al. 2018). Our results show that the so-called multimedia effect helps to integrate the new information in the cognitive system and to remember, though it is difficult and resource-intensive process, especially for non-native speakers.

## Conclusion

Developing the best compromise and form of text presentation in oral and written form becomes especially relevant in connection with the move to remote learning, in which many formats of interaction between the teacher and students involve a combination of auditory and visual modalities. Our research has shown that the text format is among the readability categories (DuBay 2004). Also, we assume that comprehension of a text depends more on the reader's factor than on text format.

## Acknowledgements

The study was supported by the research grant no. 21-18-00429 “Cognitive mechanisms of multimodal information processing: text type & type of recipient” from Russian Science Foundation. I would like to thank my student Daria Skvortsova for data collection and Vladislav Zubov for his helpful comments and suggestions regarding the experimental design.

## References

- Blinova, E., Shcherbakova, O. 2019. Cognitive Processing of Verbal Texts and Infographics: Different or the Same?//The 5th Saint Petersburg Winter Workshop on Experimental Studies of Speech and Language (Night Whites 2019). P.41.
- DuBay, W.H. 2004. *The Principles of Readability*. California. 73 p.
- Fernández-Fontecha, A., O'Halloran, K.L., Tan, S., Wignell, P.A. 2018. Multimodal approach to visual thinking: the scientific sketchnote. *Vis. Commun.* 18 5–29.
- Mayer, R.E. 2009. *Multimedia Learning*, 2nd Edn Cambridge, MA: Cambridge University Press.
- Paivio, A. 2006. *Dual Coding Theory and Education*. Draft Chapter for the Conference on Pathways to Literacy Achievement for High Poverty Children. Ann Arbor, MI: The University of Michigan School of Education.
- Zhinkin, N.I. 1982. *Rech' kak provodnik informacii*. M.: Nauka, 1982. 157 p.

# Exploring inter-dialectal mutual intelligibility and SDA in Assamese

Sabbah Qamri

School of Language Sciences, The English and Foreign Languages University, India

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Inter-dialectal Mutual Intelligibility can be defined as the rate at which speakers of different dialects understand each other. And, the process in which the speakers of a dialect (D1) acquire a different dialect (D2) of what they consider the same language is referred to as Second Dialect Acquisition. This study examines both phenomena in the context of four dialects of Assamese from the Indo-Aryan family. It also explores a possible relationship between them. The general conclusions are that inter-dialectal mutual intelligibility in Assamese is asymmetric, and a high rate of such asymmetry between a non-standard dialect (D1) and Standard Assamese (D2) results in better rates of D2 acquisition in the classroom by native speakers of D1(s).

Keywords: Assamese, dialectology, asymmetric intelligibility, second dialect acquisition, language education

## Introduction

Theoretically, mutual intelligibility (henceforth, MI) is an important criterion that distinguishes Dialect from Language. Although one might posit the occurrences of mutually unintelligible dialects in the spoken form to counter this criterion, such varieties can still be justly termed dialects of a language if they share a common written form. The many regional, social, and ethnic dialects of Assamese— spoken in north- eastern India, despite variations in morpho-phonology, syntax, and vocabulary, share a written form and their speakers are mutually intelligible. However, the degree of this MI is unknown.

In classroom Second Dialect Acquisition (henceforth, SDA), the learners are children who come to school speaking non-standard dialects as their D1. Their speech is usually markedly different from the target language of the education system, i.e., the standard variety. Moreover, since dialects are considered to be more similar and mutually intelligible than languages, the process of SDA is considered to be easier than Second Language Acquisition. But there is an immense lack of any teaching approaches or special programs designed for the acquisition of the standard variety as the D2 by educational institutions. Most classes on reading and writing in any language are conducted in the standard variety assuming prior knowledge of the D2 by all children— native speakers of the standard as well as non-standard dialects. This holds true for education imparted in Assamese as well.

The lack of existing research on either phenomenon as well as on a possible relationship between them in the context of Assamese, or in any regional language in India, led to the present study.

## Methodology

Four regional dialects— Standard Assamese (SA), Central Assamese (CA), Kamrupi (KA), and Goalparia (GA), located in a continuum from the geographical east to the west, were considered for the study. To study inter-dialectal MI, a Functional Testing approach was used. Texts were collected from 6 speakers from each dialect (1 male and 1 female speaker from 3 locations). These texts included words, sentences, and a free speech sample in the speaker's native variety along with one common passage in the standard variety. 132 listeners (2 speakers per listener) were asked to listen to these texts and their comprehension was tested. Intelligibility was calculated based on the rates of comprehension of:

- i.) For a speaker—the 11 listeners who listened to their text.
- ii.) For the overall intelligibility of a dialect—the 66 listeners who listened to texts from that dialect.
- iii.) For the intelligibility of one dialect w.r.t another— the 18 listeners from the second dialect who listened to texts from the first.

To examine the phenomenon of SDA in Assamese, due to CoVid-19 travel restrictions, only the records of marks obtained by students in the subject of Assamese from 8 schools in the four dialect-areas were collected online from educators of the schools and analyzed. A total of 480 students of the 6th grade across three academic sessions were included.

## Results and discussion

From the analysis of the data collected through the Functional Testing method, the overall rates of intelligibility of the four dialects were found to be: SA= 68%, CA=70%, KA= 58%, and GA= 54%. In other words, the dialects which are most and least intelligible among speakers of Assamese are the dialects closest to and farthest from SA—CA and GA respectively. The rates of intelligibility of the two western dialects, which vary structurally from SA, are the lowest.

The inter-dialectal MI in Assamese, as illustrated in Figure 1, is highest between the neighboring dialect-pairs of SA-CA and CA-KA. The difference in intelligibility towards each other is only 4 % in both cases. But, the inter-dialectal MI between the geographically separated pairs of dialects, which also vary greatly in vocabulary, is low and asymmetrical. In the case of SA, the asymmetry in its MI with KA and GA is 12% and 11% respectively. The rate of asymmetry, however, is the highest between CA-GA at 15 %.

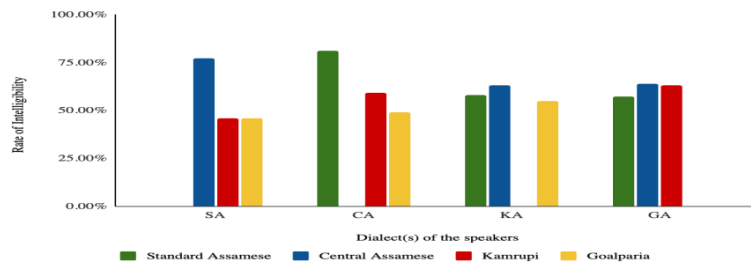


Figure 1. Rate of Inter-dialectal Intelligibility among Assamese speakers.

To examine SDA in the context of Assamese, a group analysis of the data collected from the schools was conducted. Based on their scores, the 120 students from each dialect were sorted into four groups of 0-30, 31-50, 51-79, and 80-100 marks. A comparison of the grouping of all four dialects revealed that the students who spoke the western non-standard dialects of KA and GA as their D1 showed better rates of acquisition of SA as the D2 during classroom SDA. The highest concentration of students for the dialects of CA, KA, and GA were found to be in the ranges of 31-50, 51-79, and 80-100 marks respectively. In other words, the average performance of the students in SDA seems to improve as the geographical distance between the D1 and D2 increases and the asymmetry in the rates of inter-dialectal MI between SA and the non-standard dialects becomes higher. This relationship is illustrated in Figure 2 below.

Moreover, the rate of D2 acquisition in school for speakers of different Assamese dialects seems to be inversely proportionate to the rate of intelligibility of the dialects. As seen in Figure 3 below, the dialect with the highest rate of intelligibility (CA) is the one with the worst performance in SDA whereas the dialect with the lowest rate of intelligibility (GA) is the best-performing in SDA.

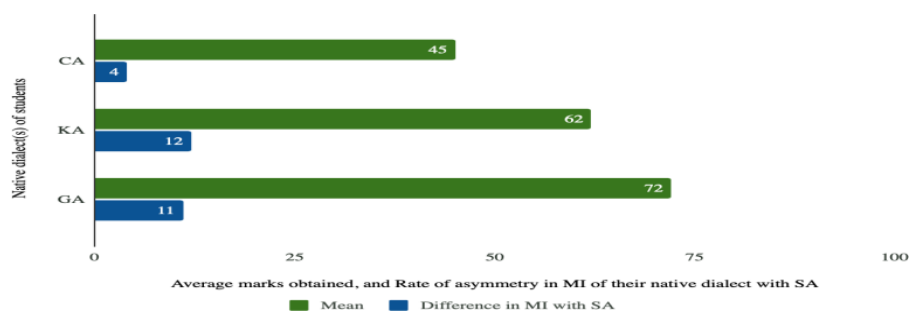


Figure 2. The relationship between inter-dialectal mutual intelligibility of non-standard dialects with SA and the process of SDA in Assamese.

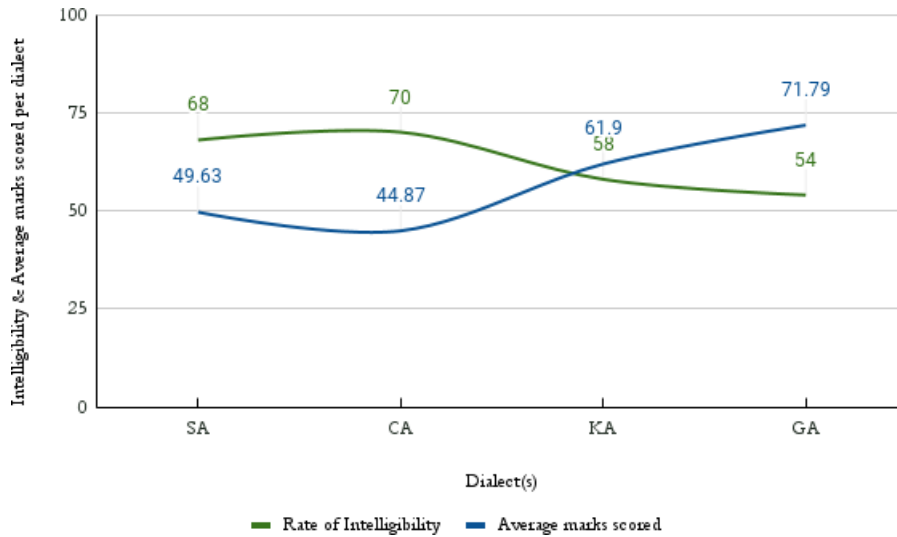


Figure 3. The relationship between overall intelligibility of Assamese dialects and SDA in Assamese.

In conclusion, the geographical location of dialects seems to play an important role in determining the rates of both inter-dialectal MI and classroom SDA. In this case, increasing geographical distance between SA and the 3 non-standard dialects implies greater structural variation from SA and hence, a lower rate of MI but a higher rate of SDA. And, the possibility of the rates of SDA in a language being related to the rates of dialectal intelligibility also opens up interesting avenues of research in Intelligibility Studies as well as in Language Education.

## References

- Das, D.N. 1990. The dialects of Goalpara and Kamrup: A Comparative Analysis (Doctoral Thesis). Guwahati: Gauhati University.
- Das, H. 2012. A Regional Dialect of Assamese Language. IJCAES, Vol. II. Gooskens, C. 2018. Dialect Intelligibility. The Handbook of Dialectology. First Edition. John Wiley & Sons, Inc. 204-218.
- Goswami, U.N. 1958. A Study on Kamrupi: a Dialect of Assamese (Doctoral Thesis). Guwahati: Gauhati University.
- Kakati, B.K. 1962. Assamese, its formation and development, 3<sup>rd</sup> edition. Guwahati: LBS Publications.
- Siegel, J. 2010. Second Dialect Acquisition. Cambridge: Cambridge University Press.

# Language teaching and assessment in the context of World Englishes

Aicha Rahal

Laboratoire Parole et Langage, Aix-Marseille University, France

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Today's English is no longer seen as a monolithic entity or a linguistic standard, rather it is distinct varieties. It is the product of 'multilingualism' and 'glocalization'. This change in the global status and usage of English leads to changing the traditional paradigm. There is, therefore, a pressing need to rethink new ways of teaching and assessing multilingual English. Theoretically, this paper gives an overview of WEs and the different issues in teaching and assessment. At the practical level, the study highlights my experience in teaching some of the varieties of English. The results of this study may have its pedagogic contributions. The lesson plan and the assessment activities can be used as a reference for a number of teachers to create a bridge between theory and practice.

Keywords: World Englishes, teaching pedagogy and assessment

## World Englishes

World Englishes (WE) refers to the new varieties that emerged in local contexts. It is defined by Kirkpatrick (2007) as “indigenous, nativised varieties that have developed around the world and that reflect the cultural and pragmatic norms of their speakers” (p. 3). In describing WE, Rahal (2019a and b) presents the multifarious aspect of English arguing English has become a multilingual language.

## Issues in teaching and assessing World Englishes

The present study raises two major issues related to teaching and assessment in the context of World Englishes. It is unrealistic to apply the monolingual approach in teaching local varieties. Galloway (2017, pp. 17- 21) listed the following issues in teaching WEs:

- persistent, exclusive exonormative orientation towards two main reference varieties of English (= BrE and AmE) in teaching and testing/assessment practices.
- Standard language ideology and native-speakerism in teacher education and teacher recruitment practices.
- Little research on if and how diversity of English is integrated into teacher training programs.

- Unawareness of linguistic variation among EFL teachers.
- Curricula provide only vague and underspecified descriptions as to language variation.
- Lack of practical teaching materials.

Jenkins (2006, p. 10) criticizes the traditional paradigms arguing that they “are unable to cope with the fact that language is messy, and lingua franca use is even messier, which renders futile the attempt to impose a present template on contingent use in diverse English contexts”.

### **A need for global paradigm**

There is a need “to invent the language we are supposed to teach” (Decke-Cornill, 2002, p. 59) and create new paradigms that aim to:

- Increasing exposure to World Englishes and ELF in language curricula
- Emphasising respect for multilingualism in ELT
- Raising awareness of Global Englishes in ELT
- Raising awareness of ELF strategies in language curricula
- Emphasising respect for diverse culture and identity in ELT (Rose and Galloway, 2019, p. 16).

Moreover, McNamara (2011) calls for a paradigm shift in language assessment that integrates the new philosophy of World Englishes and considers the new changes.

### **Lesson plan overview:**

Lesson Plan Title: Teaching local varieties: Ghanaian English, and Cameroon English  
Recommended Level(s): Intermediate and advanced students

### **Procedures**

#### **Pre-activity: Warm up Activity (5 Minutes)**

- -The teacher elicits a discussion about the different varieties of English and she asks her students whether they have been exposed or noticed the existence of new varieties of English apart from British and American English.
- -The students start talking about their experience, stating the differences they noticed and expressing their feeling.



**While-activity: Group work (15 minutes)**

- The teacher uses the sentence betting activity. She divides the students into two groups and gives them cards of money designed for activities. Then, she writes sentences from different local varieties, namely British English, American English, Ghanaian English, and Cameroon English on other cards. She provides the students with three options.
- The students in each group discuss the activity. Each group has to bet a certain amount of money. The teacher gives the correct answer and the group, that wins, gets the money.

**Small Group Discussion (10 Minutes)**

In small groups, a student reads a sentence and gives cultural clues and the other student tries to guess the country.

**Post-activity**

**Assessment activity: wrong one out activity (15 minutes)**

The teacher writes on the board a number of words, expressions that belong to the studied varieties of English and she includes some examples that do not belong to any variety. Each student reads an example and tries to find the wrong one. This goes on until all students participate.

**Assessment activity: Quizzes (15 minutes)**

Test each other on the features of Ghanaian English, and Cameroon English at the level of vocabulary, going both ways: Say a word in the Ghanaian English, or Cameroon English South and see if your partner can identify the word in British or American English.

**References**

- Decke-Cornill, H. 2002. We would have to invent the language we are supposed to teach: The issues of English as a lingua franca in language education in Germany. In Byram, M., Grundy, P. (eds.), *Context and Culture in Language Teaching and Learning* (pp. 59-70). Toronto: Multilingual Matters.
- Galloway, N. 2017. *Global Englishes and Change in English Language Teaching. Attitudes and Impact*. London and New York: Routledge.
- Jenkins, J. 2006. Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly*, 40(1), 157-181.
- Kirkpatrick, A. 2007. *World Englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press.

- McNamara, T. 2011. Managing learning: Authority and language assessment. *Language Teaching*, 44, 500-515.
- Rahal, A. 2019a. English or Englishes? A question of multilingual reality. In Christiansen, B., Turkina, E. (Eds.), *Applied psycholinguistics and multilingual cognition in human Creativity* (pp. 83-102). IGI Global.
- Rahal, A. 2019b. Transdisciplinary Approach to Linguistic Diversity: Can we co-exist without “One English”? In Wang, V.X. (Ed.), *Handbook of Research on Transdisciplinary Knowledge Generation* (pp. 383-396). IGI Global.
- Rose, H., Galloway, N. 2019. *Global Englishes for Language Teaching*. Cambridge: Cambridge University Press.

# Validation of the Moroccan Arabic version of the Token Test

Youssef Rami<sup>1</sup>, Samir Diouny<sup>2</sup>, Najib Kissani<sup>3</sup>, Assma Boudanga<sup>3</sup>

<sup>1</sup>Applied Language & Culture Studies, Chouaib Doukkali University, Morocco

<sup>2</sup>Clinical Neuroscience & Mental health, Hassan II University, Morocco

<sup>3</sup>Neurology Department, University Hospital Mohammed VI, Marrakech, Morocco

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The aim of this study was to translate and cross-culturally adapt the Token Test. 54 participants were recruited to evaluate the validity ratio (CVR) of the Token Test. A one-way between subjects ANOVA was conducted to compare MATT scores of 25 non-brain-damaged individuals, 18 individuals with aphasia and 11 individuals with epilepsy. The one-way showed a significant difference at the  $p < 0.01$  level for the three groups [ $F(2, 51) = 72.31, p < 0.01$ ]. Auditory comprehension raw score was significantly correlated with the token test score ( $r(18) = 0.836, p < 0.01$ ). The aphasic group had higher variance in terms of Token Test scores. The results support the appropriateness of using MATT to assess language skills in Moroccan Arabic-speaking subjects with cognitive deficiencies.

Keywords: adaptation, Token Test, cognitive deficiencies, Moroccan Arabic, aphasia

## Introduction

Aphasia is an acquired language disorder that results from damage to areas of the brain that produce and process language. Aphasia can be assessed using a series of screening tools, including the Token Test (De Renzi & Vignolo, 1962), a widely used tool. The Token Test is made up of tokens that come in two shapes (rectangle and circular), two sizes (large and small), and five colors. The Token Test has been translated to 40 different languages (Bastiaanse et al., 2016). Despite this, a valid assessment tool is not available in Moroccan Arabic.

## Methods and procedures

### Study population

54 right-handed individuals participated in this study. The group consisted of 25 healthy adults ranging in age from 27 to 76 with a mean of 51.60 (SD=14.13). Ages for 18 patients with brain damaged in the left hemisphere ranged from 41 to 80 with a mean of 61 (SD=12.28). The mean of 11 Epileptic patients was 42.36 (SD=11.36) with a range of 25-62. The time elapsed since onset of aphasia ranged from 1 week to 3 years. The stroke and epileptic subjects were referred to the university hospital M6 (CHU M6) and other clinics in Marrakech for rehabilitation. The epileptic patients were identified as

non-aphasic by the Moroccan version of the Montreal–Toulouse Protocol of Aphasia Linguistic Examination (TM86) (El Alaoui Faris et al., n.d), while the aphasic patients had experienced stroke and were identified aphasics by the same assessment.

### **Adaptation and procedure**

The purpose of translating a test is to create a version that is conceptually equivalent to the original. This can be accomplished by following the international gold standard translation (Efstathiou, 2019) and the translation techniques recommended by the MAPI Research (Acquadro et al., 2012). Translation steps that comply with these guidelines were forward translation from English to Moroccan Arabic, a review of forward translation, backward translation into English, a review of the backward translation, pretesting the translated version of the test, followed by final testing.

### **Statistical analysis**

To compare MATT scores of the three groups: post-stroke aphasic group, Epileptic group, and non-brain-damaged group, a one-way between subjects ANOVA was conducted. A bivariate correlation was conducted to investigate how the MATT scores relate to the raw auditory comprehension score of the Moroccan version of the Montreal–Toulouse Protocol of Aphasia Linguistic Examination (TM86).

### **Results**

Regarding Shapiro-Wilk's test ( $p > 0.05$ ), skewness and kurtosis values, our data are a little skewed and kurtotic, but they do not differ significantly from normality. We can assume that our data are normally distributed. A one-way between subjects ANOVA was conducted to compare Token Test scores of the three groups: post stroke aphasic group, Epileptic group, and healthy group. There was a significant difference at the  $p < 0.01$  level for the three groups [ $F(2, 52) = 72.31, p = .00$ ]. Post hoc comparisons using the Tukey HSD test showed that the scores of the healthy group ( $M = 32.60, SD = 2.82$ ) was significantly higher than both the Epileptic group ( $M = 30.86, SD = 2.32$ ) and post-stroke aphasic group ( $M = 17.47, SD = 6.25$ ). Table 1 summarizes the comparison of mean scores for all the participants on the MATT and the auditory comprehension test. The Epileptic group scored significantly higher than the post-stroke aphasic group ( $p = .00$ ).

As table 2 illustrates, a bivariate correlation was conducted to see how the MATT scores relate to the raw auditory comprehension score of the Moroccan version of the Montreal–Toulouse Protocol of Aphasia Linguistic Examination (TM86), which is a sentence/word-picture matching task. Participants had to point to the correct action or picture that matches the word or sentence. Auditory comprehension raw score was significantly correlated with the token test score ( $r(18) = 0.836, p = .00$ ). Compared to the two non-aphasic groups, the

aphasic group had higher variance in terms of Token Test scores, which suggests that aphasic severity can be measured through this task.

Table 1. Performance on part 1 to 6 of the MATT and Auditory Comprehension.

Group		MATT						AC	
		Part1	Part2	Part3	Part4	Part5	Part6	Total	
H	M	7.00	4.00	3.96	3.74	3.62	10.28	32.60	12.36
	SD	.00	.00	.13	.32	.415	2.28	2.82	.99
E	M	7.00	4.00	3.90	3.86	3.50	8.59	30.86	11.72
	SD	.00	.00	.30	.23	.31	2.17	2.32	1.67
A	M	6.52	3.05	2.69	2.08	1.61	1.50	17.47	7.72
	SD	.65	.98	1.15	1.16	1.19	2.07	6.25	2.49

H=Non-brain damage, E=Epileptic, A=Aphasic, A.C= Auditory Comprehension, M= Mean, SD= Standard Deviation

Table 2. Correlations of MATT and the Auditory Comprehension.

		MATT	A.C
Healthy	MATT	1	.772**
	Auditory Comprehension	.772**	1
Epileptic	MATT	1	.681*
	Auditory Comprehension	.681*	1
Aphasics	MATT	1	.836**
	Auditory Comprehension	.836**	1

\*\*  $p < 0.01$ , \* $p < 0.05$

## Discussion

This study evaluated the efficiency of available screening tools to accurately identify potential language problems in Moroccan aphasic patients. This investigation revealed subtle receptive language processing that potentially contributes to the language performance frequently reported among Moroccan-speaking aphasic patients.

The Token Test performance of the three groups were significantly different from one another. The healthy group scored higher than the other groups. The epileptic patients scored slightly lower, and the aphasic stroke patients performed the poorest of the three groups. The MATT was significantly correlated with the auditory comprehension score of the Moroccan version of the Montreal–Toulouse Protocol of Aphasia Linguistic Examination (TM86).

## Conclusion

The findings of the current study indicate that the MATT is a valid measure for the assessment and diagnosis of aphasic symptoms in Morocco, in particular, distinguishing aphasic to non-aphasic subjects, and providing information regarding aphasic severity.

## Acknowledgements

The authors thank Professors Abdelkader Marrah and Jamaa Ouchouid from the English Department of the faculty of Arts and Humanities Cadi Ayyad, Marrakech. We like to express our gratitude to all the staff of the neurology department of the University Hospital Mohamed VI.

## References

- De Renzi, E., Vignolo, L.A. 1962. The token test: A sensitive test to detect receptive disturbances in aphasics. *Brain: a journal of neurology*, 85, 665–678. <https://doi.org/10.1093/brain/85.4.665>
- Acquadro, C., Conway, K., GirouDET, C., Mear, I. 2012. Linguistic validation manual for health outcome assessments
- Bastiaanse, R., Raaijmakers, S., Satoer, D., Visch-Brink, E. 2016. The Multilingual Token Test. *Aphasiology*, 30(4), 508–508. <https://doi.org/10.1080/02687038.2015.1121710>
- El Alaoui Faris, M., Ettahiri, L., Fennich, R., & Benbelaid, F. (n.d.) Montreal–Toulouse Protocol of Aphasia Linguistic Examination: MA version. (Unpublished). Rabat: Faculty of Medicine.
- Efstathiou, G. 2019). Translation, Adaptation and Validation Process of Research Instruments. In R. Suhonen, M. Stolt, & E. Papastavrou (Eds.), *Individualized Care* (pp. 65–78). Springer International Publishing. [https://doi.org/10.1007/978-3-319-89899-5\\_7](https://doi.org/10.1007/978-3-319-89899-5_7)

# Reverse transfer of metacognitive reading strategies of Moroccan trilingual learners

Hassane Razkane<sup>1</sup>, Samir Diouny<sup>2</sup>

<sup>1</sup>Applied Language & Culture Studies Lab, Chouaib Doukkali University, Morocco

<sup>2</sup>Clinical Neuroscience & Mental health Lab, Hassan II University, Morocco

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The Linguistic Interdependence Hypothesis assumes that certain L1 language skills can be positively transferred during the process of L2 acquisition. While the transfer of linguistic skills from L1 to L2 has been studied extensively, only few studies have investigated the reverse transfer of skills. This study examined whether training poor readers in reading strategies in English (L3) would lead to an improvement in the process of reading in French (L2) and Arabic (L1). The Metacognitive Reading Strategies Questionnaire (MRSQ) (Abu-Rabia, 2018) was administered to 30 first-year baccalaureate students before and after the intervention. The participants were purposively selected and randomly assigned to an experimental group (20) and a control group (10). Findings indicated a significant improvement in the experimental group's metacognitive strategy awareness in English, French and Arabic.

Keywords: Strategy awareness, trilingual learners, CRT, reading strategy instruction

## Introduction

Reading comprehension is a complex process that requires the use of multiple (meta-)linguistic and (meta-)cognitive skills (Perfetti et al., 2005). Regarded as tools that facilitate understanding of texts, (meta)cognitive reading strategies (MCRS) are found to ameliorate reading comprehension outcomes (Wu et al., 2021). Current research has documented the use of MCRS in first language L1 and second language L2 and results have reported a forward transfer of MCRS from L1 to L2 (Abu-Rabia, 2018; Rabadi et al., 2020). These findings support Cummins' (1979) *Common Underlying Proficiency* (CUP) hypothesis, which assumes the existence of one common language proficiency that enables bilinguals to transfer acquired (meta-)linguistic skills from one language to another.

Talebi (2012) examined the effect of reading strategy training in English (L2) on improving reading strategy awareness of Iranian learners of English (L2) and Persian (L1). Results showed that the participants in the experimental group revealed a reverse transfer of reading strategies from L2 to L1. The researcher concluded that teaching reading strategies in L2 can boost students' reading strategy awareness and reading comprehension skills in both L2 and L1. Talebi's (2012) findings are supported by the Cognitive Retroactive Transfer (CRT) hypothesis, which proposes a reverse transfer of language skills from L2

to L1 after intervention (Abu-Rabia & Bluestein-Danon, 2012; Abu-Rabia et al., 2013). The present study examined whether training poor readers in reading strategies in English (L3) would lead to an improvement in the process of reading in French (L2) and Arabic (L1).

## Methods

### Participants

Data were collected from 30 first-year baccalaureate students from a High School in El Jadida, Morocco. Based on their performance in reading comprehension tests in English, French and Arabic, the participants, who scored less than 6 out of 10 in the study languages, were purposively selected and randomly assigned to an experimental group (10 females, 10 males, mean age: 17.20, SD: 0.951) and a control group (5 females, 5 males, mean age: 17.10, SD: 0.876). The participants started learning Standard Arabic as L1 in grade 1, French as L2 in grade 2 and English as an L3 in grade 9. To internalize MCRS, only the experimental group received explicit instruction in reading strategy awareness.

### Material and procedure

Both groups were administered a reading comprehension task (one in English, one in French and one in Arabic) and the MRSQ (Abu Rabia, 2018), translated into Arabic, to assess the participants' reading strategy use in the study languages before and after the intervention. The experiment group received a twelve-week explicit instruction in the use of MCRS twice a week (45 minutes in each training session). The difference between metacognitive strategy awareness means before and after the intervention was examined using one-way MANOVA.

## Results

Results regarding the participants' reading strategy use revealed significantly higher means in the experimental group's use of MCRS when doing the reading comprehension tasks after the intervention (English: 3.95, SD: .33; French: 4.06, SD: .41; Arabic: 4.01, SD: .39) than the control group (English: 1.88, SD: .19; French: 1.85, SD: .12; Arabic: 1.90, SD: .13). The results of the MANOVA test revealed no significant difference between the experimental group and the control group at the level of MCRS use before intervention: Wilks'  $\Lambda = .90$ ,  $F(3, 26) = .96$ ,  $p = .424$ , partial  $\eta^2 = .10$ . However, the MANOVA results yielded a significant difference between the two groups after the intervention: Pillai's Trace = .93,  $F(3, 26) = 124.59$ ,  $p < .001$ , partial  $\eta^2 = .93$ . Tables 1 and 2 summarize the results of the descriptive statistics.



Table 1. Means of the experimental group's MCRS use.

	English		French		Arabic	
	Means	S. D.	Means	S. D.	Means	S. D.
Before	1.89	.25	1.98	.29	2.01	.33
After	3.95	.33	4.06	.41	4.01	.39

Table 2. Means of the control group's MCRS use.

	English		French		Arabic	
	Means	S. D.	Means	S. D.	Means	S. D.
Pre-test	1.80	.18	1.83	.18	1.85	.12
Post-test	1.88	.19	1.85	.12	1.90	.13

## Discussion

Findings indicated a significant improvement in the experimental group's MCRS use in all the study languages. Regarding the use of MCRS, the experimental group in our study significantly obtained higher means of MCRS use when reading an academic text in English, French and Arabic than the control group. This is consistent with Talebi's (2012) results, which showed that the experimental group revealed a reverse transfer of reading strategies from L2 to L1 among Iranian learners of English (L2) and Persian (L1). Our finding aligns with previous research (Talebi's, 2012; Wu et al., 2021), which indicated that teaching reading strategies in L2 boosted students' reading strategy awareness in both L2 and L1. Accordingly, we argue that explicit instruction in MCRS in L3 can have a positive impact on trilingual learners' reading strategy awareness in all study languages after the intervention.

The study has pedagogical implications for language teaching/learning theory. The findings can provide language teachers with some insights into the importance of MCRS in students' reading activities. We recommend that explicit strategy instruction in MCRS can facilitate the learning process of languages and encourage learners to become autonomous. Also, a link between the different languages taught at school should be established by adopting a multilingual education approach. This way, skills and strategies learned in one language may transfer from one language to another.

One limitation of this study is that it was based only on a small sample with 10 participants in the control group and did not consider the effects of the intervention on the reading comprehension outcomes in all the study languages. Future research is needed to further investigate this area of research with a larger sample and special focus on the effects of L3 strategy training on reading strategy awareness levels and reading comprehension outcomes in L2 and L1.

## References

- Abu-Rabia, S. 2018. The effect of degrees of bilingualism on metacognitive linguistic skills. *International Journal of Bilingualism*, 23(5), 1064–1086. <https://doi.org/10.1177/1367006918781060>
- Abu-Rabia, S., Bluestein-Danon, D. 2012. A study into the results of an intervention program of linguistic skills in English (L2) and its Effect on Hebrew (L1) among poor readers: An examination of the Cognitive-Retroactive Transfer (CRT) hypothesis. *Open Journal of Modern Linguistics*, 02(04), 131–139. <https://doi.org/10.4236/ojml.2012.24017>
- Abu-Rabia, S., Shakkour, W., Siegel, L. 2013. Retroactive transfer ( CRT ) of language skills among bilingual Arabic- English readers. *Bilingual Research Journal: The Journal of the National Association for Bilingual Education Cognitive*, 36(1), 61–81. <https://doi.org/10.1080/15235882.2013.775975>
- Aghaie, R., Zhang, L. J. 2012. Effects of explicit instruction in cognitive and metacognitive reading strategies on Iranian EFL students' reading performance and strategy transfer. *Instructional Science*, 40(6), 1063–1081. <https://doi.org/10.1007/s11251-011-9202-5>
- Kong, A. 2006. Connections between L1 and L2 readings: Reading strategies used by four Chinese adult readers. *The Reading Matrix*, 6(2), 19–45.
- Perfetti, C.A., Landi, N., Oakhill, J. 2005. The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell Publishing. <https://doi.org/10.1002/9780470757642.ch13>
- Rabadi, R.I., Al-Muhaissen, B., Al-Bataineh, M. 2020. Metacognitive reading strategies use by English and French foreign language learners. *Jordan Journal of Modern Languages and Literatures*, 12(2), 243–262. <https://doi.org/10.47012/jjml.12.2.7>
- Talebi, S.H. 2012. Reading in L2 (English) and L1 (Persian): An investigation into reverse transfer of reading strategies. *English Language Teaching*, 5(3), 217–229. <https://doi.org/10.5539/elt.v5n3p217>
- Wu, L., Valcke, M., Van Keer, H. 2021. Supporting struggling readers at secondary school: an intervention of reading strategy instruction. *Reading and Writing*. <https://doi.org/10.1007/s11145-021-10144-7>

# Intonation of the Khakas fairytales

Tatiana Ryzhikova, Ilya Plotnikov, Anton Shamrin

Laboratory of Siberian and the Far Eastern Verbal Cultures, Institute of Philology SB RAS, Russia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

With the Khakas intonation contours having already been described on the material of colloquial speech, the purpose of the present paper is to verify their usage in the Khakas folklore. The data on the Khakas fairytales is found consistent with the previous studies. Both declarative and interrogative utterances have rising-falling tone and intensity, with difference in position of intensity peaks and distribution of tone changes. The texts include dialogues with reporting clauses, which are characterized by their own intonation patterns: in the beginning, the tone is rising, implying further continuation, while final remarks proceed with the general falling trend of the ending. This is consistent with the authors' previous studies on intonation of other Turkic languages.

Keywords: intonation, information structure, fairytales, Khakas language.

## Introduction

Khakas is one of the minority Turkic languages of Southern Siberia with the majority of its speakers living in the Khakas Republic (RF). It belongs to the Yugur group of the Eastern Khunnu branch of the Turkic languages. There are four Khakas dialects: Kachin, Sagay, Kyzyl, Shor. In this research, the intonation of literary Khakas (based on Kachin and Sagay dialects) is under consideration.

There are some previous investigations of the Khakas intonation made by U. Sh. Baitchura (1978) in 1950-60s and by K. N. Burnakova (Bicheldey) (2000; 2001) in 1980–2000s. While U. Sh. Baitchura studied intonation without paying special attention to the types of the utterances and their structure, K. N. Burnakova (Bicheldey) distinguished statements, different types of questions, exclamations and imperatives. She has managed to determine main intonation constructions (in line with the Russian intonation tradition). The work was performed on the material of the colloquial Khakas speech.

The purpose of the present paper is to verify the functioning of the established intonation constructions on the material of folklore. Khakas fairytales were taken from the series «Monuments of folklore of the Peoples of Siberia and the Far East» (2014), segmented into utterances with Audacity program (233 utterances in total) and further labeled and processed manually with PRAAT.

## Results and discussion

K.N. Burnakova (Bicheldey) has distinguished two main intonation patterns for declarative and interrogative utterances depending on their composition (whether they are one-component or two-component). One-component declarative and interrogative utterances are characterized by the same F0 movement (rising-falling), but differ in the position of intensity focus: in statements intensity peak coincides with the first tone rising, whereas in questions it can precede it. Two-component declarative utterances have a complex rising-falling-rising-falling intonation (Burnakova 1982).

In our research, the utterances were analyzed from two points of view: information structure and intonation. It turned out that the typical structure of a declarative utterance is theme (topic) – rheme (comment) with rising-falling tone movement and gradual declination of intensity (figure 1).

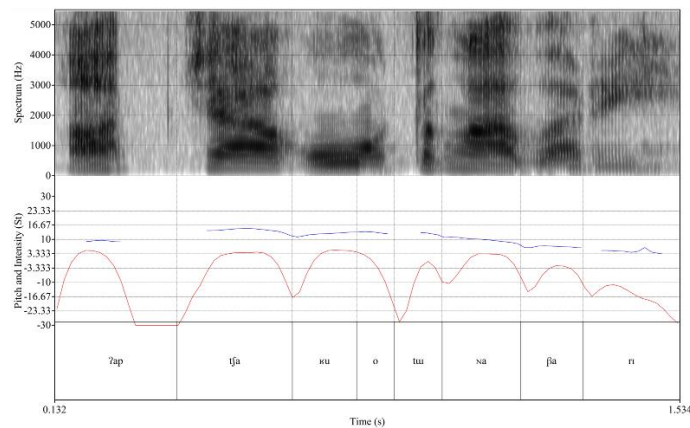


Figure 1. Spectrogram and intonogram (tone and intensity) of the Khakas utterance: *Apchagy odynga pari*. ‘The old man went for the firewood.’

The intonation of Khakas questions has been thoroughly investigated by K.N. Burnakova (2001; 2014). The modal (general, yes/no) questions in Khakas are formed by the interrogative particle *ma* (with variants *me*, *ba/be*, *pa/pe*), which is also typical for other Turkic languages. As attested by K.N. Burnakova, in colloquial Khakas such interrogative utterances are typically characterized by rising-falling intonation contour (Bicheldey 2001). In our data on Khakas folklore, F0 in questions is usually rather high with declination at the very end of the utterance (starting from the last syllable of the predicate) with intensity having a similar contour (figure 2). This is in contrast to declarative utterances, where the tone decreases more gradually.

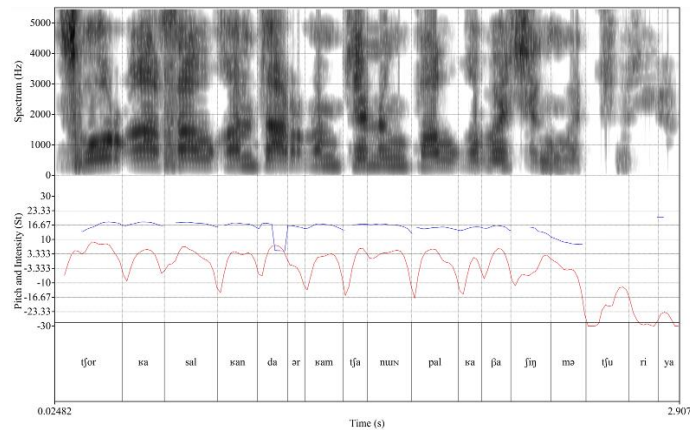


Figure 2. Spectrogram and intonogram (tone and intensity) of the Khakas interrogative utterance: «*Soorga salganda, argamngynang palgabangang ma?*» – *tipche*. ‘«Having put [it] in the sleigh, [why] don't [you] tie [it] with a rope?» – [the fox] says.’

In the fairytales under consideration there are also a lot of dialogues that are represented by the typical question–answer structure with author remarks. Reporting clauses are formed by such verbs as *tirge* ‘say, tell’, *chookhtirga* ‘say, tell’, *suringa* ‘ask’ and are often found both before and after the direct speech. When found at the beginning of an utterance, the tone is typically rising, which implies further continuation, in contrast to the ending, where the remarks are characterized by falling F0 and intensity (figure 3), denoting the end of a thought.

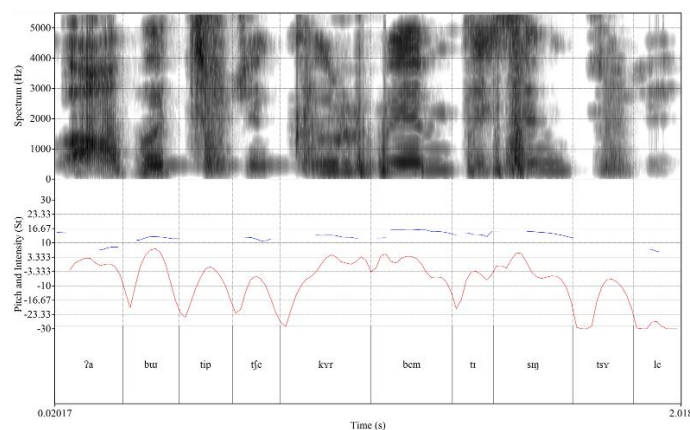


Figure 3. Spectrogram and intonogram (tone and intensity) of the Khakas utterance with direct speech: *Aba tipche: «Korbeem tizeng»* – *tipche*. ‘The bear says: «Haven't seen, say» – says.’

The idea that incompleteness of a thought correlates with the fundamental frequency rising while intensity remains rather high is consistent with our findings on intonation of the Tuvan folklore texts (in print) and T. Yanko's (1999) works on intonation of the Russian coherent text.

## Conclusion

The analysis of the Khakas fairytales showed that the intonation constructions distinguished by K.N. Burnakova (Bicheldey) for the colloquial speech are also used in folklore. Both in declarative and interrogative utterances, the intonation contour is rising-falling with different distribution of intensity and pitch movements. The idea of thought incompleteness being marked by rising tone has also been verified on the Khakas material and proved to coincide with our previous results for other Turkic languages.

## Acknowledgements

This research is carried out within the framework of the project of the Institute of Philology of the Siberian Branch of the Russian Academy of Sciences «Cultural universals of verbal traditions of the peoples of Siberia and the Far East: folklore, literature, language» supported by a grant from the Government of the Russian Federation for the promotion of research conducted under the guidance of leading scientists, contract № 075-15-2019-1884.

## References

- Baichura, U. 1978. Instrumental-phonetic data on the word stress and intonation in Khakas and Buryat. *Central Asiatic Journal* XX (1-2), 3-31.
- Bicheldey, K.N. 2000. Rhythmomelodems of simple simple sentences of the Khakass language. Moscow: RUDN Publishing house. 115 p.
- Bicheldey, K.N. 2001. Lexico-grammatical and rhythmomelodic means of expressing interrogativity in the Khakass language. Moscow: RUDN Publishing House. 282 p.
- Burnakova, K.N. 1982. Rhythmomelodems of the Khakas simple non-extended sentences: avtoref. ... Cand. Philol. Sci. Alma-Ata. 17 p.
- Burnakova, K.N. 2014. On the rhythmic-melodic investigation of the imperative-interrogative sentences (on the material of the Khakas language). *Russian Turkology* 1 (10), 13-22.
- Monuments of Folklore of the Peoples of Siberia and the Far East (2014) Novosibirsk: Nauka, 33, The Khakass Folktales. 770 p.
- Yanko, T.E. 1999. On the notions of the communicative structure and communicative strategy. *Topics in the study of language* 4, 28–55.

# Laryngeal activity in Barabian and the Surgut dialect of Khanty

Tatiana Ryzhikova<sup>1</sup>, Ilya Plotnikov<sup>2</sup>, Nikolay Urtegeshev<sup>1</sup>

<sup>1</sup>Laboratory of Experimental-Phonetic Researches, Institute of Philology SB RAS, Russia

<sup>2</sup>Department of Siberian and Far-Eastern Languages, Institute of Philology SB RAS, Russia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The paper describes a laryngeal activity in two Siberian languages of different genealogical belonging: Baraba-Tatar (Turkic) and the Surgut dialect of Khanty (Ob-Ugrian). The method of direct digital laryngoscopy has been used accompanied by acoustic data analysis. The results show that larynx actively participates in the production of the vowels in the languages under consideration. The data obtained are supported by previous MRI investigations.

Keywords: Baraba-Tatar language, Khanty language, laryngoscopy, larynx, vowels

## Introduction

The contribution of larynx into speech production process has long interested phoneticians but remained for decades poorly understood. The modern techniques such as, e.g., digital laryngoscopy, allows scientists to clarify the activity taking part in the lower part of pharyngeal cavity (tube) (see, e.g., Esling 2012; Moisiuk et al. 2010; Moisiuk et al. 2012). In the Siberian languages larynx plays an important role in the production of both consonants and vowels. In some of languages (e.g. in Shor), laryngeal activity is a constitutive-differential feature within the consonant system: consonants are classified as ejective and injective units. In others, which include Barabian and Khanty, the status of the larynx contribution into articulatory process is not quite established. The purpose of the paper is to investigate the laryngoscopic data on vowels in the idioms under consideration and to find out common and specific traits of the laryngeal activity in both of them.

Barabian is the Turkic language of the indigenous people of the Novosibirsk region, whereas Khanty is the Ob-Ugrian language of the aboriginal peoples of Khanty-Mansi and Yamalo-Nenets regions of the Russian Federation. However different may the languages and peoples seem at first sight, they do appear to have something in common. It has been established that the ancestors of the Baraba-Tatars had been Ob-Ugrian tribes that were turkified by the Turkic tribes coming from Southern-Eastern Asia into the territory of Siberia in the Middle Ages. The process of turkification was global and the majority of the

Southern Siberian Turkic peoples were formed during that period. Though some predecessors of the contemporary Siberian Turkic peoples merged with the Turks forming new tribes, some native inhabitants of the Siberian region moved northward preserving their languages and culture, and it seems to be the situation with the Khanty. Thus, the Baraba-Tatars and the Khanty people might have had some common ancestors. Similarities are found in culture, traditions and may also be present in language.

## Methods

The laryngoscopic material was obtained from the native speakers of these two idioms in the Medical centre with Pentax bronchoscope FB-18V (Pentax, Japan) and the OLYMPUS BF-3C40 bronchofibroscope (Olympus, Japan). Simultaneous video and audio recordings were made with EVIS EXERA II Video System Center Olympus CV-180 and AVER Media program. A special technique for data interpretation has been developed, but in this study we mainly used visual control of the articulatory process to determine the laryngeal role in the vowel production. The analysis of the Barabian and Surgut sound files was performed in PRAAT.

## Results and discussion

Previous studies of the vocal systems of the Barabian and Khanty languages have shown that their systems differ a lot: in Khanty (the Surgut dialect) there is an opposition of long and short (reduced) vowels (Honti 1977, Csepregi 1998, etc.), whereas the Barabian vocalism seems to be rather simple: there are only short units, the status of diphthongs is not clear (Urtegeshev et al. 2003). The so called intermittent vowels (usually long units with the glottal or laryngeal insertion in the middle of articulation (Urtegeshev, Koshkareva 2018)) have been identified in both idioms; in the Surgut dialect they have even been distinguished as a special sub-class (Urtegeshev, Koshkareva 2018).

Our previous study of the Surgut intermittent vowels by the digital laryngoscopy (Ryzhikova et al. 2021) has shown that larynx is actively engaged in the production of this type of sounds: they can be classified as injective (when larynx lowers and glottal resonator narrows) or as ejective (larynx moves upward and glottal cavity becomes wider) and are usually accompanied by changes in F0 and intensity during the vowel production (Fig. 1).

The investigations of the Baraba-Tatar vowels by magnetic-resonance imaging (MRI) have suggested that some sounds are pronounced with the ejected larynx whereas others are articulated with the injected larynx (Ryzhikova 2021). In the Barabian word-form *mitch* ‘an oven’ a rise of the larynx has been fixed both on the MRI and laryngoscopic images (Fig. 2). There can also be noticed an active work of the epiglottis moving to and fro causing additional acoustic effect of noise. One can see this noise on the waveform, but there is no change in F0 movement.



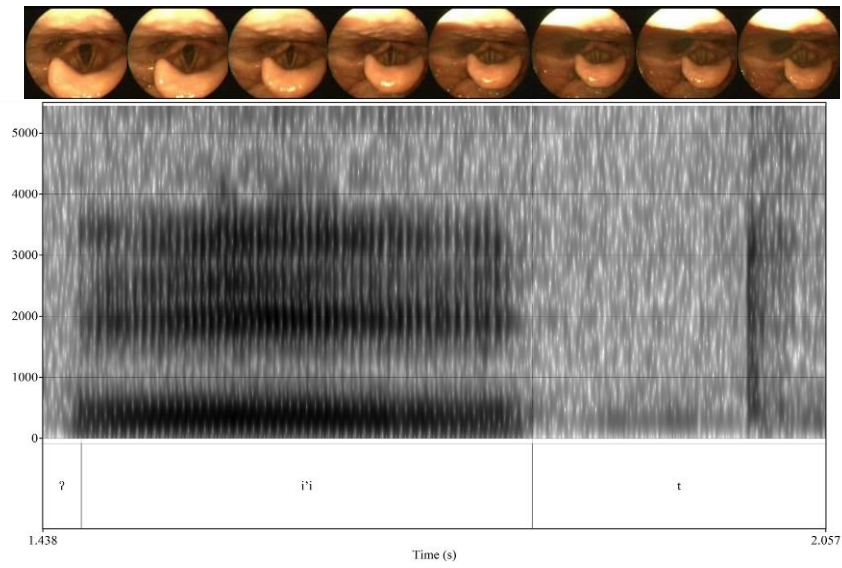


Figure 1. Laryngogram and spectrogram of the Surgut vowel *i'i* in the wordform *it* 'now'.

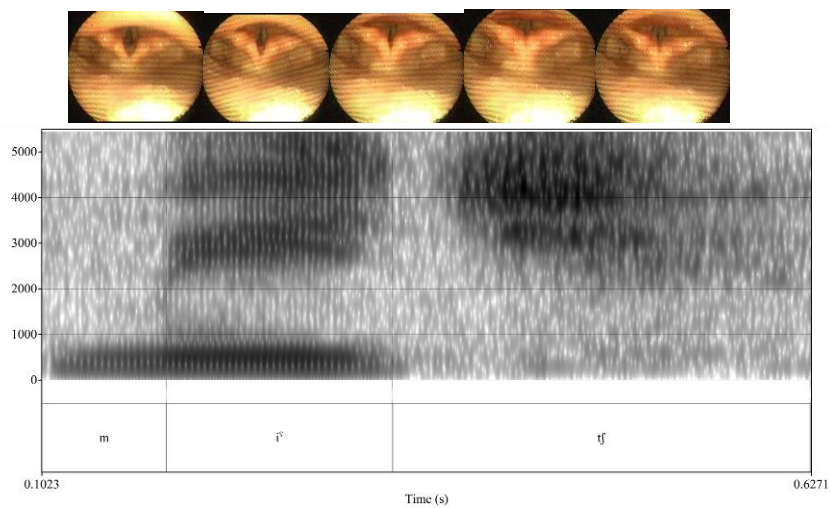


Figure 2. Laryngogram and spectrogram of the Barabian vowel *i'* in the wordform *mitch* 'an oven'.

## Conclusions

The experimental results show that in both idioms larynx actively participates in vowel production, its movement depending on the type of a vowel. When pronouncing the so-called intermittent sounds, the larynx moves down, whereas while producing long vowels in Surgut it moves upward. The short sounds are articulated with the neutral position of larynx. Epiglottis also moves actively, preventing the direct airflow from the lungs. Such complicated laryngeal work correlates with acoustic effect of tone changes: whether falling, rising, or level. In Baraba-Tatar there also exist injected and ejected vowels showing that larynx also actively participates in the vowel production, but the correlation with the tone changes has not been stated. Generally, laryngeal activity is very important for the Siberian languages of different origin.

## Acknowledgements

The research is supported by the Russian Foundation for Basic Research, the project № 19-012-00388 «Specifics of Surgut Khanty Phonetic System»)

## References

- Csepregi, M. 1998. Szurguti osztják chrestomathia. *Studia uralo-altaica supplementum*, 1-183, Szeged.
- Esling, J. 2012. The Articulatory Function of the Larynx and the Origins of Speech. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 38.
- Honti, L. 1977. Beobachtungen über die Laut- und Formenlehre gegenwärtiger Surguter Mundarten des Osjakischen. *Acta Linguistica Academiae Scientiarum Hungaricae* 27 (3 / 4), 271-286.
- Moisik, S., Czaykowska-Higgins, E., Esling, J. 2012. The Epilaryngeal Articulator: A New Conceptual Tool for Understanding Lingual-Laryngeal Contrasts. *McGill Working Papers in Linguistics*, 22.1, Winter 2012.
- Moisik, S., Lin, H., Esling, J. 2010. An investigation of laryngeal behavior during Mandarin tone production using simultaneous laryngoscopy and laryngeal ultrasound. *Proceedings of The 9th Phonetics Conference of China*.
- Ryzhikova, T. 2021. Articulatory peculiarities of the Baraba-Tatar phoneme *o* /*ö*/ (on MRI data). *Siberian Journal of Philology*, 1, 191-208.
- Ryzhikova, T., Dobrinina, A., Timkin, T. 2021. Study of the intermittent vowels of the Surgut dialect of the Khanty language through the method of direct digital laryngoscopy: preliminary results. *Bulletin of Ugric Studies*, 2021; 11 (1), 102-111.
- Urtegeshev, N., Selyutina, I., Ryzhikova, T., Vildanov, A. 2003. The Baraba-Tatar language. *Languages of the indigenous peoples of Siberia*, 10, 78-100, Novosibirsk.
- Urtegeshev, N., Koshkareva, N. 2018. An inventory of intermittent vowel phonemes in the Surgut dialect of the Khanty language. *Siberian Journal of Philology*, 1, 180-207.

# HAT - A new corpus for experimental stylometric evaluation in Arabic

Halim Sayoud, Siham Ouamour

EDT, FEI, USTHB University, Algeria

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Stylometry is a research field of NLP dealing with the identification of the actual author of a piece of text. Even though it does exist several datasets for some occidental languages, it is quite difficult to find an adequate corpus in Arabic. This problem motivated us to build a natural Arabic corpus dedicated to the task of experimental stylometric evaluation. Our corpus is composed of 100 groups of Arabic texts that are extracted from different Arabic books, where the main topic is “Travel”. The books are written by 100 different authors and each group contains 3 different texts that are written by the same author. We called this corpus "HAT" (Hundred of Arabic Travelers). Furthermore, we propose it as a free corpus that may represent a reference dataset for author style analysis in Arabic, which could be used for a purpose of experimental evaluation. To evaluate our dataset, we conducted some baseline experiments using an SVM classifier with a Leave-One-Out cross validation technique.

Keywords: natural language processing, computational linguistics, author style analysis; stylometry, digital libraries

## Introduction

As per definition, the task of author recognition can be divided into several fields that are:

- Authorship attribution (AA) or identification: it consists in identifying the author(s) of a set of different texts;
- Authorship verification: in this case, the main task is checking whether a piece of text is written or not by an author who claimed to be the writer;
- Authorship discrimination: it consists in checking if two different texts are written by a same author or not (Sayoud 2012);
- Plagiarism detection: in this research field we look for the sentences or paragraphs that are taken from another author (Küppers 2012);
- Text indexing and segmentation: the main goal is to segment the global text into homogeneous segments (each segment or paragraph contains the contribution of only one author) by giving the name of the appropriate author in each text segment (paragraph) (Forest 2006).

In this paper, a new text dataset is proposed to the scientific community for the experiments of authorship attribution in Arabic.

Although several works and datasets were cited for the English (Küppers 2012) (Juola 2006) and Greek (Tambouratzis 2003) (Tambouratzis 2004) languages, the authors did not find a lot of corpora in Arabic.

The proposed corpus contains several texts written by 100 Arabic authors on the topic of travel. The authors wrote several documents describing their travels. So a specific Arabic corpus/baseline has been built for a purpose of comparative authorship attribution.

## HAT corpus

Our textual corpus is composed of 100 groups of Arabic texts that are extracted from 100 different Arabic books. The books are written by 100 different authors and each group contains 3 different texts that are written by the same author, which means that each group belongs to only one author. This set of 300 text documents has been collected in 2019 from “Alwaraq digital library”; we called it HAT corpus (i.e. Hundred of Arabic Travelers). Furthermore, this corpus could represent a reference dataset for author style analysis in Arabic, which could be used by researchers in this field for a purpose of comparative evaluation. For concreteness, here is a piece of text belonging to Author #92 (figure 1).

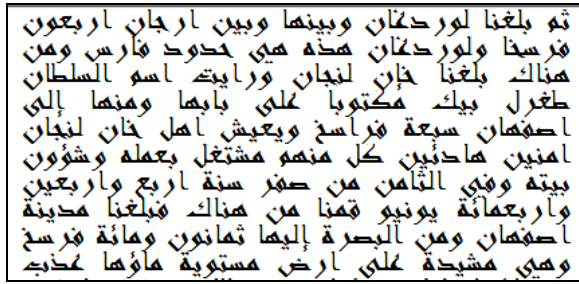


Figure 1. Example of Arabic text belonging to Author #92 (N. *Khasru*).

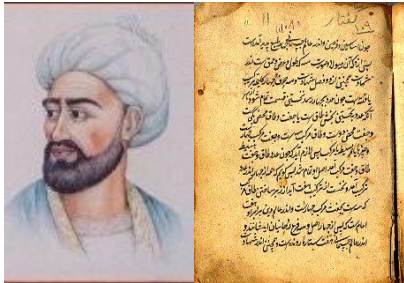


Figure 2. Ancient portrait and sheet of paper containing a text of Author #92.

The texts have a medium/short size: the average text length is about 1100 words per document and there are 3 documents per author, which corresponds to 300 documents in the total corpus. This situation involves severe experimental conditions, since it has been shown in previous research works (Eder 2010) that the minimum number of words per text should be at least 2500 words to get good attribution performances. In this investigation, the use of relatively short texts is interesting in order to evaluate the different classifiers with small documents in Arabic. In fact, when short texts are used, the AA performances decrease and it becomes difficult to make an efficient identification.

## Experiments

In order to set a baseline evaluation for our new corpus, some experiments of author style analysis have been conducted by using an SMO-SVM classifier and by employing character-N-grams as features. Furthermore, the different experiments are made using a Leave-One-Out (LOO) cross-validation for a purpose of significance and consistency. The results of this baseline evaluation are summarized in table 1.

As one can see in table 1, the authorship attribution accuracy tends to increase by increasing the size of the N-gram: hence, the best accuracy was obtained by characters 4-grams (0.95), which is higher than the accuracy of 3-grams (0.81). This last one is almost similar to the accuracy of characters 2-grams (0.83), which is again higher than the accuracy of character 1-gram (0.71).

Table 1. Baseline results of authorship attribution on the HAT corpus.

Feature	LOO Accuracy
Characters	0.71
Char bi-grams	0.83
Char tri-grams	0.81
Char tetra-grams	0.95

In particular, with character tetra-grams, the performances revealed by the LOO validation show that 95% of the texts are well attributed, which represents a good score since we deal with a large number of authors: 100 authors.

## Discussion

In this research work, we proposed a new corpus for authorship attribution, where we conducted a baseline experimental evaluation.

During this evaluation, we noticed that the best accuracy was obtained by character-tetra-grams. Right now, the first results appear interesting even though the investigation is not completely finished yet.

On the other hand, the following problems were noticed:

- The text documents are not sufficiently long to get a fair classification, as reported by previous works of Eder (Eder 2010);
- The noisy nature of those travel notes makes the task more complicated, since we did not apply any cleaning preprocessing;
- The Arabic language is very specific since it tends to employ very common introductory sentences that are commonly used by authors (eg. *Albasmala*).

However, the HAT corpus has the advantage to be quite large (100 authors) and the advantage to respect all the conditions required in stylometry. Several experiments of authorship identification were tested and the reported results have shown that one can reach an accuracy of 0.95. Since we did not reach the score of 100%, the HAT-corpus remains still interesting in order to try getting better results by the competing researchers. Finally, one of our objectives is to make this corpus a reference in experimental stylometric evaluation in Arabic.

## References

- Eder M. 2010. Does size matter? : autorship attribution, short samples, big problem. In Digital humanities 2010 conference, pp. 132-135, London.
- Forest D. 2006. Application de Techniques de Forage de Textes de Nature Prédicative et Exploratoire à des Fins de Gestion et d'Analyse Thématique de Documents Textuelles Non Structurés. PhD, Université du Québec à Montréal.
- Juola P. 2006. Authorship attribution. Foundations and Trends in Information Retrieval, pp. 233–334, Now publishing.
- Küppers R., Conrad S. 2012. A Set-Based Approach to Plagiarism Detection. PAN 2012 Lab Uncovering Plagiarism, Authorship, and Social Software Misuse held in conjunction with the CLEF 2012, 17-20 September, Rome, Italy.
- Sayoud H. 2012. Author Discrimination between the Holy Quran and Prophet's Statements. Literary and Linguistic Computing, Volume 27 Issue 4, pp.427-444.
- Tambouratzis G., Markantonatou S., Hairetakis N., Vassiliou M., Carayannis G., Tambouratzis D. 2004. Discriminating the Registers and Styles in the Modern Greek Language-Part 1. Literary & Linguistic Computing, 19(2), 197-220.

# Affective norms for Chinese characters

Thomas Schlatter

Graduate Institute of Linguistics, National Taiwan University, Taiwan

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

A database has been generated which constitutes affective ratings for single Chinese characters, viz. the Affective Norms for Chinese Characters (ANCC). This database enables researchers who study the lexico-semantic properties of Chinese characters to necessitate affective and other psycholinguistic properties in a manner that is independent and unbiased. Previous ratings for single Chinese characters, although extensive, have omitted affective properties such as valence and arousal. These factors are known to significantly influence the performance of participants in word recognition and production tasks. Close examination of the data in this study shows that affective and other psycholinguistic aspects of meaning explain a significant portion of the outcomes of previous experiments conducted on Chinese characters.

Key words: Affective Norms, Chinese Characters, Valence, Arousal, Dominance

## Introduction

The Affective Norms for Chinese Characters (ANCC) is currently the largest available dataset on single Chinese characters. This database was compiled with the goal of providing affective aspects of meaning for an extensive list of single Chinese characters, in addition to already existing dimensions such as concreteness (abstract vs. concrete) and imageability (un-imageable vs. imageable) (Liu, Shu, and Li 2007). These aspects include the affective meaning aspects of valence (positive vs. negative), arousal (intense vs. calm), dominance (dominating vs. dominated), origin (emotional vs. rational) and significance (important vs. unimportant).

Rather than dealing with words, we deal with the smallest units of meaning, which in the case of Chinese, are comprised of syllables or their visual representation, viz. characters. Chinese characters are a crucial part of the Chinese language that have been shaped by Chinese culture. For example, the Chinese character for “good”, 好, is composed of the radicals 女 (woman) and 子 (child), representing the ancient Chinese idea that it is good for a woman to have a child.

## Materials and methods

The assessment was conducted using seven Self-Assessment Manikin (SAM) scales adapted from Imbir (2016). In order to make the assessments more easily accessible to the participants, data was collected in the form of an online

survey, which could be completed by using either a computer or a mobile phone. Each word was presented in isolation and accompanied by either a combination of valence-arousal-dominance or concreteness-imageability-significance-origin scales. The Chinese characters were separated into groups of 100. Both character order and the affective or psycholinguistic dimensions were randomized. The participants were first introduced to the aim of the study followed by an instruction on how to complete the survey. First, data on age, sex, and education were collected. The SAM scales were then provided, accompanied by a brief explanation of each variable. Examples were avoided to prevent bias.

### **Participants**

Each single character in the list was rated by at least 30 participants. The participants were sourced from the Upwork network, using their native tongue (Chinese) and place of origin (China) as defining characteristics for their participation. All participants were given a one-week period to complete the online survey in a self-paced manner, with remuneration provided for their time.

The choice of the respective Chinese characters was based on frequency (Da 2004). In addition to the 3450 most frequent characters, this study added 149 gendered characters to the dataset, those of which can be found within the word list cited above. These words all contained either a 人 (person) or 女 (woman) radical. Consequently, the final list contained 3599 Chinese characters.

### **Reliability and data analysis**

To assess the reliability of the data, a comparison is made between parts of the dataset in this study and an earlier dataset. This included measurements for the dimensions concreteness and imageability. Correlations between the data in this study and the data collected by Liu (2007) ( $n=2172$ ) were strong, at 0.6 and 0.71 for imageability and concreteness, respectively.

The two models presented here include the dimensions of the ANCC in addition to the frequency (Da 2004) as the independent variables. The reaction times comprised the dependent variables. Both models were statistically significant for character recognition ( $F(1,2243)=1101.049$ ,  $p<0.001$ ,  $R^2= 0.39$ ) and character naming ( $F(1,1950)=650.688$ ,  $p<0.001$ ,  $R^2= 0.33$ ). The first regression showed that the dimensions of valence, arousal and imageability were all negatively correlated with reaction time, while significance was positively correlated with reaction times in a character recognition task. The second regression showed that the dimensions of valence, arousal, significance, imageability and concreteness were negatively correlated with reaction times, while significance is positively correlated with reaction times in both tasks. For the naming task, origin was positively correlated with reaction times. Compared to models that include frequency as the only independent variable, the complete



models, including all affective and other psycholinguistic dimensions, were able to explain a further 4% ( $\Delta R^2=0.04$  for character recognition) or 5% ( $\Delta R^2=0.05$  for character naming) of the variance in the dependent outcomes.

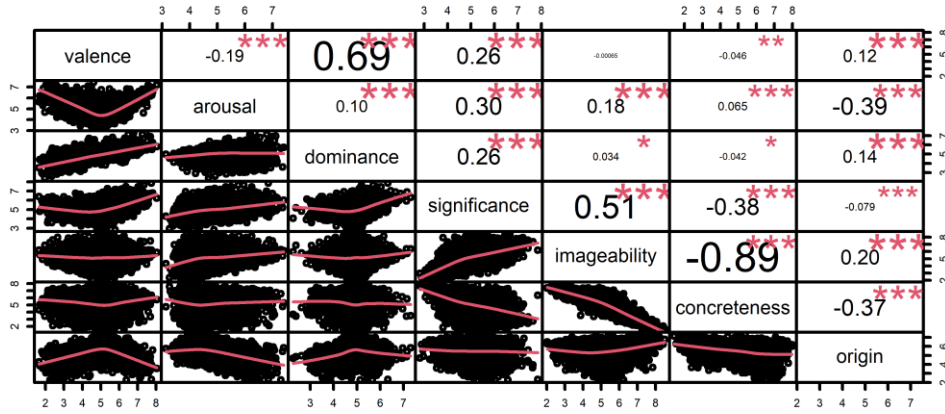


Figure 1. Correlation matrix showing the inter-relationships among the dimensions of the ANCC. The lower triangular matrix is composed by the bivariate scatter plots with a fitted smooth line. The upper triangular matrix shows the Pearson correlation plus significance level. Each significance level is associated to a symbol: p-values 0.001 (\*\*\*), 0.01 (\*\*), 0.05 (\*).

Table 1. Summary outputs of multiple regression analyses with character recognition and naming reaction times as the dependent variables. The IVs were the dimensions of the ANCC, along with frequency.

<i>Predictors</i>	<b>Recognition.RT (Liu 2007)</b>			<b>Naming.RT (Chang 2016)</b>		
	<i>Est.</i>	<i>CI</i>	<i>p-value</i>	<i>Est.</i>	<i>CI</i>	<i>p-value</i>
(Intercept)	993.27	941.77 – 1044.76	<0.001	1100.28	1032.84 – 1167.72	<0.001
frequency [log]	-24.94	-26.42 – -23.47	<0.001	-27.63	-29.75 – -25.51	<0.001
valence	-8.08	-11.18 – -4.98	<0.001	-8.65	-13.05 – -4.24	<0.001
arousal	-6.78	-10.64 – -2.92	0.001	-11.01	-16.54 – -5.47	<0.001
dominance	4.69	-0.94 – 10.33	0.103	6.74	-1.22 – 14.70	0.097
significance	8.57	4.63 – 12.50	<0.001	13.03	7.52 – 18.55	<0.001
imageability	-19.85	-24.54 – -15.15	<0.001	-21.90	-28.20 – -15.60	<0.001
concreteness	-9.40	-13.42 – -5.37	<0.001	-6.15	-11.57 – -0.73	0.026
origin	-1.07	-4.20 – 2.06	0.503	4.82	0.45 – 9.19	0.031
Observations		2252			1959	
R <sup>2</sup> / R <sup>2</sup> adj.		0.393 / 0.391			0.332 / 0.330	

## Discussion

In conclusion, the collection of emotion and other psycholinguistic norms for 3599 Chinese characters could provide researchers with a corpus that is both independent and unbiased. The ratings presented here were highly correlated with earlier studies and able to explain a significant part of the deviation in both character naming and character recognition studies.

## References

- Chang, Y.-N., Hsu, C.-H., Tsai, J.-L., Chen, C.-L., Lee, C.-Y. 2016. A psycholinguistic database for traditional Chinese character naming. *Behavior Research Methods* 48 (1): 112–22.
- Da, J. 2004. A Corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In *Proceedings of the Fourth International Conference on New Technologies in Teaching and Learning Chinese*, 501–11. Tsinghua University Press Beijing.
- Imbir, K. 2016. Affective norms for 4900 Polish Words Reload (ANPW\_r): Assessments for valence, arousal, dominance, origin, significance, concreteness, imageability and, age of acquisition. *Frontiers in Psychology* 7: 1081.
- Liu, Y., Shu H., Li, P. 2007. Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods* 39 (2): 192–98.
- Sze, W. P., Liow, S. R., Yap, M. J. 2014. The Chinese lexicon project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods* 46 (1): 263–73.

# Effects of agent position and orientation on perception and production

Judith Schlenter<sup>1</sup>, Yulia Esaulova<sup>1</sup>, Barbara Zeyer<sup>1</sup>, Martina Penke<sup>1</sup>

<sup>1</sup>Department of Special Education and Rehabilitation, University of Cologne, Germany

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

In a series of aesthetic judgement tasks, we found that speakers of German display a spatial agency bias if, and only if, a scene shows an agent performing an action in the direction of a patient. The experiments reported here replicate and extend previous findings, indicating that the position of the agent relative to the patient affects how speakers perceive a depicted event. Moreover, the experiments are the first to show that the orientation of the agent and patient (toward vs. away from other event character) is another modulating factor affecting scene perception as well as scene description.

Keywords: spatial agency bias, German, scene perception, language production

## Introduction

A primary goal during language comprehension is to find out who did what to whom. Thus, it is not surprising that it is one of the most well-researched topics in experimental linguistics. For example, researchers have investigated for a range of languages which information sources listeners use to identify the agent. Interestingly, non-verbal tasks reveal that people draw on their linguistic knowledge (word order specifics) and reading/writing habits even when perceiving or envisaging an event in which an agent acts upon a patient (e.g., Esaulova et al., 2021; Suitner et al., 2021).

Only recently, two studies reported a spatial agency bias (SAB) for German speakers. Suitner et al. (2021) found that Germans preferentially drew a given action between two characters evolving from left to right, in accordance with the reading/writing direction of their script and the linearization of subject and object in a canonical active sentence. Using an aesthetic judgement task, Esaulova et al. (2021) showed that German speakers preferred scenes with left-positioned agents over scenes with right-positioned agents, unlike a group of Arabic speakers (right-to-left script, subject-before-object order), who displayed the reverse bias. Moreover, German speakers were slower to describe scenes with right-positioned agents than scenes with left-positioned agents. Thus, a violation of the SAB resulted in longer speech planning times. Taken together, for literate speakers of German the left position in an event scene is typically associated with the position of the agent, so the mental representation of the event (from left to right) is in alignment with a linguistically preferred structure (subject/agent-verb-object/patient).

Other visual factors, besides agent position, have been found to affect agent identification. In an experiment by Dobel et al. (2007), German speakers had to identify the actions, actants, and objects in a scene after brief exposure (100 to 300 milliseconds) and had to judge whether the scene was meaningful or not. Scenes either showed a transitive or ditransitive event. Correct identification of the agent was affected by the position of the agent in the scene, in line with the SAB reported for German, but also by scene coherence. Recognition of the agent was hindered if the scene was incoherent and showed no interaction between characters, who were both turned away from each other. In the current paper, we report on a series of aesthetic judgements tasks, which investigated how agent position (left vs. right) and orientation of agent and patient (toward vs. away from other event character) affected German speakers' perception of transitive event scenes. As an outlook, we discuss recent findings from a scene description task investigating both of these factors in one experiment.

## Methodology

In four aesthetic judgement tasks, we presented adult speakers of German with two scenes that only differed in one characteristic, see Figure 1. For each contrast, they had to select one out of three response options: preference for (1) the left picture, (2) the right picture, or (3) no preference. Participants were instructed to judge which scene was more typical, natural, or better than the other. Thus, this task assessed speakers' visual preferences without requiring a motor activity or providing a verbal description that might bias participants' decision. The ordering of the scenes in each contrast, that is, which one was displayed left or right, varied.

In the first three experiments, we used the same scenes as Esaulova et al. (2021), whose data we reanalysed and henceforth refer to as Exp. 1. All scenes showed an action between two human characters. There were eight events in total. In Exp. 1 ( $N = 36$ ), the position of the agent was manipulated, while agent and patient both looked in the same direction, that is, rightwards or leftwards. In Exp. 2 ( $N = 22$ ), the position of the agent was manipulated, while agent and patient both faced each other. In Exp. 3 ( $N = 22$ ), the agent was always displayed left, but the orientation of agent and patient was manipulated, so in one scene both agent and patient looked rightwards and in the other scene both were facing each other. In Exp. 4 ( $N = 44$ ), four events were selected that most clearly depicted an activity that, however, was not directed towards another character (incoherent scenes). The four scenes appeared twice with different characters while the position of the agent was manipulated. All participants only participated in one of the experiments.

## Effects of agent position and orientation on perception and production 3

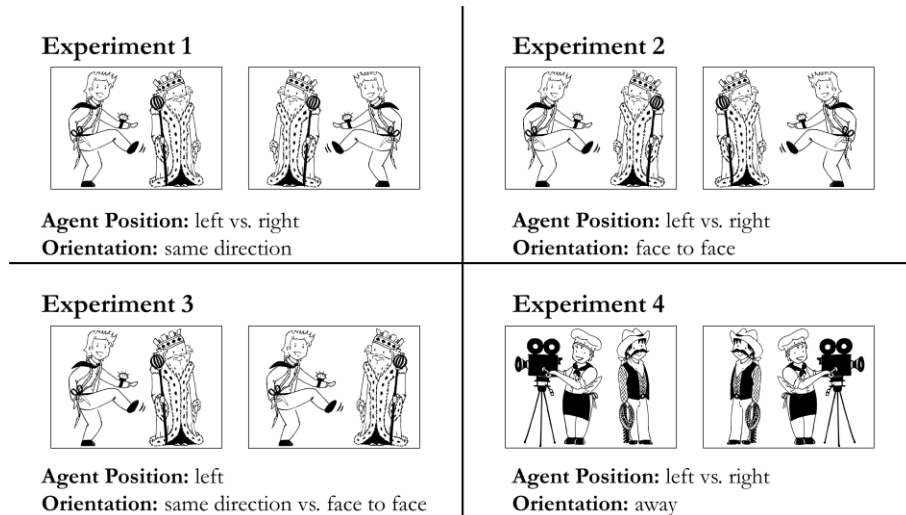


Figure 1. Experimental designs in the aesthetic judgement tasks.

## Results

Figure 2 shows the proportion of selected responses per event in each of the four experiments. For statistical analyses, we coded speakers' selections as Agent Left response (Exp. 1, Exp. 2, Exp. 4) or not, or Agent facing Patient response (Exp. 3) or not. Generalized linear mixed effects models tested whether these responses were selected above chance. Chi-Square goodness of fit tests were used to assess speakers' preferences for single events.

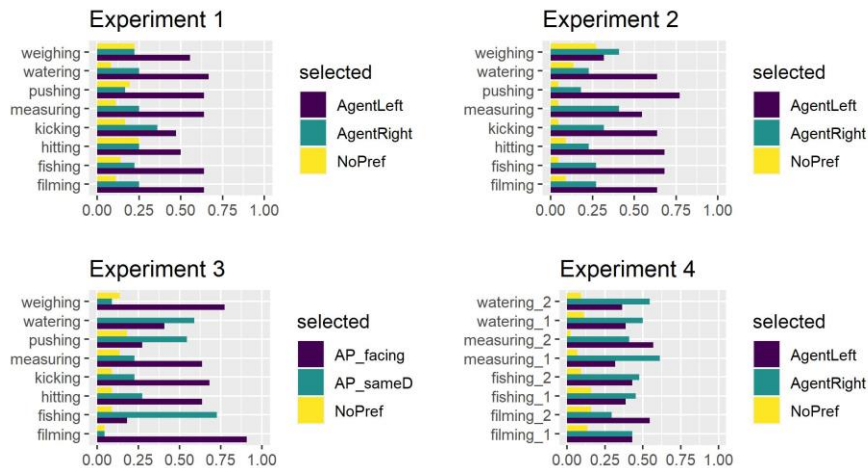


Figure 2. Bar charts showing the proportion of selected responses.

In Exp. 1 and Exp. 2, speakers were more likely to select the scene with a left-positioned agent (across experiments, Est. = 0.57, SE = 0.25,  $z = 2.31$ ,  $p = 0.02$ ) than the other two response options. No such preference was observed in Exp. 4. In Exp. 3, we observed a preference for face-to-face scenes for some (filming, hitting, kicking, measuring, weighing), but not all scenes.

## Discussion

Consistent with previous findings, our results show that German speakers display a preference for left-positioned agents. Crucially however, we could demonstrate that the SAB only holds for scenes in which the agent performs an action in the direction of the patient. Incoherent scenes do not evoke the SAB. With respect to orientation, we observed an overall preference for scenes where agent and patient face each other. However, this preference seems to be more action specific, that is, for some events (e.g., filming), the scene with both characters facing each other appears to be more meaningful than for others (e.g., pushing). In a recent scene description task we conducted, we found that the orientation of agent and patient towards each other also affects speech-production latencies (Schlenter & Penke, in prep.). Face-to-face scenes required more time for sentence planning than scenes where agent and patient were oriented in the same direction. Moreover, speakers produced more passives for face-to-face scenes compared to same-orientation scenes. Both findings suggest that a patient that faces the agent is less prototypical and, hence, more prominent. To conclude, agent position and orientation in a scene can affect perception and how the scene is described. They should, hence, be taken into account when designing materials for psycholinguistic experiments.

## Acknowledgements

This research was supported by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252.

## References

- Dobel, C., Gummior, H., Bölte, J., Zwitserlood, P. 2007. Describing scenes hardly seen. *Acta Psychologica* 125, 129-143.
- Esaulova, Y., Dolscheid, S., Reuters, S., Penke, M. 2021. The alignment of agent-first preferences with visual event representations: Contrasting German and Arabic. *Journal of Psycholinguistic Research* 50, 843-861.
- Schlenter, J., Penke, M. (in prep.). (How) Visual properties affect the perception and description of transitive events.
- Suitner, C., Maass, A., Navarrete, E., Formanowicz, M., Bratanova, B. Cervone, C., Carrier, A. 2021. Spatial agency bias and word order flexibility: A comparison between 14 European languages. *Applied Psycholinguistics* 42, 657-671.

# Effect of face mask and noise on word recognition by children and adults

Anna Sfakianaki<sup>1</sup>, George P. Kafentzis<sup>1</sup>, Devora Kiagiadaki<sup>1</sup>, George Vlahavas<sup>2</sup>

<sup>1</sup>Speech Signal Processing Laboratory, Computer Science Department, University of Crete

<sup>2</sup>Department of Informatics, Aristotle University of Thessaloniki

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Previous studies have shown that noise has a detrimental effect on speech intelligibility and language comprehension, more so for children than adults. Furthermore, the mandatory use of face masks due to the COVID-19 pandemic presents an additional communication and learning barrier. The current study investigates the effect of face mask on low frequency word recognition in quiet and noise by adults and, more importantly, by young children who just completed the first grade of primary school. Preliminary results indicate that word identification is significantly compromised when produced with a surgical face mask, with more pronounced negative effects for children listening in noise. Implications on new word recognition and learning by first graders in noisy classrooms with compulsory mask usage are considered.

Keywords: face masks, noise, word recognition, children, adults

## Introduction

A lot of research has been conducted on the speech perception abilities of children vs adults in adverse listening conditions (for a review, see Leibold & Buss, 2019), as understanding speech in noise is an everyday challenge for children in school settings and has a negative impact on learning (Erickson & Newman, 2017). There is consensus that children face greater difficulty recognising and understanding speech in background noise relative to adults, with competing speech posing a greater obstacle than white or speech-shaped noise (Corbin et al., 2016). Apart from noise, the use of face masks that has been mandated in educational institutions and other indoor spaces, presents an added barrier to speech recognition and communication (Toscano & Toscano, 2021).

Although the effect of noise and face mask on speech perception has been investigated for adults, to the best of our knowledge, there are no studies regarding children. Thus, the current paper investigates the influence of mask and noise on word recognition in adults and typically developing primary school children with normal hearing. Low frequency words were selected, as later acquired words have been shown to provide greater insight between linguistic knowledge and masked speech recognition (Leibold & Buss, 2019).

## Methodology

Twenty six listeners, fourteen adults and twelve children were recruited for the study. Adults were aged 19 to 51 years old and children were aged 6;8 to 7;6 (years;months) and had just completed the first grade of primary school. All participants underwent formal audiological testing to ensure normal hearing, and two adults were excluded due to moderate hearing loss at high frequencies. All subjects had no reported speech, language and learning difficulties. Two adults and two children took part in a pilot experiment so as to choose a suitable SNR (signal-to-noise ratio) level for each type of noise and age group. Ten adults and ten children participated in the main experiment. All testing took place in the sound proof booth of the Speech Signal Processing Laboratory of the Computer Science Department of the University of Crete and was approved by the University Research Ethics Committee.

Listener performance reported in the present paper was based on linguistic material that included 250 low frequency words selected from the lexical database Greeklex 2 (Kyparissiadis et al., 2017) which were expected to be unfamiliar to first graders. The words were embedded in the carrier phrase «Πες τη λέξη ...» (“Say the word ...”). The material was recorded twice in quiet by a female primary school teacher, once wearing a surgical face mask and once with no mask. Two types of noise were selected, real classroom noise and two-talker noise. Classroom noise (CN) was recorded inside a real classroom of a local primary school. Two-talker noise (TTN) was a mixture of two female speakers, each recorded separately reading a different children’s literature excerpt. According to the pilot experiment, children required higher SNR levels to reach adult performance (Leibold & Buss, 2019) and TTN posed a greater obstacle as documented in other studies (Corbin et al., 2016). Based on the pilot results, a single SNR level for each noise condition and age group was selected: 0 dB for adults and 2.5 dB for children for CN, and 5 dB for adults and 7.5 dB for children for TTN. A MATLAB algorithm was created in order to mix clean stimuli with the chosen SNR levels and randomise them for every listener.

For the main experiment, ten adults and ten children, five male and five female in each group, listened to 40 stimuli in each noise type, half produced with mask and half without mask, and afterwards the same words were presented in quiet, in the same mask condition as presented in noise. Thus each listener was tested on 160 words in total in three different sessions: one session in CN, one in TTN and one in quiet. A single loudspeaker was placed in front of the listener, at a radial distance of 0.9 m. Each mixture was presented at approximately 65 dB SPL. Listeners’ responses were recorded using Audacity and were rated by the first author. Any response that did not exactly match the original word was marked as incorrect.



## Results

Mixed Analysis of Variance (ANOVA) was used for the statistical treatment of the data. Participant score was analysed vs age group, gender, noise type and mask. All factors except gender were found statistically significant (age group:  $df=1$ ,  $F=7.059$ ,  $p=0.009$ ; gender:  $df=1$ ,  $F=0.163$ ,  $p=0.687$ ; noise type:  $df=2$ ,  $F=325.8$ ,  $p<0.001$ ; mask:  $df=1$ ,  $F=4.077$ ,  $p=0.0463$ ).

Both children and adults showed significantly better performance in the no mask (children: mean: 57.67%, SD: 24.89%, adults: mean: 60.67%, SD: 28.00%) than in the mask condition (children: mean: 51.50%, SD: 29.02%, adults: mean: 58.92%, SD: 27.65%). Also, children performed significantly worse than adults. Although the mask seems to impede children's performance more than that of adults (score reduction: 6.17% for children vs 1.75% for adults), the mask\*age group interaction did not reach stat. significance ( $df=1$ ,  $F=1.269$ ,  $p=0.2628$ ).

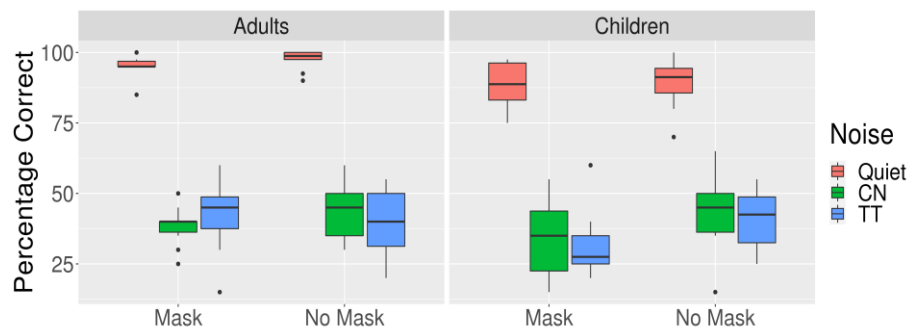


Figure 1. Percentage of correct responses for adults (left) and children (right) in the mask vs no mask condition and in quiet vs two noise types.

Among the three conditions, according to Tukey post-hoc tests, performance in quiet was significantly higher than in the two noise types (quiet vs. CN,  $p<0.001$ , quiet vs. TT,  $p<0.001$ ). No sig. differences were located between the two noise types (CN vs TT,  $p = 0.862$ ) which was expected due to the +5 SNR advantage in the TT condition. A trend for worse performance by children than adults in the mask condition, and especially in TT noise (children: mean: 31.50%, SD: 12.03%, adults: mean: 43.00%, SD: 13.58%), was manifested, although the interaction age group\*mask\*noise did not reach statistical significance ( $df=2$ ,  $F=1.161$ ,  $p=0.318$ ).

## Discussion

The pilot study confirms that children require higher SNR levels to reach adult performance (Leibold & Buss, 2019). Even with a +2.5 dB advantage, children still perform significantly worse than adults in the main experiment. TTN has been documented as a greater obstacle in relation to steady-state noise maskers

(Corbin et al., 2016); the pilot study exhibits that the same applies to TTN vs classroom noise (CN). The results of the main experiment demonstrate that recognition of low frequency words produced with a surgical mask is significantly lower for both children and adults in quiet and noise. This is an interesting finding as surgical masks have been reported to cause the least acoustic attenuation in comparison with other types of masks (Bottalico et al., 2020; Toscano & Toscano, 2021). The investigation of this matter, especially for children, is of great importance as auditory word recognition has an impact on language processing, the acquisition of new vocabulary and the development of the reading skill (Snowling et al., 1986). The choice of linguistic material (e.g. frequency of selected words or use of sentences vs individual words), as well as type of face mask are two significant factors in masked speech recognition (Vos et al., 2021). Thus, the trend for children's lower performance in the mask condition manifested in the present study should be investigated further, taking into account the aforementioned factors and including more child participants as well as participants with hearing or language impairments.

## Acknowledgements

Our special thanks to the primary school teacher for the material recordings, and all teachers and university staff who helped with the recruitment of participants.

## References

- Bottalico, P., Murgia, S., Puglisi, G.E., Astolfi, A., Kirk, K.I. 2020. Effect of masks on speech intelligibility in auralized classrooms. *J Acoust Soc Am.* 148(5), 2878-2884.
- Corbin, N.E., Bonino, A.Y., Buss, E., Leibold, L.J. 2016. Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear Hear.* 37(1), 55-63. doi: 10.1097/AUD.0000000000000201
- Erickson, L.C., Newman, R.S. 2017. Influences of background noise on infants and children. *Curr Dir Psychol Sci.* 26(5), 451-457.
- Kyparissiadis, A., van Heuven, W.J.B., Pitchford, N.J., Ledgeway, T. 2017. GreekLex 2: A comprehensive lexical database with part-of-speech, syllabic, phonological, and stress information. *PLoS ONE* 12(2), e0172493.
- Leibold, L.J., Buss, E. 2019. Masked speech recognition in school-age children. *Front. Psychol.* 10, 1981. doi: 10.3389/fpsyg.2019.01981
- Snowling, M., Goulandris, N., Bowlby, M., Howell, P. 1986. Segmentation and speech perception in relation to reading skill: A developmental analysis. *J Exp Child Psychol.* 41(3), 489–507. doi:10.1016/0022-0965(86)90006-8
- Toscano J.C., Toscano C.M. 2021. Effects of face masks on speech recognition in multi-talker babble noise. *PLoS ONE* 16(2), e0246842.
- Vos, T.G., Dillon, M.T., Buss, E., Rooth, M.A., Bucker, A.L., Dillon, S., et al. 2021. Influence of protective face coverings on the speech recognition of cochlear implant patients. *Laryngoscope* 131(6), E2038-E2043. doi: 10.1002/lary.29447.

# Global English metaphors on Greek press

Elpida Sklika

Department of Modern Greek Studies, University of Strasbourg, France

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This paper focuses on how English as a global language influences the modern Greek online journalistic discourse. The data consist of 126 Greek texts published between 2011-2015 in 14 newspapers, 4 magazines and 6 pure players. The texts are retrieved from three journalistic genres: *news articles*, *opinion articles* and *interviews*, and five semantic fields. We conducted a double analysis based on the theory of linguistic borrowing and the theory of conceptual metaphors to identify how conceptual metaphors reflect the power of English in the Greek press. According to results, these texts reflect the “effect of indexicality” and the “dramatization of discourse” to captivate the public. Metaphors are mostly in the fields of *everyday life*, *nature*, *technology*, *war*, and *disease*, and reveal the dominance of English on the journalists’ perception of the Greek society.

Keywords: Global English, Modern Greek, online press, conceptual metaphor

## Introduction

In the era of the Worldwide Net revolution, Internet transformed our society into a digital world with traditional media turning into “new media” (Manovich 2001). This paper examines how English as a global language (global English hereafter) affects Modern Greek online press and proposes an interdisciplinary study of language contact. We aim to find the means of this contact and if there are hints of linguistic globalisation.

This study is based on the theory of linguistic borrowing as the process of introducing new words in a language system through recipient language speaker’s agentivity (Winford 2005: 376). The criteria to search all types of borrowing from global English include the *frequency* of their use, their *spread* and high level of *recognition* and *acceptance*.

English has acquired the status of a global language, i.e., the dominant language for international communication. Global English is examined by many researchers (Crystal 1997, Jenkins 2000, Seidlhofer 2005, Graddol 2006, etc.), while others point out its *linguistic imperialism* and *hegemony* (Phillipson 1992, Calvet 2002, etc.). Finally, the social context in Greece today is also characterised by the extensive use of English, as the first taught foreign language, broadly used in international communication, academia, job-searching, media and culture (Mackridge 2012).

## Method

Our method is based on the theory of linguistic borrowing combined with Fairclough's discourse analysis framework (2010), and the theory of conceptual metaphors to identify the power of English in the journalists' discourse.

By using the Fairclough's framework, we followed the three-dimension model of analysing a text by examining: *discourse-as-text*, *discourse-as-discursive-practice* and *discourse-as-social-practice*. Yet, the social aspect cannot be examined in the present study (ibid.).

In addition, it is important to define the notion of metaphor. "A metaphor is a figure of speech in which a word or phrase is used to describe something that it does not literally mean" (McGlone 2007: 109). For Lakoff the context is necessary to understand metaphors since they are present in our everyday life and give meaning to our concepts (1993). They are the network of relationships between our experience of the world and our cultural perception (Lakoff & Johnson 1980). We also distinguish three components of a metaphor: the *topic/tenor* (the subject), the *vehicle* (the term used metaphorically), and the *ground* (the relationship between them) (ibid.).

## Data

The data of this study consist of 126 Greek texts published between 2011-2015 in 14 newspapers, 4 magazines and 6 digital newspapers, i.e., pure players. Our texts are retrieved from three genres: *news articles*, *opinion articles* and *interviews*, and five semantic fields, also known as columns in the press. We regrouped them into the five most read columns for a more illustrative sample, which are *art-fashion*, *finance-politics*, *world-environment*, *science-technology* and *health-sports*.

## Results and discussion

For the analysis of our corpus, we employed the theory of conceptual metaphors, evoking the idea that the metaphORIZATION is based on the experiences we have with physical objects and with our body. This type of metaphors are ontological metaphors, and they are frequent in daily discourse, as well as in the press. It is therefore important to understand that in stating a metaphor, we already have in mind a stereotype image of the terms to compare, which helps us when encoding and decoding the message. All the examples of global English metaphors in our data are translated loans from English (calques) that convey the semantic categories of *nature*, *everyday life* including *technology*, *site construction* and *finance*, *war* and *health-disease*. In Table 1, we noted some metaphors (2<sup>nd</sup> column) as indicative examples followed by the English prototype cognitive concept (3<sup>d</sup> column) and the message's metaphorical meaning (4<sup>th</sup> column). The 1<sup>st</sup> column shows the number of the text in our corpus (T: text, H: text-hyperlink), the column it is extracted from press and the semantic field that the metaphor illustrates.

Table 1. Examples of English metaphors in the Greek journalistic discourse.

N <sup>o</sup> Text/Column/se mantic field	Metaphors in Greek corpus	English cognitive concept	metaphorical meaning
T20/world- environment/nat ure	Πράσινες λύσεις	green solutions	something respectful to nature, ecological
H8.1/finance- politics/every-day life	πράσινο φως	green light	go forward, green traffic light
T56/world- environment/nat ure	φιλικά προς το περιβάλλον	environment friendly	something respectful to nature, ecological
T2/art- fashion/every-day life	Ρούχα έξυπνα	smart clothing	word play with the loan smartphone, something practical
T14/finance- politics/every-day life	ένα «μαξιλάρι» ομολόγων	cushion bond	a lifesaving cushion from drowning, a cushion for the funds' protection
T10/finance- politics/war	τα στρατόπεδα του Eurogroup	The Eurogroup's camps	conflicting countries' mapping (like war camps)
T28/science- technology/medic ine	μολυσμένα PC	infected computers	infecting with a malware/virus

In other words, we found conceptual metaphors in the fields of technology used in everyday life (e.g. *έξυπνες συσκευές* = smart devices, *έξυπνες τηλεοράσεις* = smart televisions), construction and building (e.g. *χτίζω ένα όνομα* = to build a name), finance (e.g. *μαξιλάρι ομολόγων* = a cushion bond), internet and technology (e.g., *έξυπνα κινητά* = smartphones, «*Ta social media είναι τα παράθυρα της ψυχής μας*» = social media are the windows of our soul), war (e.g. «*Ta στρατόπεδα στο τραπέζι του Eurogroup*» = the camps at the Eurogroup round table) or health and *disease* (e.g. *μολυσμένα PC* = infected computers, *τοξικούς ανθρώπους* = toxic people). Yet, this remark drove us to further distinguish this influence in two complementary types of factors that explain the use of translated English metaphors in the Greek press.

On the one hand, we suggest the *external factors* showing the implantation of anglicisms from top-to-bottom in the discourse to underline the power and the hegemonic status of English. On the other, the *internal factors* propose that the occurrences of global English influence on Greek press are linked to the bottom-up movement and show the proper journalists' choices (*word preference and personal stance in discourse*) highlighting either their language skills (*English as a marker of indexicality*) and their need to promote their skills, or rather giving a humorous tone to their texts (*symbolic function of English*) to gain the public's attention. Nevertheless, even if these occurrences are not numerous, it is worth to mention that these factors form the continuum of this contact. They reveal that the choice of each discourse strategy underlines both the hegemonic status of English in the Greek journalistic discourse, and an influence on what people read, write and what remains memorized since it becomes common knowledge.

## Conclusion

To sum up, this study examined a corpus of texts from three journalistic genres of the Greek online press using a combined methodology of a discourse and semantic-orientated framework that finally proposed a two-factors system of journalists' word choices and discourse strategies. In brief, we hope that this study will be welcome for the scientific processing of these data and will stimulate further studies so to give a larger image of the Greek online press.

## References

- Calvet, L.J. 2002. La mondialisation, les langues, les politiques linguistiques et internet. Conférence au congrès de l'AILF, 1st February 2002, 1-12.
- Crystal, D. 1997. English as a Global Language. Cambridge, Cambridge University Press.
- Fairclough, N. 2010. Critical Discourse Analysis: The Critical Study of Language. London, Routledge, Taylor & Francis Group.
- Graddol, D. 2006. English Next, Why Global English Might Mean the End of 'English as a Foreign Language'. British Council: The English Company (UK) Ltd. Online: <http://www.britishcouncil.org/learning-research-english-next.pdf>
- Jenkins, J. 2000. The phonology of English as an international language. New models, new norms, new goals. Oxford, Oxford University Press.
- Lakoff, G. & Johnson, M. 1980. Metaphors We Live By. Chicago, The University of Chicago Press.
- Lakoff, G. 1993. The contemporary theory of metaphor. In Ortony, A. (Ed.), 1993, Metaphor and Thought, 202-251, New York, Cambridge University Press.
- Mackridge, P. 2012. 'Omileite...anglika?' English influences in contemporary Greek. In Jubiläum Hans Eideneier, Hamburg, Germany, 21 June 2012.
- Manovich, L. 2001. The Language of New Media. Massachusetts, MIT Press.
- Phillipson, R. 1992. Linguistic Imperialism. Oxford, Oxford University Press.
- Seidlhofer, B. 2005. English as a lingua franca. ELT J, 59 (4), 339-341.
- Winford, D. 2005. Contact-induced changes: Classification and processes. Diachronica 22:2, 373-382.

# Extracting word-like units when two concurrent regularities collide: Electrophysiological evidence

Ana Paula Soares, Alexandrina Lages, Helena Mendes Oliveira,  
Francisco-Javier Gutiérrez-Domínguez

<sup>1</sup>Human Cognition Lab, CIPsi, University of Minho, Braga, Portugal.

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Statistical learning (SL) is a fundamental skill assumed to play a central role in the acquisition of the rule-governed aspects of language. Despite evidence that SL is present from early infancy to support the extraction of sound patterns in speech, the nature of the computations involved is unclear. Here we collected electrophysiological data while preschool children were exposed to an auditory stream in which two concurrent regularities were embedded, firstly, under accidental (implicit), and, subsequently, under intentional (explicit) conditions. Results showed that the extraction of sound patterns was enhanced by the effect of explicit instructions and, critically, that under this conditions, children seem to rely on the computation of syllable frequency rather than on transitional probabilities to extract word-like units.

Keywords: statistical learning, transitional probabilities, syllable frequency, implicit learning, explicit learning

## Introduction

We live in a world full of regularities. Language, for example, can be described by a set of rules that determine how sounds are combined into words and words into meaningful sentences. Statistical learning (SL) is a general term used in Cognitive Science to describe the action of multiple mechanisms that allow the extraction of those regularities in time and space, such as the sound patterns present in speech to determine where one word ends and the next begins.

Although most SL studies have focused on examining the sensitivity of young children to speech regularities, such as the probability of one segment (e.g., a syllable) to follow another segment, a statistic known as transitional probability (TP; Saffran et al., 1996), others have also demonstrated that children are sensitive to other statistics, such as the frequency with which each exemplar is presented in the input. For instance, Maye et al. (2002) showed that 6- and 8-months infants exposed to a bimodal distribution of the [ta-da] continuum treated exemplars occurring between the center and the endpoints of the distribution as belonging to different phonetic categories, while those exposed to a unimodal distribution of the same continuum treated these exemplars as belonging to the same category. Although all these studies demonstrate that, from a very young age, children are sensitive to different

statistics, knowledge about how the sensitivity to those regularities emerge and how the conditions under which they were presented (implicit vs. explicit) affected the reliance in one or another is limited.

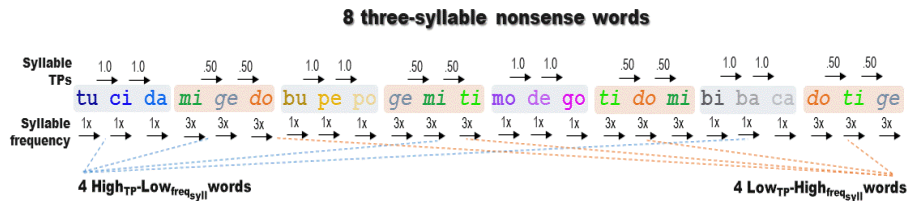
## Method

### Participants

Twenty-four children (13 female,  $M_{age} = 5;7$ ; range 5;1 to 6;5) from Portuguese kindergarten institutions participated in the study. All participants were native speakers of European Portuguese, with normal hearing and no history of language disabilities. Written informed consent was obtained from the parents. The study was approved by the local Ethics Committee (SECSH 028/2018).

### Stimuli

Thirty-two auditory European Portuguese CV syllables were drawn from Soares et al.'s (2020) study to create the 16 three-syllable nonsense words used in the implicit and explicit versions of the SL tasks (8 'words' per task). Syllables were produced and recorded by a native speaker of European Portuguese with a duration of 300 ms each. In each task, 4 of the nonsense words presented high TPs (1.0) but low frequent syllables, and the other 4 low TPs (.50) but high frequent syllables (see Fig. 1). For instance, the nonsense word 'tucida' correspond to the High TP-Low frequency syllable 'words' as the syllables they entail only appear in those 'words', hence making these 'words' to present a TP = 1.0, but a syllable frequency that is three times less than presented by the Low TP-High frequency syllable 'words' such as 'migedo', which entail syllables that appear in three different 'words' at different syllable positions ('gemit?', 'tidomi?', 'gotige?') - see Soares et al., 2020 for details.



In each SL task, the nonsense words were concatenated in a continuous stream of 8.4 min with the Audacity® software with a 50 ms interval between syllables. Each 'word' was presented 60 in random order with the restriction that the same 'word' cannot be repeated twice in a row. The stream was edited to include ~10% of the syllables a superimposed sound (a 0.1 s sawtooth wave sound from 450 to 1,450 Hz).

### Procedure

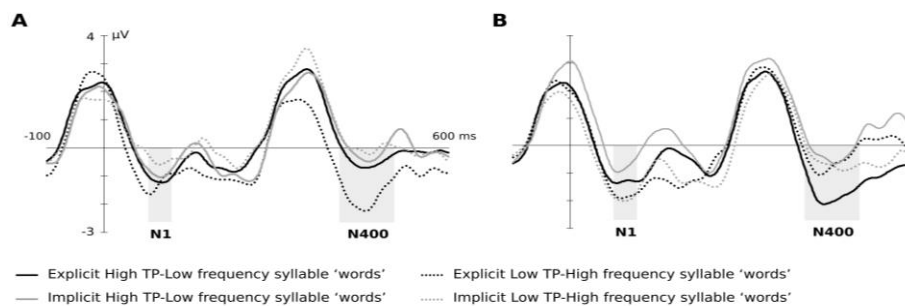
Data were collected in a shielding cabin at the Psychological Neuroscience Lab (School of Psychology, University of Minho). Participants were firstly presented



with the implicit version of the SL task and, subsequently, with the explicit version of an analogous SL task. In the implicit version, participants were instructed to pay attention to the auditory stream presented at 60 dB SPL via binaural headphones, because occasionally a deviant sound (a click) would appear, and their task would be to detect it as soon and accurately as possible, which functioned as a cover task to ensure adequate attention to the stimuli. After a brief interval, participants underwent the explicit version of the SL task that mimicked the previous one, except that each of the new ‘words’, drawn from the syllabary not used in the implicit version, was presented and participants were asked to repeat each one correctly before the familiarization phase began. The procedure took about 90 min per participant. EEG data were recorded with a 64 channels BioSemi Active-Two system according to the international 10-20 system and digitized at a sampling rate of 512 Hz. Electrode impedances were kept below 20 k $\Omega$ . EEG was re-referenced off-line to the algebraic average of mastoids. Data were filtered with a bandpass 0.1-30 Hz filter. Epochs were time-locked to the onset of the ‘words’.

## Results

EEG data processing was conducted with the Brain Vision Analyzer 2.1.1. Four participants were excluded due to artifact rejection. Mean amplitudes were calculated for the 80-120 ms (N1) and 400-500 ms event-related potentials (ERPs) taken as the neural signatures of words’ segmentation in the brain (see Soares et al., 2020) in the topographical regions where amplitudes were maximal (fronto-central and central regions, respectively). Repeated-measures ANOVAs were conducted in the 27.0 IBM-SPSS® software based on the three within-subject factors: SL task (implicit vs. explicit), Type of ‘word’ (High TP-Low frequency syllable vs. Low TP-High frequency syllable), and Time (first half vs. second half of the SL task) to further analyse how neural responses changed as exposure to the auditory stream unfolds. Figure 2 depicts the neural responses per condition in the first (A) and second-half (B) of the SL tasks.



In the N1 component, results showed a main effect of Time,  $F(1,19) = 5.22$ ,  $p = .034$ ,  $\eta_p^2 = .215$ , indicating larger amplitude in the second than in the first

half of the SL tasks. In the N400 component, the results showed a main effect of task,  $F(1,19) = 8.23$ ,  $p = .010$ ,  $\eta_p^2 = .302$ , indicating larger amplitude in the explicit than in the implicit version of the SL tasks. The three-fold SL task\*type of 'word'\*time interaction was also significant,  $F(1,19) = 4.65$ ,  $p = .044$ ,  $\eta_p^2 = .197$ . This effect revealed larger amplitude in the first half of the explicit SL task for the Low TP-High frequency syllable 'words' ( $p = .030$ ), and in the second half of the explicit SL task for the High TP-Low frequency syllable 'words' ( $p = .027$ ). Moreover, it also revealed that the Low TP-High frequency syllable 'words' elicited larger amplitude in the first half than in the second half of the explicit SL task ( $p = .022$ ), and that in the first half of the explicit SL task the Low TP-High frequency syllable 'words' elicited larger amplitude than the High TP-Low frequency syllable 'words' ( $p = .041$ ).

## Conclusion

The enhancement in the N1 component observed in the last part of the SL tasks, suggest this ERP component to index transient effects arising from the extraction of the regularities embedded in the input stream as the exposure unfolds. The results in the N400 component, taken as a putative 'marker' of words' segmentation in the brain, indicate that the extraction of word-like units was enhanced by the effect of explicit instructions and, critically, that under these conditions, children seem to rely on the computation of syllable frequency rather than syllable TPs to extract word-like units. Future research should further explore this issue by analysing how the complexity of the stream used (8 nonsense words) might have contributed to these results.

## Acknowledgements

This study was conducted at the Psychology Research Centre (PSI/01662), University of Minho, and supported by the Grant POCI-01-0145-FEDER-028212 from the Portuguese Foundation for Science and Technology and the Portuguese Ministry of Science, Technology and Higher Education.

## References

- Maye, J., Werker J. F., Gerken, L. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, 101–111.
- Saffran, J.R., Aslin, R.N., Newport, E.L. 1996. Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Soares, A.P., Gutiérrez-Domínguez, F., Vasconcelos, M., Oliveira, H. M., Tomé, D., Jiménez, L. 2020. Not all words are equally acquired: Transitional probabilities and instructions affect the electrophysiological correlates of statistical learning. *Frontiers in Human Neuroscience*, 14, 577991.

# Factive verbs and island effects in Romanian

Irina Stoica

Department of English, University of Bucharest, Romania

## Abstract

Factive verbs are generally said to induce weak island effects, allowing the extraction of arguments, but not that of adjuncts, from the post-verbal clause. However, a closer look at the data shows that not all factive verbs behave the same, which led some researchers to postulate at least two sub-types of factive verbs: strong/weak factives (Karttunen, 1977) or, in more recent studies, cognitive/emotive factives (Djarv, 2019). This difference can be seen in the case of Romanian factive verbs as well, with cognitive and emotive factives evincing distinct properties. The question this study raises is whether the two sub-types of factive verbs evince distinct behaviour with respect to island effects as well.

Keywords: factive verbs, cognitive factives, emotive factives, extraction, island effects

## Starting point

Factive verbs have been a topic of much interest ever since Kiparsky and Kiparsky's seminal work (1971), where the authors argue that many of the differences between factive verbs (*know, regret, resent*) and non-factive verbs (*think, believe*) can be accounted for via a nominal layer that exists only in the structure of the former, but not in that of the latter.

However, a closer look at the data seems to indicate that there might be a further division between factive verbs themselves, previous studies in the field distinguishing between full factives and semi-factives (Karttunen, 1977), or, in more recent studies, between cognitive and emotive factives (Djarv, 2019). These two sub-types of factive verbs are said to differ not only semantically, but also syntactically, with the former even being argued to allow extraction in some contexts, contrary to the previous treatment of factive verbs as weak island inducers, as per Szabolcsi, den Dikken, 2003.

The aim of this paper is to test whether cognitive factives and emotive factives differ in Romanian, from the point of view of syntactic islandhood.

## Cognitive and emotive factive verbs and island effects

Cognitive and emotive factive verbs are said to differ in a number of ways, both semantically (e.g. presupposition cancellation contexts) and syntactically (e.g. main clause phenomena or complementizer drop) (Djarv, 2019). These differences can be seen not only with respect to English factive verbs, but also in other languages, such as Greek (Roussou, 2010) or Bulgarian (Krapova, 2020).

In Romanian too there seems to be a split between the two sub-types of factive verbs: cognitive factives differ from emotive factives at least from the point of view of referentiality, the use of *wh*-complementizers, the acceptability of main clause phenomena. (Stoica, 2021).

With respect to extraction, Romanian factive verbs are said to induce weak island effects. However, in an experimental study (Stoica, 2019) where the sentence in (1) was used as a distractor, respondents accepted extraction out of the post-verbal clause of a cognitive factive verb.

(1) *În ce și-a amintit Ion că împachetau părinții lui cadourile?*  
In what did Ion remember that his parents used to wrap the presents?

Taking into account the distinct behaviour these sub-types of factive verbs evince, both in Romanian and in other languages, the question is whether this dichotomy is reflected into syntactic islandhood as well.

## **Factive verbs and island effects – an experimental study**

### **Aim**

The aim of the current study is two-fold: first of all, I wish to test whether factive verbs induce island effects in Romanian in the first place. Secondly, I am interested in seeing whether the distinction between cognitive factives and emotive factives is reflected into syntactic islandhood as well.

### **Materials and procedures**

This experiment combined two designs: a question after story and a truth value judgement task. Respondents listened to the experimenter tell a story to a character, who, in turn, was asked a question at the end. Respondents were asked to say whether the response given by the character was true or false.

There were 8 test sentences in total, 4 of which targeted the island effects imposed by cognitive factive verbs: *a-și aduce aminte* (remember), *a-și da seama* (realize), *a uita* (forget) and *a afla* (find out) and 4 those induced by emotive factive verbs: *a regreta* (regret), *a se enerva* (be angry), *a se bucura* (be happy) and *a se întrista* (be sad).

Given that factive verbs in general are said to induce weak island effects, I only tested adjunct extraction with *când* (when) and *unde* (where):

- The test items were balanced for length and complexity.
- There were also 8 control sentences, where respondents saw variants of the same scenarios, but which targeted short distance movement, and 8 distractors.
- Respondents were split into two groups and saw mirroring variants of the questionnaire.

Examples of these scenarios are given in Table 1 below.

Table 1. Examples of test items and their English translations.

Cognitive factives	<p>Emă se uita la Tom și Jerry în sufragerie. După câteva scene, a exclamat “Dar eu am mai văzut episodul ăsta la grădiniță”, așa că i-a zis mamei ei: “Mami, poți să îmi pui altul?”</p> <p><b>Povestitor:</b> Unde și-a adus aminte Emă că a văzut episodul?</p> <p><b>Personaj:</b> - La grădiniță</p>	<p>Emă was watching Tom and Jerry in the living room. After a couple of scenes, she said: I’ve already seen this episode at the kindergarten”, so she told her Mom: „Mommy, can I have another one?”</p> <p><b>Storyteller:</b> Where did Emă remember that she saw the episode?</p> <p><b>Character:</b> at the kindergarten</p>
Emotive factives	<p>Filip avea un motan pe nume Pișcot. Într-o zi, Pișcot a fugit în grădină, în timp ce Filip nu era acasă. Părinții lui au venit la grădiniță și i-au zis: ”Scumpule, Pișcot a fugit...” . Filip a început să plângă și i-a întrebat pe părinți: “Credeți că o să-l mai găsim pe Pișcot?”</p> <p><b>Povestitor:</b> Unde s-a întristat Filip că a fugit Pișcot?</p> <p><b>Personaj:</b> În grădină.</p>	<p>Filip had a cat called Pișcot. One day, Pișcot ran in the garden, while Filip was not at home. His parents went to the kindergarten and told him: ”Honey, Pișcot ran away...”. Filip started crying and asked his parents: “Do you think we’ll ever find him again?”</p> <p><b>Storyteller:</b> Paddington, where did Filip get sad that Pișcot had run away?</p> <p><b>Character:</b> In the garden.</p>

### Participants

74 native speakers of Romanian (mean age 36.6) took part in this task, all naïve to the aim of this experiment.

### Results

The results show that, in line with previous findings, factive verbs impose (at least) weak island effects in Romanian. As can be seen in Table 2 below, extraction of adjuncts from the post-verbal clause of factive verbs in Romanian was rejected.

Table 2. Acceptability rates of adjunct extraction.

Cognitive factives		Emotive factives	
când-extraction	unde-extraction	când-extraction	unde-extraction
5.45%	39.39%	6.75%	28.04
20%		17.94%	
18.91%			

A Welch t-test at the  $\alpha=.05$  level further shows that there is no statistically significant difference between the island effects imposed by cognitive factive verbs and emotive factive verbs ( $t(7)=1.97$ ,  $p=0.88$ ).

### Conclusions and questions for further research

The results of the experiment above show that the distinction between cognitive and emotive factives does not reflect into syntactic islandhood, extraction of adjuncts being banned overall. However, there are several questions that stem from this observation: i) can the distinctions between cognitive factives and emotive factives be accounted for structurally? and if so ii) why do cognitive factives and emotive factives behave similarly from the point of view of extraction? and, last but not least, iii) are there some aspects that improve extraction even in these cases? (i.e. see the example in (1) above). I leave all of these questions for further research.

### Acknowledgements

The research presented in this paper was funded by the ICUB Young Researchers Grant.

### References

- Djarv, K. 2019. Factive and assertive attitude reports. PhD dissertation. The University of Pennsylvania.
- Kiparsky, P., C. Kiparsky 1971. Fact. In D. Steinberg, L. Jakobovits (eds.) *Semantics*. Cambridge: Cambridge University Press.
- Krapova, I. 2020: On the semantics and syntax of factivity in South Slavic and Balkan languages. Talk given at the Linguistics in Göttingen Colloquium series
- Rroussou, Anna. 2010. Selecting complementizers. Special issue *Exploring the left periphery*. *Lingua* 120: 582-603.
- Szabolcsi, A. & Den Dikken, M. 2003. "Islands". In L. Cheng & R. Sybesma (eds.) *The Second State of the Article Book*, Berlin, Mouton de Gruyter
- Stoica, I. 2019: The syntax and the semantics of manner of speaking verbs, PhD dissertation, University of Bucharest
- Stoica, I. 2020: Why knowing something is not the same as regretting it – an overview of Romanian factive verbs. Talk given at AICED 21, University of Bucharest

# **/t/ variation in Connemara English**

Kate Tallon

Phonetics and Speech Laboratory, Trinity College Dublin, Ireland

<https://doi.org/10.36505/ExLing-2021/12>

## **Abstract**

This paper offers a preliminary investigation into the variation of /t/ in L2 English in the Connemara Gaeltacht (Irish-speaking region) Ireland. As part of a broader pilot study it investigates the anecdotal claims that, in this dialect, /t/ is realised as a plosive ([t]) in intervocalic and post-stressed word-final contexts. This pattern diverges from the typical Irish English fricated variant [tʃ] (known as ‘slit-t’) and the tapped variant [ɾ] (intervocally); the latter is popular among younger Irish English speakers. Findings show that in the investigated contexts, /t/ is not always realised as [t], which is at variance with existing accounts. In addition, /t/ variation appears to be dependent on the strength of subjects’ social ties to the community.

Keywords: Irish English, /t/-lenition, sociophonetics, language contact

## **Introduction**

A salient phonological feature of Irish English (IrE), setting it apart from other major English varieties, is the apico-alveolar fricative [tʃ] or ‘slit-*l*’, which occurs in post-stressed intervocalic position (e.g. *water* [wɑ:tʃə]) and in word-final position following a vowel or /r/ (e.g. *cat* [kɑ:tʃ], *cart* [kɑ:ɾtʃ]). This feature is prevalent throughout southern IrE in all regions and registers. Among younger IrE speakers, a tapped variant has been found to commonly occur in intervocalic position (e.g. *water* [wɑ:rə]) (Hickey 2007; Lonergan 2013, Wells 1982).

A variety of IrE which reportedly diverges from this pattern is the English spoken by L1 Irish speakers. This variety is said to display a plosive [t] rather than the mainstream IrE fricative [tʃ] in the aforementioned contexts (Hickey 1984). The production of /t/ has not yet been investigated in the English of Irish speakers and the reports of a plosive [t] realisation remain anecdotal. This study addresses this gap by quantitatively investigating /t/ variation in the Gaeltacht (Irish-speaking) region of Connemara (west of Ireland).

Based on previous findings from IrE linguistic studies, it was expected that younger speakers would produce lower levels of the traditional (and stigmatised) plosive variant and higher levels of the mainstream fricated or tapped variants (Hickey 2007, Lonergan 2013).

## **Methodology**

Five L2 English speakers were recruited as part of a broader doctoral pilot

study. All subjects were native L1 Irish speakers from the Connemara Gaeltacht and spoke both Irish and English in their daily lives. Additional details for each subject are shown in Table 1.

Table 1. Details of subjects used in study.

Participant	Gender	Age	Childhood home language	Language of education
P1	M	21	Irish	Irish
P2	F	52	Irish and English	Irish and English
P3	M	16	Irish	Irish
P4	M	25	Irish	Irish
P5	M	30	Irish	Irish

A set of 30 sentences and a reading passage were designed to elicit a wide range of phonemes in order to gain a broad picture of the phonological system of this dialect. The sentences were displayed individually on a computer screen for subjects to read. The reading passage was given to subjects in hardcopy format. Subjects were recorded in their own homes in settings with as little background noise as possible. Recordings were made with an Audio Technica AT2020 USB+ cardioid condenser microphone connected to an Apple Macbook Pro. 77 instances of /t/ were analysed in sentence-final (N = 15) and in post-stressed intervocalic (N = 62) contexts. The results of an initial auditory analysis were confirmed by acoustic analysis using Praat. Tokens of /t/ were classified as either plosives [t], taps [ɾ], or fricatives[t̪].

## Results

In intervocalic contexts (Figure 1) three out of five subjects (P3, P4, P5) demonstrated high usage (>86%) of the plosive variant and did not produce any fricative variants. The other two subjects (P1, P2) demonstrated almost exclusive use of the fricative variant (>82%). The tapped variant [ɾ] was evident in the speech of three subjects (13-18% of tokens).





Figure 1. The realisation of /t/ in intervocalic position by participant (P).

In word-final contexts (see Figure 2), three subjects (P3, P4, P5) each demonstrated 100% usage of the plosive variant. Subjects P1 and P2 each demonstrated 100% usage of the fricative variant.

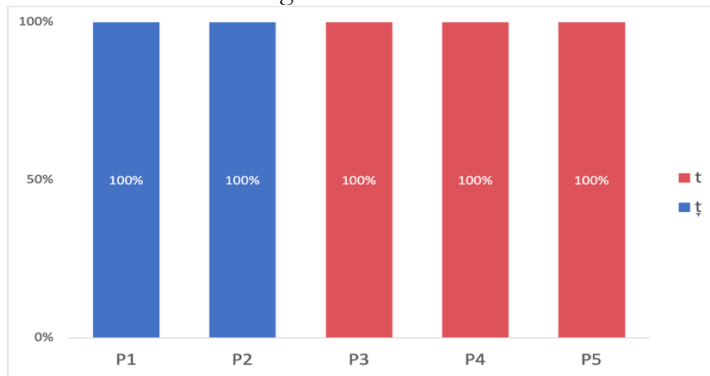


Figure 2. The realisation of /t/ in word-final position by participant (P).

## Discussion

The findings of this study are inconsistent with existing accounts of this dialect (Hickey 1984). Given the homogeneity of the subjects' linguistic backgrounds it cannot be claimed here that all Irish speakers of English display a plosive variant in the studied contexts. It therefore cannot be claimed that the production of /t/ correlates with speakers' linguistic abilities in Irish.

It was expected that younger subjects would favour the fricated or tapped variants. Although additional data is needed that includes a larger number of older speakers, this pattern is not observed in this study. Therefore, based on the data acquired here, age is not a reliable predictor for /t/ variation. A more reliable predictor may be the subjects' ties to the local community. For example,

Participant 1, who did not use the traditional plosive variant, had recently moved from rural Connemara to Galway city and had formed strong relationships there. Participant 2, who also did not use the plosive variant, had lived in London for several years and although living and working in Connemara, had strong social ties outside the local area. The other three participants (P3, P4, P5) who showed high rates of the plosive variant, all live, work and socialise in the local area. Although additional data is needed to draw more general inferences about this dialect, the collected data suggests that the more strongly a subject is associated with the local community the higher their rates of the plosive [t] variant.

### **Conclusion and recommendations**

To further investigate the patterns observed in this preliminary pilot analysis, a questionnaire could be given to subjects to elicit further details about their ties to the local community. Scores would be given to each participant (in the manner of Milroy & Milroy's (1997) network strength scores). This would facilitate a quantitative investigation into the relationship between subjects' ties to the local community and their production of /t/.

### **Acknowledgements**

Thanks are due to the Irish Research Council for funding this research via the Government of Irish Postgraduate Scholarship.

### **References**

- Hickey, R. 1984. Coronal segments in Irish English. *Journal of Linguistics*, 20(2), 233-250.
- Hickey, R. 2007. *Irish English: History and present-day forms*. Cambridge, CUP.
- Lonergan, J. 2013. *An Acoustic and Perceptual Study of Dublin English Phonology*. (unpublished PhD Thesis). University College Dublin. Dublin.
- Milroy J., Milroy L. 1997. Network Structure and Linguistic Change. In Coupland N., Jaworski A. (eds) *Sociolinguistics*. Modern Linguistics Series. Palgrave, London.
- Wells, J. C. 1982. *Accents of English*. Cambridge, CUP.

# Virtual rapport management: responses to positive online reviews

Ly Wen Taw<sup>1,2</sup>, Shamala Paravasivam<sup>2</sup>, Alan Libert<sup>1</sup>, Christo Moskovsky<sup>1</sup>

<sup>1</sup>School of Humanities and Social Science, The University of Newcastle, Australia

<sup>2</sup>Department of English, Universiti Putra Malaysia, Malaysia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The proliferation of technology has facilitated online feedback mechanisms, and has led to an increase in electronic Word-of-Mouth (eWOM). Due to the influence of eWOM, online reviews and hotel responses have a significant impact on customers' purchasing intentions. Drawing on Spencer-Oatey's (2008) Rapport Management Model (RMM), the present study explores the rapport management strategy in the RMM discourse domains that Malaysian five-, four- and three-star hotel management use to respond to positive online reviews. The data regarding hotel responses to positive online reviews were collected from TripAdvisor. Using the Genre Analysis Model (Bhatia, 1993), six major 'moves' in responding to positive online reviews were identified. The results provide insights into the cultural rapport management strategies used in Malaysia from a discourse perspective.

Keywords: online reviews, hotel responses, moves, rapport management, TripAdvisor

## Introduction

Technological development has enabled consumers to post their views about products or services via electronic Word-of-Mouth (eWOM) on social media platforms. With the widespread adoption of eWOM, customers' voices in online reviews are emerging as a powerful influence on consumers' purchasing intentions and trust, as negative reviews can be detrimental to sales growth and business performance. Thus, business management's digital engagement in social media has become significant because the effects of reviews can be mitigated.

Establishing virtual rapport via management's responses to online reviews not only builds consumer trust, but also fosters social rapport with the audience in the online review community. Given the increasing influence of eWOM, hotel management's responses can influence consumers' purchasing decisions. Research has shown that business management's engagement with customers in online communities can effectively enhance the reputation of a business online, resulting in business growth (Anagnostopoulou, Buhalis, Kountouri, Manousakis, & Tsekrekos, 2020). Hotel management's responses to positive online reviews can kindle feelings of appreciation to customers and make them feel valued. Although there has been research on move structures in hotel management's responses to negative online reviews, there is still a paucity of

studies investigating move structures in managements' responses to positive reviews on social platforms.

Drawing on Spencer-Oatey's (2008) RMM, the present study adopts a linguistic perspective to explore the rapport management strategies that Malaysian hotel management of five-, four-, and three-star hotels used to respond to positive online reviews and to establish virtual rapport with the reviewers.

The following were the research questions in this study:

- (1) What move structures did the five-, four- and three-star hotels use in their response organisation?
- (2) What were the differences in five-, four- and three-star hotels' rapport management when responding to customers' positive online reviews?

### **Theoretical framework: RMM**

Building rapport is an essential component of social relationships, and the use of language plays a crucial role, particularly in asynchronous computer-mediated communication in the absence of visual cues. This research applied Spencer-Oatey's (2008) RMM, according to which there are three inter-related rapport bases. The element of interactional goals is the first rapport base. It is believed that there are specific goals in interactions with others, which can be relational or transactional in nature, or task focused. The next rapport base is face sensitivities, which refers to an individual's sense of identity or self-concept.

Lastly, sociality rights and obligations is the rapport element base that is related to an individual's perceptions of having a range of sociality rights and obligations in relation to others. According to the RMM, there are five rapport management strategies from the linguistic perspective that are conceptualised in the discourse, illocutionary, stylistic, non-verbal and participation domains.

This research examines the discourse domain in rapport management in order to examine the structure of hotel responses when managing rapport. Therefore, the Genre Analysis Model (Bhatia, 1993) was applied to the data to identify the move structures.

### **Methodology**

The data were collected from an online travel review website, *TripAdvisor*, which is known as the world's largest travel site, with more than 830 million online reviews (TripAdvisor, 2017). The online reviews are classified according to the following five traveller ratings: *Excellent*, *Good*, *Average*, *Poor* and *Terrible*. Since this study examines hotel management's responses to positive reviews, the traveller ratings of *Excellent* and *Good* are categorised as positive reviews.

*TripAdvisor* lists the classes of hotels as five-, four-, three- and two-star. It is worth mentioning that responses from two-star hotel management were rare. Therefore, as a result of the application of purposeful sampling, a total of 72

management responses to positive reviews were collected from Malaysian five-, four- and three-star hotels in three destinations, namely Kuala Lumpur, Selangor and Pahang. As the hotel industry is correlated significantly with tourism destinations in Malaysia (Attila, 2016), the selected hotels were located in the popular tourist destinations.

Using NVivo 12, the moves in the response structures were coded according to the hotel's star rating and destinations. Herring's (2004) Computer-Mediated Discourse Analysis research approach was used to analyse the data.

## Results and discussion

The findings identified six main moves in the responses to positive reviews, namely opening, acknowledging feedback, addressing complaints, positioning the hotel brand, concluding and closing. Despite the travellers' higher ratings in the positive reviews, negative feedback was not excluded from the positive reviews. Therefore, it was found that some hotels addressed the negative comments, and the main move of dealing with complaints was identified.

Acknowledging feedback was the most frequently used main move, and four sub-moves were identified under this main move. They were *thanking for the feedback*, *recognising the value of the feedback*, *expressing positive feelings* and *acknowledging the mention of the staff*. Of these four sub-moves, *expressing positive feelings* was used most frequently. The results also revealed that satisfactory customer service was associated strongly with positive feelings (Calvert, Pathak, Ching, Trufil, & Fulcher, 2019); The following are some examples of the sub-move of *expressing positive feelings* posted by the five-, four- and three-star hotels.

### Five-star

I am elated to receive your appreciated feedback.

Your FIVE star review made our day and thank you for your kind compliments!!

We are over the moon to learn about your recent experience with us at Hotel A.

### Four-star

It is truly gratifying to read that you had such a wonderful stay with us.

We are absolutely overjoyed to hear that you had an amazing stay here in Hotel B in Kuala Lumpur.

It is indeed most heart-warming to read that our hotel reminded you so much of both Scotland and Wales.

### Three-star

We are glad that you enjoyed your stay.

We are delighted that you have enjoyed your stay with us.

I am happy to know of that you are satisfied with your stay.

The examples above show that the five- and four-star hotels were more inclined to vary their lexical choices when expressing positive feelings to

establish rapport with the reviewers. However, most of the three-star hotels were limited to the use of words such as *happy*, *glad* and *delighted* in the use of this particular sub-move in rapport management.

## Conclusion

This study provided insights into how language reflects and conveys the virtual rapport management strategies used by hotel management on digital platforms, particularly in the sociocultural context of a country in Southeast Asia, Malaysia. In light of the increasing use of online platforms in business communication, the findings could be incorporated into business-related English for Specific Purposes (ESP) courses, particularly those intended for the hotel and hospitality industries.

## Acknowledgements

The first author would like to express her sincere appreciation for the research funding provided by the University of Newcastle, Australia.

## References

- Anagnostopoulou, S.C., Buhalis, D., Kountouri, I.L., Manousakis, E.G., Tsekrekos, A. E. 2020. The impact of online reputation on hotel profitability. *International Journal of Contemporary Hospitality Management*, 32(1), 20-39.
- Attila, A.T. 2016. The Impact of the Hotel Industry on the Competitiveness of Tourism Destinations in Hungary. *Journal of Competitiveness*, 8(4), 85-104. doi:10.7441/joc.2016.04.06
- Bhatia, V.K. 1993. *Analysing Genre: Language Use in Professional Settings*. London: Longman.
- Calvert, G.A., Pathak, A., Ching, L.E.A., Truflil, G., Fulcher, E.P. 2019. Providing Excellent Consumer Service Is Therapeutic: Insights from an Implicit Association Neuromarketing Study. *Behavioral sciences (Basel, Switzerland)*, 9(10), 109.
- Herring, S.C. 2004. Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning* (pp. 338-376). New York: Cambridge University Press.
- Spencer-Oatey, H. 2008. *Culturally Speaking: Culture, Communication and Politeness Theory*. London: Continuum.
- TripAdvisor. 2017. About TripAdvisor. Media Center. Retrieved from <https://tripadvisor.mediaroom.com/US-about-us>

# Using feature selection to evaluate pathological speech after training with a serious game

Loes van Bommel<sup>1,2</sup>, Catia Cucchiarini<sup>3</sup>, Helmer Strik<sup>3,4,5</sup>

<sup>1</sup>Department of Artificial Intelligence

<sup>2</sup>Institute for Computing and Information Sciences

<sup>3</sup>Centre for Language and Speech Technology

<sup>4</sup>Centre for Language Studies

<sup>5</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

To evaluate the effectiveness of speech therapy, speech features before and after treatment can be compared, focussing on those features that changed most during treatment. In the current study acoustic features were automatically extracted from speech of patients affected by Parkinson's Disease who had received speech treatment. Praat and openSMILE were used for feature extraction. Through feature selection, the top ten most characterizing features for pre vs. post-treatment were found. Further analysis of these features confirmed that after treatment the speakers spoke louder with lower pitch, which were the goals of the treatment.

Keywords: Parkinson's disease, pathological speech, feature selection

## Introduction

Evaluating the effectiveness of speech therapy is a complex issue. First, because there are many measures and procedures to obtain these measures from speech data. Second, because the evaluations are generally based on human ratings, which are time-consuming, error-prone and may contain an element of subjectivity. Objective metrics derived from acoustic measurements would seem to be an interesting alternative, but high quality objective speech measures are difficult to obtain, non-trivial to interpret, and the differences between before and after treatment might be small and non-significant. The extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) (Eyben et al., 2015) is a standardized set of features that were chosen for their demonstrated theoretical relevance, potential to distinguish important aspects of speech production, and ease in automatic computation. In addition, extra features were calculated with a Praat script. In the current study we investigate the usability of these acoustic features to evaluate the effectiveness of speech therapy that was provided through a serious game. We address the following research questions:

- 1) Can we employ automatic feature selection to find the most characterizing features for further analysis out of a large number of measures?
- 2) Do these features indicate the effectiveness of the therapy by showing a decrease in pitch and an increase in loudness?

## Methodology

### Speech data

The speech data were collected from eight native Dutch adults suffering from Parkinson's Disease (PD) who read seven Dutch sentence prompts. The prompts are taken from the story "Papa en Marloes" (Van de Weijer and Slis, 1991) and from apple pie recipes also used by Ganzeboom et al. (2018). There were three recording times: T1 four weeks pre-experiment, T2 pre-treatment, and T3 post-treatment after four weeks of training. The treatment in this experiment is the second version of the serious speech training game "Treasure Hunters" (<https://waag.org/project/chasing> and Ganzeboom et al, 2018). The main goal of the game is to improve intelligibility of speech through Pitch Limiting Voice Treatment, thus increasing loudness but reducing pitch.

### Acoustic features and feature selection

Praat (Boersma and Weenik, 2020) and openSMILE (Eyben et al., 2010) were used to automatically extract 103 features per recording. The 15 Praat features are duration, four formants, pitch variance, gravity center, and the mean, minimum, maximum and standard deviation of pitch and intensity. OpenSMILE was used to extract the 88 eGeMAPS feature set (Eyben et al., 2015). The 103 features were extracted on three different levels: full, word and phoneme. Segmentations on these levels were obtained by a forced aligner. (<https://webservices.cls.ru.nl>).

Classification with Support Vector Machine and feature selection with Recursive Feature Elimination (RFE; Guyon et al., 2002) were used to determine which of the features changed most between pre and post-treatment, and thus showed the highest scores for pre vs. post-treatment classification. The top ten features were analysed in more detail using classification and statistical analysis. Precautions were taken to reduce the risk of overfitting on the small dataset: using a simple model (linear SVM), speaker-based normalization and Leave One Subject Out cross validation (Sakar et al., 2013), and Matthew's Correlation Coefficient (MCC) instead of accuracy.

## Results

Using the feature ranking gained by RFE for the pre vs. post-treatment contrast, the MCC scores per time contrast and segmentation level for linear SVM classification led to the results shown in Figure 1. Six out of the top ten T2 vs. T3 features were found to have a significant time effect with  $p < 0.05$  when tested with a GLM analysis, shown in Table 1.



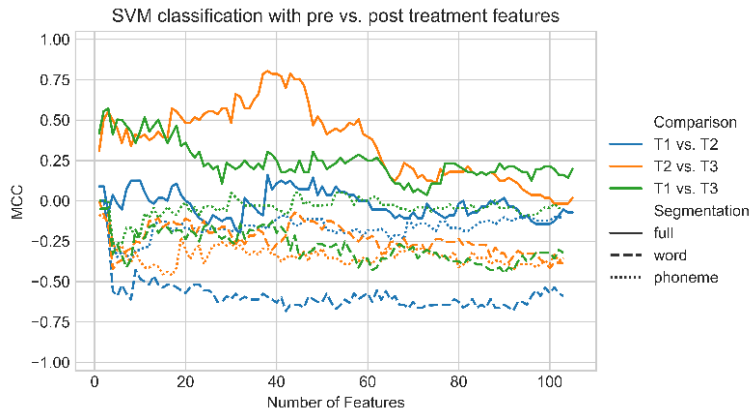


Figure 1. Classification scores of the pre-experiment (T1) vs. pre-treatment (T2) vs. post-treatment (T3) features on three linguistic segmentation levels using SVM classification with a linear kernel and the T2 vs. T3 ranked features.

Table 1. GLM p-values for T2 vs. T3 time effects and mean values for the top ten full-segmentation T2 vs. T3 features.

Rank	Feature	Mean T2	Mean T3	p
1	F3amplitudeLogRelF0_sma3nz_stddevNorm	-0.672	-0.708	**
2	slopeV0-500_sma3nz_amean	0.028	0.030	**
3	F0semitoneFrom27.5Hz_sma3nz_percentile50	29.503	29.140	**
4	HNRdBACF_sma3nz_amean	7.011	6.980	
5	mfcc1_sma3_amean	15.799	17.089	**
6	hammarbergIndexUV_sma3nz_amean	8.822	9.041	
7	mfcc2_sma3_stddevNorm	2.006	-2.635	
8	mfcc2_sma3_amean	4.645	3.989	**
9	mfcc3_sma3_stddevNorm	1.177	1.155	
10	mfcc3V_sma3nz_stddevNorm	0.902	0.957	*

Note: \*\* p < 0.01, \* p < 0.05

## Discussion

The ranked pre vs. post-treatment features reach the highest classification scores around 40 features for the full-segmentation level and for the T2 vs. T3 contrast, while other segmentations and time contrasts result in lower scores. Subsequent analysis of the top ten features, both statistically and manually, confirms that training with “Treasure Hunters” significantly changes the loudness and pitch for speakers in ways that were expected. Aside from the final feature evaluation the entire process is automatic. Future research could include more data, more features, and more classification methods.

## Conclusions

We succeeded in obtaining a feature ranking by using a Recursive Feature Elimination method based on Support Vector Machine classification, which answers our first research question. Statistically significant differences were found for the top ten measures. After treatment, the eGeMAPS features HammarbergIndex and slopeV0-500 had higher values, indicating that loudness had increased, and F0semitone-Median was lower, indicating that pitch was lower. These results provide a positive answer to our second research question.

## References

- Boersma, P. and Weenik, D.. 2020. Praat: doing phonetics by computer (version 6.1.22). <http://www.praat.org>
- Eyben, F., Wöllmer, M., Schuller, B.. 2010. Opensmile: the munich versatile and fast open audio feature extractor. Proceedings of the 18th ACM international conference on Multimedia. 1459-1462.
- Eyben, F., Scherer, K.R., Schuller, B.W., Sundberg, J., André, E., Busso, C., Devillers, L.Y., Epps, J., Laukka, P., Narayanan, S.S.. 2015. The Geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. IEEE transactions on affective computing. 7(2), 190-202.
- Ganzeboom, M., Bakker, M., Beijer, L., Rietveld, T., Strik, H.. 2018. Speech training for neurological patients using a serious game. British Journal of Educational Technology. 49(4), 761-774.
- Guyon, I., Weston, J., Barnhill, S., Vapnik, V.. 2002. Gene selection for cancer classification using support vector machines. Machine learning, 4(6), 389-422.
- Sakar, B.E., Isenkul, M.E., Sakar, C.O., Sertbas, A., Gorgen, F., Delil, S., Apaydin, H., Kursun, O.. 2013. Collection and analysis of parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics. 17(4), 828-834.
- Van de Weijer, J., Slis, I.. 1991. Nasaliteitsmeting met de nasometer. Logopedie en Foniatrie 63(97), 101.

# What is difficult in second language acquisition of syntax

Jiaojiao Yao

Center of Linguistics, University of Lisbon, Portugal

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

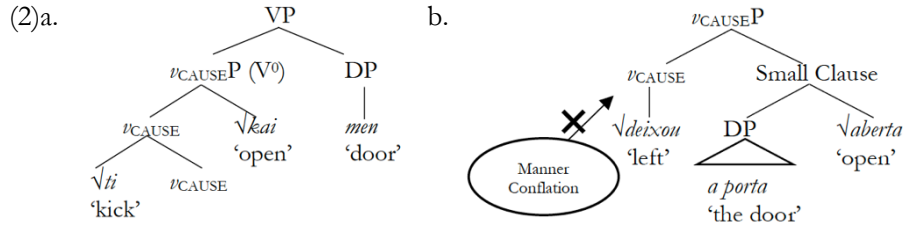
This study investigates the second language (L2) acquisition of Chinese Causative Resultative V-Vs (CR V-Vs) by Portuguese-speaking learners through a Semi-Inducted Production Task (SPT) and a Grammaticality Judgment Task (GJT). While the results of the SPT indicate a proficiency effect in the CR V-V production, the results of the GJT are somewhat mixed. Some CR V-V constraints have been successfully acquired by L2 learners at least from the intermediate level, but others remain difficult even for advanced learners. The results show that with more L2 exposure, parameters may be reset by L2 learners, but successful L2 acquisition requires more than that. The bottleneck is beyond the parameter resetting.

Key words: second language acquisition, syntax, resultative, Chinese, Portuguese

## Introduction

The Chinese Causative Resultative V-Vs (CR V-Vs), as in (1a), express caused-result events, with the Manner-denoting component (V1) and the result-denoting component (V2) in adjacency. Within the Minimalist Program and Distributed Morphology framework, we claim that Chinese CR V-Vs involve the head  $v_{\text{CAUSE}}$ . The Manner-denoting root ( $\sqrt{1}$ ) conflates to this head as an adjunct, and the Result-denoting root ( $\sqrt{2}$ ) is incorporated into it as its Complement, as illustrated in (2a). In contrast, Portuguese allows simple resultatives with light verbs but not true resultatives with Manner (1b). We claim that Portuguese simple resultatives involve embedded Small Clause, as illustrated in (2b). However, since Manner Conflation is not allowed, the Manner can only be expressed externally, as in (1c).

- (1) a. Ta ti kai le men.  
he kick open ASP door  
'He kicked the door open.'  
b. O João deixou/\*pontapeoua porta aberta.  
the John left/\*kicked the door open  
'John left/\*kicked the door open.'  
O João abriu a porta pontapeando-a.  
the John opened the door kicking it  
'John opened the door (by) kicking it.'



Therefore, for L1 Portuguese learners to acquire L2 Chinese CR V-Vs, parameter resetting will be required. In particular, the L2 learners should allow  $\nu_{\text{CAUSE}}$  to directly select a root (instead of Small Clause) and “switch on” the Manner Conflation option.

Our research questions are: Are the L1 Portuguese L2 Chinese learners successful in acquiring Chinese CR V-Vs? What can account for the variation?

## Methods and materials

### Participants

The participants in our experiment include 27 L1 Portuguese L2 Chinese learners (18 intermediate and 9 advanced learners) and 27 Chinese native controls. All the participants are adults.

### Materials and procedures

The experiment includes a Semi-Induced Production Task (SPT) and a Grammaticality Judgment Task (GJT).

The stimulus set in the SPT is composed of 20 target video clips and 10 distracting clips. Each target video clip depicts a caused-result event and is accompanied by a set of keywords as cues. After watching each clip, the participants were asked to say a sentence in Chinese using the provided keywords. All the responses were recorded and transcribed. Notes were taken based on the construction type that was used for each video clip. The frequency of each structure type in each group was calculated.

The GJT includes 35 target items (20 grammatical and 15 ungrammatical) and 17 distracting items. All the target items contain CR V-Vs, including pairs of items that only differ minimally – one is grammatical, and the other violates a particular CR V-V constraint (e.g., the semantic constraint of V2, the transitivity alternation, the V-V adjacency, the V-V integrity, and the “small size” constraint). The participants were asked to rate the sentences focusing on grammaticality by choosing a value on a Likert scale from 1 (completely unacceptable) to 5 (completely acceptable).

All materials were pilot-tested prior to the actual gathering of the data. The items in each task were randomized so that similar items would not be adjacent to each other. All the instructions were given in the participants’ mother

tongues. The participants were required to complete the tasks independently. There was no limitation of time for the test completion.

### Results

The CR V-V frequency by each group in the SPT is presented in the boxplots in Figure 1. The native controls (M=91.85%) outperformed both L2 learner groups, and the advanced L2 group (M=46.67%) outperformed the intermediate group (M=20.83%). Through Welch's t-tests (see Table 1), we found that the difference is significant between all groups ( $p < 0.05$ ).

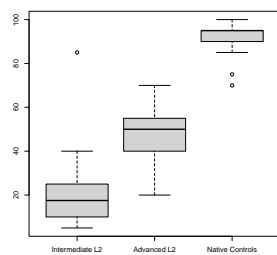


Figure 1 SPT boxplots.

Table 1 CR V-V frequency comparison.

Welch's t-test	t	df	p-value
Inter. × Adv.	-3.8821	20.149	0.0009162
Adv. × Native	-8.8211	9.3054	8.063e-06
Inter. × Native	-3.8821	20.149	0.0009162

Results of the GJT show that the contrast between the acceptable and unacceptable items is the biggest in the native control group and the smallest in the intermediate L2 group (see Figure 2).

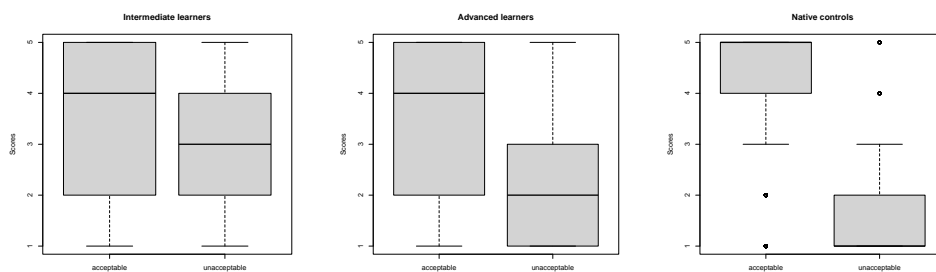


Figure 2 Boxplots of grammatical and ungrammatical items.

We ran Welch's t-tests and found that there was a significant difference between grammatical and ungrammatical items in all groups (Inter:  $t=6.3025$ ,  $p=5.839e-10$ ; Adv:  $t=7.4094$ ,  $p=1.268e-12$ ; Native:  $t=45.305$ ,  $p<2.2e-16$ ).

Looking at specific constraints, we found that both the intermediate and the advanced L2 groups were sensitive to the semantic constraint of V2 ( $t=3.3629$ ,  $p=0.001438$ ;  $t=4.1626$ ,  $p=0.0003056$ ), the V-V integrity ( $t=6.0678$ ,  $p=6.287e-07$ ;  $t=4.0817$ ,  $p=0.0007769$ ), and the transitivity alternation constraints (causative:  $p=0.006667$ ,  $0.006275$ ; anticausative:  $p=0.009485$ ,  $0.01159$ ). As to the "small size" constraint, although neither L2 group showed strong sensitivity ( $t=-1.1578$ ,  $p=0.2548$ ;  $t=0.4165$ ,  $p=0.6823$ ), there was a weak tendency that the score difference in the acceptable-unacceptable pairs is bigger in the advanced group than in the intermediate group (Inter×Adv:  $t=-1.0521$ ,  $p=0.2993$ ). However, neither L2 group showed knowledge of the V-V adjacency ( $t=0$ ,  $p\approx 1$ ;  $t=-0.10422$ ,  $p=0.9182$ ), and no proficiency effect was detected (Inter×Adv:  $t=0.092348$ ,  $p=0.9271$ ). An interesting result was also found in the Type VIII constraint: while the intermediate L2 group showed sensitivity ( $t=3.4236$ ,  $p=0.001592$ ), the advanced group did not ( $t=0.82328$ ,  $p=0.4217$ ).

## Conclusions

The overall results of our experiment show that there is a general tendency that the L2 learners gained more knowledge of Chinese CR V-Vs as proficiency increased. However, when looking at the results in more depth, we found that different CR V-V constraints were not acquired at the same speed. Some constraints have been acquired successfully at least from the intermediate level (e.g., the semantic constraint on V2), while constraints such as the V-V adjacency seem to impose great difficulty even for advanced L2 learners. Therefore, different aspects of L2 grammar form a hierarchy of acquisition difficulty.

The L2 acquisition process is more complicated than selecting the correct values for parameters. As Lardiere (2008, 2009) hypothesized, the difficulty exists in Feature Assembly, namely, learning how the bundles of features should be realized in L2. While the results of our study showed a generally positive learning curve regarding parameter resetting, they also showed that successful acquisition requires more than that – the bottleneck is beyond parameter resetting.

## References

- Lardiere, D. 2008. Feature assembly in second language acquisition. In Liceras, J., Zobl, H., Goodluck, H. (eds.) 2008, *The Role of Formal Features in Second Language Acquisition*, 106–140. New York, Lawrence Erlbaum.
- Lardiere, D. 2009. Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25(2), 173–227.



interference effects). Meanwhile, Warren & Gibson (2002, 2005)'s complexity rating study, observed that the parser is sensitive to the gradient status of a distractor in discourse, following the Givenness Hierarchy ("GH": Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993). For example, a distractor that is most central in the discourse (e.g. *pronoun*) causes the least processing cost, followed by less central NPs on the hierarchy (e.g. *definite, indefinite descriptions*) [pronouns > first names > full names > definites > indefinites] (Warren & Gibson, 2002: p.87). Gordon et al (2001)'s experiment therefore does not allow for teasing apart the givenness hierarchy effects from similarity-based effects. On this view, this paper explores whether the interference effect of a distractor is truly a similarity effect or is in fact a more fine-grained discourse-level of the semantic hierarchy, or both.

## Experiment 1

This experiment had a 2 x 3 design, crossing two NP types of the filler in the clefted (NP1) position and three NP types of a distractor in the embedded NP (NP2) position: [definite descriptions, indefinite descriptions] x [definite descriptions, indefinite descriptions, pronouns].

- (2) It was {the actor/an actor} who {we /the director/a director} graciously thanked before the show.

The similarity-based interference hypothesis predicts that matching NP2s which featurally match with NP1s should lead to increased processing difficulty, as reflected in longer reading times at the main verb (e.g. *thanked*). The discourse hierarchy, in contrast, predicts a main effect of NP2 type: the fastest reading times when NP2 is a pronoun (e.g. *we*) condition, longer when it is a definite description (e.g. *the director*), and longest when it is an indefinite description (e.g. *a director*).

Thirty-six native speakers of English were recruited using Amazon's Mechanical Turk. The task was a self-paced reading using a moving window display. Participants read items like (2), each of which was followed by a comprehension question. The experiment lasted approximately 25 min. Experimental materials consisted of 24 sets of 4 items in each of the 6 conditions, plus 26 filler sentences.

A linear mixed-effects model revealed no main effect of NP1 type ( $t=-0.62$ ,  $p=.54$ ) at the critical verb region (e.g., *thanked*). In terms of the NP2 type, the pronoun condition was approximately 60ms faster than the indefinite condition and 90ms faster than the definite condition. The model revealed a significant effect of NP2, in that the pronoun condition was read significantly faster than the averaged definite and indefinite conditions ( $t=-3.60$ ,  $p < .001$ ). Surprisingly, the overall reading times of definiteness conditions in NP2 type averaged 29ms faster than the overall reading times of indefiniteness conditions. A marginal



effect of definiteness was found between the definite and indefinite conditions ( $t=1.78, p=.07$ ).

These findings cannot be not fully accounted for either by the GH or the similarity-based interference effect. The GH predicted more processing difficulty of indefinite descriptions than definite descriptions. The similarity-based interference effect predicts higher reading times in the definite-definite condition and, the indefinite-indefinite condition, which we did not observe. We conclude that the difficulty of definiteness found by Gordon et al. is due *neither* to similarity *nor* givenness and hypothesize that it instead arises from the absence of a prior context which would provide a (unique) referent for a definite NP, thus making difficult to satisfy the uniqueness presupposition for the (English) definite determiner (Löbner, 1985). We test this in Experiment 2.

## Experiment 2

Experiment 2 focused exclusively on the definite-indefinite contrast, aiming to examine whether the difficulty of definiteness in Experiment 1 is attenuated with the right contextual support. We included contexts favoring either definite intervenors (unique referent) or indefinite intervenors (two possible referents) and crossed this with NP2 type (definite or indefinite), in a 2x2 design in which NP1 was always definite, as in Table 1.

Table 1. Sample item set for Experiment 2

	Indefinite NP2	Definite NP2
<u>Context favoring <b>DEF</b> NP2:</u> John, an actor, and Matt, <b>a director</b> , were at the coffee shop. They met Andy, <b>a writer</b> , who graciously thanked John for his incredible performance in the latest movie.	Target TRUE/FALSE sentence:  It is the actor [who <b>a</b> writer graciously thanked for the incredible performance].	Target TRUE/FALSE sentence:  It is the actor [who <b>the</b> writer graciously thanked for the incredible performance].
<u>Context favoring <b>INDEF</b> NP2:</u> John, an actor, and Matt, <b>a writer</b> , were at the coffee shop. They met Andy, <b>another writer</b> , who graciously thanked John for his incredible performance in the latest movie.		

We created 16 sets items, presented in 4 lists according to Latin square. Participants (n=33, recruited on Murk) read the context (as a whole) and then the target TRUE/FALSE sentence, (region-by-region). At the verb, there were no main effects either of context type or of NP2 (i.e., distractor) type on RTs ( $p_s > .25$ ), but the interaction was significant ( $t= -2.5; p= .013$ ), meaning that the

effect of definiteness of NP2 differed according to context type. Planned comparisons showed no effect of NP2 type in the contexts favoring **definites** ( $p = .4$ ) but in the contexts favoring **indefinites**, the indefinite distractor condition was read significantly faster than the definite intervenor condition (431ms vs 523ms,  $t=-2.5$ ,  $p=.007$ ). We interpret this finding to indicate that the processing difficulty of definite distractors, which need to accommodate the presupposition, arises from the absence of a unique referent (as in indefinite-favored contexts). This effect disappears when contextual support provides an appropriate referent (as in definite-favored contexts).

### General discussion

The overall findings show that NP types of distractors modulate the processing of filler-gap dependencies. However, neither (i) similarity nor (ii) givenness underlies the difficulty of processing definite distractor NPs. More fine-grained discourse of an intervenor, involving the accommodation of uniqueness presupposition, mediates filler-gap dependency formation of this type.

### References

- Ariel, M. 1990. *Accessing Noun-phrase Antecedents*. London, Routledge.
- Gordon, P.C., Hendrick, R., Johnson, M. 2001. Memory interference during language processing. *Journal of Experimental Psychology* 27, 1411-1423.
- Gundel, J. K., Hedberg, N., Zacharski, R. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69, 274-307.
- Lewis, R.L., Vasishth, S. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29, 375-419.
- Löbner, S. 1985. Definites. *Journal of Semantics* 4, 279-326.
- Warren, T., Gibson, E. 2002. The influence of referential processing on sentence complexity. *Cognition* 85, 79-112.
- Warren T., Gibson, E. 2005. Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes* 20, 751-767.

# The Production of Emphatic Stress Test for Korean

Jihyeon Yun<sup>1</sup>, Jee-Hye Chung<sup>2</sup>, Yeong-Wook Kim<sup>3</sup>, Il-Young Jung<sup>2,3</sup>

<sup>1</sup>Department of Linguistics, Chungnam National University, Republic of Korea

<sup>2</sup>Department of Rehabilitation Medicine, Chungnam National University Hospital, Republic of Korea

<sup>3</sup>Department of Rehabilitation Medicine, Chungnam National University Sejong Hospital, Republic of Korea

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

The Production of Emphatic Stress Test in the Right Hemisphere Language Battery is revised for the Korean version to elicit prosodic prominence more effectively using corrective focus. The mean score obtained by 80 Korean speakers without language disorders was 4.88 ( $\pm 0.33$ ) for the revised subtest and 4.56 ( $\pm 0.67$ ) for the subtest in its original format out of 5 points for each. The variety of the speakers' responses in the current study anticipates the intricacy of assessing prosody for clinical evaluation.

Keywords: focus, prominence, phrasing, RHLB, Korean

## Introduction

The Right Hemisphere Language Battery (RHLB; Bryan 1995) is a tool developed to test language comprehension and production that are pertinent to the function of the right cerebral hemisphere. The RHLB contains a test for prosody. The original Production of Emphatic Stress Test, however, has a limitation regarding direct application to the Korean language because of cultural and linguistic differences.

Unlike English, which is a head-prominence language, phrasal boundaries mark prominence in Korean (Jun 2011). Language-specific strategies exploiting morpho-syntactic alternatives can also come into play to express contrastive information. Examining the production of focus prominence remains a practical procedure to assess (semi-)spontaneously produced speech prosody in clinical settings. Hence, the original emphatic stress test was retained with a revision in the Korean version of RHLB (K-RHLB, in development). We describe the structure of the K-RHLB's emphatic stress test and report the test scores of speakers without language disorders as a reference. Finally, observed intonation patterns and empirical criteria for assessing focus prominence are discussed.

## Methods

### Materials

The Production of Emphatic Stress Test of the K-RHLB consists of the following two parts, with five items for each.

#### Emphatic Stress Test Ga (A)

This subtest has five items for inducing corrective focus, with one preceding practice item at the beginning. For each item, a picture is provided for reference, and an examiner asks a yes or no question including a word that does not conform to the picture. Participants are asked to respond with *ani-o* ‘no’ and then correct the faulty component. An (expected) example sentence for an examiner and a test subject is given below, respectively, in Yale Romanization of Korean.

(1) Practice Item

Examiner:	yeca-ka	koki-lul	sa-le	ka-yo?
	woman-nom	meat-obj	buy-part	go-hon
	‘Does the woman go for buying meat?’			
Subject:	ani-o,	yeca-nun	yachay-lul	sa-le ka-yo.
	neg-hon	woman-top	vegetable-obj	buy-part go-hon
	‘No, the woman goes for buying vegetables.’			

#### Emphatic Stress Test Na (B)

This subtest consists of five items, preceded by one practice item, translated from the original RHLB. In each item, two pictures are presented side by side. After the examiner says the first half of the sentence, referring to the first picture, the participant completes the second half of the sentence based on the contrastive components depicted in the second picture. The participant is expected to produce the rest of the sentence, introducing two or three words with new information. A pair of example (target) phrases for an examiner and a subject is as follows:

(2) Practice Item

Examiner:	namca-nun	cenhwa-lul	ha-ci	anh-ko,
	man-top	telephone-obj	do-part	neg-conj
	‘The man does/did not make a phone call, but’			
Subject:	pyenci-lul	sse-ss-e-yo.		
	letter-obj	write-past-decl-hon		
	‘wrote a letter.’			

### Procedures

Each participant was tested individually by an examiner trained in linguistics and phonetics. A series of pictures for eliciting speech was presented on a monitor, and the participant’s response was recorded. Evaluation scores collected from 80 native speakers of Korean without language disorders were analyzed [57 female and 23 male; mean age: 36.1 ( $\pm$ 12.6)]. They are from the

Seoul/Gyeonggi (26) or Daejeon /Chungcheong (54) region where they lived while under the age of 15. The speakers' dialects have no lexical tone or lexical pitch accent.

## Results

Each speech item received one point for the appropriate stress. The mean score for subtest A is 4.88 ( $\pm 0.33$ ) and that for subtest B is 4.56 ( $\pm 0.67$ ), out of 5 for each. The total mean score for both subtests is 9.44 ( $\pm 0.74$ ) out of 10. The stress pattern varied by sample, but the most salient acoustic correlate for stress was the fundamental frequency in the first or second syllable of the targeted unit or in the right boundary of the Accentual Phrase (AP; cf. Jun 2011 for the K-ToBI convention, i.a.). The particle, rather than the target morpheme itself, was accentuated frequently (cf. Fig. 1). Dephrasing after the focused word and drastic lowering of the pitch after the focused phrase were also confirmed.

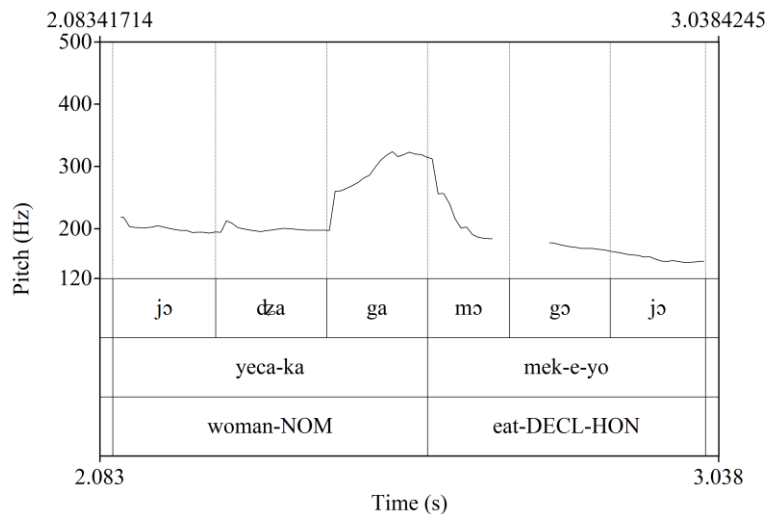


Figure 1. Prominence of the particle *-ga* [nominative] following the corrective target word *yeca* 'woman'.

## Discussion

Speakers were better at subtest A with corrective focus than at subtest B with contrastive focus. However, the overall score (mean = 9.44) is lower than that for the English-speaking control group's (30 neurologically normal subjects) emphatic stress test in the original battery (Bryan 1995), the mean of which was 10 ( $\pm 0$ ) out of 10, indicating perfect performance.

Whereas many experimental studies on the production of focus prosody in Korean are conducted either based on reading tasks or with repetitive practice, the extemporary responses in the present test varied from utterance to

utterance, both in morpho-syntactic composition and in prosodic realization. Emphasis was expressed through a combination of three different manners in the speakers' utterances.

First, speakers omitted components with given information. It is usual for Korean speakers to omit the sentence's subject or postpositions, especially in colloquial speech. Meanwhile, the answer ellipsis is likely to bring the most emphasized word toward the utterance-initial position. Although subtest A is devised in order to elicit corrective focus prominence within a sentence with about 3-4 APs, speakers sometimes had difficulty maintaining the original sentence structure given in the examiner's question. This is possibly not only due to cognitive load but also to the unnaturalness of repeating the components that are out of focus.

Second, speakers used morpho-syntactic composition to highlight the target unit. One simple example is that a speaker says *pakke-ey-nun pi-ka wa-yo* 'It is raining outside.' instead of *pakke-ey pi-ka wa-yo*. Here, the insertion of the particle *-nun* [topic], without being focused itself, can have the effect of emphasizing the succeeding constituent with regard to *pakke-ey* 'outside'. Another type of syntactic modulation is the change of word order.

Third and finally, the focus prominence could be expressed and perceived in prosody. Pitch was a salient cue signaling phrasing among other phonetic cues within an Intonation Phrase (Jun 2011, i.a.), in part by definition. Often, the prominence was found in the postposition rather than the target word itself, as in Fig. 1 above.

The current result shows that emphatic stress can be expressed in various ways. The prosodic realization of focus in Korean may not be as clear and consistent as in English (cf. discussion on the perception of focus prominence by Korean listeners in Lee, Cho 2020). The same fact also implies difficulty in testing prosody for Korean. This calls for objective criteria for the evaluation of intonation in clinical settings.

## Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF-2021R1C1C1005188) to IJ and Chungnam National University Sejong Hospital Research Fund 2021 (2021-S2-005) to YK.

## References

- Bryan, K. 1995. The Right Hemisphere Language Battery. London, Whurr.
- Jun, S.A. 2011. Prosodic markings of complex NP focus, syntax, and the pre-/post-focus string. In Washburn M. B. et al. (eds.), Proc. 28<sup>th</sup> West Coast Conference on Formal Linguistics, 214-230, Somerville, MA.
- Lee, Y.C., Cho, S. 2020. Focus Prosody Varies by Phrase-Initial Tones in Seoul Korean: Production, Perception, and Automatic Classification. *Languages* 5(4), 64.

# Profiling the central auditory processing of speech sounds: a MMN multifeature study

Maria Ioanna Zavogianni<sup>1,2</sup>, Ferenc Honbolygó<sup>1,3</sup>

<sup>1</sup>Brain Imaging Centre, Research Centre for Natural Sciences, Hungary

<sup>2</sup>Multilingualism Doctoral School, University of Pannonia, Hungary

<sup>3</sup>Institute of Psychology, ELTE Eötvös Loránd University, Hungary

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

Multifeature paradigms record fast Mismatch Negativity (MMN) ERP responses and investigate the central auditory processing. Our speech multifeature paradigm included changes in vowel, consonant, word stress, frequency and intensity. We investigated adults' involuntary discrimination of speech sound features which are phonetically or acoustically altered. Participants watched a film without sound while pseudowords were presented through headphones. The deviants elicited MMN responses with different amplitudes. The vowel and consonant elicited the largest MMNs, while the intensity and pitch deviants the smallest. Word stress deviants elicited a double negativity. This multifeature paradigm could be utilized to record MMN, and allows to apply the MMN as assessment tool for profiling the central auditory processing of speech sounds.

Keywords: MMN, multifeature paradigm, speech, central auditory processing

## Introduction

The Mismatch Negativity (MMN) is an Event-Related Potential (ERP) component which is elicited in an auditory environment. According to Näätänen et al. (2007), the MMN is evoked by a detectable change (i.e., a 'deviant' stimulus) during the presence of regular features (i.e., a 'standard' stimulus). The experimental design which records MMN responses is the speech multifeature paradigm, during which one 'standard' stimulus and several 'deviant' stimuli are presented (Pakarinen et al., 2009). Previous studies have applied the multifeature paradigm with focused alterations on the phonetic or acoustic parameters of the syllables of the certain stimuli; these studies reported that all changes in the sound stimuli allowed the statistically significant elicitation of the MMN responses (Pakarinen et al., 2009; Pakarinen et al., 2013). The aim of our study was to investigate the MMN correlates of speech sound processing in a paradigm that allows the fast recording of ERP responses to different kind of speech features.

In the current study, we constructed a multifeature paradigm with one standard stimulus and five deviant stimuli with phonetic (i.e., vowel, consonant, prosody) or acoustic (i.e., frequency and intensity) changes. This study is the first to involve a change in prosody to the multifeature paradigm. We expected

to find that these speech sound features were detectable in an involuntary manner.

## **Materials and methods**

### **Participants**

Twenty-five native Hungarian speakers (14 female) participated in our study. Their mean age was 25 years (SD:  $\pm 5.6$ , range: 19-38); none of them reported any neurological issue, hearing or speech-related issue.

### **Stimuli and procedure**

The stimuli were CVCV pseudowords. The standard stimulus was a four-syllabic Hungarian pseudoword (/ 'keke/). The deviant stimuli were formulated through the alterations to the standard stimulus at the first syllable. The deviants consisted of a vowel (/ 'kike/), a consonant (/ 'peke/) and a stress deviant, where the stress was on the second syllable (/ 'keke/); additionally, the deviant stimuli consisted of a frequency ( $F_0 \pm 8\%$ ) and an intensity deviant ( $\pm 6$  dB).

Stimuli were presented in a multifeature paradigm with a stimulus onset asynchrony (SOA) of 750 ms. Participants were auditorily presented with 315 standard and 300 deviant stimuli; the experiment began with fifteen standard stimuli and the deviants were randomly presented to participants. During the EEG experiment, participants were required to listen to the stimuli while watching a movie with no sound, which they were asked not to be attentive to. The total duration of the recording time was 17 min.

### **EEG Recording and data analysis**

The EEG was recorded using a 64-channel recording system (BrainAmp amplifier and BrainVision Recorder software, BrainProducts GmbH). The sampling rate was 1000 Hz and electrode Fz was used as a reference. The EEG data were analysed with Brain Vision Analyzer. To test the significant appearance of the MMN component, one-sample t-tests were performed between 176 and 500 ms from the deviation onset at Fz electrode; difference curves of the ERP waveforms were calculated by subtracting the ERPs of the standard condition from the ERPs of the deviant condition.

## **Results**

All deviant conditions elicited significant MMN peaks (Figure 1) in different time windows. A detailed statistical description of the ERP results can be found in Table 1.



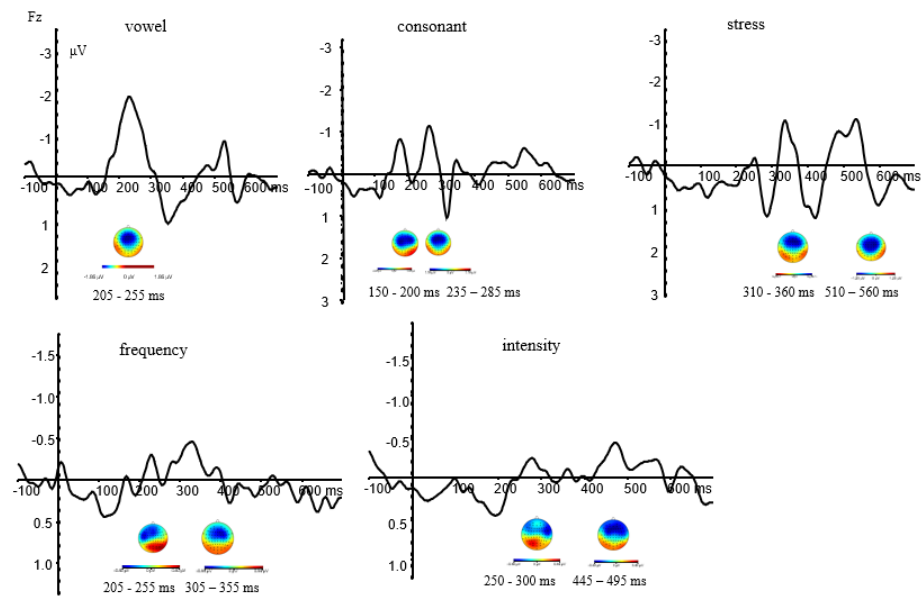


Figure 1. Deviant -standard subtraction waveforms for the five deviant conditions (vowel, consonant, stress, frequency, intensity) at Fz electrode.

Table 1. One-sample t-tests of the EEG difference curves.

Conditions	Time windows	Mean	Confidence Interval		t
			Lower	Upper	
vowel	205-255 ms	-1.66	-1.96	-1.35	-11.16***
early consonant	150-200 ms	-.529	-.86	-.19	-3.28**
late consonant	235-285 ms	-.79	-1.22	-.36	-3.80**
early stress	310-360 ms	-.61	-.95	-.26	-3.64**
late stress	510-560 ms	-.81	-1.09	-.52	-5.81***
early pitch	205-255 ms	-.14	-.45	.17	-.92 n.s.
late pitch	305-355 ms	-.37	-.61	-.13	-3.22**
early intensity	250-300 ms	-.16	-.47	.16	-1.02 n.s.
late intensity	445-495 ms	-.33	-.53	-.12	-3.28**

Note: \*\* $p < .01$ . \*\*\* $p < .001$ . n.s.: non significant

## Discussion

The results showed that all conditions elicited MMN responses, each of which were shown at different time windows. Like in previous studies (Pakarinen et al., 2009), vowel and consonant deviants elicited the most significant MMN

responses; these stimuli facilitated the auditory processing of the participants since they could discriminate them in a more significant way than the rest of the stimuli. Furthermore, as previously reported, the stress deviant elicited a double negativity. Pitch and intensity deviants did not evoke significant MMN responses at the early time windows, a finding which is inconsistent with previous literature (Pakarinen et al., 2009; Pakarinen et al., 2013). Pitch and intensity alterations might not be as detectable by Hungarian speakers as by Finnish. The significant activation of the late time windows indicates the activation of later cognitive processes. To conclude, our speech multifeature paradigm can be applied to the investigation of adults' central auditory processing; it can also provide a further insight to underlying deficits (e.g., phonological processing) of reading disorders.

### **Acknowledgements**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant agreement No.813546.

### **References**

- Näätänen, R., Paavilainen, P., Rinne, T., Alho, K. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology*, 118(12), 2544-2590.
- Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., Kujala, T. 2009. Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biological psychology*, 82(3), 219-226.
- Pakarinen, S., Teinonen, T., Shestakova, A., Kwon, M. S., Kujala, T., Hämäläinen, H., Näätänen, R., Huotilainen, M. 2011. Fast parametric evaluation of central speech-sound processing with mismatch negativity (MMN). *International Journal of Psychophysiology*, 87(1), 103-110.

# Thematic roles in dementia: the case of psychological verbs

Eleni Zimianiti<sup>1</sup>, Maria Dimitrakopoulou<sup>2</sup>, Anastasios Tsangalidis<sup>2</sup>

<sup>1</sup> Department of Languages, Literature and Communication, Utrecht University, Netherlands

<sup>2</sup> Department of Theoretical and Applied Linguistics, Aristotle University of Thessaloniki, Greece

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This study investigates the difficulty of people with Mild Cognitive Impairment (MCI), mild and moderate Alzheimer's disease (AD) in the production and comprehension of psychological verbs, as thematic realization may involve both the canonical and non-canonical realization of arguments. More specifically, we aim to examine whether there is a deficit in the mapping of syntactic and semantic representations in psych-predicates regarding Greek-speaking individuals with MCI and AD, and whether the linguistic abilities associated with  $\theta$ -role assignment decrease as the disease progresses. Moreover, given the decline of cognitive abilities in people with MCI and AD, we explore the effects of components of memory (Semantic, Episodic, and Working Memory) on the assignment of thematic roles in constructions with psychological verbs.

Keywords: thematic roles, psychological verbs, Mild Cognitive Impairment (MCI), Alzheimer's disease (AD), sentence-picture matching

## Introduction

Assignment of  $\theta$ -roles has been shown to be problematic for populations with AD, especially in the case of psychological predicates which seem to add a further level of difficulty (see Manouilidou et al., 2009, a.o., for studies conducted on English-speaking populations). However, the number of studies regarding such populations is limited and does not involve a wide range of languages. Even fewer studies have examined  $\theta$ -role assignment in populations with MCI, which in most cases is the precursor of a dementia type (Campbell et al., 2013). Previous work examining linguistic deficit in people with Mild Cognitive Impairment (MCI) demonstrated that language impairment begins before entering dementia.

This study aims to address the following research questions: (1) Is there a deficit in the mapping between syntactic and semantic representations in populations with MCI and AD and especially in psychological verbs? (2) Is there the same degree of severity in assigning  $\theta$ -roles among MCI and AD groups? (3) Are the linguistic difficulties of the population with MCI and AD related to components of memory (working, semantic and episodic)?

## Methodology

### Participants

Twenty monolingual Greek-speaking people with MCI, sixteen with mild AD, eleven with moderate AD, and forty-seven matched healthy individuals participated in the study (see Table 1).

Table 1. Participant profile. Group means for the 3 experimental groups.

	MCI (N=20)	AD mild (N=16)	AD moderate (N=11)
Age	74.70	75.00	78.55
Education	11.30	9.69	8.18
MMSE	27.55	20.38	14.09
CDR	1.175	4.850	10.591
FUCAS	43.40	50.73	0
MoCA	23.70	22.14	73.36

### Tests

Two Sentence-Picture Matchings Tasks (SPMTs) aiming to assess the comprehension and the production of thematic roles were administered to the participants. They included verbs that follow the canonical thematic hierarchy (e.g. *The man pushes the woman*) and verbs that follow non-canonical thematic *Theme < Experiencer* argument realization (e.g. *frighten, surprise*) resulting in 24 stimulus sentences. The Greek version of the Rivermead Behavioral Memory Test (RBMT; Efklides et al., 2002), and the Working Memory Test with digit forward and backwards tasks (Wechsler, 1987) were also administered prior to the main experimental tasks aiming to investigate memory effects. In particular, RBMT measured the episodic and semantic components of memory and the digit tasks measured the working memory.

### Results

Data analysis identified a deficit in the assignment of thematic roles with greater difficulty in sentences involving the *Experiencer* thematic role or otherwise involving psychological verbs. When compared to the healthy population group, scores of population with MCI and AD revealed that there is a gradual decrease in performance in both production and comprehension, as the disease progresses (see Figures 1 and 2). Moreover, the performance of dementia groups has proven to be even worse in sentences involving a psychological verb in both levels of production and comprehension, as shown in Figures 3 and 4. With regard to the possible effects of the components of memory on language performance, correlation analyses showed a strong correlation between language performance and both semantic and episodic memory, but a weak one as regards working memory.

Production

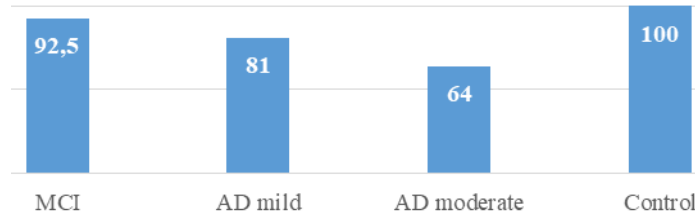


Figure 1. Accuracy of Production.

Comprehension

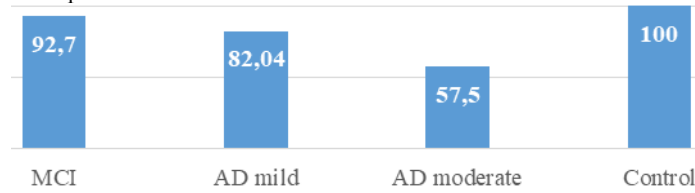


Figure 2. Accuracy of Comprehension.

Production

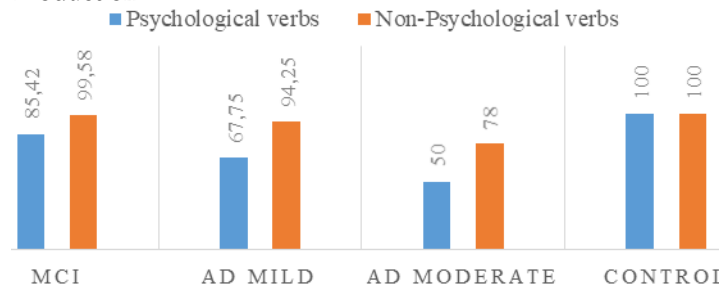


Figure 3. Production of Psychological and Non-Psychological verbs.

Comprehension

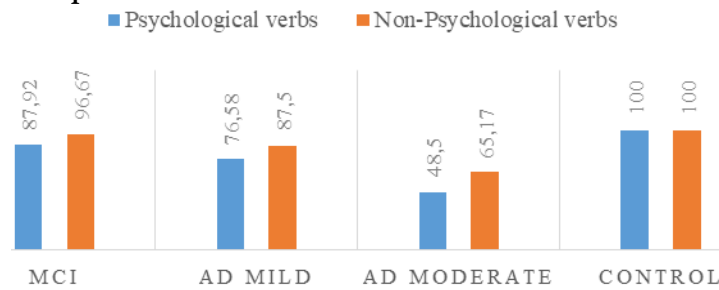


Figure 4. Comprehension of Psychological and Non-Psychological verbs among groups.

## Discussion

The findings of this study suggest that there is gradual degradation of assigning thematic roles as the disease progresses, especially in sentences involving psychological verbs. The overall picture of the participants with MCI and AD reveals that  $\theta$ -role assignment in sentences deviating from canonical thematic hierarchy (psychological verbs) creates confusion in dementia populations. Their evident semantic deficit impacts on the syntax-semantics interface, which leads to the observation that the deficit is more of a general one in mapping semantic participants onto syntactic structures rather than a limited one.

With regard to the possible effects of the components of memory on language performance, the weak relationship between linguistic abilities and working memory contradicts previous work in this field. This result may be due because of the memory training that participants with MCI and AD have been receiving, leading to higher performance than expected. The strong relationship of language performance and both semantic and episodic memory supports the significant involvement of semantic memory in processing argument structure (Grossman and White-Devine, 1998) and provides further evidence for the necessity of more efficient and in-depth co-examination.

## Acknowledgements

We would like to thank the scientific personnel of the Day Care Centers “Agia Eleni” and “Agios Ioannis” of the Greek Association of Alzheimer’s Disease and Related Disorders (Alzheimer Hellas) in Thessaloniki for providing access to their pool of participants and performing their neuropsychological evaluation.

## References

- Campbell, N.L., Unverzagt, F., LaMantia, M.A., Khan, B.A., Boustani, M. A. 2013. Risk factors for the progression of mild cognitive impairment to dementia. *Clinics in geriatric medicine*, 29(4), 873–893. <https://doi.org/10.1016/j.cger.2013.07.009>
- Efklides, A., Yiulsi, E., Kangellidou, T., Kounti, F., Dina, F., Tsolaki, M. 2002. Wechsler Memory Scale, Rivermead Behavioral Memory Test, and Everyday Memory Questionnaire in healthy adults and Alzheimer's patients. *European Journal of Psychological Assessment*, 18(1), 63.
- Grossman, M., White-Devine, T. 1998. Sentence comprehension in Alzheimer’s disease. *Brain and Language*, 62, 186-201.
- Manouilidou, C., de Almeida, R.G., Schwartz, G., Nair, N. P.V. 2009. Theta roles in Alzheimer’s disease: Hierarchy violations in psychological predicates. *Journal of Neurolinguistics*, 22, 167-186.
- Wechsler, D. 1987. *Manual for the Wechsler Memory Scale-Revised*. The Psychological Corporation, San Antonio, TX.

# Oral and silent reading in adolescents: evidence from Russian

Vladislav Zubov<sup>1</sup>, Tatiana Petrova<sup>2</sup>, Svetlana Alexeeva<sup>2</sup>

<sup>1</sup>Ludmila Verbitskaya Department of General Linguistics, St Petersburg University, Russia

<sup>2</sup>Institute for Cognitive Studies, St Petersburg University, Russia

<https://doi.org/10.36505/ExLing-2021/12>

## Abstract

This study aimed to answer two questions: what type of reading contributes to more successful text processing and understanding, and what are the main characteristics of silent and oral reading for Russian-speaking secondary school children. Results show that while reading orally, participants with reading disorders comprehend the text poorly. They have a limited amount of attentional resources available for any given cognitive tasks, and thus, the greater the amount of attention they paid to text processing, the less is available for text understanding. However, in silent reading, the same participants demonstrate good comprehension if they read slowly, i.e. make more fixations and regressions. Their cognitive resources are spent on text comprehension, which is eventually what reading is aimed at.

Keywords: reading aloud, silent reading, Russian, adolescences, eye-tracking

## Introduction

According to Adams (1990), the ability to read text fluently is one of the essential requirements for successful reading comprehension. Fuchs et al. (2001) gave theoretical and experimental arguments for supposing that oral reading fluency may reflect overall reading competence. Prior et al. (2011) examined comprehension after oral and silent reading in elementary- and middle-school students. They revealed that silent reading emerged as the better mode for text comprehension only in the seventh grade. Differences between reading modes were explained by additional processing demands imposed by concurrent articulation and eye voice coordination when reading aloud. Kim et al. (2012) suggested that silent reading fluency predicted reading comprehension better for skilled readers than for average readers. Price et al (2015) revealed that oral and silent reading fluency are separate constructs, though only oral reading fluency contributed to reading comprehension. Vocabulary was found to contribute a lot to comprehension even after controlling for reading fluency. It was also shown that silent reading was stronger for retelling narratives, but there was no significant difference for comprehension questions. The expository passages revealed no difference between the reading modes (Schimmel & Ness 2017).

Our study investigates how readers process and understand information when reading texts orally and silently. In particular, the relationship between reading fluency and reading comprehension was explored.

## **Experiment 1. Reading aloud**

### **Participants**

32 Russian-speaking adolescents with speech disorders (experimental group, N=32, Mage=15.3) and 28 Russian-speaking adolescents with normal reading skills (control group, N=28, Mage=15.6) participated in two experiments.

### **Methodology**

Both groups of participants read aloud two texts of the same length and the same level of readability, and afterwards answered 8 comprehension questions.

The reading fluency was assessed using the reading rate — the number of correctly read words per minute (CW/min). A number of correct answers to the questions showed reading comprehension (reading understanding).

Using CW/min, in the experimental group we identified ‘poor readers’ (N=15) — participants who showed a result less than 3SD from the average value in the control group. Other participants of the experimental group were identified as ‘good readers’ (N=17).

For the good readers, poor readers and control group the Spearman correlation between reading fluency and reading comprehension was measured.

### **Results**

In the group of ‘poor readers’, there was a negative correlation  $r = -0.708$ ;  $p = 0.003$  between CW/min and the number of points for text comprehension. In the group of ‘good readers’ and control group, we didn’t find any correlations between reading fluency and text understanding.

## **Experiment 2. Silent reading**

### **Participants**

The same participants as in Experiment 1 took part in the Experiment 2. Using the same procedure as Experiment 1 they were divided into three groups: the control group, poor readers and good readers.

### **Methodology**

This time participants read silently. Two texts of the same length and the same level of readability but different from the ones of the Experiment 1 was used as reading material. After each text 10 questions were presented.

To assess reading fluency in the silent mode we recorded participants eye-movement using an eye-tracker EyeLink 1000 plus by SR Research. We



measured the reading time (RT), total fixations count (TFC), average saccade amplitude (ASA), and regressions count (RC). Text understanding was evaluated by the number of correct answers to the comprehension questions. To explore the relationship between reading fluency and reading comprehension we performed the Spearman correlation.

### Results

In the group of 'poor readers' we found a correlation between text understanding and RT ( $r = 0.67$ ;  $p = 0.006$ ), TFC ( $r = 0.747$ ;  $p = 0.001$ ), RC ( $r = 0.598$ ;  $p = 0.019$ ), as well as a negative correlation between comprehension level and ASA ( $r = -0.534$ ;  $p = 0.04$ ). In the control group and the group of the good readers there was no significant correlation between any reading fluency measure and reading comprehension accuracy (all  $ps > 0.05$ ).

### Discussion

Overall results show that while reading orally, participants with reading disorders comprehend the text poorly. They have a limited amount of attentional resources available for any given cognitive tasks, and thus, the greater the amount of attention they paid to text processing, the less is available for text understanding. However, in silent reading, the same participants demonstrate good comprehension if they read slowly, i.e. make more fixations and regressions, and have smaller amplitude of saccades. Their cognitive resources are spent on text comprehension, which is eventually what reading is aimed at.

### Acknowledgements

The study was supported by the research grant no. ID 75288744 "Text processing in L1 and L2: Experimental study with eye-tracking, visual analytics and virtual reality technologies" from St Petersburg University.

### References

- Adams, M. J. 1990. *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Fuchs, L.S., Fuchs, D., Hosp, M.K., Jenkins, J.R. 2001. Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239-256.
- Kim, Y.S., Wagner, R.K., Foster, E. 2011 Relations Among Oral Reading Fluency, Silent Reading Fluency, and Reading Comprehension: A Latent Variable Study of First-Grade Readers. *Sci Stud Read.* 2011 Jul, 15(4), 338-362.
- Price, K.W., Meisinger, E.B., Louwerse, M.M., D'Mello, S. 2016. The contributions of oral and silent reading fluency to reading comprehension. *Reading Psychology*, 37(2), 167-201.

- Prior, S. M., Fenwick, K. D., Saunders, K. S., Ouellette, R., O'Quinn, C., Harvey, S. 2011. Comprehension after oral and silent reading: Does grade level matter? *Literacy Research and Instruction*, 50(3), 183-194.
- Schimmel N., Ness M. 2017. The effects of oral and silent reading on reading comprehension. *Reading Psychology*, 38(4), 390-416.

## Index of names

---

Aldhafyan, R. ....1	Gutiérrez-Domínguez, F.J. ....1	Pota, S. ....1
Aldholmi, Y. ....1	Hitz, J. ....1	Qamri, S. ....1
Alexeeva, S. ....1	Honbolygó, F. ....1	Rami, Y. ....1
Alotaibi, S. ....1	Hu, W. ....1	Razkane, H. ....1
Alqahtani, A. ....1	Hubers, F. ....1	Ryzhikova, T. ....1
Alrouqi, M. ....1	Hurtado, I. ....1	Sayoud, H. ....1
Anderson, T. ....1	Iliopoulos, I. ....1	Schlatter, Th. ....1
Andreou, G. ....1	Jung, I.-Y. ....1	Schlechter, J. ....1
Barattieri di San Pietro, Ch. ....1	Kachkovskaia, T. ....1	Segklia, M. ....1
Baroncini, I. ....1	Kafentzis, G.P. ....1	Sfakianaki, A. ....1
Boudanga, A. ....1	Kampaki, V. ....1	Shamrin, A. ....1
Bulgheroni, M. ....1	Karpava, S. ....1	Shields, R. ....1
Chamorro, G. ....1	Katsarou, D. ....1	Simonetti, V. ....1
Chen, T.-Y. ....1	Kholiavin, P. ....1	Sklika, E. ....1
Chernova, D. ....1	Kiagiadaki, D. ....1	Slioussar, N. ....1
Cho, J. ....1	Kiapekaki, A. ....1	Soares, A.P. ....1
Chung, J.-H. ....1	Kim, Ch.S. ....1	Song, G. ....1
Corbett, G.G. ....1	Kim, Y.-W. ....1	Spinelli, E. ....1
Cordero-Rull, M. ....1	Kissani, N. ....1	Stoica, I. ....1
Crocama, C. ....1	Kiyama, S. ....1	Strik, H. ....1
Cucchiari, C. ....1	Lages, A. ....1	Tallon, K. ....1
del Rosso, G.A. ....1	Li, Y. ....1	Taw, L.-W. ....1
Demina, Y. ....1	Liu, M. ....1	Tollan, R. ....1
Dimitrakopoulou, M. 1	Mamushina, A. ....1	Tomioka, S. ....1
Diouny, S. ....1	Martin, Ph. ....1	Torregrossa, J. ....1
El Ouardi, L. ....1	McCabe, C. ....1	Tsangalidis, A. ....1
Esaulova, Y. ....1	Mendes Oliveira, H. ..1	Urtegeshev, N. ....1
Fang, Sh. ....1	Menshikova, A. ....1	van Bommel, L. ....1
Felser, C. ....1	Meunier, F. ....1	Vlahavas, G. ....1
Franjeh, M. ....1	Nasukawa, K. ....1	Witecy, B. ....1
Frolova, M. ....1	Nirgianaki, E. ....1	Yao, J. ....1
Galantomos, I. ....1	Novozhilov, A. ....1	Yoo, M.-H. ....1
Georgiou, G.P. ....1	Ouamour, S. ....1	Yuan, F. ....1
Goertz, G. ....1	Özay Demircioğlu, A.G. ....1	Yun, J. ....1
Goertz, G. ....1	Penke, M. ....1	Zavogianni, M.I. ....1
Grandison, A. ....1	Petrova, T. ....1	Zeyer, B. ....1
Greco, M. ....1	Plotnikov, I. ....1	Zhang, Z. ....1
		Zimianiti, E. ....1
		Zubov, V. ....1



ExLing 2021

Ebook ISSN: 2529-1092

Ebook ISBN: