



International Society of Experimental Linguistics

## **ExLing 2022**

### **Proceedings of 13<sup>th</sup> International Conference of Experimental Linguistics**

17-19 October 2022  
Paris, France

Edited by Antonis Botinis



Université  
Paris Cité





International Society of Experimental Linguistics

## **ExLing 2022**

Proceedings of 13<sup>th</sup> International Conference of  
Experimental Linguistics

17-19 October 2022

Paris, France

Edited by Antonis Botinis

ExLing 2022  
Proceedings of 13<sup>th</sup> International Conference of Experimental Linguistics

Published by ExLing Society  
Electronic edition  
ExLing 2022  
Athens, Greece  
ISSN: 2529-1092  
ISBN:  
DOI: 10.36505/ExLing-2022/13  
Copyright © 2022 ExLing Society

## Foreword

---

Welcome to the 13<sup>th</sup> International Conference of Experimental Linguistics ExLing 2022 Paris. This year, our hybrid set up enables us to meet in-person thanks to the hospitality of the Université Paris Cité, enjoying at the same time the excitements of the City of Light, Paris.

As an international forum for generations of linguists, our Society is collectively devoted to analyzing language through the development of experimental methods in Linguistics. We encourage new and established researchers to participate in and discuss developments in linguistic research and related experimental methodologies.

It all 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 cities such as Paris, Saint Petersburg, Heraklion, Lisbon and has been established since 2019 as the annual International Conference of Experimental Linguistics.

ExLing 2022 has a recurrent connection to Paris - this year marks our third conference here. We are excited to return to this wonderful city once more, this time with the added benefit of virtual conference attendance which opens new horizons and allows us to reach a wider audience.

This volume includes the proceedings of ExLing 2022. In addition to the main conference, ExLing 2022 is hosting three special sessions encompassing diverse and energetic research domains focusing on experimental methods and current topics in Experimental Phonetics, Language Education and Language Pathology.

The launching of the Language Technology Exhibition presents a new aspect of the ExLing 2022 Paris conference, which is expected to unify diverse language applications and promote interdisciplinary advances in major areas of technological language aids.

We would like to thank all ExLing 2022 participants and our keynote speakers, Ann Bradlow, Edward Gibson, Frank Guenther, and Charles Hulme, as well as colleagues from the International Advisory Committee, and the Review and Organization Committees for their contribution to the successful outcome of the Conference.

Antonis Botinis  
ExLing Society

## Contents

---

### Keynote paper

- Tempo, information rate, and communicative efficiency in L2 speech* ..... 1  
Ann R. Bradlow

### Research papers

- Speech rate and intelligibility of Arab Airlines' safety instructions* ..... 5  
Manal Aldhafeeri, Ghadi Alqahtani, Yahya Aldholmi
- Variable Adaptation of /v/ and /tʃ/ in English Loanwords in Saudi Arabic* ..... 9  
Areej Alenazi, Sam Hellmuth
- Using entropy to determine vowel spaces* ..... 13  
Terese Anderson, Grandon Goertz
- Relative clauses in Down Syndrome: evidence from the Greek language* ..... 17  
Georgia Andreou, Efstathia Chartomatsidou
- Evaluation and reading time of predicate agreement with conjuncts* ..... 21  
Daria Belova
- Romanian children are not able not to derive actuality entailments* ..... 25  
Simona Bilea, Adina Camelia Bleotu
- Language environmental analysis (LENA) of three Cypriot Greek-speaking children* ..... 29  
Paris Binos, Kakia Petinou
- The role of presentation modality in sentence processing* ..... 33  
Daria Chernova, Artem Novozhilov, Natalia Sliousar
- Image complexity in the tracking of DLD* ..... 37  
Leticia Corrêa, Erica Rodrigues, Marina Augusto
- The annotation of information structure in spoken Chinese: a pilot study* ..... 41  
Emanuela Cresti, Shuai Luo, Massimo Moneglia, Alessandro Panunzi
- Disentangling word order and function assignment preferences in Modern Greek* ..... 45  
Yanis da Cunha, Ioanna Chorai, Anne Abeillé
- Iconicity in fictional and natural languages: an experimental study* ..... 49  
Varvara Davydova, Elena Shamina
- The adaption to online synchronous teaching and voice fatigue: acoustic and clinical data* ..... 53  
Karina Evgrafova, Natalia Sokolova, Nikolay Shvaley
- Shared syntactic representations in non-native languages* ..... 57  
Chiara Facipieri, Michaela Vann, Giulia Bencini
- Sentence repetition as a function of episodic buffer: a pilot study in Croatian* ..... 61  
Mateja Gabaj, Jelena Kuvač Kraljević

---

<i>Exploring the prosody of affective speech</i> .....	65
Anna M. Giovannini, Ailbhe Ní Chasaide, Christer Gobl	
<i>Comparing number and gender agreement processing in Russian: an experimental study</i> .....	69
Anastasiia Ivanova, Natalia Slioussar	
<i>Dominance shift during the stay-abroad experience: a tip-of-the-tongue experiment</i> .....	73
Victoria Kamp, Maud Péliissier	
<i>Narrative skills of Russian heritage children in Cyprus</i> .....	77
Sviatlana Karpava	
<i>Negative vaccine voices in Swedish social media</i> .....	81
Dimitrios Kokkinakis, Mia-Marie Hammarlin	
<i>Exploring the role of L1 reading ability when reading in L2</i> .....	85
Anastasiia Konovalova, Vladislav Zubov	
<i>Implicit causality in Romanian interpersonal verbs</i> .....	89
Sofiana-Iulia Lindemann, Mădălina Matei	
<i>The L1 influence on the processing of L2 tense-aspect by L2 learners</i> .....	93
Yu Liu	
<i>Tracking participants' behaviour when performing linguistic tasks</i> .....	97
Maria do Carmo Lourenço-Gomes, Cecília Castro, Ana Paula Amorim, Gitanna Brito Bezerra	
<i>Testing the Endoskeletal Hypothesis in English/German bilingual code-switching</i> .....	101
Toby S. Lowther	
<i>Topicalization out of English and Mandarin if-clauses and that-clauses</i> .....	109
Ruihua Mao, Anne Abeillé, Barbara Hemforth, Edward Gibson	
<i>Brain waves and syllabic stress clash</i> .....	113
Philippe Martin	
<i>Timing of high pitch in Munster Irish</i> .....	117
Connor McCabe	
<i>The neural machine translation of dislocations</i> .....	121
Behnoosh Namdarzadeh, Nicolas Ballier	
<i>Comparing pre-linguistic normalization models against US English listeners' vowel perception</i> .....	129
Anna Persson, T Florian Jaeger	
<i>The effect of sound quality on attention and load in language tasks</i> .....	133
Kilian Seeber, Dongpeng Pan	
<i>Does the number morphology determine the agreement strategy?</i> .....	137
Kseniia Studenikina	
<i>DRESS-raising in Connemara English: shedding light on the role of Irish</i> .....	141
Kate Tallon	

vi Contents

---

<i>Listeners' categorisation behaviour correlates with gradient changes in exposure statistics</i> .....	145
Maryann Tan, Rachel Sabatello, Iva Savic, T. Florian Jaeger	
<i>Romanian lexical blending: from zero to hero</i> .....	153
Monica Vasileanu, Anabella-Gloria Niculescu-Gorpin, Cristina-Andreea Radu-Bejenaru	
<i>Word recognition in Developmental Language Disorders in Greek</i> .....	157
Konstantinos Vasileiou, Georgia Andreou	
<i>Testing variables of linguistic prestige in sociolinguistic questionnaires</i> .....	161
Tobias Weber	
<i>Four-stage evolution of Jin Chinese entering tones: from Lüliáng to Bingzhōu</i> .....	165
Yang Wei	
<i>Cue reweighting in Shanghainese sandhi patterns</i> .....	169
Zihao Wei	
<i>Mandarin and Spanish basic color terms vis-à-vis</i> .....	173
Mingshan Xu, Jingtao Zhu, Antonio Benítez-Burraco	
<i>Experimental investigation of Mandarin lexical tone production</i> .....	177
Yuyu Zeng, Robert Fiorentino, Jie Zhang	
<i>Classifier versus taxonomic relations in a similarity judgement task</i> .....	181
Jiahuan Zhang	



# Tempo, information rate, and communicative efficiency in L2 speech

Ann R. Bradlow

Department of Linguistics, Northwestern University, USA

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

This presentation seeks links between speech tempo, information transmission rate and communicative efficiency in conversations involving L2 participants. First, we show that L2 speech tempo is influenced by both “state” and “trait” characteristics. Specifically, while L2 speech is invariably slower than L1 speech (L1 vs. L2 “state”), within bilingual individuals L1 rate is a significant predictor of L2 rate (individual “trait”). Next, we show that the slower rate of L2 speech (fewer syllables/second) is compounded by lower L2 information density (more syllables/text) yielding a very low L2 information transmission rate (fewer bits of information transmitted/second). Finally, we show variation in conversation-based cooperative task completion time depending on degree of language background overlap between task participants. Together, these studies provide a scaffold for eventually linking acoustic level temporal variation to discourse level communicative efficiency in L2 speech communication.

Keywords: bilingual speech communication; speech rate; information transmission rate, communicative efficiency.

## Introduction

One of the most salient features of speech communication between interlocutors from different language backgrounds is its relatively slow pace. At every step along the speech chain - from lexical selection to speech articulation to word recognition – second-language (L2) speech production and perception proceed more slowly than first-language (L1) speech production and perception. As a consequence, conversational interactions in which one or both of the participants are L2 speakers are generally less efficient than interactions that include only L1 speakers. In this presentation, I will discuss a series of studies of the temporal domain of L2 speech with an emphasis on its communicative consequences. Specifically, these studies will show possible links between speech rate at the phonetic level, density of information coding at the semantic level, and communicative efficiency at the discourse level.

## “State” and “trait” characteristics in speech tempo

At the phonetic level, one of the most salient differences between L2 and L1 speech production is the reduced tempo of L2 speech in terms of syllables produced per second. This reduced L2 speaking rate is well established across

L1 and L2 talkers of a given language as well as across languages within bilingual individuals (e.g. Guion et al, 2000; Baese-Berk & Morrill, 2015; Bradlow et al, 2017). Moreover, slower L2 than L1 speaking rates have been demonstrated in several L2s, including Spanish, French, Dutch, and German (García Lecumberri et al, 2017; De Jong et al, 2015; Trouvain & Möbius, 2014). Data from a large corpus of both read and spontaneous speech recordings in both the L1 and L2 of a group of bilingual individuals (n=86) from various language backgrounds (n=10) show that this language-general, group-level difference between L2 and L1 speaking rates exhibits systematicity at the individual talker level. Using automatically extracted speaking rate measurements (syllables/second), a comparison of L1 and L2 speaking rates within individual bilinguals revealed that, while speaking rate was always slower in L1 speech than in L2 speech, L1 speaking rate significantly predicted L2 speaking rate. That is, relatively fast or slow talkers in L1 were also relatively fast or slow in L2, respectively. These results indicate a persistent influence of a talker-specific articulatory setting, or “trait” characteristic, that combines with, rather than is overwhelmed by language dominance (i.e. a L1 versus L2 “state” characteristic) in the global temporal structure of bilingual speech production (Bradlow et al, 2017; see also De Jong et al, 2015 and Derwing et al, 2009).

This combination of “trait” and “state” characteristics in speech tempo suggests that information about both talker-specificity (who is talking) and mode-specificity (L1 versus L2 speech mode) are available to listeners as part of the context for message interpretation. Indeed, studies of foreign-accented speech detection have shown that listeners can determine whether an utterance was produced in L1 speech (i.e. sounds “unaccented”) or L2 speech (i.e. sounds “foreign-accented”) based on very short snippets of speech in a known language (Flege, 1984) as well as in an unknown language (Major, 2007), suggesting that listeners are sensitive to language-general, “state” characteristics of L2 speech even without knowledge of language-specific phonetic contrasts. Moreover, studies of talker identification have shown that listeners can identify specific bilingual talkers based on speech samples from both of their languages (e.g. Winters et al, 2008) indicating that talker-specific “trait” characteristics are accessible and potentially useful for listeners. Global speech tempo is very likely one of the acoustic dimensions that conveys this indexical information about the language “state” (L1 or L2) and talker “trait” (talker identity).

### **Information transmission rate in L1 and L2 speech**

To explore the communicative impact of the slow L2 tempo, we examined the relationship between speaking rate (syllables/second) and information density (number of syllables for a given text) in L1 and L2 recordings of a standard reading passage (NWS, North Wind and the Sun passage) in three languages, English, French, and Spanish. This analysis followed the reasoning and approach of cross-language comparisons indicating an inverse relationship

between speaking rate (syllables/second) and information density (number of speech units for a given meaning) yielding relative cross-language consistency in information conveyed per second (information rate) (Coupé et al, 2019). Extending this approach to L2 speech, we see that L2 productions of the NWS passage in all three languages involved both slower rates and lower information density than their respective L1s. A follow-up comparison of the number of acoustic syllables (intensity peaks in the signal) versus orthographic syllables (dictionary-based counts of phonological syllables in the text) indicated substantial syllable reduction for L1 speech (number of acoustic syllables < number of orthographic syllables) in contrast to either substantial syllable epenthesis (number of acoustic syllables > number of orthographic syllables) for L2 speech (L2 English and L2 French) or no reduction (L2 Spanish). Thus, compared to L1 speech, L2 speech involved information-sparse syllables (i.e. more syllables were produced to convey the same meaning/text) as well as slower speaking rates (fewer syllables per second) yielding a very slow information transmission profile (fewer bits of information conveyed per second) that, at an extreme, may fall outside the optimal range for human information processing of dynamic signals (Bradlow, 2022).

### **Communicative efficiency in a cooperative task**

Finally, to assess the impact of a language barrier in a task with communicative intent rather than in decontextualized, laboratory-based, monologue recordings, we compared task-completion time in a conversation-based, cooperative, picture-matching task, the diapix task (Van Engen et al, 2010; Baker & Hazan, 2011), across pairs of talkers representing various combinations of L1 and L2 talkers. These data showed that, while all pairs successfully completed the task, L2 talker pairs were substantially less efficient than L1 talker pairs in terms of both task-completion time and number of word repetitions (word type-to-token ratio). However, the L2 completion time disadvantage was mitigated when the L2 conversation partners shared L1 as compared to L2 partners from different L1 backgrounds (Van Engen et al, 2010).

### **Summary and conclusions**

Taken together, these studies demonstrate an accumulation of time-related influences of a language barrier on speech communication. At the phonetic level, L2 speech is distinguished from L1 speech by a slow tempo (fewer syllables per second). This relatively slow L2 rate is compounded by a relatively low L2 information density (less syllable-level reduction) yielding a speech signal with a very low information rate for a given text/meaning. While a direct link between information rate at the phonetic level and communicative efficiency at the discourse level remains elusive and is undoubtedly compounded by lexical, syntactic, and other linguistic and cognitive components of L2 speech production and perception, we can speculate that the

temporal character of L2 speech communication is not just a matter of global slowing but instead also involves a distinctive pattern of information encoding.

## References

- Baese-Berk, M., Morrill, T. 2015. Speaking rate consistency in native and non-native speakers of English. *J. Acoust. Soc. Am.*, 138:3, EL223-EL228.
- Baker, R., Hazan, V. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior research methods*, 43(3), 761-770.
- Bradlow, A. 2022. Information encoding and transmission profiles of first-language (L1) and second-language (L2) speech. *Bilingualism: Language and Cognition*, 25(1), 148-162.
- Bradlow, A.R., Kim, M., Blasingame, M. 2017. Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *J. Acoust. Soc. Am.*, 141(2), 886-899.
- Coupé, C., Oh, Y. M., Dediu, D., Pellegrino, F. 2019. Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science advances*, 5(9), eaaw2594.
- De Jong, N.H., Groenhout, R., Schoonen, R., Hulstijn, J. H. 2015. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Appl. Psycholing.*, 36(2), 223-243.
- Derwing, T.M., Munro, M.J., Thomson, R.I., Rossiter, M.J. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31(4), 533-557.
- Flege, J.E. 1984. The detection of French accent by American listeners. *J. Acoust. Soc. Am.*, 76, 692-707.
- García Lecumberri, M.L., Cooke, M., Wester, M. 2017. A bi-directional task-based corpus of learners' conversational speech. *Intl. J. of Learner Corpus Rsch*, 3(2), 175-195.
- Guion, S.G., Flege, J.E., Liu, S.H., Yeni-Komshian, G.H. 2000. Age of learning effects on the duration of sentences produced in a second language. *Appl. Psycholing*, 21(2), 205-228.
- Major, R.C. 2007. Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition*, 29(4), 539-556.
- Trouvain, J., Möbius, B. 2014. Sources of variation of articulation rate in native and non-native speech: comparisons of French and German. In *Proc. Speech Prosody 7*, 275-279.
- Van Engen, K.J., Baese-Berk, M., Baker, R.E., Choi, A., Kim, M., Bradlow, A.R. 2010. The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, 53(4), 510-540.
- Winters, S.J., Levi, S.V., Pisoni, D.B. 2008. Identification and discrimination of bilingual talkers across languages. *J. Acoust. Soc. Am.*, 123(6), 4524-4538.

# Speech rate and intelligibility of Arab Airlines' safety instructions

Manal Aldhafeeri, Ghadi Alqahtani, Yahya Aldholmi

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

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

We first analyzed and reported the speech rate (SR) of safety instructions (SIs) for nine Arab airlines, including Saudia, Flynas, EgyptAir, and Oman Air. EgyptAir had the slowest SR and Omani Air the fastest, a statistically significant difference (Kruskal-Wallis,  $p < 0.001$ ). Then, we assessed the impact of normal and accelerated (by 5%, 10%, 15%, and 20%) SR in passenger-boarding noise (-6 dB) on speech intelligibility evaluation. Listeners' ratings demonstrated that intelligibility diminished to 48% at the original SR and to 7% at the 20% accelerated SR and that SR was a statistically significant factor (repeated-measures ANOVA,  $p < 0.01$ ). Arab airlines are advised to consider this finding and consult the International Civil Aviation Organization on the recommended SR.

Keywords: speech rate, speech intelligibility, noise, safety instructions, Arab airlines

## Introduction

Within the plentiful scholarship on talker-listener interaction and speech intelligibility, the concept of speech rate (SR) has been examined consistently in numerous studies, yet experiments on SR combined with and/or compared to noise have been conducted only in a few such studies (e.g., Adams & Moore, 2009; Rosenhouse & Kishon-Rabin, 2003; Rosenhouse et al., 2006). The aim of these studies was to probe and understand the impact both of production conditions (i.e., SR) and listening conditions (i.e., noise) separately and collectively on intelligibility. Among such diverse conditions are aircrafts and air traffic control (ATC) towers, which are two critical environments for intelligibility (and safety). Taylor and colleagues (1994) reported that SR exerts a negative impact on aircraft pilots' ability to execute simulated ATC instructions, while Nitta and colleagues (2018) observed that American ATC officers who were native speakers of English did not slow their SR when communicating with nonnative pilots. The current study adds to the existing literature, first, by establishing a baseline for the SR (syllables per second [sps]) of safety instructions (SIs) for nine Arab airlines (e.g., Saudia, Flynas), and, second, by assessing the impact of SR in noise on intelligibility evaluation. The findings should enhance our understanding of the interplay between SR and noise and its pertinence to intelligibility, especially in Arabic, a language that presently suffers a scarcity of studies on speech rate (e.g., Aldholmi et al., 2021a).

## Part I: SR analysis of SIs

### Methods

We obtained YouTube video clips containing Modern Standard Arabic (MSA) SIs for nine Arab airlines (Saudia [S], Flynas [N], Qatar Airways [Q], Oman Air [O], Etihad Airways [Et], Emirates Airlines [Em], EgyptAir [Eg], Royal Jordanian [J]) and Royal Air Maroc [M], representing four major Arabic varieties: Arabian Peninsula (Gulf), Egypto-Sudanic, Levantine, and Maghrebi Arabic. We made efforts to select only clips that appeared to be original, complete, and unnoisy. We then extracted .WAV audio from each clip, segmented and syllabified the sentences, and removed long pauses. The pause removal process was based on a criterion of 200 ms and guided by pause classification and methodological consideration in previous research (e.g., Sakamoto et al., 2018; Tanka et al., 2011). Finally, SR was calculated according to the equation below.

$$\text{Equation 1. } SR \text{ (sps)} = \frac{\# \text{ of syllables}}{\text{Utterance Duration (seconds)} + \text{Pauses (200 ms or less)}}$$

### Results

The analysis of 485 sentences containing SIs yielded that EgyptAir had the slowest SR (*Median* = 5.8 sps, *Mean* = 5.8 sps) whereas Omani Air had the fastest (*Median* = 7.1 sps, *Mean* = 6.9 sps). The SR medians for the other airlines (Figure 1A) ranged between 6.0 sps (*Mean* = 5.9) and 6.7 sps (*Mean* = 6.9). A Kruskal-Wallis test showed a statistically significant difference between the medians for all airlines,  $H(8) = 91.495$ ,  $p < 0.001$ . As shown in Figure 1B, the pairwise comparisons revealed (a) that all differences among Gulf airlines, with the exception of the difference between Oman Air and Qatar Airways, were not statistically significant,  $p > 0.05$ ; and (b) that the statistically significant differences occurred between EgyptAir or Royal Air Maroc and most other airlines,  $p < 0.05$ , but not between EgyptAir and Royal Air Maroc,  $p > 0.05$ .

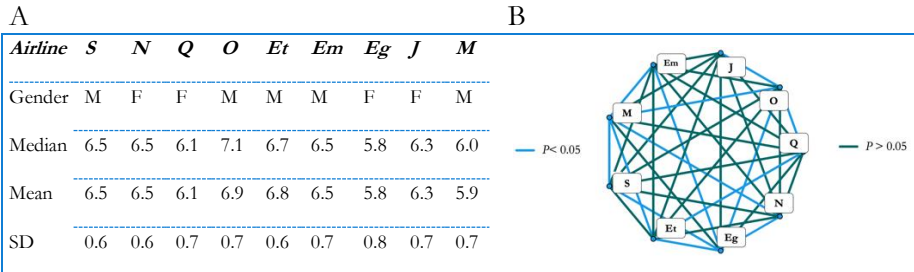


Figure 1. (A): Gender of talker and median, mean, and SD of SR for all airlines (B): Pairwise comparisons (B) between airlines (*note: each node shows a SR median*).

## Part II: SR and intelligibility of SIs in Noise

### Methods

Twenty imperative SIs were randomly selected from the Saudi Airlines (Saudia's in a male voice and Flynas's in a female voice) and aurally presented in a random order to 54 Saudi participants at the original SR (6.5 sps) and at an accelerated rate of the just-noticeable-difference (JND) level 5% (see Quené, 2007) as well as accelerations by 10%, 15%, and 20%. As in similar previous studies (e.g., Aldholmi, 2021b), the PSOLA (Pitch-Synchronous Overlap-and-Add) feature in Praat (Boersma & Weenink, 2022) was used for the acceleration process. All sentences were first normalized at 60 dB and then mixed with a passenger-boarding noise at -6 dB and resampled at 44100 hz. Fifty-four (22 male and 32 female) Saudi native speakers of Arabic (age,  $Mean = 30.6$   $SD = 6.8$ ) used the online platform *Phonic* (Phonic, 2022) to rate the intelligibility of each sentence on a scale from zero to 100 (zero = completely unintelligible, 100 = completely intelligible). The participants were allowed to listen to each stimulus as many times as they wished, but no sentence appeared in more than one condition to each participant.

### Results

The results showed that, in the simulated aircraft boarding noise, (a) the intelligibility diminished to 48% at the original SR and to 7% at the 20% accelerated SR (Figure 2B) and (b) that SR was a statistically significant factor (repeated-measures ANOVA,  $F[2.9] = 52.6$ ,  $p < 0.01$ ) in intelligibility rating. Pairwise multiple comparisons revealed that the differences between every two pairs of SR conditions were statistically significant,  $p < 0.05$ , with the exception of the comparison between 15% Acc and 20% Acc,  $p = 0.847$  (Figure 2A).

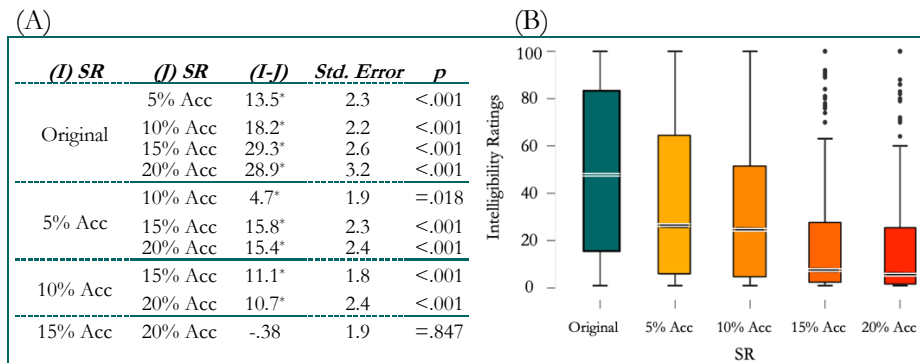


Figure 2. (A): Pairwise comparisons for all SR conditions (normal, 5%, 10%, 15%, and 20%), (B): Intelligibility ratings for all SR conditions.

### Conclusion

Most Arab airlines seem to have a fast SR, especially considering that SIs are meant for both native and nonnative passengers. Most SR medians exceeded the average reported for Saudi, Egyptian, and Moroccan dialects (e.g., Aldholmi et al., 2021b; Vaane, 1982). A median SR of 6.1 sps for English ATC officers has been described as a worrying and unsafe SR (Nitta et al., 2018). The results also showed that a slight increment in SR coupled with noise can lead to a large decrement in intelligibility rating. Should Arab airlines be apprised of this finding, they should consult the International Civil Aviation Organization on the recommended SR and collaborate on further research in this regard.

### References

- Adams, E. M., Moore, R. E. 2009. Effects of speech rate, background noise, and simulated hearing loss on speech rate judgment and speech intelligibility in young listeners. *Journal of the American Academy of Audiology*, 20(01), 028-039.
- Aldholmi, Y., Alqahtani, A., Aldhafyan, R. 2021a. Teen production and perception of dialectal Arabic speech rate. In *ExLing 2021*, Athens, Greece. (pp. 13-16).
- Aldholmi, Y., Aldhafyan, R., Alqahtani, A. 2021b. Perception of Standard Arabic Synthetic Speech Rate. In *Interspeech* (pp. 1704-1707).
- Boersma, P., Weenink, D. 2022. Praat: doing phonetics by computer [Computer program]. Version 6.2.15, retrieved 14 March 2022 from <http://www.praat.org/>
- Phonic - The Complete Toolkit for Unmoderated Research. (2022). Retrieved 13 March 2022, from <https://www.phonic.ai>
- Quené, H. 2007. On the just noticeable difference for tempo in speech. *Journal of Phonetics*, 35(3), 353-362.
- Rosenhouse, J., Kishon-Rabin, L. 2003. The effect of speech rate and noise on bilinguals' speech perception: the case of native speakers of Arabic in Israel. In the 8<sup>th</sup> Eurospeech. 2089-2092.
- Rosenhouse, J., Haik, L., Kishon-Rabin, L. 2006. Speech perception in adverse listening conditions in Arabic-Hebrew bilinguals. *Int. Journal of Bilingualism*, 10(2), 119-135.
- Sakamoto, S., Cui, Z., Miyashita, T., Morimoto, M., Suzuki, Y., Sato, H. 2018. Effects of inter-word pauses on speech intelligibility under long-path echo conditions. *Applied Acoustics*, 140, 263-274.
- Tanaka, A., Sakamoto, S., Suzuki, Y. 2011. Effects of pause duration and speech rate on sentence intelligibility in younger and older adult listeners. *Acoustical Science and Technology*, 32(6), 264-267.
- Nitta, H., Okazaki, H., Klinger, W., Kido, K. 2018. Speech Speed Rates of American Air Traffic Controllers in Communications with Non-Native English-Speaking Pilots. *Studies in the Humanities*, 103, 351-360.
- Taylor, J. L., Yesavage, J. A., Morrow, D. G., Dolhert, N., Brooks III, J. O., Poon, L.W. 1994. The effects of information load and speech rate on younger and older aircraft pilots' ability to execute simulated air-traffic controller instructions. *Journal of gerontology*, 49(5), P191-P200.
- Vaane, E. 1982. Subjective estimation of speech rate. *Phonetica*, 39(2-3), 136-149.



# Variable Adaptation of /v/ and /tʃ/ in English Loanwords in Saudi Arabic

Areej Alenazi, Sam Hellmuth

Language and Linguistic Science, University of York, UK

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

This study seeks to explore the impact of the input modality, language exposure, context, and gender on the production patterns of two non-native sounds, /tʃ/ and /v/, in Saudi Arabic. A production task was conducted to test 67 Saudi speakers in three conditions: aural-only (auditory inputs), written-only (orthographic inputs), and aural-written (auditory-orthographic inputs). Language exposure had a main effect on the production of the two sounds. Context was a major factor influencing the production accuracy of /v/ but not /tʃ/; /v/ was more likely to be devoiced in word-final position. The written input resulted in a decrease in the production accuracy for /tʃ/ but not /v/, suggesting that the effect of the input type varies for different non-native sounds.

Keywords: borrowing, variable adaptation, input modality, language exposure

## Introduction

When one language borrows a word from another language, a non-native sound may be preserved or substituted (e.g., Kang et al. 2016). This means that substitution doesn't necessarily occur when a sound doesn't exist in the native language. Previous studies of English loanwords in Arabic (e.g., Al-Athwary, 2017) showed variations between speakers in the production of /tʃ/ and /v/: /tʃ/ was either preserved or replaced by /ʃ/; /v/ was sometimes preserved and sometimes replaced by /f/. Prior research has attributed these variations to Arabic phonology, but less attention has been paid to the role of perception or orthography.

Research on L2 acquisition suggests that the input modality (with acoustic and/or orthographic information) in which a lexical item is initially encountered can influence the production of non-native sounds, especially if they are realised differently or do not occur in the native language (e.g., Bassetti, 2017). The production patterns of non-native sounds may thus vary with how borrowers are first exposed to a source word.

In this study, we consider the possible impact of input modality and a range of other factors (context, gender, and level of English exposure) on production accuracy of the two sounds.

## Methodology

The participants were 67 Saudi speakers, split by gender (31 male/ 36 female) and level of English exposure (high/medium/ low). Adapting the design of Vendelin and Peperkamp (2006), in this study, participants' oral productions of target words were elicited in three conditions: aural-only, written-only and aural-written. In aural condition, they heard non-words pronounced by an English native speaker, and in written condition they saw the word on screen. In aural-written condition, they heard the non-words while viewing the written forms on screen. Gorilla (<https://gorilla.sc>) was used to present stimuli and record participants' responses. The stimuli were 24 CVCVC non-words with target segments embedded in three positions: initial, intervocalic and final. The vowels were held constant with /æ/ in the first syllable and /ɪ/ in the second syllable. In each condition, participants were requested to produce six non-words twice. Written stimuli were presented in English orthography, with /v/ spelled with the letter <v> and /tʃ/ with the digraph <ch>; /æ/ was spelled with <a> and /ɪ/ with <i>. Recordings of auditory stimuli were produced by a female native speaker of British English, normalised to the same intensity level (75 dB). Participants' productions were auditorily coded as one of either [v~f] or [ʃ~tʃ], by the first author, with reference to the waveform and spectrogram in Praat.

## Results

Mixed effects logistic regression models were run for each target segment to estimate the probability of the binary outcome based on the different predictors. Figures 1-2 visualise the predicted probability of substitution for each target. Overall, production accuracy (i.e. preservation of the loan segment) was lower for /v/ than /tʃ/. Language exposure had a main effect on the production accuracy of /v/; the high exposure group ( $\beta = -2.1889$ , SE = 0.4171,  $z = -5.247$ ,  $p = 1.54e-07$  \*\*\*) and the medium exposure group ( $\beta = -1.0199$ , SE = 0.4059,  $z = -2.513$ ,  $p = 0.01198$  \*) are each significantly different from the low exposure group. There was also a main effect of context: /v/ was more likely to be produced in both word-initial context ( $\beta = 1.5524$ , SE = 0.4960,  $z = 3.130$ ,  $p = 0.00175$  \*\*) and word-intervocalic context ( $\beta = 0.8802$ , SE = 0.4316,  $z = 2.040$ ,  $p = 0.04139$  \*). Female participants showed higher production accuracy than male participants for /v/ ( $\beta = -1.9756$  (SE = 0.4181,  $z = -4.725$ ,  $p = 2.30e-06$  \*\*\*) as well as an interaction of gender with condition ( $\beta = 1.2801$ , SE = 0.4087,  $z = 3.132$ ,  $p = 0.00174$  \*\*), whereby male participants were less likely to produce /v/ whenever aural input is available. The performance of the female participants did not differ between conditions. Input modality had a main effect on production accuracy of /tʃ/. The results showed that /tʃ/ was less likely to be produced in the aural-written condition ( $\beta = 1.1494$ , SE = 0.3543,  $z = 3.245$ ,  $p = 0.00118$  \*\*) and in the written condition

( $\beta = 3.1923$ ,  $SE = 0.3534$ ,  $z = 9.033$ ,  $p = < 2e-16$  \*\*\*) than in the aural condition. The effect of language exposure was also significant; the high exposure group differed significantly from the low exposure group ( $\beta = -2.4332$ ,  $SE = 0.6027$ ,  $z = -4.037$ ,  $p = 5.4e-05$  \*\*). The production accuracy of /tʃ/ was not affected by context or gender.

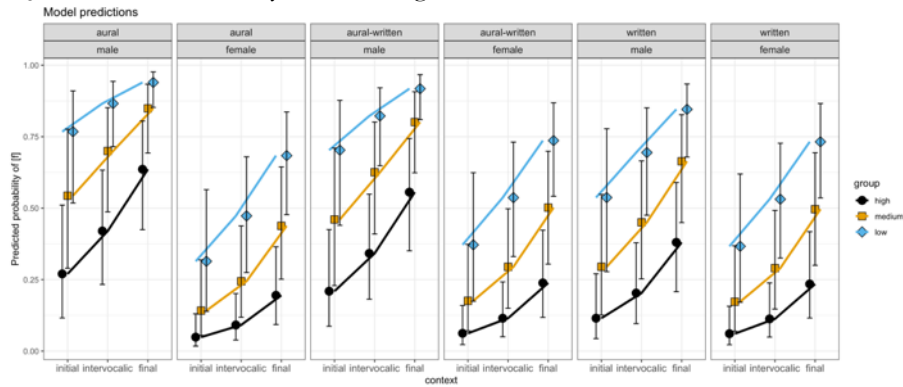


Figure 1. Predicted probability of [f] production in target non-words by context, group (level of language exposure), condition and gender.

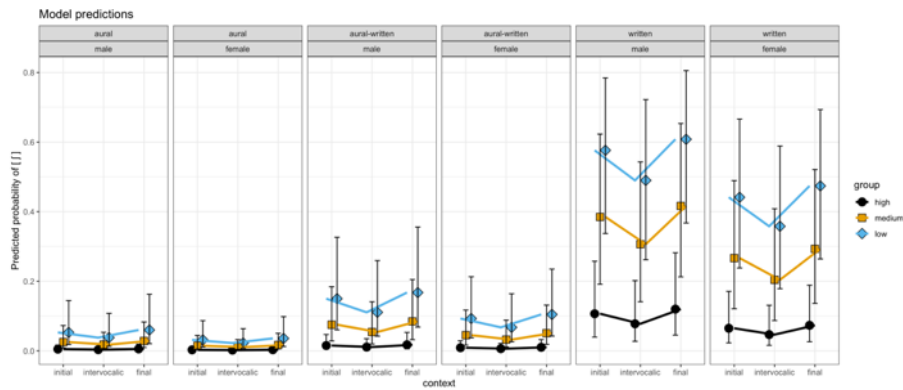


Figure 2. Predicted probability of [ʃ] production in target non-words by context, group (level of language exposure), condition and gender.

## Discussion

The influence of level of English exposure is most evident in how input condition contributes differently to the production of /v/ and /tʃ/. Overall, as already noted, the presence of written-only input supported the production accuracy of /v/, but this effect was greater with more exposure to English. Participants in the high exposure group were more likely to utilize the orthographic information whereas written-only input had reduced impact on production for the participants in the low exposure group. In contrast, the

absence of aural input resulted in lower production accuracy for /tʃ/, though participants in the high exposure group had the highest production accuracy in written condition, perhaps as they are more familiar with English orthography.

These results contrast with some previous studies which found that orthographic input along with auditory input can improve production accuracy of non-native sounds (e.g., Davidson, 2010), but are in line with other studies which showed that orthographic input accompanying auditory input is not helpful (e.g., Vendelin and Peperkamp, 2006). It is worth noting that target /tʃ/ was spelled as <ch> in the stimuli, which in English can represent not only /tʃ/ but also /ʃ/ and even /k/ (e.g., chair, chef, and character). This spelling inconsistency contrasts with the transparency of Arabic orthography where one letter represents one sound. Context influenced production of /v/ but not /tʃ/: participants were more likely to replace /v/ with [f] in word-final position. It is worth noting that the final /v/ was partially devoiced in the auditory stimuli, which is a feature of English word-final fricatives, so /v/ was even more similar to the participants' closest native sound /f/ in word-final position. The source of devoicing in the participants' own productions cannot be attributed to their native language since has a word-final voicing contrast (e.g., [dəz] 'push' and [dəs] 'hide'). Thus markedness, rather than native language transfer, may have influenced participants' production of /v/ (Eckman, 1991).

In conclusion, the findings demonstrate a primary role of language exposure in influencing variable adaptation of loan segments; participants with high English exposure were more likely to produce target-like /v/ and /tʃ/. Nevertheless, all participants, regardless of level of English exposure, struggled with production of /v/ in word-final context. Finally, written input supported production accuracy for /v/ but not for /tʃ/, suggesting that the effect of the input type (auditory, written, or both) varies for different non-native sounds.

## References

- Al-Athwary, A. 2017. The Phonotactic adaptation of English loanwords in Arabic. *Arab World English Journal* 8, 392-406.
- Bassetti, B. 2017. Orthography affects second language speech. *Learning, memory, and cognition*, 43(11), 1835-1842.
- Davidson, L. 2010. Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics*, 38(2), 272-288.
- Eckman, F.R. 1991. The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in second language acquisition*, 13(1), 23-41.
- Kang, Y., Pham, A.H., Storme, B. 2016. French loanwords in Vietnamese. In Albright, A and Fullwoods, M, Proc. The Annual Meetings on Phonology. Linguistic Society of America, Washington DC.
- Vendelin, I., Peperkamp, S. 2006. The influence of orthography on loanword adaptations. *Lingua*, 116(7), 996-100.

# Using entropy to determine vowel spaces

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

<sup>1</sup>Health Sciences Center, University of New Mexico, US

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

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Vowel spaces are often illustrated with the first and second formants plotted against each other on X-Y plots. The numerical values of vowel F1, F2, and F3 were evaluated using Pearson correlations and results often show inconsistencies, indicating that some F1-F2 plots might not provide the best vowel space depiction. The same formant values were examined using conditional entropy methodology and determinations were made which formants should be the independent X variable, and which formants should be the dependent Y variable for plotting. This research describes the examination of relationships between formants to determine optimal, more accurate vowel space plots.

Keywords: vowel space, entropy, correlation, formant

## Introduction

Formant values are traditionally plotted on an X-Y vowel space diagram to show how the formant values are related to each other (Story & Bunton, 2017). This research investigates how F1, F2, and F3 values compare.

This research is motivated by two research questions: Are vowel space plots a viable method to display depict vowel characteristics? A second research question is: Is it possible to measure the relationship between formants, prior to plotting, and determine which formant plots would produce optimal results and which plots would not be the best for comparison purposes?

## Correlations of vowel formants

The speech data for this research comes from three sources: 1) Linguists provide examples of the 28 IPA vowels on several academic websites (see references) and a total of 112 vowel examples were used in this evaluation. 2) A total of 908 CVC and CV words read by native English speakers (two female, 1 male) using a Røde N microphone in a GretschKen™ Industries acoustic sound booth. 3) The third portion of the speech data was 426 speech tokens from 29 languages downloaded from the IPA and UCLA websites.

All vowels and vowel portions of the words were evaluated using PRAAT (standard settings of 5,500 Hertz ceiling and 5 formants), to produce F1, F2, and F3. Formants were then evaluated using Pearson correlation methods.

Correlations varied widely for the 112 vowels found that approximately 15% have strong correlations and 50% have low or negative correlations. Many correlations indicated F1-F3 formants would be preferable.

The word list correlations varied from very low values such as -0.263 (man) and -0.067 (key) to higher values 0.901 (pet) and 0.842 (put). Some F1-F3 correlations were found to be stronger than F1-F2 correlations.

The correlations for the tokens of the 29 languages also varied from very low to high values and did not support a general F1-F2 plotting scheme.

Correlations did not produce clear indications for vowel space plotting, prompting the investigation to the suitability of conditional entropy.

### Conditional entropy for vowel formant pairs

Conditional entropy was developed from Information Theory. Entropy is used to detect patterns in data, and to establish their relationships by providing numerical values that shows their relative degree of overlapping information (see MacKay 2003, Haglund, Jeppsson, & Strömdahl 2010, Goodfellow et al. 2016). Conditional entropy is the amount of data needed for one variable to describe another variable: their suitability for plotting one against the other. The conditional entropy was calculated with the Mutual Info 0.9 cross-platform program package (Peng 2002) operating in Matlab.

### Results

The entropy data shows which formant is the independent variable (X) and which formant is the dependent variable (Y) for plotting.

Table 1. Entropy values of vowel portions of speech samples. Data from female speakers are indicated in italics. The Bulgarian word *nem* translates to ‘five’ and the Czech word *světlo* translates to ‘light’. Other words are English.

	<i>kid</i>	boat	cot	out	<i>nem</i>	<i>světlo</i>	/æ/	<i>ship</i>
<b>F1F2</b>	<i>0.3185</i>	<b>0.024</b>	<b>0.0606</b>	0.1086	<b>0</b>	<i>0.3419</i>	<b>0</b>	<b>0</b>
<b>F2F1</b>	<i>0.1757</i>	0.0481	0.101	<b>0.0869</b>	<i>0.0952</i>	<i>0.0408</i>	0.1609	<b>0</b>
<b>F1F3</b>	<i>0.2471</i>	0.1204	0.1414	0.1304	<b>0</b>	<b>0</b>	0.1006	<b>0</b>
<b>F3F1</b>	<i>0.1757</i>	0.0481	0.101	<b>0.0869</b>	<i>0.0952</i>	<i>0.0408</i>	0.1379	<b>0</b>
<b>F2F3</b>	<b>0.0952</b>	0.1204	0.1414	0.1304	<b>0</b>	<b>0</b>	0.1236	<b>0</b>
<b>F3F2</b>	<i>0.1667</i>	<b>0.024</b>	<b>0.0606</b>	0.1086	<b>0</b>	<i>0.3419</i>	<b>0</b>	<b>0</b>

The left bolded column shows the six pairs of the formants that were compared for the analysis. F1F2, for example, is notation for the test of the suitability of F1 as the dependent variable and F2 as the independent variable. Lower entropy values show better suitability for vowel space plotting. In the case of ‘*kid*’, plotting F2-F3 would provide optimal results. ‘*Boat*’ and ‘*cot*’

indicate the dominance of F2. The results for ‘out’ shows that F1 should be plotted on the X-axis. The zero values of *nem* shows that F2 and F3 are independent, *světlo* shows that F3 is the independent variable, and the UCLA vowel /æ/ spoken in isolation shows that F2 is the independent variable. *Ship* data, having all zero values, indicates that the vowel formants are combined or related to each other.

Some vowel formant pairs have zero entropy values as seen in the four right columns of Table 1. Conditional entropy has a unique defined mathematical property: *if and only if* conditional entropy is equal to zero *then* the value of one variable (dependent) is a direct function of the input variable (independent). This property was noted in samples taken from both male and female speakers, and sometimes does not appear consistently in words that speakers repeated words. These zero or near zero values mean that two formants are linked and perhaps one formant generates the other. Rather than being distinct formants, the two zero formants could be considered a combined wide band of formants.

## Summary

Conditional entropy determines which formants would be most accurately displayed on the x-axis and which formant would be plotted on the y-axis. In some cases, F2-F3 or F3-F2 plots would convey more complete information than F1-F2 plots. Assumptions about vowel depictions may have to be modified, according to entropy data that defines the appropriate X-Y vowel space plot.

The discovery of zero entropy values of some formant pairs should be investigated further. Hopefully additional research will show if zero value formant pairs should be plotted against each other, or against a third value.

## References

- Goodfellow, I., Bengio, Y., Courville, A. 2016. Deep Learning. Cambridge, Massachusetts: The MIT Press.
- Haglund, J.,F. Jeppsson, H. Strömdahl. 2010. Different Senses of Entropy – Implications for Education. Entropy 12, 490-515. Doi: 10.3390/e12030490.
- Hayes, G./SOUND%20FILES/Bruce%20Hayes%20words%20S%20T%20C%20E International Phonetic Association. 2022. IPA Handbook. <https://www.internationalphoneticassociation.org/content/ipa-handbook>
- Ladefoged, P. <http://www.phonetics.ucla.edu/course/chapter1/vowels>.
- MacKay, D. 2003. Information Theory, Inference and Learning Algorithms. Cambridge: Cambridge University Press.
- Peng, H. 2022. Mutual Information computation. ([www.mathworks.com/matlabcentral/fileexchange/14888-mutual-information-computation](http://www.mathworks.com/matlabcentral/fileexchange/14888-mutual-information-computation)),
- Story, B., Bunton, K. 2017. Vowel space density as an indicator of speech performance. The Journal of the Acoustical Society of America **141**, EL458. <https://doi.org/10.1121/1.4983342>





# Relative clauses in Down Syndrome: evidence from the Greek language

Georgia Andreou, Efstathia Chartomatsidou

Department of Special Education, University of Thessaly, Greece

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

The aim of the present study is to investigate the comprehension and production of Relative Clauses (RCs) by a group of Greek-speaking individuals with Down Syndrome (DS) and their mentally age-matched peers. The possible asymmetry in performance between Subject-Gap (SG) and Object-Gap (OG) relatives was also examined and an error analysis was conducted. The results indicated that DS individuals encounter difficulties in both comprehending and producing RCs compared to Typically Developing Children (TDC), as well as that both groups performed significantly better in SG than in OG relatives and presented the same pattern of error types. The same pattern of performance observed gives support to the hypothesis that language development in DS could be characterized as more delayed than deviant.

Keywords: Down Syndrome, relative clauses, language development, delay hypothesis

## Introduction

Many cross-linguistic studies focus on the language development of individuals with DS and characterize it as either delayed or deviant compared to that of typical population. However, most of them conclude that morphosyntax is severely impaired in this population (Katsarou & Andreou 2022). For the Greek language, very few studies to date have been conducted that have examined morphosyntactic phenomena in this impaired population. The purpose of this study is to investigate both the comprehension and production of RCs by Greek-speaking individuals with DS and to compare their performance to that of mentally age-matched peers. In particular, it was examined whether the individuals with DS comprehend and produce RCs in the same way as the TDC, as well as the asymmetry in their performance between SG and OG relatives, an issue which is studied in typical and atypical populations (McKee & McDaniel 2001; Stavrakaki 2001). Moreover, a detailed error analysis has been conducted.

## Methodology

### Participants

15 Greek-speaking individuals with DS aged from 9 to 37 with moderate retardation participated in the current study. In addition, a control group of

TDC aged from 6 to 7 years, without any reported learning or neurological disabilities, also participated in the study. The two groups were matched on mental ages through Raven and WAIS-IV tests, as well as on the raw scores of two subtests which assess their expressive language skills, namely the Action Picture Test (T-value=-1,88,  $p=0,071$ ) and their receptive vocabulary, namely the Peabody Picture Vocabulary Test (T-value=-0,89,  $p=0,383$ ). Detailed information of the participants and scores on language tests is presented as mean  $\pm$  standard deviation (SD) in Table 1.

Table 1. Participant information and scores on the language tests for the examined groups.

	DS group (N=15)	Cntrl. group (N=15)
Gender (Male-Female)	46,7 %-53,3%	40%-60%
Chronological Age	19,77 $\pm$ 8,66	6,12 $\pm$ 0,245
Mental Age	6,37 $\pm$ 0,48	
Action Picture Test	69 $\pm$ 4,80	72,40 $\pm$ 5,12
PPVT	70,33 $\pm$ 7,18	72,67 $\pm$ 7,24

## Materials

Two experiments were conducted. The first experimental task was an elicited production task with twenty trial sentences (10 Subject -10 Object relatives). The second experiment was a picture selection task with twenty trial sentences on five RC types: Subject head-Subject gap (SS), Subject head-Object gap (SO), Object head-Subject gap (OS), Object head-Object gap (OO) and Object head-Object gap with clitics (OOcl).

## Results

Descriptive statistics analysis revealed that the mean correctness score of the DS group is statistically significant lower (Production = 2,86, SD 1,46/Comprehension = 8,93 SD 2,22) than that of the TDC for both tasks (Production = 17,67, SD 2,32/ Comprehension = 15,87, SD 1,68). Since the data were not normally distributed non-parametric tests were employed. Mann-Whitney U Test analysis revealed a highly significant difference between the two groups' median scores on production (3,000  $\neq$ 19,000,  $p=0,000$ ) and on comprehension (9,000 $\neq$ 16,000,  $p=0,000$ ). TDC showed almost equal performance in the median scores of the two tasks ( $p=0,0310$ ), while the DS participants showed higher performance in the comprehension task than in the production one ( $p=0,000$ ).

Regarding the production task, both groups showed better performance in subject RCs than in object RCs. However, there was a statistically significant difference between the two types of RCs only in the DS group ( $p=0,0006$ ), while TDC did not present a statistically significant difference ( $p=0,1524$ ). On

the other hand, in the comprehension task, examining the SG (SS+OS relatives) and OG (SO+OO relatives) asymmetry, we found that both groups comprehend the SG relatives better than the OG relatives (DS  $p=0,0001$  / TDC  $p=0,0001$ ). One way ANOVA revealed that the TDC exhibited the highest level of performance in OS (M=0,267 SD=0,458) and SS (M=0,467 SD=0,516) relatives and the lowest level in SO (M=1,800 SD=0,775) and OO (M=0,800 SD=0,561) relatives. The DS group exhibited the highest level of performance in SS (M=1,133 SD=0,743) and OS (M=1,933 SD=0,884), and the lowest level in SO (M=3,067 SD=0,884) and OO (M=2,400 SD=0,632) relatives. Detailed error analysis has been conducted in order to examine the pattern of the responses of the two groups. In the production task we noticed that both groups presented the same pattern of errors. DS participants preferred to produce main clauses (Subject RCs: M=2,733 SD=0,884 / Object RCs: M=2,667 SD=0,900) or elliptical responses (Subject RCs: M=2,333 SD=0,900 / Object RCs: M=3,000 SD=1,134) instead of the RCs and also made reversal errors (Subject RCs: M=2,333 SD=1,234 / Object RCs: M=2,533 SD=0,743). Respectively, TDC produced main clauses (Subject RCs: M=0,467 SD=0,743 / Object RCs: M=0,800 SD=0,862) and elliptic responses (Subject RCs: M=0,200 SD=0,414 / Object RCs: M=0,467 SD=0,640). Few lexical/semantic errors in the DS's responses were also found (Subject RCs: M=0,333 SD=0,488 / Object RCs: M=0,667 SD=0,617).

In the comprehension task we noticed that in SS and OO RCs the most frequent error in the DS group is the reversal error (M=0,600 SD=0,507 / M=1,200 SD=0,676), in SO (M=1,333 SD=0,976) and OS RCs (M=1,000 SD=0,535) is the agent error, while in OOcl the percentage is the same for reversal (M=1,067 SD=0,799) and agent errors (M=1,067 SD=0,704). TDC presented the same pattern of error types in the examined RCs categories.

## Discussion

The results indicate that DS individuals encounter difficulties both in comprehending and producing RCs, as they performed at a significant lower level than the TDC. The DS group performed better in the comprehension task than in the production, while TDC performed almost equally well in both tasks. Moreover, both groups performed significantly better in SG relatives than in OG relatives, with the overall performance of the DS group being at a lower level in both tasks. The same pattern of performance observed in both groups, as well as the same pattern of errors produced gives support to the hypothesis that language development in DS could be characterized as more delayed than deviant. Our findings are in line with previous cross-linguistic studies the results of which revealed difficulties with RCs on the part of individuals with DS (Stathopoulou 2007, Witecy & Penke 2017). As the RC construction has not been extensively studied yet, especially in the Greek language, our findings help to fill in this research gap. However, more research is needed in this field in

order to elucidate further the delay hypothesis especially in the morphosyntactic domain and understand the kind of difficulties Greek speaking individuals with DS present so as to help them improve their morphosyntactic abilities.

### **Acknowledgements**

We acknowledge a research grant from the Greek State Scholarship Foundation (I.K.Y.) awarded to the second author.

### **References**

- Katsarou, D., Andreou, G. 2022. Morphosyntactic abilities in young children with Down syndrome: Evidence from the Greek language. *International Journal of Language & Communication Disorders*, 1-11.
- McKee, C., McDaniel, D. 2001. Resumptive Pronouns in English Relative Clauses. *Language Acquisition* 9, 113-156.
- Stathopoulou, N. 2007. Producing Relative Clauses in Greek: Evidence from Down Syndrome. *Essex Graduate Student Papers in Language & Linguistics* 9, 104-125, Department of Language & Linguistics, University of Essex.
- Stavrakaki, S. 2001. Specific Language Impairment in Greek: Aspects of Syntactic Production and Comprehension. PhD Thesis, Aristotle University of Thessaloniki.
- Witecy, B., Penke, M. 2017. Language comprehension in children, adolescents, and adults with Down syndrome. *Research in Developmental Disabilities* 62, 184-196.

# Evaluation and reading time of predicate agreement with conjuncts

Daria Belova

Department of Theoretical and Applied Linguistics, Lomonosov Moscow State University, Russian Federation  
Interdisciplinary Scientific and Educational School of Moscow University “Brain, Cognitive Systems, Artificial Intelligence”, Russian Federation

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

This paper aims to investigate possible patterns of predicate agreement with coordinated pronominal subjects in Russian. We have conducted two experiments to examine the effect of two factors on acceptability and reading time of different verbal forms, namely: the order of conjuncts exhibiting different grammatical features (‘1sg – 2sg’ or ‘2sg – 1sg’) and the order of subject and verb (SVO or OVS). The experimental results showed the differences in the agreement patterns related to the word order as the respondents more often allowed the less frequently encountered closest adjunct agreement in OVS-stimuli than in SVO-ones.

Keywords: Closest Conjunct Agreement, Personal Hierarchy, Coordinate Subject

## Strategies of agreement with coordinated subject

With respect to phi-features, a coordinated subject may contain more than one value, therefore, there is a need for some sort of a strategy for choosing the form of an agreeing predicate. Three basic strategies are described in the literature. The first one is *resolution*: the value is calculated depending on what values the conjuncts have. For example, the person feature can be chosen based on the *person hierarchy* (Zwicky 1997); in the meantime, the number feature can be simply copied from the one that the coordinate phrase has, which is always plural. The second strategy is *partial agreement*: a predicate gets all the values from one of the conjuncts, for example, the linearly closest (closest conjunct agreement, CCA (Al Khalaf 2015)), and the second conjunct is ignored. Finally, if none of these strategies is applied, the default agreement can appear as a last resort option (Nevins, Weisser 2018). Predicate agreement with coordinated subjects has also been investigated by means of the experimental methods, mainly speech production ones, and great variability has been observed, cf. (Timmermans et al. 2004) for German and Dutch and (Marušič et al. 2015) for Slovenian.

Russian prescriptive grammars provide only one strategy of personal agreement with a coordinated subject – the one that is based on the person hierarchy. However, corpus research show that the CCA in person and gender

is possible and facilitated by the postverbal subject position (Corbett 1985). Taking into consideration the experimental findings in other languages, we suppose that the acceptability of different person agreement strategies in Russian should be experimentally investigated.

### Research design

Two experiments that differ in the word order of stimuli (SVO and OVS) were conducted using the 1–7 Likert scale and the self-paced reading task. Both experiments shared the same lexicalizations and the 4×2 AJ experimental design that featured two independent variables: the verb form representing all three agreement strategies (1pl, 1sg/2sg, and 3pl, all in non-past tense) and the conjunct order (*ja i ty* ‘me and you’ and *ty i ja* ‘you and I’). Each experiment consisted of eight experimental lists containing 32 target stimuli and 32 grammatical and ungrammatical fillers. In (1) one can see the stimuli structures of the SVO and VSO experiments respectively.

- (1) a. [*ja i ty / ty i ja*] V Obj P NP  
 b. Obj P NP V [*ja i ty / ty i ja*]

### Results: the ratings

The results of the acceptability judgment task are presented in fig.1 below. All the ratings were z-transformed and then statistically processed with the use of linear mixed modeling (LMM) and a posterior Tukey’s HSD test.

The SVO-experiment involved 107 native Russian speakers (19–72 y.o., *mean*=38.59, *sd*=11.77). The final LMM (formula: z-scores ~ 1 + verb\_form) included respondent’s ID as random effect (formula: ~1 | id). The model’s total explanatory power is substantial (conditional R<sup>2</sup>=0.42) and the part related to the fixed effects alone (marginal R<sup>2</sup>) is of 0.39. The pairwise comparisons show the significant difference between all four verb forms: 1pl is rated highest, followed by 3pl, 1sg, and 2sg. There is no difference between two conjunct orders within each verb form (the blue and the green lines in the plot). The OVS-experiment involved 126 respondents (17–76 y.o., *mean*=34.54, *sd*=11.5). The final LMM (formula: z-scores ~ 1 + conjunct\_order \* verb\_form) included respondent’s ID and sentence ID as random effects (formula: (~1 | id) + (~1 | sentence\_id)). The model’s total explanatory power is substantial (conditional R<sup>2</sup> = 0.39) and the part related to the fixed effects alone (marginal R<sup>2</sup>) is of 0.28. The conjunct order is significant for 1sg and 2sg but not for 1pl and 3pl.

The overall low level of the target ratings in comparison to the grammatical fillers we relate to a pragmatically unusual context of the stimuli. Given that, the superiority of the resolution form 1pl is obvious. In the meantime, the CCA in the OVS-order is observed: the conjunct order ‘1sg – 2sg’ is rated higher when following the 1sg verb form, and the same is true for the 2sg forms. The non-partial agreement strategies are rated the same independently of the conjunct order or the word order.

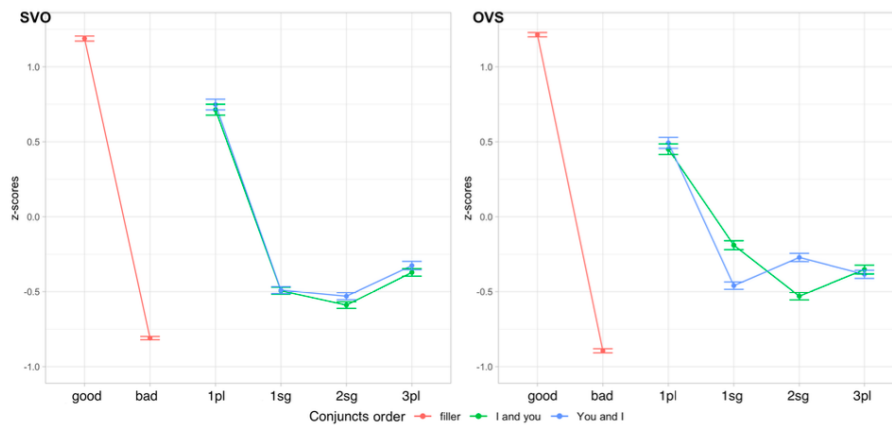


Figure 1. Normalized ratings from SVO & OVS experiments by the verb form.

### Results: the reading time

The starting hypotheses concerning the self-paced reading component of the experiments were the following: (i) in the SVO-experiment the verb forms differing from the basic 1pl would be read slower; (ii) in the OVS-experiment the first conjunct mismatching the verb person feature would be read slower than the matching one.

Neither of these predictions is borne out by the results. In the SVO-experiment the pairwise comparison with the use of the Student's t-test does not show any differences in the reading time of verbs between the two conjunct orders for any of the verb forms. In the VSO-experiment there is no difference found in the reading time of first nor second conjuncts, however, conjunctions are read faster when the conjunct order is *ja i ty* for the stimuli with every verb form except the 2sg, see fig.2 below.

These results are unexpected. The SVO-experiment shows that although different verb forms are clearly on different levels of acceptability, they all take the same amount of time to read and process. In turn, the OVS-experiment was presumably more likely to demonstrate the signs of the reanalysis. The alternative hypothesis, leading to other predictions, is that it is only after the first conjunct that a reader starts processing the whole coordinated phrase. In the stimuli with the singular verb forms where the verb stands close to the matching pronoun the conjunction *i* should be read slower because this pronoun can constitute a subject by itself. Meanwhile, in the stimuli with the plural verb forms, there should be no such pattern as the second pronoun is needed to match the verb number. Since this hypothesis is not confirmed either, the most suitable explanation for the data seems to be methodological. The pronouns *ja* and *ty* and the conjunction *i* are monosyllable and overall short words, which were presented separately, and it could influence a reader's pace of passing the experiment.

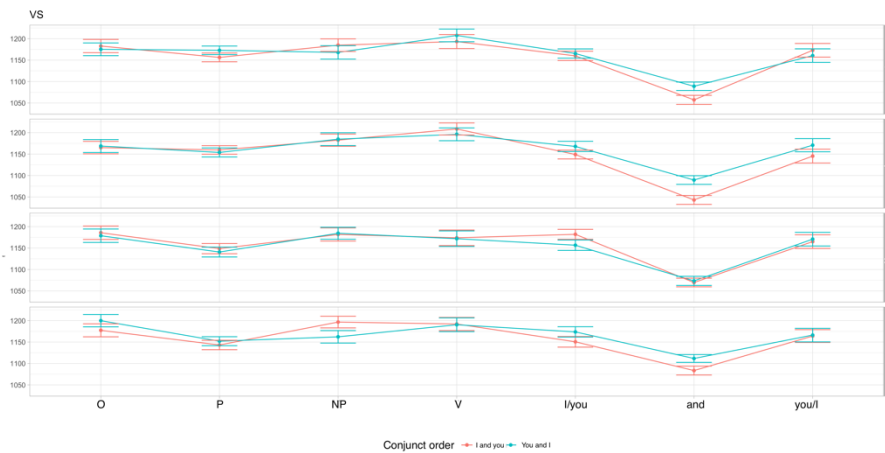


Figure 2. The mean reading time (in ms) of each element in OVS-stimuli.

## Conclusion

All three agreement strategies (person hierarchy resolution, CCA, and default agreement) are possible in Russian in the sense that they all are rated higher than the ungrammatical fillers. The postverbal subject position does indeed facilitate the partial agreement strategy compared to the preverbal position, hence, the previous conclusions based on the corpus data are confirmed. Another interesting result is the acceptability of the default agreement which is even higher than the partial agreement.

## Acknowledgements

This research is supported by Russian Science Foundation, RSF project 22-18-00037, Lomonosov Moscow State University, <https://rscf.ru/en/project/22-18-00037/>.

## References

- Zwicky, A.M. 1997. Hierarchies of Person. *Chicago Linguistics Society*, 714-733.
- Al Khalaf, E. 2015. Coordination and linear order. PhD thesis, University of Delaware.
- Marušič F.L., Nevins A., Badecker W. 2015. The grammars of conjunction agreement in Slovenian. *Syntax* 18(1), 39-77.
- Nevins A., Weisser P. 2018. Closest Conjunction Agreement. *Annual Review of Linguistics*, 5(1), 219-241.
- Corbett G. 1985. Hierarchies, Targets and Controllers. *Agreement Patterns in Slavic*. London & Canberra, Croom Helm
- Timmermans, M., Schriefers, H., Dijkstra, T., Haverkort, M. 2004. Disagreement on agreement: person agreement between coordinated subjects and verbs in Dutch and German. *Linguistics* 42(5), 905-929.



# Romanian children are not able not to derive actuality entailments

Simona Bilea<sup>1</sup>, Adina Camelia Bleotu<sup>1,2</sup>

<sup>1</sup>Department of Foreign Languages and Literatures, University of Bucharest, Romania

<sup>2</sup>Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Actuality entailments (AEs) are inferences from ability readings (“X could/was able to do Y”) to implicative readings (“X did Y”). They have been considered semantic inferences (Mari, Martin, 2007), arising from the speaker’s presupposition, or pragmatic implicatures (Bhatt 1999, Hacquard 2020) interacting with aspect. We investigate experimentally AEs and (double) negation in child Romanian. Like adults, Romanian 5-year-olds derive AEs with affirmative *Minnie a putut să coacă o piZZă* “Minnie was able to bake a pizza”, which suggests AEs are semantic in nature or pragmatic implicatures different from “some, not all”. Unlike adults, children derive few AEs with double negation in *Minnie nu a putut să nu coacă o piZZă* “Minnie was not able not to bake a pizza”.

Keywords: first language, actuality entailments, implicatures, negation, DN

## Introduction

Actuality entailments (AEs) are inferences from “X could/was able to do Y” to “X did Y”. A sentence such as *Minnie could make a cake* is ambiguous in English between (i) an ability reading (“She was able to”), and (ii) an implicative reading (“She actually did it”). Such readings have been considered either (a) semantic interpretations (Mari, Martin 2007, Mari 2015), arising from the speaker’s presupposition, or (b) pragmatic interpretations/implicatures (Bhatt 1999; Hacquard 2020), arising from the interaction with aspect (i.e., AEs arise only with perfective aspect in Romance). In the current paper, we examine AEs in child and adult Romanian, where AEs have not been investigated. We look at how children interpret sentences containing a main verb such as *a putea* ‘can’ followed by another predicate, when in interaction with negation or double negation (DN), e.g. “Bucle Aurii spune: *Minnie nu a putut să nu coacă o piZZă*. Donald se întrebă: *A copt Minnie o piZZă?*” (“Goldilocks says: *Minnie was able to bake a piZZă*. Donald wonders: *Did Minnie bake a piZZă?*”).

Previous studies on (related) telicity/completion entailments indicate an adult-like behaviour on the part of children (Stoicescu, Dressler 2022). However, previous studies on implicatures (Noveck 2001, Stoicescu et al. 2015) show that 5-year-olds derive fewer implicatures than adults in truth value judgment tasks (TVJTs), though there is task variation. Consequently, if AEs

are semantic, we expect adult-like rates of AEs in children, while, if they are pragmatic, we expect fewer AEs. Regarding negation, previous studies suggest that DN is challenging for children (Thornton et al. 2016, Tagliani 2019). Thus, we expect Romanian children to have difficulties with DN in Romanian, a negative concord language, where DN is interpreted as negative in meaning.

## Methodology

### Participants

The study involved 40 native Romanian speakers: a test group consisting of 20 pre-school children between ages 5 and 6 (Mean age: 5;5), recruited from a kindergarten in Bucharest, and a control group of 20 adults between 24 and 37 (Mean age: 33), of various academic backgrounds.

### Procedure

The experiment involved a truth-value judgement task, where participants had to help Donald figure out what Goldilocks was saying about other Disney characters: e.g., “Bucle Aurii spune: *Minnie nu a putut să nu coacă o pișcă.* Donald se întreabă: *A copt Minnie o pișcă?*” (“Goldilocks says: *Minnie was able to bake a pișcă.* Donald wonders: *Did Minnie bake a pișcă?*”). Participants could answer with *Yes/No/I don't know*. The task employed 16 critical items in 4 conditions: Affirmative (*Minnie a putut să* “Minnie was able to”), Negated Main Verb (*Minnie nu a putut să* “Minnie was not able to”), Negated Verbal Complement (*Minnie a putut să nu* “Minnie was able not to”), DN (*Minnie nu a putut să nu* “Minnie was not able not to”). The test conditions employed the implicative verbs *a putea* (“be able to”) and *a reuși* (“manage to”) followed by creation verbs (e.g., *a face*, “to make”) and change of state verbs (e.g. *a strica*, “break”). There were also 16 control items preventing a possible yes-bias. All items were used with perfective aspect to encourage AEs.

## Results

Just like adults, children interpreted affirmative sentences as giving rise to AEs, while they gave a negative interpretation to sentences in the Negated Main Verb condition and in the Negated Verbal Complement condition. However, while adults mostly interpreted DN affirmatively, deriving AEs, children mostly interpreted it negatively.

The statistical data were fitted into a generalized linear-mixed (logit) model (*glmer*) with Answer as the Dependent Variable (coded as 1 if accurate, and 0 otherwise), Age as fixed effect and random slopes per Item and Participant. There was a significant difference between Groups only for DN ( $\beta = 0.08043$ , SE = 0.02506,  $Z = 3.209$ ,  $p < .01$ ).

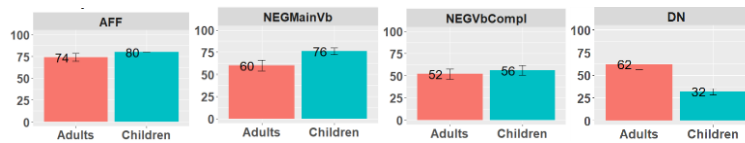


Figure 1. Rate of accurate answers (with SE) per group.

Legend: AFF=Affirmative, NEGMainVb=Negated Main Verb, NegVbCompl=Negated Verbal Complement, DN= Double negation

## Discussion

The current study shows that 5-year-old Romanian children are adult-like in their ability to derive AEs (“X did Y”) with implicative verbs in the affirmative condition and their ability to derive the opposite inference (“X did not do Y”) in conditions involving one single negation (either on the main verb or on the complement). If AEs were similar to *some not all* implicatures, then we would have expected them to derive fewer such inferences than adults, since children are known to have difficulties with *some not all* implicatures (Noveck, 2001, Stoicescu et al., 2015). However, we find the opposite result. Children’s ease with AEs suggests that AEs are best treated either as a semantic inference, which cannot be cancelled, or as a type of pragmatic implicature different from “some, not all”, perhaps of a more conventionalized nature (Bhatt, 1999). Nevertheless, children may have used different cognitive strategies than adults when agreeing with a sentence, such as *premature closure*, i.e., the tendency to choose one alternative out of several (Ozturk & Papafragou, 2015; Bleotu, Benz & Gotzner, 2021).

Regarding DN, our prediction that children will not derive AEs in DN contexts is met. Interestingly, children’s failure with AEs in DN contexts is not because of an inability to handle AEs, but because of interpreting DN negatively rather than affirmatively. This is in line with a general cross-linguistic negative-concord preference in child language.

## Conclusion

Our study represents the first experimental study on AEs in child language in Romanian. We find that, just like adults, children are able to derive AEs in the affirmative and the opposite inferences in the negative, when negation is either on the implicative or on its complement. However, unlike adults, children are unable to interpret DN affirmatively and, consequently, they do not derive AEs in these contexts. Thus, children are at a developmental stage where they are not fully adult-like with respect to DN and entailment readings in such contexts. Further research is needed to clarify the exact semantic/pragmatic nature of AEs.

## References

- Bhatt, R. 1999. Ability modals and their actuality entailments. University of Pennsylvania/Massachusetts, Institute of Technology. Dissertation.
- Bleotu, A. C., Benz, A., Gotzner, N. 2021. Shadow-playing with Romanian 5-year-olds. Epistemic adverbs are a kind of magic! *Experiments in Linguistic Meaning* 1, 59–70.
- Hacquard, Valentine. 2020. Actuality entailments. In Gutzmann, D., Matthewson, L., Meier, C., Rullmann, H., Zimmermann, T.E. (eds.) *Companion to Semantics*. Wiley.
- Mari, A. 2015. Actuality entailments: When the modality is in the presupposition. *LACL* 2016, vol. 10054. 191–210.
- Mari, A., Martin, F. 2007. Tense, abilities and actuality Entailment. In Aloni, M., Dekker, P., Roelofsen, F. (eds.) *Proceedings of the Sixteenth Amsterdam Colloquium*, 151–157. Amsterdam, Palteam.
- Noveck, I. 2001. When children are more logical than adults. *Cognition* 78(2), 165–188.
- Ozturk, O., Papafragou, A. 2015. The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development* 11(3), 191–214.
- Stoicescu, I., Sevcenco, A., Avram, L. 2015. The acquisition of scalar implicatures in child Romanian. In Burada, M., Tatu, O. (eds.) *Proceedings of Conference on British and American Studies*, 141–155. Newcastle upon Tyne, Cambridge Scholars Publishing.
- Stoicescu, I., Dressler, W.U. 2022. On the acquisition of semantic vs pragmatic telicity in child Romanian. *Italian Journal of Linguistics* 33 (2).
- Tagliani, Marta. 2019. The Acquisition of double negation in Italian. In Emonds, J., Janebová, M., Veselovská, L., *Proceedings of the Olomouc Linguistics Colloquium* 2018, 99–115.
- Thornton, R., Notley, A., Moscati, V., Crain, S. 2016. Two negations for the price of one. *Glossa, a journal of general linguistics* 1(1), 45, 1-30.

# Language environmental analysis (LENA) of three Cypriot Greek-speaking children

Paris Binos, Kakia Petinou

Department of Rehabilitation Sciences, Cyprus University of Technology, Cyprus

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

This is the first study that explores the quantity and quality of language input in Cypriot Greek-speaking families using highly specialised software. The purpose of this longitudinal case study was to record, analyse and evaluate the longitudinal changes in conversational turns, input and participant vocalisations in Cypriot Greek-speaking families using the computerised LENA (Language Environment Analysis) software. The study involved two typically developing children (TD) and a child with a Language Disorder (LD). TDs' performance and language input confirmed the norms of typically developing scores of the LENA system. In contrast, LD differed quantitatively and qualitatively from the mean number of CVC and CTC. The mean numbers of LD measured below the <25th percentile in any measurement and the quantity of language input. Recent research supports the findings of the present study.

Keywords: LENA, language input, vocalisation, prelinguistic, language disorder

## Introduction

LENA system is a new tool combining a wearable audio recorder with automated vocal analysis software (LENA Research Foundation, 2014). LENA was developed to overcome the difficulties in obtaining data from young children and conducting detailed analyses of language samples. LENA system investigates various linguistic aspects (Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011; Sunsking et al., 2016).

Recently, Wang et al. (2020) reported a statistically reliable correlation between adults' words and children's language skills. Thus, the LENA system estimates of in-home language input since the 30-million-word gap project showed that parents were encouraged to increase the quality and quantity of their turns. Another milestone study by Gilkerson et al. (2018) recently demonstrated long-term associations between early language exposure and subsequent child language and IQ through late middle school through LENA and their naturalistic methodology. Thus, they have proved that turn-taking predicts crucial outcomes by age 14. The authors affected the clinical practice recommendations of the American Academy of Pediatrics.

## **Aim of the study**

This study aims to record for the first time in Cypriot-Greek dialect and Greek language how much exposure is needed for language acquisition since a language is acquired through exposure. Also, are there any recorded differences among TDs and LD participants?

## **Method**

This is a longitudinal study of three cases included in-home recordings where two Cypriot Greek-speaking families and their typically developing children (6 and 46 months old) (TDs) and a Cypriot Greek-child (81 months old) (LD) diagnosed with a language disorder participated in the study. TDs were born full-term without neurodevelopmental or hearing disorders. All participants were recruited through the Cyprus University of Technology University Rehab Clinic and their participation approved by the Research Ethics Board of the Clinic and the University.

LENA's hardware includes a digital language processor (DLP), a USB port for data transfer, and a display screen that is held in a t-shirt worn by the participant with a pocket on the front. The device records in 16-bit, and when it is full, the researcher can upload it to the software since a full day consists of 20 to 50K segments (VanDam et al., 2016). The software classifies each segment based on volume and suprasegmental speech features as one of seven categories: adult male/female, another child, essential child, TV, noise, silence and overlapping sounds. Other sounds, defined as "vegetative sounds" (burping, sneezing) or "fixed signals" (crying, laughing) measured and classified based on their prosodic features (intonation, duration) (Oller et al., 2010). The segments classified as "vocalizations" counted as CVC and adult speech as AWC. Vocalization turns produced within five seconds of an adult's production without interruption are classified as the contiguous, linguistic turn-taking CTC. LENA system generates an estimated expressive language standard score calculated on the quality of the child's productions called automatic vocalization assessment (AVA) and provides a comparison to typically developing peers.

The software requires recordings to be at least 10 hours long to complete a full analysis, including percentile rankings. The audio data (>150 hours overall) was processed and classified by the LENA Pro software into seven sound categories. The speech and turn-takings of the six months old TD was recorded and analysed based on recordings in two months that fulfilled the system's criteria for analysis. The analysis of 46 months old TD was based on four recordings for two months. Accordingly, the analysis for LD child was based on 46 hours of three recordings in two months. An SLP explained the proper wearing and use of the DLP to the parents asking parents to have the child wear the recorder in the unique LENA clothing. Parents received an instruction sheet and were asked to provide information on whether the recording day seemed to represent a typical day for the child.

## Results

### Auditory environment

Results for the auditory environment by the categories are divided into five categories: 1) TV and electronic sounds, 2) silence and background noise, 3) noise, 4) meaningful speech, and 5) distant speech. The TV and electronic sounds time for TDs ranged from 3 to 9 hours. The overall time for LD child was 4 hours respectively. The amount of exposure to meaningful speech was found 4 ½ to 8 hours for TDs, while the corresponding time for LD was only 4 hours. The time of the distant speech was measured between 7 ½ to 16 hours for TDs, while the distant speech time of LD was found 16 ½ hours. The noise was measured for all the participants for almost 1 hour. Finally, the silence and background time for TDs was measured long since, ranging from 21 to 27 hours overall.

### Vocalizations and turns

The mean number of vocalizations of LD (46 months old) was 990 words classified at the 50<sup>o</sup> percentile, while the mean number of LD (6 months old) was 529 vocalizations classified at the 10<sup>o</sup> percentile of the LENA system. The performance of LD was measured at 313 words and classified at the 10<sup>o</sup> percentile of the norms. The adult words of TDs were classified at 75-90<sup>o</sup> and 70<sup>o</sup> percentile accordingly. Regarding the concern, the mean conversational turns were found at 270 (50<sup>o</sup> percentile) for six months old TD and 223 (10<sup>o</sup> percentile) for the other TD child. The LD child had only 85 turns classified at the 10<sup>o</sup> percentile of the system since the 50<sup>o</sup> percentile is measured at 200 turns.

## Discussion

The purpose of this study was to present in detail data for the first time in Cypriot-Greek and Greek about the informal environment of TD and LD participants. The present study is part of a more extensive study involving more participants from both groups. We are particularly interested in revealing subtle communicative differences among TDs and LDs, especially in Cypriot-Greek dialect, so we can differentiate our clinical protocols in Speech-Language Pathology, providing the parents with the opportunity to be more “communicative” qualitatively and quantitatively. AWC, CT and CV variables of the LENA system proved reliable in many studies and linked the performance to possible LDs (Ganek & Eriks-Brophy, 2018; Thiemann-Bourque et al., 2014).

According to our results, the adult words that LD child received were few, only at the 10<sup>o</sup> percentile. The same stands for the turns since the mean number was calculated at 85 and the 10<sup>o</sup> percentile. Thus, there is a significant mean difference between the adult words that LD child received (6.800) and

the vocalizations that TDs received (16.000) from their parents. The results revealed a delay for the turns as well. The LD child received only a mean number of 85, while TDs ranged from 223 to 270 turns. In sum, the study also revealed the need for intensive counselling of the parents by changing clinical protocols in Speech Pathology, following other studies that revealed similar needs (Suskind et al., 2013).

### Acknowledgements

The authors warmly thank all families that participated in this study. Our special thanks go to our SLT students who helped and participated when necessary.

### References

- Ganek, H., Eriks-Brophy, A. 2018. Language environment analysis (LENA) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders*, 72, 77–85.  
Doi: 10.1016/j.jcomdis.2017.12.005
- Gilkerson, J., Richards, J., Steven, W., Oller, K., Russo, R., Vohr, B. 2018. Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, 142(2), e20174276.
- Greenwood, C., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, 32(2), 83-92.
- Oller, K., Niyogi, P., Gray, S., Richards, J., Gilkerson, J., Xu, D., Yapanel, U., Warren, S. 2010. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences of the USA*, 107(30), 13354-13359.
- Suskind, D., Leffel, K. R., Hernandez, M. W., Sapolich, S. G., Suskind, E., Kirkham, E., Meehan, P. (2013). An exploratory study of "quantitative linguistic feedback." *Communication Disorders Quarterly*, 34(4), 199–209.  
Doi: 10.1177/1525740112473146.
- Thiemann-Bourque, K. S., Warren, S. F., Brady, N., Gilkerson, J., & Richards, J. A. 2014. Vocal interaction between children with down syndrome and their parents. *American Journal of Speech-Language Pathology*, 23(3), 474–485. Doi: 10.1044/2014\_ajslp-12-0010.
- VanDam, M., Warlaumont, A., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, R., MacWhinney, B. 2016. HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech & Language*, 37(2), 128-141.
- Wang, Y., Williams, R., Dille, L., Houston, D. 2020. A meta-analysis of the predictability of LENA automated measures for child language development. *Developmental Review*, 57(September).  
Doi: 10.1016/j.dr.2020.100921



# The role of presentation modality in sentence processing

Daria Chernova<sup>1</sup>, Artem Novozhilov<sup>2</sup>, Natalia Sliouzar<sup>3</sup>

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

<sup>2</sup>Institute for Cognitive Studies, St.Petersburg State University, Russia

<sup>3</sup>School of Linguistics, Higher School of Economics, Russia

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

The question whether listening comprehension is more or less costly than reading comprehension remains open. We investigate whether the effects of grammar complexity are same in reading and listening sentences. We show that presentation mode has no effect on comprehension accuracy but affects response times. In both modes the most difficult constructions to comprehend and the longest constructions to process were the same which demonstrates that the cost of syntactic processing is not affected by presentation modality.

Keywords: syntactic complexity, sentence comprehension, Russian language.

## Introduction

The question whether listening comprehension is more or less costly than reading comprehension remains open. From one point of view, reading a more complicated and costly activity than oral language comprehension because reading ability is acquired later in ontogenesis than oral speech, and orthographic processing is mapped onto phonological representations, which; moreover, in listening comprehension we can use prosodic cues which we do not have in reading comprehension. From the other point of view, reading is less costly than listening because in most languages word boundaries are demarcated in written text unlike in spontaneous speech, and readers can process a word as long as they choose, in natural reading they also can refixate and make regressions, so oral language comprehension is more demanding because the processing pace is regulated by the speaker, not by the hearer.

Some research has been done on the effect of modality for grammatical judgment tasks. Vetter, Volovecky and Howell (1979) compared visual and auditory presentation, both normal and monotone, as well as simultaneous visual and auditory presentation, and found no overall effect of modality. As for neurocognitive research on the modality of sentence processing, there is neurological evidence for supramodal language system that integrates linguistic input from speech to print and activates a common code (Braze et al. 2011, Constable et al. 2004).

Our study was aimed to investigate whether the effects of grammar complexity are same in reading and listening modes of the sentence comprehension task.

## Material

The task included syntactically complex 60 target sentences of 6 types: object relative clauses, locative constructions, temporal constructions, sentences with high and low modifier attachment to a complex noun phrase and complex comparative constructions, as well as 40 fillers with simpler syntactic structure (see Chernova, Novozhilov, Slioussar 2021).

For every sentence, we created a comprehension question with a choice of two answers aimed to assess syntactic structure comprehension. Both response options were mentioned in the sentence but only one was the correct answer to the question. All the sentences were semantically reversible and unbiased, so both response options referred to equally plausible interpretations, so the syntactic structure of the sentence had to be analyzed to give a correct answer.

The test consisted of two parts: reading mode and listening mode. In the first part the sentences were presented in a non-cumulative self-paced reading mode. In the second part the sentences were presented as audio recordings (recorded by a male native-Russian speaker who read with a natural and consistent pace and volume and was unaware of the purpose of the study). The test had a within-subject design: all participants read and heard all the sentences, which were randomized in each protocol.

## Participants

98 native speakers of Russian (48 male/ 50 female) aged 19–63 volunteered to take part in the experiment which was run online using PCIBex platform, <https://farm.pcibex.net/> (Zehr & Schwarz 2018). All the participants provided informed consent and were unaware of the purpose of the study.

## Procedure

The participants were asked to read or listen to the sentence and then answer a comprehension question. It was possible to read or listen to the sentence only once. After that the participant was presented a comprehension question and chose a response option by mouse click. Answer accuracy and response times were registered.

## Results

We analyzed participants' question answering times and accuracy. No differences in comprehension accuracy between reading and listening mode was found:  $\beta=0.1$ ,  $z=0.8$ ,  $p=0.3$ . As for construction types, low attachment sentences were processed the least accurately both in reading and listening

mode,  $p < 0.01$  for most pairwise comparisons, while temporal sentences were the easiest to interpret in both modes,  $p < 0.01$  for most pairwise comparisons.

It took significantly more time to answer the questions in listening mode compared to reading mode:  $\beta = 703.1$ ,  $t = 4.5$ ,  $p < 0.001$ . Complex comparative constructions and locative constructions were the longest to give an answer both in reading and listening modes,  $p < 0.01$  for most pairwise comparisons. The shortest answering times were registered for high and low attachment and temporal constructions.

## Discussion

As we see, presentation mode has no effect on overall comprehension accuracy. However, it affects response times: it takes significantly more time to answer a question if the sentence was presented orally than if the sentence was presented in the written form. This effect may be caused by the procedure of the experiment: as soon the comprehension questions were always presented in the written form in both parts of the experiment, the increasing response times may reflect the modality switch effect.

What is more important, the most and the least difficult constructions to comprehend are the same in both modes: constructions with low attachment of the modifier to a complex noun phrase provoked significantly more errors while temporal constructions provoked significantly less errors both in reading comprehension task and listening comprehension task. Complex comparative and locative constructions took the longest time to give a response also in both modalities. These data gives evidence that the same syntactic processing system seems to be is used in both modalities.

## Acknowledgements

The study is supported by RSF, project № 21-18-00429

## References

- Braze, D., Mencl, W. E., Tabor, W. et al. 2011. Unification of sentence processing via ear and eye: An fmri study. *Cortex*, 47:4, 416–431. doi:10.1016/j.cortex.2009.11.005
- Chernova D., Novozhilov A., Slioussar N. 2021. Sentence comprehension assessment in Russian. In: A. Botinis (ed.), *Proceedings ExLing 2021*, 45-48. Doi: 10.36505/ExLing-2021/12/0012/000485.
- Constable, R.T., Pugh, K.R., Berroya, E., Mencl, W.E., Westerveld, M., Ni, W., Shankweiler, D. 2004. Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *NeuroImage*, 22(1), 11–21. <https://doi.org/10.1016/j.neuroimage.2004.01.001>
- Vetter, H.J., Volovecky, J., Howell, R.W. 1979. Judgments of grammaticalness: A partial replication and extension. *Journal of Psycholinguistic Research*, 8, 567-583.
- Zehr, J., Schwarz, F. 2018. PennController for Internet Based Experiments :IBEX.doi:10.17605/OSF.IO/MD832



# Image complexity in the tracking of DLD

Letícia Corrêa<sup>1</sup>, Erica Rodrigues<sup>1</sup>, Marina Augusto<sup>2</sup>

<sup>1</sup>Department of Letters, Pontifical Catholic University of Rio de Janeiro, Brazil

<sup>2</sup>Department of Language Studies, Rio de Janeiro State University, Brazil

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Image complexity in a picture-identification task (PI-task) is shown to affect children's comprehension of relative clauses (RCs), object RCs, in particular. Images depicting reversible thematic relations led to more errors than non-reversible images for 5-7-year-olds. Schoolchildren identified as at risk of Developmental Language Disorder (DLD) had poorer performance than their age-matched control in the comprehension of RCs and which-questions (particularly in object sentences with complex images). Results of a correlation between children's scores on a language assessment test by means of a PI-Task and two non-verbal inhibitory control tasks suggest low inhibitory capacity makes it more difficult for children at risk of DLD to cope with costly sentences.

Keywords: relative clauses; sentence comprehension; image complexity; executive functions; reversibility.

## Introduction

This paper is concerned with the task demands in the assessment of children's comprehension of costly sentences and their implications for the tracking and understanding of language impairment in the syntactic domain. It focuses on the comprehension of relative clauses (RCs) and which-questions (WQ) in a picture identification task (PI-task). The asymmetry between subject and object RCs (SRCs/ORCs) with the latter being particularly demanding for both children and adults is widely attested across languages (Lau & Tanaka, 2021). The ability to comprehend ORCs has been presented as a late acquisition and as a possible index of language impairment in the syntactic domain due to Extended Relativized Minimality (ERM), an overextension of a language-specific principle (Friedmann et al., 2009). There is, nevertheless, the possibility of underestimation of children's language abilities. Language-independent factors may contribute to children's difficulties in the task.

The comprehension of RCs has been traditionally assessed by means of a PI-task in which children have to choose one out of two images in which the same characters perform the same action described by the RC, though assuming reverse thematic roles. Hence, children have to decide on the correct picture by keeping in mind the thematic roles of the subject and of the object of the RC while distinguishing in which picture the same characters assume the respective actor and patient roles. The competing demands of sentence and image processing in the comprehension of costly structures were characterized

(Rodrigues & Correa, 2013). In Correa & Augusto (2019), an effect of image complexity (IC) was obtained in the number of correct responses given by schoolchildren to RCs and Who-questions, suggesting that the difficulty imposed by the comparison of similar reversible semantic relations may confound the factors predominantly affecting children's performance: a domain-specific effect of ERM or inefficient inhibitory control. The aims of this paper are: to report an experiment in which the effect of IC (reversible actor-patient relations) is verified; to evaluate the impact of IC on the performance of children identified as in risk of DLD (R-DLD) and the extent to which coping with a PI-task may rely upon inhibitory control abilities.

### Reversible and non-reversible images

In this experiment, the same setup of Correa & Augusto's study (2019) was used. A background picture shows two different characters of the same type, corresponding to the subject/object of the RC. The background scene presents the two characters of the same type performing the same action with reversed roles (eg. one lion pushes a bear; a bear pushes the other lion, reversible actions (RA)) or one of them as the actor-patient of a different action (non-reversible action (NRA))<sup>1</sup>. Children have to choose among three options (the two characters of the same type and another corresponding to the subject/object of the RC), the one corresponding to the head noun (see Figs. 1 and 2). *Type of RC* (subject and object) and *image complexity (IC)* (RA – NRA) were the independent variables (within and between factors, respectively).

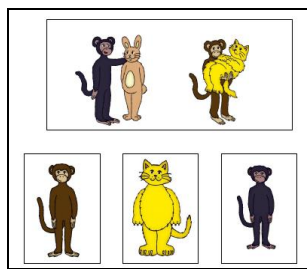


Figure 1. NRA (...the monkey that carried the cat).

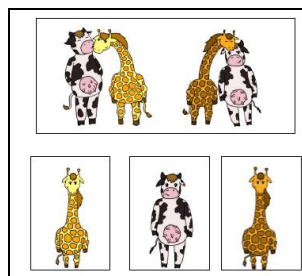


Figure 2. RA(...the giraffe that the cow kissed).

The participants were 34 middle-class Brazilian Portuguese (BP)-speaking children (age range 5.0 to 7.5), equally divided into two groups (NRA: 7 girls, mean age 6.0 years old; RA: 7 girls, mean age 5.8 years). 4 sentences of each type (1-2) (randomized in 4 lists) were presented, with 3 distractor sentences (subject who-questions) in a fixed order. The same sentences were orally presented to the two groups, with images in PowerPoint slides. Children were tested in an isolated room of the school. The procedure took around 10 min.

1. Mostra o macaco que carregou o gato. (Subject RC)  
[Show (me) the monkey that carried the cat]
2. Mostra o elefante que o urso puxou. (Object RC)  
[Show (me) the elephant that carried the bear pulled]

## Results

Correct responses were analysed with R Version 4.2.1(2022-06-23), and the *lme4* package (Bates et al., 2015). A GLMM, using the maximum likelihood method in R (Laplace Appr.) was fitted, using binomial (logit) distribution, with *Type of RC* and *IC* as fixed effects, and participant as a random effect (Baayen et al., 2008). This analysis revealed that children performed significantly better with SRCs than ORCs with complex images (estimate = 1.2442, SE = 0.4016,  $z = 3.098$ ,  $\Pr(|z|) = 0.00195$ ) and ORCs were impacted by IC (estimate = 1.8687, SE = 0.4684,  $z = 3.990$ ,  $\Pr(|z|) = 6.61e-05$ ).

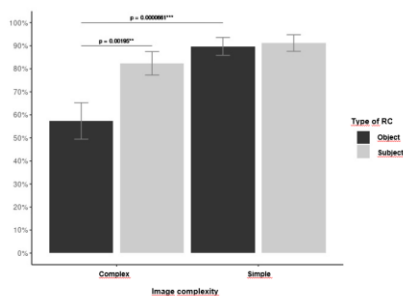


Figure 3. Percentage of correct answers as a function of conditions.

The results show that reversible images add difficulty to the task, particularly in the most demanding sentences, and reveal that the SRC-ORC asymmetry can be dissolved when RCs can be integrated with a non-reversible image. What is the impact of IC in the tracking of DLD by means of a PI-task?

## Image effects, DLD and inhibitory control abilities

Two groups were created from a sample of 263 BP-speaking schoolchildren previously submitted to the MABILIN battery:<sup>2</sup> 15 whose overall performance was suggestive of risk of DLD (6 girls; mean age 7.2 years) (R-DLD group); 15 with high scores (6 girls, mean age 6.7 years), in the matched control (CTL) group). Subject/object RCs and WQs with simple (SI) (see note 1) and complex images (CI) were selected for the present analysis. There were 4 sentences in each condition. Additionally, the performance of 24 different children in the whole MABILIN battery was correlated with their performance in two inhibitory control tests Go/Nogo test and Flanker test.

## Results

The number of correct responses in a Mann-Whitney test differed significantly in the R-DLD and the CTL group both in the simple ( $U=55.500$ ,  $z=-2.403$ ,  $p=.016$ ) and complex image conditions ( $U=22.00$ ,  $z=-3.784$ ,  $p=.0001$ ). For SI, which is more likely to capture difficulties in syntactic processing, there were significant differences between R-DLD and CTL only for ORCs ( $U=69.000$ ,  $z=-2.004$ ,  $p=.045$ ). For CI, more likely to capture the additional burden of a reversible image, the groups differed significantly for SRCs ( $U=60.000$ ,  $z=-2.480$ ,  $p=.013$ ); ORCs ( $U=45.500$ ;  $z=-2.860$ ,  $p=.004$ ) and OWQs ( $U=41.500$ ,  $z=-3.029$ ,  $p=.002$ ). There was a positive correlation between the total score on MABILIN and the total of correct responses on Flanker ( $\rho(24)=0.48$ ,  $p=.02$ ) and on Go/Nogo test ( $\rho(24)=0.41$ ,  $p=.04$ ). No correlations were obtained for response time.

## Final remarks

The assessment of children's language comprehension abilities in the tracking of DLD has to take into account the specific task demands. The differential demands of simple and complex images in a PI-task can contribute to distinguishing children whose difficulty is more likely to stem from linguistic factors (difficulty with simple images) from those who are additionally sensitive to the demands of reversible images on inhibitory control abilities.

## Notes

1. In Corrêa & Augusto (2019), the simple image had one of the characters of the same type standing by the scene.
2. The MABILIN battery for the tracking of DLD (Corrêa, 2000) includes simple reversible actives; non/reversible passives; right-branching and centre-embedded subject/object RCs (with in/transitive verbs), who and which questions.

## Acknowledgements

FAPERJ: E-26/200.903/2021; CNPq 312123/2020-5; CNPq 311422/2019-5; Prociência 2020-FAPERJ.

## References

- Baayen, R. H. et al. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *J. Mem. Lang.*, 59(4), 390-412.
- Corrêa, L., Augusto, M. 2019. Image effects and long-distance dependencies, *Proceedings of the 10th Int. Conf. of Exp. Linguistics*, 53-56, Lisbon, Portugal.
- Friedmann, N., Belletti, A., & Rizzi, L. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119(1), 67-88.
- Rodrigues, E., Corrêa, L. 2013. Controle executivo na resolução de tarefas linguísticas, II Enc. Intern. de Psicolinguística da ANPOLL, JP, Brazil.



# The annotation of information structure in spoken Chinese: a pilot study

Emanuela Cresti, Shuai Luo, Massimo Moneglia, Alessandro Panunzi

LABLITA research unit, University of Florence, Italy

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

The paper sketches a pilot aimed at verifying the consistency of the L-AcT principles and tagging methodology for the annotation of Chinese speech. The pilot confirms both the correspondence between Prosodic Units and Information Units foreseen in this theory and the applicability of its tag-set to Chinese speech.

Keywords: Information Structure, Speech Segmentation, Chinese

## Introduction

According to the Language into AcT theory (L-AcT, Cresti & Moneglia 2018), the speech continuum is parsed into prosodic units (PUs) by boundary phenomena that correlate with pauses, pitch reset, lowering of intensity, and lengthening. Prosodic Boundaries are highly perceivable in connection to intentional movements on stressed syllables ('t Hart 1990) and define the edges of PUs, which signal information functions. Terminal prosodic breaks (TB) mark the utterance, i.e. the minimal unit which can be pragmatically interpreted. Non-terminal prosodic breaks (NTB) signal the information units (IUs).

This paper sketches a pilot study aimed at verifying the consistency of the above principles and tagging methodology specifically with respect to the relation between prosody and information structure in Chinese spontaneous speech. The pilot is the preliminary work of a doctoral project which foresees the collection and analysis of a new corpus of spoken Chinese "C-ORAL-CIN" (10 hours, with diaphasic variation) implementing Chinese in the IPIC Information structure data base. In 2. we will show the correspondence with PUs of the IUs of Comment (COM), Topic (TOP), Appendix (APC), Locutive introducer (INT), Parenthesis (PAR) and Bound Comments (COB) and the consistency of their informational value according to L-AcT's tag-set (see. Moneglia & Raso 2014 and the definitions therein).

## The annotation of the data set according to L-AcT

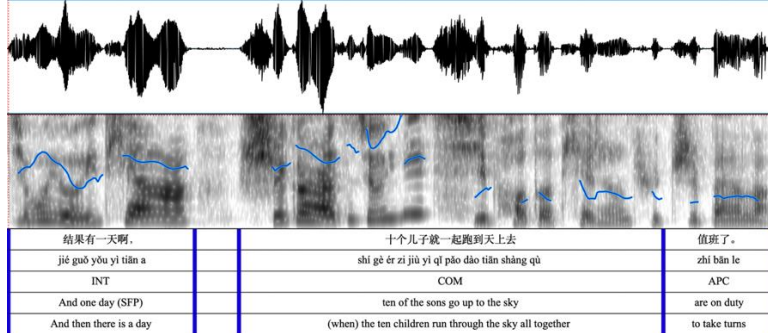
The pilot deals with the annotation of a story told to a child by his mother (the Chinese fairy tale of Hou Yi). The data set has been transcribed in characters and transliterated in pinyin. The prosodic parsing of the stretch of speech marking TB and NTB is provided following the perceptual recognition by a

mother tongue. In parallel, the sound wave is analysed through WinPitch and PRAAT, allowing the verification of perceptual judgements. Finally, the word-by-word translation of each chunk in English, and translation in current language is added. In agreement with L-AcT's principles, dedicated layers are provided for the segmentation of the wave in connection with TB and NTB.

The correspondence of terminated sequences with reference units such as *utterances* and *stanzas* has been verified (Izre'el et al.2020) dividing the recording into 90 reference units. To each PU is assigned an IU tag according to the L-AcT tagset. (1) allows to verify the main principle of the L-AcT approach: within an utterance compound by more than one PU, one IU is at the core (the COM), since it bears the illocutionary information. The COM can be pragmatically interpreted for this reason even if the other IUs (in this case INT and APC) are erased.

(1) \*LYF: 结果有一天啊, 十个儿子就一起跑到天上去, 值班了。

jié guǒ yǒu yì tiān a/<sup>INT</sup>shí gè ér zǐ jiù yì qǐ pǎo dào tiān shàng qù/<sup>COM</sup>zhí bān le//<sup>APC</sup>  
 And one day (SFP)/<sup>INT</sup>ten of the sons go up to the sky/<sup>COM</sup>are on duty//<sup>APC</sup>  
 [then there is a day (when) the ten children run through the sky together to take turns]

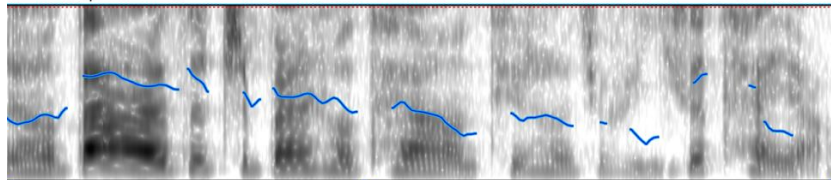


It is worth noticing that, in Chinese, from a semantic point of view, the locutive content in the APC (值班了-*zhí bān le*) could be in principle a sentence. But this is not the case considering its actual flat prosodic performance, which is not consistent with any possible pragmatic interpretation of the unit. According to L-AcT definition, the APC just *adds unnecessary information to the COM* (4 APC in this text). So, prosody defines the structure of the speech flow, segmenting the same utterance into three IUs instead of parsing it in two utterances.

Although simple utterances occupy an essential role in Chinese speech (43% in the tale), the IUs of TOP, COB and PAR have been retrieved. (2) is an example of TOP/COM utterance, very frequent in Chinese (Li & Thompson 1981), which records 38% of terminated units in mother's speech.

(2) \*LYF: 然后大儿子值班的时候, 太阳天上就是一个太阳。

rán hòu dà ér zǐ zhí bān de shí hou /<sup>TOP</sup> tài yáng [ / ] tiān shàng jiù shì yí gè tài yáng // <sup>COM</sup>  
 then elder son makes turn when / <sup>TOP</sup> sun [ / ] sky over exactly be one (CL) sun // <sup>COM</sup>  
 [When the elder son makes his turn, the sun [ / ] there is in the sky only one sun.]



然后大儿子值班的时候,	太阳天上就是一个太阳。
rán hòu dà ér zǐ zhí bān de shí hou	tài yáng [ / ] tiān shàng jiù shì yí gè tài yáng
TOP	COM
then elder son make turn when	sun [ / ] sky over exactly be one (CL) sun
When the elder son makes his turn,	the sun [ / ] there is in the sky only one sun.

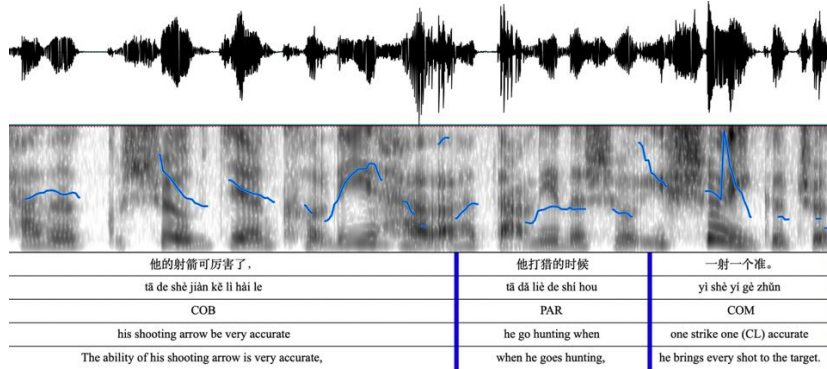
In (2) the TOP is a conditional subordinate sentence, and the COM is a nominal predicate conveying an assertive illocution. Coherently with L-ACI definition, the TOP works as *the pragmatic domain relevant for the interpretation of the illocutionary act accomplished by the COM*.

As it is frequent in monologues, terminated units go beyond the utterance limits. In our tale too (19% in mother's speech) sequences ending with a TB are made by Comments bound the one to the other in an adjunctive process (COB). These sequences are called *Stanzas*.

(3) \*LYF: 他的射箭可厉害了, 他打猎的时候, 一射一个准。

tā de shè jiàn kě lì hài le / <sup>COB</sup> tā dǎ liè de shí hou / <sup>PAR</sup> yì shè yí gè zhǔn // <sup>COM</sup>  
 his shooting arrow be very accurate / <sup>COB</sup> he go hunting when / <sup>PAR</sup> one strike one (CL)  
 accurate // <sup>COM</sup>

[The ability of his shooting arrow is very accurate, when he goes hunting, he brings every shot to the target.]



In (3) the first COB is in principle interpretable, but it does not end with a terminal break. On the contrary prosody marks the continuation of the discourse. It corresponds to an assertion which frames the utterance, while the final COM adds an explanation (his shooting-arrow ability is strong).

As it also happens in Romance corpora COBs in a stanza are interpolated by IUs, which add incidental information. These units (PAR), 4 instances in our tale, correspond to prosodic units with less  $f_0$  variation, performed at a lower-mid  $f_0$  values. Accordingly, in (3) PAR, which is a conditional subordinate sentence, shows a  $f_0$  profile at a lower frequency level than the other IUs. The PAR frames the situation, but prosody signals that the circumstance (hunting) is reported “on a different plan” and even if the speaker omits this part, the meaning of the overall terminated sequence still holds.

In conclusion, the research shows that in the Chinese pilot: (a) reference units (*utterance, stanza*) are marked by TB; (b) IUs correspond to PUs; (c) the IU typology of the L-Act framework are retrieved in spoken Chinese.

## References

- Cresti, E., Moneglia, M. 2018. The illocutionary basis of Information Structure. Language into Act Theory (L-Act), in: Adamou, E., Haude, K., Vanhove, M. (eds.) Information structure in lesser-described languages: Studies in prosody and syntax, 359-401. Amsterdam, Benjamins
- 't Hart, J., Collier, R., Cohen, A. 1990. A Perceptual Study on Intonation. Cambridge, Cambridge University Press.
- Izre'el, S., Mello, H., Panunzi, A., Raso, T. (eds.) 2020. In Search of Basic Units of Spoken Language: A Corpus-Driven Approach, Amsterdam, Benjamins.
- Li, C. N., & Thompson, S. A. 1981. Mandarin Chinese: A Functional Reference Grammar. Berkeley: University of California Press.
- Moneglia, M. & Raso, T. 2014. Notes on the Language into Act Theory, in: Raso, T., Mello, H. (eds.), Spoken corpora and linguistics studies. 468–494 Benjamins,

# Disentangling word order and function assignment preferences in Modern Greek

Yanis da Cunha, Ioanna Chorai, Anne Abeillé

LLF, Université Paris Cité, France

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Construction alternations are influenced by several conceptual properties of referents, including animacy or definiteness, known as “prominence features” (Aissen 1999, Haspelmath 2020). However, it remains unclear whether these features influence construction alternations through word order or function assignment. To disentangle these effects, we conducted a quantitative corpus study in Greek, comparing a word order alternation (SOV/OVS) and a function assignment alternation (active/passive). We annotated a random sample of 600 sentences from the Corpus of Modern Greek (Arkhangelskiy, Kisilier 2018) and fitted a Bayesian categorical model to construction (SVO/OVS/passive). We show that animacy and definiteness features have an effect on function assignment, not on word order, but we found a short-before-long length effect.

Keywords: Modern Greek, construction alternations, corpus study, Bayesian modelling

## Introduction

Construction alternations are influenced by several factors, including animacy, definiteness or length, as in dative alternation in English (Bresnan, Ford 2010). However, it remains unclear whether so-called “referential prominence features” (Aissen 1999, Haspelmath 2020) like animacy and definiteness affect construction choice through word order (eg. animate-first preference) or function assignment (eg. animate-subject preference) (Thuilier et al. 2021). Using sentence recall tasks, experimental studies both in English (Bock, Warren 1985, McDonald et al. 1993) and French (Thuilier et al. 2021) have tried to disentangle these preferences by comparing animacy effects on active/passive alternation (a function assignment alternation) and NP coordinations (a word order alternation). The authors report an animacy effect on active/passive alternation, but not on word order in coordinations, suggesting that animacy plays a role only for function assignment.

However, this conclusion is challenged by Feleki & Branigan (1999), who found an animate-first effect in Greek using the same sentence recall task with the SVO/OVS alternation. The authors thus claim that animacy only affects word order and that animate-subject preferences observed in English are just due to the tight connection between subject and first position in this language. But as Feleki and Branigan (1999) only took into account the SVO/OVS

alternation, word order and function assignment preferences have not yet been directly compared in Greek.

As for length, it is sometimes included among prominence features (Bresnan, Ford 2010), while McDonald et al. (1993) treat it separately, as it did not show any effect in sentence recall in English. Feleki and Branigan (1999) study did not include this factor in Greek.

## Methodology

We aim at disentangling word order and function assignment preferences in Greek by conducting a quantitative corpus study. We used newspaper texts from the morphologically annotated Corpus of Modern Greek (Arkhangelskiy, Kisilier 2018). We randomly extracted a sample of 600 sentences, containing 200 active SVO (1), 200 active OVS (2) and 200 passive constructions with a preverbal subject and an expressed by-phrase agent (3). Verbs were all transitive and passivizable, and the arguments were only NPs, because clitic object pronouns are preverbal and subject pronouns mostly drop.

- (1) Oi ispanoi, prosperoun 1.500 eisitiria  
the Spanish offer 1500 tickets  
'The Spanish people offer 1500 tickets' (Makedonia, 2007/09/26)
- (2) Ipotheseis ekane o Oli Pen  
hypotheses made the Oli Pen  
'Oli Pen made assumptions'
- (3a) I synchroni techni empneetai apo ti vizantini (Makedonia, 2010/10/20)  
the contemporary arts inspire.PASSIVE by the byzantine  
'Contemporary art is inspired by Byzantine art' (To Vima, 2012/01/26)
- (3b) Tessera paidia travmatistikan apo tin puravliki epithesi  
four children injure.PASSIVE by the missile attack  
'Four children were injured by the missile attack' (To Vima, 2011/05/10)

Using Arg1 for active subject and passive by-phrase, and Arg2 for active object and passive subject, we annotated argument animacy, definiteness and length (number of words). With these three constructions (SVO/OVS/Passive), we can tease apart word order from function assignment preferences: in SVO/OVS, functions are the same but order differs, in OVS/Passive, there is the same Arg2-Arg1 order but function mapping varies (object Arg2 in active vs. subject Arg2 in passive).

For statistical analysis, we fitted a Bayesian categorical model to construction as a function of animacy, definiteness and length, using the *brms* package in R (Bürkner 2017). We used OVS as reference level for the outcome, as it minimally differs from both SVO and Passive. Verb lemma was used as a random variable. Length was included using a *Length difference* variable:  $\log(\text{Arg1 Length}) - \log(\text{Arg2 Length})$  (Bresnan & Ford, 2010). We used weak

informative priors for regression coefficients (normal distribution with  $\mu=0$ ,  $\sigma=1$ ) and default brms priors for other parameters (Student's t-distribution with  $\nu=3$ ,  $\mu=0$ ,  $\sigma=2.5$ ). Four sampling chains ran 4000 iterations with a warm-up period of 1000 iterations.

## Results

Results from the Bayesian categorical model are reported in Figure 1. We examine 6 hypotheses: preferences for human-subject, human-first, definite-subject, definite-first, shorter-subject, shorter-first. In each case, we report mean parameter value ( $\beta_{\text{Passive}}$  for Passive/OVS contrast and  $\beta_{\text{SVO}}$  for SVO/OVS contrast), 95% credible interval and the probability  $P$  of  $\beta$  differing from 0. We find compelling evidence for human-subject preference for both Arg1 (1) ( $\beta_{\text{Passive}}=-1.00$ ,  $\text{CI}=[-1.48;-0.53]$ ,  $P(\beta_{\text{Passive}}<0)=100\%$ ) and Arg2 (3b) ( $\beta_{\text{Passive}}=1.84$ ,  $\text{CI}=[1.13;2.58]$ ,  $P(\beta_{\text{Passive}}>0)=100\%$ ). However, we do not find evidence for human-first preference for neither Arg1 (2) ( $\beta_{\text{SVO}}=-0.20$ ,  $\text{CI}=[-0.6; 0.21]$ ,  $P(\beta_{\text{SVO}}>0)=21\%$ ) nor Arg2 ( $\beta_{\text{SVO}}=0.56$ ,  $\text{CI}=[-0.14;1.27]$ ,  $P(\beta_{\text{SVO}}<0)=10\%$ ). Animacy thus has an effect on function assignment, not on word order. The same holds for definiteness: there is evidence for a definite-subject preference for both Arg1 (1) ( $\beta_{\text{Passive}}=-1.85$ ,  $\text{CI}=[-2.47;-1.24]$ ,  $P(\beta_{\text{Passive}}<0)=100\%$ ) and Arg2 (3a) ( $\beta_{\text{Passive}}=2.70$ ,  $\text{CI}=[2.14;3.28]$ ,  $P(\beta_{\text{Passive}}>0)=100\%$ ). But no evidence is provided for definite-first preference with Arg1 ( $\beta_{\text{SVO}}=-0.15$ ,  $\text{CI}=[-0.77;0.47]$ ,  $P(\beta_{\text{SVO}}>0)=35\%$ ) or Arg2 (2)(3b) ( $\beta_{\text{SVO}}=0.71$ ,  $\text{CI}=[0.31;1.11]$ ,  $P(\beta_{\text{SVO}}<0)=0\%$ ). Finally, we found a shorter-first preference ( $\beta_{\text{SVO}}=-1.32$ ,  $\text{CI}=[-1.57; -1.09]$ ,  $P(\beta_{\text{SVO}}<0)=100\%$ ) but no shorter-subject preference ( $\beta_{\text{Passive}}=-0.12$ ,  $\text{CI}=[-0.36;0.12]$ ,  $P(\beta_{\text{Passive}}>0)=20\%$ ).

## Conclusion

Statistical modelling on Greek corpus data shows that prominence features (animacy and definiteness) only affects function assignment, which can be viewed as argument coding efficiency (Haspelmath 2020). By contrast, length only affects word order, which can be explained by dependency length minimization (Temperley, Gildea 2018). This difference brings evidence to distinguish the status of prominence features and length. While prominence features represent referential/conceptual information, length is more tied to phrasal/formal properties (McDonald et al., 1993). We failed to replicate an animate-first preference as found by Feleki, Branigan (1999) in Greek, which may be due to task effects: ecological production may not involve the same cognitive processes as sentence recall.

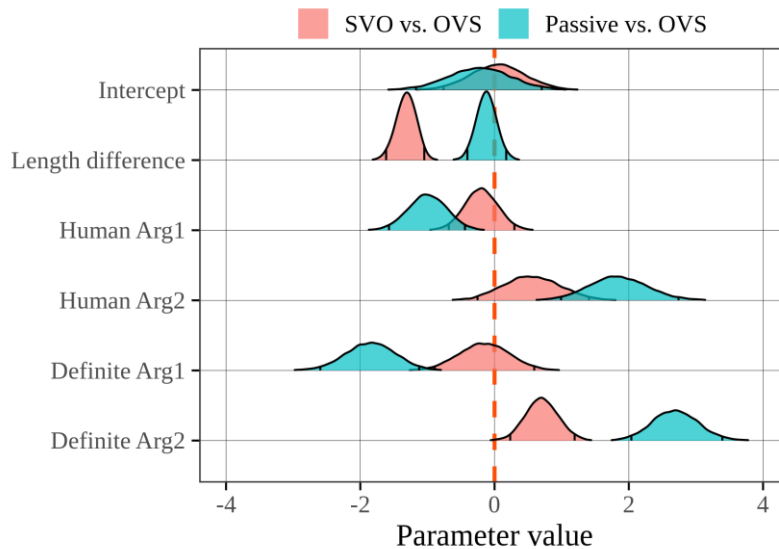


Figure 1. Fixed-effects posteriors in the categorical model. Vertical black lines represent 95% credible intervals.

## References

- Aissen, J. 1999. Markedness and Subject Choice in Optimality Theory. *Natural Language & Linguistic Theory*, 17(4), 673-711.
- Arkhangelskiy, T., Kisilier, M. 2018. Corpora of Modern Greek: Achievements and goals. *Indo-European Linguistics and Classical Philology*, 50-59.
- Bresnan, J., Ford, M. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language*, 168-213.
- Bock, J.K., Warren, R.K. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21(1), 47-67.
- Bürkner, P.-C. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, 80, 1-28.
- Feleki, E., Branigan, H. 1999. Conceptual accessibility and serial order in Greek speech production. *Proceedings of the 21st Annual Cogsci Conference*, 961-101.
- Haspelmath, M. 2020. Role-reference associations and the explanation of argument coding splits. *Linguistics*, 45.
- McDonald, J.L., Bock, K., Kelly, M.H. 1993. Word and World Order: Semantic, Phonological, and Metrical Determinants of Serial Position. *Cognitive Psychology*, 25(2), 188-230.
- Temperley, D., Gildea, D. 2018. Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Review of Linguistics*, 4, 67-80.
- Thuilier, J., Grant, M., Crabbé, B., Abeillé, A. 2021. Word order in French: The role of animacy. *Glossa: a journal of general linguistics*, 6(1), 55.



# Iconicity in fictional and natural languages: an experimental study

Varvara Davydova<sup>1</sup>, Elena Shamina<sup>2</sup>

<sup>1</sup>Department of Foreign Languages, St Petersburg Electrotechnical University, Russian Federation

<sup>2</sup>Department of Phonetics, St Petersburg state university, Russian Federation

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Language iconicity has lately become an important issue in cognitive studies. The phonetic form of iconic linguistic signs is supposed to be motivated and aligned with some of the physical properties of the signified, thus making it more understandable and recognisable for the language user. The hypothesis laying the foundation of the present research is that iconic words of fictional languages are constructed in accordance with the same or similar principles as iconic words in natural languages. The paper discusses the results of a series of psycholinguistic experiments that are part of a broader study of iconicity in fictional languages and whose aim is to test the hypothesis.

Keywords: psycholinguistic experiment, iconicity, phonosemantics, fictional and natural language

## Introduction: language iconicity

Iconic language signs (unlike indices or symbols in C. S. Peirce's classification (Peirce 1994)) are those that represent by their sound form certain properties of the signified directly, by means of synesthetic associations (Ramachandran, Hubbard 2001). They fall into two categories: sound imitative and sound symbolic, and have been shown to have cognitive potential that can ensure instantaneous understanding of nonsense (Abelin 2006) or foreign (Шамина 2018) words. It is only logical to suppose that literary fictional languages abound in phonosemantic vocabulary.

## Background: fictional languages

Fictional (invented) languages are those constructed for use in literary writings describing thought-up worlds and their non-human inhabitants. Often, the sound of their speech is the major means of their characterization, and the readers should be given an opportunity to understand what is going on without constantly referring to the dictionary. Iconic features in such languages have been pointed out (Fimi, Higgins 2016; Peterson 2015), and an extensive study of the phenomenon have been conducted for 4 of them (Davydova 2016). That was performed with the help of the phonosemantic analysis procedure (Воронин 1982), and a special emphasis was placed on the phonemic inventory

and the phonotactics of the lexis not being copied from English, the native tongue of the inventors: cf. Klingon *tlhup* /tʰupʰ/ ‘whisper’, Na’vi *bi’i* /’hi.ʔi/ ‘small’.

## Perception of iconic lexis of fictional and natural languages

### Material and methodology

The material for the perceptual experiment with fictional lexis (ДАВЫДОВА, ШАМИНА 2022) is extracted from two invented languages found in popular fantasy novels: Klingon by Mark Okrand and Na’vi by Paul Frommer who supplied phonetic transcription and detailed pronunciation instructions. The 20 words constitute designations of acoustic events of different types, as well as size, shape, luminosity, etc., i.e., their semantics is typical for phonosemantic items in natural languages. The methodology used is a variation of the associative experiment with 3 options offered to the respondents to choose from when trying to determine the meaning of the target stimulus. These were presented to the participants in their audio form. The subjects were native English and Russian speakers of both sexes and various ages.

For comparison, the data from a different series of phonosemantic experiments (Shamina 2019) are used. The methodology and the respondents’ characteristics are very similar, but the word-stimuli come from the author’s corpus of English and Russian slang and the onomatopoeic vocabulary of the languages mentioned. About 5000 responses were elicited and analysed for each of the experiment series and the data statistically evaluated. Any observed discrepancies between reactions of respondents in different age and gender groups are statistically insignificant.

### Results and discussion

The analysis of the data obtained points to the respondents’ reactions to the word-stimuli, whether they are fictional items or existing foreign words, being based on their sound iconic properties, and to the speakers of both non-closely related languages demonstrating similar perception strategies. The overall results show that in the case of fictional languages, for 17 out of 20 lexical items both groups of subjects are inclined to choose the meaning originally conceived by the inventor. The average correct interpretation of the language units under consideration, both fictional and natural, varies around 70 - 80 %, peaking to more than 90 % in some instances, e.g. fictional *slele* /’slɛ.lɛ/ ‘swim’ for the Russian speakers and Russian *пентюх* /’pɛn’tɯx/ ‘lout’ for the English speakers.

In the experiment with the fictional language material, the subjects’ “correct” choices embrace both sound imitations and sound-symbolic names referring to non-acoustic properties of the denotatum. Speaking of the latter, designations of small size are known to be associated with front high vowels and fore-lingual

consonants, and Na'vi *lini* /'li.ni/ 'young of an animal' follows suit, being recognized as such by 80 % of the Russian and 83 % of the English subjects. Iconic words denoting fast sharp movement usually have stops in their form, thus Na'vi *takuk* /'ta.kuk/ 'rush forward' is perceived as expected by about 80 % of the Russians and about 72 % of the English. The Klingon word *gho* /ʎo/ comprising a rounded vowel triggers its perception as 'circle' and not 'sharp-angled figure' or 'dotted line' by about 70 % of all the respondents. The same percentage of both groups of the subjects are sure that Na'vi *hufwe* /hu.'fwe/ means 'wind' and not 'stone' because denotations of air movement and breathing regularly contain fricatives and rounded vowels.

In respect to onomatopes, all the subjects identified the Na'vi word *nguwai* /'ŋu.wai/ as meaning 'howl' in more than 60 % of cases and the Na'vi word *rurur* /ru.'rur/ as meaning 'water flowing among the rocks' in about 80 % of cases. This happens due the well-established tendency (Voronin 2005) for tone continuants (words imitating prolonged and fluid tones) to have long vowels and sonorants in their phonetic structure while frequentatives (words imitating vibrations, disruptive noises) tend to include rhotic consonants. Instants (imitations of pulses) are better recognised than other types of onomatopes, e.g., Na'vi *pxek* /p'ɛk/ 'sound of strike' and Klingon *baS* /baʂ/ 'sound of a sheet of metal' are correctly perceived in about 80 % and 90 % of cases, respectively. It is worthy of mention that the same observation was made when analysing the data of the experiment with the onomatopoeic lexis of the natural languages.

## Conclusion

The results of the experimental series discussed confirm the hypothesis about the non-random selection by inventors of fictional languages of sound forms for the transmission of various meanings, which, apparently, is determined by the universal iconic principles of linguistic systems. The findings can be, at least partly, accounted for by the supposition that literary fictional languages do not function as a means of communication between (non-human) literary characters, but serve the purpose of transferring the authors' ideas to the readers who are speakers of natural languages.

## Acknowledgements

The authors are deeply grateful to all the participants in their experiments.

## References

- Abelin Å. 2006. Experiments in investigating sound symbolism and onomatopoeia. In Proceedings of the 1st International Conference of Experimental Linguistics. Athens, Greece. 61-64.

- Davydova, V. A. 2016. Sound Symbolism in Invented Languages. In *Anglistics of the XXI century*, vol.2. St. Petersburg: St. Petersburg state university Press. 32–39.
- Fimi, D., Higgins, A. (eds.) 2016. Tolkien, J. R. R. *A Secret Vice: Tolkien on Language Invention*. London: Harper Collins.
- Frommer, P. M. 2022. Na’vi – English Dictionary v. 14.4. Compiled by Mark Miller. URL: <https://files.learnnavi.org/dicts/NaviDictionary.pdf> (date of access: 16.08.2022).
- Okrand, M. 1992. *The Klingon Dictionary*. New York: Pocket Books.
- Peirce, C. S. 1994. In James Hoopes (ed.). *Peirce on Signs: Writings on Semiotic*. Chapel Hill, NC, University of North Carolina Press.
- Peterson, D. 2015. *The Art of Language Invention: From Horse-Lords to Dark-Elves, the Worlds behind World-Building*. New York: Penguin Books.
- Ramachandran, V. S., Hubbard, E. M. 2001. Synaesthesia - a window into perception, thought and language. In *Journal of Consciousness Studies*, Vol. 8. 3–34.
- Shamina, E. A. 2019. The Natural Environment for the Experimental Study of Phonosemantics. In *Proceedings of the 10th International Conference of Experimental Linguistics*. Lisbon, Portugal. 193 – 196.
- Voronin, S.V. 2005. The Universal Classification of Onomatopes 25 Years On. In Voronin, S.V. *Iconicity. Glottogenesis. Semiosis*. St.Petersburg: St.Petersburg State University Press. 30 – 37.
- Воронин С.В. 1982. *Основы фоносемантики*. Ленинград, Изда-во ЛГУ.
- Давыдова В.А., Шамина Е.А. 2022. Экспериментальное исследование фоносемантического потенциала вымышленной лексики. In *Материалы Международной филологической конференции памяти проф. Л. А. Вербицкой*. Санкт-Петербург: СПбГУ. 431 – 447.
- Шамина Е.А. 2018. Когнитивная валентность звукоподражательной лексики (экспериментальное исследование). In *Когнитивные исследования языка, XXXIY. Cognitio и Communicatio в современном глобальном мире*. Москва – Тамбов: Ин-т языкознания РАН, 338-342.

# The adaption to online synchronous teaching and voice fatigue: acoustic and clinical data

Karina Evgrafova<sup>1</sup>, Natalia Sokolova<sup>2</sup>, Nikolay Shvaleyev<sup>3</sup>

<sup>1</sup>Phonetica Department, St Petersburg University, Russian Federation

<sup>2</sup>Department of English Philology and Cultural Linguistics, St Petersburg University, Russian Federation

<sup>3</sup>St Petersburg State Pediatric Medical University, Russian Federation

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Due to the COVID-19 pandemic, all educators has faced an unprecedented challenge to their professional skills and wellbeing. [1, 6] In 2020 teachers around the world had to shift to online teaching. In 2021 we presented research on the impact of this new professional reality on the vocal load of Saint Petersburg university professors. The online synchronous teaching caused the significant increase in the focal fatigue in university professors in comparison with the pre-pandemic studies. [6] We continued our study during the post-pandemic semester (winter-spring 2022). The goal of this study was to find out whether adaptation mechanisms during the pandemic period were developed to adjust to the different types of teaching mode. The acoustic and clinical data are presented.

Keywords: vocal fatigue, teacher's voice, voice load. online synchronous teaching, COVID-19 pandemic

## Introduction

Vocal fatigue in voice professionals has been studied intensely for decades, especially regarding symptoms and risk factors. It is particularly self-reported by teachers as a sense of increased vocal effort and a sensation of laryngeal and pharyngeal constriction. The clinical analysis performed through laryngoscopy can detect symptoms associated with vocal disorders. Besides, vocal fatigue is also shown in tonal range, dynamic range, vocal quality, intensity and fundamental frequency changes. The acoustical aspect of the phenomenon allows its objective evaluating in terms of degree and dynamics.

We performed the acoustic, auditory and clinical analysis of vocal fatigue symptoms in the professors of Saint Petersburg state university (pronunciation teachers and lecturers) in a number of previous studies in the COVID-19 pre-pandemic years. [3-5] Due to the COVID-19 pandemic, there was a dramatic change in the work mode of all voice professionals. In 2020 university professors around the world had to shift to online teaching. In 2021 we presented the research on the impact of this new professional reality on the vocal load of Saint Petersburg university professors. The online synchronous

teaching caused the significant increase in the focal fatigue in university professors in comparison with the pre-pandemic studies.

We continued our study during the post-pandemic semester (winter-spring 2022). During that period our participants either returned to classroom teaching or switched to hybrid mode of teaching (consisting of a mixture of distant and classroom activities). The goal of this study was to find out whether adaptation mechanisms during the pandemic period were developed to adjust to the different types of teaching mode.

## Methodology

We followed the protocol used in our *pre-pandemic* and *pandemic* vocal fatigue studies. All the participants were the professors of Saint Petersburg State University with average work experience of 5 years. The participants were involved in different types of teaching activities: teachers delivering lectures on linguistics; English teachers running practical classes; pronunciation coaches.

The minimum workload a day was 3 hours while the maximum was 6 hours. No one had pathological voice problems. 10 female teachers recorded themselves *before* and *after* classroom/online synchronous teaching using their mobile phones. The participants read a four minute phonetically representative text.

The participants also filled in the self-reporting questionnaire which they had faced in the 2021 study. They graded their physical state, mood and a level of activity at the moment of the recording. We used the **WAM** questionnaire to evaluate psychoemotional state of the teachers before and after their work. **WAM** (wellbeing, activity, mood) is used to assess the mental state of patients and healthy people, their psychoemotional response to loading. [2]

Besides, the participants had the laryngoscopy of vocal cords done regularly during the period of 2021-2022.

## Results

Thus we obtained acoustic data (*objective evaluation*), self-reports (*subjective evaluation*) and laryngoscopy results (*clinical evaluation*) which can show the impact of different types of teaching mode on vocal fatigue.

### Acoustic data

We calculated a number of acoustic parameters which had been significant for detecting voice fatigue in the previous studies (mean F0, vowel duration and laryngealization) in non-fatigued (NF) and fatigued (F) speech samples. The values of these parameters in pre-pandemic, pandemic and post-pandemic recordings are presented in Table 1 below.

Table 1. Mean F0, vowel duration increase and the percentage of laryngealized segments in non-fatigued/fatigued speech (pre-pandemic, pandemic and post-pandemic material).

	Pre-pandemic	Pandemic	Post-pandemic
	Mean F0, Hz		
NF	185	178	182
F	188	210	186
	Vowel Duration Increase, ms		
F	4.3	7.2	5.2
	Laryngealization, %		
NF	1.5	1.8	1.4
F	1.2	2.3	1.9

F0 tends to be higher in the fatigued speech across all types of the recordings. However, the post-pandemic values are closer to the pre-pandemic ones. The vowel duration increase in the fatigued speech is still significant, although it has decreased in the post-pandemic period.

Laryngealization which is marked by significant decrease in pitch value and pitch breaks is associated with a creaky voice quality. The symptom was frequently reported by the teachers during the self-assessment of voice quality.

The mean duration of laryngealized speech segments is the longest during the pandemic and has also reduced in the post-pandemic period.

### Clinical data

During the extended period of online teaching the clinical picture showed hypotonic dysphonia that was potentially caused by overuse or excess voice use. One of the patients excess voice use during the pandemic period resulted in the pre-nodule condition of vocal cords (pic. 1). The relief in voice fatigue came with both developing adaptation mechanisms and partly switching to in-class teaching in post-pandemic. The picture 2 illustrates certain ease in the condition of vocal tract.



Pic. 1 Pre-nodule condition of vocal cords. Pic. 2 Hypotonic dysphonia.

## Discussion and conclusion

Although the vocal quality improved as well as the clinical picture, but neither the post-pandemic voice nor the laryngoscopic data yet resemble the pre-pandemic condition. The return to the regular working environment (with the absence of the necessity of the microphone use and visible audience follow-up and reaction) has had a positive effect on the vocal functions and reduced possible pathological changes in the larynx.

The analysis of the acoustic and clinical data confirmed the effectiveness of the adaption mechanisms in terms of vocal fatigue decrease. The self-reports showed that during the pandemic period teachers had to develop new strategies to avoid voice overstraining such as slowing the pace, taking frequent pauses, putting an emphasis on diction and consonants rather on increasing the loudness.

## References

- Besser, A., Lotem S., Zeigler-Hill V. Psychological Stress and Vocal Symptoms Among University Professors in Israel: Implications of the Shift to Online Synchronous Teaching During the COVID-19 Pandemic. *Clinics (Sao Paulo)*. 2021; 76: e2641. Published online 2021 Mar 19. doi: 10.6061/clinics/2021/e2641
- Doskin, V. A. Test differentsirovannoy samootsenki funktsional'nogo sostoyania (The Test of Differentiated Self-Assessment of Functional Status)/ Doskin, V.A. Lavrenteva, N.A. Miroshnikov, M.P. Sharay V.B. // *Voprosy psikhologii* 1973 №6, 141-145.
- Evgrafova, Evdokimova, K, V. Acoustic analysis of vocal fatigue in professional voice users. *MAVEBA* 2011, 153-156.
- Evgrafova, K., Evdokimova, V., Skrelin, P., Chukaeva, T. Vocal fatigue in voice professionals: collecting data and acoustic analysis. DOI:10.36505/ExLing-2016/07/0011/000270
- Evdokimova, V.V., Evgrafova K.V., Skrelin P.A, Chukaeva, T.V. The database of normal and pathological singers' voices: an approach to collecting data. *MAVEBA* 2017, 23-24.
- Evgrafova, K. V., Sokolova, N. S., Shvaley N. V. The Impact of online teaching during the COVID-19 on vocal fatigue in university professors: self-reports and acoustic evaluation. *Maveba* 2021, 167-170. DOI: 10.36253/978-88-5518-449-6.



# Shared syntactic representations in non-native languages

Chiara Facipieri, Michaela Vann, Giulia Bencini

Department of Linguistics and Comparative Cultural Studies, Ca' Foscari University of Venice, Italy

<https://doi.org/10.36505/ExLing-2022/>

## Abstract

How do speakers represent and process syntactic information in their second and third language? We investigated this question with cross-linguistic syntactic priming of passives from Spanish (L3) to English (L2). We found that cross-linguistic structural priming occurs between two non-native languages in intermediate to advanced speakers of English as an L2 and Spanish as an L3. We take these results to show that multilingual speakers have one integrated shared abstract representation of passives. The findings are compatible with the extension of the bilingual shared-syntax model to multilingual syntactic processing.

Keywords: multilingualism, L3 processing, structural priming, proficiency

## Introduction

According to the shared-syntax model of bilingual language production (Hartsuiker et al. 2004, Hartsuiker, Bernolet 2015) syntactic representations are shared between an L1 and later acquired languages as much as possible. Most of the evidence for this model comes from studies that examine processing in an L1 and a later acquired L2. The question of how speakers represent, and process subsequently acquired languages (L3 and beyond) is still unanswered. To our knowledge, Hartsuiker and colleagues (2016) is the only study to have investigated priming between two non-native languages. Their findings support the shared-syntax model: if the representation of a structure is shared between languages, then all languages can activate it with equal strength.

Our study tests the shared syntax model by examining how multilingual speakers of Italian as an L1 represent and process passive sentences in later acquired languages, specifically English (L2) and Spanish (L3). Our research questions were:

1. Do late multilinguals (acquiring an L2 and L3) share syntactic information between their L2 and L3? To answer this question, we examined whether cross-linguistic syntactic priming of passive sentences occurs between Spanish L3 and English L2.
2. Is the strength of priming modulated by language proficiency (Bernolet et al. 2013, Hartsuiker, Bernolet 2017)?

3. Is cross-linguistic structural priming between L2 and L3 influenced by conceptual information?

If the shared syntax model also applies to additional languages, the prediction is that, provided sufficient proficiency, priming should be possible between an L3 and an L2 (and vice-versa). In this study, we looked at priming from the L3 to the L2.

## Methods

### Participants

26 Italian native speakers (24 Female, 2 Male,  $M_{age} = 28$ ) took part in a cross-linguistic priming experiment (data collection is ongoing). Participants acquired English as an L2 and Spanish as an L3. Their linguistic background was assessed using a Language Profile Questionnaire. In the survey, participants were asked to rate their language level according to the criteria of the Common European Framework (CEFR) in all four language modalities (writing, listening, speaking, and reading) in their L2 and L3 using a Likert-scale from 1 to 6.

### Design and procedure

The study investigated primed production of passive sentence structures from Spanish to English using a within-subjects 2 (structure) x 2 (animacy) factorial design. Example materials are shown in Table 1. Participants read a prime sentence in Spanish and typed in a written description to a target picture in English. The experiment was administered remotely, using Psychopy (Peirce et al., 2019) on the Pavlovia.org platform.

Table 1. Experimental conditions. The animacy condition of the agent (inanimate) was kept constant across conditions.

	Prime structure - Active	Prime structure - Passive
Animate patient	El sol cegó al hombre. The sun blinded the man.	El hombre fue cegado por el sol. The man was blinded by the sun.
Inanimate patient	El huracán arruinó la cosecha. The hurricane ruined the crop.	La cosecha fue arruinada por el huracán. The crop was ruined by the hurricane.

### Analysis and results

All descriptions to target pictures were scored for syntactic structure. To be considered a passive, descriptions had to contain the patient as syntactic subject, followed by the auxiliary *be* in any tense, followed by the agent introduced by the preposition *by*. Passives without an explicit agent (e.g., truncated passives) were scored *other*, and excluded from statistical analyses.

The data were analysed with generalised linear models (McCullagh & Nelder, 1989) predicting the log odds (logit) of a passive sentence. The best-fit model is summarised in Table 2.

There was a significant effect of prime structure, indicating that structural priming occurs from an L3 (Spanish) to an L2 (English). There was also a main effect of animacy: participants on average produced 29% of passive responses after Animate patient primes as opposed to 17% after Inanimate patient primes. Animacy effects were independent of prime structure, suggesting that semantic features may drive production in multilingual speakers more than syntactic structure.

Proficiency in the L2 and L3 alone was not significant (this is likely to be due to lack of power). In order to investigate the relationship between priming and self-reported language proficiency, we computed a condensed measure of additional language dominance (i.e., which additionally acquired non-native language is more dominant?). This measure was defined as the difference between the average proficiency scores - negative language dominance scores equal Spanish dominance, whereas positive scores reflect English dominance. The model suggests that Language dominance did not have a significant independent effect on the production of passive responses. However, there was a significant positive Prime Structure x Language dominance score interaction, indicating that as the language dominance score increases, the production of passive responses after passive primes increases as well.

Table 2. Model results.

Predictors	Coefficient	SE	z value	p
(Intercept)	-1.30	0.09	-14.01	<.001
Prime Structure	0.20	0.09	2.40	.0339
Animacy	0.36	0.09	3.82	.0001
Language dominance	-0.09	0.1	-1.48	>.1
Prime Structure x Animacy	-0.07	0.09	-1.15	>.1
Prime Structure x Language dominance	0.24	0.1	3.54	.0093
Animacy x Language dominance	0.03	0.1	0.80	>.1
Prime Structure x Animacy x Language dominance	-0.10	0.1	-1.70	>.1

## Discussion

Our study confirms that priming can occur between two non-native languages (Hartsuiker et al. 2016), in line with the extension of the bilingual shared-syntax model (Hartsuiker et al. 2004) to multilingual syntactic processing: provided that speakers have a high enough proficiency in L2 and L3, they conveniently merge abstract representations of similar syntactic structure. Our results add to these findings that if two non-native languages are involved, in the case of L3 to L2 priming, target language proficiency may need to be higher than prime

language proficiency to observe significant evidence of shared-syntactic representations. Having only tested intermediate to advanced speakers of English and Spanish may have prevented us from seeing a clear modulating effect of only target language proficiency like the one found by Bernolet et al., (2013) and Hartsuiker & Bernolet (2015).

Furthermore, our results confirm our initial prediction that participants would produce more passive responses after animate patient primes regardless of prime structure. This suggests that the patient's inherent accessibility, more than its derived accessibility, allowed for the animate entity to be more easily retrievable and to be encoded in a prominent sentential position, namely the subject of a passive sentence. These findings lend support to the hypothesis that syntactic priming and animacy influence the choice of syntactic structure independently of each other (Pickering & Ferreira, 2008).

## References

- Bernolet, S., Hartsuiker, R.J., Pickering, M.J. 2013. From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition*, 127(3), 287–306.
- Hartsuiker, R.J., Bernolet, S. 2015. The development of shared syntax in second language learning. *Bilingualism: Language and Cognition*, 20(2), 219–234.
- Hartsuiker, R.J., Beerts, S., Loncke, M., Desmet, T., Bernolet, S. 2016. Cross-linguistic structural priming in multilinguals: Further evidence for shared syntax. *Journal of Memory and Language*, 90, 14–30.
- Hartsuiker, R.J., Pickering, M.J., Veltkamp, E. 2004. Is syntax separate or shared between languages? *Psychological Science*, 15(6), 409–414.
- McCullagh, P., Nelder, J.A. 1989. *Generalized Linear Models*. 2nd Edition, Chapman and Hall, London.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Lindeløv, J.K. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203.
- Pickering, M.J., Ferreira, V.S. 2008. Structural priming: A critical review. *Psychological Bulletin*, 134(3), 427–459.

# Sentence repetition as a function of episodic buffer: a pilot study in Croatian

Mateja Gabaj<sup>1</sup>, Jelena Kuvač Kraljević<sup>2</sup>

<sup>1</sup>Postgraduate doctoral study “Speech, Language and Hearing Disorders”, Faculty of Education and Rehabilitation Sciences, University of Zagreb, Croatia

<sup>2</sup>Department of Speech-Language Pathology, Faculty of Education and Rehabilitation Sciences, University of Zagreb, Croatia

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

In sentence repetition tasks, the function of the episodic buffer of working memory is to adjust semantic and syntactic information stored in long-term memory. To explain how working memory uses linguistic knowledge about words and constraints on their order in sentences, *the constrained sentence span task* was adapted and developed. The aim of the study was to determine how many words in sentences children could repeat under both conditions with and without articulatory suppression. Sixteen ten-year-old children participated in the study ( $M = 10;04$ ). The results showed that success in repeating sentences under both conditions increased up to sentences with six words. This experimental task accounts for the function of the episodic buffer when controlling various linguistic aspects.

Keywords: sentence repetition, long-term linguistic knowledge, episodic buffer, working memory

## Introduction

In Baddeley's model of working memory, the episodic buffer has the task of communicating with long-term memory while integrating and storing information from the other two components of working memory – the phonological loop and/or the visuospatial sketchpad, depending on the modality of the information (Baddeley, 2000).

In verbal tasks, the episodic buffer allows syntactic and semantic information from long-term memory to interact with that from working memory. In immediate sentence recall tasks, it is thought to act as a system that adapts to the contribution of long-term linguistic knowledge, explaining the ability to repeat a larger number of related units of meaning in a particular order. Other theories of immediate recall emphasise that linguistic knowledge is interwoven with memory for immediate sentence repetition (Acheson & McDonald, 2009).

Baddeley et al. (2009) have developed a sentence repetition task that reflects the functioning of, or measures the capacity of, the episodic buffer of working memory. This task can be used to examine the span of immediate sentence memory when demands on working memory are increased while controlling for linguistic knowledge of words and constraints on their order in sentences. The

aim of the present study is to test this adapted task with Croatian-speaking children.

## Methods

### Participants

Sixteen typically developing school-aged children participated in this pilot study (10;00 - 10;10 years old;  $M = 10;04$  years,  $SD = 3.26$  months). Children's non-verbal cognitive abilities were assessed using the Raven's Progressive Matrices (Raven et al., 2000). All children scored normal on the nonverbal IQ, confirming that they had no diminished intellectual abilities (SR mean = 104; SR range = 90 - 128.75;  $SD = 12.25$ ) and had no language difficulties.

### Experimental task

For this study, a specific experimental sentence repetition task – *the constrained sentence span task* – was designed to investigate sentence memory span and to determine how working memory uses linguistic knowledge about words and constraints on their order in sentences stored in long-term memory (Baddeley et al., 2009). In this task, sentences are constructed from a closed set of words (nouns, verbs and adjectives). Examples of 6-word sentences formed from a closed set of words are:

- 1) Susjed Matej prodaje dobar stari auto. [Neighbour Matthew sells (a) good old car].
- 2) Novi susjed otvara stari zeleni auto. [(The) new neighbour opens (the) old green car].

This task is similar to a standard memory span procedure in which a smaller number of words are used repeatedly in different places in sentences, taking into account the linguistic regularities and constraints of the Croatian language. Selecting a closed set of words and using a limited range of syntactic structures increases the demands on working memory during sentence repetition and forces subjects to focus on their current memory, minimising the variability that may arise from their individual language knowledge or the form and content of the sentences. Furthermore, this allows for proactive interference of similar linguistic material, reducing the contribution of gist-based long-term episodic memory. The proactive interference paradigm ensures that participants focus on temporary binding in working memory.

### Design and procedure

A 6x2 repeated measures design was used in which sentence length was manipulated by gradually increasing noun phrases (so that sentences were 3-8

words long) and secondary tasks (performance without or with articulatory suppression to disrupt the phonological loop). Ten different sentences were designed for each sentence sequence, and each experimental condition was counterbalanced using a Latin square. As linguistic knowledge has previously been shown to influence immediate sentence repetition, word order, semantic plausibility, word length and word frequency were controlled. Therefore, all words were 2-3 syllables long, of medium to high frequency based on data from the Children's Frequency Dictionary of the Croatian Language (Kuvač Kraljević et al., 2021) and the Croatian Lexical Database (Kuvač Kraljević & Olujić, 2018), constructed in sentences in the canonical SVO order, and semantically plausible, as subsequently confirmed by children's rating on a Likert scale of 1-5 ( $M = 4.45$ ). The experiment was conducted at school, in a quiet room. After listening to the audio recording of the sentences through headphones, the participants had to repeat each sentence as accurately as possible. Under conditions of articulatory suppression, they continuously repeated 1-2-3-4 while listening to the sentences.

## Results and discussion

Since this study focused on the number of correct words that could be repeated in a sentence, each word that the children could recall was quantified. The repetition was correct if (1) the words were repeated in the correct order, i.e. there was no change of position between two adjacent words, (2) all inflexions were preserved and the number, gender and case of the content words matched the target word.

Inter-rater reliability was assessed using the intraclass correlation coefficient (ICC). Thirty percent of the samples (results from five children) were randomly selected and evaluated by the first author and an independent rater. Overall, the results showed excellent agreement between the two raters ( $ICC = .98$ ).

The mean proportion of words correctly repeated by the children for each sentence sequence and in relation to performance in the conditions without and with articulatory suppression (AS) is shown in Table 1.

As expected, the results showed significant effects of sentence length,  $F(1.76, 52.65) = 132.70$ ,  $p < .01$ ,  $\eta_p^2 = .82$ , but no effects of secondary tasks,  $F(1, 30) = 1.78$ ,  $p = .192$ ,  $\eta_p^2 = .06$ , and no interaction between these two effects,  $F(1.76, 52.65) = 1.24$ ,  $p = .295$ ,  $\eta_p^2 = .04$ . Sentence recall performance increased up to sentences of six words, after which it began to decline. For example, the mean number of words in the retrieved 6-word sentences was 5.69 ( $SD = .34$ ) and in the articulatory suppression condition was 5.54 ( $SD = .46$ ) (Table 1).

	Secondary tasks	
	Without AS	With AS
3-word sentences	2,99 (.05)	2,93 (.09)

4-word sentences	3,98 (.06)	3,96 (.09)
5-word sentences	4,78 (.25)	4,78 (.25)
6-word sentences	5,69 (.34)	5,54 (.46)
7-word sentences	5,63 (.88)	5,24 (.99)
8-word sentences	5,45 (.97)	4,99 (1.04)

Table 1. Descriptive statistics data – mean proportion of words for each sentence sequence (with standard deviation) in relation to the two conditions

Articulatory suppression did not interfere with the binding of words to sentences during retrieval. These results support the fact that articulatory suppression, which limits the involvement of the phonological loop that supports repetition, is not cognitively demanding enough to significantly impair sentence memorisation and repetition. This study contributes to the assumptions that the constrained sentence span task accounts for the function of the episodic buffer when controlling for different linguistic aspects from the word to the sentence level. Furthermore, the type of errors the children made could provide additional information about underlying language processing problems.

## References

- Acheson, D.J., MacDonald, M.C. 2009. Verbal working memory and language production: common approaches to the serial ordering of verbal information. *Psychological Bulletin*, 135, 50–68.
- Baddeley, A.D. 2000. The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423.
- Baddeley, A.D., Hitch, G.J., Allen, R.J. 2009. Working memory and binding in sentence recall. *Journal of Memory and Language*, 61(3), 438–456.
- Kuvač Kraljević, J., Olujić, M. 2018. Imageability and subjective frequency of the 500 rated nouns in the Croatian Lexical Database. *Suvremena lingvistika*, 44(85), 73–90.
- Kuvač Kraljević, J., Hržica, G., Štefanec, V. 2021. *Čestotni rječnik hrvatskoga dječjeg jezika – natuknice*. Zagreb: Naklada Slap.
- Raven, J., Raven, J.C., Court, J.H. 2000. *Manual for Raven's progressive matrices and vocabulary scales, Section 3: The standard progressive matrices*. Oxford. UK: Oxford.



# Exploring the prosody of affective speech

Anna M. Giovannini, Ailbhe Ní Chasaide, Christer Gobl

School of Linguistic, Speech and Communication Sciences, Trinity College Dublin,  
Ireland

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

This paper introduces a research project on voice quality and affect expression. It explores affective prosody by investigating the relationship between voice source parameter changes and perceived affect. Firstly, it aims to examine the relative contribution of voice source shifts occurring globally across an utterance and shifts that are aligned to the prosodic structure of the utterance. Secondly, it aims to formulate a simple model for affect expression that could, in principle, be applied to text-to-speech synthesis systems for Irish (Gaelic) dialects. The analytic methods to be used include voice source and intonation analysis of utterances produced to portray a range of emotions, and perception experiments with stimuli varying in terms of global vs. local, structured source manipulations.

Keywords: emotion, synthesis, prosody, voice source, affect

## Introduction

Current linguistic prosodic research largely avoids analysis of affective prosody. In fact, most research in the area of affective speech has been conducted by psychologists such as Scherer and colleagues (Scherer, 2003; Juslin and Scherer, 2005). However, although there is an emerging consensus that voice quality is central to the communication of affect, most analyses have tended to focus on global shifts in fundamental frequency ( $f_0$ ), intensity and tempo, while the crucial parameters of voice quality (tone of voice) are largely absent (but see, for example, Gobl and Ní Chasaide, 2003). Furthermore, beyond global average values, there is little account of whether and to what extent the changes to the voice source are in fact global shifts affecting entire utterances, or within-utterance shifts that take account of their known prosodic structure.

Building on earlier research on voice quality and affect expression (e.g., Murphy, Yanushevskaya, Ní Chasaide and Gobl, 2022), the present project sets out to explore affective prosody through analysis, and ultimately through synthesis, to investigate the relationship between voice source parameters (including  $f_0$ ) and perceived affect. A first objective is to examine to what extent affective changes may be cued by global, utterance-wide shifts in the source parameters and/or local shifts that are aligned to the prosodic structure of the utterance.

The second goal of this research is to build a simple model for affect expression that can, in principle, be exploited in speech synthesis. This goal is

motivated by practical considerations, as it would be desirable to be able to implement basic affective shifts in the synthetic speech output of an Irish (Gaelic) text-to-speech (TTS) system ([www.abair.ie](http://www.abair.ie)) which is increasingly being used in applications for education and for users with disabilities. We aspire to enable a synthetic utterance to be produced with some basic affective modulation, e.g., to render the narration of a story more engaging to a young user, or to allow a disabled user to modify the voice of speech-based systems they use to communicate.

The research question for this project would therefore be:

### **Research questions**

What are the roles of global (utterance-wide) and local (within-utterance, prosodically structured) changes in voice parameters (i.e., involving voice quality and fundamental frequency) in signalling affective states?

### **Hypothesis**

The signaling of affect is ultimately more a matter of prosodically structured shifts in the voice source parameters than global shifts through the entirety of an utterance.

### **Methodology**

The project involves the analysis of newly elicited production data as well as drawing on past research in the area as a basis for the construction of synthetic stimuli which incorporate either global or local voice source changes, as well as stimuli where both types of manipulations are incorporated.

### **The analytic study**

Analysis is currently ongoing on recordings carried out in the semi-anechoic chamber of the Phonetics and Speech Laboratory at Trinity College Dublin. The subject is a young male speaker of Kerry Irish and a professional actor. He was chosen based on the fact that the current Kerry Irish male synthetic voice (available at [www.abair.ie](http://www.abair.ie)) is based on his voice, and this TTS system is envisaged as a potential testbed for the emerging model of affect manipulation. Declarative, semantically neutral Irish sentences with long open vowels were created, containing two or three (potentially) accented syllables. The subject was then asked to read the sentences so as to portray different affects, including *neutral*, *angry*, *sad*, *happy/excited*, *interested*, *relaxed/contented* and *bored*. Multiple repetitions were recorded, and the subject was advised to keep repetitions consistent in terms of intonation patterns.

From the recordings, a few exemplars were chosen on the basis of an informal listening session conducted at the Phonetics and Speech Laboratory involving the authors and colleagues. The selection of the two final exemplars was based on the criteria that (1) the utterances were perceived as good

portrayals of a desired affect, and (2) the intonation pattern of the recording matched the expected intonation pattern which is typically found in the output of the TTS system, i.e., sentences had the typical falling contour of this dialect, and did not involve shifts in nuclear placement.

### Analysis methods

The chosen exemplars were then manually inverse-filtered and the resulting estimate of the voice source signal was modelled using the Liljencrants-Fant (LF) model (Fant, Liljencrants and Lin, 1985). For details on the software system and the techniques used, see Ní Chasaide, Gobl and Monahan (1992) and Gobl and Ní Chasaide (2010). The LF model is a mathematical glottal flow model that allows extraction of voice source parameters. The main parameters of interest to this research are  $f_0$ ; EE (excitation strength) which is closely related to the overall intensity of the signal; and RD, a measure relating to the perceived tension in the voice, where RD is typically high for lax voice and low for tense/harsh voice (see Laver, 1980, for an in-depth description of voice quality). Samples were then segmented and annotated using the Praat software (Boersma and Weenink, 2020) for parameter visualization relative to the prosodic structure of the utterances (example shown in figure 1).

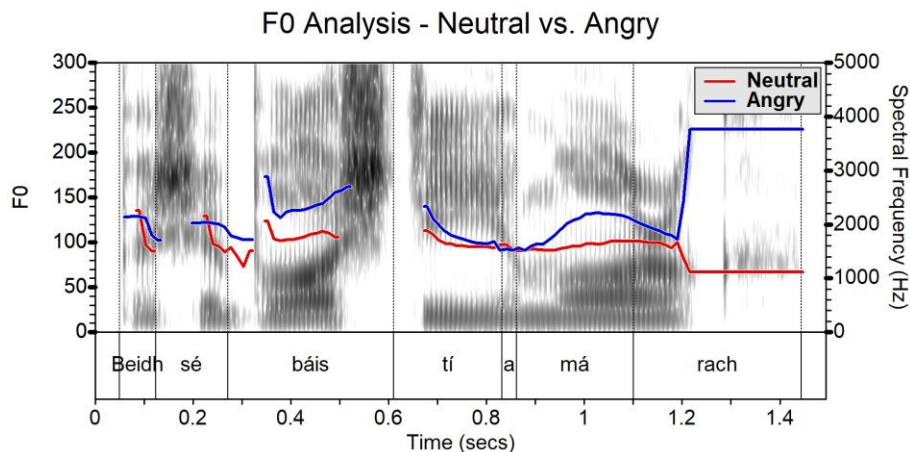


Figure 1. Preliminary visualization of the manual analysis comparing  $f_0$  measurements between *neutral* and *angry* exemplars for the sentence *Beidh sé báistí amárach* (“It will rain tomorrow”).

### Next steps

The construction of global and prosodically structured stimuli will be carried out by modifying the basic voice source characteristics of a synthesized utterance produced by the TTS voice for the speaker previously described using the system for voice analysis and synthesis GlórCáil (Murphy, Yanushevskaya,

Ní Chasaide and Gobl, 2020). Recordings and resynthesis of a parallel set of stimuli will be carried out for a female synthetic voice for the Connemara dialect, whose basic intonation contours are roughly similar to the Kerry renditions. These resynthesized stimuli will then be used on perception tests in order to evaluate the effectiveness of the manipulations in the perception of affect by listener judges.

### Acknowledgements

The research conducted in this publication was funded by the Irish Research Council under award number 17000, project number 211752.

### References

- Boersma, P., Weenink, D. 2020. Praat: doing phonetics by computer. Version 6.2.08.
- Fant, G., Liljencrants, J., Lin, Q. 1985. A four-parameter model of glottal flow. *STL-QPSR* 26(4), 1-13, Speech, Music and Hearing, Royal Institute of Technology, Stockholm.
- Gobl, C., Ní Chasaide, A. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40(1), 189-212.
- Gobl, C., Ní Chasaide, A. 2010. Voice source variation and its communicative functions. In Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds.) 2010, *The Handbook of Phonetic Sciences* (Second Edition), 378-423. Oxford, Blackwell.
- Juslin, P.N., Scherer, K.R. 2005. Vocal Expression of Affect. In Harrigan, H., Rosenthal, R. Scherer, K.R. (eds.) 2005, *The New Handbook of Methods in Nonverbal Behaviour Research*, 65-135. Oxford, Oxford University Press.
- Laver, J. 1980. *The phonetic description of voice quality*. Cambridge, Cambridge University Press.
- Murphy, A., Yanushevskaya, I., Ní Chasaide, A., Gobl, C. 2020. Testing the GlórCáil System in a Speaker and Affect Voice Transformation Task. 10<sup>th</sup> International Conference on Speech Prosody, 950-954. Tokyo, Japan.
- Murphy, A., Yanushevskaya, I., Ní Chasaide, A., Gobl, C. 2022. Affect Expression: Global and Local Control of Voice Source Parameters. *Speech Prosody 2022*, 525-529. Lisbon, Portugal.
- Ní Chasaide, A., Gobl, C., Monahan, P. 1992. A Technique for Analysing Voice Quality in Pathological and Normal Speech. *Journal of Clinical Speech and Language Studies* 2(1), 1-16.
- Scherer, K.R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40(1), 227-256.

# Comparing number and gender agreement processing in Russian: an experimental study

Anastasiia Ivanova<sup>1</sup>, Natalia Slioussar<sup>1, 2, 3</sup>

<sup>1</sup>School of Linguistics, Higher School of Economics, Russian Federation

<sup>2</sup>Faculty of Liberal Arts and Sciences, St Petersburg University, Russian Federation

<sup>3</sup>Institute for Cognitive Studies, St Petersburg University, Russian Federation

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

We conducted a self-paced reading experiment comparing attraction effects in number and gender agreement in Russian. Only one previous comprehension study (Tucker et al., 2021) compared them in Standard Arabic, getting distinct profiles. In Russian, larger reading time delays were associated with gender errors compared to number errors, but attraction effects with both features were similar.

Keywords: Russian, language comprehension, agreement attraction, grammatical gender, grammatical number

## Introduction

Agreement attraction, as in the sentence “The key to the cabinets were rusty”, in which the verb erroneously agrees not with the subject, but with a dependent noun, *an attractor*, has been extensively studied in many languages. Attraction effects were observed both in gender and in number agreement, both in production and in comprehension. However, very few studies tried to compare gender and number attraction effects, especially in comprehension: this was done only by Tucker et al. (2021) on Standard Arabic. We conducted a comprehension experiment on Russian, getting partially different results.

Russian language has three genders (masculine, feminine, neuter) and two numbers (singular, plural), as well as six cases. Adjectives and participles agree with nouns in case, number and gender (in singular). Verbs agree with subjects in person and number in the present and future tense and in number and gender (in singular) in the past tense.

Previous comprehension studies on Russian found attraction effects both in number and in gender agreement (Slioussar, Malko 2016; Slioussar 2018; Slioussar et al., 2022), but never compared them. In production, Lorimor et al. (2008) studied both features in one experiment, eliciting some number errors, but no gender errors (such errors were observed by Slioussar and Malko (2016), but their study did not include number agreement). Lorimor et al. concluded that gender agreement is more resistant to attraction, and experiments on Romance languages (e.g. Vigliocco et al., 1995) point to the same conclusion.

## Experiment

### Method

130 Russian native speakers (18-70 years old, mean age 22,4) took part in the experiment. We constructed 28 target sentence sets, as in (1). In all sentences, the head of the subject noun phrase was a feminine or neuter noun in nominative singular<sup>1</sup>. The number and gender of the accusative dependent noun and the predicate varied across seven experimental conditions shown in Table 1. All heads and dependent nouns were inanimate and syncretic (their accusative forms coincided with their nominative forms) to maximize attraction effects (see Slioussar 2018; Slioussar et al. 2022 for the role of syncretism in agreement processing in Russian).

- (1) Zapis' pro povest' / povesti / rasskaz  
 entry<sub>F.NOM.SG</sub> about novel<sub>F.ACC.SG</sub> / novels<sub>F.ACC.PL</sub> / story<sub>M.ACC.SG</sub>  
 byla/\*byl/\*byli najdena/\*najden/\*najdeny v dneknike pisatelja.  
 was<sub>F.SG/M.SG/PL</sub> found<sub>F.SG/M.SG/PL</sub> in writer's diary

	i.	ii.	iii.	iv.	v.	vi.	vii.
Dependent noun	N/F.Sg	N/F.Sg	N/F.Pl	N/F.Pl	N/F.Sg	M.Sg	M.Sg
Predicate	N/F.Sg	N/F.Pl	N/F.Sg	N/F.Pl	M.Sg	N/F.Sg	M.Sg

Table 1. Seven experimental conditions.

Three conditions were grammatically correct (i, iii, vi in Table 1), four contained number (ii, iv) and gender (v, vii) agreement errors. In the conditions in which the ungrammatical features of the predicate matched the relevant features of the dependent noun (iv, vii) attraction effects could be expected. Target sentences in different conditions were distributed across 7 experimental lists. There were also 66 grammatical filler sentences.

The word-by-word self-paced reading methodology on the IbeXFarm platform was used in the experiment. After one third of the sentences participants were asked a comprehension question with a choice of two answers to ensure that they were paying attention. No participants were excluded based on low accuracy, and below, only RTs are discussed.

### Results

Firstly, we excluded all RTs that exceeded a threshold of 2.5 standard deviations, by region and condition, from further analysis. This led to the exclusion of 0.8% of the data. Mean reading times in different conditions are shown in Figure 1 (we checked that the picture was very similar for the sentences with feminine and neuter subject nouns and analyzed them together).

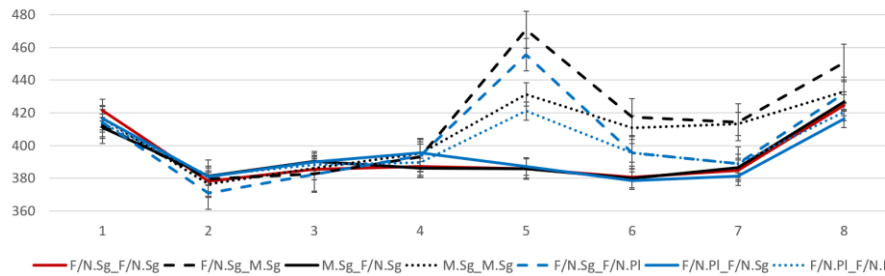


Figure 1. Average RTs per region (in ms) in the seven experimental conditions. Regions: N1 (head)—preposition—N2 (dependent)—copula (*byt'* 'to be')—adjective/participle—three words modifying the predicate.

The data were modelled with a mixed-effects regression in the R software. Random intercepts and random slopes by participant and by item were included in the model. Three factors were considered: *number* (number agreement error: yes/no), *gender* (gender agreement error: yes/no), *match* (the features of the subject match the features of the dependent noun: yes/no).

Statistically significant differences between conditions were found only on the fifth word (a participle, like 'found' in (1)). The number factor ( $\beta = -29,29$ ,  $SE = 6,88$ ,  $z = 4,25$ ,  $p < 0,01$ ) and the gender factor ( $\beta = -45,05$ ,  $SE = 6,83$ ,  $z = 6,59$ ,  $p < 0,01$ ) were significant, showing that readers slowed down on agreement errors. As we can see from the  $\beta$ -values, the effect of gender errors was more noticeable. The interaction between number and match ( $\beta = -43,44$ ,  $SE = 10,42$ ,  $z = 4,17$ ,  $p < 0,01$ ) and gender and match ( $\beta = -40,77$ ,  $SE = 10,47$ ,  $z = 3,89$ ,  $p < 0,01$ ) also reached significance, i.e. similar attraction effects were detected for number and gender agreement. Other comparisons were not significant.

## Discussion

Tucker et al.'s (2021) study on Standard Arabic was the only comprehension study comparing number and gender agreement attraction. They found that number attraction effects were smaller in comparison to gender agreement, but reached significance earlier (on the verb, not on the following word). At the same time, Tucker et al. noted that number attraction in Arabic appeared to differ from what was observed in other languages. Therefore, it was important to look at other languages to see if the comparison would yield similar results.

Our experiment demonstrated that in Russian, number and gender agreement attraction profiles are very similar in comprehension, both in terms of the effect size and timing. However, the delay provoked by gender agreement errors (both with and without attraction) was more pronounced in comparison to number errors. How to explain these findings and to reconcile them with the results of production studies (Lorimor et al., 2008) showing that number attraction errors are more frequent?

Stronger reaction to gender errors may have the following explanation. Gender is a property of the lexeme, while number is a property of a particular form, most nouns have both singular and plural, but only one gender. Maybe, this also explains why number attraction errors are more frequent in production — gender agreement may be in general more resistant to errors. However, if an error is present — and in comprehension studies, they are purposefully introduced in the stimuli — the same retrieval mechanisms are activated to check number and gender features, so attraction profiles are similar.

Further experiments are necessary to check whether these results would generalize to other constructions in which attraction is observed, most importantly, to relative clauses that were studied by Tucker et al. (2021), and to other languages with number and gender agreement. Our results are also interesting in the light of recent findings on Russian showing that in some other tasks that invoke memory, the number feature is more salient than the gender feature (Antropova et al., 2022). We can conclude that different properties of features become relevant depending on the nature of the task.

## Acknowledgements

The study was supported by the Russian Ministry of Science and Higher Education (the research project 075-15-2020-793).

## Notes

1. In their comprehension study on Russian, Slioussar and Malko (2016) observed no attraction effects for masculine heads, and crosslinguistically, no attraction effects were found for plural heads (on Russian, see Slioussar (2018)).

## References

- Antropova, D., Chernova, D., Slioussar, N. 2022. To forget or not to forget: processing gender, number and case features. A talk given at the AMLaP 2022 conference.
- Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., Beard, R. 2008. Agreement and attraction in Russian. *Language and cognitive processes*, 23, 769-799.
- Slioussar, N. 2018. Forms and features: The role of syncretism in number agreement attraction. *Journal of Memory and Language*, 101, 51-63.
- Slioussar, N., Magomedova, V., Makarova, P. 2022. The role of case syncretism in agreement attraction: A comprehension study. *Frontiers in Psychology*, 13, art. 829112, 1-13.
- Slioussar, N., Malko, A. 2016. Gender agreement attraction in Russian: production and comprehension evidence. *Frontiers in Psychology*, 7, art. 1651, 1-20.
- Tucker, M. A., Idrissi, A., Almeida, D. 2021. Attraction effects for verbal gender and number are similar but not identical: Self-paced reading evidence from Modern Standard Arabic. *Frontiers in psychology*, 11, art. 586464, 1-27.
- Vigliocco, G., Butterworth, B., Semenza, C. 1995. Constructing subject-verb agreement in speech. *Journal of Memory and Language*, 34, 186–215.



# Dominance shift during the stay-abroad experience: a tip-of-the-tongue experiment

Victoria Kamp, Maud Pélissier

Université Paris Cité, CLILLAC-ARP, F-75013, Paris France

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Previous research found that changes in the linguistic and socio-linguistic environment of bilinguals can provoke language dominance shifts. Investigating vocabulary loss permits to find differences in the quality and quantity of linguistic knowledge that deteriorates through attrition processes. This longitudinal study evaluated the effects of study abroad programs on L1 cognate and non-cognate word retrieval. Subjects were French-English bilinguals spending one to two semesters in an English-speaking country. Results of a picture-naming task show that the participants experienced less tip-of-the-tongue states in their L1 than in their L2 and retrieved cognates more easily than non-cognates. Profiling questionnaires helped to evaluate the correlation between those results and their monthly exposure to English.

Keywords: attrition; tip-of-the-tongue experiment; lexical retrieval; language dominance

## Introduction

The phenomenon of non-pathological loss of a native or first language (L1) is called attrition and can occur in a bilingual's (or multilingual's) life when he or she acquires and uses an additional language (L2), leading to an impoverishment of the first language (Schmid and Köpke, 2009). This provokes changes in the structural aspects of the language and the speaker's proficiency in the first language. Previous research indicates that vocabulary is one of the more vulnerable linguistic abilities (Köpke and Schmid, 2011). The attrition of the L1 is not only a consequence of the lack of use of it, but also a consequence of translinguistic influences that are entailed by the contact with the second language (Kroll & Stewart, 1994).

During studies abroad, a shift of dominance can ensue: the L1 is used less frequently, and lexical access becomes harder, while L2 proficiency increases. The competition between both languages as well as a dominance switch may lead to L1 attrition. Lexical access has been found to be one of the most vulnerable linguistic skills; and tip-of-the-tongue occurrences hint at a deterred access to the phonological form of the word, while the concept is activated. To account for TOTs in L2, two main frameworks have been used: the "weaker link" hypothesis (Gollan, Montoya, Cera & Sandoval, 2008), which postulates that the more restricted use of L2 leads to weakened connections between form and concepts, and language competition or cross-language interference, which

assumes that impaired retrieval comes from competition from the other language. This study focused on investigating the effects of studies abroad on L1 vocabulary loss. As it becomes harder to retrieve words in the neglected language, more tip-of-the-tongue states should occur.

## Methods

### Participants

19 native speakers of French aged 19 to 25 took part in this study. All were university students completing a bachelor or master's degree and had intermediate to advanced proficiency in English. All participants were spending one ( $n=3$ ) or two semesters ( $n=16$ ) abroad for studies in an English-speaking country. All participants filled out a questionnaire investigating their background as bilinguals. This was then followed by monthly forms to record the frequency of use for both languages.

### Materials and procedure

The participants were tested on L1 word retrieval on three occasions during their stay. A list of 420 stimuli were selected, comprising 180 English non-cognate words, 180 French non-cognate words and 180 French-English cognates. All stimuli were low-frequency words (frequency inferior to 50 per million occurrences) and were associated with black line drawings taken from the IPNP (Szekely et al., 2004) or drawn in the same style by the authors. Six balanced sets of 120 words were created, three in each language. Cognates were the same in English and French equivalent sets. The pictures were shown one by one on the screen and the participant had to name them as fast as possible while the experimenter recorded possible TOTs.

We predicted that participants would experience more TOTs in French by the end of the year, fewer TOTs in English, and name cognates more accurately than non-cognates.

### Analysis and results

Analyses were run in R (R Core Team, 2022) with linear mixed-effect models and generalised linear mixed-effect. Accuracy data was analysed with a binomial linear mixed-effect model with a logit link with Accuracy as a dependent variable and Language, Cognate status, Session and their interaction as fixed effects. Random effects included an intercept by Participant and an intercept by Item.

Analyses on the TOT accuracy data showed a significant effect of Language ( $\beta = -0.58$ ,  $SE = 0.07$ ,  $z = 7.69$ ,  $p < .001$ ), as participants experienced more TOTs in English than in French (English:  $M = 9.31$  %; French:  $M = 5.67$  %). The effect of Cognate status (cognate or non-cognate) was also significant ( $\beta = 0.86$ ,  $SE = 0.12$ ,  $z = 6.95$ ,  $p < .001$ ) as TOTs were experienced more frequently when naming non-cognate pictures than cognate ones (cognates:  $M = 5.31$  %;

non-cognates:  $M = 9.70\%$ ). Finally, there was also a significant effect of Session: performance in session 1 was significantly different from session 2 and session 3 ( $\beta = -0.43$ ,  $SE = 0.06$ ,  $z = 7.31$ ,  $p < .001$ ). The analysis showed that there was a significant interaction between Session and Cognate status ( $\beta = 0.41$ ,  $SE = 0.12$ ,  $z = 3.49$ ,  $p < .001$ ; see Figure 1). Pairwise comparisons showed that the cognate effect was present in each session (all  $ps < .001$ ), but that the effect was twice as big in session 3 ( $\beta = 1.269$ ) than in session 1 ( $\beta = 0.594$ ) (see Figure 1).

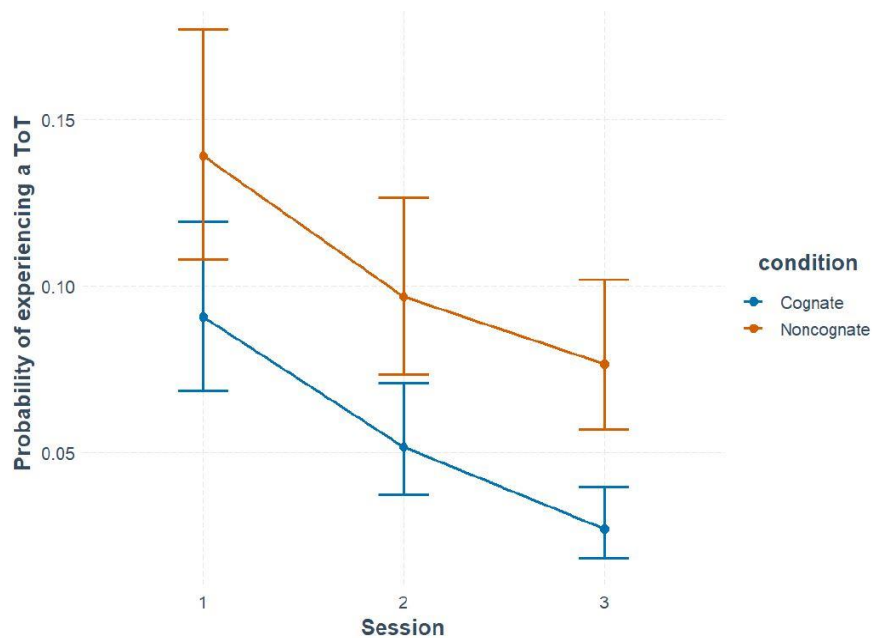


Figure 1. TOT rate as a function of Cognate status and Session (in both languages)

## Discussion

This study investigated the L1 vocabulary loss of French-English bilinguals spending a year abroad. Results of this exploratory experiment are in line with previous studies on two points: there was a facilitation effect linked to the cognate status of the stimuli and participants experienced more TOTs in their L2.

However, the results were inconclusive in relation to the starting hypothesis, given that the number of TOTs decreased over the course of the year more for cognates than for non-cognates, no matter the language. Several reasons can explain the results that contradict the starting hypothesis. For one, the pool of participants was not ideal in the sense that three of them only stayed for one

semester abroad and that three participants were French Language teachers in their host country. Additionally, two semesters have proven to be too short a time to have any conclusive results regarding loss of lexical access in L1, especially when the reality of study abroad programs is that students tend to speak more of their L1 than anticipated.

As for the difference between the first session and the two that followed, one explanation can be that the experimenter was not as precise in the way she registered the answers as she was getting used to the task and the process and tended to confuse real tip-of-the-tongues with a “feeling of knowing”. If the experiment was to be redone, it would also be important to have the option to record quickly resolved TOTs so as to have more accurate data.

The fact that the TOT rate diminished more for cognates than for non-cognates could point to habituation to the task. It could also suggest that there is increased non-selective activation (Marian & Spivey, 2003), due to more balanced bilingualism as the participants’ proficiency in French does not decrease while their proficiency in English increases. Additional research is needed to investigate this possibility.

## References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. 2008. More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of memory and language*, 58(3), 787-814.
- Köpke, B., & Schmid, M. S. 2011. L’attrition de la première langue en tant que phénomène psycholinguistique. *Language, Interaction and Acquisition*, 2(2), 197-220.
- Kroll, J. F., & Stewart, E. 1994. Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, 33(2), 149-174.
- Lenth, R., V., 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.3. <https://CRAN.R-project.org/package=emmeans>
- Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: language and cognition*, 6(2), 97-115.
- R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Schmid, M. S., & Köpke, B. 2009. L1 attrition and the mental lexicon. *The bilingual mental lexicon: Interdisciplinary approaches*, 209-238.
- Szekely, A., Jacobsen, T., D’Amico, S., Devescovi, A., Andonova, E., Herron, D., ... & Bates, E. 2004. A new on-line resource for psycholinguistic studies. *Journal of memory and language*, 51(2), 247-250.

# **Narrative skills of Russian heritage children in Cyprus**

Sviatlana Karpava

Department of English Studies, University of Cyprus, Cyprus

<https://doi.org/10.36505/ExLing-2022/13>

## **Abstract**

The present study investigated narrative skills of 40 Russian heritage children (HL) in Cyprus (age: 4;0-6;0), with the focus on macro-structure (story structure, structural complexity and internal states terms) and grammaticality, i.e. the extent to which utterances follow the grammatical rules of a language. The analysis of the data revealed that the level of language proficiency, mode of narration (telling/retelling) affected their narrative abilities, as the measures of proficiency in HL were correlated with the macro-structure. The grammaticality analysis revealed that functional categories were the most vulnerable domain for Russian heritage children, which may have been due to cross-linguistic interference or developmental/language acquisition pattern in the language-contact setting. The number of errors was correlated with the language proficiency measures.

Keywords: Russian Heritage children, narrative skills, macro-structure, grammaticality

## **Heritage language development and narrative skills**

Heritage language maintenance and development depend on family language policy, language dominance, frequency of use, linguistic distance and similarity between the minority and the majority languages or (dia)lects of the society, multi-directionality of cross-linguistic influence and accommodation (Rothman et al., 2019). The present study investigated narrative skills of Russian heritage children in Cyprus, with the focus on macro-structure (story structure, structural complexity and internal states terms), micro-structure and grammaticality, i.e. the extent to which utterances follow the grammatical rules of a language. Grammaticality in narrative tasks can predict language development (Bedore et al., 2010).

Narratives can measure cognitive, linguistic and social skills of bilingual and multilingual children in a less biased way than standardized assessments of language. Narratives can help to identify linguistic, cognitive, semantic and social abilities, communicative competence and cultural awareness of a child. Narrative skills are essential for children's success at school. There is a relationship between oral language (experience with and exposure to discourse) and literacy (Westby, 2012). Cultural communities, language environment, home language use, parental attitudes towards bilingual and bi-cultural learning,

the level of language proficiency are some of the factors that can affect the development of narrative abilities (Fiestas, 2004).

## The Study

The participants of the study were 40 Russian–Cypriot Greek simultaneous bilingual children. Their age ranges from 4;0 to 6;0 (mean 5;2), and they attend kindergarten and primary CG schools, where the language of instruction is Greek. The LITMUS-MAIN, the multilingual assessment instrument for narratives (Gagarina et al., 2015) was used for data collection. Their language proficiency in Russian was measured by the Russian Proficiency Test for Multilingual Children (RPTMC) (Gagarina et al., 2010). Background information was collected using parental questionnaires and interviews. The narratives were recorded, transcribed and analyzed in terms of macro-structure and grammaticality. Grammaticality percentages were calculated for each child.

## Results

The analysis of the data revealed that Russian heritage speakers had better productive skills than perceptive lexical skills, and that they had superior performances with regard to noun production in comparison to verb production, which was not the case for lexical comprehension. The difference between the two age groups regarding such language proficiency measures of their heritage Russian language as the comprehension of grammatical structures, morphological marking on the verb, and receptive and productive lexicon confirmed that there was a developmental pattern regarding their heritage language, see Table 1.

Table 1. Russian Proficiency Test for Multilingual Children: Results.

RPTMC/ Mean scores	Productive lexicon total (52)	Noun production (26)	Verb production (26)
4;0-5;0	28.65	18.2	10.45
5;1-6;0	29.5	17.6	11.9
RPTMC/ Mean scores	Receptive lexicon total	Noun perception (10)	Verb perception (10)
4;0-5;0	14.65	7.35	7.3
5;1-6;0	15.8	8	7.8
RPTMC/ Mean scores	Case (6)	Comprehension of grammatical structures (22)	Morphological marking on the verb (12)
4;0-5;0	3	14	7.1
5;1-6;0	3	15.55	8.4

It was found that the mode of narration (telling/retelling) influenced the macro-structure; in particular, the telling mode triggered the use of more ISTs by the four-year-old bilingual children and improved/more coherent story structure by the five-year-old bilingual children, whereas the retelling mode led to higher scores for story structure, structural complexity and comprehension

questions for the younger age group and to structural complexity, ISTs and comprehension questions for the older group.

Table 2. Macro-structure and comprehension measures.

MAIN/Mean scores	Story structure (17)	Structural complexity (9)	Internal state terms	Comprehension questions (10)
Retelling: 4;0-5;0	6.5	2.5	2.8	7.55
Telling 4;0-5;0	5.8	2	3.55	4.85
Retelling 5;1-6;0	6.5	2.2	3.15	8.45
Telling 5;1-6;0	6.7	2	2.6	6.7

The results revealed that the mode of narration influenced narrative production in Russian. Overall, this is in line with previous findings that the retelling mode elicits longer and more detailed narratives with a more complex story structure than does the telling mode (e.g. Otwinowska et al., 2020). The ages of the participants appeared to affect their narrative skills, particularly their use of ISTs, story structure and their comprehension of the story. The level of language proficiency affected their narrative abilities, as it was found that the measures of proficiency in the heritage language were correlated with the macro-structure measures. There was an obvious effect of chronological age and language proficiency, both for production and comprehension, which is in line with previous studies (e.g. Lindgren, 2019).

Errors/Mean Non-target production	Retelling 4;0-5;0	Telling 4;0-5;0	Retelling 5;1-6;0	Telling 5;1-6;0
Mean number of utterances	10.3	9.65	9.7	9.75
Mean number of words	43.65	46.75	49.45	56.5
MLU	4.39	4.85	5.14	5.73
One-word sentence/utterance	1.1	0.55	3.3	0.2
Null subject pronoun	0.95	0.45	0.6	0.3
Aspect: imperfective instead of perfective	0.25	5	0.45	0.5
Aspect: perfective instead of imperfective	0.1	0	0	0.1
Wrong verb form/innovation	0.55	1.3	0.9	0.85
Infinitive instead of subjunctive	0.05	0.05	0.05	0
Infinitive instead of finite verb	0.05	0	0.05	0
Direct object omission	1.05	1.1	0.7	0.9
Repetition of conjunction	0.7	0.7	0.8	0.75
Case error	0.85	0.65	0.5	1.15
Subject-Verb agreement: Gender	0.25	0.3	0.45	0.35
Subject-Verb agreement: Number	0.05	0.05	0	0.1
Pronoun + Noun agreement: Gender	0.05	0.05	0	0
Word order	0.25	0.3	0	0.05
Wrong noun form: Innovation	0.1	0.15	0	0.01

<b>GAP verb/Circumlocution</b>	0.05	0.35	0.1	0.3
<b>Wrong preposition</b>	0.25	0	0.35	0.25
<b>Omission of preposition</b>	0.05	0	0.2	0.15
<b>Code-switching</b>	0	0.05	0.05	0.1
<b>Wrong conjunction</b>	0	0.05	0.05	0
<b>Total errors</b>	6.7	6.8	8.55	6.15

The grammaticality analysis revealed that the most vulnerable domains for Russian heritage children were morphological agreement, gender, case and aspect features, functional categories such as subject and object pronouns, conjunctions and prepositions, as well as word order, word formation and morphology, which may have been due to cross-linguistic interference or developmental/language acquisition pattern in the language-contact setting, see Table 3. The number of errors was correlated with the measures of language proficiency and not with the measures of macro-structure, with the exception of ISTs.

## References

- Rothman, J., Alonso, J.G., Puig-Mayenco, E. 2019. *Third Language Acquisition and Linguistic Transfer*. Cambridge University Press.
- Bedore, L. M., Peña, E.D., Gillam, R.B., Ho, T. 2010. Language sample measures and language ability in Spanish English bilingual kindergarteners. *Journal of Communication Disorders* 43(6), 498–510.
- Westby, C. 2012. Assessing and remediating text comprehension problems. In A.G. Kamhi, H. Catts (eds.), *Language and Reading Disabilities*, 163-225. Boston, MA: Pearson.
- Fiestas, C., Peña, E. 2004. Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools* 35, 155-168.
- Gagarina, N., Klassert, A., Topaj, N. 2010. Sprachstandstest Russisch für mehrsprachige Kinder [Russian language proficiency test for multilingual children]. *ZAS Papers in Linguistics*, 54.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., Walters, J. 2015. Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, N. Meir (eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment*, 243-276. Bristol: Multilingual Matters.
- Otwinowska, A., Mieszkowska, K., Bialecka-Pikul, Opacki, M., Haman, E. 2020. Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism* 23, 1083-1107.
- Lindgren, J. 2019. Comprehension and production of narrative macrostructure in Swedish: A longitudinal study from age 4 to 7. *First Language* 39, 412-432.



# Negative vaccine voices in Swedish social media

Dimitrios Kokkinakis<sup>1</sup>, Mia-Marie Hammarlin<sup>2</sup>

<sup>1</sup>Department of Swedish, Multilingualism, Language Technology, University of Gothenburg, Sweden

<sup>2</sup>Department of Communication and Media, University of Lund, Sweden

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Vaccinations are one of the most significant interventions to public health, but vaccine hesitancy creates concerns for a portion of the population in many countries, including Sweden. Since discussions on vaccine hesitancy are often taken on social networking sites, data from Swedish social media are used to study and quantify the sentiment among the discussants on the *vaccination-or-not* topic during phases of the COVID-19 pandemic. Out of all the posts analyzed a majority showed a stronger negative sentiment, prevailing throughout the whole of the examined period, with some spikes or jumps due to the occurrence of certain vaccine-related events distinguishable in the results. Sentiment analysis can be a valuable tool to track public opinions regarding the use, efficacy, safety, and importance of vaccination.

Keywords: vaccine hesitancy, sentiment analysis, Swedish, exploratory study

## Introduction and background

According to the World Health Organization (2019) vaccine hesitancy, “the reluctance or refusal to vaccinate despite the availability of vaccines”, was one of the top-10 threats to global health even before the pandemic. Although vaccinations are considered as one of *the* most significant interventions to public health, vaccine hesitancy and resistance creates serious concerns for a significant part of a population in many countries, including Sweden. Vaccine discussions are often taken online in various social media platforms. For many people, the use of such platforms is the major source for information related to health issues including vaccinations (Klimiuk et al., 2021). Consequently, the availability of such digital content enables researchers to rapidly analyze and monitor large amounts of data, to e.g., identify and better understand the vaccine-deniers’ arguments against vaccinations which in turn, can rapidly be spread as *rumours* to an even wider audience. The aim of this study is to investigate the Swedish public sentiment reflected in discussions related to the COVID-19 vaccines (Scannell et al., 2021; Yousefinaghani et al., 2021). Sentiment analysis could guide us to identify, classify and quantify posts related to vaccine hesitancy and provide useful information to assist public health experts and pro-vaccine organizations to formulate even more targeted policies and strategies to reduce anti-vaccine reactions. This work is part of a project

that investigates vaccination skepticism, and how it can be understood as an expression of civic engagement in the present digital times.

### Data collection and preparation

The data in the study comes from a popular Swedish open discussion platform *Flashback Forum*. Fourteen threads have been chosen that discuss topics related to vaccination, such as *Vad ska man göra om man vägrar ta corona-vaccin?* ‘What to do if you refuse to take the corona vaccine?’; collected prospectively between Sept. 2020 – Jan. 2022; 21,000 unique posts from over 2000 unique users. For each post, we stored the text and relevant metadata such as the date of publication and the number of posts produced by each user. During preprocessing, duplicate posts were removed and transformed to lower case. The dataset was further tokenized (separating punctuation and metadata from words), while emoticons were converted to a text description (☺: <smiling face with sunglasses>). Multiword expressions and phrasal verbs were also recognized, and their contiguous components were joined with an underscore prior to further processing (*spanska sjukan* ‘the Spanish flu’; *spruta in* ‘to inject’).

### Sentiment analysis and sentiment shifters

Sentiment analysis (Liu, 2012) is an application that infers subjective aspects of writing such as opinions, feelings, and classify the polarity of a text as *positive*, *negative*, or *neutral*. Sentiment analysis is a technology that allows us to gain a wider public opinion behind certain topics and a picture of how public opinion develops over time. There are many ways to approach the task (Birjali et al., 2021). Here, we apply a “classical” dictionary-based approach having a Swedish word list as its main component (*SenSaldo*; Rouces et al., 2018). Each tokenized text fragment in each post, is assigned a sentiment value according to *SenSaldo*. Moreover, since the polarity of individual words can change, when these are used in specific contexts, we also identify *sentiment shifters*, i.e., modifiers that change the sentiment orientation of words, e.g., via negation. Finally, the sum of the words’ sentiment values for each post are summarized and scored, as the ratio of words’ sentiment values by the total number of tokens in a post. E.g., for the fragment: *jo det är ni idioter som är hysteriska och de verkliga träskallarna* ‘well it’s you idiots who are hysterical and the real numbskulls’ there are 3 underlined negative words found out of the 12, which returns a negative score of -25.

### Exploratory analysis

Sentiment analysis is the practice of extrapolating the sentiment of an idea or event by automatically classifying written texts as some polarity value. In the specific context of *vaccine hesitancy/skepticism*, the people’s attitudes present a dominant negative polarity throughout the whole covered period with 56.2% negative, 25.1% positive and 18.7% neutral posts. Figure 1 (left) shows the 25 most frequent positive (left; the most frequent is: ‘fullt frisk’ *completely healthy*) and negative (right; the most frequent is: ‘allvarlig sjukdom’ *serious disease*)

*multiwords* in the data as word clouds. Figure 1 (right) shows the proportion of positive/negative users in thread t3342731: *Vaccinerade, när kan ni tänka er ta en tredje dos?* ‘Vaccinated, when can you imagine taking a third dose?’. Using sentiment analysis, negative voices can be identified and further analyzed using *qualitatively driven mixed methods* such as the sociologically-grounded narrative typology (Frank, 2013), and the exploration of the civic aspects of ethnological research. These are research directions that we have already started to explore in this project (cf. the Acknowledgements).

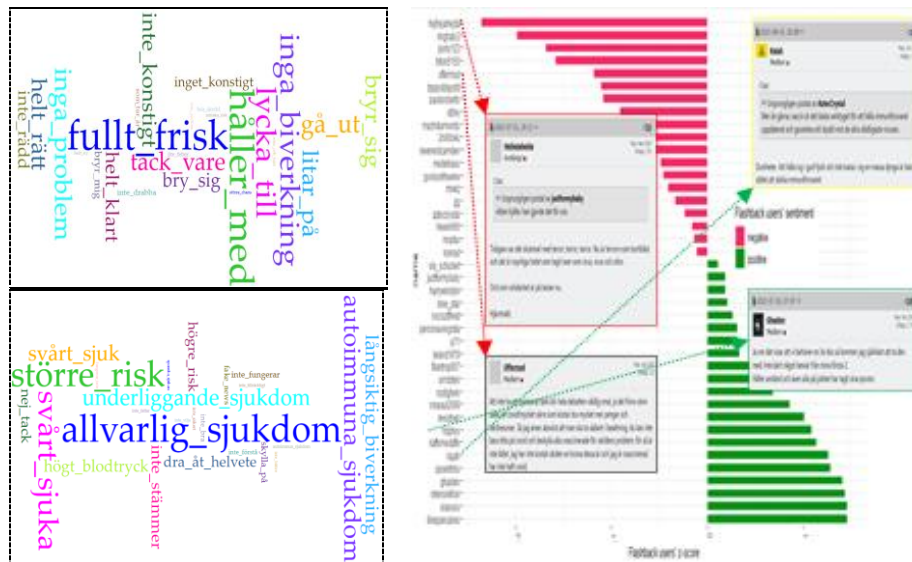


Figure 1 (left). The 25 most frequent *multiword* sentiments in the whole data (positive at the top and negative at the bottom). Figure 1 (right). The proportion of positive/negative users in thread t3396531.

## Conclusions, limitations and future work

It is well established that the Internet provides valuable resources to fight vaccine hesitancy, providing insights to vaccine-undecided individuals concerns. Digital technologies such as sentiment analysis are important tools which can be used for gaining helpful insights into various complex health related topics (cf. Sandhya et al., 2021). Since this is an exploratory study there are many questions left for future work, for instance to investigate whether the positive sentiment messages are the pro-vaccination messages while the negative sentiment messages are the voice of the opposite. The key motivation of sentiment analysis is to analyze a body of text for understanding the opinion expressed by it. Public sentiment is important to e.g., determining appropriate interventions and policies, and therefore these techniques have been used in

many scientific and commercial applications. Nevertheless, sentiment analysis has its limitations with respect to accuracy, which can be lower for certain types of creative language use such as sarcasm and irony (Hernández Fariás & Rosso, 2017). In the future, we would like to improve the accuracy of the sentiment analysis tool by employing more advanced techniques that can also capture subtleties of opinion and emotions beyond simple polarity (Izard, 2009). The dictionary-based approach, perhaps is not the best technique (Atteveldt et al., 2021). Thus, to demonstrate the effectiveness of this approach, a comparison between different methods (e.g., dictionary *vs* machine learning) is left for future research. Another limitation of this work is the use of a small set of social media posts, and from only one available site. This might cause a sampling bias, therefore, soon we plan to enhance the dataset with posts from two other major social media sites, such as *Familjeliv* ([www.familjeliv.se/](http://www.familjeliv.se/)).

## Acknowledgements

This work is part of the project *Rumour Mining* (MXM19-1161:1) financed by the *Bank of Sweden Tercentenary Foundation* (Riksbankens Jubileumsfond).

## References

- Atteveldt, et al. 2021. The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches & ML. *Com Meth Meas.* 15:2
- Birjali, M., Kasri, M. and Beni-Hssane, A. 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Syst.* 226.
- Frank, AW. 2013. *The wounded storyteller: body, illness, and ethics.* 2nd ed., UCP.
- Hernández Fariás, DK and Rosso P. 2017. Irony, Sarcasm, and Sentiment Analysis. In *Sentiment Analysis in Social Networks.* Pozzi F. (et al.) (eds). 113-128. M Kaufm.
- Izard, CE. 2009. Emotion theory and research. *Ann Rev Psych.* 60:1-25.
- Klimiuk, K., et al. 2021. Vacc. misinf. on social media—topic-based content & sentiment analysis of Polish vaccine-deniers’ comments on FB. *Hum Vacc & Immunoth* 17:7.
- Liu, B. 2012. *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers.
- Rouces, J., Tahmasebi, N., Borin, L., and Rødven Eide, S. 2018. SenSALDO: Creating a Sentiment Lexicon for Swedish. 11th Lang Res and Eval (LREC). 4192-98. Japan.
- Sandhya, A., et al. 2021. Information Extraction & Sentiment Analysis to gain insight into the COVID-19 crisis. In *Advances in Intelligent Syst & Comp* 1387. Springer.
- Scannell, D., et al. 2021. COVID-19 Vaccine Discourse on Twitter: A Content Analysis of Persuasion Techniques, Sentiment and Mis/Disinformation. *J Health Com.* 26:7.
- Yousefinaghani, S., et al. 2021. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *J of Inf Dis.* 108, 256-262.

# Exploring the role of L1 reading ability when reading in L2

Anastasiia Konovalova, Vladislav Zubov

Ludmila Verbitskaya Department of General Linguistics, St Petersburg University,  
Russian Federation

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

The present research is devoted to the study of the role of L1 (Russian) reading ability when reading in L2 (English). The text comprehension was assessed using a question and answer technique and a keywords method (Murzin, Stern 1991). Q&A results analysis did not reveal differences between the number of correct answers to questions for different texts in the same language and the same texts in different, however, a significant difference in comprehension of one of the texts in L1 and another one in L2 was found. It is assumed that this difference is due to a combination of text content and text language factors. Comparison of a set of keywords will allow us to evaluate this contribution of these factors more specifically.

Keywords: reading, text comprehension, L2, question and answer technique, keywords method

## Introduction

The study of the mechanisms of reading and text comprehension is one of the main topics of experimental linguistics. This topic is essential for the construction of theoretical models of speech perception and in the development of reading theory. Text comprehension includes the construction of an abstract meaning based on the decoding of linguistic material, as well as the integration of new information and background knowledge into the mental text representation (Borisenko, Shulekina 2021). The success of text comprehension is influenced by many factors, such as reader characteristics (age, reading experience, reading disorders, etc.), text characteristics (complexity, text format, genre), etc. The role of the language system in text processing is well studied in psycholinguistics, pedagogy, psychology, neurophysiology, however, the role on language in text comprehension needs further exploration. L2 reading studies claim that L2 language knowledge, L1 reading ability, strategic knowledge, and background knowledge contribute to reading comprehension, emphasizing the need to study each factor as independent (McNeil 2012).

## Materials

The material of the study was two pairs of non-fiction prosaic texts from (Kuperman et al. 2022), one text tells the story of the origin and meaning of the shaka gesture ('Shaka'), the second is about the god Janus ('Janus').

Russian and English texts readability indices were close (see Table 1 below). The text readability was assessed using [readable.com](https://readable.com) (English) and [readability.io](https://readability.io) (Russian) services.

## Participants

112 native speakers of Russian (93 female, age 18–59,  $M_{\text{age}} = 22.91$ ) took part in the experiment. All participants speak English at the level B2–C2 (B2 – 128 participants, C1 – 74 participants, C2 – 14 participants). The participants determined the language level themselves, based on their competencies. In addition, participants reported the number of years they have been learning English ( $M = 13.14$ ).

## Method

Participants were asked to read two texts in L1 and L2, write out 10 keywords from the texts, and answer 8 questions on the content of the text: 2 multiple choice questions, 2 questions that require a detailed answer, and 4 general questions. While answering the questions the participant could not return to the text. The keywords were defined for participants as 'words from the text that are most important in terms of the text content'.

Each participant assessed the subjective complexity of the text on a scale from -3 to 3 (for ease of analysis, the scale was converted to seven points, where - the easiest text, 7 – the most difficult).

The method of keywords extraction and analysis is described in [Murzin, Shtern 1991]. A group of participants is asked to write out keywords from the text. Each of the participants will extract their own set of keywords. Some words will be common, some will be different, which is due, on the one hand, to the same understanding of the text, and on the other hand, to the individual differences in the understanding of both the content of the text and the task of indexing.

For each word that appears in the answers, you can determine the absolute frequency of occurrence ( $m$ ), as well as the relative frequency of occurrence ( $p$ ):  $p = m / n$ , where  $n$  is the number of participants. The group of words that received the highest relative frequency of occurrence constitutes the "true" set of keywords.

## Results and discussion

In this paper, only quantitative data will be analyzed (see Table 1).

Table 1. Characteristics of texts and average values of participants' answers.

Text	Language	Flesch-Kincaid Grade level	SMOG formula	Correct answers	Subjective complexity rate
Janus	English	10.40	12.70	6.46	3.07
Janus	Russian	8.98	9.73	6.91	2.39
Shaka	English	14.00	15.90	6.72	2.93
Shaka	Russian	14.51	13.42	7.20	2.16

A correlation was found between the number of correct answers to questions to texts on L2 and the level of L2 proficiency (Spearman's  $\rho = 0.411$ ,  $p < 0.001$ ), as well as the number of years of studying L2 (Spearman's  $\rho = 0.199$ ,  $p < 0.038$ ). Some studies suggest that there is a positive correlation between L2 proficiency and academic performance, others say that the relationship is not significant (see Al-Busaidi 2021 for a review). The discrepancy in findings may be since there is no clear definition of the concepts of language proficiency and academic achievement, variations in research design and data collection techniques and other reasons.

All texts in a foreign language were evaluated more difficult than texts in their native language ( $t = 4.38$ ,  $p < 0.001$ ). No statistically significant correlation was found between the subjective text complexity assessment and readability indices (Spearman's  $\rho = -0.07$ ,  $p = 0.304$  for Flesch-Kincaid; Spearman's  $\rho = 0.059$ ,  $p = 0.381$  for SMOG), which can be explained by two reasons: either the shortcomings of the readability index formulas for different languages, or the fact that the complexity of the text is a more voluminous concept than readability and is not directly related to readability (Hiebert 2011).

Analysis did not reveal differences between the number of correct answers to questions for different texts in the same language ( $p_{\text{Tukey}} = 0.493$ ,  $p_{\text{Tukey}} = 0.671$ ) and the same texts in different languages ( $p_{\text{Tukey}} = 0.292$ ,  $p_{\text{Tukey}} = 0.449$ ), however, a significant difference in comprehension of one of the texts in L1 and another one in L2 ( $p_{\text{Tukey}} = 0.021$ ) was found. It is assumed that this difference is due to a combination of text content and text language factors. A similar result was obtained for subjective assessments of complexity: the most complex text was Janus in English, and the easiest was Shaka in Russian. Janus in Russian and Shaka in English were rated the same ( $p_{\text{Tukey}} = 0.104$ ).

In a further study, it is planned to analyze the keywords extracted by the participants when reading the text on L1 and L2. Comparison of sets of keywords can, firstly, assess the text comprehension skill on L1 and L2 in more detail, and secondly, describe more specifically the contribution of language processing to text comprehension.

## Acknowledgements

The study has been conducted within the research project “Text processing in L1 and L2: Experimental study with eye-tracking, visual analytics and virtual reality technologies” and supported by the research grant No. ID92566385 from Saint Petersburg State University.

## References

- Al-Busaidi, S. 2021. Language Proficiency and Academic Performance. In Mohebbi, H., Coombe, Ch. (eds.) 2021, *Research Questions in Language Education and Applied Linguistics*, 213-218. Springer Cham.
- Borisenko, N.A., Shulekina, Yu.A. 2021. Semantic reading. In Melentyeva Yu. P. (ed.) 2021, *Chtenie. Enciklopedicheskij slovar'* [The encyclopedic dictionary of reading], 231. Moscow, Nauka. [in Russian].
- Hiebert, E.H. 2011. Beyond Single Readability Measures: Using Multiple Sources of Information in Establishing Text Complexity. *Journal of Education* 191(2), 33-42.
- Kuperman, V., Siegelman, N., Schroeder, S., Acartürk, C., Alexeeva, S., Amenta, S., . . . Usal, K. 2022. Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus. *Studies in Second Language Acquisition*, 1-35.
- Murzin, L.N., Shtern, A.S. 1991. *Tekst i ego vosprijatie* [Text and its perception]. Sverdlovsk, Ural Univ. Publ. [in Russian].
- McNeil, L. 2012. Extending the compensatory model of second language reading. *System* 40(1), 64-76.



# Implicit causality in Romanian interpersonal verbs

Sofiana-Iulia Lindemann, Mădălina Matei

Department of Theoretical and Applied Linguistics, Transilvania University of Brasov,  
Romania

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Natural languages display a great variety of devices that may be used to speak of causal relations, ranging from prepositions, sentence connectives and verbs. This paper focuses on the way in which different classes of verbs affect the subsequent discourse in terms of implicit causality. We report on an offline sentence-continuation study that tested next mention preferences triggered by four Romanian classes of verbs and compare them these results with verbal biases observed in other languages.

Keywords: implicit causality, verbs, pronoun resolution, next mention, Romanian

## Introduction

Natural languages display a great variety of devices that may be used to speak of causal relations, such as prepositions, sentence connectives (e.g. *because, so*), or verbs, as for example *impress* or *hit*. The present paper focuses on implicit causal relations expressed inter-clausally by verbs. Implicit causality (IC) refers to the property of interpersonal verbs to relate two human or animate entities in such a way that one of the entities is “implicated as the assumed locus of the underlying cause of the action or attitude (Garvey and Caramazza 1974: 460)”.

The literature distinguishes between three classes of verbs in terms of their effect on the subsequent discourse: some, such as *telephone* or *approach* assign the cause of the event to the first noun phrase (NP1), the subject, while other verbs, such as *fear* or *praise* attribute the cause to the second noun phrase (NP2), the object. The third class of verbs, such as *see* or *bear*, are considered neutral towards the assignment of the cause. The proportion of continuations picking up the subject or the object referent is referred to as IC bias (e.g. Crinenan & Garnham 2006). These biases towards the subject or the object referent are manifestations of an expectation about who will be mentioned next in the discourse, but they can also affect the resolution of a subsequent pronoun or the coherence relation used in the next sentence (Garvey et al. 1974, Kehler, Kertz, Rohde & Elman 2008, Hartshorne & Snedeker 2014, Chiriacescu 2011, Lindemann & Homană 2019). It was argued that this property is rooted in the argument structure properties of verbs. In the current paper, we focus on next-mention biases triggered by interpersonal verbs.

One central question in the literature on IC pertains to the cross-linguistic and cross-cultural validity of these biases. Moreover, those studies comparing IC biases in different languages found that even though IC effects display similar patterns, the strength of these effects differs among languages. This paper reports on an off-line sentence-continuation study in which we tested the implicit causality biases of four classes of interpersonal verbs in Romanian.

### **The experimental study**

The aim of the study was to examine the implicit causality biases of 48 Romanian verbs in terms of their next-mention preferences. We furthermore investigated the coherence relations used to introduce the event in the immediately following sentence.

### **Participants**

56 monolingual native speakers of Romanian from the Transilvania University of Brasov, Romania, took part in the experimental study (age range: 18–45 years, mean age 28 years, 29 female). It took about twenty minutes to complete each version of the study.

### **Design, materials and procedure**

We selected 48 verbs according to two criteria: (i) they belong to one of the four classes of verbs discussed in Rudolph & Försterling (1997): Agent-Patient (AP), Patient-Agent (PA), Stimulus-Experiencer (SE), Experiencer-Stimulus (ES); and (ii) they were tested in other languages with respect to their IC biases (e.g. Goikoetxea, Pascual & Acha 2008; Bott & Solstad, 2014).

Two same-gendered human referents were mentioned in each experimental item. We manipulated the verb class, by choosing verbs from the four main classes of verbs discussed in the literature on implicit causality (IC): AP, PA, ES, SE verbs, as illustrated in Table 1. Participants were instructed to write one sentence continuation to each item, which did not end in a (pronoun or connective) prompt. The resulting corpus of 1886 continuations was annotated with respect to choice of subsequent mention, i.e. whether participants opted to continue the next sentence re-mentioning the initial subject (NP1) or the initial object (NP2).

### **Predictions**

In light of previous findings (e.g. Goikoetxea et al. 2008; Hartshorne & Snedeker 2013), we expect IC to affect the choice of subsequent mention, such that we obtain more subject continuations for SE verbs (i.e., NP1 biasing) and more object continuations for ES verbs (i.e., NP2 biasing). For AP and PA verbs, we expect more heterogeneous results (e.g. Ferstl et al., 2011).

Table 1. Sample experimental items with English translations

Agent-Patient	Marian l-a iertat pe Sorin. Marian forgave Sorin.
Patient-Agent	Adela a recompensat-o pe Irina. Adela recompensated Irina.
Stimulus-Experiencer	Paul l-a speriat pe Mihai. Paul scarred Mihai.
Experiencer-Stimulus	Diana a plăcut-o pe Gabriela. Diana liked Gabriela.

### Preliminary results

Results show that implicit causality influences reference in terms of subsequent next mention. First, as expected, SE verbs were significantly more subject-biased (at least 70%) than the ES verbs, which showed a clear preference for the initial direct object (at least 81%). The AP verbs showed a similar pattern, as the initial subject referents were more prone to be picked up in the immediately following matrix clause.

Second, the distribution of the discourse relations associated with the four verb types following a full stop reveals differences in terms of strength and coherence type. Overall, participants preferred more explanation relations than any other type of continuations. PA verbs prefer explanations in 92% of cases, while ES and SE verbs prefer them in 66% and 52% of cases respectively. For AP verbs, participants preferred elaborations (47%) and results (23%) rather than explanations (25%). We will discuss these results in light of cross-linguistic findings on implicit causality.

### Acknowledgements

This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P4-ID-PCE-2020-0731, within PNCDI III.

### References

- Bott, O., Solstad, T. 2014. From verbs to discourse – a novel account of implicit causality. In Hemforth, B., Mertins, B., Fabricius-Hansen, C. (eds.), 2014, *Psycholinguistic approaches to meaning and understanding across languages*, 213–251. New York, Springer.
- Chiriacescu, S.I. 2011. Effects of reference form on frequency of mention and rate of subsequent mentio. In Devi, S.L., Branco, A., Mitkov, R. (eds.), 2011, *Anaphora Processing and Applications. Lecture Notes in Computer Science*, 132-142. Heidelberg, Springer.
- Crinenan, M., Garnham, A. 2006. Implicit causality, implicit consequentiality and thematic roles. *Language and Cognitive Processes* 21, 636-648.

- Garvey, C., Caramazza J., Yates, F. 1974. Factors influencing assignment of pronoun antecedents. *Cognition* 3, 227-243.
- Goikoetxea, E., Pascual, G., Acha, J. 2008. Normative study of the implicit causality of 100 interpersonal verbs in Spanish. *Behavior Research Methods* 40, 760-772.
- Hartshorne, J. K., Snedeker, J. 2013. Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes* 28 (10), 1474-1508.
- Ferstl, E. C., Garnham, A., Manouilidou, C. 2011. Implicit causality bias in English: a corpus of 300 verbs. *Behavior Research Methods* 43(1), 124-135.
- Kehler, A., Kertz, L., Rohde, H., Elman, J. L. 2008. Coherence and coreference revisited. *Journal of Semantics* 25(1), 1-44.
- Lindemann, S.I., Homană, S. 2019. The implicit causality in verbs. *Scientific Bulletin of the Politehnica University of Timisoara* 18 (1), 14-22.
- Rudolph, U., Forsterling, F. 1997. Zur impliziten Kausalität in Sprache: Kriterien zur Selektion von Stimulusmaterial. *Zeitschrift für Experimentelle Psychologie* 44, 293-304.

# The L1 influence on the processing of L2 tense-aspect by L2 learners

Yu Liu

Department of Education, University of York, UK

## Abstract

The present study aims to investigate the effect of L1 on the processing of L2 tense-aspect by adult Chinese and Arabic learners of English. A visual-world eye-tracking task, a sentence-matching task, and an acceptability judgement task were used to measure the learners' L2 implicit and explicit knowledge. The preliminary results have suggested that L2 learners' performance in on-line tasks was largely influenced by their respective L1s, which suggests that late L2 learners may not be able to fully acquire L2 morphosyntactic structure which are different in their L1.

Keywords: cross-linguistic influence, L2 tense-aspect, L2 processing

## Introduction

The amount of research into the acquisition of L2 tense-aspect has been increasing over the past two decades. Many previous studies have found that some learners have persistent difficulties or exhibit little progress in their learning and usage of L2 tense-aspect even they are advanced learners. This often happens when the relevant L2 tense-aspect structures are different or absent in the learners' L1, which suggests that learners' L1 may have a strong or even permanent influence on the acquisition of L2 grammatical knowledge. Some recent studies (e.g., Roberts & Liszka, 2013) also found that even though some learners are able to display explicit and correct knowledge of L2 tense-aspect, they are not able to apply their knowledge in real-time comprehension. A great majority of previous studies have focused on L2 learners from European language backgrounds. Learners with Mandarin or Arabic L1 background are still underrepresented, and to my knowledge, these two learner groups have not been paired together so far. Considering the two languages are very different in how they encode tense and aspect, it is worth examining the influence of L1 by measuring L2 knowledge displayed by Chinese and Arabic learners.

## Methodology

### Participants

The study includes one control group of native English speakers and two experimental groups of Chinese and Arabic native speakers who learn English as a second language. All the participants were recruited at a university in

England, and they are all university students. All the L2 learners have achieved an IELTS score above 6.5 or equivalent.

### Experiment design

The present study employed two online measures and one offline measure to examine the participants' implicit and explicit knowledge. Firstly, a visual-world eye-tracking paradigm was adopted to look at the participants' anticipatory processing. Specifically, the participants are presented with a spoken sentence (e.g., *The little girl washed her hands with soap after playing.*) while viewing a visual image containing two pictures (see Figure 1). It is anticipated that if the participants are able to process the temporal information encoded in the sentence, they would look more to the left picture area (completed event area) than the right picture area (ongoing event area) after hearing the morphological form -ed. A total of 54 critical images and 54 fillers were created, and for each critical visual image, three target sentences (one simple past, one present perfect, and one present progressive) were recorded.



Figure 1. Example visual image used in the eye-tracking task.

Another sentence-matching task were used to look at participants' online sensitivity towards ungrammaticalities based on their reaction time. In this task, participants are asked to judge a pair of sentences which are sequentially presented for them and decide whether the second sentence is identical to the first one in form. The response latency (RT) for each pair is analysed by focusing on the possible differences in the mean RTs between the grammatical and ungrammatical pairs. A total of 48 critical items and 54 filler items were created. The set of 48 items comprised 16 items from each of the three grammatical constructions. Each test item contains two versions: the grammatical version and ungrammatical version. The ungrammatical version (the tense/aspect violation) was designed by having a mismatch between the temporal adverbial in the topic position and the following verb (1, 2, 3).

#### (1) Simple Past

Two weeks ago/\* For the last two weeks, Mike went to the birthday party of his best friend.

(2) Present Perfect

For the last six months/\*Six months ago, my two brothers have studied Italian at a school in Rome.

(3) Present Progressive

Right now/\*Last night Ella, is dancing to her favourite song in her bedroom.

In combination of the two implicit measures, an acceptability judgement task (AJT) was also included to measure the participants' explicit knowledge. The test items are identical with the previous two tasks, and the participants need to judge the sentences regarding their grammatical acceptability using a scale from 1 (completely unacceptable) to 7 (completely acceptable). An oxford quick placement English test was also administered to the L2 learners.

### Results

*Eye-tracking results*, the average proportion of looks were computed in every 50ms time window over a period of 1800ms following the onset of the critical verbs (see Figure 1. An example time-course graph for two groups). For the simple past items, L1 English and L1 Arabic groups showed similar preferential looks towards the completed event area at the beginning. Even though L1 Chinese group started looking more 550ms after the onset of the verb, the difference between their looks to the two areas was not significant ( $p=.091$ ). For the present perfect, both L2 learners showed delayed predication effect compared to the L1 English group, and all three groups started looking more to the ongoing event picture over time when hearing the verbs in the present progressive sentences, which was confirmed by the time cluster analysis ( $p<.001$ ).

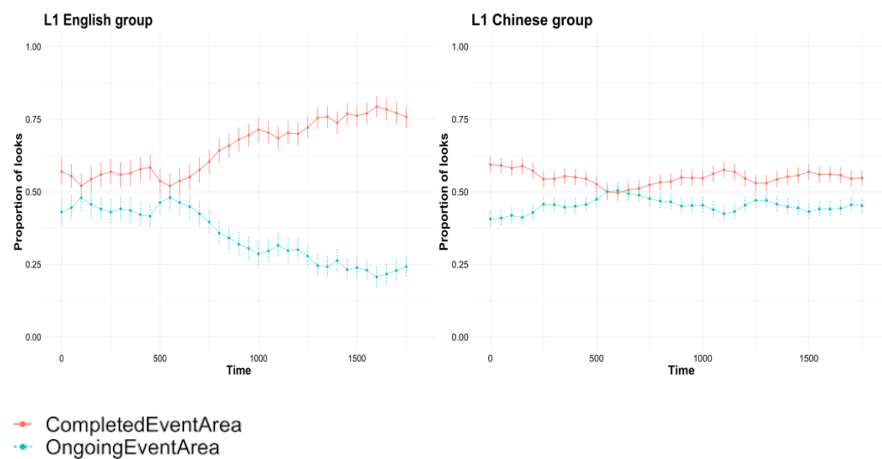


Figure 2. Mean proportion of looks to the two event areas for simple past items.

*Sentence-matching results*, for the simple past items, all three groups read the ungrammatical pairs slightly slower than the grammatical pairs, but the differences are not significant (L1 English group:  $b = 45.79$ ,  $p=0.44$ ; L1 Chinese group:  $b = 75.55$ ,  $p=0.16$ ; L1 Arabic group:  $b = 99$ ,  $p=0.09$ ). For the present perfect construction, L1 English group read the ungrammatical pairs faster than the grammatical pairs ( $b = -55.8$ ,  $p=0.32$ ), whereas the Chinese and Arabic groups read them slower, but no statistical significance were found (L1 Chinese group:  $b = 30.9$ ,  $p=0.55$ ; L1 Arabic group:  $b = 39.26$ ,  $p=0.48$ ). Interestingly, all three groups showed sensitivity towards the ungrammaticality in the present progressive sentences (L1 English group:  $b = 161.57$ ,  $p=0.007$ ; L1 Chinese group:  $b = 119.36$ ,  $p=0.03$ ; L1 Arabic group:  $b = 121.42$ ,  $p=0.03$ ).

*AJT result*, for the simple past items, all three groups rated ungrammatical sentences as less acceptable than grammatical sentences (L1 English group:  $b = -1.07$ ,  $p<0.001$ ; L1 Chinese group:  $b = -1.22$ ,  $p<0.001$ ; L1 Arabic group:  $b = -0.7$ ,  $p<0.001$ ). Similar results were found for present perfect items (L1 English:  $b = -1.46$ ,  $p<0.001$ ; L1 Chinese:  $b = -1.18$ ,  $p<0.001$ ; L1 Arabic:  $b = -0.59$ ,  $p < 0.001$ ), and for present progressive items (L1 English:  $b = -2.22$ ,  $p<0.001$ ; L1 Chinese:  $b = -2.01$ ,  $p<0.001$ ; L1 Arabic:  $b = -2.06$ ,  $p<0.001$ ). This suggests that all three groups were able to distinguish the incorrect use of tense-aspect.

## Discussion

Both Chinese and Arabic learners displayed correct explicit knowledge about English tense and aspect in the AJT task. However, unlike the L1 English and L1 Arabic groups, the L1 Chinese group did not show any predicative processing towards simple past items in the eye-tracking task which could be explained by the reason that past tense is not grammaticalized in Chinese language. Both L2 learner groups were sensitive to the ungrammaticality of present progressive in the sentence-matching task, and the usage of the progressive marker in Mandarin and the imperfective form (express progressive meaning) in Arabic might have facilitated both L2 groups to show anticipatory processing in the eye-tracking task. For the present perfect, both L2 learner groups failed to show sensitivity to grammatical violations in the sentence-matching task and displayed reduced predication effect in the eye-tracking task, which could also be the L1 influence that the perfective form in Arabic conveys both past meaning and present perfect meaning, and the perfect marker in Mandarin is still not obligatory in some cases, and adverbials can be used to achieve the perfect meaning.

## References

- Roberts, L., Liszka, S.A. 2013. Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research*, 29(4), 413–439.



# Tracking participants' behaviour when performing linguistic tasks

Maria do Carmo Lourenço-Gomes<sup>1</sup>, Cecília Castro<sup>2</sup>, Ana Paula Amorim<sup>2</sup>,  
Gitanna Brito Bezerra<sup>3</sup>

<sup>1</sup>Centre for Humanistic Studies (CEHUM), University of Minho, Portugal

<sup>2</sup>Department of Mathematics, University of Minho, Portugal

<sup>3</sup>Graduate Program in English, Federal University of Santa Catarina, Brazil

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

This study aims to identify behaviours that accompany the psychological processes involved in reading sentences and deciding on a response. The response times of different participants' actions while they completed a questionnaire were measured. We focus on two datasets from the sample: cases in which participants changed their responses and cases in which they did not. The results show that changing the response has an impact on response times. More importantly, this impact occurs not only during the decision-making process but also throughout the task. This research may contribute to a better understanding of response time data in off-line techniques and to fine-tuning experimental designs.

Keywords: Psycholinguistics; off-line task; plausibility judgement; acceptability judgement; decision-making process

## Introduction

The experimental approach has contributed to a better understanding of the cognitive mechanisms involved in language comprehension and production from on-line and off-line techniques (cf., e.g., Fernández & Cairns, 2018). However, serious concerns have been raised in the literature regarding methodological issues of the experiments, sometimes subtle but which can influence in many ways the results. Off-line data from judgments about the well-formedness of linguistic stimuli are also subject to debate on methodological issues (e.g., Langsford et al., 2019, Leivada & Westergaard, 2020, Schütze & Sprouse, 2013).

Regarding data analysis and interpretation, in the experimental approach, the association of question-answering times and the variables of interest can be tricky because time measures are influenced by multiple variables and unknown factors (Whelan, 2008). While this is true for outputs of both on-line and off-line techniques, researchers have prioritised the use of the former over the latter. On-line techniques are supposed to capture participants' automatic processes in response to a linguistic stimulus. In contrast, off-line techniques capture the participants' conscious and reflexive processes in response to it.

They are thus more likely to be influenced by extra-linguistic variables (e.g., individual differences in working memory, attention, and degree of engagement with or during the task).

In this work, we defend a more systematic attention to off-line techniques and the value of the time measures extracted through them in linguistic and psycholinguistic research. Our interest lies in participants' behaviours during task performance to identify signals regarding the degree of confidence, hesitation and engagement while deciding on a response. One of these behaviours refers to changes in responses and will be addressed in this paper.

## Method

We analysed a dataset with 7,791 observations from questionnaires applied to 163 native Brazilian Portuguese speakers (M age=27.45, SD=10.82; 118 females) with an undergraduate or a postgraduate degree in progress or completed. The dataset is part of another study carried out by the authors (Lourenço-Gomes et al., in prep.).

The materials consisted of 16 experimental sentences and 32 fillers and participants were instructed to indicate, on a 7-item Likert-type scale, the extent to which the situation described in each sentence seemed acceptable or realistic (1=not at all plausible; 7=totally plausible). Participants had to fill in a consent form and a demographic and linguistic data form to complete the questionnaire.

The questionnaires were implemented in a *JavaScript* web-based application that records (1) the responses, (2) any changes in responses, (3) how many times they are changed, and the time spent on (4) reading the stimuli, (5) marking the responses, and (6) submitting them before proceeding to the following sentence. The study is part of an ongoing Project (Lourenço-Gomes, 2019) and had the approval of the ethics committee.

## Results

The response times were categorized into five groups with probability quantiles for the statistical analyses. This approach was used as an alternative to the more commonly used data treatment since the distribution of response times is highly skewed positive, and we have chosen not to exclude any observations (considered outliers) nor proceed with a transformation in the data (cf. Whelan, 2008, for a discussion). Our focus in this study is on general participant behaviours accompanying the performance of a linguistic task. Thus, we consider all time measures to be informative. R (version 4.2.0) and IBM SPSS (version 27.0.1.0) software were used for the analyses.

With regard to the overall behaviour of the sample, the following was observed: (i) 550 cases (7.06%, of 7,991) of response changes in 142 participants (87.12%); (ii) in most cases (434, 78.91%), two responses were marked for the same item, and less frequently three (92, 16.73%) and between

four and six (24, 4.36%); (iii) the number of items with response changes in the same questionnaire varied between subjects (Min. = 1, Max. = 12, M = 3.87, SD = 2.73).

For the cases with no response change (NRC) we found a uniform distribution of response times across the categorised time groups and no significant difference was reached. In contrast, in the response change (RC) cases, a clear effect on the distribution of response times across time classes was observed throughout the task (Figure 1).

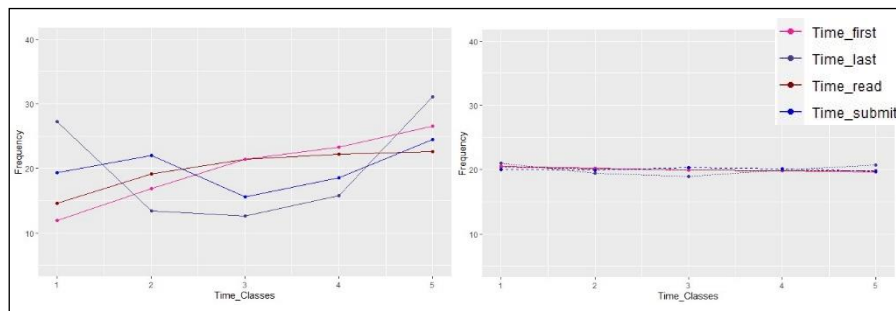


Figure 1. Graphic representation of the contrast in the data distribution across time classes in cases with response changes (RC, on the left) and with no response changes (NRC, on the right)

The results of the RC cases are as follows: The analysis showed significant differences in (i) time to read a sentence ( $\chi^2=11.302$ ,  $df=4$ ;  $p=0.023$ ). A lower percentage of observations (14.6%) were in the class with the lowest reading times; (ii) time to mark a first response ( $\chi^2=33.762$ ,  $df=4$ ;  $p<0.001$ ). A higher percentage was clustered in the highest time class (26.5%) while 11.9% were in the lower time class; (iii) time to mark the last response ( $\chi^2=53.763$ ,  $df=4$ ;  $p<0.001$ ). A high percentage of observations (31.1%) was found in the highest and lowest time classes (27.2%); and (iv) time to submit the response ( $\chi^2=11.925$ ,  $df=4$ ;  $p=0.018$ ). The distribution of observations in the most extreme time classes did not vary markedly in this parameter compared to the others (19.3%, 22%, 15.6%, 18.5% and 24.5%, from the lowest to the highest time classes, respectively).

Considering all cases (RC and NRC), a significant association was found among: (i) reading time and time to mark the last response ( $\chi^2=639.502$ ,  $df=16$ ,  $p<0.001$ ). Lower and higher reading times were associated, respectively, with lower and higher times to mark the last response; (ii) reading time and time to submit a response ( $\chi^2=478.251$ ,  $df=16$ ,  $p<0.001$ ). A high concentration of cases (51.3%) in the lowest time class was found; (iii) time to mark last response and time to submit a response ( $\chi^2=2641,102$ ,  $df=16$ ,  $p<0.001$ ), where the concentration of cases was in the lowest time classes.

### **Final remarks**

There is a justified concern among researchers with the object of study, experimental design, materials, and data treatment and method of analysis. Although assumed, participants' general behaviours that may influence outcomes seem underexplored. By tradition, they have been included in a generic category as "extra-linguistic variables". Perhaps, as a result, much is lost in interpreting the data. The work presented is part of an endeavour to pursue clues as to the degree of confidence and hesitation in participants' responses and engagement while performing linguistic tasks. Although still exploratory, the topic is hoped to motivate further reflection and inspire future work.

### **Acknowledgements**

Research at the Centre for Humanistic Studies (CEHUM) was funded by FCT-Foundation for Science and Technology (CEECIND/04331/2017). FCT has funded the research partially at CMAT (UIDB/00013/2020, UIDP/00013/2020) and the research at NIPE (UIDB/03182/2020).

### **References**

- Fernández, E.M., Cairns, H.S., Wiley, J. (Eds.). 2018. The handbook of psycholinguistics. John Wiley & Sons.
- Ferreira, F., Yang, Z. 2019. The problem of comprehension in Psycholinguistics. *Discourse Processes*, 56(7), 485-495.
- Langsford, S., Stephens, R.G., Dunn, J.C., Lewis, R.L. 2019. In search of the factors behind naïve sentence judgments: A state trace analysis of grammaticality and acceptability ratings. *Frontiers in Psychology*, 10, 2886.
- Leivada, E., Westergaard, M. 2020. Acceptable ungrammatical sentences, unacceptable grammatical sentences, and the role of the cognitive parser. *Frontiers in Psychology*.
- Schütze, C., Sprouse, J. 2013. Judgment data. In Podesva, R, Sharma, D. (eds.) *Research Methods in Linguistics*. R. Podesva and D. Sharma. Cambridge University Press), 27–50.
- Whelan, R. 2008. Effective analysis of reaction time data. *Psychological Record*, 58, 475-482.

# Testing the Endoskeletal Hypothesis in English/German bilingual code-switching

Toby S. Lowther

Faculty of Linguistics, Philology and Phonetics, University of Oxford, UK

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

In this paper, I report an experiment which provides preliminary evidence in favour of the Endoskeletal Hypothesis with respect to word order constraints on subordinate clauses in English-German bilingual sentences. The debate between exoskeletal (structure-driven) and endoskeletal (lexically-driven) models of syntax remains a contentious issue in contemporary syntactic theory, yet there has been no theory-neutral attempt to distinguish between the predictions of these two models. In this study, I utilise the unique opportunity presented by bilingual code-switched sentences to test the predictions of these two hypothesis against speaker judgements. The results of this experiment provide preliminary evidence in favour of the Endoskeletal Hypothesis (EnH).

Keywords: Syntax, bilingualism, code-switching, Endoskeletal Hypothesis

## Introduction

Within syntax, the merits of exoskeletal approaches (associating grammatical properties with structures) and endoskeletal approaches (associating grammatical properties with words) to syntax remain a contentious debate (Borer 2003). However, no prior experimental work has tested these hypotheses in bilingual contexts, which provide a unique opportunity to draw apart their empirical predictions.

In the present study, I examined these differing experimental predictions through the lens of speaker judgements regarding word order in subordinate clauses in English-German bilingual sentences. In subordinate clauses, English exhibits VO word order, while German has OV word order. Participants' preferences regarding word order can therefore tell us which set of syntactic constraints is being used in a given bilingual construction. The null hypothesis ( $H_0$ ) was that word order preference is unconditioned by linguistic context, varying by individual idiolect. I had two alternative hypotheses: that word order preference varied with the language of matrix clause and complementiser ( $H_1$ ), as predicted by functional exoskeletal approaches; or that word order preference varied with the language of the embedded verb ( $H_2$ ), as predicted by endoskeletal approaches. The results of the study indicate that the Endoskeletal Hypothesis (EnH) holds true with respect to word order constraints in subordinate clauses in English-German bilingual code-switched (CS) sentences.

## Methodology

Participants were recruited using the online recruitment service Prolific ([www.prolific.co](http://www.prolific.co)) [accessed 24.06.2021] and paid for their participation. A total of 79 participants completed the survey. Of these participants, 25 (31.6%) reported at least one parent speaking English at home, with 56 (70.9%) reporting at least one German speaking parent.

The study was composed of a sentence judgement task and a forced choice task (see Schütze and Sprouse 2013). The sentence judgement task involved rating written bilingual English-German sentences on a 5-point Likert scale ranging from 1 (a fully acceptable sentence) to 5 (a fully unacceptable sentence). In the forced choice task, participants were presented with the matrix clause, complementiser, and embedded subject of an English-German bilingual sentence, and asked to select one of two continuations or ‘neither of the above’.

A total of 94 stimuli were prepared for the sentence judgement task (10 practice sentences, 56 test sentences, and 28 filler sentences). For the forced choice task, a total of 70 stimuli were prepared (10 practice sentences, 32 test sentences, and 28 filler sentences). Stimuli were pseudo-randomly allocated to four lists per task using a Latin square method. Each participant was presented with a single list for each task, resulting in 98 responses per participant (a total of 7,742 responses, of which 1,738 were test sentences).

The experiment was scripted in HTML and JavaScript for Ibex, and hosted on IbexFarm ([spellout.net/ibexfarm](http://spellout.net/ibexfarm)) [accessed 01.07.2021, now defunct]. Tasks were completed successively. All sentences were presented visually, and participant responses as well as response times were recorded.

The data for the sentence judgement task were analysed using a linear mixed model fit by REML, modelling response score by linguistic condition, controlling for participant and test item. Interaction effects were examined with pairwise comparisons, using the Kenward-Roger degrees-of-freedom method and the Tukey method for p-value adjustment. In addition, summary statistics show significant outliers in response time, so I excluded high outliers for RT, using a  $Q3+(1.5*IQR)$  level filter as proposed by Tukey (1977).

For the forced choice task, an initial generalised linear mixed model failed to converge. I therefore used a multinomial logistic regression model fit with a neural network, modelling categorical response value by linguistic conditions. P-values were calculated with Wald tests, and this model was analysed under pairwise comparisons.

## Results

### Sentence Judgement Task

Visualisation showed a preference for VO order with English embedded verbs, and no clear word order preference with German embedded verbs.

The linear model showed a significant main effect for word order ( $t = -5.186$ ,  $d.f. = 1,014$ ,  $p < 0.0001$ ), and interaction effects for matrix clause language by word order ( $t = -5.354$ ,  $d.f. = 1,014$ ,  $p < 0.0001$ ) and embedded clause language by word order ( $t = -4.258$ ,  $d.f. = 1,014$ ,  $p < 0.0001$ ). Pairwise comparisons confirmed that participants preferred a VO order with embedded English verbs ( $p < 0.01$ ), and no significant preference for word order with a German embedded verb ( $0.345 < p < 0.834$ ). In ‘language-heavy’ contexts where both the matrix clause language and dependent language variables differed from the embedded verb, the inverse pattern was observed: an English embedded verb in a German-heavy context showed no word order preference ( $p = 0.438$ ), while a German embedded verb in a German-heavy context showed a preference for VO order ( $p < 0.0001$ ).

### Forced Choice Task

Visualisation showed a similar pattern to the previous task, except German embedded verbs seem to show a preference for OV order (see figure 1).

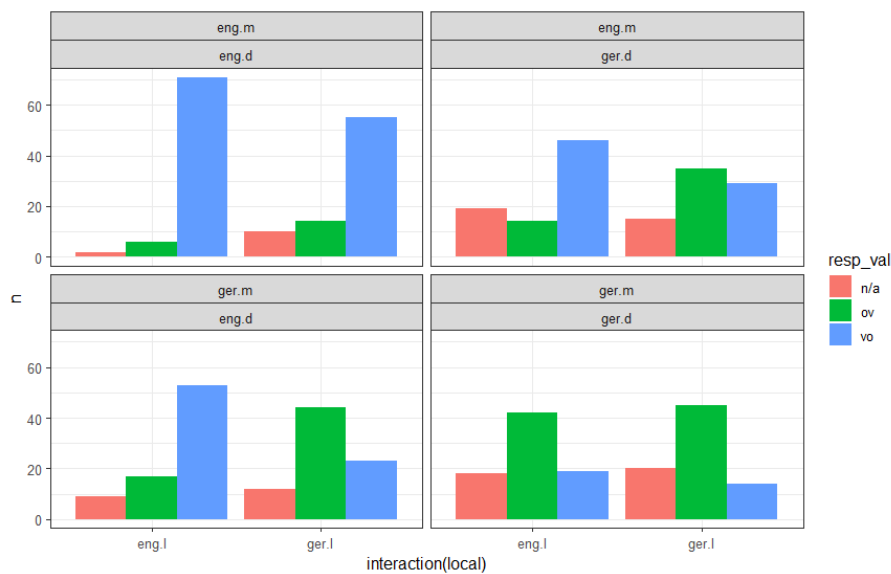


Figure 1. Bar chart for forced choice task frequency of response by condition, grouped by matrix clause language and dependent language.

Pairwise comparisons on the multinomial logistic regression model confirmed the patterns observed in figure 1. With an English embedded verb, participants preferred a VO word order ( $p$ -values ranging from 0.0084 to  $< 0.0001$ ). In the German case, the data are less clear – a clear preference for OV is observed in the German monolingual case ( $p = 0.0001$ ), while the results for cases with German embedded verb and either a German matrix clause or

German dependents are not significant ( $p = 0.509$ ,  $p = 0.988$ ). Similarly, to the sentence judgement task, 'language-heavy' contexts show the opposite pattern, with a strong preference for VO in English-heavy contexts despite a German embedded verb ( $p = 0.0005$ ) and no preference in German-heavy contexts despite an English embedded verb ( $p = 0.198$ ).

## Discussion

The results of this study suggest that verb phrase word order constraints in subordinate clauses in English-German bilingual sentences are typically determined by the embedded verb that heads the verb phrase, i.e., that the EnH holds true. The data from 'language-heavy' contexts suggest that speakers treat a single lexical item in an otherwise monolingual sentence as a case of spontaneous borrowing, rather than true CS, with predictable syntactic differences. As participant judgements appear to follow an endoskeletal pattern, and assuming an appropriate relationship between linguistic judgements and linguistic competence, these results suggest that exoskeletal theories of syntax (such as neoconstructivist and cartographic approaches) are not psychologically plausible: they model linguistic competence as acting in a way which is not replicated by the psychological behaviour of speakers in the context of the present experiment. Thus, this study provides an important contribution, bringing experimental data and quantificational methods to bear on a fundamental question of syntax through the lens of bilingualism.

## Acknowledgements

This research was funded with support from the ESRC Grand Union Doctoral Training Partnership and Balliol College, Oxford. My particular thanks to my supervisor, Dr. Louise Mycock, and to Drs. Matthew Husband and Diego Krivochen. This work would not have been possible without the critical support of Lara Scheibli, my native speaker informant in preparing stimuli.

## References

- Borer, H. 2003. Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. In Moore, J. C., Polinsky, M. (eds.) 2003, *The Nature of Explanation in Linguistic Theory*, 31-67. Stanford, CSLI.
- Schütze, C.T., Sprouse, J. 2013. Judgement data. In Podesva, R. J., Sharma, D. (eds.) 2013, *Research methods in linguistics*, 27-50. Cambridge, Cambridge University Press.
- Tukey, J.W. 1977. *Exploratory data analysis*. Boston Mass., Addison-Wesley.



# Corpus linguistic tools and terminology translation tasks

Fausto Luna<sup>1</sup>, César Aguilar<sup>2</sup>

<sup>1</sup>Doctorado en Investigación Educativa, Universidad Veracruzana, Mexico

<sup>2</sup>Instituto de Investigaciones en Educación, Universidad Veracruzana, Mexico

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Despite the progress that have been made regarding the use of corpus linguistics (CL by proposals developed by well-known authors (Aijmer, 2009; Reppen, 2010, Granger, 1998), in the Hispanic sphere there are few cases that make it, e.g. in Spain Alcina & Gamero (2002), and in Chile, Parodi (2010) to mention some. On the other hand, except for the case of the Universidad Autónoma de Baja California, in Mexico corpus linguistics are not used as an electronic resource for learning English, particularly in Language Teaching or Translation. Thus, we show here a working method applied to a translation course for undergraduate students of English Language at the Universidad Veracruzana (Mexico). We consider the use of Sketch Engine and One Click term extractor, in order to assess what kind of impact (positive, negative or null) such tools might have when used to solve translation tasks in a set of engineering and medicine texts.

Keywords: corpus linguistics, translation, terminology, teaching, language.

## Introduction

The CL is a research topic that has been gaining relevance in the most recent years in relation to its usage in language teaching. Therefore, it might be highlighted that Granger has been one of its main exponents and the Pioneer in the research field focusing on the usage of CL for foreign languages teaching. In this regard, the corpus has been distinguished as a useful tool for teaching foreign languages, in this particular case of the English teaching. Notwithstanding, this methodology, as it is defined by McEnery et al (1996), might not be limited to be exclusively employed for teaching English in a general way, but also in the translation field. This, with the aim to teach and clarify specialized lexicon, which is denominated by Cabré (2011) as terminology.

In this sense, a methodology, a strategy or a tool have not been implemented for students and professors' performance regarding teaching as well as acquiring and managing with terminology adequately in English so as to translate specialized texts. Additionally, the hypothesis of this paper states that the use of CL turns out to be a useful tool for translation students in order to have a better understanding of terminology in English and enhancing their

translation performances. Notwithstanding, it also exists the possibility of a contra hypothesis which might argue that the CL might not have any type of effect on the students' work when developing their translations.

Furthermore, it is valuable to highlight that this paper shows the quantitative data gathered by means of a questionnaire and two units of analysis in a preliminary study by means of a pilot translation whose main aim is to analyse if the use of digital tools such as SketchEngine to create a CL and the extractor of terms OneClick — through the translating processes of two abstracts from specialized areas — impacts on the students' translation performance.

### **Research methodology**

This exploratory and empirical research work is based on a quantitative method. This means that the main characteristics that stand out in this paper relies on statistics as well as graphs, the verification of the hypothesis, delimitation of approaches and measurement of phenomena. Also, this type of approach follows a deductive, sequential, and probative procedure, as well as the objective analysis of a reality. Moreover, the data collection technique employed in this research is a digital questionnaire designed in Google Forms which consists of 21 open-ended questions.

Additionally, a couple of units of analysis were necessary to gather the valuable information regarding the impact of the implementation of corpus linguistics as a tool for managing with terminology when translating.

The participants considered in this research study are 19 translation students from different semesters (sixth, eighth, and ninth) enrolled in the class *Taller de Traducción* corresponding to the sixth period of the B.A. Lengua Inglesa at the Languages School of the Universidad Veracruzana during the period February-July, 2022. The population contemplated for this study is constituted by 13 women and 6 men who are in a range of ages between 18-23. Similarly, it is worth mentioning that the informants have an upper-intermediate level of English (B2) since some of them have already concluded the class of *Inglés Avanzado* and some others are currently taking it.

### **Results**

This last section is devoted to discuss the main results regarding the impact that the implementation of corpus linguistics had in the students' translation performance.

Firstly, this work discusses about the number of accurate, incorrect and omitted terms in the informants' translation performance of the pilot translation. The table 1 and the graphs from figure 1 show that the average of accurate terms translated from the medicine abstract is 15.0 out of 17 terms from the text having 88.2% of accuracy; the average of incorrect terms translated is 1.4 out of 17 with the 8.2%; and the average of omitted terms is 0.5 out of 17 with the 2.9%. Further, the results also show the average of

accurate terms translated from the engineering abstract which is 13 out of 15 terms with the 88.6% of accuracy; the average of incorrect terms translated is 0 out of 15 with the 0%; and the average of omitted terms is 1.7 out of 15 with the 11.3%.

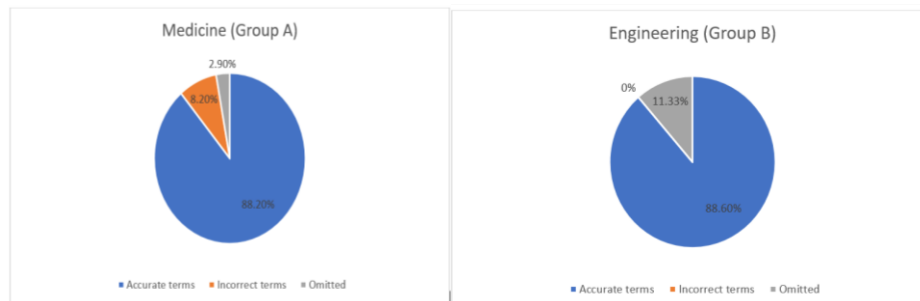


Figure 1. Graphs of the pilot translation.

Secondly, so as to identify if there was an incidence after the implementation of Sketch-Engine and One-click as supportive digital tools for translating as well as some exercises for terminology acquisition, it was important to test students' performance with a translation performance evaluation. In order to allow informants to experience different areas, the informants who carried out the translation of the medicine abstract in the pilot, worked on translating the engineering text in the evaluation. And the informants who worked with the engineering text in the pilot, worked on medicine in the evaluation.

This unit of analysis figures out the number of accurate, incorrect and omitted terms in the informants' translation performance of the evaluation. The table 2 and the graphs from figure 2 show that the average of accurate terms translated from the engineering abstract is 17.6 out of 19 terms having 92.6% of accuracy; the average of incorrect terms translated is 0.7 out of 19 with the 3.6%; and the average of omitted terms is 0.5 out of 19 with the 2.6%. Moreover, the results also show the average of accurate terms translated from the medicine abstract shows an average of 17.6 out of 18 terms with the 97.7%; the average of incorrect terms translated is 0.2 out of 18 with the 1.1%; and the average of omitted terms is 0.3 out of 18 with the 1.6%.

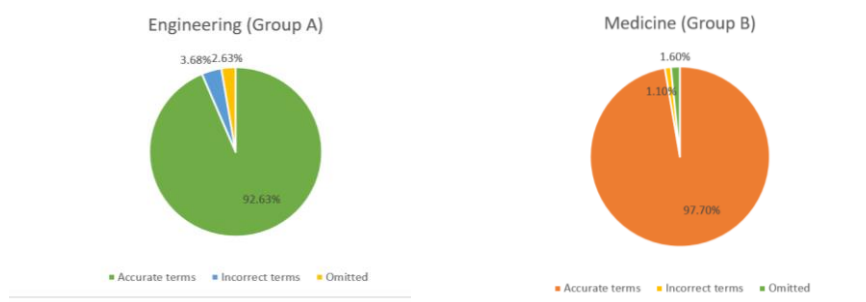


Figure 2. Graphs of the evaluation.

## Conclusion

So as to conclude, it is possible to notice a slight improvement on the students' translation performance. Based on the results, the informants' performance of the group A show an improvement with the 4.4% regarding the accurate translated terms from the pilot (88.2%) to the evaluation (92.6%); a decrease in incorrect translated terms with the 4.6% from the pilot (8.2%) to the evaluation (3.6%); and the omitted terms decrease 0.3% from the pilot (2.9%) to the evaluation (2.6%). With respect to the participants who belong to group B, the results show an improvement with the 9.1% regarding the accurate translated terms from the pilot (88.6%) to the evaluation (97.7%); surprisingly, there was a negative but slight increase in the incorrect translated terms with the 1.1% from the pilot (0%) to the evaluation (1.1%); and the omitted terms decrease 9.7% from the pilot (11.3%) to the evaluation (1.6%).

## References

- Aijmer, K. 2009. *Corpora and Language Teaching*, Amsterdam/Philadelphia, John Benjamins.
- Alcina, A., Gamero, S. 2002. *La traducción científico-técnica y la terminología en la sociedad de la información*, Castelló de la Plana, Spain, Publicacions de la Universitat Jaume I.
- Cabré, M.T. 2011. *La Terminología, una disciplina en evolución: pasado, presente y algunos elementos de futuro*. Universitat Pompeu Fabra, España.
- Granger, S. 1998. *Learner English on Computer*, London, UK, Longman.
- McEnery, T., Wilson, A. 1996. *Corpus Linguistics an Introduction*. Edinburgh University Press. Edinburgh, Second Edition.
- Parodi, G. 2010. *Academic and Professional Discourse Genres in Spanish*, Amsterdam/Philadelphia, John Benjamins.
- Reppen, R. 2010. *Using Corpora in the Language Classroom*, Cambridge, UK, Cambridge University Press.

# Topicalization out of English and Mandarin if-clauses and that-clauses

Ruihua Mao<sup>1</sup>, Anne Abeillé<sup>1</sup>, Barbara Hemforth<sup>1</sup>, Edward Gibson<sup>2</sup>

<sup>1</sup>Laboratoire de Linguistique Formelle, Université Paris Cité, France

<sup>2</sup>Department of Brain and Cognitive Sciences, MIT, USA

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

There are three main approaches to the (un)acceptability of long-distance dependencies (LDD): syntactic theories, processing theories, and discourse theories. Syntactic theories argue constraints on LDD are universal and purely syntactic, predicting these constraints should hold across languages and constructions. This study probes into the acceptability of topicalization from English and Mandarin if-clauses and that-clauses through contextualized acceptability judgment experiment, finding no adjunct island effect in either language, which supports the discourse approach, arguing that the discourse status of the extracted element in the construction is at play (focus status of the extracted element depends on the construction). If there is a focus status conflict between the element and construction, the penalty will appear.

Keywords: syntactic islands, topicalization, experimental syntax, information structure.

## Introduction

Constraints on LDD in natural languages are a prominent issue in linguistic theory since 1960s (Ross 1967), and there are three main approaches to their (un)acceptability. The syntactic approach argues they are syntactic and generalize across languages and across constructions (Huang, 1982). The processing approach claims processing factors such as frequency and high working memory load are at play given that sentences with islands may exhibit gradient acceptability and individual variation (Hofmeister & Sag 2010, Hofmeister & Sag 2010, Liu et al. 2022). The discourse approach assumes that semantic and discourse factors can explain the unacceptability of extraction from so-called islands in terms of their felicity in context. Based on the “Backgrounded Constructions are Island” principle proposed by Goldberg (2006, 2013), Abeillé et al (2020) proposed that the discourse function of the construction also plays a role: extracting an element out of a backgrounded constituent is worse when the extracted element is focalized. Extractions do not necessarily make extracted elements become focus. Specifically, the extraction elements are focalized in *wh*-questions and *it*-clefts, but not in topicalizations. This paper will focus on the so-called adjunct island constraint, comparing the acceptability of topicalization from *if*-adjuncts and *that*-complements in English and Mandarin by corresponding native speakers.

Kush et al. (2018, 2019) and Bondevik et al. (2021) observed an if-adjunct penalty in Norwegian wh-questions, but not in topicalizations. Gibson et al. (2021) and Abeillé et al. (2022) replicated the adjunct penalty in English wh-questions without context, but neither in wh-questions with context, nor in relative clauses. Myers (2012) tested Mandarin topicalizations out of if-adjuncts and because-adjuncts, finding a lack of adjunct island effect in Mandarin. In corpus studies on Danish and English, Müller & Eggers (2022) found extractions from adjunct clauses and relative clauses happen at a non-trivial rate in relativization and topicalization structure (Danish), but not in wh-questions.

## Method

We conducted contextualized acceptability judgment experiments on IBEX, testing topicalization from English and Mandarin if-clauses and that-clauses. We had a 2x2 design crossing Topicalization (extraction vs. non-extraction) and Clause-type (if-clause vs. that-clause), generating four conditions in both languages, as illustrated in (1) and (2). 16 experimental items and 80 filler items compose the English experiment. Mandarin experiment comprises 20 experimental items and 37 fillers. Participants were presented with sentence pairs and asked to rate the second sentence on a 1-7 Likert scale, followed by corresponding yes/no comprehension questions. Only data from participants with an accuracy rate above 80% were analyzed, and the effective data comprise judgment from 46 US natives (recruited through prolific), and 60 Mandarin natives living in Mainland China (recruited through social media).

### (1) English example set

Context: Paul cares about my music training.

- a) **Topic & if:** This concert, he worried [if I missed\_].
- b) **Topic & that:** This concert, he worried [that I missed\_].
- c) **Notopic & if:** He worried if I missed this concert.
- d) **Notopic & that:** He worried that I missed this concert.

### (2) Mandarin example set

Context: Zhangshan de baba xiai wenxue.  
Zhangshan GEN father be.fond.of literature

- a) **Topic & that:** zheben shu, ta feichang gaoxing ta renzhen yuedu le.  
this-CL book, he very happy she carefully read-ASP  
'This book, he was elated that she read carefully.'
- b) **Topic & if:** zheben shu, ta hui feichang gaoxing, ruguo ta renzhen yuedu.  
this-CL book, he will very happy, if she carefully read  
'This book, he will be elated if she reads carefully.'
- c) **NoTopic & that:** ta feichang gaoxing ta renzhen yuedu le zheben shu.  
he very happy she carefully read-ASP this-CL book  
'he was elated that she read this book carefully.'
- d) **NoTopic & if:** ta hui feichang gaoxing ruguo ta renzhen yuedu zheben shu.  
he will very happy if she carefully read this-CL book  
'He will be elated if she reads this book carefully.'

## Predictions

According to the traditional syntactic approach, topicalization from if-clauses should be rated worse than extraction from that-clauses because of the “adjunct island” constraint. However, if topicalizations are not extractions in Chinese (there is a null pronoun) (Xu & Langendoen, 1985), the adjunct penalty may only arise in English. Regarding processing approaches, distance-based processing theories predict no differences between topicalization from if-clauses and that-clauses, and the higher frequency of topic structures in Mandarin (Pan & Paul 2018) should favor these compared with English. Since topicalization is not a focalizing construction, the discourse approach predicts no differences between the two clause types in both languages.

## Results and discussion

We computed Bayesian models in R. The results of English (Figure 1a, 1b) and Mandarin (Figure 2a, 2b) show a high probability for a main effect of topicalization (higher acceptability in non-topicalized sentences) and no effect of clause type. However, there is a relatively higher probability for an interaction in the Mandarin experiment where the difference between extraction and non-extraction is bigger for if-clauses.

The lack of distinction between clause types challenges the traditional syntactic approach because there was no island effect in either language. Mandarin participants rated the topicalization conditions much higher than the corresponding English ones, which may be explained by easier processing due to frequency since Chinese is a topic-prominent language (Li & Thompson, 1981), using topic structures more frequently. Our results also refute the putative cross-linguistic difference between English (island sensitive) and Mandarin (island insensitive) (Zenker & Schwartz, 2017) at least for the constructions investigated here.

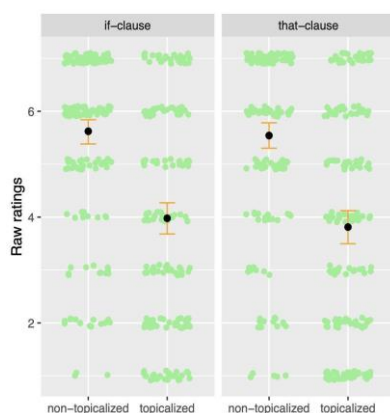


Figure 1a. English.

Probability of interaction:  $p(\beta > 0) = .60$   
 Probability of main effect topicalization:  $p(\beta < 0) = 1.00$

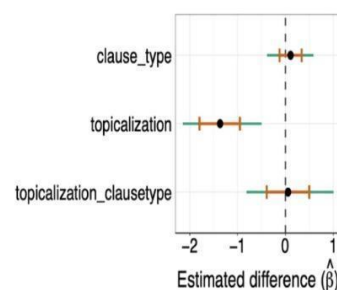


Figure 1b. Analysis.

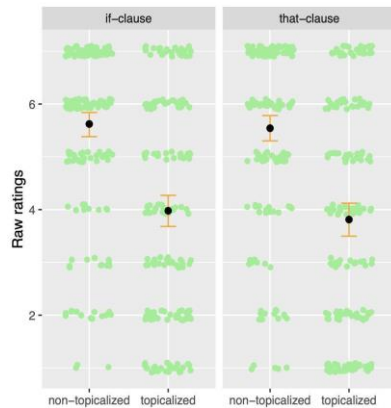


Figure 2a. Mandarin.

Probability of main effect clause type:  $p(\beta > 0) = .67$   
 Probability of main effect topicalization:  $p(\beta < 0) = 1.00$   
 Probability of interaction:  $p(\beta > 0) = .88$

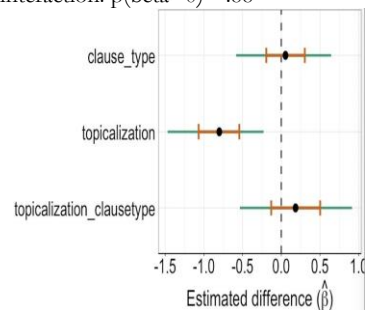


Figure 2b. Analysis.

## References

- Abeillé, A., Hemforth, B., Mao, R., Gibson, E. 2022. Acceptability of extraction out of English adjunct clauses: the role of topicalization. 35th Annual Conference on Human Sentence Processing, University of California, Santa Cruz, USA.
- Abeillé, A., Winckel, E., Hemforth, B., Gibson, E. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204.
- Bondevik, I., Kush, D., Lohndal, T. 2021. Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics*, 44(3), 223-254.
- Goldberg, A. E. 2013. Backgrounded constituents cannot be “extracted”. in Hornstein & Sprouse (eds) *Experimental syntax and island effects*, 221-238. Cambridge, UK: Cambridge University Press.
- Gibson, E, Hemforth, B., Winckel, E., Abeillé, A. 2021. Acceptability of extractions out of adjuncts depends on discourse factors. 34<sup>th</sup> CUNY conference, UPenn, USA
- Hofmeister, P., Sag, I. A. 2010. Cognitive constraints and island effects. *Language*, 86 (2), 366-415.
- Myers, J. 2012. Testing adjunct and conjunct island constraints in Chinese. *Language and Linguistics*, 13(3), 437-470.
- Müller, C., Eggers, C. U. 2022. Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English. *Languages*, 7(2), 125-149.
- Zenker, F., Schwartz, B. D. 2017. Topicalization from adjuncts in English vs. Chinese vs. Chinese-English interlanguage. 41st BUCLD Conference, 806-819, Boston University, USA.
- Xu, L., Langeonden, D. 1985. Topic Structures in Chinese. *Language* 61, 1-27.



# Brain waves and syllabic stress clash

Philippe Martin

UFRL, Université Paris Cité, France

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

The resolution process for stress clash condition, where two consecutive syllables are both (morphologically) stressed, has been well documented for lexically stressed languages such as English or Italian. However, non-lexically stressed languages such as Korean or French present more variations in their mechanism of stress clash resolution. Acoustic analysis examples taken from the spontaneous speech corpus Orfeo show that there is no stress shift of the first stressed syllable when the gap between the phonologically stressed vowels exceeds some 250 ms. This gap depends on the speech rate, a slow rate leading to maintain the morphologically stressed syllable in place, whereas a fast speech rate would reduce the gap between the stressed vowels below 250 ms and will lead to a stress shift. This 250 ms gap value is linked to the minimum interval required by delta brain oscillations to synchronize the perception of stressed syllables.

Keywords: Syllabic stress, stress clash, French, Delta brain oscillations.

## Introduction

Non-emphatic syllabic stress in lexically stressed languages such as English or Italian is assumed to be located on some syllable of lexical words (verbs, nouns, adjectives and adverbs), categories opposed to grammatical words (pronouns, conjunctions, etc.). However, non-lexically stressed languages (e.g., French, Korean) place syllabic stress on some word final syllable according to a rhythmic rule separating consecutive stressed syllables by an interval between 250 ms and 1250-1350 ms (Martin, 2018).

This property of stress in French leads to the definition of an accent phrase as a sequence of words, the last of which is stressed on its final syllable. As lexical words do not have to be necessarily stressed as in English or Italian, an accent phrase can contain more than one lexical word, as in *la ville de Stras**bour**g* or *l'air du **tem**ps*, etc.

## Stress clash

The so-called stress clash condition has been analyzed for some time, whether for lexically stressed languages like English (Lieberman and Prince, 1997), or for non-lexically, rhythmic, stressed language like French. (Garde, 1968). In both cases, the question pertains to a stress shift when two consecutive stressed syllables occur.

Two often-cited examples include an oxytonic first word, such as *thirteen* and *kangaroo* leading to realizations such as *thirteen men* and *kangaroo saddle*.

In contemporary French, the word stress is oxyton, and paroxyton if the last pronounced syllable includes the vowel [ə]. A stress clash configuration can therefore only occur if the second word involved is monosyllabic, such as *noir* or *verte* in *soulier noir* or *la ceinture verte*, with both words belonging to the same syntactic unit.

Most of the recent (Jun and Fougeron, 2002) or old (Dell, 1984) analyses either propose rules accounting for these stress shifts, or describe them as possible but not obligatory while remaining undecided on their conditions of realization. We propose here a unique explanatory principle based on the temporal properties of sequences of stressed syllable.

### Stress and brain waves

Electroencephalographic (EEG) analysis, exploring the relationship between brain waves and the perception of syllables, has shown that theta brain waves, varying in the range 4 Hz to 10 Hz, i.e., between 100 ms and 250 ms, synchronize syllabic perception in listeners (Ghitza, 2013). Without this synchronization, syllabic perception becomes less efficient, since it is triggered by unorganized theta waves on the temporal axis.

On the other hand, delta brain waves, whose periods vary between approximately 250 ms and 1250-1350 ms, have been shown to be synchronized by speech stressed syllables (Martin, 2015). They also trigger the perception of syllable as stressed even in the lack of relevant acoustic parameters of stress. In the absence of speech, the delta waves oscillate freely in their frequency range, but the occurrence of effectively stressed syllables provoke their temporal alignment. Therefore, stressed syllables ensure the regulation of accent phrases linguistic processing, necessarily taking place in the range of delta waves variation, i.e., between 250 ms and 1250-1350 ms. To satisfy this condition, the intervals between consecutive stressed syllables must therefore fall within this interval.

As a result, stressed syllables must imperatively be spaced apart by a minimum of 250 ms and a maximum of 1250-1350 ms, whether in production or in perception, even in silent reading, and independently of their linguistic function. Experimental data from spontaneous speech recordings in French presented below do support clearly this hypothesis, showing that consecutive stressed syllables remain in place while separated by a gap of more than some 250 ms.

### Experimental data

The analyzed experimental data presenting potentially stress clash conditions were extracted from the ORFEO corpus, which brings together various corpora of spontaneous speech in French. Examples were easily located thanks

to the integrated concordancer of the WinPitch software (2022). This software automatically displays the speech sound corresponding to the selected concordancer text, together with its acoustic analysis, spectrogram and melodic and intensity curves. In the examples given below, the numbers between potentially stress clashed words give the measured gap in ms, taken from the peak of the stressed syllable vowels. Parentheses indicate the syntactic group involved, and the speech rate is given in number of syllables per second.

### Gap < 250 ms

#### 1.1 First word of the sequence with 1 syllable

*et je crois que ce qu'il a voulu d- expliquer à ce moment-là c'est que c'était pas du tout du tout **noir** (et tout 200 **blanc**) mais que au contraire par exemple justement il y avait des des druides qui étaient pythagoriens 5.29 syl/s (13madmc110913)*

*il a commis l'erreur de venir me chercher une **veille** (de bac 163 **blanc**) 8.2 syl/s (07madmc110912)*

#### 1.2 First word of the sequence with 2 syllables or more with stress shift

*il y a deux ampoules blanches et une ampoule bleue ms donc (les **ampoules** 195 **blanches**) sont là pour apporter de la lumière et euh faire ressortir la couleur des euh des poissons 3.27 syl/s (bres\_m1\_08) **recul**  
mais elles prédisent toujours (**L'avenir** 229 **noir**) 6.42 syl/s (ago\_ram\_07)*

#### 1.3 First word of the sequence with 2 syllables or more unstressed

*la reine prend le linge elle le plonge dans l'eau et à peine l'a-t-elle plongé dans l'eau que le **linge** (devient 163 **blanc**) 5.04 syl/s (nataf\_041-3\_le\_roi\_des\_corbeaux)  
parce qu'il y a un **gros** (machin 226 **noir**) qui est passé c'est un corbeau 4.84 syl/s (alg\_jac\_06)*

### Gap > 250 ms

*regarde-le si tu aimes (l'**humour** 250 **noir**) regarde-le 8.62 syl/s (04gicl110912)  
et une (**ampoule** 261 **bleue**) 4.83 syl/s (bres\_m1\_08)*

*et tu as ton premier (**cheveu** 280 **blanc**) là 280 ms 5.5 syl/s (03\_mv\_cd\_100222)*

*il y a deux (**ampoules** 306 **blanches**) et une ampoule bleue ms donc les ampoules blanches sont là pour apporter de la lumière et euh faire ressortir la couleur des euh des poissons 3.27 syl/s (bres\_m1\_08)*

*genre le vieux bâti**ment** 316 **noir** 5,36 syl/s (unine08a03m)*

These data show clearly that the stress clash condition does not involve a first stress shift when the gap between stressed vowels is greater than 250 ms.

Moreover, the complexity of the consonant clusters potentially present between consecutive stressed syllables does by itself avoid a stress shift when the distance between the two syllables is below 250 ms.

We also note, at least on these samples of spontaneous speech, that stress drop in the event of a collision is not at all the rule when the first word is not monosyllabic (cases 1.2). Moreover, when a stress shift occurs, the first stress is not necessarily located on the first syllable of the first plurisyllabic word (when it exists), but could be on the final syllable of a preceding word (cases 1.3). Whether by this mechanism or by shifting the stress, the speaker creates stress groups of longer duration, so as to ensure eurythmy between successive stress groups (Martin, 2018).

## Conclusion

The minimum and maximum gap between two successive stressed syllables, linked to the range of delta brain waves periods, is the key which gives a proper account of stress clash resolution in French, and potentially for other non-tonal languages as well.

## References

- Dell, F. 1984. « L'accentuation dans les phrases en français » dans Dell, F. 1984. D. Hirst, J.-R. Vergnaud, *Forme sonore du langage*, Paris, Hermann, pp.65-122.
- Fónagy, I. 1980. L'accent français : accent probabilitaire, dans Fonagy et Léon, *L'accent en français contemporain*, Studia Phonetica, Didier: Montréal (15) 123-233.
- Garde, P. 1968, 2013. *L'Accent*, L'Accent. Paris Presses universitaires de France, 1968, 2e éd. revue et corrigée Limoges, Lambert-Lucas, 2013.
- Ghitza, O. 2013. The theta-syllable: a unit of speech information defined by cortical function, *Frontiers in psychology*, 2013/3/20, (4) 138.
- Jun, S-A., Fougeron, C. 2002. The Realizations of the Accentual Phrase in French Intonation, *Probus* 14, 147-172.
- Liberman, M. and Prince, A. 1977. On stress and linguistic rhythm, *Linguistics Inquiry* 8: 249-336.
- Martin, Ph. 2015. *The Structure of Spoken Language. Intonation in Romance*, Cambridge: Cambridge University Press, 306 p.
- Martin, Ph. 2017. Ondes cérébrales et accent syllabique, *Komunikace řeči, komunikace lidí...*, Marii Bořek-Dohalskou ed. / Praha, Karolinum 2017, 24-36.
- Martin, Ph. 2018. *Intonation, structure prosodique et ondes cérébrales*, London : ISTE.
- Orfeo, *Le Corpus d'Etude du Français Contemporain - Le projet ORFEO* [www.projet-orfeo.fr/corpus/le-corpus-d-etude-du-francais-contemporain/14-orfeo](http://www.projet-orfeo.fr/corpus/le-corpus-d-etude-du-francais-contemporain/14-orfeo)
- WinPitch 2022. Logiciel d'analyse acoustique de la parole [www.winpitch.com](http://www.winpitch.com)

# Timing of high pitch in Munster Irish

Connor McCabe

Phonetics & Speech Laboratory, Trinity College Dublin, Ireland

<https://doi.org/10.36505/ExLing-2022/>

## Abstract

A turning-point analysis of intonation in Munster Irish (Gaelic) investigated alignment between regions of increased F0 with metrically strong syllables defined by increased intensity. The data comprised recordings of story reading and retelling from 20 L1 speakers in 1928 (archived and digitised) and 14 in 2020-21, with 8487 F0 contours in total. Results show geographic and diachronic variability between speakers who achieve high pitch within +/-50ms of the strong-syllable vowel-midpoint, and those who prefer delays of 100ms or more. Greater delay characterises more conservative varieties in 1928, and this is reflected in the patterns of change between then and 2020-21.

Keywords: phonetics, prosody, turning points, bottom-up

## Introduction

Munster Irish (MI) is considered prosodically distinct from the other regional macrovarieties of Irish by virtue of phonological lexical prominence (variably termed ‘stress’ or ‘accent’) occurring outside of initial position subject to weight sensitivity. This has been hypothesised to derive historically from high pitch occurring one or more syllables after lexical prominence (Blankenhorn 1981).

As part of a broader investigation of phonetic exponents of lexical and phrasal prominence, the timing relationship between metrically strong syllables and high pitch was examined. Naturalistic data from story reading and retelling by speakers of MI from two eras were collected for analysis. The Doegen Records (Royal Irish Academy 2009) provided digitised 1928 wax-cylinder recordings of 20 L1 speakers from Counties Kerry, Clare, Cork, Waterford, and Tipperary with birth years 1846-1892. Corresponding modern data (2020-21) were collected from 14 L1 speakers from the surviving Kerry, Cork, and Waterford varieties with birth years 1932-2001.

## Background on Munster Irish prosody

Regarding intonation itself, previous work on MI comprises both British School descriptions and Autosegmental-Metrical (AM) approaches.

There is a general preference for falls and rise-falls described for MI varieties, generally given H\*+L labels in AM work. A minority of L\*+H accents have been described for Kerry and Cork, and may align with British School descriptions for Cork of high peaks systematically occurring out of phase with metrically strong syllables (Ó Cuív 1944).

Recent findings from a systematic statistical investigation in the author's PhD of prominence correlates across di- and trisyllabic words of different weight-structures, and of nonwords of controlled structure, indicate that F0- and intensity-prominence often diverge. Location of intensity-prominence generally varied as a function of weight structure, suggesting this to be a better diagnostic of lexical prominence location. Changes in F0 height were unrelated to item weight structure, consistent with proposals in the literature that in MI this is a property of phrasal prosody alone (Windsor *et al.* 2018).

Frequent occurrence of high pitch outside of metrically strong syllables is directly relevant for Blankenhorn's (1981) hypothesis of non-initial lexical prominence in MI deriving from intonational (mis)alignment. This further relates to issues of identifying the location of lexical prominence, and of the historical timeline for the purported MI system's development. For the present data, it was hypothesised that more conservative 1928 data would show more frequent late alignment of high pitch, whereas modern data with more English influence would show closer coordination between the two.

## Methods

To investigate the timing relationship between hypothesised phrase-level F0 movement and lexical-level prominence, a modified form of Ahn *et al.*'s (2021) Points, Levels and Ranges (PoLaR) intonational transcription system was used. This is based on turning points (TPs) in the F0 contour, and was selected (i) for its modular nature, and (ii) to avoid prematurely attributing phonological labels (e.g. H\*+L or L\*+H) to contours. All annotations and extractions were carried out using Praat (Boersma & Weenink 2021).

In a PoLaR transcription, the midpoint of a metrically strong syllable's vowel is labelled with <\*>, after which individual TPs in the F0 contour are identified in a separate tier. On a third tier, a local F0 range is specified in Hertz. A script is then used to automatically assign each TP a 'level' within the F0 range specified; any number of levels is possible, five – the system default – were chosen for this analysis.

Two slight modifications were made to this system to facilitate operationalisation of annotation. First, on the basis of the aforementioned findings for lexical-level prominence, increased intensity was selected as the main criterion for identifying metrically strong syllables. Second, a maximum TP-count of 6 was set for individual contours. If a described contour was found to require more than 6 TPs for adequate definition of its shape, it either required division into two parts, or one or more TPs was superfluous. Intervals of high pitch were labelled in each contour, either as single TPs for 'sharp' peaks or as a stretch between two TPs for longer plateaux; these were either two TPs of the same PoLaR level, or within one level of each other.

Measurements for the annotated contours were extracted automatically. Each contour was then assigned to one of 7 timing categories based on where

the labelled high interval began relative to the strong-syllable vowel-midpoint: ‘On-Time’ (+/-10ms), and anticipations (-) or delays (+) of magnitudes ‘Slight’ (10-50ms), ‘Moderate’ (50-100ms), or ‘Large’ (>100ms).

## Results

A total of 8487 F0 contours were transcribed and measured. The vast majority of described contours are declarative (rise-)falls, and patterning is roughly equivalent between prenuclear (non-final) and nuclear (final) position. The distribution of the contours across the seven timing categories is summarised in Tables 1-2, with colour-coding to indicate frequency of occurrence.

Table 1. Distribution of intonation contours across timing categories for the five regional varieties attested in the 1928 data (4782 contours).

		Region					Freq.
Timing		Kerry	Cork	Clare	Tipperary	Waterford	
Antic.	>100ms	3%	1%	7%	11%	8%	LEAST
	50-100ms	4%	3%	12%	8%	15%	
	10-50ms	10%	9%	23%	22%	25%	
	+/-10ms	8%	8%	14%	14%	14%	
Delay	10-50ms	15%	18%	18%	24%	19%	
	50-100ms	17%	20%	9%	15%	7%	
	>100ms	43%	40%	18%	5%	11%	MOST

Table 2. Distribution of intonation contours across timing categories for the three regional varieties attested in the 2020-21 data (3705 contours).

		Region			Freq.
Timing		Kerry	Cork	Waterford	
Antic.	>100ms	10%	18%	8%	LEAST
	50-100ms	12%	7%	14%	
	10-50ms	34%	19%	35%	
	+/-10ms	8%	7%	13%	
Delay	10-50ms	11%	8%	15%	
	50-100ms	6%	5%	6%	
	>100ms	20%	37%	9%	MOST

Two notable findings emerge. First, in the 1928 data (Table 1), there is a divide between Cork/Kerry and Waterford/Tipperary/Clare. Cork and Kerry exhibit a clear preference for large delays between strong-syllable vowels midpoints and the onset of high pitch (100ms or more). The remaining three subregions show closer coordination between high pitch and metrical strength, with most high intervals beginning within 50ms of the strong-syllable vowel midpoint. Second, the 2020-21 Cork and Kerry data (Table 2) show a change

towards the latter close-coordination tendency, while Waterford is virtually unchanged from 1928. Both Cork and Kerry retain a substantial minority of large delays, but these are now outnumbered or roughly matched by contours in the +/-50ms range.

## Discussion and conclusion

Comparison of high-pitch timing in equivalent Munster Irish data from 1928 and 2020-21 shows subgrouping of regional varieties and change between eras.

I suggest that the variation evident in these findings relates to relative conservatism in these regions and eras. In 1928, Cork and Kerry supported robust Irish-speaking communities, including a limited number of remaining monolinguals. By contrast, Waterford, Clare, and Tipperary speaker-communities were already under heavy pressure from English, with only Waterford surviving into the 21<sup>st</sup> century. Meanwhile changes in the Cork and Kerry data in 2020-21 are in the latter's direction, correlating with the geographical and numerical decline of Irish as a robust community language.

Finally, I speculate that ambiguity from conservative phrase-level high-pitch misalignment has complicated identification of lexical-level prominence-location in previous accounts of MI. This calls phonological accounts of MI prominence-assignment based on traditional dialect descriptions into question. Systematic, critical examination of the phonetic basis for phonological accounts of lexical and phrasal prosody is relevant cross-linguistically beyond Irish.

## Acknowledgements

This work was supported by a 2021-2022 Government of Ireland Postgraduate Scholarship from the Irish Research Council (GOIPG/2021/216).

## References

- Ahn, B., N. Veilleux, S. Shattuck-Hufnagel & A. Brugos. 2021. PoLaR Annotation Guidelines, Version 0.9.8. Accessed 18 October 2021. <https://www.polarlabels.com>.
- Boersma, P. & D. Weenink. 2021. Praat: doing phonetics by computer, Version 6.1.54.
- Ó Cuív, B. 1944. *The Irish of West Muskerry*, Co. Cork. Dublin: A. Thomas Ltd & Co.
- Royal Irish Academy. 2009. The Doegen Records Web Project. <https://www.doegen.ie>.
- Windsor, J.W., S. Coward & D. Flynn. 2018. "Disentangling Stress and Pitch Accent in Munster Irish". Proc. 35<sup>th</sup> West Coast Conference on Formal Linguistics, 430-437. University of Calgary, Canada.



# The neural machine translation of dislocations

Behnoosh Namdarzadeh, Nicolas Ballier

School of English Studies, CLILLAC-ARP, Université Paris Cité, France

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

This paper investigates neural machine translation (NMT) outputs for dislocated constructions from French into English. Dislocations are often considered to be “substandard in formal registers” (Lambrecht 1994: 12). In French, multiple copies of the subject are licit in spoken data, whereas translations into English preclude them (De Cat 2007). We analysed 436 translations of French dislocated segments in the novel *Voyage au bout de la nuit* (Céline 1932) and a contemporary corpus for spoken data from Corpus de Français Parlé Parisien (CFPP) (Branca-Rosoff & Lefevre 2016) by DeepL and Google. Beyond prototypical X, *c'est* dislocations, translation toolkits continue to misfire, and this might be due to the lack of spoken data in training sets of NMT.

Keywords: dislocation, language pairs French English, neural machine translation

## Introduction

This paper replicates the challenge set approach proposed by Pierre Isabelle and colleagues for English into French (Isabelle et al. 2017). The idea is to target the difficult linguistic features and observe the output of machine translation systems like Google Translate and DeepL. Our focus for our challenge set is dislocations; structures where double constituents are licit in the source text (like two grammatical subjects in French) but not in English.

Dislocations are universal (Lambrecht 1994) and all languages seem to have identical forms of topic-marking. Syntactically, two positions can be considered. One is called *theme* or *Left Dislocation* (LD), a clause and a constituent to its left. In the example “[Peter] I’ve known him for a long time” (Westbury 2016), Peter is a dislocated segment which occurs before the clause. The other is right dislocation (RD). In the example “He lived in Africa, [the wizard] (Lambrecht 1994) right dislocation has the wizard at the right edge of the sentence. Pragmatically, a linear arrangement of linguistic elements in a sentence affects information packaging. Our dislocation challenge set of French examples encompasses a pragmatic need (usually expressed by thematization) and a syntactic constraint (only one subject) for the translations into English.

## Methodology

Our corpus includes *Voyage au bout de la nuit* from the INTERSECT parallel corpus (Salkie 2022). The reference translation in English of the French sentences in which dislocation occurs was obtained using AntPConc software.

In addition to this classic subcorpus, a contemporary corpus for spoken data from the Corpus de Français Parlé Parisien (CFPP) des années 2000 (Branca-Rosoff & Lefevre 2016) was also searched using the Universal Dependency (UD) annotation to retrieve the possible dislocations with the dependency relation. Annotation was performed with the {UDpipe} package in R. The corpora yielded 2,546 occurrences, out of which we analysed 436 translations by DeepL and Google translate. We briefly report the discrepancies in the translations of dislocations observed in our corpus.

## Results

While we do not report recall for the automatic detection of dislocations with UD, the precision of our retrieval method was pretty accurate for the 218 analysed dislocations (91% and 98 %). Few false positives were detected but more frequently for the written data (appositions and parentheticals mistaken as dislocations) than for the spoken data (repairs and repetitions).

Table 1. Distribution of main dislocation types and success rate for the *c'est* dislocation in our data.

corpora	multiple	c'est	“subject copy” in the translation
Voyage (n=109)	15	62	Google = 17, DeepL = 11
CFPP (n=109)	22	73	Google = 30, DeepL = 32

The detailed typology of dislocations observed in the data is beyond the remit of this paper. We describe the complexity of the dislocations in Table 1 by reporting multiple cases of dislocations within sentences, much more frequent in our spoken data. We focus on the dominant type of dislocations (<left dislocated item>, *c'est* dislocations) and compare the two toolkits on their ability to produce translations that avoid the repetition of the subject (we call it “subject copy”). *C'est* constructions often have *ça* as a left dislocated constituent like *ça c'est vrai* (Céline 1932), which is translated as *that is true*. Despite the high number of occurrences of *c'est* construction in both of our corpora (62 occurrences out of 109 sentences for the novel and 73 occurrences out of 109 sentences for CFPP), there is still a deficiency in translating this construction by the toolkits. Google tends to produce more “subject copies” in typical examples. The picture is more blurred for more complex cases like *l'amour c'est elle la misère ...* (Céline 1932), Google outperforms DeepL and translates it as *love is misery ...* with the suppression of the extra subject which is not required in English, whereas DeepL output is closer to the ST translation suggesting *the love it is it the misery... .*

Left dislocated items can also be stacked as instances with more than two constituents for topic-marking function (Raquel 2002). For example, *in Lui, le père, je l'apercevais...*(Céline 1932), which includes double topicalization, all the initial subjects are translated by the two translation toolkits as *him, the father, I*

.... . In the other example, *moi ça m'a toujours semblé... normal...* (Branca-Rosoff, Lefevre 2016), Google keeps the structure of double topicalization as in the French source text and translates it into *me it always seemed to me... normal*, whereas, DeepL omits the extra subject and translates it into *I always thought it was... normal*.

Many dislocations have a tonic pronoun as the left dislocated item, prototypically the *moi, je* construction in French sentences. It can be tricky for the MT systems. In translating the source text *mais autrement non moi je trouve j'aime bien* (Branca-Rosoff, Lefevre 2016), Google follows word-by-word strategy translating it into *but otherwise not me I find I like*. DeepL omits this part of the source text and suggests *but otherwise I like it*.

## Discussion and conclusion

We retained the original absence of punctuation (commas) of the CFPP, which is even more distinct from the canonical training data of the toolkit. For instance, translating the French ST *le parc Mabille[,] c'est parc des Beaumonts maintenant[,] ils l'ont bien aménagé* (Branca-Rosoff & Lefevre 2016), DeepL suggests *the park Mabille it is park of Beaumonts now they arranged it well*. The zero article for *Parc de Beaumonts* may account for the absence of recognition of the pattern. We revised the transcription of CFPP and added commas where appropriate to check the ability of the MT toolkits. Re-punctuating the sentences (see our [,]) did not solve the *it* subject copy issue.

Analysing only 436 translations of dislocated constructions produced by Google and DeepL still outlines meaningful patterns for the toolkit translations of this tricky structure. The partial success with *c'est* dislocation suggests training data is crucial for the results. Les frequent structures tend to be mistranslated, especially for spoken data. Overall, the challenging dislocated segments mainly originate from spoken language and this might suggest that more spoken data should be included in the training sets of neural machine translation.

While parentheticals seem to ease the translation of dislocations, more complex structures with stacking remain an issue for NMT toolkits. The topic-marking function and some topicalized object constructions can also be challenging.

The overall patterns differ in the two toolkits. On the one hand, Google seems to keep the paratactic structure of the French source sentence, i.e., to produce a word-by-word translation of the structure and to reiterate the words. If anything, Google tends to be more source-based for the translation of dislocations and more systematically preserves the original punctuation. On the other hand, DeepL outputs are hypotactic in a sense that even for double constituents, the toolkit suppresses reiterations or links the constituents using subordinating conjunctions.

## Acknowledgements

Part of this research was financed within the SPECTRANS project, funded under the 2020 émergence research project, under the ANR grant (ANR-18-IDEX-0001, Financement IdEx Université de Paris).

## References

- Branca-Rosoff, S., Lefevre, F. 2016. Le Corpus de Français Parlé Parisien des années 2000: Constitution, outils et analyses. Le cas des interrogatives indirectes. *Corpus* 15, 265–284. <https://doi.org/10.4000/corpus.3043>
- Céline, L.-F. 1932. *Voyage au bout de la nuit*. Paris, Gallimard.
- De Cat, C. 2007. *French Dislocation: Interpretation, Syntax, Acquisition*. Oxford, Oxford University Press.
- Isabelle, P., Cherry, C., Foster, G. 2017. A Challenge Set Approach to Evaluating Machine Translation. *Association for Computational Linguistics*, 2486–2496. <https://doi.org/10.18653/v1/D17-1263>
- Lambrecht, K. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge, Cambridge University Press.
- Raquel, H. 2002. Establishing topic in conversation: A contrastive study of left-dislocation in English and Spanish. *Circle of Linguistics Applied to Communication*, 31–50, Department of Spanish Philology, Universidad Complutense de Madrid.
- Salkie, R. 2022. INTERSECT. <http://arts.brighton.ac.uk/staff/raf-salkie/intersect>
- Westbury, J. 2016. Left dislocation: A typological overview. *Stellenbosch Papers in Linguistics Plus*, 50(0), 21-45. <https://doi.org/10.5842/50-0-715>

# Lexical access to unadapted English loanwords in Croatian: evidence from translation priming

Eva Pavlinušić Vilus<sup>1</sup>, Irena Bogunović<sup>1</sup>, Bojana Čoso<sup>2</sup>

<sup>1</sup>Faculty for Maritime Studies, University of Rijeka, Croatia

<sup>2</sup>ISPP, Cambodia

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

The influence of English as *lingua franca* has resulted in an influx of unadapted English words into numerous languages, including Croatian. However, cognitive processing of such words has not been studied yet. A translation priming experiment was designed to examine the processing of 64 unadapted English loanwords in Croatian. Croatian native speakers with different levels of English proficiency participated in the study. A significant translation priming effect was observed in both language directions, while language direction and proficiency did not show a significant effect on participants' reaction times. This suggests that the most frequent unadapted English loanwords are well integrated into Croatian despite their unadapted forms and processed highly efficiently by Croatian speakers of variable English proficiency levels.

Keywords: bilingualism, lexical access, English loanwords, translation priming, lexical decision task

## Introduction

In the study of bilingual lexical access, research has generally supported the non-selective approach, at least in the early stages of word recognition. A number of models of bilingual word recognition recognize the importance of factors such as age of acquisition (AoA), language exposure and use. Exposure to English has been rapidly increasing due to its role as a global language. Many Croatian speakers are daily exposed to English, which is why it has been recognized as a part of societal as well as individual bilingualism (e.g., Mihaljević Djigunović et al., 2006). English has also become the dominant donor language for many languages, including Croatian. As a result, many unadapted English words (e.g., 'shopping', 'e-mail') have become part of Croatian informal communication. In other words, all Croatian speakers are exposed to them, regardless of their proficiency in English.

Research on cognitive processing of unadapted English loanwords is still very scarce. This study aims to contribute to this line of research by investigating lexical access to unadapted English loanwords in Croatian speakers with different levels of language proficiency.

## Methods

A translation priming experiment was designed to examine the processing of 64 unadapted English loanwords in Croatian. The words were selected based on their frequency (Bogunović et al., 2022), and the existence of adequate Croatian translation equivalents. All English words were concrete nouns. Concreteness was established using ratings in Brysbaert et al. (2013).

A total of 37 Croatian native speakers (18 female, 19 male), students from the University of Rijeka, participated in the study. To test the participants' vocabulary knowledge, Lexical test (LexTale) (Lemhöfer & Broersma, 2012) was used.

Translation priming effect was examined with a lexical decision task. The participants completed the task in both language directions (L1-L2/L2-L1), each presented in a separate experimental block. List of experimental stimuli consisted of 128 prime-target pairs. The ratio of pseudowords was .50. Identical sets of words were used for primes and targets, with four lists counterbalanced across participants. There were 16 items in each of the four experimental conditions: L1-L2/equivalent; L1-L2/non-equivalent; L2-L1/equivalent; L2-L1/non-equivalent. The stimulus onset asynchrony was set to 250 ms.

## Results

To examine the effect of translation equivalence, language direction and proficiency on reaction time, a 2 (proficiency) x 2 (equivalence) x 2 (language direction) ANOVA with repeated measures on equivalence and direction variables was used. Three participants were excluded from further analysis due to low accuracy of answers.

ANOVA showed a significant effect of equivalence on reaction time,  $F(1,31) = 9.79$ ,  $p = 0.004$ . The participants responded significantly faster when primes and targets were translation equivalents ( $M = 557.49$ ,  $SD = 19.72$ ) compared to pairs which were not translation equivalents ( $M = 652.49$ ,  $SD = 37.60$ ). At the same time, language direction did not have a significant impact on reaction time,  $F(1,31) = 1.99$ ,  $p = .168$ . In other words, L1-L2 pairs ( $M = 629.92$ ,  $SD = 36.12$ ) had similar reaction times as L2-L1 pairs ( $M = 580.06$ ,  $SD = 25.70$ ). Finally, proficiency in English did not have a significant effect on reaction time,  $F(1,28) = .77$ ,  $p = .387$ ; i.e., participants with higher proficiency ( $M = 582.26$ ,  $SD = 36.07$ ) had similar reaction time as those with lower level of proficiency ( $M = 627.73$ ,  $SD = 37.18$ ).

In addition, the interaction between proficiency, equivalence and language direction was significant,  $F(1,31) = 4.25$ ,  $p = .048$ , but mainly due to one proficiency group. No significant difference between two proficiency groups and two language directions for equivalent pairs was found ( $p > .05$ ). In the case of non-equivalent pairs, participants with lower proficiency needed more time to respond to L1-L2 pairs ( $M = 791.66$ ,  $SD = 86.81$ ) compared to L2-L1

pairs ( $M = 598.91$ ,  $SD = 47.85$ ),  $p = .004$ . No such difference was observed in participants with higher proficiency,  $p > .05$ .

## Discussion and conclusions

A significant priming effect was observed in both language directions, which is in line with previous research (e.g., Basnight-Brown & Altarriba, 2007; Schoonbaert et al., 2009; Smith et al., 2019). However, the effect of language direction was not significant, meaning that the effects in both directions were of similar strength. This is unexpected, since previous studies have generally found a smaller translation priming effect in the L2-L1 direction (Chaouch-Orozco et al., 2022; review in Wen & van Heuven, 2017), even in balanced bilinguals (Basnight-Brown & Altarriba, 2007; see Duñabeitia et al., 2010, for different results).

Lexical representations of L2 words are thought to be comparable to those of L1 words only in balanced bilinguals. However, the participants in the present study were not balanced bilinguals. This suggests that the Croatian speakers are highly familiarised with the most frequent unadapted English loanwords. These words seem to be well integrated into the mental lexicon of Croatian speakers, regardless of their proficiency in English. This assumption is further supported by the fact that the effect of proficiency on reaction time was not significant. The interaction between proficiency, equivalence and language direction showed that lower proficiency speakers responded faster to L2-L1 non-equivalent pairs compared to L1-L2 pairs. One possible explanation could be that unadapted English loanwords are used in specific contexts, which is why L2 word recognition might have been delayed by semantically unrelated L1 primes. On the other hand, the fact that no such interaction was observed in higher proficiency speakers could be due to their exposure to English words in both L1 and L2 environments. Another possible explanation for the observed asymmetry could be a more frequent usage of English loanwords in spoken than in written language in Croatian.

Taken together, the results of this study show that the processing of unadapted English loanwords might follow different patterns compared to the processing of L2 non-loanwords. However, this line of research is still relatively underdeveloped, so future research could reveal more about the specific aspects of loanword processing.

## Acknowledgements

This work has been supported by Croatian Science Foundation under the project “English words in Croatian: Identification, affective-semantic norming and investigation into cognitive processing via behavioural and neuroscientific methods” (UIP-2019-04-1576).

## References

- Basnight-Brown, D., Altarriba, J. 2007. Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition*, 35 (5), 953-965.
- Bogunović, I.; Jelčić Čolakovac, J., Borucinsky, M. 2022. The Database of English words and their Croatian equivalents. [Baza]. repository.pfri.uniri.hr/*islandora/object/pfri%3A2518*  
doi.org/10.6084/m9.figshare.20014364.v1
- Brysbaert, M., Warriner, A. B., Kuperman, V. 2013. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904-911.
- Chaouch-Orozco, A., González Alonso, J., Duñabeitia, J., Rothman, J. 2022. The elusive impact of L2 immersion on translation priming. *Studies in Second Language Acquisition*, 1-23.  
doi:10.1017/S0272263122000249
- Duñabeitia, J. A., Perea, M., Carreiras, M. 2010. Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology*, 57, 98–107.
- Lemhöfer, K., Broersma, M. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44, 325-343.
- Mihaljević Djigunović, J., Cergol, K., Qingmin, L. 2006. Utjecaj medija na nenamjerno usvajanje engleskog vokabulara. In J. Granić (Ed). *Jezik i mediji – Jedan jezik: više svjetova*, pp. 445-452, Zagreb: Hrvatsko društvo za primijenjenu lingvistiku.
- Schoonbaert, S., Duyck, W., Brysbaert, M., Hartsuiker, R.J. 2009. Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition*, 37(5), 569-586.
- Smith, Y., Walters, J., Prior, A. 2019. Target accessibility contributes to asymmetric priming in translation and cross-language semantic priming in unbalanced bilinguals. *Bilingualism: Language and Cognition*, 22, 157–176.
- Wen, Y., van Heuven, W. J. 2017. Non-cognate translation priming in masked priming lexical decision experiments: A meta-analysis. *Psychonomic Bulletin & Review*, 24, 879–886.



# Comparing pre-linguistic normalization models against US English listeners' vowel perception

Anna Persson<sup>1,2</sup>, T Florian Jaeger<sup>2</sup>

<sup>1</sup>Swedish Language and Multilingualism, Stockholm University, Sweden

<sup>2</sup>Brain and Cognitive Sciences, University of Rochester, USA

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

We investigate the role of pre-linguistic normalization in the perception of US English vowels. We train Bayesian ideal observer (IO) models on unnormalized or normalized acoustic cues to vowel identity using a phonetic database of 8 /h-VOWEL-d/ words of US English. We then compare the IOs' predictions for vowel categorization against L1 US English listeners' 8-way categorization responses for recordings of /h-VOWEL-d/ words in a web-based experiment. Results indicate that pre-linguistic normalization substantially improves the fit to human responses from 74% to 90% of best-possible performance.

Keywords: speech perception, vowel normalization, computational model

## Introduction

One of the central challenges for human speech perception is that talkers differ in pronunciation – i.e., how they map linguistic categories and meanings onto the acoustic signal. While this challenge is always present, it is most evident when listeners first encounter talkers with unfamiliar pronunciations. What mechanisms allow listeners to overcome this challenge – often rapidly, even after brief exposure – remains unclear.

One highly influential hypothesis holds that inter-talker differences are removed via low-level pre-linguistic auditory normalization of acoustic cues. There is now at least a dozen of competing normalization proposals (e.g., Lobanov, 1971; Nearey, 1978). Previous work has found that normalization reduces inter-talker variability due to, e.g. anatomical or physiological factors (e.g., Adank et al., 2004; Disner, 1980; Labov, 2010). This leaves open whether listeners actually employ normalization, and which normalization approach best explains listeners' vowel categorization. Only a relatively small number of studies has addressed these questions (e.g. Richter et al., 2017, for US English). Here, we contribute to this line of research by comparing normalization accounts against novel data on the perception of US English vowels.

## Methods

### Predicting speech perception from phonetic databases

To compare how well different normalization approaches explain listeners' vowel perception, we employ a model of Bayesian inference, ideal observers (IOs, see e.g., Kleinschmidt & Jaeger, 2015). To provide predictions about human perception, IOs need estimates of the (1) the prior probability of the vowels in the current context and (2) vowel-specific cue distributions.

Since we use the IOs to provide predictions for an 8-way forced choice categorization experiment (see below), we set (1) to a uniform prior of .125 for each of the eight vowels. We obtained (2) from a phonetically annotated database of L1 US English vowel productions (Xie & Jaeger, 2020). This assumes—as do all major theories of speech perception—that listeners acquire implicit knowledge of the category-specific distribution of phonetic cues. The database includes 1,240 recordings of eight VOWEL-d words (*heed, hid, bead, bad, odd, but, hood, who'd*, N=9 tokens per word from each of 17 female and male talkers). All words are annotated for the first three formants (F1-F3) as well as the mean fundamental frequency (F0). IOs were trained on the unnormalized or normalized F1 and F2 cues, the primary cues to US English vowel identity.

Specifically, we considered nine types of normalization. The first four transform F1 and F2 from the untransformed *acoustic* space (Hz) into one of four *perceptual* spaces hypothesized to underlie human auditory perception (Mel, Bark, ERB, and semitones). The remaining five approaches constitute normalization in a narrower sense: they center and/or standardize F1 and F2 based on their marginal distribution *across* all eight vowels (e.g., Gerstman, 1968; Lobanov, 1971; Miller, 1989; and two approaches in Nearey, 1978).

While the Xie & Jaeger database is comparatively large for a phonetically annotated corpus, it is small compared to the amount of input that human learners receive during language acquisition. To avoid over-fitting IOs to the database, we used 5-fold cross-validation: we trained five different IOs for each of the 10 different unnormalized, transformed, or normalized approaches. Each IO was trained on 80% of the recordings from each vowel of each talker in the database. The predictions of each of the 5 \* 10 IOs were then compared against human responses from a perception experiment described next.

### Vowel categorization experiment

We exposed L1 US English listeners (N=22) to the h-VOWEL-d productions of one female L1 US English talker from the Xie & Jaeger database. The experiment was administrated on Amazon Mechanical Turk and consisted of 144 trials (9 recordings per vowel \* 8 vowels \* 2 repetitions). On each trial, participants saw all 8 h-VOWEL-d words displayed on screen (order counter-balanced across participants) and then heard one of the recordings (in randomized order, grouped by repetition of the recording into two blocks). Participants were instructed to click on the word they heard the talker say.

## Results and discussion

### Human performance

Participants' responses matched the vowel intended by the talker on 71.1% of all trials. This illustrates the challenge posed by cross-talker variability and individual differences in listeners' language backgrounds: without informative exposure to the unfamiliar talker and in the absence of disambiguating context, listeners categorize recordings incorrectly in at least 1 of 4 cases! We then calculated, for each of the 72 recordings, how much listeners agreed on its categorization. On average, the most frequent response for a recording was given on 72% of all trials (out of 2 trials for each of the 22 participants). This provides an important reference against which to compare model performance: 72% recognition accuracy is what one would achieve in predicting human performance if one employs the accuracy-maximizing decision rule (criterion choice), and always categorizes recordings based on the most frequent responses given by listeners (henceforth *expected ceiling performance*).

### Model performance

The performance of the IOs was assessed by comparing their predictions for human responses, i.e., their posterior probability of inferring human categorization responses (Figure 1). We make five observations. (1) All models overall perform substantially above chance (Figure 1, left panel). (2) Transformations from the acoustic space into perceptual spaces does not improve model performance, but (3) normalization can: IOs trained on normalized cues, perform significantly better than the IO trained on unnormalized cues (53.1%, SE=0.3%,  $p < 2e-16$ ), except for Gerstman normalization (mean accuracy 50.3%, SE=0.5%). The two highest performing IOs employ Nearey's log-mean (mean accuracy 64.9%, SE=0.5%) or Lobanov normalization (mean accuracy 63.6%, SE=0.5%). The high performance of general standardizing procedures, such as Nearey and Lobanov, replicates previous findings, both from studies comparing against human responses (e.g., Richter et al., 2017) and simulated responses (e.g., Escudero & Bion, 2007). Gerstman normalization, however, still outperformed untransformed models in previous studies (unlike here). The improvements due to normalization are substantial: the unnormalized IO achieves 74% of the best possible performance (the expected ceiling performance), whereas the best performing normalization IOs achieve 90% of the best possible performance.

(4) Points 1-3 also hold for each vowel is separately (right panel). Finally, (5) no single normalization procedure outperforms all others normalization procedure on *all* vowels (right panel, Figure 1). Even for the two best-fitting IOs, there is at least one vowel for which they are not among the best models.

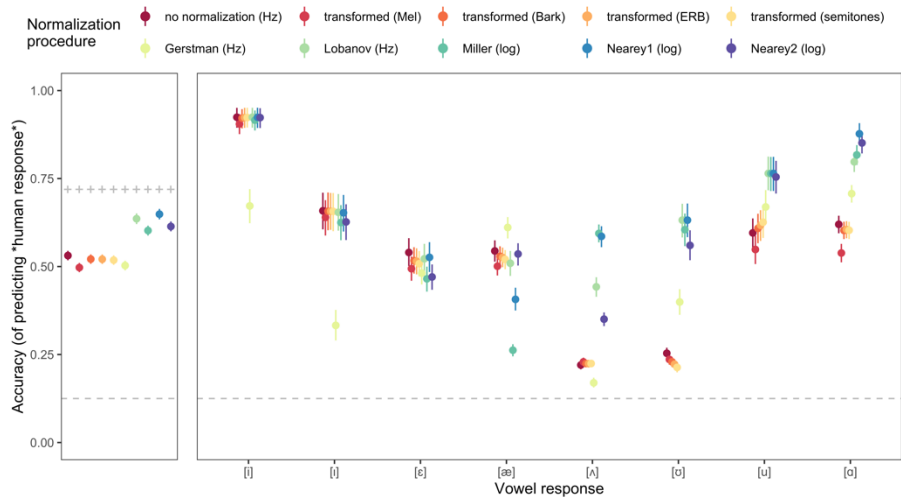


Figure 1. Prediction accuracy of 10 ideal observers for human vowel responses. **Left:** Overall accuracy across vowels. Plus line indicates expected ceiling performance (see text). **Right:** per-vowel accuracy. Dots indicate mean accuracy across the five folds. Intervals show average bootstrapped 95% confidence intervals across the five folds, thus indicating uncertainty about model’s accuracy in predicting human performance. Grey line indicates chance.

Our results suggest that pre-linguistic normalization (or computationally similar algorithms) provide a plausible explanation for the remarkable adaptive abilities of human speech perception. We find that models based on normalized F1 and F2 cues can achieve up to 90% of the achievable accuracy. Future work should determine whether the remaining 10% can be achieved by adding additional cues (e.g., F3 or vowel duration), or whether they point to additional mechanisms (e.g., representational changes or changes in decision-making, Xie, Jaeger, & Kurumada, 2022). Similarly, it is possible that the effects we observed for normalization could be accounted for by alternative mechanisms (ibid).

## References

- Escudero, P., Bion, R.A.H. 2007. Modeling vowel normalization and sound perception as sequential processes. *ICPhS* 16. 1413-16.
- Richter, C., Feldman, N.H., Salgado, H., Jansen, A. 2017. Evaluating low-level speech features against human perceptual data. *ACL* 5, 425-40.
- Xie, X., Jaeger, T.F. 2020. Comparing non-native and native speech: Are L2 productions more variable? *The Journal of the Acoustical Society of America* 147. 3322-47.

# The effect of sound quality on attention and load in language tasks

Kilian Seeber, Dongpeng Pan

Faculty of Translation and Interpreting, University of Geneva, Switzerland

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

In this paper, we report the first part of the results of a study on the effect of sound quality on attention and load during different language processing tasks. Forty-two professional conference interpreters completed three tasks (comprehension, production, and simultaneous interpreting of sentences) in high-quality and low-quality audio conditions. The first sentence (of the type, “The man/woman is at the *location*”) was accompanied by a visual array with three images: the smallest object (Target X), the bigger object (Foil) and the biggest object (Target Y). Gaze patterns and pupil diameter were measured during the second sentence (of the type, “The *Target X* is next to the *Target Y*”) in a blank-screen paradigm. There were 12 visual arrays for each task. We found that participants shifted their attention to the mentioned target earlier in adverse sound conditions in both the comprehension and the interpreting task, and cognitive load is higher during interpreting than it is during comprehension.

Keywords: visual world paradigm, attentional shift, cognitive load, language comprehension, simultaneous interpreting.

## Introduction

This study aims to investigate the effect of sound quality on attention and cognitive load in the comprehension, production, and simultaneous interpreting of sentences by analysing participants’ eye movements towards sequentially occurring targets. To that end, we designed a blank screen visual-world eye-tracking experiment to measure attentional shift patterns. Pupil diameter was used as an indicator of real-time cognitive load. In this paper, we report and briefly discuss results for sentence comprehension and simultaneous interpreting.

## Methodology

Forty-two Geneva-based professional conference interpreters participated in the study. All interpreters were L1 speakers of either French, German, Spanish, Italian or Russian and had English as an L2 language.

Thirty-six experimental visual arrays were created; 12 for each experimental task, i.e., comprehension, production, and simultaneous interpretation. Each array contained three images representing objects of different sizes: the smallest

object (Target X), the bigger object (Foil) and the biggest object (Target Y), e.g., a shrimp, a jellyfish, and a whale.

For the comprehension and the interpreting task, participants heard an introductory sentence and a critical sentence. The introductory sentence provided context (e.g., “The man is at the aquarium”). The critical sentence described the location of the two targets in relation to each other (e.g., “the shrimp is next to the whale). In the comprehension task, participants were asked to look at the screen and listen to the sentences. In the interpreting task, participants were asked to look at the screen and simultaneously interpret all sentences into their respective L1.

The experiment was based on the blank screen paradigm (Altmann 2004), where images of objects are not visible during the critical part of the sentence. Consequently, participants only saw the visual stimuli during the first, introductory sentence, while during the critical sentence the screen was blank. The audio stimuli for half of the trials had a frequency range of 125Hz to 15KHz (i.e., high quality: HQ), the other half of 300Hz to 3400Hz (i.e., low quality: LQ). Tasks were blocked and randomly rotated.

To illustrate participants’ shifts of visual attention among objects over time, bootstrapping was used. Resampled datasets were created, and a one-sample t-test on fixation proportions was conducted after each resample, without assumptions about the population distribution (Stone et al. 2021). A divergence point represents the first time point followed by at least ten consecutive time points with significant t-values. Then, we used a non-parametric bootstrap to create new datasets by resampling participant, timepoint, and object type (Target X vs Target Y, Target X vs Foil, and Target Y vs Foil). The bootstrap comprised 2000 iterations.

Pupil diameter was used as an indicator of real-time variations in cognitive load during sentence comprehension (Just and Carpenter 1993) and simultaneous interpreting (Seeber 2013).

## Results

Bootstrapped means and CIs for HQ and LQ of the comprehension task are plotted in Figure 1. In the HQ condition, during the PePOI “the X”, participants tend to look at Target X, the mentioned object, more than at Target Y, the significant divergence beginning 933.78 [820, 990] ms after sentence onset. Similarly, during the POI “the Y”, participants look at Target Y, the mentioned object, more than Target X, albeit with a very late divergence onset at 3131.516 [2644.57, 3554.5]. In the LQ condition, during the POI “the X” participants’ preference of looking at Target X, the mentioned object, over Target Y, is no longer significant and merely a trend. During the POI “the Y”, participants look at Target Y, the mentioned object, more than at Target X, with a significant divergence onset at 2555.244 [2494.57, 2574.57]. Participants’

shift from Target X to Target Y occurs about 300ms earlier during the LQ condition.

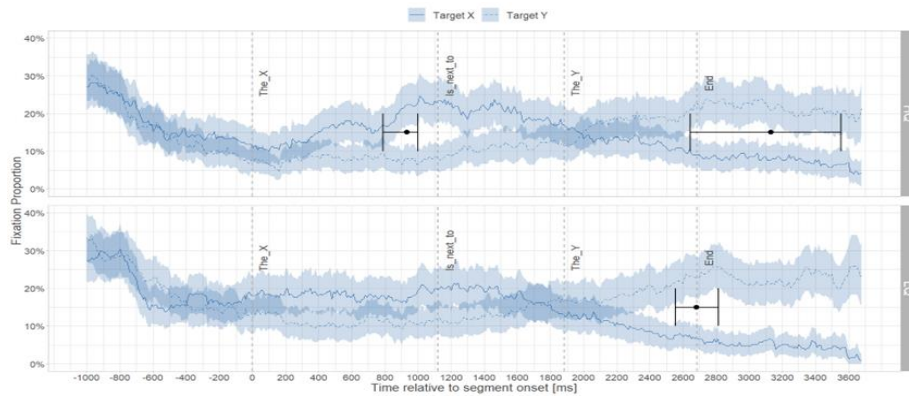


Figure 1 Divergence points and 95% confidence intervals superimposed on the fixation curves, depicting fixation to Target X and target Y of the comprehension task.

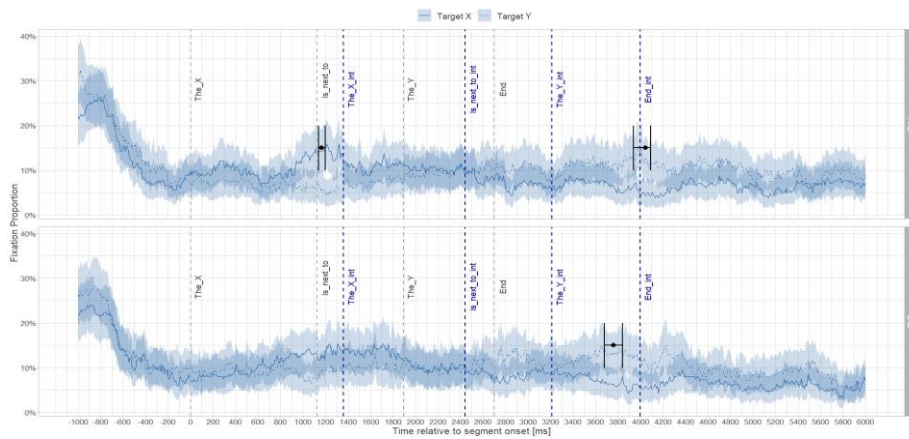


Figure 2 Divergence points and 95% confidence intervals superimposed on the fixation curves, depicting fixation to Target X and target Y of the interpreting task.

Bootstrapped means and CIs for HQ and LQ of the interpreting task are plotted in Figure 2. In the HQ condition, during the POI “the X”, participants tend to look at Target X, the mentioned object, more than at Target Y, although the first significant divergence occurs at 1165.071[1136.5, 1196.5] ms from sentence onset, and thus during the POI “is next to”. During the POI “the Y” participants tend to look at Target Y, the mentioned target, more than Target X, although the significant divergence onset does not start before 4045.1

[3940.1, 4090.1], the thus after the POI “the *Yinterpretation*”. In the LQ condition, the only significant divergence between looks at Target X and Target Y starts at 3760.1[3680.1, 3840.1], during the POI “the *Yinterpretation*”, with participants looking at Target Y more than at Target X. Like in the comprehension task, visual attention to Target Y shifted earlier in the LQ condition.

The analysis of participants’ pupil size across tasks and sound quality conditions shows significantly larger pupil diameter during POI “Target Y” as compared to “Target Y” in both tasks. During comprehension, LQ sound did not significantly affect pupil diameter. During interpreting, pupil size was significantly larger than during comprehension, and LQ sound engendered significantly larger pupil diameter than HQ sound.

## Conclusion

Sound quality modulates the shift of visual attention (measured as proportion of fixations) to mentioned objects during comprehension and simultaneous interpreting, suggesting a quantitative and a temporal effect. On the one hand, low sound seems to decrease the likelihood of shifting visual attention to the first object mentioned. On the other hand, it appears to increase the speed with which visual attention is allocated to the second object mentioned.

Cognitive load (measured as pupil diameter) is significantly higher during interpreting than it is during comprehension. Also, while the low-quality sound used in this experiment did not modulate load during comprehension, it generated significantly higher load during simultaneous interpreting. Regardless of the task, load significantly increases towards the end of the sentence.

## References

- Altmann, Gerry T.M. 2004. Language-mediated eye movements in the absence of a visual world: the ‘blank screen paradigm’. In *Cognition* 93 (2), B79-B87. DOI: 10.1016/j.cognition.2004.02.005.
- Just, M.A., Carpenter, P.A. 1993. The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 47(2), 310–339. DOI : 10.1037/h0078820.
- Seeber, K.G. 2013. Cognitive load in simultaneous interpreting: Measures and methods. In *Target* 15 (1), 18-32. DOI: 10.1075/target.25.1.03see.
- Stone, K., Lago, S., Schad, J. 2021. Divergence point analyses of visual world data: applications to bilingual research. In *Bilingualism* 24 (5), pp. 833–841. DOI: 10.1017/S1366728920000607.



# Does the number morphology determine the agreement strategy?

Kseniia Studenikina

Department of Theoretical and Applied Linguistics, Lomonosov Moscow State University, Russian Federation  
<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Parameterizing agreement variation requires the analysis of multiple language factors. This paper investigates the role of number morphology in determining agreement strategy. In present study, I consider the variation in Russian noun phrases with coordinated modifiers, also known as nominal right node raising constructions (NRNR). Previous studies pointed out semantic, discourse and syntactic factors parameterizing the agreement variation. Using experimental study, I show that the type of noun number inflection imposes certain restrictions on the possible agreement strategy in Russian NRNR. I argue that considering noun morphology enables to opt for the theoretical approach explaining variation in this construction.

Keywords: number morphology, agreement, variation, coordination, right node raising

## Introduction

In order to parameterize agreement variation in a certain construction, it is necessary to list the variety of factors which can be identified as predictors for choosing one or another agreeing option. The present paper deals with the influence of morphological factors on the agreement strategy. The study investigates variation in Russian noun phrases with coordinated modifiers, also known as nominal right node raising constructions (NRNR, Shen 2018).

## Nominal right node raising constructions

Russian nominal right node raising constructions are noted to demonstrate number agreement variation: both distributive strategy (singular noun) and summative strategy (plural noun) are acceptable (1). Previously semantic, discourse and syntactic factors parameterizing the agreement variation were considered (Kodzasov 1987). However, the role of noun morphology in Russian has not been studied before.

- |              |     |           |                         |
|--------------|-----|-----------|-------------------------|
| (1) vysok-ij | i   | nizk-ij   | student/ student-y      |
| tall-SG      | and | short-SG  | student.SG / student-PL |
| 'the tall    | and | the short | student'                |

The correlation between the morphology and the agreement was described for Bulgarian (Harizanov & Griбанова 2015). The regular number morphology of the noun enables only summative strategy (2) while the irregular number morphology enables only distributive agreement (3).

(2) bālgarsk-i-ja		i	rusk-i	narod-i/ * narod
Bulgarian-SG.M-the	and		Russian-SG.M	nation-PL/ nation.SG
'the Bulgarian	and		Russian	nations'

(3) naj-nisk-o-to		i	naj-visok-o	dete/ * deca
most-short-SG.N-the	and		most-tall-SG.N	child/children
'the shortest	and		tallest	child'

The syntactic analysis proposes an ATB-movement, number feature is assumed to be privative. The difference between regular and irregular morphology is explained by early insertion of lexical roots and late insertion of inflection morphemes (Embick & Halle 2005). The root is inserted before syntactic movements, but the plural feature arises in syntactic derivation. Therefore, contextual allomorphy is impossible.

### Experimental study

In order to examine the relation between the number morphology and the agreement in Russian, I conducted a self-paced acceptability experiment (Likert scale 1-7). It tested 2 factors: the noun number (singular/ plural) and the morphological noun type: suppletion and stem alternations (4), suffix alternations (5), syncretic forms (6), namely *singulare tantum* (6a) and *plurale tantum* (6b), regular nouns (7). Statistical analysis was conducted with linear mixed models.

(4) vysok-ij	i	nizk-ij	čelovek / ljudi
tall-SG	and	short-SG	man / people
'the tall	and	the short	man'
(5) bur-yj	i	bel-yj	medvež-onok / medvež-ata
brown-SG	and	white-SG	bear-SG / bear-PL
'the brown	and	the white	bear'
(6) a. morsk-aya	i	sukhoputn-aya	pekhota
naval-SG	and	ground-SG	infantry
'the naval	and	the ground	infantry'
b. elektronn-ye	i	mekhaničesk-ie	vesy
electronic-PL	and	mechanical-PL	scales
'the electronic	and	the mechanical	scales'

- (7) požil-oj i molod-oj prepodavatel' / prepodavatel-i  
 elderly-SG and young-SG teacher.SG / teacher-PL  
 'the elderly and the young teacher'

The results of the experiment are as follows. The syncretic nouns receive the highest scores and show no delay in the reading task ( $p=0.5992$ ). The nouns with stem alternations demonstrate the preference of singular number. The plural forms receive significantly lower scores ( $p<.0001$ ) and cause reading delays ( $p=0.0039$ ). There is no significant difference between acceptability of singular and plural forms for nouns with suffix alternations ( $p=0.2131$ ). Plural forms with regular morphology receive significantly higher scores ( $p=0.0085$ ), but singular regular forms are still more acceptable than ungrammatical fillers.

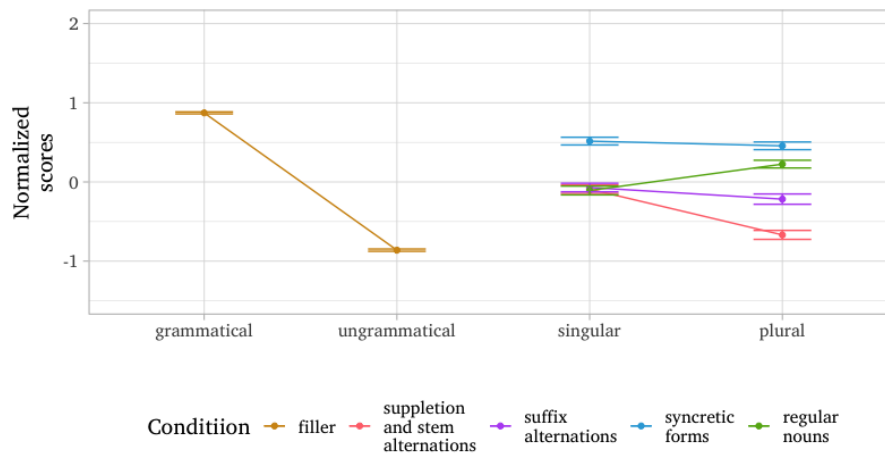


Figure 1. The interaction plot for acceptability judgments

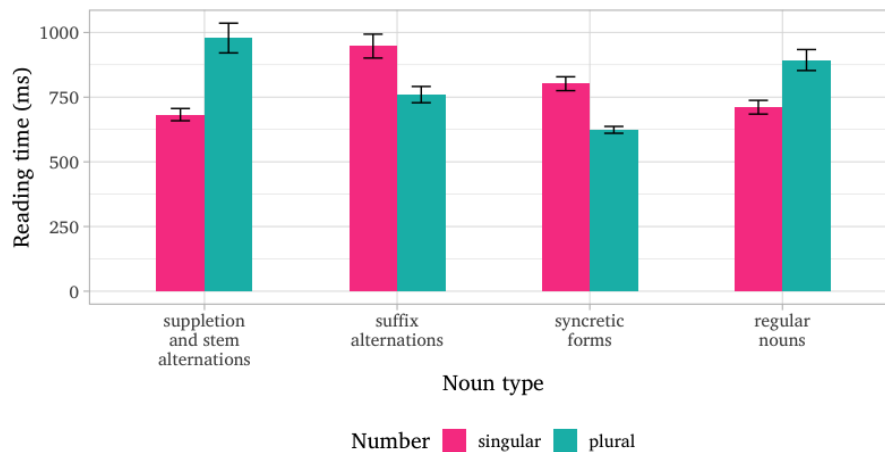


Figure 2. The barplot for reading time of a noun

Thus, the analysis with ATB-movement explains the preference of singular forms with stem alternations. However, the acceptability of both singular and plural forms with suffix alternations and with regular morphology is impossible if the number feature is privative. I argue that multidominance analysis should be applied (Shen 2018). According to this approach, the number feature is equipollent, which explains the acceptability of singular and plural forms with suffix alternations and with regular morphology. Since the multidominant structure is derived after the insertion of lexical roots, the contextual allomorphy becomes impossible on this stage. Thus, it explains the unacceptability of plural forms with suppletive morphology and stem alternations.

### **Acknowledgements**

This research is supported by Russian Science Foundation, RSF project 22-18-00037 realized at Lomonosov Moscow State University, <https://rscf.ru/en/project/22-18-00037/>

### **References**

- Embick, D., Halle, M. 2005. On the status of stems in morphological theory. In Geerts, T., van Ginneken, I., Jacobs, H. (eds.) *Romance Languages and Linguistic Theory 2003: Selected papers from 'Going Romance' 2003*, Nijmegen, 20-22 November, 37–62. Amsterdam: John Benjamins.
- Harizanov, B., Gribanova, V. 2015. How across-the-board movement interacts with nominal concord in Bulgarian. In *Proceedings from the Annual Meeting of the Chicago Linguistics Society 49*, University of Chicago, IL. Chicago Linguistics Society.
- Kodzasov, S. 1987. Chislo v sochinitel'nyx konstrukcijax [Number in coordinated structures]. In Kodzasov, S., Laufer, N., Savina, E. (eds.) *Modelirovanije jazykovej dejatel'nosti v intellektual'nyx sistemax*, 204-219, Moscow, Russia.
- Shen, Z. 2018. Feature arithmetic in the nominal domain. PhD thesis, University of Connecticut, Storrs, CT.

# DRESS-raising in Connemara English: shedding light on the role of Irish

Kate Tallon

Phonetics and Speech Laboratory, Trinity College Dublin, Ireland

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Pre-nasal DRESS-raising (/ε/ to /ɪ/) is a salient feature of western Irish English. However, its origins are little understood. Accounts differ on whether it is an Irish transfer feature or an Early Modern English retention. To investigate the possibility of this being a synchronic transfer feature from Irish 448 DRESS tokens recorded from bilingual L1 Irish speakers were acoustically analysed and compared with 450 DRESS tokens from monolingual English speakers from the same region. L1 Irish speakers of English showed a higher degree of pre-nasal raising than monolingual English speakers. However, an equivalent analysis of the bilingual speakers' Irish (113 tokens) shows no evidence of pre-nasal raising. These findings cast doubt on the possibility of an Irish origin for DRESS-raising in Irish English.

Keywords: Irish English, DRESS-raising, PEN-PIN merger, contact Irish English

## Introduction

Phonological accounts of western Irish English describe raising of the vowel in the DRESS lexical set (Kallen 2012, Peters 2016). Historically occurring in a wide variety of phonological contexts, it is now largely a pre-nasal phenomenon, hence its being termed the PEN-PIN merger. Peters' 2016 study of Galway City English found this merger to be a recessive feature, occurring mainly in the speech of the older population.

The origins of DRESS-raising in Irish English are subject to speculation. Kallen (2012) and Henry (1957, p. 76) have drawn parallels with Early Modern English (brought to Ireland during the 17<sup>th</sup> century), which exhibited /e/-raising before /s/, /l/ or /n/ + consonant. Others (Ó Baoill 1997, O'Sullivan 1942 as cited in Peters 2016) have attributed DRESS-raising to phonological transfer from Irish, which is said to demonstrate raising of /ε/ to /ɪ/ in nasal environments in Connacht and Munster (Ó Siadhail 1989). Ó Baoill (1997) draws parallels with Irish mid-vowel raising in both Connacht and Munster.

Varieties of Irish English spoken by L1 Irish speakers provide a setting for investigating the extent of phonological transfer from Irish to English. Such varieties are reported to exhibit particularly extensive Irish influence (Ó hÚrdail 1997) and, as put by Harris, (1991, p. 45) are an "obvious research site for testing assumptions about the influence of Irish on the development of Irish English".

To shed light on the plausibility of an Irish source for Irish English DRESS-raising, this study addresses the following research question: Do L1 Irish speakers of English have a particularly extensive presence of dress-raising when compared to monolingual English speakers from the same region?

## Methodology

### Data collection

Two sets of participants from the Connemara Gaeltacht (Irish-speaking region) in the west of Ireland were recorded. Group 1 consisted of ten L1 Irish speakers (Irish-English bilinguals); Group 2 consisted of ten monolingual English speakers. Each group comprised five male and five female speakers, giving a total of twenty participants. Participants were recorded individually in their own homes. Two English datasets were recorded from each participant: a short interview and a series of sentences. This resulted in a total of 448 DRESS tokens for bilingual speakers (191 interview tokens, 257 sentence tokens) and 450 DRESS tokens for monolingual speakers (198 interview, 252 sentences). Additionally, an Irish dataset (interview and sentences) was collected from the bilingual speakers, giving 113 /ɛ/ tokens. Recordings were made using an AT2020 USB microphone and an Apple Macbook Pro.

### Data analysis

F1 and F2 measurements were extracted for all vowel tokens via Praat. To facilitate investigation of a PEN-PIN merger 753 (total) KIT tokens and 144 tokens of /ɪ/ in Irish were also measured. These measurements were Z-score normalised in R (R Core Team 2022) to facilitate interspeaker comparison.

The ggplot function in R was used to visualise the data. DRESS tokens were subdivided into pre-oral and pre-nasal (PEN) contexts to investigate pre-nasal raising. To quantify the degree of raising for each group of speakers, Pillai scores were calculated via MANOVA tests in R as a measure of overlap between vowel groups.

## Results

Figure 1 demonstrates that L1 Irish speakers showed greater overlap of PEN and KIT vowels than monolingual speakers. The degree of vowel overlap for each speaker is represented by Pillai scores given in Table 1. These results confirm that bilingual speakers have a larger degree of overlap between PEN and KIT vowels than monolingual speakers. In addition, monolingual speakers DRESS and PEN vowels show little separation when compared to those of bilingual speakers.

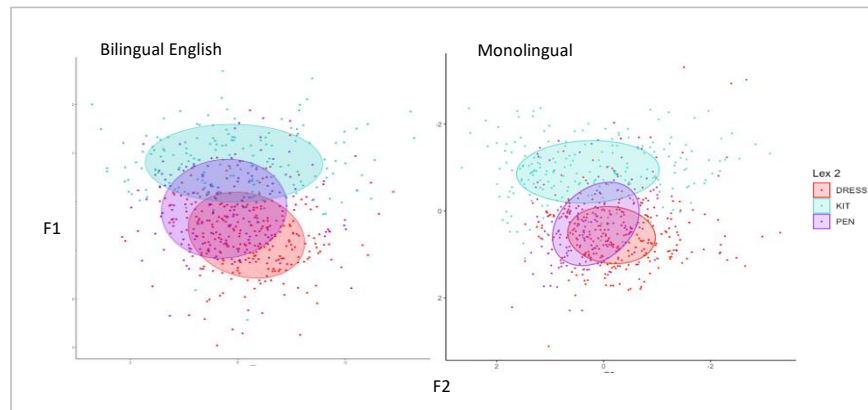


Figure 1. Scatterplot of DRESS, KIT AND PEN vowels for bilingual (L1 Irish) English and monolingual English.

Table 1. Summary of Pillai scores for vowel groups of each dataset.

	Monolingual		Bilingual		Irish	
	kit	pen	kit	pen	/ɪ /	/ɛ/
pen	0.32		0.16		0.24	
dress	0.5	0.06	0.4	0.15	0.48	0.04

Irish /ɛ/ tokens were similarly divided into pre-nasal and pre-oral contexts and were plotted with Irish /ɪ/ tokens for contextualisation. Figure 2 shows that the bilingual speakers' production of Irish /ɛ/ vowels did not mirror that of their English, i.e. no evidence of expected /ɛ/ raising was present in the Irish tokens. In addition, the Pillai scores for Irish were closer to those for monolingual English than for bilingual English.

## Discussion and conclusion

L1 Irish speakers showed a greater degree of pre-nasal DRESS-raising than monolingual English speakers from the same region. The role of L1 Irish transfer in such a discrepancy is naturally considered. However, bilingual speakers did not show equivalent raising in their Irish (Figure 2). Thus, the PEN-PIN merger in the bilingual speakers' English does not appear to be a

synchronic phonological transfer from Irish. This finding casts doubt on an Irish language source for Irish English DRESS-raising/PEN-PIN merger.

Although preliminary, these findings highlight that an analysis of Irish data alongside Irish English is necessary when investigating origins of particular phonological features of Irish English; anecdotal claims should be supported by quantitative analyses.

## References

- Harris, J. 1991. Ireland. In J. Cheshire (Ed.), *English around the world: Sociolinguistic perspectives*: Cambridge University Press.
- Henry, P. 1957. An Anglo-Irish dialect of North Roscommon. Phonology, accent, syntax/Patrick Leo Henry. Dublin: Department of English, University College. 1957.– 193 p.
- Kallen, J.L. 2012. The English Language in Ireland: An Introduction. *International Journal of Language, Translation and Intercultural Communication*, 1, 25-41.
- Ó hÚrdail, R. 1997. Hiberno-English: Historical Background and Synchronic Features and Variation. In H. Tristram (Ed.), *The Celtic Englishes* (pp. 180-199).
- Ó Siadhail, M. 1989. *Modern Irish: Grammatical structure and dialectal variation*: Cambridge University Press.
- Peters, A. 2016. *Linguistic change in Galway City English* / Arne Peters.



# Listeners' categorisation behaviour correlates with gradient changes in exposure statistics

Maryann Tan<sup>1,2</sup>, Rachel Sabatello<sup>2</sup>, Iva Savic<sup>2</sup>, T. Florian Jaeger<sup>2</sup>

<sup>1</sup>Centre for Research on Bilingualism, Stockholm University, Sweden

<sup>2</sup>Brain & Cognitive Sciences, University of Rochester, US

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Listeners can understand talkers despite cross-talker variability in the mapping from phonetic cues to linguistic categories. The mechanisms that underlie this adaptive ability are not well understood. We test to what extent listeners can adapt their interpretation of speech based on the distribution of phonetic cues in the recent input, and whether prior expectations about how talkers typically sound guide and constrain this process.

Keywords: speech perception, speech adaptation, distributional learning

## Introduction

Spoken language is highly variable – a given talker's /d/ in “din” may sound more like another's /t/ in “tin”. These differences arise from multiple sources including physiological, linguistic, and extra-linguistic factors. Despite the absence of invariant acoustic cues to a talker's intended category, listeners usually comprehend talkers with apparent ease. When faced with unfamiliar accents, listeners often adapt with little exposure (e.g., Bradlow, Bent 2008).

The mechanisms underlying such adaptation are not yet fully understood. One hypothesis holds that as listeners encounter a talker, they incrementally learn the statistics of that talker's input and integrate it with their prior expectations of how talkers should sound (cf. ideal adapter, Kleinschmidt, Jaeger 2015). We exposed US American English (AE) listeners to recordings of AE that was phonetically manipulated between participants. All listeners heard word recordings starting with /d/ or /t/ (e.g., “dill” or “till”). Recordings varied in the primary phonetic cue to the /d/-/t/ contrast (voice onset time, VOT). Between participants, an initial exposure phase shifted the VOT distributions for /d/ and /t/ by +0, +10, or +40msecs. We assessed the consequences of those shifts during subsequent test phases that were identical across all participants.

## Methods

Our approach closely follows previous work (Clayards et al. 2008, Kleinschmidt, Jaeger 2016, Theodore, Monto 2019) but extends these paradigms in ways intended to increase the ecological validity of the stimuli and

exposure distributions. Our design choices were also motivated by intentions to computationally model the incremental changes in listener behaviour at each phase of exposure. Here, however, we report empirical observations.

## Materials

Previous work employed stimuli that sounded robotic (Clayards et al.; Kleinschmidt & Jaeger, 2016) and/or exhibited unnatural cue correlations (Theodore & Monto, 2019). We used a Praat script (Winn, 2020) to create three human-sounding minimal pair VOT continua (dill-till, dip-tip, and din-tin) from original voice recordings of a 23-year-old female AE native speaker. The continua ranged from -100ms to +130ms VOT in 5ms steps. To avoid unnatural correlations with secondary cues to onset stop voicing in AE, we set the F0 at vowel onset to follow its natural correlation with VOT, as observed in the original recordings. Similarly, the duration of the vowel was set to follow the natural trade-off relation with VOT reported in Allen & Miller (1999).

## Design

To assess incremental changes in listeners' categorization functions, we employed a multi-block exposure-test design (Fig. 1). Exposure was manipulated between participants. We first estimated listeners' expectations for a typical talker's VOT means and variances of /d/ and /t/. These estimates were based both on a norming experiment with our stimuli (N=24) and a phonetic database of AE onset stop voicing (Chodroff & Wilson, 2017). We then made three exposure conditions that shifted the VOT distributions for /d/ and /t/ by +0ms, +10ms, or +40ms relative to our 'typical talker' estimate (Fig. 2). Previous work set exposure distributions in the voiced and voiceless categories to have equal variance and to be distributed symmetrically around the category mean. Neither is the case for natural language. We thus sampled stimuli (see Fig. 2) from distributions with *unequal* variances observed in natural language (e.g., Chodroff & Wilson, 2017).

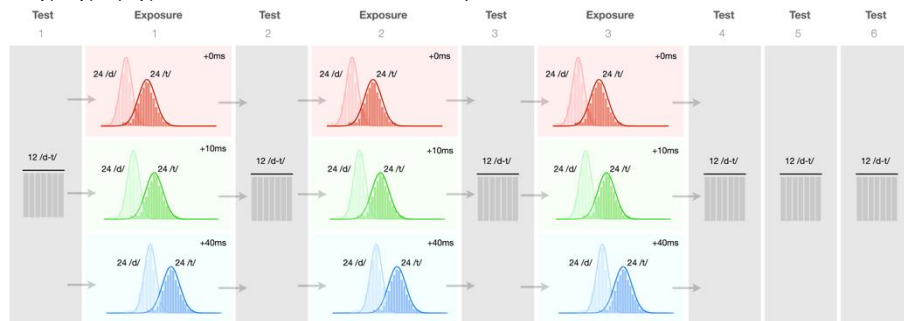


Figure 1. Block design of experiment with number of critical trials in each block. Test blocks were identical within and between conditions. Lines show the underlying distributions of the exposure and test stimuli. Block transitions were concealed from participants.

Test blocks consisted of 12 VOT-items (from -5ms–70ms), counter-balanced by minimal-pairs. In total there were 234 trials (including 18 catch trials that served to assess participant attention).

### Participants

122 AE listeners (male = 61; mean age = 37.6 years, SD = 12 years) were recruited from the Prolific crowdsourcing platform, and randomly assigned to one of the three exposure conditions (+0ms, +10ms, and +40ms shift).

### Procedure

Participants first underwent a headphones test and were given instructions to answer as quickly and as accurately as possible before the experiment began. On each exposure trial, participants clicked on a green button to play the recording of the word. Simultaneously, written forms of possible responses were displayed on the top left and top right of the screen. As shown in Fig. 2, half of the exposure trials *labelled* the voicing category: e.g., if a recording was intended to be *voiced*, both displayed words started with “d” (e.g., displaying “dill” and “dip” for a “d/til” recording). The other half of the trials were *unlabelled* (e.g., displaying “dill” and “till” for a “d/til” recording). Upon clicking on the word heard the next trial commenced. Test trials were always unlabelled. The order of trials and the left-right placement of responses was randomized for each participant, and counter-balanced across participants.

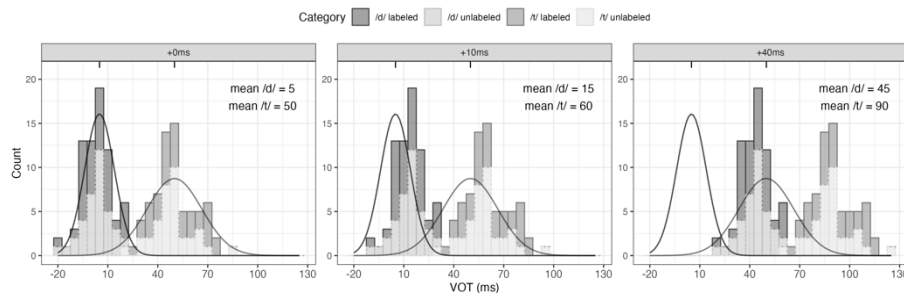


Figure 2. Histograms of distributions of /d/ and /t/ for the three exposure conditions. Black density lines are identical across panels and show the VOT distribution that the +0ms condition is sampled from.

### Results and discussion

Figure 3 shows listener categorisation behaviour at each test block. We focus on the estimated category boundary (point of subjective equality, PSE). Block 1 indicates participants' PSE *before* informative exposure, indicating participants' *prior* expectations about the VOT distribution of /d/ and /t/ for this talker. Our previous estimate of this PSE (24.5, which determined the +0ms exposure distributions) proved to be about 20 ms lower than the PSE observed for Block 1 (44.7 ms).

By Block 2, the PSE of all exposure conditions had shifted. The direction and magnitude of these shifts qualitatively follow the predictions of an ideal adaptor (or similar theories of incremental adaptation). Specifically, (1) the PSEs for the three conditions order in the same way as the means of exposure distributions ( $+0\text{ms} < +10\text{ms} < +40\text{ms}$ ); (2) the PSE of Block 2 shifted leftwards relative to Block 1 for the  $+0\text{ms}$  and  $+10\text{ms}$  conditions, in line with the observation that the *prior* PSE was actually about 20ms to the right of what we intended to be the  $+0\text{ms}$  condition (so that  $+0\text{ms}$  is actually  $-20\text{ms}$  exposure and  $+10\text{ms}$  is actually  $-10\text{ms}$  exposure); and, finally, (3) the PSE of Block 2 shifted rightwards relative to Block 1 for the  $+40\text{ms}$  conditions (which is actually  $+20\text{ms}$  exposure once the correct prior PSE is considered).

These PSEs remained fairly unchanged through Blocks 3-6: the remaining 96 exposure trials had only minor effects that showed mostly in the most extreme shift ( $+40\text{ms}$  exposure). This suggests that participants learned the distributions of the talker quickly—after exposure to 48 trials (2<sup>nd</sup> panel). Of note is that, despite these rapid changes in PSEs, the *extent* to which PSEs changed was greatly limited: even though the  $+0\text{ms}$  and  $+40\text{ms}$  exposure distributions differed by 40ms, the PSEs for those two conditions only differed by 10-14ms.

While these results are broadly consistent with exemplar and Bayesian theories of incremental adaptation, they also raise novel questions. In particular, it is unclear why shifts are both so quick—which would seem to imply weak weighting of prior expectations—and yet strongly constrained—which would seem to imply strong weighting of prior expectations. To the best of our knowledge, this tension has not previously been discussed but strikes as an important characteristic of speech adaptation to be understood.

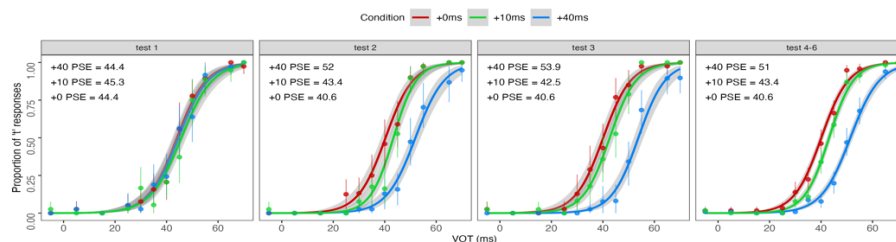


Figure 3. Mean categorisation functions by exposure condition. The last panel combines the final three post-exposure test blocks into one.

## References

- Clayards, M., Tanenhaus, M.K., Aslin, R.N., Jacobs, R.A. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3).
- Kleinschmidt, D.F., Jaeger, T. F. 2016. What do you expect from an unfamiliar talker? *CogSci*.
- Theodore, R.M., Monto, N.R. 2019. Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review*, 26(3), 985-992.

## Idiomaticity as a tool to explore automaticity and control in bilinguals and translators

Giulia Togato<sup>1</sup>, Teresa Bajo<sup>2,3</sup>, Pedro Macizo<sup>2,3</sup>

<sup>1</sup>RGRL Department, California State University Long Beach, US

<sup>2</sup>Mind, Brain and Behavior Research Center (CIMCYC), Spain

<sup>3</sup>University of Granada, Spain

### Abstract

We explored the impact of translation expertise and untrained bilingualism on the automaticity in retrieving translations of idiomatic units across languages. We predicted a differential access to idiomatic representations due to different bilingual experiences. Untrained bilinguals and professional translators were compared on the availability of cognitive resources to process idioms for later translation, employing a dual task paradigm. Our hypothesis was that translators would count on more available cognitive resources (i.e., more automatic behavior). Contrary to our predictions, results showed that translators, compared to bilinguals, exerted higher levels of cognitive control over the task, possibly to guarantee the high-quality standards required by professional translation. Findings are discussed in light of theoretical models of bilingual idiomatic processing and professional practice.

Keywords: idioms, figurative language, bilingualism, translation.

### Introduction

Idioms (e.g., *hit the road*) are “multiword conventional expressions whose semantic meaning cannot be derived from the comprehension of the individual words in the sentence” (Cacciari, Tabossi, 1988). Scholars have investigated how idioms are processed and represented in the native (L1) and second language (L2). Studies have shown that idioms are processed more quickly than non-idiomatic matched literal expressions by native speakers and by some proficient L2 speakers who, due to greater exposure, can use a direct access route to retrieve the idiom entry instead of computing the individual words included in it (e.g., Carrol, Conklin 2014). The processing differences between idiomatic and non-idiomatic expressions have been accounted for by theoretical models that explain how idiomatic expressions are processed. These models vary as to the degree of the idiomatic figurative meaning being directly (or less directly) related to the literal meaning of its individual components. Specifically, hybrid models (e.g., Sprenger et al. 2006) posit that idiomatic entries are connected to/and retrieved through both the idiom’s unitary lexical concept as well as the simple lemmas comprised in it. Hybrid models of idiomatic processing have been successfully applied to the bilingual domain, but whether they hold for professional translation is still unknown. The present study explored the predictions of the hybrid models regarding a differential access to

the idiom entry depending on whether translation expertise modulates the strength of the associative links amongst the idioms' constituent lemmas (i.e., a direct route).

Interestingly, it has been demonstrated that bilingual idiomatic processing is modulated by idiomatic cross-linguistic (in)congruency; specifically, authors have shown that the degree of idiomatic similarity across languages modulates the facilitation effect observed for idiomatic expressions in bilinguals (e.g., Titone et al. 2015). Hence, (in)congruency will be critically considered in the present study.

The only studies that have established a link between idiomatic processing and (early stages of) translation practice have focused on 1) language brokering experience (i.e., informal translation practice) and 2) formal training in translation (e.g., López et al. 2017). This is the first study that explores bilingual idiomatic processing as a function of professional translation expertise.

## **Hypothesis and methodology**

Our hypothesis was that professional translators, relative to untrained bilinguals, would be able to map L1 and L2 idiomatic expressions (at the idiom entry level) in a relatively automatic way. We based our prediction on the idea that training in translation tasks allows comprehension and retrieval processes to be fast and automatic, which plays a key role in the allocation of task-relevant cognitive resources. We employed a translation task where we manipulated the cross-language congruency between idioms (i.e., congruent vs. incongruent), and compared them to matched control expressions. To measure the degree of automaticity we asked participants to detect a tone while processing the idioms for posterior translation. In the tone detection methodology, the more cognitive resources are needed to process the idiomatic expression, the fewer resources will be left to process the tone. RTs to the tone will provide an index of the cognitive resources required to process idioms by bilinguals and translators, and the ease to accomplish the task (i.e., automaticity). All the materials used in the experiment underwent normative studies to control relevant sociolinguistic and demographic variables.

## **Results**

RTs for congruent idioms showed the “usual” idiomatic superiority effect for both bilinguals and translators; in fact, RTs in response to the secondary task were faster while processing congruent idioms relative to control sentences. However, translators were slower than bilinguals in tone detection when the tone was placed on incongruent idioms. This was not due to an inferior translation performance since accuracy analyses showed that translators outperformed bilinguals in correct translations, both for incongruent and congruent idioms.

Hence, the locus of the differences between the groups in the availability of cognitive resources might come from the degree to which they activated the L1 and L2 equivalent idiomatic entries relative to the activation of the individual lemmas across the two languages. All participants needed to activate the equivalent idiomatic entries to translate the idioms correctly; however, untrained bilinguals' errors while translating incongruent forms might be due to a word-by-word translation approach (i.e., the use of co-activated individual words in the L1 and L2), relative to the activation of the equivalent idiomatic entries across languages. This interpretation was supported by additional analyses indicating that untrained bilinguals experienced significantly more word-by-word translations [i.e., *He kicked the bucket* = *Pateó el cubo* (literal) instead of *Estiró la pata* (idiomatic)] compared to the translators. These results suggest that translators, despite needing more cognitive resources to process incongruent forms (slower RTs to tone detection), have more direct access to translation equivalents at the superlemma level, and adopt a chunked approach to translation. Additionally, translators might be activating the figurative meaning of sentences even when they should not (i.e., while processing control units). Therefore, we might interpret the relatively slower RTs previously observed in the processing of incongruent forms as due to the concurrent co-activation of the word-by-word meaning and the idiomatic meaning of the sentence. The translators might be increasing the time needed to detect the tone possibly because they need more cognitive resources to control for the concurrent coactivation.

## Discussion and conclusions

Our study indicated that the facilitation effect for congruent idioms did not depend on professional practice. However, differences between bilinguals and translators emerged when they processed incongruent idioms: less automatized behavior and higher levels of cognitive control were put into motion by translators to guarantee high-quality standards in translation. These results confirm the impact of cross-linguistic (in)congruency on bilingual idiomatic processing. This pattern can be accommodated within the hybrid models of idiomatic processing: these models posit that the lemmas of an idiom are bound together by one common entry in the mental lexicon and activated twice. Thus, until the appropriate meaning is selected, both the individual literal meanings and the figurative one would be activated and kept available. In addition, results of the study suggested that congruency was boosted by representational and linguistic overlap across languages, resulting in faster

availability of the idiom, and larger lemma/idiom competition for incongruent idioms.

Why were the simultaneous co-activation and related interference caused by incongruent idioms detected in translators and not in bilinguals? Although inhibitory control has been suggested as the way bilinguals select the required language, this mechanism does not seem appropriate to perform professional translation tasks (e.g., Ibáñez et al. 2010; Togato et al. 2017). Our study is in line with hybrid models of idiomatic processing and those studies based on the idea that bilinguals and translators may differ as to the way they negotiate their two languages.

### Acknowledgements

This research was supported by the Spanish Ministry of Education (Grants PSI2012–33625, PSI2015–65502–C2–1–P and PCIN–2015–132 to Teresa Bajo and Grant PSI2012–32287 to Pedro Macizo); Proyecto de Excelencia de la Junta de Andalucía P12–CTS–2369, Fondos Feder; FPI grant to Giulia Togato by the Spanish Ministry of Education (BES2009–025776, EDU2009–01111).

### References

- Cacciari, C., Tabossi, P. 1988. The comprehension of idioms. *Journal of Memory and Language*, 27, 668–683.  
[https://doi.org/10.1016/0749-596X\(88\)90014-9](https://doi.org/10.1016/0749-596X(88)90014-9)
- Carroll, G., Conklin, K. 2014. Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, 17, 784–797.  
<https://doi.org/10.1017/S1366728913000795>
- Ibáñez, A., Macizo, P., Bajo, M. T. 2010. Language access and language selection in professional translators. *Acta Psychologica*, 135, 257–266.  
<https://doi.org/10.1016/j.actpsy.2010.07.009>
- López, B. G., Vaid, J., Tosun, S., Rao, C. 2017. Bilinguals' Plausibility Judgments for Phrases with a Literal vs. Non-literal Meaning: The Influence of Language Brokering Experience. *Frontiers in psychology*, 8, 1661.  
<https://doi.org/10.3389/fpsyg.2017.01661>
- Sprenger, S., Levelt, W., Kempen, G. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54, 161–184.  
<https://doi.org/10.1016/j.jml.2005.11.001>
- Titone, D., Columbus, G., Whitford, V., Mercier, J., Libben, M. 2015. Contrasting bilingual and monolingual idiom processing. *Bilingual figurative language processing*, 171–207.  
<https://doi.org/10.1017/cbo9781139342100.011>
- Togato, G., Paredes, N., Macizo, P., Bajo, T. 2017. Syntactic processing in professional interpreters: Understanding ambiguous sentences in reading and translation. *Applied Linguistics*, 38(4), 581–598.  
<https://doi.org/10.1093/applin/amv054>



# Romanian lexical blending: from zero to hero.

Monica Vasileanu<sup>1,2</sup>, Anabella-Gloria Niculescu-Gorpin<sup>1,3</sup>, Cristina-Andreea Radu-Bejenaru<sup>1,3</sup>

<sup>1</sup>The 'Iorgu Iordan – Alexandru Rosetti' Institute of Linguistics, Romania

<sup>2</sup> Faculty of Letters, University of Bucharest, Romania

<sup>3</sup> Faculty of Foreign Languages and Literatures, University of Bucharest, Romania

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

In an elicitation experiment, 109 Romanian native subjects were presented with 50 pictures of hybrid objects, e.g. a half-orange, half-pear fruit, and were asked to name each object using a single word either in Romanian or in English. 64.66% of the elicited English words were blends, thus confirming the expected tendency towards using blends in English; unexpectedly, the same tendency was observed for Romanian with 68.50% blends. In Romanian linguistics, lexical blending has been seldom mentioned, interest resting only in speech errors, and not in the deliberate phenomenon as a fully-fledged word formation process of Romanian morphology. We believe that our data analysis suggests a shift in Romanian word formation tendencies, possibly under the pervasive influence of English.

Keywords: lexical blending, elicitation experiment, Romanian language, word formation, language change

## Introduction

Lexical blending, i.e. the process of coining a new lexeme out of (at least) two other source lexemes with some material loss from one or both source-words (Bauer, Lieber, Plag 2013, Renner 2022, a.o.) is a minor word formation process (Bauer 2021) with a sharp increase in productivity not only in present-day English (Kemmer 2003, Mattiello 2019), but also in other European languages, such as Polish (Konieczna 2012) or Bulgarian (Stamenov 2015). For present-day Romanian, we have observed a similar tendency that we wanted to study more closely, especially since the phenomenon did not receive attention in Romanian linguistics before, not being included among Romanian word formation reference works.

Since lexical blending is specific to English and the English influence seems to be the driving force for the increased productivity of lexical blending in other languages, and since little, if anything has been said about Romanian lexical blending as a word formation process, thus implying that compounding would be used in naming hybrid objects, our hypothesis was that Romanian native speakers with an above-average knowledge of English would produce more blends in English, but more compounds and fewer blends (if any) in

Romanian. To test our hypothesis, we used a hybrid-object naming task (Borgwaldt, Kulish, Bose 2012).

## Methods

### Participants

109 (92 BA and 17 MA) students from the English department, the University of Bucharest, aged 18-45 (mean = 21.2), with a self-assessed level of English as advanced (86) or intermediate (23), with formal English education of 2 to 29 years (mean = 12.31) participated in the experiment. 78% used English on a daily basis, 14% every two to three days, 6.5% weekly and only 1.5% occasionally. Students received course credits for their participation.

### Stimuli and task

The stimuli, 50 images of hybrid non-existent objects (e.g. a half-orange, half-pear fruit), balanced to represent animals, fruits, objects and mixtures (e.g. a half-frog, half-broccoli entity), retrieved online (e.g. from Pinterest) or manipulated by the authors, were chosen to display different degrees of phonetic similarity between the expected source-words in Romanian, ranging from no common features (e.g. *banană* 'banana' + *kiwi* 'kiwi') to a large segment of common phonemes (e.g. *banană* 'banana' + *ananas* 'pineapple').

Subjects were asked to name each object using a single word in Romanian for 25 images and the same for English. The order of the language blocks was counterbalanced across participants and subjects were able to self-pace the images. 57 subjects recorded themselves, and 52 subjects wrote down the answers. Due to the pandemic conditions, the experiment was carried out online.

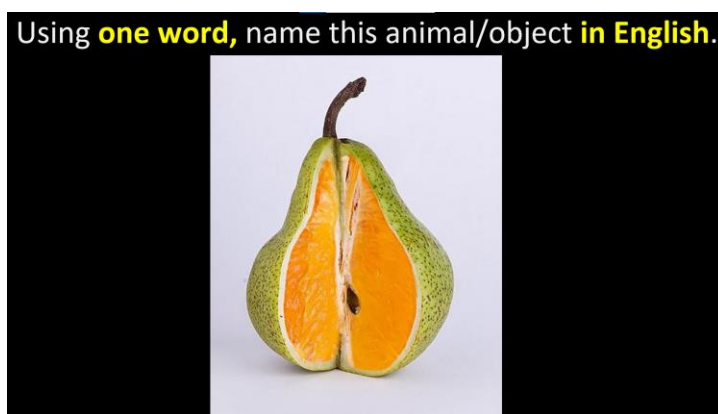


Figure 1. An image displaying a half-orange, half-pear hybrid fruit. The task summary was repeated on each slide.

The 109 subjects produced 5488 words, sometimes creating two names for the same image. 59 words presented a language mixing effect (e.g. *mandarlic* < Ro. *mandarină* ‘tangerine’ + En. *garlic*) so we could not assign them to any language and were discarded from the final analysis that comprised 2727 Romanian and 2702 English words.

Subjects used lexical blending, compounding, broadening (participants used one of the source-words to name the hybrid), affixation or a random proper name not connected with the objects as naming strategies. Some words were coded *ambiguous* either because the word formation process used was opaque or because they can have multiple interpretations.

Table 1. The naming strategies used by subjects in Romanian and English.

	Romanian	English
Blends	68.50%	64.66%
Compounds	12.21%	15.62%
One source word	13.71%	13.92%
Ambiguous	3.15%	3.03%
Other means	2.42%	2.78%

## Discussion

In both languages, lexical blending was the preferred strategy used for naming hybrid objects. While this is in line with the literature for English as the morpho-semantic iconicity of blending plays a major part (Renner 2020), it contradicts the expectations for Romanian: the traditional naming strategy for hybrid objects is compounding as shown in the literature or (academic) dictionaries that record already-established words for hybrid objects (see, for instance, *struțocămilă* ‘ostrich-camel’, a literary creation that has come to be used ironically for any mixture of two different things whose outcome is considered pointless).

Our results mirror the ones obtained for Ukrainian by Borgwaldt, Kulish & Bose (2012). The authors argue that their subjects invested more effort in the naming task because they were friends of the researchers. But this explanation does not hold in our case, since our subjects were students who fulfilled the task for course credits, were only told that they need to follow the instructions and that there were no correct or expected answers, and did the task in their own pace.

Another unexpected result is the fact that the subjects produced slightly more blends in Romanian than in English, perhaps because students are more used to creating new words in Romanian, their mother tongue, than in English. Nevertheless, the English and Romanian blends obtained are comparable, suggesting that, contrary to the non-existent Romanian literature on the subject, the process seems to be just as alive in Romanian as in English, at least for our subjects.

## Conclusion and further research

Our results suggest that, most probably under the English influence, lexical blending has become the preferred strategy for naming hybrid objects at least for the Romanian native speakers in our study both in English and Romanian, possibly indicating a shift in the Romanian word formation patterns. Further testing is necessary to establish whether lexical blending is restricted to Romanian native speakers of a particular age or/and with a good command of English or not.

## Acknowledgements

This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2019-0894, within PNCDI III.

## References

- Bauer, L. 2021. Compounds and minor word formation types. In Aarts, B., McMahon, A., Hinrichs, L. (eds.) *The Handbook of English Linguistics*, 463-482. Hoboken, NJ, Wiley.
- Bauer, L., Liebe, R., Plag, I. 2013. *The Oxford Reference Guide to English Morphology*. Oxford, Oxford University Press.
- Borgwaldt, S. R., Kulish, T., Bose A. 2012. Ukrainian blends: Elicitation paradigm and structural analysis. In Renner, V., Maniez, F., Arnaud, P. (eds.), *Cross-disciplinary perspectives on lexical blending*, 75-92. Berlin, Boston, De Gruyter Mouton.
- Kemmer, S. 2003. Schemas and lexical blends. In Cuyckens, H. et al. (eds.), *Motivation in Language: Studies in Honor of Günther Radden*, 69-97. Amsterdam, John Benjamins.
- Konieczna, E. 2012. Lexical blending in Polish: A result of the internationalisation of Slavic languages. In Renner, V., Maniez, F., Arnaud, P. (eds.), *Cross-Disciplinary Perspectives on Lexical Blending*, 51-73. Berlin, Walter de Gruyter.
- Mattiello, E. 2019. A corpus-based analysis of new English blends. *Lexis* 14, <https://doi.org/10.4000/lexis.3660>.
- Renner, V. 2020. An ecosystem view of English word-formation. *The Mental Lexicon* 15(1), 4-20.
- Renner, V. 2022. Blending. In Ackema, P., Bendjaballah, S., Bonet, E., Fábregas, A. (eds.), *The Wiley Blackwell companion to morphology*. Hoboken, Wiley.
- Stamenov, H. 2015. Borrowing word-formation: -ing suffixation and blending in Bulgarian. *Съпоставително езикознание/Сопоставительное языкознание*, 3, 163-197.

# Word recognition in Developmental Language Disorders in Greek

Konstantinos Vasileiou, Georgia Andreou

Department of Special Education, University of Thessaly, Greece

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

To this day, characteristics of the online visual perception in DLD remain quite unexplored. This study aimed to investigate the dynamic recognition of spoken words for nouns and verbs in children with Developmental Language Disorder whose spoken language is Greek. 9 children with DLD participated in this pilot study and their eye movements were recorded as they searched among an array of pictures for the target picture in response to hearing a single. The results showed that children with DLD presented a stronger performance of recognition for verbs than nouns while the more the word information the quicker the recognition.

Keywords: Developmental Language Disorder, eye-tracking, word recognition

## Introduction

The term Developmental Language Disorder (DLD) refers to persistent difficulties in acquiring, understanding, producing or using written and spoken language that arise during development and cause significant limitations in the communication abilities of the person. The language deficits of the disorder are not consistent with the individual's mental age and are not attributable to another neurodevelopmental or sensory disorder or other neurological condition (ICD-11, 2018). The areas of deficits cover the whole range of language, namely semantics, syntax, morphology, phonology and pragmatics (Leonard, 2014).

While language processing based on verb argument structure in children with DLD has been studied, this has been done mainly with off-line methodologies. *Visual world paradigm* (Tanenhaus et al., 1995) has made the online investigation of language feasible. This model basically makes use of eye-tracking while the child simultaneously sees an image and listens to the corresponding naming thus offering real-time measurements of language processing.

The aim of the present work is to investigate the dynamics of spoken word recognition for nouns and verbs in children with Developmental Language Disorder (DLD) as part of a pilot study which attempts to adapt and replicate the study of Andreu, Sanz-Torrent and Guardia-Olmos (2012). In particular, it will be studied if children with DLD show better performance in identifying verb and noun targets in response to auditory/visually presented words in the

Greek language and whether verbs with a larger number of arguments take longer to retrieve due to more complex encoding in the semantic lexicon.

## Methods

All participants had Greek as their mother tongue, normal vision, hearing, mental ability and lack of neurological, sensory and emotional deficits. Nine (9) children (6 boys – 3 girls) with DLD and age ranging from 4;6 to 7;11 years took part in this study. A language assessment based on the standard criteria given by ICD-11 for the diagnosis of DLD was addressed by the researcher. The assessment included intelligence scales, language tests, temperament and socio-economic questionnaires, tests of psycho-emotional development and assessment of oral structure and motor function.

18 nouns and 18 verbs (six one-argument, six two-argument, and six three-argument) made up the target words. Care was taken to weight number of syllables, frequency of occurrence and age of acquisition in words of the same category. Each word was paired with a corresponding image rated as highly visual by a group of adults. Each target image was then combined with 3 other images (distractors) so that the resulting set of four images always included 2 object images and 2 action images. The words were uttered and recorded by a native Greek speaker while they were placed in four square quadrants on the computer screen. The duration of the spoken target word was constant (i.e. 1 sec) and was embedded in a 4-sec mp4 file.

Participants were instructed to look at the picture that would be heard until it disappears from the screen. Familiarization tests preceded. A region of interest was defined for each target image based on screen coordinates to coincide with the rectangle surrounding the image.

## Results

Figure 1 demonstrates that the proportion of no fixations is smaller to the more complex verbs. The proportions were calculated by excluding any trial track losses (i.e. events where the participants did not show any gaze to the picture) and then dividing the rest of the events to the total number of trials per word.

In figure 2. the proportion of successful trials over time is presented. 3 time windows are noted (i.e. 200-1000ms, 1000-2000ms, 2000-4000ms). Success was defined as a first fixation to the correct stimulus. The rate of increase was similar and linear across the different word types during the first and second time window. Nevertheless, for times larger than 2000ms the rate was smaller, implying that the time to first fixation was likely to be during the 2000ms and if the child failed to fixate on the correct picture, it was unlikely that would do so for times larger than 2000ms.

A Poisson regression accounting for child-specific clustering using random effects was performed. The main outcomes were fixation times and time to first fixation. 19% increased fixations in verbs of type 1 (1-argument) compared to

nouns was found. This increase was marginally significant. Another 21% of increased fixations on verbs of type 2 and 28% of increased fixations on verbs of type 3 compared to nouns were found and both increases were significant. Respectively, overall children needed approximately 5% less time to first fixation when it came to verb type 1 or verb type 2 compared to nouns, whereas same time to first fixation between a noun and a verb of type 3 was noted. In addition, there was a 17% decrease in the fixation counts when the word consisted of 3 or 4 syllables compared with 2-syllables words. This result was significant. Respectively, children with DLD needed 32% and 21% more time to first fixation to a word with 3 or 4 syllables compared with 2-syllable words. These results were significant.

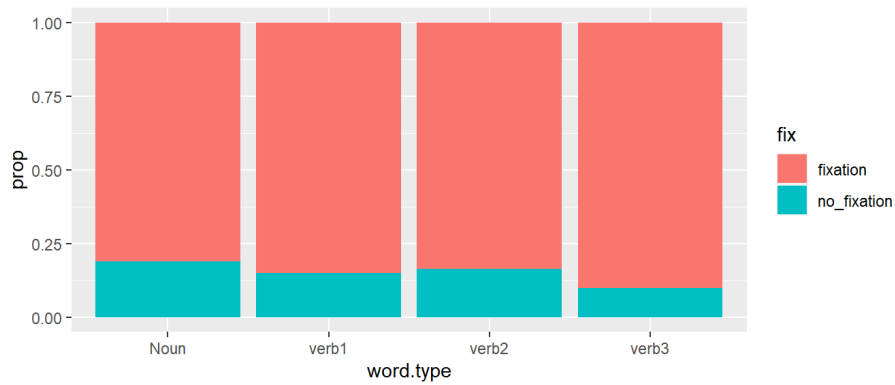


Figure 1. Proportion of fixations and no-fixations for each word type.

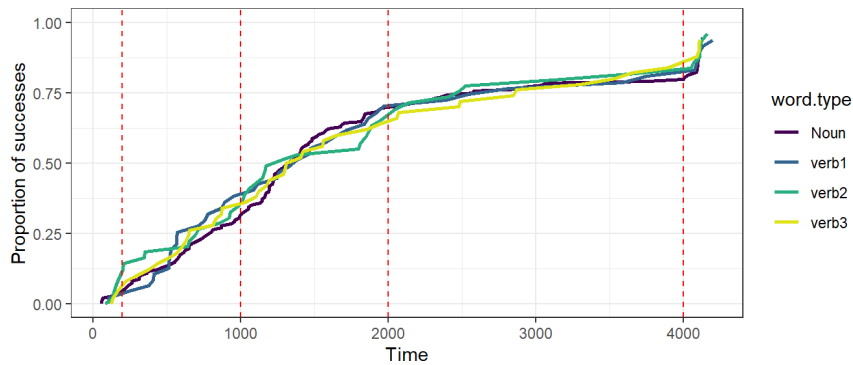


Figure 2. Proportion of fixations to the nouns (no), one-argument (v1), two-argument (v2) and three-argument verbs over time from the video onset.

Accounting for age, sex, duration of speech therapy, Raven's PM scores and language scores, no effect of any was found. That meant that both the number of fixations and the time to first fixation did not change for any one unit in the language, cognitive or demographic scores.

## Conclusion

The current work intended to collect data about the way Greek speaking children recognize nouns and verbs as measured by an on-line eye-tracking experiment. As a result, children with DLD demonstrated a tendency to recognize verbs in a faster way than nouns, while as the number of the verb arguments increased, the recognition was becoming faster. These findings not only did they not replicate the results the original study in Spanish speaking children with DLD offered, but also showed quite the contrary. Since Greek literature has not still offered rich data on the online understanding of words in the DLD population, this pilot study suggested that in some way the verb argument information helped children with DLD to achieve a quicker recognition of more complex words. That would be attributed to the more concrete semantic nature of verbs with more arguments (i.e. sell) as opposed to the more abstract nature of verbs with fewer arguments (i.e. cry) that were used in this study. Since the more the word information the quicker the word recognition, a significant role of the semantic links the children own may be arose. A control group as well as the enrollment of a bigger population will shed light on the initial research question invigorating the observation power of the current study.

## References

- World Health Organization. 2018 (June 18). International statistical classification of diseases and related health problems (11th Revision). <https://icd.who.int/browse11/l-m/en>
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.C. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Andreu L, Sanz-Torrent M, Guàrdia-Olmos J. 2012. Auditory word recognition of nouns and verbs in children with Specific Language Impairment (SLI). *Journal of Communication Disorders*. 2012; 45(1):20–34.
- Leonard, B.L. 2014. *Children with Specific Language Impairment*. Massachusetts: the MIT Press Cambridge.



# Testing variables of linguistic prestige in sociolinguistic questionnaires

Tobias Weber

Graduate School Language & Literature, Ludwig-Maximilians-Universität München, Germany

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

This paper outlines a questionnaire to investigate the influence of sociolinguistic descriptions on language attitudes and prestige. It follows a design commonly employed in customer choice and market research: A (semi-)fictitious entity is presented to different respondents with a range of descriptors or values to test the effect of particular variables of interest. Respondents are then asked to rank or share their perception of the entity's different attributes. This design can be transferred to sociolinguistic descriptions regarding speaker numbers, degree of urbanity, institutional support, or historical importance. The goal is to learn more about sociolinguistic factors that influence language attitudes and to be able to quantify their effects. This knowledge can be used by activists and stakeholders when discussing language vitality.

Keywords: sociolinguistics, prestige, language attitude, ideology, survey

## Introduction

Perceptions about languages, their prestige, formulated as attitudes or ideologies are difficult to operationalise. These concepts touch upon abstract understandings of language, identity, or speakerhood, which are equally complex in their formalisation. At the same time, many social sciences dealing with preferences and attitudes have long-standing traditions of measuring these for their research. This paper outlines a questionnaire type inspired by customer and market research that can be used to understand the formation of prestige and to quantify the importance of different factors. A pilot of the questionnaire is currently circulating.

## Background

In the sociolinguistic research tradition, language regard or attitude research has been largely carried out through qualitative, ethnographic study on the one hand, or judgements of grammaticality and perception (e.g. the matched-guise test) on the other (cf. Garrett 2010). As a parallel development in the sociology of language (Fishman 1991), researchers using an abstract perspective on language and its users have attempted to describe the same phenomena based on quantitative measurements, such as speaker numbers, enrolment in language courses, or economic benefits. A prominent example can be seen in Abram de

Swaan's *Q-value* or *communication value* calculated for a range of official state languages (2001). While the communication value is not equal to linguistic prestige, qualitative sociolinguistic research confirms that speakers consider the notion of "usefulness" of a language when planning language acquisition or transmission (Vigouroux & Mufwene 2020). The approach presented in this paper aims to reconcile these research strands by offering a tool to operationalise and measure effect sizes for different variables that influence linguistic prestige and the perception of languages based on individual responses to a questionnaire.

Market researchers investigating consumer choice face a similar situation where they can either ask customers about particular products and brands or attempt to generalise decisions based on abstract attributes of the industry and its goods. While a company might be very interested to hear about its comparative prestige among competitors, the reasons for these differences may remain opaque because it is difficult to control for all variables in real-world examples. Likewise, a perception test or attitude questionnaire on a set of languages can produce very insightful reports on particular sociolinguistic settings, yet leave the underlying generative mechanism in parts to the investigators' interpretation. To provide deeper understanding of these mechanisms and the interaction between variables, market research also uses fictional cases to test hypotheses or experiment with particular variables – the same method can also be applied to sociolinguistic research.

## Methodology

As indicated above, this research is based on a questionnaire, yet its central ideas may also be applied to interviews: A (semi-)fictitious sociolinguistic description is presented to the respondents who are asked to describe their attitudes towards the language and its speakers afterwards. The latter part can be conducted through the evaluation of declarative statements, estimation of similarity and social distance, likeliness to interact with the language (e.g. as a learner, in social activism, financial support), or through open-ended interview questions. The descriptions can be varied between different groups, in accordance with the purpose of the study. For the pilot study, these variables were speaker numbers, institutional support, Global North vs. Global South, and urban lifestyle. While some of these variables like speaker figures could be coded with their real values (for linear regression), these were treated as ordinal or binary data in the pilot to keep the number of experimental groups manageable (i.e. [larger] vs. [smaller], [+institutional support] vs. [-institutional support]). In addition, the marginal effect of speakers, schools, or village counts on perception and prestige cannot be calculated on this atomistic level beyond extreme cases such as languages with 0 or 1 speaker. Similarly, de Swaan's *Q-value* also uses a ratio of speakers for the calculation of communicative value.

It is important to note that the texts are purposely phrased in a way that can trigger stereotypes, emphasise aspects of ethnolinguistic vitality, or present generalisations about the language and its speakers. Many linguists, myself included, would contest this essentialist presentation of a sociolinguistic situation for any language. Subsequently, a stereotyped description of real-world examples can be considered unethical, as they replicate historical injuries. With fictitious examples, these descriptions exist within a grey area of ethics: the general style of presentation may be deprecated but the examples serve as a blank canvas for the experiment. There are certain boundaries that must be respected, even with fictitious languages; racial slurs, religious discrimination, or colonial stereotypes are to be avoided in all contexts. The text should rather evoke the impression of a travel guide or encyclopedia entry. Importantly, all participants are informed about the fictitious nature of the examples after completing the questionnaire, with further information on how existing language communities may be supported.

Table 1. Stereotyped statements about the sociolinguistic situation of the fictional case studies.

	Positive presentation	Negative presentation
Stereotypes of modernity	Most of the ___ live in <b>suburban areas</b> of the ___ <b>metropolitan area</b> . Although fishing had been of central importance in the traditional way of living, few community members follow this <b>traditional lifestyle</b> .	The ___ inhabit <b>forested areas</b> in ___ a part of a national park. Most of the ___ engage in <b>subsistence economy</b> , i.e. they produce most of the products they need themselves.
Institutional support and means of transmission	The ___ language is taught in <b>25</b> K-12 schools and <b>supported by the local government</b> .	The ___ language is mostly <b>transmitted at home</b> but also taught in <b>7</b> local schools.
Historical importance	Recent <b>archaeological findings</b> suggest that, in [before the colonisation], the ___ lived in a network of smaller city states.	—

## Discussion

This section will discuss some insights from the pilot and participant feedback that will be used to improve the experimental design further. First, participants generally found this novel approach to be engaging, as it led to reflections about their own language use and linguistic attitudes. At the same time, the abstract nature of the questions caused some difficulties for some respondents who stated that “it was difficult for me to answer questions because I have not

heard of [these] languages”. There is no straightforward solution, as most non-linguists will not be aware of many languages outside of their everyday lives. The presented approach using fictitious examples circumvents availability bias but requires more explanation of the task in the questionnaire.

Second, the questionnaire drew some criticism levelled at the ranking or explicit formulation of attitudes. This social desirability bias was especially prominent among student respondents stating that “no language is better or prettier than another [...] I did not want to answer those questions”. While I wholeheartedly agree with the sentiment (“I would love to see more people speak minority languages”), most respondents have studied a foreign language themselves and thus made a decision for or against some of the world’s languages. This link to decision making could be used for an improved version of the questionnaire, where participants estimate the probability of learning the (fictitious) language or attending a cultural event of the community. Furthermore, a different presentation of the ranking questions may reduce reluctance to respond honestly.

Last, it was uplifting to see that respondents (mostly students in Germany) expressed an interest in learning more about linguistic diversity and minority languages. The experimental approach may help to broaden the range of languages studied in attitude surveys, as a set of languages the respondents know will always favour languages of wider communication and global languages. While the fictitious examples do not count as languages, they provide a reference point in the data for the vast majority of languages that are unfamiliar to the participants. To this end, the experiment aims to increase their visibility in research.

## References

- De Swaan, A. 2001. *Words of the World: The Global Language System*. Cambridge, Polity Press.
- Fishman, J. 1991. Putting the ‘socio’ back into the sociolinguistic enterprise. *International Journal of the Sociology of Language* 92, 127-138.
- Garrett, P. 2010. *Attitudes to Language*. Cambridge, Cambridge University Press.
- Vigouroux, C., Mufwene, S. 2020. Do Linguists Need Economics and Economists Linguistics? In Vigouroux, C., Mufwene, S. (eds.) 2020, *Bridging Linguistics and Economics*, 1-55. Cambridge, Cambridge University Press.

# Four-stage evolution of Jin Chinese entering tones: from Lǚliáng to Bìngzhōu

Yang Wei

Chinese Language Centre, Shenzhen MSU-BIT University, China  
Division of Humanities, The Hong Kong University of Science and Technology, Hong Kong Special Administrative Region, China

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

This paper demonstrates the evolution of Jin Chinese entering tones. The “flip-flop” phenomenon has been observed in Jin entering tones. Both Lǚliáng and Bìngzhōu Jin have two entering tones, namely, Yīnrù tone and Yáng-rù tone. The Bìngzhōu entering tones have a “Yīn-low vs. Yáng-high” contrast, which is opposite to the Lǚliáng “Yīn-high vs. Yáng-low” contrast. 34 native Lǚliáng and Bìngzhōu Jin speakers’ recordings have been acoustically and phonologically analyzed. This study reconstructs a four-stage path that shows how this “flip-flop” phenomenon, namely, the Lǚliáng “Yīn-high vs. Yáng-low” entering-tone contrast changed to the Bìngzhōu “Yīn-low vs. Yáng-high” contrast.

Keywords: Jin Chinese, entering tones, flip-flop, tone evolution

## Introduction

Jìn is a major variety of Chinese, which is spoken in Shānxī province and neighboring regions. Bìngzhōu and Lǚliáng are two main dialects of Jin spoken in Shānxī and Shǎnxī provinces. These two dialects both have two entering tones, namely, Yīnrù tone and Yáng-rù tone (Shen 2006).

It is generally assumed that Yīn tone has a higher pitch than Yáng tone within the same Middle Chinese tone category. However, we find that the Bìngzhōu entering tones have a “Yīn-low vs. Yáng-high” contrast, which is opposite to the Lǚliáng “Yīn-high vs. Yáng-low” contrast. Wang (1967) first proposed the linguistic concept “flip-flop” to describe alternations of two tones, in which the high tones become low tones and the low tones become high tones in certain linguistic environments.

By analysing 34 native Lǚliáng and Bìngzhōu Jin Chinese speakers’ recordings, and demonstrating the pitch patterns of Yīnrù tone (T4a) and Yáng-rù tone (T4b), we identify four stages with regard to the development of the two entering tones. We argue that the contrast of Bìngzhōu entering tones has evolved from that of Lǚliáng.

## Methodologies

The “multi-register and four-level” tonal model (Zhu 1999, 2005), the universal tonal inventory, and the evolutionary comparative method (Zhu 2014, 2018) are the main methodologies applied in this study.

34 native Lǚliáng and Bìngzhōu Jìn Chinese speakers’ recordings have been acoustically and phonologically analyzed. The software Praat was used to reduce noise on some recordings, to annotate the recorded tokens for tonal measurements, to extract pitch values, and to draw up spectrograms, etc. The method in Zhu (2010) was adopted to measure the tone-bearing part of each token and to normalize the inter- and intra-speaker variations.

## Data analyses

In this section, we demonstrate the evolution of Jìn entering tones, to reconstruct a path that shows how the Lǚliáng “Yīn-high vs. Yáng-low” entering-tone contrast changed to the Bìngzhōu “Yīn-low vs. Yáng-high” contrast. We think the “flip-flop” that happened in Jìn entering tones is a gradual phonetic change rather than a sudden reversal.

### Stage one: high T4a vs. low-dipping T4b

The first stage is that T4a is a short high tone while T4b is a mid-short low-dipping tone. Five tonal systems belong to stage one as shown in Figure 1 and Table 1. All the five systems are Lǚliáng dialect.

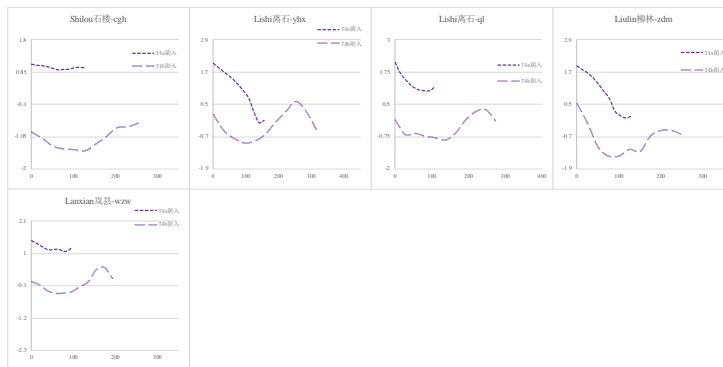


Figure 1. Stage one: high T4a vs. low-dipping T4b.

Table 1. Stage one: high T4a vs. low-dipping T4b.

Dialect sites	Yīnrù T4a	Yánggrù T4b
Shílóu-cgh; Líshí-yhx; Líshí-ql;	{55/54/53}	{323/303}
Lǚlín-zdm; Lánxiàn-wzw	short, high-level	mid-short, low-dipping

The pitch height, especially the height of tonal head of T4a, is as high as “5” on the scale of modal register under the framework of the multi-register-four-level model. T4b, as a low-dipping tone, is obviously lower than T4a.

### Stage two: lowering T4a vs. low-dipping T4b econd-level

In the second stage, similar to the first stage, T4b is a low-dipping tone. However, T4a is gradually lowering its pitch height. Stage two has 10 systems, of which nine are Lüliáng dialect, and one is Bingzhōu dialect.

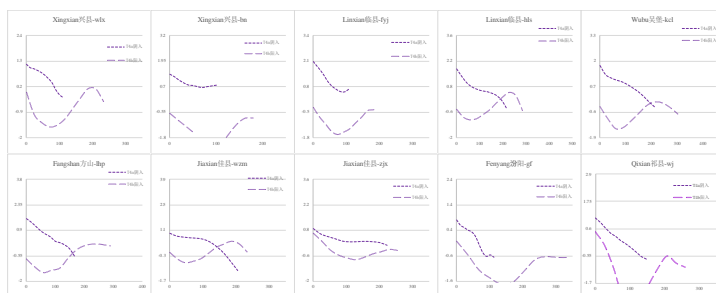


Figure 2. Stage two: lowering T4a vs. low-dipping T4b.

### Stage three: lowering T4a vs. back-dipping T4b econd-level

In stage three, when T4a continues lowering, T4b starts to change as well. Stage three includes 11 systems, of which seven are from Lüliáng and four are from Bingzhōu.

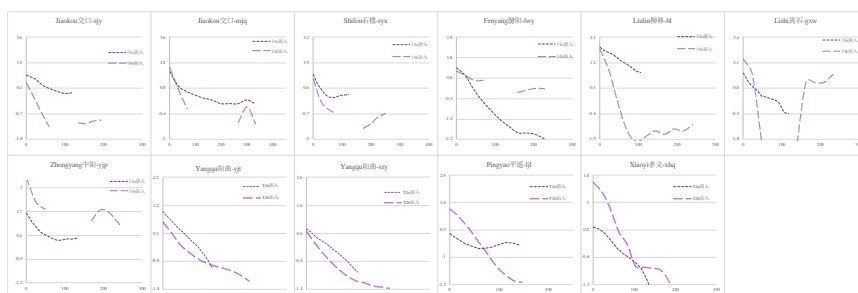
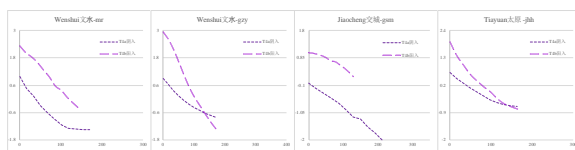


Figure 3. Stage three: lowering T4a vs. back-dipping T4b.

### Stage four: low T4a vs. high T4b

Stage four is the final stage, in which T4a is lower than T4b. This stage includes eight tonal systems which are all from Bingzhōu dialect.



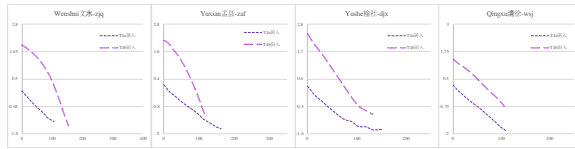


Figure 4. Stage four: low T4a vs. high T4b.

## Conclusions

This study, by providing the illustrative examples of Lǚliáng and Bingzhōu entering tones, exhibits a four-stage evolutionary path to explain how the Lǚliáng “Yīn-high vs. Yáng-low” contrast finally becomes the Bingzhōu “Yīn-low vs. Yáng-high” contrast. The process of change is shown schematically in Figure 5.

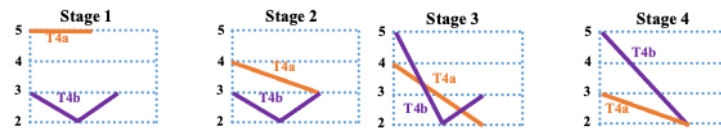


Figure 5. A schematic process of the “flip-flop” of Lǚliáng and Bingzhōu entering tones.

At the first stage, the T4a is a short high tone while the T4b is a mid-short low tone with a dipping contour. In stage two, the T4a begins to lower its height and lengthen its duration, to gradually become a mid-tone with a falling contour. From stage three, as the T4a continues lowering, the T4b starts changing as well to avoid merging with each other. The T4b, originally a low-dipping tone, becomes a back-dipping tone by raising its tonal head. For the final stage, the T4a becomes a mid- or low-falling tone, and the T4b becomes a high-falling tone. By the end of these four stages, the two entering tones have finished the “flip-flop” process from Lǚliáng “Yīn-high vs. Yáng-low” to Bingzhōu “Yīn-low vs. Yáng-high”.

## References

- Shen, Ming. 2006. Jinyu de fenqu (gao) [Re-classification / Re-distribution of Jin group]. *Fangyan [Dialect]* 4, 343-356.
- Wang, William S-Y. 1967. Phonological features of tone. *International Journal of American Linguistics* 33.2, 93-105.
- Zhu, Xiaonong. 1999. *Shanghai Tonetics*. Munich: Lincom.
- Zhu, Xiaonong. 2005. *Shanghai Shengdiao Shiyuan Lu [An experimental study on Shanghai tones]*. Shanghai: Shanghai Educational Publishing House.
- Zhu, Xiaonong. 2014. *Shengdiao leixingxue dayao [An outline of tonal typology: A study of tonal types]*. *Fangyan [Dialect]* 3, 193-205.
- Zhu, Xiaonong. 2018. *Yuyin Dawen [On phonetics]*. Shanghai: Xuelin Press.



# Cue reweighting in Shanghainese sandhi patterns

Zihao Wei

Department of linguistics, University of Ottawa, Canada

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

The role of cue reweighting in reshaping tone sandhi patterns is rarely discussed. The consensus is that in the disyllabic prosodic words in Shanghainese, F0 differences in lexical tones are neutralized in the second syllable and that VOT and closure duration are the primary cues in maintaining the voicing/register contrast. However, a cue reweighting process in reshaping Shanghainese sandhi patterns: F0 is started to replace VOT and closure duration as the primary cue for voicing/register contrast in the second syllable position. This process of F0 gradually acquiring the status of the primary cue in the sandhi tone is very similar to tone split in initial positions in other language.

Keywords: Tonogenesis, Cue weighting, Sound change, Sandhi pattern, Shanghainese.

## Introduction

It is well-known that obstruents affect the F0 of the following vowel, and the vowels following a voiceless onset tend to have relatively higher F0 in many languages (e.g., Kirby & Ladd, 2015). Furthermore, in tonal languages where F0 differences indicate lexical contrasts, the relation between voicing and F0 is often manifested as co-occurrence restrictions between tone and the onset consonant of tone-bearing syllables.

In Shanghainese, a disyllabic prosodic word gets its surface sandhi tone by delinking the tone of the second syllable and spreading the tonemes of the first syllable over the two syllables. For example, the prosodic word /tsɔ<sup>3a</sup> vɛ<sup>3b</sup>/ ‘fried rice’ that could be autosegmentally represented as /ts<sup>hɔ</sup><sup>MH</sup> vɛ<sup>LH</sup>/ has a surface form [ts<sup>hɔ</sup><sup>M</sup> vɛ<sup>H</sup>] (e.g., Yip, 1995). In Shanghainese, as traditionally described, the functional cost of neutralizing tone in second syllables is mitigated by the preservation of vocal fold vibration. Some previous studies suggested that vocal fold vibration is no longer the primary cue (e.g., Chen & Wang, 2012), while others show that the two categories have multidimensional acoustic and articulatory correlates, including a significant difference in Closure of Duration of onset stop (CD) (Shen, Wooter & Wang, 1987) and in breathiness concentrated mainly at the onset of the following vowel (e.g., Ren, 1992).

The lexical contrasts can still be maintained when the primary cue changes. Such a change is reminiscent of tonogenesis, a type of cue shift that has been studied extensively (e.g., Hyman, 1973). In this work, we want to explore how

the cue reweighting of F0, VOT, CD, and various phonation-related measurements reshapes Shanghainese sandhi patterns.

## Experiment

### Stimuli

Four disyllabic names, /pɔ<sup>3a</sup> tɛ<sup>1a</sup>/, /pɔ<sup>3a</sup> dɛ<sup>3b</sup>/, /bɔ<sup>3b</sup> tɛ<sup>1a</sup>/, and /bɔ<sup>3b</sup> dɛ<sup>3b</sup>/, were used as target words (XY) in a frame sentence /ŋu<sup>3b</sup> ʔiɔ<sup>3a</sup> zɪn<sup>3b</sup> fiəʔ<sup>4b</sup> ɲin<sup>3b</sup> tɛiɔ<sup>3a</sup> XY/ (The person whom I am finding is XY). In total, 180 tokens were used in the experiment (4 disyllabic prosodic words × 3 repetitions × 15 speakers).

### Speakers

Fifteen native speakers of Shanghainese participated in this work and were divided into three age groups: younger than 40 (7 participants), 40 – 60 (5 participants), and older than 60 (3 participants). All participants were recorded on a Komplete Audio 6 recorder with a C544L Vocal microphone.

### Data analysis

The data were normalized using z-score. Principal Component Analysis (PCA) (Jackson, J. E., 1991) was used to reduce the dimensionality of phonation-related measures. Linear Discriminant Analysis (LDA) (Schertz, 2020) was used to determine the weights of each acoustic property in production.

## Results

A principal component analysis was first run on all phonation-related spectral tilt (H1\*-H2\*, H2\*-H4\*, H4\*-H2K\*, H2K\*-H5K\*, H1\*-A1\*, H1\*-A2\*, H1\*-A3\*) and noise measures (CPP, HNR 0-500 Hz, HNR 0-1500 Hz, HNR 0-2500 Hz, HNR 0-3500 Hz). The main results are that the first principal component (PC1) accounted for 49.2% of the variance of the measures, and the second principal component (PC2) accounted for 27.1%. PC1 was highly correlated with HNR 0-1500 Hz ( $r = -0.502$ ), HNR 0-2500 Hz ( $r = -0.501$ ), HNR 0-500 Hz ( $r = -0.495$ ) and HNR 0-3500 Hz ( $r = -0.494$ ), and PC2 was highly correlated with H1\*-A2\* ( $r = -0.59$ ), H1\*-A2\* ( $r = -0.526$ ), and H1\*-A3\* ( $r = -0.515$ ). Therefore, PC1 and PC2 were used to represent phonation in the LDA models below.

According to Figure 1 and 2, each age group has its own strategy. In /pɔ<sup>3a</sup> tɛ<sup>1a/3b</sup>/, VOT is the primary cue only for elder speakers, and F0 becomes the primary cue in /pɔ<sup>3a</sup> tɛ<sup>1a/3b</sup>/ in young speakers. In /bɔ<sup>3b</sup> tɛ<sup>1a/3b</sup>/, VOT is always the primary cue, but its relative weight has been greatly diminished for middle-aged and young speakers. In /bɔ<sup>3b</sup> tɛ<sup>1a/3b</sup>/, there was no such increase of F0. However, since the relative weight of VOT is greatly weakened, we can likewise observe an increase of F0 in the transition from the middle-aged to the young generation.

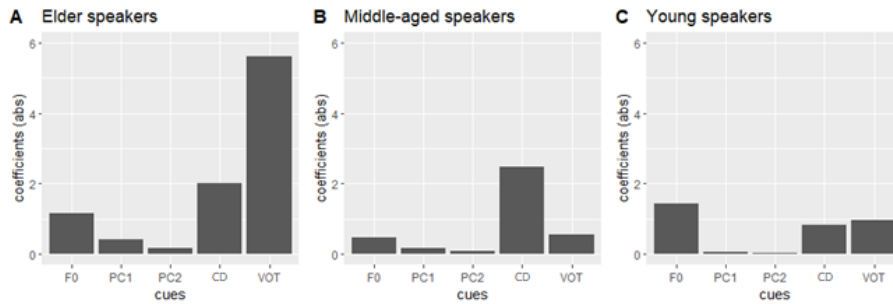


Figure 1. Visualized LDA results for /pɔ<sup>3a</sup> te<sup>1a</sup>/ and /pɔ<sup>3a</sup> de<sup>3b</sup>/.

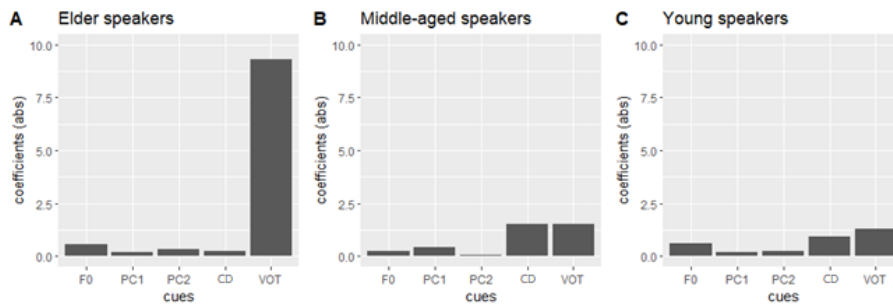


Figure 2. Visualized LDA results for /bɔ<sup>3b</sup> te<sup>1a</sup>/ and /bɔ<sup>3b</sup> de<sup>3b</sup>/.

Overall, the LDA results exhibit a negative trend in that the height of the bars reflecting CD went down as much as F0 went up. Whether it keeps its status as a secondary cue in each generation depends on the relationship between F0 and VOT. Meanwhile, the relative weights of PC1 and PC2 are lower than that of the previously mentioned F0, VOT, and CD on the second syllable in both sandhi patterns, suggesting the relative unimportance of phonation for the voicing/register contrast.

## Discussion and conclusion

The first conclusion we can draw from the results is the relative unimportance of phonation. It is well known that the effect of obstruents on F0 is manifested mainly through the F0-onset co-occurrence restriction on the tone-bearing syllables (e.g., Hyman, 1973), and both the obstruent properties and F0 are controlled by laryngeal adjustment (e.g., Ohala, 1978). Many studies investigate the mechanisms of consonant articulation that account for the local F0 perturbation (e.g., Halle & Stevens, 1971). Since our experiment is done within the sandhi domain, whether this inconsistency that exposed the F0-phonation relationship behaves differently in different contexts requires further study.

In addition, the relative weight of VOT undergoes a significant decrease, and the relative importance of CD exhibits a negative correlation with F0. Chen's

research (2011) on the consonant-F0 interaction on the second syllables in Shanghaiese suggests that there was usually a significant F0 perturbation caused by vocal fold vibration, which is generally consistent with the present study's findings. Still, the consonant-F0 interaction does not result in a direct trade-off between the two cues: the decrease in the absolute weight value of VOT between elders and middle-aged speakers is not mirrored by an increase in F0 or CD. On the contrary, the absolute value of F0 rises more in 3a-1a/3b and finally achieves predominant cue status in the younger group's production of 3a-1a/3b. The results indicate an interesting fact: the essence of this sound change at the production level is that F0 is gradually gaining its importance.

The results suggest that the tone sandhi pattern in Shanghaiese is changing: in the second syllable of a sandhi domain, F0 is overtaking CD and VOT as the primary cue in younger generations. From a wider perspective, looking at the phonetic nature of the cue reweighting in Shanghaiese sandhi patterns will enable us to understand better the tone split that widely happens among Eastern and Southeastern Asian languages.

## Acknowledgements

This work is based on a section of the author's memoir in uOttawa, supervised by Professor Marc Brunelle.

## References

- Chen, Z., Wang, Y. 2012. Roles of F0 and closure in the voiced-voiceless distinction for initial stops in Wu dialects—take Shanghaiese for example. In *Tonal Aspects of Languages-Third International Symposium*.
- Halle, M., Stevens, N. 2002. A note on laryngeal features (pp. 45–61). Berlin: Mouton de Gruyter. 1971 (Reprinted in M. Halle, *From memory to speech and back: Papers on phonetics and phonology*).
- Hyman, L. (Ed.) 1973. *Consonant types and tone: Southern California occasional papers in linguistics*. Los Angeles: The Linguistic Program, University of Southern California.
- Kirby, J.P., Ladd, D.R. 2015. Stop voicing and F0 perturbations: Evidence from French and Italian[C]/ICPhS.
- Ohala, J. 1978. Production of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). New York: Academic Press. 1978.
- Jackson, J.E. 1992. *A user's guide to principal components[M]*. John Wiley & Sons.
- Ren, N. 1992. *Phonation types and consonant distinctions: Shanghai Chinese*. Ph.D. Dissertation. The University of Connecticut.
- Schertz, J., Clare, E.J. 2020 *Phonetic cue weighting in perception and production[J]*. *Wiley Interdisciplinary Reviews: Cognitive Science*, 11(2): e1521.
- Shen, Z., Wooters, C., Wang, W.S.Y. 1987. Closure duration in the classification of stops: A statistical analysis. *OSU Working Papers*, 35, 197–209.
- Yip, M. 1995. Tone in east Asian languages[J]. *The handbook of phonological theory*, 476: 494.

# Mandarin and Spanish basic color terms vis-à-vis

Mingshan Xu<sup>1</sup>, Jingtao Zhu<sup>2,3</sup>, Antonio Benítez-Burraco<sup>1</sup>

<sup>1</sup>Departamento de Lengua Española, Lingüística y Teoría de la Literatura (Lingüística), Universidad de Sevilla, Spain

<sup>2</sup>ClicAsia, Centre d'Estudis Orientals, Spain

<sup>3</sup> Departament de Filologia Catalana, Universitat Autònoma de Barcelona, Spain

## Abstract

It is still debated how many basic color terms (BCTs) Mandarin has and how they are used compared to Western languages like Spanish. For clarifying this, we analysed the performance of 21 Mandarin speakers and 21 European Spanish speakers in two related tasks: the list task and the color naming task. Our results suggest that Mandarin has 9 BCTs in line with Gao and Sutrop (2014), while Spanish has 11 BCTs, in line with Lillo et al. (2007). We further found that Mandarin speakers use some of the common BCTs less consistently than Spanish speakers. Overall, both languages partition the color space similarly. Our findings reinforce the view that there are universal constraints on color naming, that are compatible with subtle cross-cultural differences in how colors are used.

Keywords: basic color terms, Mandarin, Spanish, list task, color naming, cross-cultural variation

## Introduction

Despite numerous prior studies, important questions persist about basic color terms (BCTs) in the world's languages, including their exact number in many languages. According to Berlin and Kay (1969), Mandarin is a stage-V language with six BCTs. However, subsequent research using corpora and perceptual-cognitive methods has increased this number to nine (Gao and Sutrop, 2014), or perhaps eleven (Hsieh et al., 2020). According to Sun and Chen (2018), this uncertainty stems from a lexical problem: compared to other languages, Mandarin has more synonyms for each color category, so that basic color categories are referred to in more diverse (and thus inconsistent) ways (see also Hsieh et al., 2020 on dialectal differences). A second pending issue is to determine whether BCTs in Mandarin are distributed across the spectrum like in other languages, as already found in several many pairs on non-related languages, such as English vs. Japanese (Kuriki et al., 2017). In this paper, we contribute to this debate around the exact number and use of BCTs in Mandarin. We use Spanish for comparisons.

## Method

Two experiments were conducted: the color list task and the color naming task (Davies & Corbett, 1994). For the color list task, we recruited 19 native

speakers of Mandarin from Mainland China ( $F = 10$ , age range 23-30, mean age = 27.2,  $SD = 2.3$ ) and 19 native speakers of European Spanish ( $F = 9$ , age range 21-44, mean age = 25.5,  $SD = 5.0$ ). In this task, participants were asked to list and write down all the color names from their native language they could remember. The maximum time allowed for this was 5 minutes. For the color naming task, we included 2 additional participants, up to a total of 21 Mandarin speakers ( $F = 11$ , age range 20-30, mean age = 26.7,  $SD = 2.8$ ) and 21 European Spanish speakers ( $F = 10$ , age range 21-44, mean age = 25.7,  $SD = 4.8$ ). In this task, participants were asked to name the 330 standard chips from the Munsell color chart, as used in the World Color Survey ([www.icsi.berkeley.edu/wcs/](http://www.icsi.berkeley.edu/wcs/)), in an unconstrained way. Chips were presented in a random order, which changed from one subject to another. It took approximately 30-40 minutes for each participant to complete the task. Subjects with color blindness or vision problems were excluded from both tasks.

### Data treatment and analysis

In the list task, for each of the colors that were mentioned by at least 5 subjects, we calculated its cognitive Saliency index (S) (Sutrop, 2001), which measures the perceptive/cognitive bias that predisposes individuals to focus on it. In the naming task, for each of the dominant terms (i.e. terms used for at least half of the participants for a given tile), we calculated its Specificity Index (SI) (Gao & Sutrop, 2014), which measures the consensus among subjects for using it. To establish the number and the identity of the BCTs in each language, we relied on the results from both tasks (following the minimal criteria proposed by Gao and Sutrop, 2014), as well as the criteria for BCTs proposed by Berlin and Kay (1969)). Finally, in order to clarify how BCTs are used in each language, we used a multidimensional scaling to assess individual differences in naming for each BCT. We also used the Kernel Density Estimation (KDE) to determine the distribution of BCTs across the Munsell color chart.

### Results

In the list task, Mandarin speakers generated 78 different color terms ( $M = 27.3$ ), most of which were compounds and modified terms (e.g. *anhong* ‘dark red’), while Spanish speakers generated 81 different color terms ( $M = 18.7$ ), with most of them being single-word terms (e.g. *teja* ‘brick red’). The highest cognitive Saliency values were found for *hong* ‘red’ ( $S = .39$ ), *huang* ‘yellow’ ( $S = .23$ ) and *lan* ‘blue’ ( $S = .21$ ) in Mandarin, and for *azul* ‘blue’ ( $S = .33$ ), *rojo* ‘red’ ( $S = .31$ ) and *verde* ‘green’ ( $S = .22$ ) in Spanish. In the color naming task, Mandarin gave more different color names to each chip ( $M = 9.69$ ,  $SD = 3.78$ ) than Spanish speakers ( $M = 8.34$ ,  $SD = 3.75$ ) and this difference was significant ( $t(658) = 4.743$ ,  $d = .36$ ,  $p < .001$ ). However, the highest SI value was found for ‘black’ in both languages. Overall, our results suggest that Mandarin has 9 BCTs: *hei* ‘black’, *bai* ‘white’, *zi* ‘purple’, *liu* ‘green’, *huang* ‘yellow’, *lan* ‘blue’, *hui*

‘gray’, *fen* ‘pink’ and *hong* ‘red’, whereas Spanish has 11 BCTs: *negro* ‘black’, *rojo* ‘red’, *naranja* ‘orange’, *blanco* ‘white’, *amarillo* ‘yellow’, *verde* ‘green’, *morado* ‘purple’, *rosa* ‘pink’, *gris* ‘gray’, *azul* ‘blue’ and *marrón* ‘brown’. Finally, the multidimensional scaling revealed that Mandarin speakers showed more individual differences than Spanish speakers when using ‘grey’ ( $F(1,40) = 12.61, p = .001$ ), ‘red’ ( $F(1,40) = 4.29, p = .045$ ), ‘blue’ ( $F(1,40) = 6.15, p = .02$ ) and ‘purple’ ( $F(1,40) = 11.39, p = .002$ ). By contrast, Spanish-speakers showed more variability when referring to ‘red’, ‘blue’, and ‘purple’, as illustrated by the additional peaks in the density plots (Fig. 1): one additional peak (*burdeos* ‘bordeaux’) for the ‘red’ category; two additional peaks (*celeste* ‘light blue’ and *turquesa* ‘turquoise’) for the ‘blue’ category, and two additional peaks for the ‘purple’ category (*violeta* ‘violet’ and *lila* ‘lilac’). Despite these differences, both languages partition the color space similarly (see Fig. 2).

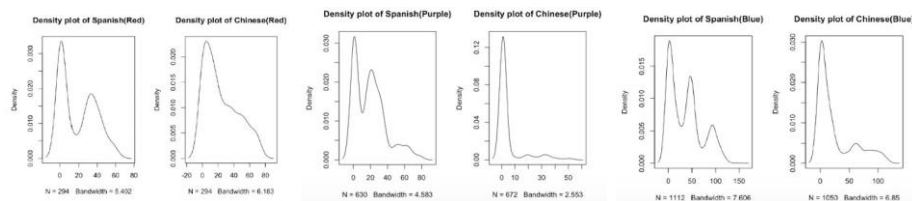


Figure 1. Density plots for red, purple and blue category in Spanish and Mandarin Chinese.

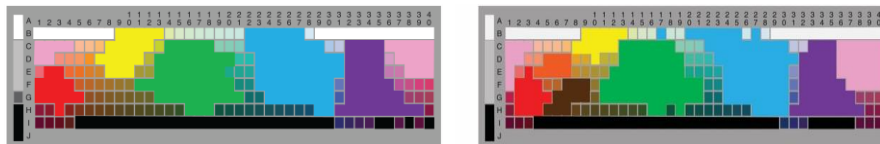


Figure 2. Map of the distribution of basic color categories for Mandarin speakers (left) and Spanish speakers (right).

## Discussion and conclusion

Conway et al. (2020) have recently suggested that the BCTs in the Berlin-Kay scheme should be clustered according to their information content and communicative efficiency: warm colors (red, orange, yellow, brown), intermediate colors (purple, pink) and cool colors (blue, green), with cool colors being the less informative and efficient. Nonetheless, in our sample, we have found an opposite trend, with *hong* ‘red’, *huang* ‘yellow’ and *lan* ‘blue’ in Mandarin, and *azul* ‘blue’, *rojo* ‘red’ and *verde* ‘green’ in Spanish, being the most salient colors. At the same time, we have found that ‘black’ is used similarly in both languages, which parallels previous research (e.g. Ratliff et al., 2010), seemingly because blacks are prevalent in natural images.

Regarding the exact number of BCTs in both languages, we suggest that Mandarin has 9 BCTs, in line with Gao and Sutrop (2014). Contrary to some

other studies (e.g. Sun & Chen, 2018; or Hsieh et al., 2020), we found no evidence of *cheng* ‘orange’ and *zong* ‘brown’ being dominant colors in Mandarin. This can be due to methodological concerns (we relied on a free-choice paradigm, but not on a fixed-choice paradigm), or to sampling concerns (we tested Continental Mandarin speakers, but not Taiwanese Mandarin speakers). With regards to Spanish, our results support the view that it has 11 BCTs, in line with Lillo et al. (2007). Interestingly, in our experiment, the term *celeste* ‘light blue’ obtained a high S value ( $S=0.07$ ) and possessed an SI value ( $SI=0.14$ ). This suggests that this term might be emerging as a 12th BCT in Spanish. Likewise, we found that Spanish-speakers used more than one term for ‘purple’ and for ‘blue’, with these two terms, *violeta* ‘violet’ and *celeste* ‘light blue’, respectively, showing S values that were similar to the BCT *marrón* ‘brown’. This suggests that additional color categories might be emerging in Spanish, in the line of Levinson’s (2000) emergence hypothesis.

Overall, despite some differences between Mandarin and Spanish in the variability of BCT use, both languages partition the color space similarly. These are suggestive of some universal constraints on color naming, probably due to the common psychophysics of human color perception.

## References

- Berlin, B., Kay, P. 1969. Basic color terms. Berkeley: University of California Press.
- Conway, B.R., Ratnasingam, S., Jara-Ettinger, J., Futrell, R., Gibson, E. 2020. Communication efficiency of color naming across languages provides a new framework for the evolution of color terms. *Cognition*, 195, 104086.
- Davies, I, Corbett, G. 1994. The basic color terms of Russian. *Linguistics*, 32(1). 65–89.
- Gao, J., Sutrop, U. 2014. The BCTs of Mandarin Chinese: A theory-driven experimental study. *Studies in Language*, 38(2), 335-359.
- Hsieh, T.J.T., Kuriki, I., Chen, I. P., Muto, Y., Tokunaga, R., Shioiri, S. 2020. Basic color categories in Mandarin Chinese revealed by cluster analysis. *Journal of vision*, 20(12), 6.
- Kuriki, I., Lange, R., Muto, Y., Brown, A.M., Fukuda, K., Tokunaga, R., ..., Shioiri, S. 2017. The modern Japanese color lexicon. *Journal of vision*, 17(3), 1-1.
- Levinson, S.C. 2000. Yéí Dnye and the theory of BCTs. *Journal of Linguistic Anthropology*, 10(1), 3-55.
- Lillo, J., Moreira, H., Vitini, I., Martín, J. 2007. Locating Basic Spanish Colour Categories in CIE L\* u\* v\* Space: Identification, Lightness Segregation and Correspondence with English Equivalents. *Psicológica: International Journal of Methodology and Experimental Psychology*, 28(1), 21-54.
- Ratliff, C. P., Borghuis, B. G., Kao, Y. H., Sterling, P., Balasubramanian, V. 2010. Retina is structured to process an excess of darkness in natural scenes. *Proceedings of the National Academy of Sciences*, 107(40), 17368-17373.
- Sutrop, U. 2001. List task and a cognitive salience index. *Field methods*, 13(3), 263-276.
- Sun, V. C., Chen, C. C. 2018. Basic color categories and Mandarin Chinese color terms. *PLoS ONE*, 13(11), e0206699.



# Experimental investigation of Mandarin lexical tone production

Yuyu Zeng<sup>1,2</sup>, Robert Fiorentino<sup>1</sup>, Jie Zhang<sup>1</sup>

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

<sup>2</sup>Department of Communication Sciences and Disorders, University of Wisconsin-Madison, USA

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Via two picture naming experiments, we investigated the role of lexical tone in Mandarin spoken word production. In both experiments, target words are monosyllabic, and naming latency is the dependent variable. Experiment 1 (pictures as distractors) examines whether lexical tone is relevant during lexical selection. Trials with exact homophone distractors (zhū vs. zhǔ) were named significantly more slowly than trials with simple homophone distractors (zhū vs. zhù), suggesting that lexical tone mediates lexical selection. Experiment 2 (pseudo-word transcriptions as distractors) examines the relative timing of lexical tone encoding. It was found that: (1) lexical tone encoding is no later than syllable encoding and precedes the encoding of within-syllable segments; (2) lexical tone encoding takes less time than syllable encoding.

Keywords: spoken word production, lexical tone, Mandarin Chinese, lexical selection, form encoding.

## Introduction

The spoken word production process has four identifiable stages: conceptualization, lexical selection, form encoding, and articulation. Much has been learned about the spoken word production process by studying Indo-European languages such as English. Mandarin Chinese, a Sino-Tibetan language, differs from Indo-European languages in many crucial aspects. Research into the Chinese language has yielded valuable insights regarding how language-specific characteristics shape spoken word production. One well-known example is the syllable-retrieval hypothesis of Chinese (O'Seaghdha et al., 2010), which states that form encoding in Chinese spoken word production starts with accessing the syllable, followed by the specification of within-syllable segments; in contrast, in languages such as English, form encoding starts with accessing the segments, which are then assembled into a syllable.

The use of contrastive lexical tone, whose primary acoustic correlate is the pattern of fundamental frequency (F0) realization over a syllable, is one salient characteristic of Chinese. However, little is known regarding the role of lexical tone in Chinese spoken word production. The present study addresses this gap

by examining the role of lexical tone in Mandarin spoken word production, focusing on the lexical selection stage and the form encoding stage.

### Experiment 1: lexical tone during lexical selection

Experiment 1 investigates whether lexical tone is relevant during lexical selection in Mandarin spoken word production. If lexical tone is used during lexical selection, exact homophones (zhū vs. zhū) and simple homophones (zhū vs. zhù) should behave differently.

#### Methods

Thirty-four Mandarin speakers (8 males; mean age = 27.25, SD = 5.52) participated in Experiment 1.

Figure 1 illustrates example items in Experiment 1. The targets are 82 monosyllabic Mandarin words. Each target is paired with an exact homophone distractor, a simple homophone distractor, and two unrelated distractors, corresponding to the three conditions in Experiment 1: Exact Homophone, Simple Homophone, and Unrelated.

Experiment 1 consists of three sessions, which took place over three days over the internet. Session 1 familiarized the participants with the stimuli. Session 2 verified participants' mastery of the stimuli and collected control measurements for statistical modelling. The main picture naming task using picture distractors (Picture-picture Interference) occurred in Session 3, where each unique target-distractor combination was repeated twice.

The naming latency of each trial was manually obtained from the audio recordings of the Picture-picture Interference task. Mixed-effects modelling was conducted using the *lme4* package (version 1.1-27.1) in R (version 4.1.1). The dependent variable was the log-transformed naming latency, and the independent variables include Condition (also included as a random slope) and five control measurements (Sassenhagen & Alday, 2016). The control measurements are not detailed in this paper due to space limitations. The best model was automatically selected using the *buildmer* package (version 2.1). The raw data and analysis scripts can be accessed from this link.

The naming latency of each trial was manually obtained from the audio recordings of the Picture-picture Interference task. Mixed-effects modelling was conducted using the *lme4* package (version 1.1-27.1) in R (version 4.1.1). The dependent variable was the log-transformed naming latency, and the independent variables include Condition (also included as a random slope) and five control measurements (Sassenhagen & Alday, 2016). The control measurements are not detailed in this paper due to space limitations. The best model was automatically selected using the *buildmer* package (version 2.1). The raw data and analysis scripts can be accessed from this link.

#### Results

Condition was significant in the selected model ( $p = 0.011$ ). The Exact Homophone condition (914.60 ms) yielded a longer naming latency than the Simple Homophone condition (908.14 ms;  $p = 0.061$ ). There was also a significant Condition by Stimulus Repetition interaction ( $p = 0.038$ ). In

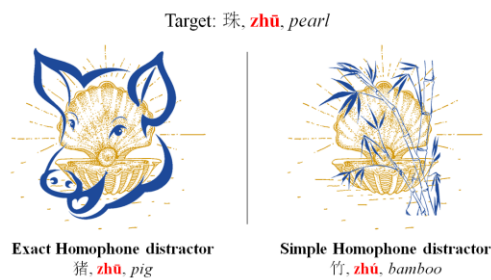


Figure 1. Experiment 1 example items.

Repetition 1, the Exact Homophone condition (963.39 ms) was named significantly more slowly than the Simple Homophone condition (945.66 ms;  $p = 0.001$ ), but this difference was not significant in Repetition 2 ( $p = 0.60$ ).

### Discussion

Experiment 1 revealed evidence for the differentiation of exact homophones and simple homophones, suggesting that lexical tone mediates lexical selection in Mandarin spoken word production. During lexical selection, the competition between exact homophones is stronger than between simple homophones, resulting in longer naming latency in the Exact Homophone condition.

The difference between the Exact Homophone condition and the Simple homophone Condition was significant in Repetition 1 but not in Repetition 2. This is best explained by continuous information flow from lexical selection to form encoding. The phonological overlap between a target and its distractor in the Exact Homophone condition is larger. The facilitation from the form encoding stage (phonological facilitation, henceforth) is thus expected to be stronger in the Exact Homophone condition. In Repetition 1, participants were less familiar with the stimuli, and the phonological activation of the distractors was relatively weak, so the results were dominated by lexical competition. In Repetition 2, participants were more familiar with the stimuli, and the corresponding phonological activation of distractors was higher. Consequently, phonological facilitation overcame lexical competition, leading to no difference between the Exact Homophone and Simple Homophone conditions.

### Experiment 2: lexical tone during form encoding

Experiment 2 examines the relative timing of lexical tone encoding by comparing it to syllable encoding, which has been found to precede segmental encoding in previous studies (e.g., O'Seaghdha et al., 2010).

### Methods

Twenty-six Mandarin speakers (11 males; mean age = 27.90, SD = 4.84) participated in Experiment 2. The distractors in Experiment 2 are sound transcriptions of Mandarin pseudo-words. This way, the distractors primarily influence the form encoding stage of target naming. The targets are 108 monosyllabic Mandarin words (e.g., 麦, *mài*, *wheat*), each paired with three types of distractors: (1) Syllable-related (*māi*); (2) Tone-related (*sùn*); (3) Unrelated (*sún*). There are three levels of SOA (stimulus-onset asynchrony) regarding the timing of a target and its distractor: (1) -100 ms, where a distractor precedes its target by 100 ms; (2) 0 ms, where a target and its distractor appear simultaneously; (3) +100, where a target precedes its distractor by 100 ms. The procedure, data processing, and statistical analysis were similar to Experiment 1.

## Results

The Syllable-related condition was named significantly faster than the Unrelated condition at all SOA levels ( $p$ -values  $< 0.028$ ). The Tone-related condition (886.43 ms) was named significantly faster than the Unrelated condition (889.21 ms) at the -100 ms SOA level ( $p = 0.032$ ), and marginally faster than the Unrelated condition at the 0 ms SOA level (921.08 ms vs. 921.74 ms;  $p = 0.071$ ). The Tone-related condition tended to be named more slowly than the Unrelated condition at the +100 ms SOA level, but the difference was not significant (927.19 ms vs. 924.19 ms;  $p = 0.10$ ).

## Discussion

The Experiment 2 results suggest that lexical tone encoding is no later than syllable encoding, as both Syllable-related and Tone-related conditions were named faster than the Unrelated condition at the -100 ms SOA level. This finding also implies that lexical tone encoding precedes segmental encoding (cf., O'Seaghdha et al., 2010). Moreover, all SOA levels show phonological facilitation in the Syllable-related condition, suggesting that lexical tone encoding takes less time than syllable encoding.

## Conclusion

Our study demonstrates that lexical tone participates in lexical selection and form encoding in Mandarin spoken word production. First, lexical tone mediates lexical selection, so the competition between exact homophones is stronger than between simple homophones (Experiment 1). Moreover, after being utilized in lexical selection, lexical tone is also encoded early during form encoding, potentially concurrent with the access of the syllable, followed by the encoding of within-syllable segments (Experiment 2). The Experiment 2 data also suggests lexical tone encoding takes less time than syllable encoding.

## Acknowledgments

This work was supported by a Linguistics Research Scholarship awarded to Yuyu Zeng from the Department of Linguistics at the University of Kansas. This work also received support from the Summer Research Scholarship at the University of Kansas.

## References

- O'Seaghdha, P. G., Chen, J.-Y., Chen, T.-M. 2010. Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, 115(2), 282–302.
- Sassenhagen, J., Alday, P.M. 2016. A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language*, 162, 42–45.

# Classifier versus taxonomic relations in a similarity judgement task

Jiahuan Zhang

Department of Theoretical and Applied Linguistics, The University of Cambridge, UK  
Faculty of Education, The University of Hong Kong, Hong Kong SAR, China

<https://doi.org/10.36505/ExLing-2022/13>

## Abstract

Compared with non-classifier speakers, classifier speakers often showed more sensitivity to classifier-oriented parameters (also referred to as conceptual saliency) such as animacy, shape and function – the purported “classifier effect”. This study thus further testified the classifier effect through a similarity judgement task. Chinese speakers (N=41) rated significantly lower than English speakers (N=41) in taxonomic pairs, but not in thematic, classifier or filler pairs. However, both groups consistently rated highest in the thematic pairs, followed by the taxonomic, classifier and filler pairs. Chinese speakers also spent longer time than their English counterparts in each pair condition. A subset analysis of taxonomic pairs indicated a varying degree of disparity on conceptual saliency. Findings together suggested an implicit classifier effect from Chinese speakers, but classifier relation itself was not employed as a predominant parameter for object perception.

Keywords: Chinese classifiers, similarity judgement, conceptual structure, object categorisation, conceptual saliency

## Introduction

Classifiers are grammatical category. As a reference-tracking device, Chinese classifiers are unique in their semantic association with the internal property of head nouns, which leads to the discussion of the relationship between language and cognition. Much research adduced evidence to the purported classifier effect from Chinese speakers through a range of conceptual categorisation tasks (e.g., Gao & Malt, 2009; Saalbach & Imai, 2007; Speed et al., 2021). However, classifier types (e.g., shape, animacy, and function) as the representative of conceptual saliency can also affect categorisation decision (e.g., Wang & Zhang, 2014), which have been limitedly considered insofar. This study thus aims to further scrutinise whether the classifier effect is modulated by classifier types.

## Methodology

### Participants

Participants were 41 native Chinese speakers (female=29,  $M_{age}=24.16$ ,  $SD_{age}=2.58$ ) and 41 native English speakers (female=26,  $M_{age}=22.31$ ,  $SD_{age}=2.71$ ), all were university students.

## Materials

The majority of the targets and items were selected from previous classifier-related similarity judgement tasks (Gao & Malt, 2009; Saalbach & Imai, 2007), with one modification being the counterbalance of classifier types, i.e., shape, animacy, and function across feature conditions. Seventy-two trials were constructed, including 18 quintuplets of objects. Each quintuplet consisted of one target and four objects representing each of the four features to the target. The first feature shares the same Chinese classifier but does not overlap with taxonomic or thematic features (e.g., river-scarf). The second feature is the taxonomically correlated item with the target (e.g., river-sea), and the third one is thematically correlated (e.g., river-water). The last one is the filler pairs. All trials in each version were automatically randomised across the three blocks and all participants by Gorilla.

## Procedure

In each trial of the task, participants were shown an image (with its written word in red shown beneath) and a written word of another item simultaneously. A Likert-scale (1 for strongly dissimilar and 7 to strongly similar) was presented on the lower half of the screen along with the test item. Participants were asked to judge the similarity between the two items. To avoid confirmation bias, participants were told that there was no clear definition for “similarity”, so that they could not follow any prescribed rules and have to define the concept themselves. Figure 1a and 1b illustrated an example for the English and Chinese version, respectively. Reaction times (RTs) on each trial were automatically recorded. Participants were encouraged to respond as quickly and as accurately as possible, without time limit imposed.



Figure 1a. English example.



Figure 1b. Chinese example.

## Results

### The analysis of responses & RTs

A mixed-effects OLR model was fit to the data, formula: response  $\sim$  L1\*feature+(1|participant)+(1|item), Hess=TRUE. Figure 2a demonstrates the results. There was a significant interaction between L1 and feature. The

drop from taxonomic to classifier pairs was less by 1.022 ( $p < 0.001$ ) with Chinese group in comparison to the English one. Referring to Chinese group as the base level, the estimated  $\beta$  for English group was significantly positive at 1.266 ( $p < 0.001$ ) in taxonomic pairs. Chinese and English groups gave comparable ratings for other feature conditions.

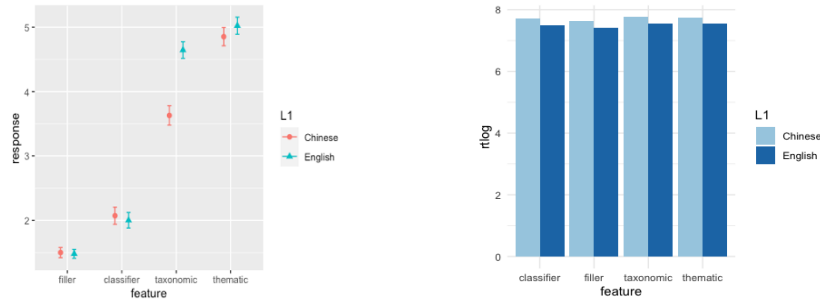


Figure 2a. Responses across L1 and features.

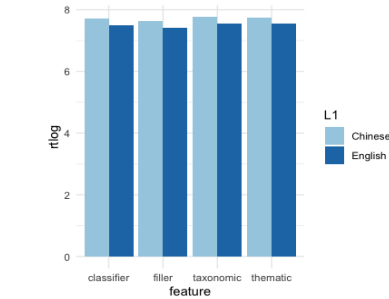


Figure 2b. RTs (ms) across L1 and feature.

RTs were log transformed to ensure normality. A linear mixed-effects model was fitted, formula:  $\text{lmer}(\text{rtlog} \sim \text{L1} * \text{feature} + (1 | \text{participant}) + (1 | \text{item}))$ . Overall, there was no interaction (Figure 2b). Chinese group needed significantly longer time than the English group on each critical feature condition: classifier ( $\beta = 0.204, p < 0.05$ ); taxonomic ( $\beta = 0.194, p < 0.05$ ) and thematic ( $\beta = 0.191, p < 0.05$ ).

### Subset analysis – the analysis of taxonomic pairs

There was a significant interaction between L1 and type ( $\beta = -0.830, p < 0.05$ ). Pairwise comparisons showed that the English group rated significantly higher in each type than the Chinese group, and the significance levels are: animacy ( $\beta = 2.254, p < 0.001$ ), followed by function ( $\beta = 1.424, p < 0.01$ ), and shape ( $\beta = 1.251, p < 0.05$ ), respectively.

### Discussion and conclusions

In contrast with many previous studies (e.g., Schmitt & Zhang, 1998), the present study did not obtain the classifier effect. One possibility could be the research design. This study deliberately paired classifier items across the taxonomically superordinate boundary (cf. Saalbach & Imai, 2012), such as *river-scarf* for classifier *tiao2*, to avoid potential bias on taxonomic categorisation. Another explanation is that incorporating taxonomic and thematic pairs probably skewed the judgements towards apparent semantic relation, which is different from when classifier pairs were only pitted against with fillers (see Schmitt & Zhang, 1998). Still, what can be rest assured is that the classifier-related feature was noticeable to English speakers although not exposed to

Chinese before (see also Saalbach & Imai, 2012; Speed et al., 2021). These together support that classifier relation is reflective of universal conceptual structure, not linguistically restricted (Speed et al., 2021).

Intriguingly, Chinese speakers rated significantly lower in taxonomic pairs compared with English speakers. My interpretation is that the habitual use of classifiers may have underlyingly formulated a more fine-grained scale for taxonomic categorisation, in line with the hypothesis in Wang and Zhang (2004). The longer RTs may also account that the Chinese speakers showed more sensitivity to processing object categorisation because of the use of classifiers.

The group differences in taxonomic pairs potentially offered us a unique window to discuss the classifier type effect underpinning by conceptual saliency. The more salient features, reversely, attracted a lower rating because of a higher level of corresponding sensitivity. The varying magnitude of significances showed convergence with Chinese classifier acquisition studies (e.g., Zhang, Gnevsheva 2022) and cognitive research (e.g., Gao, Malt 2009).

Taken together, the above findings suggest that Chinese speakers did not capitalise on classifier categorisation rule in their object conceptual reasoning, at least not in a moment-to-moment fashion (Saalbach, Imai 2007, Speed et al. 2021). Synthesising the discussion, the indirect, elusive classifier effect obtained in this study is concluded as an “implicit classifier effect”.

## Acknowledgements

I would like to express my very deep gratitude to Professor John Williams for his valuable and constructive suggestions on the conceptualisation of this research work.

## References

- Gao, M. Y., Malt, B. C. 2009. Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, 24, 1124-1179.
- Saalbach, H., Imai, M. (2007). The scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General*, 136, 485-501. doi:10.1037/0096-3445.136.3.485
- Speed, L. J., Chen, J., Huettig, F., Majid, A. 2021. Classifier categories reflect but do not affect conceptual organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(4), 625-640.
- Wang, R., Zhang, C. 2014. Effect of classifier system on object similarity judgement: A cross-linguistic study. *Journal of Chinese Linguistics*, 42(1), 188-217.
- Zhang, J., Gnevsheva, K. (2022). The effects of L1, task, and classifier type in Chinese-L2 learners' use of classifiers. *Chinese as a Second Language Research*. <https://doi.org/10.1515/caslar-2022-0002>
- Zhang, S., Schmitt, B. 1998. Language-dependent classification: The mental representation of classifiers in cognition, memory, and ad evaluations. *Journal of Experimental Psychology: Applied*, 4, 375-385. doi:10.1037/1076-898X.4.4.375.



## Index of names

---

- Abeillé, A.....45, 105  
Aldhafeeri, M.....5  
Aldholmi, Y .....5  
Alenazi, A.....9  
Alqahtani, Gh.....5  
Amorim, A.P. .... 97  
Anderson, T..... 13  
Andreou, G.....17, 153  
Augusto, M. .... 37  
Bajo, T. .... 145  
Ballier, N. .... 117  
Belova, D. .... 21  
Bencini, G. .... 57  
Benítez-Burraco, A.  
..... 169  
Bezerra, G.B. .... 97  
Bilea, S. .... 25  
Binos, P. .... 29  
Bleotu, A.C. .... 25  
Bogunović, I. .... 121  
Bradlow, A.R. ....1  
Castro, C..... 97  
Chartomatsidou, E. 17  
Chernova, D. .... 33  
Chorai, I. .... 45  
Corrêa, L. .... 37  
Ćoso, B. .... 121  
Cresti, E..... 41  
da Cunha, Y..... 45  
Davydova, V..... 49  
Evgrafova, K. .... 53  
Facipieri, Ch..... 57  
Fiorentino, R. .... 173  
Gabaj, M..... 61  
Gibson, E..... 105  
Giovannini, A.M. .... 65  
Gobl, Ch..... 65  
Goertz, G..... 13  
Hammarlin, M.M. ... 81  
Hellmuth, S. ....9  
Hemforth, B. .... 105  
Ivanova, A..... 69  
Jaeger, T.F..... 141  
Jaeger, T.P..... 125  
Kamp, V..... 73  
Karpava, S..... 77  
Kokkinakis, D. .... 81  
Konovalova, A. .... 85  
Kuvač Kraljević, E. 61  
Lindemann, S.-I. .... 89  
Liu, Y. .... 93  
Lourenço-Gomes,  
M.C. .... 97  
Lowther, T.S. .... 101  
Luo, Sh. .... 41  
Macizo, P..... 145  
Mao, R. .... 105  
Martin, Ph. .... 109  
Matei, M. .... 89  
McCabe, C. .... 113  
Moneglia, M..... 41  
Namdarzadeh, B. .. 117  
Ní Chasaide, A. .... 65  
Niculescu-Gorpin,  
A.G. .... 149  
Novozhilov, A.....33  
Pan, D..... 129  
Panunzi, A.....41  
Pavlinušić Vilus, E.121  
Pélissier, M.....73  
Persson, A..... 125  
Petinou, K.....29  
Radu-Bejenaru, C.A.  
..... 149  
Rodrigues, E. ....37  
Seeber, K..... 129  
Shamina, E.....49  
Shvaley, N.....53  
Slioussar, N.....33  
Slioussar, N.....69  
Sokolova, N.....53  
Studenikina, K..... 133  
Tallon, K. .... 137  
Togato, G..... 145  
Vann, M.....57  
Vasileanu, M. .... 149  
Vasileiou, K. .... 153  
Weber, T. .... 157  
Wei, Y..... 161  
Wei, Z..... 165  
Xu, M..... 169  
Zeng, Y..... 173  
Zhang, J.....173, 177  
Zhu, J..... 169  
Zubov, V.....85



ExLing 2022

Ebook ISSN: 2529-1092

Ebook ISBN: