ExLing 2024 Paris

Proceedings of 15th International Conference of Experimental Linguistics

23-25 October 2024
Université Paris Cité
Paris, France

Edited by Antonis Botinis

INTERNATIONAL SOCIETY OF EXPERIMENTAL LINGUISTICS

International Society of Experimental Linguistics

ExLing 2024

Proceedings of 15th International Conference of Experimental Linguistics

Université Paris Cité

23-25 October 2024
Paris, France

ExLing 2024
Proceedings of 15th International Conference of Experimental Linguistics

Published by ExLing Society
Electronic edition
ExLing 2024 Paris
Athens, Greece
ISSN: 2529-1092
ISBN: 978-618-84585-5-0
DOI: 10.36505/ExLing-2024/15

Foreword

Welcome to the 15th International Conference of Experimental Linguistics (ExLing 2024 Paris). This year, we are delighted to bring you a hybrid conference experience, enabling us to meet in person thanks to the generous hospitality of Université Paris Cité. At the same time, we can revel in the vibrancy of Parisian life and the richness of French culture.

As an international forum spanning generations of linguists, our Society remains dedicated to advancing the study of language through the development and application of experimental methods in linguistics. We warmly invite both emerging and established researchers to join us in exploring and discussing the latest advancements in linguistic research and experimental methodologies.

ExLing's journey began in 2006 in Athens, with the inaugural ExLing Conference, originally an International Speech Communication Association (ISCA) ExLing Workshop. Since then, the conference has been hosted in remarkable cities such as Paris, Saint Petersburg, Heraklion, and Lisbon, becoming an annual fixture since 2019 as the International Conference of Experimental Linguistics. This year, we are thrilled to return to the enchanting city of Paris for ExLing 2024.

This volume features the proceedings of ExLing 2024 Paris. Alongside the main conference, we are proud to present three special sessions that span dynamic and diverse research areas, focusing on Experimental Phonetics, Language Education, and Language Pathology.

In addition, we are excited to launch the Language Application Exhibition, a new initiative of ExLing 2024 Paris. This exhibition aims to unify various applications of language technologies, fostering interdisciplinary innovations in key areas of technological language aids.

We extend our heartfelt gratitude to all participants of ExLing 2024 Paris, as well as our keynote speakers, Mirjam Ernestus, Valerie Shafer, and Jenny Thomson. A special thanks to our colleagues on the International Advisory Committee, along with the Review and Organisation Committees, for their invaluable contributions to the success of this conference.

Antonis Botinis
ExLing Society

Contents

Chinese speakers' production of russian word stress

Marina Agafonova, Anastasia Mishina
Peter the Great St. Petersburg Polytechnic University, Russian Federation

https://doi.org/10.36505/ExLing-2024/15/0001/000626

Abstract

From the perspective of phonetic features, the Russian language and Chinese language vary greatly, due to which it is impossible to avoid an accent when the Chinese learn the Russian language. This article is devoted to the study of phonetic features of erroneous realizations of the accentual-rhythmic pattern of a Russian word by native Chinese speakers. The discrepancy between the rules for placing the word stress is the main reason for the emergence of stable interference, which has a negative impact on the communication process in most cases. The study investigated the interaction between two languages: Russian, which is an accent language, and Chinese, which is a non-accent language.

Keywords: Russian language, Chinese language, accent, word stress

Realization and perception of word accent in the speech of Chinese people in the Russian language

Speakers, material and methodology

Today, studies of the Chinese language and Russian-Chinese interference are actively conducted by Russian and Chinese researchers. A phenomenon of the Russian-Chinese interference as a whole, as well as manifestations of interference at the segmental level are investigated in a number of papers. However, studies of stress features are based on theoretical material and supported by little comparative material (D.Wang & S.Lee 2021, Oglezneva 2007, Yaming 1996). Each native speaker of some or other language has a notion of the integral appearance of a word, which implies both phonemic compatibility and accent-rhythmic organization. Accordingly, a person learning a foreign language and who at the same time is not familiar with the phonetic features of this language will try to use the accent-rhythmic models of his native language, which are already familiar to him.

The text "It was a quiet grey evening" created in the phonetics laboratory of the St. Petersburg State University was selected for analysis in the study (Smirnova & Khitrov 2013). The phonetically representative text "It was a quiet grey evening..." comprises 456 words (111 monosyllables, 146 disyllables, 96 trisyllables and 65 polysyllables). Eight second- and third-year Chinese students of the Linguistics Department of the Peter the Great St. Petersburg Polytechnic

University, who are native speakers of Mandarin with a Beijing dialect (four men and four women, the period of language study is 2-3 years, the level of language proficiency is A2- B1) were enrolled in the experiment. Recordings of 64 minutes and 22 seconds duration, comprising 3,648 words, were analysed. The recordings analysis employed auditory and instrumental methods. The analysis of this study has been conducted using PRAAT software, which identified four main characteristics that were considered during the analysis of the recorded speech of respondents: the duration of the stressed vowel, the intensity of the stressed vowel, vowel reduction and the presence or absence of a vowel insertion. The particular cases of incorrect realization of the accent-rhythmic pattern of a Russian word by Chinese students is provided below.

Phonetic characteristics of mistakes

The duration of a stressed vowel is one of the most important features of a stressed syllable in the Russian language (Shcherba 1983). Deviations in the realization of the stressed vowels' duration and intensity are a frequent error in the speech of native Chinese speakers. The last characteristic is the presence or absence of a vowel insertion. The study showed that a vowel insertion most often appears after occlusive-plosive or fricative consonants at the end of a word, and more often after a soft consonant. Vowel insertions occur most often when a word is pronounced slowly, because a certain overtone that resembles a vowel sound appears. Among isolated errors, the loss of a stressed syllable in some words and the placement of two stresses in one word can be emphasized.

The detailed analysis of the recordings produced the statistics giving an impression of some patterns and propensities for certain types of errors in realizing the accentual-rhythmic pattern of a word in the Russian language. First of all, it was identified how many syllables a word should have so that a native Chinese speaker made a mistake. For example, errors were made more often in three-syllable words (26.8% of erroneous realizations), when realizing the accentual-rhythmic pattern of words. For two-syllable words 14.4% , and for one-syllable words 6.3%, were realized incorrectly.

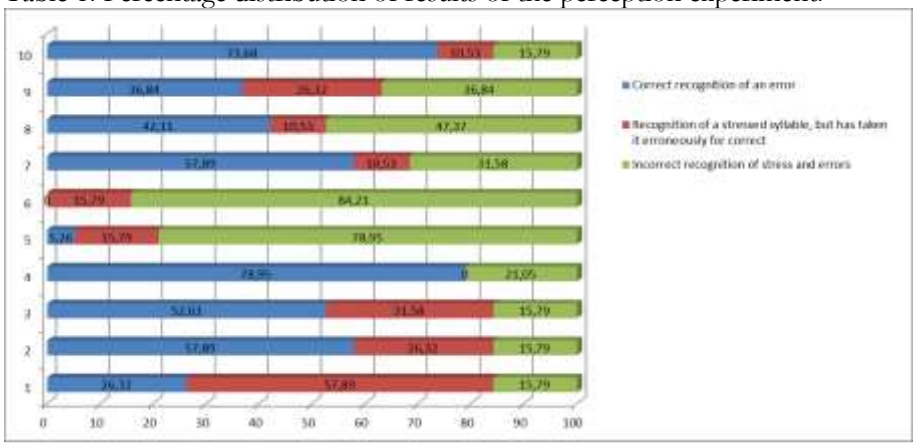
The study showed that a native Chinese speaker is most likely to shift the stress to an overstressed syllable (65.9% of errors are transferred to an overstressed syllable) when making an error in the vowel duration. If the intensity was realized incorrectly, Chinese speakers emphasized the pre-stressed syllable in 51.4% of cases and the over-stressed syllable in 48.6% of cases.

Perception experiment

An issue has arisen from the analysis of this research: whether Chinese speakers would be able to recognize a stressed syllable in a Russian word pronounced by another Chinese speaker and determine whether an error was made. A

perception experiment was conducted, during which the third-year students were asked to listen to 19 words selected from the speakers' speech recordings used in the previous experiment. The researchers selected words in which an error was made in different parameters, for example, in the realization of duration and intensity. Also, they selected words, in which such errors as double stress or the same duration and intensity for stressed and unstressed vowels were made. The results of the survey distributed the results into the categories: "Correct recognition of an error", "Recognition of a stressed syllable, but has taken it erroneously for correct", "Incorrect recognition of stress and errors".

Table 1. Percentage distribution of results of the perception experiment.



The perception experiment showed that the task, in which it was required to phonetically mark the stress in a word spoken by another native Chinese speaker, turned out to be very difficult for some respondents. In addition, when determining stress as correct or incorrect, no dependence on the realization of the duration or intensity of the stressed vowel was found, but the respondents correctly noted most errors in words that are more common in everyday speech and did not notice an error when realizing less common words. According to calculations, 44.2% of realized stimuli were correctly identified as erroneous; in all other stimuli, the stress location was determined incorrectly or the error was not recognized.

Conclusion

Taking all the aforesaid into consideration, it can be stated that the following results were obtained:

1. Native Chinese speakers mostly make errors in realizing the accentual-rhythmic pattern of words in three-syllable words (26.8%);
2. If during the realization of the accentual-rhythmic pattern of a Russian word by a native Chinese speaker an error consisting of the erroneous duration of the stressed vowel, the stress will most likely be shifted to the overstressed syllable (65.9%);
3. If during the realization of the accentual-rhythmic pattern of a Russian word by a native Chinese speaker an error consisting of the erroneous duration of the stressed vowel, the stress can be shifted equally to the pre-stressed (51.4%) and the post-stressed syllable (48.6%);
4. Native Chinese speakers can identify an error in realizing the accentual-rhythmic pattern of a Russian word in 44.2% of cases. Most likely, the listeners did not rely on phonetic correlates of word stress in the Russian language, but were guided by their own knowledge of vocabulary and the familiar phonetic appearance of the word.

Thus, we can conclude that the realization of the accent-rhythmic pattern of a word in the Russian language by native Chinese speakers has certain regularities, but due consideration should be placed on such factors as the period of language learning and the native speaker's involvement in the language environment (as shown by the results of a perception experiment, which turned out to be too simple for some listeners and too complex for others). In addition, the number of incorrect realizations may be due to knowledge of the vocabulary found in the text, or the presence or absence of homographs in the text.

References

Oglezneva, E.A. 2007. Russian-Chinese pidgin: experience of sociolinguistic description. Blagoveshchensk: AmSU Publishing House.

Smirnova, N.S. & Khitrov M.V. 2013. Phonetically representative text for fundamental and applied research of Russian speech. News of Higher Educational Institutions: Instrument manufacture, 2, 5–10.

Shcherba, L.V. 1983. Russian vowels in terms of qualitative and quantitative analysis. Leningrad: Nauka (Leningrad department).

Yaming, J. 1996. Word stress as a fragment of the ethnic type of native speakers of Russian and Chinese languages: thesis of PhD in Philology Sciences. M.

Wang, D. & Lee, S. 2021.Theoretical foundations of studying the rhythmic pattern of the Russian word in th e aspect of RFL. News of the Voronezh State Pedagogical University, 2(291), 210–219.

Disentangling typical and atypical Russian acquisition patterns: an automated approach

Katsiaryna Aharodnik[1], Isabelle Barrière[1,3], Irina A. Sekerina[1,2], Valerie L. Shafer[1]

[1]CUNY Graduate Center, US
[2]CUNY College of Staten Island, US
[3]St Elizabeth University, US

https://doi.org/10.36505/ExLing-2024/15/0002/000627

Abstract

The present study investigated a machine learning approach for disentangling Developmental Language Disorder (DLD) from typical development in Russian-speaking children. We explored the BiSLI Corpora from CHILDES and conducted two studies: 1) with monolingual Russian children with typical development (TD) versus DLD; 2) with bilingual Russian-Dutch TD children and Russian monolingual children with DLD. The efficacy of using computational features was compared to that of traditional clinical measures for diagnosis of DLD. The results revealed that computational features, such as tf-idf vectors and n-gram models, performed at or above the clinical threshold for acceptable and good performance (using the "F1" measure) and demonstrated significantly better results than the traditional language sample analysis models.

Keywords: Developmental Language Disorder, machine learning, narrative skills, Russian children, bilingualism

Introduction

Reliable identification of children with Developmental Language Disorder (DLD) remains a significant challenge for researchers and clinicians. DLD is a language disorder with a high prevalence rate of approximately 7% among English-speaking preschoolers (Tomblin et al., 1997), and its language deficits often persist beyond early childhood. The pervasive nature of this disorder is evident in its impact on various linguistic domains, affecting both comprehension and production, despite the absence of generalized cognitive impairment. DLD profiles vary across languages, with highly inflectional languages posing particular challenges for clinical identification due to the subtle nature of errors in these languages (e.g, Armon-Lotem, 2014; Kornev & Balciuniene, 2021; Leonard, 2014).

Identifying DLD among bilingual children presents an additional and distinct challenge. Previous research has highlighted significant overlaps between the error profiles of typically developing (TD) bilingual children and those with DLD. Distinguishing between differences in first or second language

production due to lower proficiency and the deficits indicative of DLD can be particularly difficult. Currently, no assessment tools are available that can accurately differentiate DLD from typical bilingual language development. To address this challenge, new approaches to bilingual assessment must be explored, taking into account the heterogeneity within the bilingual population (e.g. Blume et al., 2019) and the overlaps with clinical population.

The primary goal of the current study is to evaluate the effectiveness of using a machine learning technique for DLD identification in Russian-speaking children. A secondary goal is to identify computational linguistics features that outperform traditional Language Sample Analysis (LSA) approaches (Thordardottir, 2015). To achieve these objectives, two binary classification studies were conducted: 1) distinguishing between narratives produced by Russian-speaking 4-to-9-year-old monolingual children with typical development (TD) and those with DLD, and 2) distinguishing between 4-to-9-year-old Russian-Dutch bilingual TD children and Russian monolingual children with DLD.

The study tested two hypotheses: 1) computational features (e.g., tf-idf vectors and n-grams) used alone or in combination with additional narrative microstructure features would outperform the traditional LSA approach; 2) n-gram features (POS n-grams and word n-grams) would demonstrate superior performance compared to other features.

Methodology

The BiSLI Bilingual Corpus (Gülzow & Gagarina, 2007), available on CHIDLES (MacWhinney, 2000), served as the data source for this study. For Study 1, we analyzed data from the Monolingual (ML) Russian and ML Russian DLD corpora. For Study 2, we examined the data from the ML Russian DLD and Bilingual (BL) Russian corpora, focusing on simultaneous BLs with Russian-speaking mothers residing in the Netherlands. Two age groups were considered: a younger group (children aged 3;11 to 6;11 years), and an older group (children aged 7;0 to 9;11 years). Descriptive statistics for both studies are provided in Table 1.

We presented narrative microstructure features and computational features to the machine learning classifiers. The narrative microstructure feature set included language productivity and lexical diversity measures. The features traditionally employed for LSA and identified as markers of DLD in prior research (Botting, 2002) were defined as a reference point. These features were the mean length of utterance (MLU) in words, the number of different words (NDW), the total number of words (TNW), and the total number of utterances (TNU). In contrast, the computational features included tf-idf vectors (word-based and part-of-speech) and n-gram (word-based and part-of-speech)

features. Part-of-speech (POS) n-grams represented extended POS tags with grammatical categories (e.g., case, gender, number) mentioned within tags.

Two supervised machine learning classifiers, the Logistic Regression and Support Vector Machines classifiers were employed in the study. We used a training-validation-test split for training and testing phases. Additionally, we implemented a k-fold (k=6) cross-validation to check the stability and generalizability of the data. Leave-one-out-cross-validation (LOOCV) was also performed on the entire training set. The LOOCV output was used to calculate statistical difference between the classifier models with McNemar Test.

Table 1. Descriptive statistics for Study 1 and Study 2.

Study	Age	N participants		Mean age		Mean number of tokens per child	
		TD	DLD	TD	DLD	TD	DLD
ML	4-6	78	64	5;5	5;6	120.3	124.7
ML	7-9	60	58	8;2	8;2	170.7	160.7
BL	4-6	77	64	5;6	5;5	62.8	61.6
BL	7-9	77	58	8;1	8;2	71.3	80.8

Results

The results of both studies were interpreted in relation to diagnostic criteria employed in speech pathology research. The performance of a measure ≥ 0.8 was considered acceptable, and ≥ 0.9 as good (Plante & Vance, 1994; Spaulding et al., 2013). The results of both studies supported the initial hypotheses. For the ML study, computational features reached the threshold for clinically acceptable performance for the older age group, with word tf-idf vectors (p = 0.03389) and POS n-grams (p = 0.000355) performing significantly better than the LSA models. In the BL study, all computational features, word tf-idf vectors (p = 0.00142), word n-grams (p = 0.00865) and POS n-grams (p = 0.011117) performed significantly better than the LSA models for the younger age group.

The results of the first study demonstrated that the machine learning classifiers were able to disentangle between ML TD and DLD based on word-based features, which could be attributable to differences in pronunciation, frequency of mazes, vocabulary choices, and lexical errors among the monolingual groups. The POS n-grams demonstrated the best and most stable result for the older age group, suggesting that it was easier for the algorithms to classify the groups with more mature syntactic patterns. This is in line with the research showing differences in complex syntax for the older DLD and TD children (Reilly et al., 2004).

The high classification performance for both word and POS n-grams in the BL study could be attributed to BLs' lower proficiency in Russian. Specifically, the differences were found in the use of Subject-Verb agreement, and in morphological and morphosyntactic markers. Further error analyses should reveal whether there are additional differences in error profiles of the groups.

Conclusion

The current study validated the machine learning approach for the DLD identification for monolingual and bilingual children and highlighted its promise as a valid, and potentially automated, tool for DLD diagnosis in a highly inflectional language. The computational features outperformed a traditional LSA approach and reached clinically acceptable thresholds.

Future follow-up studies should consider including errors as feature-predictors and explore classifiers other than traditional statistical machine learning algorithms. In addition, future studies should determine whether the current approach can be applied to the diagnoses of DLD in children speaking languages other than Russian.

References

Armon-Lotem, S. 2014. Between L2 and SLI: Inflections and prepositions in the Hebrew of bilingual children with TLD and monolingual children with SLI. Journal of Child Language, 41(1), 3-33.

Botting, N. 2002. Narrative as a tool for the assessment of linguistic and pragmatic impairments. Child language teaching and therapy, 18(1), 1-21.

Blume, M.; Barrière, I.; Dye, C.; Kang, C. 2019. Challenges for the Development of Linked Open Data for Research in Multilingualism. A. Pareja, M. Blume, B.C. Lust, M.R. Melanson (Eds.) Development of Linguistic Linked Open Data Resources for Collaborative Data-Intensive Research in the Language Sciences. (185-200) Boston: MIT press.

Kornev, A.N., Balčiūnienė, I. 2021. Lexical and grammatical errors in developmentally language disordered and typically developed children: The impact of age and discourse genre. Children, 8(12), 1114.

Reilly, J., Losh, M., Bellugi, U., Wulfeck, B. 2004. "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. Brain and language, 88(2), 229-247.

Spaulding, T.J., Hosmer, S., Schechtman, C. 2013. Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. International Journal of Speech-Language Pathology, 15(5), 453-462.

Thordardottir, E. 2015. The relationship between bilingual exposure and morphosyntactic development. International journal of speech-language pathology, 17(2), 97-114.

Entropy shows: is it real speech, or a clone?

Terese Anderson[1], Grandon Goertz[2], Evan Ashworth[2]
[1]University of Chicago, US
[2]University of New Mexico, US

https://doi.org/10.36505/ExLing-2024/15/0003/000628

Abstract

This is an exploratory study that compares the conditional entropy formant values of naturally spoken words to the conditional entropy formant values of cloned vowels. It was hypothesized that cloned vowel formants, F1, F2, and F3, would have measurable and distinct differences compared to the formants of natural speech. This study shows that there are indeed variations in the entropy of Artificial Intelligence (AI) cloned vowels. These differences would be useful for forensic analyses and for distinguishing natural speech from AI generated imitations.

Keywords: entropy, cloned speech, forensics, formants, AI

Introduction

Voice cloning is a product of AI technology that creates a digital copy of a person's voice. In some cases, clones have been noted as sounding very close to or even identical to naturally spoken speech. This research compares naturally spoken English words to their cloned copies to answer the question: Do the entropy values of the formants of cloned vowels show measurable, identifiable differences compared to entropy values of the formants of natural speech?

Entropy concepts and calculations are used extensively in Information Theory and machine learning, and entropy computational methods can be used in evaluating phonetic phenomena. Entropy calculations are used to detect patterns in data that are related, and to show data relationships by providing numerical values that show the relative degree of overlapping information (see Haglund, Jeppsson, & Strömdahl, 2010; Goodfellow et al. 2016). Conditional entropy values are used to evaluate the mathematical relationship of one variable set to another variable set and entropy values indicate the suitability for plotting one formant against the formant, for example, F1 vs. F2 or F1 vs. F3 plots.

The conditional entropy values were calculated with a modified version of Mutual Info 0.9 cross-platform program package (Peng 2002) operating in Matlab ™. Conditional entropy calculations provided numerical data that shows the relationship of formant values to: F1 to F2, F1 to F3, and F2 to F3.

Materials and methods

This experiment was designed to compare the entropy values of the vowel portions of target words that were produced in clearly spoken sentences. This experiment used words that were produced for a previous perception-production study that compared English and Greek vowel spaces (Botinis, et al., 2022). Male and a female native New Mexico English speakers, using carrier sentences, recorded the words: *bit, beat, bet, bat, boot, butt, bought,* and *bot*, producing the monothongs /iː, ɪ, e, æ, uː, ʌ, ɔ, ɑː/, which represent the corner vowels and edges of their respective dialect vowel space. Speech recordings were produce using a Røde N microphone in a GretchKen™ Industries acoustic sound booth. The sentences were spoken multiple times clearly and the key words in the sentence were spoken with brief silence before and after each word. Formants were next created using PRAAT, (standard settings of 5,500 Hertz ceiling and 5 formants) and the formant data was recorded in a spreadsheet. Entropy calculations were completed for natural speech and the data was displayed in a self-populating datasheet for comparisons.

The Speechify ™ cloning program was trained on the New Mexico English carrier sentences. Speechify™ was used then to produce the same cloned sentences. The cloned vowel formants were determined again using PRAAT, and the values were copied into data sheets. Entropy calculations were completed for the cloned speech and the data was displayed in a self-populating datasheet for comparisons.

Results

The vowels spoken by the female participant (/iː, ɪ, e, æ, uː, ʌ, ɔ, ɑː) were examined and it was found that in the eight vowels, at least one formant was dominant and at least one formant was dependent. This is a common pattern observed in our numerous entropy calculations for natural speech. The dominant formant would be best plotted on the x-axis and the dependent formant would be best plotted on the y-axis.

For the cloned speech samples of the female speaker, the entropy values of each vowel portion have significantly altered dominant formants, compared to the formants of natural speech. Entropy relationships between cloned formants are significantly weaker than the entropy values of natural speech formants. This also indicates that plotting cloned formants would not produce reliable graphs.

The vowels spoken by the male participant (/iː, ɪ, e, æ, uː, ʌ, ɔ, ɑː) were examined and it was found that in the six vowel pairings, at least one formant was dominant and at least one formant was dependent. Again, the dominant formant would be best plotted on the x-axis and the dependent formant would be best plotted on the y-axis.

For the male speaker cloned speech, the entropy values of the target vowels from the words *bat, beat, bet*, and *bit* showed that each cloned vowel formant has altered dominant formants. The conditional entropy values increased for cloned speech, indicating a weakening of the relationship between formants. For example, the cloned vowels derived from words *boot* and *bot* showed an entropy increase in formant relationships, weakening and changing F3 to the dominant formant. The cloned vowels derived from the words *bought* and *butt* showed reversed dominant formants with the dependent formants, compared to natural speech.

Table 1. Conditional entropy differences between speech and cloned speech for the female speaker.

Female speaker	bat /æ/	beat /ɪ/	bet /e/	bit /i:/	boot /u/	bot /ɑ:/	bought /ɔ/	butt /ʌ/
F1-F2 conditional	0.068966	0.095238	0.083333	0	0.095238	0.181818	0.442684	0
F2-F1 cond	0.137931	0	0	0	0.095238	0.090909	0.076923	0.1
F1-F3 cond	0.137931	0	0	0	0	0	0.076923	0
F3-F1 cond	0.137931	0	0	0	0.095238	0.090909	0.076923	0.1
F2-F3 cond	0.137931	0	0	0	0	0	0.076923	0
F3-F2 cond	0.068966	0.095238	0.083333	0	0.095238	0.181818	0.442684	0
Cloned words	**/æ/**	**/ɪ/**	**/e/**	**/i:/**	**/u/**	**/ɑ:/**	**/ɔ/**	**/ʌ/**
F1-F2 conditional	0.295385	0.094118	0.179775	0.144928	0.1	0.133333	0.294872	0.094118
F2-F1 cond	0.306119	0.806227	0.384996	0.659562	0.673429	0.133333	0.557818	0.361822
F1-F3 cond	0.178921	0.164706	0.269663	0.144928	0.184436	0.222222	0.34785	0.188235
F3-F1 cond	0.32336	0.806227	0.384996	0.659562	0.673429	0.133333	0.557818	0.361822
F2-F3 cond	0.178921	0.164706	0.269663	0.144928	0.184436	0.222222	0.34785	0.188235
F3-F2 cond	0.312626	0.094118	0.179775	0.144928	0.1	0.133333	0.294872	0.094118

Table 2. Conditional entropy differences between speech and cloned speech for the male speaker.

Male speaker	bat /æ/	beat /ɪ/	bet /e/	bit /i:/	boot /u/	bot /ɑ:/	bought /ɔ/	butt /ʌ/
F1-F2 conditional	0	0	0.321661	0.26087	0	0	0.133333	0.1
F2-F1 cond	0.173913	0.222222	0.190476	0.173913	0.086957	0.086957	0	0
F1-F3 cond	0	0.222222	0	0	0.347826	0.347826	0.066667	0.1
F3-F1 cond	0.173913	0.222222	0.190476	0.173913	0.086957	0.086957	0	0
F2-F3 cond	0	0.222222	0	0	0.347826	0.347826	0.066667	0.1
F3-F2 cond	0	0	0.321661	0.26087	0	0	0.133333	0.1
Cloned words	**/æ/**	**/ɪ/**	**/e/**	**/i:/**	**/u/**	**/ɑ:/**	**/ɔ/**	**/ʌ/**
F1-F2 conditional	0.153846	0.210526	0.173913	0.111111	0	0	0	0
F2-F1 cond	0.076923	0.210526	0.347826	0.222222	0.08	0.166667	0.125	0.222222
F1-F3 cond	0.076923	0.105263	0.173913	0.111111	0.16	0.166667	0.27359	0
F3-F1 cond	0.076923	0.210526	0.347826	0.222222	0.08	0.166667	0.125	0.222222
F2-F3 cond	0.076923	0.105263	0.173913	0.111111	0.16	0.166667	0.27359	0
F3-F2 cond	0.153846	0.210526	0.173913	0.111111	0	0	0	0

Conclusions

Conditional entropy has been used to detect patterns in data by providing numerical values that show their relative degree of overlapping information, showing how dependent one formant is on another.

The clones of the target words were perceptually similar to the original spoken words, but the entropy values for the clones of showed considerable changes. Generally, conditional entropy values were higher for cloned speech, which is an indication of increased randomness in the sound signal. Higher clone entropy values indicate weaker relationships between the formants for the cloned speech, which would translate to weaker or less clearly defined harmonics.

This study is limited, and it is exploratory, but strong positive results indicate that the vowel formants undergo entropy and often formant priority changes when they are cloned. More testing will be used to duplicate the method we have described using a variety of speech samples as we look for predictable patterns in the cloned speech entropy values.

We theorize that identifying entropy changes due to cloning may contribute to linguistic forensic examinations. Cloned speech has the potential to be used fraudulently and identifying cloned speech by its entropy values should be a way to determine the authenticity of speech.

References

Botinis, A., Goertz, G., Kontostavlaki, A., Anderson, T. 2023. Vowel discrimination of American English. ExLing 2023 Athens: Proceedings 14th International Conference of Experimental Linguistics, October 13-16. Athens, Greece.

The MathWorks Inc. 2023. MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com.

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.

Peng, H. 2022. Mutual Information computation. (www.mathworks.com/matlabcentral/fileexchange/14888-mutual-information-computation),

Perceptual analysis of charismatic speech characteristics

Juliana Andreassa
Pontifical Catholic University of São Paulo, Brazil

https://doi.org/10.36505/ExLing-2024/15/0004/000629

Abstract

Charisma can be described in terms of semantic pragmatic features related to trustworthiness, motivation, and persuasion. In this paper, we are concerned with the multidimensionality of charismatic speech because it depends on the production of the speaker and the perception of the listener. Our aim, in this study, is to compare the voice quality profiles of two YouTube communicators whose manners of speaking were judged as the most charismatic and the least charismatic in a perceptual test. Minimized Mandibular Range, Minimized Tongue Body Range, Nasality and Interrupted Speech were judged as less charismatic than Extensive Mandibular Range and Fast Speech.

Keywords: multimodality, charisma, charismatic speech, voice quality

Introduction

The term charisma was used in Ancient Greece to indicate a gift, and an extraordinary leadership quality. From the sociological point of view, charisma has been defined throughout the centuries as an extraordinary quality of a person (Weber 1920) and it can be related to other semantic pragmatic features such as trustworthiness, motivation, and persuasion.

As a result, the term was also attributed as a type of leadership that presents a relation to the leader and to the followers since a charismatic leader is a product of his/her followers (Weber 1920; Boss 1976). According to Signorello (2021), charisma is not related to leadership and convincing at most, but also to the way the listener feels when receiving the message. For that reason, the charismatic speech is a multidimensional construction that depends on the production of the speaker and the perception of the listener.

A charismatic leader sounds convincing and motivating because he/she is able to make the public feel the message that is expressed. Kouzes and Posner (2001) and Militaru (2013) argue that charisma is directly related to the social relationships involving leadership and followers. Therefore, charisma is not only associated with speakers' attributes, but it is also an attribute associated to the listeners and how they perceive them.

The way the listeners perceive charismatic speakers can be analyzed from the perspective of voice quality characteristics. Voice quality is described as a product of articulatory and phonatory settings that "colors" the speaker's voice

by Laver (1980). Niebuhr et al. (2018) points out that voice quality is a charisma-relevant feature to investigate how charismatic speech is produced by the speaker. Speakers with some specific voice quality settings such as extensive mandibular range, extensive labial range, and fast speech are considered to be confident, charming, engaging, and persuasive (Signorello 2014).

Our aim is to compare the voice quality profiles of two YouTube communicators whose manners of speaking were judged as the most charismatic and the least charismatic in a perceptual test. Our hypothesis is that speaker's voice quality settings contribute to the evaluation of the listener.

Methodology

The methodological procedures comprised the application of a perceptual semantic questionnaire, and the perceptual analysis of voice quality and prosodic settings with the Voice Profile Analysis (VPA).

The perceptual test was conducted with speech samples from ten male investment counselor speakers from São Paulo, Brazil, and it was answered by a group of 53 judges also from São Paulo, Brazil using the SurveyMonkey platform. The judges had to evaluate on a linear scale how charismatic, convincing, trustworthy, and motivating the speaker was.

Two speakers were selected: the best rated and the worst rated speakers judged by the listeners.

The perceptual analysis of voice quality and prosodic settings using VPA was performed by a group of 7 trained judges and the type of assessment was consensual.

Results

For the perceptual semantic test, the results showed that the best rated speaker was considered 68,7% charismatic, 69,3% convincing, 69,7% trustworthy, and 67,1% motivating, while the worst rated speaker was considered 47,4% charismatic, 45,8% convincing, 50,3% trustworthy, and 37,3% motivating.

Figure 1 presents how the best and the worst rated speakers were judged in the perceptual test.

Concerning the vocal profiles, the two speakers shared the following non-neutral voice quality and prosodic settings: extensive labial range, lowered tongue body, raised larynx, tense vocal tract, open jaw, tense larynx, high pitch, extensive pitch range and high loudness. The worst rated speech production was described with the following non-neutral voice quality and prosodic settings: lip spreading, minimized mandibular range, fronted tongue body, minimized lingual body range, nasality, and interrupted speech.

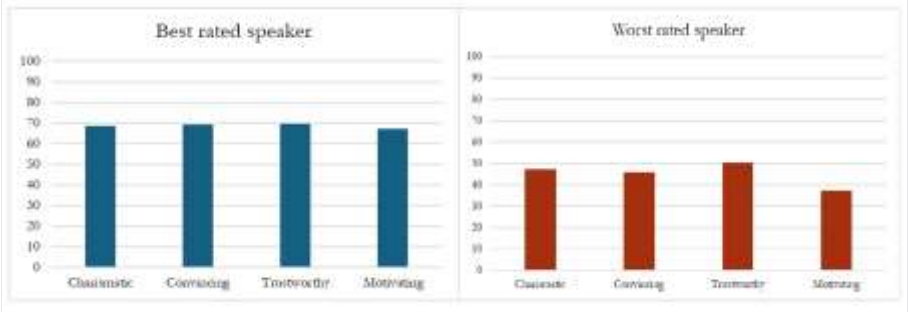


Figure 1. Best and worst rated speakers in the perceptual test

Figure 2 presents the best and the worst rated speakers' voice quality settings.

Best rated speaker's voice quality settings	Worst rated speaker's voice quality settings
	Lip Spreading
Extensive Labial Range	Extensive Labial Range
Open Jaw	Intermitent Open Jaw
Protruded Jaw	
Extensive Mandibular Range	
	Minimised Mandibular Range
	Fronted Tongue Body
Lowered Tongue Body	Lowered Tongue Body
	Minimised Tongue Body Range
	Nasal
Raised Larynx	Raised Larynx
Tense Vocal Tract	Intermitent Tense Vocal Tract
Tense Larynx	Tense Larynx
Intermitent Creaky	Intermitent Creaky
	Interrupted Continuity
Fast Rate	

Figure 2. Voice quality settings – similarities and differences

Discussion and conclusions

The types of voice quality settings characterizing the speakers' voice profiles can be interpreted in terms of speech style, speech task and idiosyncratic features.

In terms of speech style, clear speech is a valued feature in YouTube talk channels and voice quality settings such as extensive labial range, open jaw, lowered tongue body shared by both speakers can contribute for the speaker to sound clear, and convincing.

Other voice quality settings also shared by both speakers, such as tense vocal tract, tense larynx, high pitch, raised larynx can be viewed as related to the kind of speech tasks that require effort from the part of the speaker to communicate effectively.

Otherwise, voice quality settings such as protruded jaw and nasality can be related to idiosyncratic features. Previous studies presented that nasality can be associated with negative perceptual impressions (Zuckerman and Miyake 1993; Lukkarila et al. 2012).

Features such as minimized mandibular range, minimized tongue body range, nasality and interrupted speech may have contributed to judgements of a less charismatic hesitant speech, whereas extensive mandibular range and fast speech to a more charismatic, and fluent speech.

In conclusion, potential features determining rating differences between the two speakers in perceptual semantic questionnaires can be motivated by the kinds of settings that characterize the speakers' vocal profiles.

References

Lukkarila, P., Laukkanen, A. M., Palo, P. 2012. Influence of the intentional voice quality on the impression of female speaker. Logopedics Phoniatrics Vocology, 37(4), 158-166.

Niebuhr, O., Skarnitzl, R., Tylečková, L. 2018, June). The acoustic fingerprint of a charismatic voice-Initial evidence from correlations between long-term spectral features and listener ratings. In 9th International Conference on Speech Prosody 2018 (pp. 359-363). International Speech Communication Association (ISCA).

Signorello, R. 2020. The Vocal Attractiveness of Charismatic Leaders. In Benjamin, Weiss, Jürgen Trouvain, Melissa Barkat-Defradas, John J. Ohala (ed.). Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speakers. Prosody, Phonology and Phonetics, 41-54.

Signorello, R. 2014. The biological function of fundamental frequency in leaders' charismatic voices. The Journal of the Acoustical Society of America, 136(4_Supplement), 2295-2295.

Zuckerman, M., Miyake, K. 1993. The attractive voice: What makes it so?. Journal of nonverbal behavior, 17(2), 119-135.

The face, sound and expressiveness of voice quality settings

Juliana Andreassa, Alice Crochiquia, Rafael Scarpelli, Mateus Pires
LAEL, PUC-SP, Brazil

https://doi.org/10.36505/ExLing-2024/15/0005/000630

Abstract

In this paper, we investigate associations between vocal and facial gestures and meaning effects in utterances produced by an actor with labial, mandibular and tongue-tip phonetic settings of voice quality. The corpus comprises 12 videos. Each video refers to an utterance produced with a different voice quality setting. Four kinds of analysis were performed: a perceptual analysis of voice quality settings and prosodic features by four experts on the use of the VPA system; an automatic analysis of facial AUs and related basic emotions, valence, and arousal states; an analysis of acoustic parameters extracted with the Prosody Descriptor Extractor Script; and a perceptual test to investigate 30 listeners' judgements of semantic criteria based on their listening to the audio stimuli. The Exploratory multivariate statistical analysis was applied to correlate the quantitative and qualitative variables concerned.

Keywords: voice quality, facial expression, multimodality.

Introduction

Voice qualities are expressive gestures. For centuries, metaphors have described them based on the sensations (Fónagy 2000) accompanying sound production (proprioceptive, tactile, motor sensations) and the auditory impressions they give rise to.

Facial gestures are especially relevant to consider alongside voice in the attribution of paralinguistic and extralinguistic meaning in speech as the changes in the facial plan can directly affect the acoustic properties in speech (Banse and Scherer 1994), and the use of particular voice quality settings, vocal prosody and their acoustics effects can lead to the correlation to facial expressions, such as Lip Spreading and the aural perception of a smile and emotions with neutral or positive valence (Madureira and Fontes 2023). Unsurprisingly, the perception of emotions is far more efficient in bimodal situations, where people have both auditory and visual information. However, some expressions of emotions may be perceived somewhat more accurately in one channel in isolation than the other, like anger in the visual modality and surprise with audio (Abelin 2008).

The experiment presented in this paper has the following aim: to investigate associations between vocal gestures and their acoustic output, facial gestures,

and meaning effects in utterances produced by an actor with labial, mandibular and tongue tip phonetic settings of voice quality.

Material and methods

Corpus and subject

The corpus contains 12 utterances. They were recorded with a Canon 60D camera connected to a lapel camera with a 24mm lens in full HD files with lightning quality. The speaker was always in a frontal position.

The research subject is a 33-year-old actor from the State of São Paulo, Brazil. The speaker was told to produce the sentence: "*O ignorante rejeita, o sábio duvida, o sensato reflete*" (The ignorant one rejects, the wise one doubts, the sensible one reflects), using a different setting for each repetition. Each utterance was produced with a distinct voice quality setting. The chosen settings were (1) Minimised Labial Range; (2) Extensive Labial Range; (3) Lip Rounding; (4) Lip Spreading; (5) Labiodentalization; (6) Closed Jaw; (7) Open Jaw; (8) Protruded Jaw; (9) Extensive Mandibular Range; (10) Minimised Mandibular Range; (11) Advanced Tongue Tip/Blade; (12) Retracted Tongue Tip/Blade.

Perceptual-semantic analysis test: corpus, descriptors and judges

Only the audios were selected to be included in the perceptual-semantic test. The test was presented entirely online to 30 Brazilian Portuguese native speakers, male and female, aged between 20 and 60. The judges were asked to listen to 12 audios to evaluate the speaker using the following adjectives: pleasant/unpleasant, fragile/strong, relaxed/tense. The semantic descriptors chosen by the judges for each utterance were rated in positive, negative, and neutral levels from -3 to 3 on a scale, and the measures related to these levels were weighted and yielded a value for each of the polar semantic descriptors.

Acoustic analysis

The Prosody Descriptor Extractor script (Barbosa 2020) for Praat (Boersma and Weenink 2021) was used for the acoustic analysis. It extracts 22 acoustic measures related to frequency and intensity acoustic parameters.

Perceptual voice quality analysis

Four phoneticians carried out the perceptual voice quality analysis using the VPA protocol (Laver and Mackenzie Beck 2007). In this study, we have included just the first 12 settings described in the protocol, which cover labial, mandibular, and tongue-tip phonetic settings.

Facial expression analysis

For the facial analysis, the software FaceReader 8.1 was used. This software characterizes the facial Action Units (AUs) based on the FACS system (Ekman

and Friesen, 1971) and associated basic emotions, Valence, and Arousal. For this study, 7 emotions (Happiness, Sadness, Anger, Surprise, Fear, Disgust, and Contempt), 20 AUs, Arousal (activated and passive), and Valence (positive and negative) were considered.

Principal component analysis

The variables were analyzed using the Multi-Factor Analysis (MFA) (Josse, Pagès and Husson 2008). All the variables were normalized by z-score.

Results

Table 1 shows the significant variables and the degree of correlation of the variables in Dimensions 1, 2, 3, and 4. Positive values are highlighted in light orange, and negative values in light blue. Based on the percentages of correlation between variables with positive values and variables with negative values, variables were associated.

Dim. 1		
quanti	co.	p. value
AU26	0,8222	0,001
AU10	0,8027	0,0017
AU27	0,7799	0,0028
AU25	0,7737	0,0032
Disgust	0,7528	0,0047
f0base	0,7419	0,0057
Arousal	0,7131	0,0092
AU07	0,7009	0,0111
Laxness	0,6726	0,0165
f0min	0,6592	0,0197
AU09	0,6164	0,0328
Surprise	0,5869	0,0449
Happiness	0,5816	0,0473
sdf0peak	-0,5889	0,0439
Tenseness	-0,6148	0,0334
AU24	-0,6157	0,0331
quali	R2	p.value
MEJ	0,3687	0,0363

MEJ = Mandibular Extensive Range

Dim. 2		
quanti	co.	p. value
AU02	0,7556	0,0045
f0SAQ	0,6983	0,0115
df0posmean	0,6698	0,0172
f0sd	0,664	0,0185
AU05	0,6601	0,0195
Pleasantness	0,5945	0,0415
AU01	0,5929	0,0422
f0peak_rate	-0,61	0,0353
Unpleasantness	-0,6157	0,0331
quali	R2	p.value
MOJ	0,4963	0,0105

MOJ = Open Jaw

Dim. 3		
quanti	co.	p. value
AU12	0,791	0,0022
Valence	0,7664	0,0036
AU06	0,682	0,0146
Happiness	0,6238	0,0302
Sadness	-0,619	0,0317
AU43	-0,65	0,022
AU15	-0,684	0,0142
AU17	-0,747	0,0053
quali	R2	p.value
MPJ	0,3908	0,0297

MPJ = Protruded Jaw

Dim. 4		
quanti	co.	p. value
jitter	0,7093	0,0098
sll.TAShigh	0,6972	0,0117
Strength	0,6896	0,0131
emph	-0,63	0,028
Fragility	-0,816	0,0012
quali	R2	p.value
TIA	0,3424	0,0457

TIA = Advanced Tongue Tip/Blade

Table 1. Significant variables, percentages of correlation (co.) and p.values.

Conclusions

Based on the correlation percentages and the voice quality settings of the stimuli, six frames containing the variables related to the 3 semantic descriptors included in the Perceptual Test (Pleasantness, Strength, and Tenseness) were derived.

Frame 1 - Pleasantness: Labial Extensive Range, Lip Spreading, Open Jaw, Happiness, Surprise, AU06 (Cheek Raiser), AU12 (Lip Corner Puller), Positive Valence, Interquartile Semi-Amplitude of f0 (f0SAQ), Mean of f0 Positive First Derivative (df0posmean), and Standard Deviation of f0 (f0sd). Features related to utterances 2, 4, and 7.

Frame 2 - Unpleasantness: Labial Minimized Range, Close Jaw, Mandibular Extensive Range, Advanced Tip Blade and Retracted Tip Blade, AU15 (Lip Corner Depressor), AU17 (Chin Raiser), AU24 (Lip Pressor), AU43 (Eyes Closed), AU 07 (Lid Tightner), Sadness, Disgust, Negative Valence, Rate of f0 peaks (f0peak_rate). Features related to utterances 1, 6, 9, 11, and 12.

Frame 3 - Fragility: Labial Minimized Range, Labiodentalization, Mandibular Minimized Range, Advanced Tip Blade, AU06 (Cheek Raiser), AU12 (Lip Corner Puller), and Spectral Emphasis (emph). Features related to utterances 1, 5, 10, and 11.

Frame 4 - Strength: Labial Extensive Range, Lip Rounding, Mandibular Protruded Jaw, Retracted Tip Tongue, AU10 (Upper Lip Raiser).Features related to utterances 2, 3, 6, 8, and 12.

Frame 5 -Laxness: Lip Rounding, Open Jaw, Lingual Advanced Tip, AU25 (Lip Part), AU26 (Jaw Drop), and f0 Minimum (f0min). Features related to utterance 3, 7 and 11.

Frame 6 - Tenseness: Labial Minimized Range, Labial Extensive Range, Mandibular Minimized Range, Close Jaw, AU24 (Lip Pressor) and Standard Deviation of F0 peaks (sdF0peaks), Jitter and (sILTAShigh) and. Features related to utterances 1, 2, 6, and 10.

References

Abelin, Å. 2008. Seeing glee but hearing fear? Emotional McGurk effect in Swedish. Proceedings of Speech Prosody. May 6, 9.

Banse, R., Scherer, K. R. 1994. Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology, vol. 70, no. 3, pp. 614–636.

Barbosa, P. A. 2020. Prosody Descriptor Extractor (for Praat). [Online].

Boersma, P., Weenink, D. 2021. Praat: doing phonetics by computer. [Online].

Ekman, P., Friesen, W. V. 1971. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, vol. 17, no. 2, pp. 124–129.

Ekman, P., Friesen, W. V., Hager, J. C. 2002. Facial Action Coding System.

Fónagy, I. 2000. Languages within languages: an evolutive approach. John Benjamins.

Josse, J., Pagès, J., Husson, F. 2008. Testing the significance of the rv coefficient. Computational Statistics amp; Data Analysis, vol. 53, no. 1, pp. 82–91.

Laver, J., Mackenzie Beck, J. 2007. Vocal profile analysis scheme-VPAS. Queen Margareth University College-QMUC, Speech Science Research Centre, Edinburgh, Handout.

Madureira, S., Fontes, M.A.S. 2023. Multimodal impressions of voice quality settings: the role of vocal and visual symbolism. Frontiers in Communication, vol.8. [Online].

Experimenting Free Choice through palm up gestures and prosody

Marion Bonnet[1], Cornelia Ebert[2], Kurt Erbach[2], Markus Steinbach[1]
[1]University of Goettingen, Germany
[2]University of Frankfurt, Germany

https://doi.org/10.36505/ExLing-2024/15/0006/000631

Abstract

In the present experimental studies, we investigated the interaction of co-speech palm up gestures and prosody in sentences typically triggering free choice (FC) interpretation. Results show that multimodal sentences boost the exclusive interpretation compared to utterances displayed only in the spoken modality, which only received the FC interpretation (inclusive reading). On the basis of this evidence, we suggest that visual cues should more systematically be used as a tool to shed light on linguistic theories, and enrich experimental methods.

Keywords: co-speech gestures, palm up, multimodality, Free Choice, semantics.

Background

Sentences containing disjunction in the scope of a possibility modal give rise to the FC interpretation (Meyer, 2018). Therefore, a sentence like *Alex can have ice cream or cake* implies that Alex can freely choose to have one of the two options or both. It is still under debate which theory would best account for such data. Tieu et al. (2019) used a ternary judgment task involving pictures and a colour code representing the available choices to empirically investigate which of the two main approaches – the implicature (Bar-Lev & Fox, 2017) vs. the homogeneity approach (Goldstein, 2018) – makes the best predictions.

In addition, it is well-known that prosody plays a role in the interpretation of disjunctive sentences: Pruitt & Roelofsen (2013) have shown that final contour and pitch accents allow to disambiguate between alternative and yes-no questions. Similarly, in the context of coordinated names such as 'Nino or Willi or Mila', phrasing variations mirror the syntactic parsing (Kentner & Féry, 2013).

Moreover, recent studies, looking at sign languages and/or gestures, show that visual cues, like prosodic cues, affect the interpretation and argue that theoretical accounts of disambiguation and enrichment should consider the semantic and pragmatic impact of the visual modality. More specifically, iconicity and alignment of speech and gestures (Schlenker, 2023; Ebert & Ebert, 2014) could provide interesting insight to such topic. Nonetheless, there is still an open question regarding the extent to which visual information can impact the interpretation of disjunctive utterances. Following Cooperrider et al.

(2018), the resort to palm up gestures to express possibility is widely attested in sign languages and gestures accompanying speech. One interesting question in our context is therefore whether the standard FC interpretation could be modified when visual information and phrasing suggest a different reading?

Methods

We investigated the contribution of palm up co-speech gestures to German FC sentences in three experiments. We used a picture selection task and a colour code (as in Tieu et al. 2019) targeting the two classical interpretations at stake: inclusive (FC) vs. exclusive reading. Four answers were available, each represented by a set of two pictures in which each disjunct is either framed in green (allowed) or red (not allowed) (*Picture 1*). It resulted in the four following choices: both disjuncts are allowed (inclusive reading), no disjunct is allowed (dual prohibition), only disjunct 1 is allowed (exclusive reading), only disjunct 2 is allowed (exclusive reading). Crucially, we paid attention to the items we created so that the two disjunct were not mutually exclusive based on common knowledge and logic

Experiment 1 used audio only items and served as a baseline, verifying the FC effect and testing two phrasing patterns, adapted from Pruitt and Roelofsen (2013). The first pattern, "disjunctive phrasing" separates the disjuncts by a short break before the word *oder* ('or'), and displays the main pitch accent on the first disjunct. The second pattern, "conjunctive phrasing", does not show any pause and bears the main accent on the second disjunct.

We then implemented two multimodal experiments testing four gesture conditions: three types of palm up gestures, differing in the alignment between the hands' movements and the disjuncts (*Picture 2*) and a no gesture condition. Audio and video were recorded separately and then merged. Experiment 2 used the disjunctive phrasing pattern and Experiment 3 the conjunctive pattern.

Picture 1. Screenshot of the task item "*Alex can wear a scarf or a beany*".

Picture 2. Illustrations of the four gesture conditions.

Results

Results from Experiment 1 (audio only) verify the FC effect since nearly 100% of the items were interpreted with the inclusive (FC) reading. No significant difference appears between the two phrasing patterns in this experiment (Figure 1).

Experiment 2 (multimodal items combined with the disjunctive phrasing) presents notably different results. Target items received the exclusive reading in more that 50% of the trials. The type of palm up gesture does not create a significant difference in the interpretation and more surprisingly, the no gesture condition behaves like the other gesture conditions, that is boosting the exclusive reading (p-value = 0.03887) (Figure 2).

Finally, Experiment 3 (multimodal items combined with the conjunctive phrasing) present results very similar to Experiment 2. Target items received the exclusive reading in more than half percent of the trials. There is no significant difference between the no gesture, the alternating palm up and the holding palm up conditions. However, although it is a minor difference, the both hands palm up triggers slightly more inclusive reading compared to the other three conditions (p-value = 0.03345) (Figure 2).

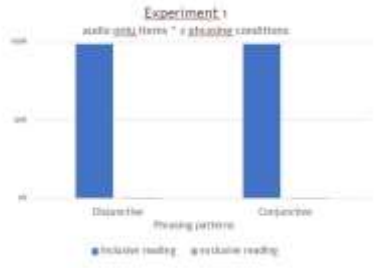


Figure 1. Interpretation of audio only items per phrasing pattern.

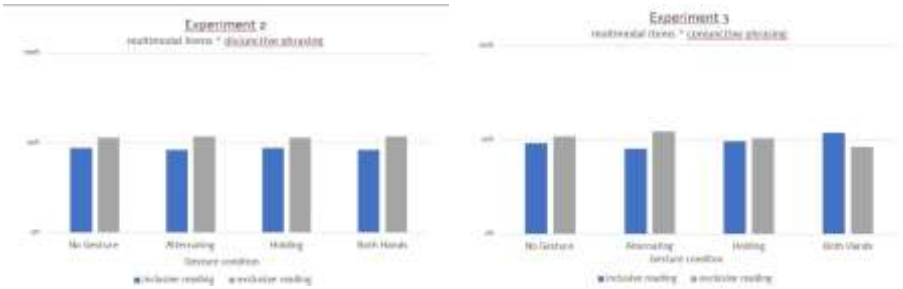


Figure 2. Left: multimodal items interpretation per gesture condition, combined with the disjunctive phrasing (exp. 2). Right: multimodal items interpretation per gesture condition, combined with the conjunctive phrasing (exp. 3).

Discussion

As expected, the FC interpretation appears very robust, but only when the sentences are displayed in the audio only condition. The difference between the audio only and the multimodal items suggests that visual cues tend to nuance the interpretation across participants. Surprisingly, it does not seem to be the gestures per se that boost the exclusive reading but rather the overall multimodal setting since the no gesture condition gives rise to similar results as any other gesture conditions. One possible explanation could be that participants presented with a multimodal input throughout the experiment were expecting a gesture. At a more formal level, as in Tieu et al. (2019), our results do not clearly disentangle between the two main approaches but they seem to be more straightforwardly favouring the homogeneity approach. We believe that experimental studies involving gestures could be a promising tool for shedding light on this debate. Finally, we propose to analyse multimodal disjunctive sentences as complex disjunctions (Spector, 2014) in which the visual component can nuance the FC interpretation. We conclude that experimental linguistics could greatly benefit from multimodal research and give rise to a boomerang effect whereby the study of multimodality / visual modality would profit from more frequent experimentation.

References

Bar-Lev, M.E., Fox, D. 2017. Universal free choice and innocent inclusion. Semantics and Linguistic Theory, 95-115.

Cooperrider K., Abner N., Goldin-Meadow S., 2018. The Palm-Up Puzzle: Meanings and Origins of a Widespread Form in Gesture and Sign. Front. Commun. 3:23.

Ebert, C, Ebert, C. 2014. Gestures, demonstratives, and the attributive / referential distinction. Handout of a talk given at Semantics and Philosophy in Europe (SPE 7), Berlin 28 (2014).

Goldstein, S. 2019. Free choice and homogeneity. Semantics and Pragmatics 12, 23-1.

Kentner, G., Féry, C. 2013. A new approach to prosodic grouping. The Linguistic Review 30(2), 277-311.

Meyer, M.C. 2018. An apple or a pear: free choice disjunction. Wiley's semantics companion, Wiley and Sons.

Pruitt, K., Roelofsen, F. 2013. The interpretation of prosody in disjunctive questions. Linguistic inquiry 44(4), 632-650.

Schlenker, P. 2023. On the typology of iconic contributions. Theoretical Linguistics 49(3-4), 269-290.

Spector, B. 2014. Global positive polarity items and obligatory exhaustivity. Semantics and Pragmatics 7, 11-1.

Tieu, L., Bill, C., Romoli, J. 2019. Homogeneity or implicature: An experimental investigation of free choice. Semantics and linguistic theory 29, 706-726).

Experimental cognitive-linguistic assessment in Italian learners of L2 Chinese with dyslexia

Nicoletta Cirota
University of Venice, Italy

https://doi.org/10.36505/ExLing-2024/15/0007/000632

Abstract

This study develops an experimental battery to assess cognitive-linguistic skills in Italian dyslexic and non-dyslexic learners of Chinese as a second language (L2). Adapting tasks used for Chinese native children with dyslexia, the battery targets phonological awareness, Rapid Automatized Naming (RAN), orthographic awareness, working memory, morphological awareness, and literacy skills. The goal is to identify predictors of Chinese character acquisition, focusing on challenges faced by dyslexic learners. Early findings indicate how cognitive-linguistic skills interact to influence L2 Chinese acquisition. This research aims to address gaps in dyslexia and L2 Chinese acquisition, offering implications for educational strategies, particularly for dyslexic learners in multilingual contexts.

Keywords: Dyslexia, Chinese L2 Acquisition, Cognitive-Linguistic Skills, Educational Interventions

Introduction

This study investigates the cognitive-linguistic predictors of Chinese character acquisition in Italian dyslexic and typically developing learners (TDLs) of Chinese as a second language (L2). Despite extensive research on dyslexia in alphabetic languages, little attention has been paid to its manifestation in learners of non-alphabetic L2s such as Chinese. This gap is particularly critical as Chinese is increasingly being introduced as a curricular subject in Italian secondary schools (Favaloro, 2017), where tailored pedagogical frameworks and assessments are lacking (Verzì, 2023). The research uses the Hypothesis of Granularity and Transparency (HGT) (Wydell & Butterworth, 1999) as a theoretical framework, which posits that dyslexic deficits are influenced by the transparency (grapheme-phoneme regularity) and granularity (size of orthographic units) of the orthographic system. Using a custom-designed experimental battery, this study aims to identify predictors of Chinese character acquisition, compare the cognitive-linguistic profiles of dyslexic learners (Zhang et al., 2023) and TDLs, and explore how deficits manifest in Italian (a transparent, alphabetic L1) and Chinese (an opaque, morpho-syllabic L2). The main research questions are how cognitive-linguistic deficits affect Chinese L2 learners with dyslexia, whether these deficits are influenced by L1 characteristics, and whether TDLs present dyslexia-like challenges in Chinese

acquisition. Early diagnosis and intervention are crucial as dyslexia significantly affects not only cognitive skills but also social-emotional well-being (Wydell & Fern-Pollak, 2012), with unique challenges in Chinese including poor character retention, radical inversions, and confusion with visually similar characters (Siok et al., 2009).

Methodology

This study uses a quantitative testing approach, combining standardised and experimental assessments, to assess the cognitive-linguistic skills of Italian secondary school learners of Chinese as a second language (L2). Participants include dyslexic learners and typically developing learners (TDLs) recruited from language-focused high schools in northern Italy. Inclusion criteria ensured that all participants were actively learning Chinese and were at an intermediate level (B1.1-B1.2 CEFR). Exclusion criteria, such as incomplete schooling in Italy or concurrent developmental conditions, were applied to maintain homogeneity.

Experimental battery

The experimental battery includes tasks designed to assess a wide range of cognitive-linguistic skills that are critical for the acquisition of Chinese characters. These include phonological awareness, orthographic processing, morphological awareness, working memory, Chinese reading and writing fluency, Rapid Automatised Naming (RAN) and visual attention. The tasks in the experimental battery were adapted from previous research on native Chinese dyslexic learners (Pan et al., 2024; Liu et al., 2015) and L2 learners of Chinese (Chang et al., 2022; Yang, 2021). The Chinese-specific section of the battery targets key cognitive-linguistic domains. For phonological awareness, participants completed an *Onset Detection Task*, identifying the differing onset in sets of Pinyin syllables, and a *Rime Detection Task*, distinguishing the rime in other sets of Pinyin syllables. Additionally, a *Pinyin Writing Task* required participants to write Pinyin accurately, including tones, after listening to words.

Orthographic awareness and working memory were assessed using the *Chinese Orthographic Choice Task*, in which participants chose the more realistic character from two artificial options, and the *Chinese Delayed Copying Task*, where they reproduced Chinese characters from memory after a brief presentation. Morphological awareness was measured through two tasks: *Compound Word Production with Pictures*, where participants formed compound words based on images, and *Compound Word Production with Sentences*, requiring word formation after reading descriptive sentences in Italian (L1). For Chinese literacy, three tasks were included. The *Chinese Character Reading Task* assessed the ability to read high-frequency characters aloud, while the *Chinese Word Spelling Task* required participants to write characters after listening to them. Lastly, the

Chinese Word Segmentation Task involved segmenting words from a continuous chain of unspaced Chinese characters. Additionally, the test battery administered included a standardised Italian battery, the *Nuova Batteria per Studenti Universitari e Adulti LSC-SUA* (Cornoldi & Montesano, 2020), to assess participants' L1 literacy skills. This provided a baseline for understanding how deficits in Italian may affect Chinese L2 acquisition.

Data collection

Italian and Chinese tasks were administered individually in quiet school settings for accuracy. Each task was introduced with a PowerPoint presentation and practical examples, with responses recorded using standardised protocols. Oral performances were recorded with parental consent, and the OpenDyslexic font was used to enhance material accessibility (Rello, Baeza-Yates 2013).

Preliminary findings

Initial findings reveal that dyslexic learners struggle with slower reading speeds, higher error rates, and greater variability in RAN tasks, alongside difficulties in rime identification and tone accuracy in Pinyin writing. They also exhibit issues with letter order, substitution, and phoneme inversion when forming compound words. Typically developing learners perform faster with fewer errors and better rime recognition but share challenges in character order. Both groups show morphological processing difficulties, with dyslexic learners experiencing more frequent syllable/phoneme inversions or additions. These results underscore the need for targeted strategies to support dyslexic learners in Chinese character acquisition.

Discussion

This study developed an experimental battery to assess cognitive-linguistic skills in Italian dyslexic and non-dyslexic learners of Chinese as a second language (L2), focusing on phonological awareness, orthographic processing, morphological awareness, and Rapid Automatized Naming (RAN). Initial findings highlight significant challenges in rime recognition and tone accuracy among dyslexic learners, with greater performance variability compared to typically developing peers. The battery offers insights into cognitive-linguistic predictors of Chinese character acquisition and aims to bridge research gaps in dyslexia and second language acquisition, ultimately supporting tailored intervention strategies for dyslexic learners in multilingual contexts.

Acknowledgements

I would like to express my sincere gratitude to the students, teachers and headmasters of the schools involved in this study, without whom this research would not have been possible.

References

Cornoldi, C., Montesano, E., 2020. Nuova Batteria per Studenti Universitari e Adulti LSC-SUA.

Chang, L.Y., Tseng, C.C., Perfetti, C.A., Chen, H.C., 2022. Development and validation of a Chinese pseudo-character/non-character producing system. Behavior Research Methods, 54(2), pp.632-648.

Favaloro, F.B., 2017. L'insegnamento della cultura cinese nell'istruzione secondaria. Osservazioni sullo stato dell'arte e costruzione disciplinare. In: T. Pellin, G. Trentin, eds. 2017. Atti del XV Convegno 2015. Venezia: Libreria Editrice Cafoscarina. Associazione Italiana di Studi Cinese.

Liu, D., Chen, X., Chung, K.K.H., 2015. Performance in a visual search task uniquely predicts reading abilities in third-grade Hong Kong Chinese children. Scientific Studies of Reading, 19, pp.307-324.

Pan, D.J., Meng, X., Lee, J.R., et al., 2024. Correction: The cognitive-linguistic profiles and academic performances of Chinese children with dyslexia across cultures: Beijing, Hong Kong, and Taipei. Annals of Dyslexia, 74, p.243.

Rello, L., Baeza-Yates, R., 2013. Good fonts for dyslexia. Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, 14, pp.1-8.

Siok, W.T., Niu, Z., Jin, Z., Perfetti, C.A., Tan, L.H., 2008. A structural-functional basis for dyslexia in the cortex of Chinese readers. Proceedings of the National Academy of Sciences, 105(14), pp.5561-5566.

Verzì, I., 2023. Il ruolo della consapevolezza fonologica nell'acquisizione della lettura dei caratteri cinesi in studenti sinofoni dislessici: un'analisi preliminare. In: Percorsi in Civiltà dell'Asia e dell'Africa III. Quaderni di studi dottorali alla Sapienza. pp. 153-157.

Wydell, T.N., Butterworth, B., 1999. A case study of an English-Japanese bilingual with monolingual dyslexia. Cognition, 70(3), pp.273-305.

Wydell, T.N., Fern-Pollak, L., 2012. Dyslexia - A comprehensive and international approach. BoD - Books on Demand.

Yang, Y., 2021. An analysis of word segmentation strategies in Chinese L2 learners. Journal of Chinese Linguistics, 49(1), pp.92-108.

Zhang, L., Xia, Z., Zhao, Y., Shu, H., Zhang, Y., 2023. Recent advances in Chinese developmental dyslexia. Annual Review of Linguistics, 9(1), pp.439-461.

Vocal stereotypes as characterisation in animated films

Alice Crochiquia[1,2], Anders Eriksson[2], Sandra Madureira[1]
[1]PUC-SP, Brazil
[2]Stockholm University, Sweden

https://doi.org/10.36505/ExLing-2024/15/0008/000633

Abstract

Humans have long associated voice with personality and physical traits, with consistent vocal stereotypes evident across cultures. This paper examines how vocal stereotypes influence character perception in animated films, focusing on the multilingual dubbing of *Zootopia* in English, Brazilian Portuguese and Swedish. Animated media provides a unique lens for studying these correlations, as character voices are intentionally matched to stereotypical features, and dubbing allows cross-language comparisons. This study considers Ohala's Frequency Code, which links vocal pitch to size and dominance, rooted in both human and animal behaviour. The analysis includes acoustic parameters, voice quality, and listener perceptions. Results show the use of higher *f*0 to portray smaller characters, in contrast with larger ones, which display lower *f*0 across all languages.

Keywords: dubbing, animated films, vocal stereotypes, voice quality, voice analysis.

Introduction

Vocal stereotypes are present not only in daily, face-to-face interactions but also permeate media and entertainment, as they function as effective shortcuts to establish a character's features to the audience. Animated films and TV series provide a unique lens for investigating vocal stereotypes, as they require voices to be matched to the physical and psychological features stereotypically correlated to them. In addition, animated features are often dubbed in different languages for international distribution, thus allowing for the comparison of how the same character sounds in various languages.

This paper outlines materials and methods chosen for a multilingual study of how vocal features are used to reflect physiological, psychological and social cues in an animated film. It leverages theories like the Frequency Code (Ohala 1984), which describes the association of *f*0 and body size and extends those correlations of size to social behaviours such as dominance and submission. The paper also presents the results and discussion of the experiments and analyses carried out so far in the study.

Material and methods

Material

The material for this study comprises the dialogue lines of the four main characters of the animated film Zootopia, in its original version in English and its dubbings in Brazilian Portuguese and Swedish. These characters are Judy Hopps (a female bunny), Nick Wilde (a male red fox), Chief Bogo (a male African buffalo) and Assistant Mayor Bellwether (a female sheep).

For the analyses, one audio sample was produced for each character in each language, edited together from dialogue lines from different scenes in the film. For Bellwether, two audio samples were produced, as the character is revealed as the film's villain and presents two different personas during the narrative. Each audio sample was around 27 seconds in length.

Perceptual experiment

The perceptual experiment for the audio samples in Brazilian Portuguese was conducted entirely online, with 77 Brazilian Portuguese native speakers, male and female, aged between 20 and 50. The judges were asked to listen to the five audios to evaluate the speaker using 14 bipolar scales speakers' size, age, temper, attitudes, character, and social and vocal features. This was done using a sliding tool to indicate the scale point they felt best fit the character, which was translated into a score from 0 to 100.

The descriptors chosen for the test were based on features displayed by the characters in the film and on descriptions given by the writers, directors, and voice actors in interviews, discussions, and promotional material.

Acoustic analysis

A modified version of the script Prosody Descriptor Extractor (Barbosa 2020) for Praat (Boersma & Weenink 2021) was used for the acoustic analysis of the Brazilian Portuguese, Swedish and English samples. The script automatically extracts measures of parameters related to *f*0, intensity, and long-term spectrum, among others.

Perceptual voice quality analysis

Voice quality analysis of the Brazilian Portuguese samples was carried out by seven phoneticians, using the VPA protocol (Laver & Mackenzie Beck 2007).

Results

Table 1 shows selected results from the perceptual experiment with audio samples from the Brazilian Portuguese dubbing. It displays the average ratings of each character on the scales Small – Big, Docile – Aggressive, Submissive – Dominant. Higher values indicate ratings closer to the descriptors listed in the table, and lower values ratings closer to the opposite descriptor. Highlighted in

bold are the strong scores (<25, >75) each character received in these scales. Bellwether 1 represents the character's friendly façade, while Bellwether 2 represents her true scheming personality.

As Table 2 shows, Chief Bogo had the lowest *f*0 values across all languages, while Bellwether 1 had the highest.

Bellwether 1 and Bellwether 2 present different voice quality settings in Brazilian Portuguese, as seen in Table 3. Chief Bogo, in addition to low *f*0, also uses voice quality settings that help lower the overall tone of voice even further.

Table 1. Average ratings for the characters in the perceptual experiment.

Characters	Big	Aggressive	Dominant
Bellwether 1	29.97	12.64	44.36
Bellwether 2	24.09	68.99	88.13
Chief Bogo	86.71	86.42	90.69
Judy Hopps	34.40	48.51	68.70
Nick Wilde	53.48	42.81	65.65

Table 2. Mean *f*0 values in the three languages.

Characters	English	Brazilian Portuguese	Swedish
Bellwether 1	237 Hz	222 Hz	242 Hz
Bellwether 2	199 Hz	178 Hz	202 Hz
Chief Bogo	109 Hz	127 Hz	124 Hz
Judy Hopps	194 Hz	203 Hz	189 Hz
Nick Wilde	148 Hz	135 Hz	150 Hz

Table 3. Voice quality and vocal dynamics settings of Brazilian Portuguese voices in Zootopia

Characters	Voice quality settings	Vocal dynamics settings
Bellwether 1	Lip Spreading, Whispery Voice	High Mean Pitch, High Mean Loudness
Bellwether 2	Harshness, Extensive Mandibular Range	High Mean Pitch, High Mean Loudness
Chief Bogo	Lip Rounding, Lowered Larynx	Low Mean Pitch, High Mean Loudness
Judy Hopps	Extensive Labial Range, Raised Larynx	High Mean Pitch, Extensive Pitch Range & High Pitch Variability
Nick Wilde	Pharyngeal Expansion, Lax Vocal Tract	Extensive Pitch Range & High Pitch Variability

Conclusions

The use of different voice quality settings for the friendly Bellwether and her conniving persona supports the use of distinct vocal features to signify different psychological features in characters. Bellwether 1's use of the Lip Spreading setting, which often correlates to perception of a smile, also raises the overall tone of a speaker's voice, as it shortens the length of the vocal tract.

The listeners' perception of body size matched the descriptions in Ohala's Frequency Code hypothesis. Chief Bogo, who was accurately perceived as the largest of the characters, had the lowest *f*0 across the three languages. In contrast, the two smallest characters, Judy and both iterations of Bellwether, had the highest values (the film inaccurately depicts Bellwether, who's a sheep, as smaller than Nick, who's a red fox).

On the other hand, the perception of aggressiveness and dominance were also influenced by voice quality settings and loudness. The combination of a low *f*0, Lowered Larynx, Lip Rounding and High Loudness in Chief Bogo's voice signalled both aggressiveness and dominance. Bellwether 2, who was perceived as dominant but not aggressive, had a combination of a high *f*0, Harshness, Extensive Mandibular Range and High Loudness.

Acknowledgements

The first author acknowledges grants from CAPES #88887.630904/2021-00 and #88881.689915/2022-01. The second author acknowledges a grant from The Swedish Foundation for International Cooperation in Research and Higher Education (STINT) (IB2015-6488). The third author acknowledges a grant from CNPq #302194/2019-3, and a grant from PIPEq-PUCSP #211463.

References

Crochiquia, A., Eriksson, A., Barbosa, P. A., Madureira, S. 2022. A perceptual and acoustic study of dubbed voices in an animated film. In Frota, S., Cruz, M., Vigário, M. (eds.), Proc. 11th Intern. Conf. on Speech Prosody, 565-569, Lisbon, Portugal. https://doi.org/10.21437/SpeechProsody.2022-115

Crochiquia, A., Eriksson, A., Madureira, S., Barbosa, P. A. 2023. Animated film character profile: the roles of voice and lexical content. In Skarnitzl, R., Volín, J. (eds.), Proc. 20th ICPHS 466-47, Prague, Czech Republic. https://guarant.cz/icphs2023/862.pdf

Liu, W., Zhang, X., Liang, C. 2023. An acoustic study on character voices of dominators and subordinates: A case study on male characters in Empresses in the Palace. In Frontiers in Communication 7. https://doi.org/10.3389/fcomm.2022.1088170

Ohala, J. J. 1984. An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice. In Phonetica, 41(1), 1–16.
. https://doi.org/10.1159/000261706

Can character viewpoint gesture guide pronoun resolution in German?

Noémi Ecsedi[1], Cornelia Ebert[1], Kurt Erbach[1, 2], Sebastian Walter[1]
[1]Department of Linguistics, Goethe University Frankfurt, Germany
[2]Department of English, Saarland University, Germany

https://doi.org/10.36505/ExLing-2024/15/0009/000634

Abstract

Co-speech gestures can affect ambiguous pronoun resolution (Goodrich Smith & Hudson Kam 2012). We report the findings of two forced choice studies investigating the hypothesis that performing a *character viewpoint gesture* from the perspective of the subject or the object influences the resolution of p- (Exp. 1) and d-pronouns (Exp. 2) in German. The results, however, do not corroborate our hypothesis.

Keywords: perspective in gesture, character viewpoint gesture, pronoun resolution in German

Introduction

Since German has a rich pronominal system, there are several pronominal forms that can in principle refer to the same singular entity, among them personal pronouns (p-pros), as for example *er* 'he$_{\text{p-pro}}$' and *sie* 'she$_{\text{p-pro}}$' and so-called d-pronouns (d-pros), for example: *der* 'he$_{\text{p-pro}}$' and *die* 'she$_{\text{d-pro}}$'. Previous research on anaphora resolution in German revealed a functional difference of the two forms regarding their coreference distribution (Bittner & E. Dery 2015). P-pros refer back to the most salient antecedent, i.e., normally the subject. However, they can also refer to the object referent (1a) and are therefore ambiguous. In contrast, d-pros are marked for non-topical reference, thus typically avoiding subject antecedents (1b) (Patil et al. 2023).

(1) Der Koch$_i$ hat den Piloten$_j$ erschossen. 'The cook$_i$ shot the pilot$_j$.'
   a. Er$_{i/j}$ war sehr jung. 'He$_{\text{p-pro } i/j}$ was very young.'
   b. Der$_j$ war sehr jung. 'He$_{\text{d-pro } j}$ was very young.'

Previous research suggests that there are different cues guiding ambiguous pronoun resolution, among them *gender*, *grammatical function*, *order of mention*, *thematic prominence*, and *givenness* (Arnold 2010). Goodrich Smith & Hudson Kam (2012) found that the preference for the second-mentioned character as referent for the ambiguous p-pro increased, when the coreferential gesture was conflicted with order-of-mention of the referents.

Perspective can be encoded in speech-accompanying gestures, as well, the distinction relevant for the present work being the one between *character*

viewpoint gestures (CVGs) on the one hand and *observer viewpoint gestures* (OVGs) on the other (McNeill 1992). CVGs depict an event from a first-person perspective and their entire body is normally involved in the gesture production. OVGs, by contrast, depict an event from a third-person perspective and only hands and arms are involved in the gesture production. Besides the obvious differences in size, CVGs have been shown to be more informative than OVGs (Beattie & Shovelton 2002). In addition, CVGs and OVGs have different preferences what they typically express (Parrill 2010).

Considering (1), one could either perform a CVG from the subject's (= cook's) perspective (henceforth, subject CVG), i.e., a shooting gesture, or a CVG from the object's (= pilot's) perspective (henceforth, object CVG), i.e., backward tumbling.

Building on Goodrich Smith & Hudson Kam's (2012) findings that gestures influence how comprehenders interpret pronouns, we investigated whether the perspective expressed through a CVG serves as a cue in ambiguous pronoun resolution in German.

Methodology

Experiment design and procedure

Two forced choice studies with a one-factorial design were conducted (GESTURE). Participants first saw a written context. Then, they saw a videotaped target utterance. Afterwards, there was a neutral written sentence with an ambiguous pronoun. Finally, participants had to decide whether the pronoun refers to the subject or object of the target sentence. The completion took approximately 15 minutes and participants were compensated with £2.70.

Materials and participants

Each experiment consisted of 15 target items, 15 unrelated filler items and 2 training items. Each written context consisted of 3 sentences in order to describe the event. The videotaped target sentence introduced two masculine discourse referents, one in the subject and one in the object position (cf. (1)). Target sentences occurred in three conditions: aligned with a subject CVG, aligned with an object CVG, or with no gesture (factor GESTURE). The neutral sentence containing an ambiguous pronoun started either with a p-pro (*er* '$he_{p\text{-}pro}$'; Exp. 1) or a d-pro (*der* '$he_{d\text{-}pro}$'; Exp. 2). The test items were constructed such that both the subject referent and object referent were in principle plausible antecedents for the pronoun. An example for an experimental item can be found in (2).

(2) **Context**: *Gestern habe ich etwas Unheimliches miterlebt. Ich war in der Bank, um Geld abzuheben. Du glaubst nicht, was passiert ist.* 'Yesterday, I witnessed something scary. I went to the bank to withdraw some money. You would not believe what happened.'
Target sentence: *Der Wachmann$_{subj}$ hat den Geldtransportfahrer$_{obj}$ angeschossen.* 'The security$_{subj}$ shot the cash transport driver$_{obj}$.'
Sentence with ambiguous pronoun: *Er/Der war übrigens noch sehr jung.* 'He$_{(p\text{-}pro)}$ /$_{(d\text{-}pro)}$ was still very young, by the way.'
Question: *Wer war noch jung?* 'Who was young, by the way?'

For each experiment 21 monolingual native speakers of German were recruited via Prolific as participants.

Predictions and results

Since p-pros preferably pick up subjects as antecedents in German and d-pros, by contrast, avoid topical reference, we hypothesized that performing an object CVG results in higher rates of choosing the object as antecedent of a p-pronoun (Exp. 1). For d-pros (Exp. 2), we hypothesized that the preference to pick up the non-topical referent (= the object) is boosted if an object CVG is performed. Thus, a main effect for GESTURE was predicted for both experiments. In Experiment 1, the object was chosen as a referent of the p-pro in 11.4% of the trials when an object CVG was performed, compared to 9% in the subject CVG and 14.3% in the no gesture control condition. A logistic regression model was fitted to the data, showing no main effect (cf. Table 1).

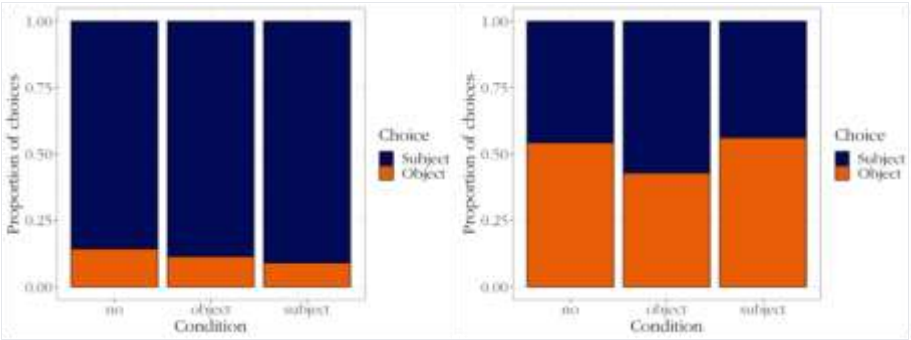


Figure 1. Proportion of choices of Exp. 1 (p-pro) and Exp. 2 (d-pro)

For Experiment 2, the object was chosen as antecedent of the d-pro in 42.9% of the trials when an object CVG was performed, compared to 56.2% in the subject CVG and 54.3% in the no gesture control condition. A logistic regression model showed a main effect for GESTURE in the pairwise comparison between the no gesture and object CVG condition (cf. Table 2).

Table 1. Logistic regression model with Gesture as fixed effect and participants and items as random intercepts (Exp. 1: p-pro).

	Estimate	**Std. error**	**z value**	**Pr(> |z|)**
(Intercept)	-2.938	.418	-7.036	1.98e-12***
No vs. obj	.428	.358	1.197	.231
Obj vs. subj	.284	.385	.736	.462

Table 2. Logistic regression model with Gesture as fixed effect and participants and items as random intercepts (Exp. 2: d-pro).

	Estimate	**Std. error**	**z value**	**Pr(> |z|)**
(Intercept)	.23	.474	.48	.962
Subj vs. no	-.095	.347	-.274	.784
No vs. obj	.872	.352	2.475	.013*

Discussion and conclusion

Against our hypothesis, the results of Exp. 1 show that object CVGs do not increase the choice of the object as the antecedent of p-pros in German. This is in line with previous findings that p-pros pick up the most salient antecedent, the subject. The results of Exp. 2 suggest that the preference of d-pros to pick up non-topical referents is not boosted when an object CVG is performed. Instead, somewhat surprisingly, the preference of d-pros to pick up a non-topical referent decreased when an object CVG was performed. Overall, the results do not corroborate our hypothesis that CVGs can influence the way that ambiguous pronouns are interpreted in German.

References

Arnold, J. E. 2010. How speakers refer: The Role of Accessibility. Language and Linguistics Compass 4, 187-203.

Beattie, G., Shovelton, H. 2002. An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. British Journal of Psychology 93(2), 179–192.

Bittner, D., E. Dery, J. 2015. The Information structural effects of German P- and D-pronouns in discourse. Papers in Linguistics 58, 49-71.

Goodrich Smith, W., Hudson Kam, C. L. 2012. Knowing 'who she is' based on 'where she is': the effect of co-speech gesture on pronoun comprehension. Language and Cognition 4, 75-98.

McNeill, D. 1992. Hand and Mind: What gestures reveal about thought. UCP.

Parrill, F. 2010. Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure. Language and Cognitive Processes 25(5), 650-668.

Patil, U., Hinterwimmer, S., Schumacher, P. B. 2023. Effect of evaluative expressions on two types of demonstrative pronouns in German. Glossa 8, 1-29.

The development of phonological awareness across different grade levels

Ibtissam El Ouakyly, Mohamed Yeou
Chouaib Doukkali University, Morocco

https://doi.org/10.36505/ExLing-2024/15/0010/000635

Abstract

Over the past three decades, phonological awareness has emerged as one of the strongest predictors of reading. Although phonological awareness has been found to correlate with reading in cross-linguistic studies of Romance languages, much is still unknown about how phonological awareness is manifested in linguistically complex countries such as Morocco. This study examines how phonological awareness is developed during early literacy stages. To assess children's phonological awareness, phoneme and syllable awareness tasks were administrated to 45 pupils. The results revealed that performance in syllable awareness tasks was higher than performance in phoneme awareness tasks across the three grades. The results are discussed in light of the Arabic orthography and the Moroccan educational system.

Keywords: Phonological awareness, children, Arabic, orthography

Introduction

Phonological awareness has been recognised as an important component of reading. Not only does phonological awareness indicate an awareness among children of the internal structure of words, but it also underlies individuals' ability to operate on and manipulate the phonemic and syllabic structure of spoken words (National Reading Panel, 2000). The importance of phonological awareness in reading is especially apparent in children's early stages of learning in which their ability to manipulate the sound structure of words is crucial for understanding letter-sound correspondences in written words. Accordingly, phonological awareness is the foundation upon which many literacy skills are based as well as one of the best predictors of children's reading development.

While research on phonological awareness has significantly improved researchers' understanding of the cognitive and linguistic factors involved in reading development, much is still unknown about the factors that enable or restrain the development of phonological awareness in linguistically complex countries such as Morocco. Thus, this study seeks to examine the development of phonological awareness in the Moroccan setting. To this end, the following research question was formulated: How is phonological awareness manifested across three grade levels?

Methodology

The present study examined the development of phonological awareness in Morocco. Data collection took place in a public elementary school in Casablanca-Settat, Morocco. Elementary schools in Morocco allocate a significant portion of the syllabus to teaching languages. In first grade, children start learning Modern Standard Arabic (MSA) and they start learning French in the second grade. The sample of the study consisted of 15 first graders (Mage = 7), 15 second graders (Mage = 7.6) and 15 third graders (Mage = 9). The selected participants speak Moroccan Arabic as their native language, and they have learnt MSA in school.

Phonological awareness was assessed through phonemic awareness tasks and syllable awareness tasks. These tasks tested children's abilities to blend, segment, and delete phonological units. The test items for the phonological awareness tasks were adopted from Saiegh-Haddad et al. (2020). 15 test items were administrated per task; words across the six tasks were matched based on phonemic length, syllabic structure, morphological structure, and orthographic length. The test items progressed in difficulty from a simple syllabic structure to a structure consisting of consonantal sequences (i.e., geminates). Cronbach's alpha corresponding to the phonological awareness tasks provided a value of .773, ensuring the reliability of the instruments used.

Results

The present study examined the development of phonological awareness across three grade levels. Phonological awareness was measured as the mean of phoneme awareness and syllable awareness tasks. Table 1 below provides descriptive statistics of the performance of participants on all phonological awareness tasks used in the study.

Task	Mean	SD
Phonological awareness	9.3860	2.02975
Phoneme Awareness	9.1373	2.27032
Syllable Awareness	9.6191	2.19871

Table 1. Descriptive statistics of all tasks.

In order to compare children's performances in phonological awareness tasks, analysis of variance (ANOVA) was performed to determine whether changes in the independent variable (i.e., grade) had significant effects on the dependent variable (i.e., phonological awareness). ANOVA revealed no statistically significant differences in phonological awareness across the three grades ($F(2,3) = 4{,}407$, $p = .128$).

Although our analysis did not find a statistically significant results, the graphical representation of the data revealed a potential trend where

participants across the three grades performed better in syllable awareness tasks than in phoneme awareness tasks (see Figure 1). In particular, Children across the three grades performed poorly in both syllable and phoneme deletion tasks in comparison to the blending and segmentation tasks. Adding to that, second graders were the highest performing group in phonological awareness tasks while first graders were the lowest performing group.

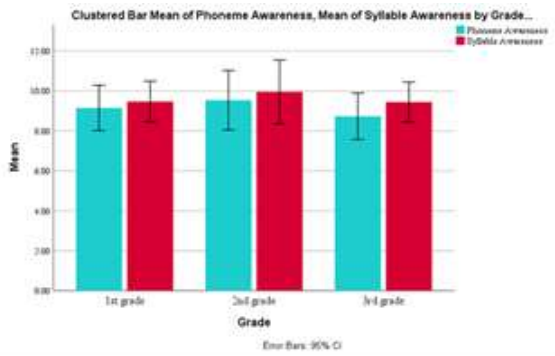


Figure 1. Children's performances in phoneme awareness and syllable awareness tasks.

Discussion

Although ANOVA found no statistically significant differences, results still do show interesting tendencies. It is important to note that second graders were the highest performing group in phonological awareness. This odd developmental curve could be explained by recent educational reforms in Morocco. Recently, The Moroccan government has introduced a new reading reform that focuses on learning phonological components through different tasks instead of focusing on text and sentence reading. Educational experts, such as Khalid Chouker, elaborated that phonological awareness is an initial but essential step in learning to read as it helps develop learners' linguistic competencies necessary to acquire reading.

Although all participants have been taught to read through the syllable-based reading method, first graders were the lowest performing group in phonological awareness. Their low performance may be due to the Arabic orthography and/or their lack of familiarity with/awareness of orthographic information. These results are consistent with Castles et al. (2003) who argue that to obtain high phonological awareness children need to first acquire orthographic information. Similarly, Ziegler et al. (2010) showed that the impact of phonological awareness in reading was regulated by the transparency of the orthography. Thus, our results as well as similar results from Castles and Coltheart (2004) highlight phonological awareness as a variable that is influenced by the orthography of the language being learnt.

Results from this study lend support to the psycholinguistic grain size theory that argues that larger phonological units are acquired earlier than smaller ones. According to the grain size theory, when learning to read, children first acquire larger grain sizes, and they gradually start to recognise phonemes as they grow older. Accordingly, Moroccan children across the three grades performed better in syllable awareness tasks than in phoneme awareness tasks. One interpretation for their high performance in syllable awareness tasks is that syllable awareness is dependent on the syllabic structure of the language being learnt. In this study, children's high performance in syllable awareness tasks may be associated with the simple syllabic structure of MSA that does not allow for complex consonant clusters. In this line of thought, Duncan et al. (2006) discovered that children who speak a language with a simple syllabic structure were more phonologically aware than children who speak a language with a complex syllabic structure. Nevertheless, it is important to highlight that the acquisition of reading among Arabic-speaking children may also be influenced by the morphological structure of Arabic. In this regard, it is recommended that future studies consider phonological, orthographic, and morphological processing skills as to acquire a more comprehensive look into reading.

Acknowledgements

This work was supported by the Centre Nationale de la Recherche Scientifique et Technique (CNRST) as part of the "PhD-Associate Scholarship- PASS" program.

References

Castles, A., Coltheart, M. 2004. Is there a causal link from phonological awareness to success in learning to read?. Cognition 91(1), 77-111.

Castles, A., Holmes, V.M., Neath, J., Kinoshita, S. 2003. How does orthographic knowledge influence performance on phonological awareness tasks?. The Quarterly Journal of Experimental Psychology Section A, 56(3), 445-467.

Duncan, L.G., Colé, P., Seymour, P.H.K., Magnan, A. 2006. Differing sequences of metaphonological development in French and English. Journal of Child Language, 33(2), 369–99. doi:10.1017/S030500090600732X

Saiegh-Haddad, E., Shahbari-Kassem, A., Schiff, R. 2020. Phonological awareness in Arabic: the role of phonological distance, phonological-unit size, and SES. Reading and Writing: An Interdisciplinary Journal 33(6), 1649–1674.

National Reading Panel. 2000. Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Jessup, MD: National Institute for Literacy.

Ziegler, J.C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., Blomert, L. 2010. Orthographic depth and its impact on universal predictors of reading: a cross-language investigation. Psychological science 21(4), 551–559.

Testing the noun-to-measure development path

Kurt Erbach
Saarland University, Germany
Goethe-University Frankfurt, Germany

https://doi.org/10.36505/ExLing-2024/15/0011/000636

Abstract

The experiment described herein tested unit words (e.g. *bunches, heaps, lots*) in various contexts thereby testing theories of how their use evolves. Participants rated the naturalness of naturally occurring sentences with *bunches, heaps,* and *lots* and sentences created by swapping the unit words in the original sentences (e.g. *lots* replaced with *heaps*). Swapping unit words across naturally occurring sentences allows testing the hypothesis that unit words grammaticalize from noun to measure, assuming *bunches* is a noun denoting objects, *lots* is a nominal denoting a unit of measure, and *heap*(*s*) is used in both ways. Superficially, our results support the idea of a uni-directional development path, but linear models fail to support this hypothesis, so further work is necessary.

Keywords: measures, pseudo-partitives, development path, grammaticalization

Introduction

Partitive structures like *a lot of the land* are said to grammaticalize over time (e.g. Koptjevskaja-Tamm 2001). In particular, it has been said that unit words follow a specific path in their evolution, namely from denoting concrete objects like portions of land (e.g. *lot*), piles of a certain size/shape (e.g. *heap*), or a kind of closely-oriented cluster (e.g. *bunch*) to eventually denoting a measure unit of large quantity (e.g. *I drank heaps/lots/bunches of coffee*). Crucially, this is argued to be a one-directional path meaning that a word will go from primarily concrete use, to mixed use, and finally to primarily measure use; once a unit word loses its concrete object denotation and only has measure uses, it will not develop concrete uses again (Brems 2011, 2012). While corpus studies support a one-way path (Brems 2011, 2012) the acceptability of 'backward' change has not been tested on this particular grammaticalization path. What follows describes a test of this path. The results appear to support a unidirectional path, but the possibility of backwards change cannot be ruled out.

Experimental design

In a dialect simulation experiment, participants are told they are rating the acceptability of sentences that come from a particular dialect. The difference in judgments between experimental conditions is assumed to be representative of the diachronic development of the phenomenon investigated (Gergel et al.

2021). This paradigm was used to test the directionality of the noun-to-measure development path.

Sentences in (1) and (2) exemplify concrete unit uses of *bunches* and *heap*, where *dense* and *low* respectively indicate the status of these unites as nouns: the bunch is dense and the heap is low so the unit words must be nouns rather than measures. (3) and (4) exemplify measure uses of *heaps* and *lots,* given *bake* and *eat* respectively indicate the status of these units as measures—e.g. the pumpkin is not baked in a heaps, rather an amount equal to heaps is baked.

(1) Dense bunches of pollen grains cling to each other in gnarled clumps.
(2) The thief considered the low heap of gems on the table.
(3) We bake heaps of pumpkin and freeze it to have on hand.
(4) People are eating lots of fruits and vegetables.

The sentences in (1)-(4) were collected from the Corpus of Contemporary American English and edited minimally be complete sentences with clear contexts. These original items were then used to make modified items: each unit nominal was swapped across contexts, yielding sentences like (5)-(16). "#" marks sentences as "odd" and "##" as "very odd."

(5) A dense heap of pollen grains cling to each other in gnarled clumps.
(6) #Dense heaps of pollen grains cling to each other in gnarled clumps.
(7) #Dense lots of pollen grains cling to each other in gnarled clumps.

(8) The thief considered the low bunches of gems on the table.
(9) #The thief considered the low heaps of gems on the table.
(10) #The thief considered the low lots of gems on the table.

(11) ##We bake bunches of pumpkin and freeze it to have on hand.
(12) ##We bake a heap of pumpkin and freeze it to have on hand.
(13) We bake lots of pumpkin and freeze it to have on hand.

(14) ##People are eating bunches of fruits and vegetables.
(15) ##People are eating a heap of fruits and vegetables.
(16) People are eating heaps of fruits and vegetables.

Brems (2011, 2012) shows that *bunches* is primarily used as a noun, *lots* as a measure, and *heap*(*s*) is one or the other depending on number. *Bunches, heap*(*s*), and *lots* therefore exemplify a noun-to-measure development path. Brems (2011, 2012) predicts one-way change from (concrete) noun to (abstract) measure. We assume that when concrete units are swapped (i.e. *bunches* and *heap*), they will receive relatively high acceptability ratings; the same is expected

for measure units (i.e. *heaps* and *lots*). However, swapping across category (e.g. *bunches* and *lots*) should result in relatively low ratings. At the same time, going from concrete unit to measure unit should be more permissible than vice versa, so the former should be less marked in the latter context than vice versa. For example, *heap* is still largely a concrete unit so using it to replace *bunches*, which is also a concrete unit, should be acceptable in (5). Because *heaps* and *lots* are largely and entirely measure units respectively, they should not be acceptable replacements for *bunches* in (6) and (7). The same should be true when *bunches* replaces *heap* (8) and *heaps and lots* replace *heap* (9) and (10). Conversely, when the measure unit *heaps* is replaced with the concrete units *bunches* (11) and *heap* (12), this should be the most marked, but when *heaps* is replaced with the measure unit *lots* (13) this should be acceptable. The same should be true when the measure unit *lots* is replaced with the concrete units *bunches* (14) and *heap* (15)—i.e. this should be the most marked—but when *lots* is replaced with the measure unit *heaps* (16), the sentence should be acceptable. Summarizing, (1)-(4) should be the most acceptable, because they are original items. (5), (8), (13), and (16) should have similarly high acceptability since they are made by swapping unit nominals of similar kind, (6), (7), (9), and (10) should be marked because they are made by swapping unit nominals of different kinds albeit along the diachronic path, and (11), (12), (14), and (15) should be the most marked because they are made by swapping unit nominals of different kinds in the opposite direction of the diachronic path.

Given three nominals are swapped across each others' contexts, we have a 3x3 Latin Square design where the factor MODIFIER has three levels (bunches, heap_s, and lots) and the factor CONTEXT has three levels (concrete, flexible, and abstract). Importantly *heap*(*s*) is in the process of change: *heap* is more concrete in use and *heaps* is more abstract in use. Because of *heap*(*s*)' status, in the experiment they are treated as a single MODIFIER (heap_s) that naturally occurs in a single CONTEXT (flexible). In practical terms, this means that half of the *heap_s* items will be singular (concrete) and half will be plural (abstract), and when *heap_s* is used in the *concrete* CATEGORY—i.e. in the *bunches* sentences—half will be *heap* and half will be *heaps*. Given the differences in acceptability predicted across all categories, a linear mixed effects model of sentence ratings should show effects of UNIT, CONTEXT, and MODIFIER* CONTEXT interaction. 60 native English participants were recruited via Prolific, given a similar study with 53% power had that many (Erbach & Gergel, forth).

Results

Most predictions were borne out: original sentences (*bunches*-concrete, *heap_s*-flexible, *lots*-measure) were generally rated highest in modifier categories, though not for *bunches*, which is unexpected. Backwards steps in the

development path were all rated lower than the present stage (heap_s-concrete<heap_s-flexible, lots-concrete<lots-flexible<lots_measure), and, for *bunches*, the first step on the path, flexible, was rated higher than the second step, measure. Results were analyzed in R with a linear mixed effects model using MODIFIER and CONTEXT as fixed effects, and PARTICIPANT and ITEM as random effects. Despite the results generally being in line with predictions, linear models compared with ANOVA did not indicate any effect of MODIFIER ($p = 0.1738$), CONTEXT ($p = 0.3237$), or MODIFIER*CONTEXT interaction ($p = 0.2924$). This may be due to the low power of the experiment (35%), which suggests that more participants or items are necessary. In summary, while the results appear to be in line with the developmental path of unit modifiers from noun to measure, the ANOVA between a linear models do not show the robustness of these results. Therefore, backwards change cannot be ruled out and further work is needed.

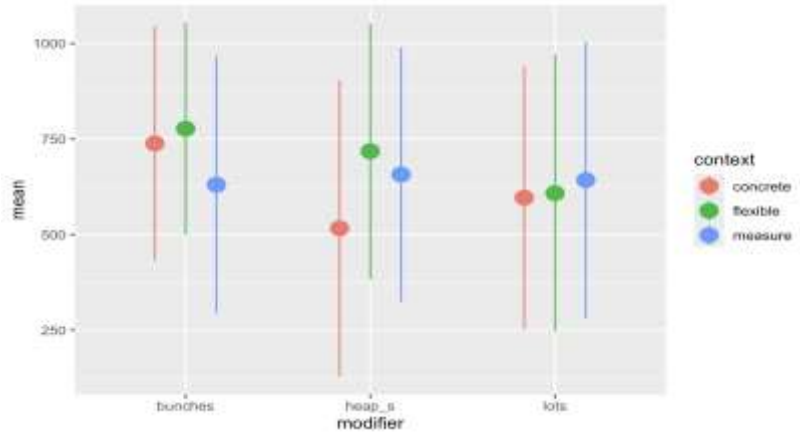


Figure 1. Average rating (point) and SD (line) of items grouped by modifier*context.

References

Brems, L. 2011. The layering of size noun and type noun constructions in English. Berlin: Mouton de Gruyter.

Brems, L. 2012. The establishment of quantifier constructions for size nouns: A diachronic case study of heap(s) and lot(s). Journal of Historical Pragmatics 13(2).

Erbach, K., Gergel, R. Forthcoming. Testing dialects with simulations: The status of pseudo-partitives in US English. Linguistica Brunensia.

Gergel, R., Kopf-Giammanco, M., Puhl, M. 2021. Simulating semantic change: a methodological note. Experiments in Linguistic Meaning, 1, 184-196.

Koptjevskaja-Tamm, M. 2001. "a piece of the cake" and "a cup of tea". In: Östen D., Koptjevskaja-Tamm, M. (eds.), The Circum-Baltic languages: typology and contact. Vol. 2 Grammar and typology, 523–568. Amsterdam: John Benjamins.

Does story-type affect the assessment of age-related differences in discourse production?

Giulia Gasparotto, Francesco Petriglia, Silvia D'Ortenzio, Andrea Marini
University of Udine, Italy

https://doi.org/10.36505/ExLing-2024/15/0012/000637

Abstract

Assessing discourse production is essential for understanding the effects of healthy aging on language across the lifespan. The choice of task for collecting speech samples is critical, as different tasks can elicit varying types of discourse. Building on prior research (Marini, 2022), we examined potential effects of story type on age-related differences in discourse production among healthy adults. Speech samples from 143 participants across various ages were collected using both single pictures and vignettes. A multilevel discourse analysis revealed age-related differences and story-type effects on measures of productivity and discourse organization. This study is relevant for clinical practice, highlighting the importance of story type in discourse production assessment.

Keywords: image; vignette; discourse analysis; story-type; cognition.

Introduction

Discourse is a complex structure that includes several levels of processing (i.e., phonological, lexical, grammatical, and suprasegmental traits). The assessment of narrative language sheds light on aspects such as cohesion and coherence, derailments and extraneous utterances that make discourse production vague and ambiguous but that are not always detectable form structured tests (Marini et al., 2011). Different methods of language elicitation (i.e., unstructured conversation, semistructured conversation, monologue with a single picture, monologue with a sequence of pictures) can lead to variations in the type of discourse produced (Leaman and Edmonds, 2023; Schnur and Wang, 2004; Mayer and Murray, 2003). For example, when describing a story depicted on an image, people tend to produce sentences with greater syntactic complexity compared to semi-structured interviews. Conversely, during spontaneous storytelling, patients tend to produce more syntactically complex sentences compared to when they describe a story using images. However, the potential effect of story type (i.e., single picture versus sequential images) on eliciting speech samples remains unclear. Single images and sequential pictures possess distinct characteristics that may result in varying performances (Fasnacht, 2023; Cohn 2012, 2020). Single images can be interpreted independently, while vignettes are connected to one another. This suggests that single images require less cognitive effort, while describing a vignette demands identifying the elements that connect the images cohesively.

This study focused on the difference in the language sample elicited through a single picture and a sequence of pictures. Based on previous findings (Marini, 2022), we hypothesized that narrative tasks elicited through single pictures or image sequences would affect specific aspects of discourse production, i.e., the ability to generate coherent links between the utterances. This issue is highly relevant for clinical practice, as speech and language pathologists (SLPs) commonly evaluate discourse by eliciting language samples using either a single image or a vignette (image sequence).

Material and methods

143 healthy adults formed 4 groups: 35 young adults (aged 20-39); 38 middle-aged adults (aged 40-59); 35 young-elder adults (aged 60-74) and 35 old-elder adults (aged 75-89). They all scored within normal range on the Montreal Cognitive Assessment (MOCA; Nasreddine et al., 2005) and the Naming subtest of the Aachener Aphasie Test (AAT; Huber et al., 1983). Their narrative production skills were assessed with a multilevel procedure of discourse analysis (Marini et al., 2011) applied to speech samples obtained with a picture-story description task using 2 single pictures and 3 vignettes. This analysis allowed us to obtain information about the participants' productivity levels, lexical informativeness, grammatical completeness, and discourse organization. Group-related differences were found on the level of education ($F(3, 139)=17.789$, $p<.001$) with the oldest group showing significantly lower levels of education than all other groups (all $ps<.01$).

Results

As group-related differences were found on the level of formal education, correlation analyses were performed between this variable and the target linguistic measures. These analyses showed that this demographic variable correlated with all target variables: productivity measures (words, $r=.342$; $p<.001$; Speech Rate, $r=.335$; $p<.001$), lexical informativeness ($r=.400$; $p<.001$), grammatical completeness ($r=.376$; $p<.001$), and the two measures of discourse Patilerrors, $r=-.286$; $p<.001$). For this reason, potential age-related and story-type differences were explored with a series of mixed ANCOVAs with age-groups as between-subject factors (1. Young adults; 2. Middle-aged adults; 3. Young elder adults; 4. Old elder adults), story type as within-subject factors (1. Single pictures; 2. Cartoon-stories) and level of formal education as covariate. Tukey's post-hoc analyses were also performed to further explore group-related differences. As for productivity, no age-related differences were found in the production of words or Speech rate. Only for words a significant story-type effect was registered ($F(1, 138)=13.073$, $p<.001$) as in all groups cartoon-picture stories elicited speech samples with more words. As for lexical

informativeness, a significant group-related effect was found ($F(3, 128)=24.852$, $p<.001$) with no story-type effect. Post-hoc analyses showed that the oldest group produced significantly fewer informative words than all other groups (all ps <.001). Similarly, group differences were found in the % of grammatical completeness ($F(3, 128)=5.638$, $p<.001$) with no story-type effect. Post-hoc analyses showed that the oldest group produced significantly fewer complete sentences than all other groups (all ps <.007). Finally, group-related differences were found also in the % both local ($F(3, 128)=16.801$, $p<.001$) and global ($F(3, 129)=20.456$, $p<.001$) Coherence Errors. In both cases a significant effect of story-type was found ($F(1, 128)=4.301$, $p<.040$) and ($F(1, 129)=16.025$, $p<.001$), respectively. In neither case a story_type*group effect was found.

Discussion and conclusion

Our results support previous evidence of age-related differences in discourse production (Hilviu et al., 2024), highlight the relation between level of education and discourse skills, and suggest the presence of story-type effects on measures of productivity and discourse organization. Indeed, after controlling for education, older participants produced stories with fewer informative words, lower levels of grammatical completeness, and increased production of both local and global coherence errors. Furthermore, in all participants the carton-picture sequences triggered the production of more words and more local and global coherence errors supporting the hypothesis that different tasks elicit different speech samples from the same speakers.

In conclusion, this study highlights the need to use varied tasks to elicit speech samples from individuals with or without brain damage. Relying on a single type of stimulus may bias the results in studies involving healthy individuals and, even more critically, affect the interpretation of linguistic difficulties that patients may exhibit during clinical assessment.

Acknowledgments

This research was supported by PRIN 2022 PNRR, Prot. n. P2022M9JCM, project title: "Standardization of the Multilevel prOcedure for discOurse analysis and Training program for narrative production in Healthy adults - SMOOTH". Avviso pubblico n. 1409 del 14/09/2022 – PRIN 2022 PNRR M4C2 Inv. 1.1. Ministero dell'Università e della Ricerca (Financed by EU, NextGenerationEU) – CUP G53D23007250001

References

Cohn, N. 2012. Comics, Linguistics, and Visual Language: The Past and Future of a Field (pp. 92–118).

Cohn, N. (2020). Your Brain on Comics: A Cognitive Model of Visual Narrative Comprehension. Topics in Cognitive Science, 12(1), 352–386.

Fasnacht, H. (2023). The Narrative Characteristics of Images. The British Journal of Aesthetics, 63(1), 1–23.

Hilviu, D., Parola, A., Bosco, F. M., Marini, A., & Gabbatore, I. (2024). Grandpa, tell me a story! Narrative ability in healthy aging and its relationship with cognitive functions and Theory of Mind. *Language, Cognition and Neuroscience*, 1–19. (in press).

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). Aachener Aphasie Test (AAT). Hogrefe.

Leaman, M. C., & Edmonds, L. A. (2023). Analyzing Language in the Picnic Scene Picture and in Conversation: The Type of Discourse Sample We Choose Influences Findings in People With Aphasia. American Journal of Speech-Language Pathology, 32(4), 1413–1430.

Marini, A., Galetto, V., Zampieri, E., Vorano, L., Zettin, M., & Carlomagno, S. (2011). Narrative language in traumatic brain injury. Neuropsychologia, 49(10), 2904–2910.

Marini, A. (2022). Cognitive and linguistic characteristics of narrative discourse production in healthy aging. In Coelho, C., Cherney, L., R., Shadden, B. (eds.) Discourse analysis in adults with and without communication disorders: A resource for clinicians and researchers (pp. 15-31). Plural Publishing, Inc.

Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. Aphasiology, 17(5), 481–497.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. Journal of the American Geriatrics Society, 53(4), 695–699.

Schnur, T. T., & Wang, S. (2024). Differences in Connected Speech Outcomes Across Elicitation Methods. Aphasiology, 38(5), 816–837.

On the intonation of tag questions in Italian: samples from central varieties

Rosa Giordano
University of Calabria, Italy

https://doi.org/10.36505/ExLing-2024/15/0013/000638

Abstract

The paper presents the phonetic description of the prosodic form of tag-questions in Rome and Perugia Italian. The analysis is based on a sample of confirmation-seeking questions in check dialogic moves in semi-spontaneous speech selected from the CLIPS Corpus of Italian. The phonetic-acoustic analysis shows that rising or high boundaries (L)H, preceded by a HL accent configuration, frequently occur. This shape is similar to the prosodic pattern of confirmation-seeking questions in the same varieties, identified in previous studies; therefore, data support the hypothesis of a homogeneous prosodic form for the class of confirmation-seeking questions.

Keywords: intonation, Italian, tag questions, prosody, spontaneous speech

Introduction

Syntax and pragmatics represent the main areas of intersection between the intonational tier of prosodic structure and language structure or use. Italian is a Romance language relevant to the study of pragmatics-prosody relations (Rossi 1998, Gili Fivela et al. 2015, Frota 2007), because, in several cases, speech act distinction and the related distinction of typologically correspondent sentence types exclusively rely on intonation. In particular, the opposition between declarative and interrogative sentences is conveyed by intonation; only *wh*-questions present morphosyntactic markers of sentence type.

Phonological and phonetic studies on Italian deal with regional varieties, also focussing on different kinds of questions in many varieties (Rossi 1998, Grice et al. 2005, De Dominicis 2002, Giordano 2006, Gili Fivela et al. 2015).

The present study aims at the investigation of one class of questions in central varieties of Italian, basing on semi-spontaneous speech. The analysis is limited to *tag-questions* occurring as confirmation-seeking questions in *check* pragmatic moves in task-oriented dialogues (e.g. Siemund 2001). In Italian, it is expected that A) question tags are isolated in a specific tone group on mid or high levels of f0, ending with rising or high boundary; on the other hand, B) differences among regional varieties of Italian may occur, with divergence in the phonetic/phonological realizations between Rome and Perugia Italian. The broad phonetic transcription is provided and discussed, avoiding to directly address a phonological interpretation.

Corpus and methodology

The CLIPS Corpus of Italian is composed of read and semi-spontaneous speech representing different regional varieties of Italian, based on a sample of speakers homogeneous as for sociolinguistic criteria[1]. 62 items have been selected from the samples of speech of Rome and Perugia. They occur in 12 task-oriented dialogues, 6 map-task and 6 spot the difference dialogues, performed by 12 speakers. All the items are confirmation-seeking questions, positively oriented, composed by a statement followed by a polar question (Fava 2001a: 123ff.); on the syntactic level, they correspond with a declarative clause followed by an interrogative alternative clause, in relation of juxtaposition (Fava 2001a: 123) (Fava 2001v: 43), like in ex. (1) (DGmtB01O_p2G#180).

(1) Ex.: ... poi ci dovrebbe essere un cuore giallo <sp> *ce l'hai, sì?*
... then there should be a yellow heart <sp> *you have it, don't you?*

The latter element, defined *tag-question* or *parenthetical question*, conveys the pragmatic value of confirmation of the truth of the statement it refers to (Fava 2001a: 123-124). In Italian, it can occur at the end or in internal position with respect to the sentence it refers to (Fava 2001a: 125) and is supposed to be separated from the statement sentence on the prosodic tier (Fava 2001a: 125).

The whole sample analysed in this study consists of tag-questions in final position. Tag-words are monosyllables (***no*** [no], ***sì*** [yes], equivalent for the English *isn't it*?) or, rarely, disyllables (about 20%) (***vero*** [true], ***giusto*** [right]).

The phonetic analysis of prosody depends on three parameters: prosodic phrasing; intonational features; metrical structure. Prosodic phrasing detection is based on phonetic-acoustic criteria generally marking prosodic cohesion: final lengthening and global trends of f0 and energy, like baseline, slope continuity or presence of resetting. Acoustic analysis is also always matched with uditive parsing.

The intonational shape of prosodic phrases identifies pitch accents and boundary configurations functional to communicative matters; to the aim of this work, their tonal sequence is intended to be the phonetic realization of the f0 target points within the syllable, either different in values form preceding ones, then configuring rising or falling stretches, or similar, then configuring a *plateaux*. F0 values are analysed in order to evaluate the target points within the speaker's range and broad phonetic transcription is made using the symbols H (High), M (Mid) and L (Low); synchronization with the syllabic tier is reported.

Metrical structure is analysed following the general phonological interpretation accepted for Italian, in which metric foots are mono-, di- or trisyllabic and left-headed, while prosodic words are right-headed.

Results and discussion

The scheme of this specific type of confirmation-seeking questions consists of two prosodic groups, the former corresponding to the statement and the latter to the tag-question. They can be separeted, on the temporal axis, by a pause, although it is not frequent neither necessary in order to obtain the uditive effect of distinct prosodic units. Example (2) (turn DGmtA02R_ p2F#62) and figure 1 presents the case of the presence of pause between the two groups.

(2) Ex.: ... cioè, io son passato intorno alla macchina rossa <lp> *no*?
... well, I turned around the red car, <lp> *isn't it*?

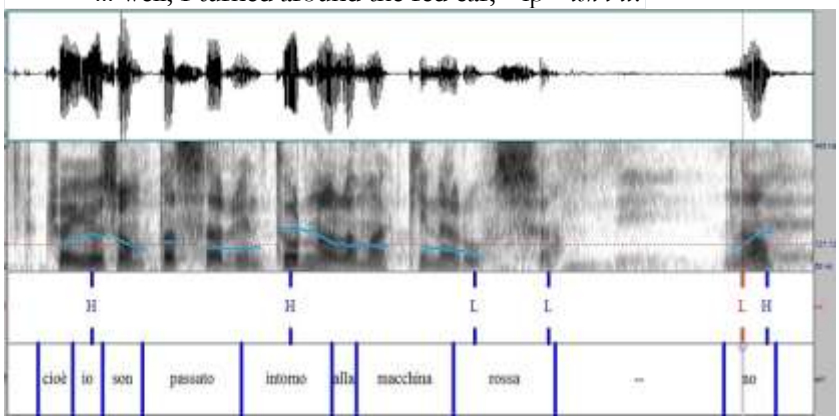


Figure 1. Ending part of the turn DGmtA02R_p2F#62 (Rome).
Tier 1: phonetic labels for target points; tier 2: orthographic transcription.

Prosodic groups corresponding to the statement always present the complete intonational pattern of a broad or a narrow focus statement in Italian, as in example (2) and figure 1 shown above; sometimes continuation boundaries can occur at the right edge. In most cases, the final accent conveys the pragmatic function of statement; its configuration is a sequence of M and L targets, both aligned within the last strong syllable at the word level on the metrical grid; in the current debate this accent is represented as H+L* in most varieties. Prenuclear accents usually identify new topic or highlight the activated part of knowledge; their configuration is a peak on high levels of the speaker's range, synchronized with strong syllables at the word level, usually transcribed in ToBI as H* (Grice *et al.* 2005, Gili Fivela *et al.* 2015).

The phonetic pattern of the second prosodic phrase, corresponding with the tag-question, can vary depending on the underlying metrical grid.

Monosyllabic words (***no***, ***sì***) almost always (about 90%) present a rising configuration from mid to high levels of the speaker's range, although the target point H never reaches the topline. The phonetic stylization is LH (Ex. 2).

Disyllabic words beginning with the strong syllable (i.e. *vero*, *giusto*), on the contrary, present three main tonal patterns. The first one consists of monotonic stretches in which both strong and weak syllables are associated with plateaux at mid-high level of the range, whose stylization would be H – H; the second pattern shows a sequence L - LH, in which mid-level accent configuration is followed by slightly rising edge contour; the third one is the most articulated pattern, a falling-rising contour on mid-high level, consisting of a falling accent and a rising edge contour, whose stylization is HL – LH.

Although these are just few cases, they consistently provide evidence of the variability of realization of the intonational forms for this group of questions. The last case is particularly interesting, because it matches the general pattern found for confirmation-seeking questions in these varieties in previous works.

Notes

[1]URL: https://www.clips.unina.it
File format: .wav; sampling rate: 22050; quantization: 16 bit. Dialogue labels: DGmtA01R, DGtdA01R, DGmtA02R, DGtdA02R DGmtB04R, DGtdB04R, DGmtB01O, DGtdB01O, DGmtB03O, DGtdB03O, DGmtA04O, DGtdA04O. Phonetic analysis: Praat (Boersma P., Weenink D. © 1992-2011, PRAAT. Doing Phonetics by Computer, http://www.praat.org)

References

Rossi, M. 1998. Intonation in Italian. In Hirst, D., Di Cristo, A. (eds.) 1998. Intonation Systems. A survey, 219-238. Cambridge: CUP.

Gili Fivela, B. *et al.* 2015. Intonational phonology of the regional varieties of Italian. In Frota S., Prieto P. (eds) 2015. Intonation in Romance, 140-197. Oxford: OUP.

De Dominicis, A. 2002. Assertive e interrogative a Bologna e Roma. In Regnicoli A. (ed.) Atti XII GFS 2001, 129-136. Roma, Il Calamo.

Giordano, R. 2006. The intonation of polar questions in two central varieties of Italian. In Hoffmann, R., Mixdorff D. (eds.) Speech Prosody, 2006, 155. Dresden TUD.

Fava, E. 2001[2]a. Il tipo interrogativo. In Renzi L., Salvi G., Cardinaletti A. (eds.) 2001[2], Grande grammatica italiana di consultazione (3) 70-126. Bologna, Il Mulino.

Fava, E. 2001[2]b. Tipi di atti e tipi di frasi. In Renzi L., Salvi G., Cardinaletti A. (eds). Grande grammatica italiana di consultazione (3) 19-48. Bologna, Il Mulino.

Grice, M., D'Imperio, M., Savino, M., Avesani, C. 2005. Strategies for intonation labelling across varieties of Italian. In Jun, S.A. (ed.) 2005, Prosodic Models and Transcription: Towards Prosodic Typology, 55-83. Oxford: OUP.

Frota, S., D'Imperio, M., Elordieta, G., Prieto, P., Vigário, M. 2007. The phonetics and phonology of intonational phrasing in Romance. In Prieto P. *et al.* (eds.) 2007, Segmental and prosodic issues in Romance phonology. Amsterdam, Benjamins.

Siemund, P. 2001. Interrogative constructions. In Haspelmath M., König E., Oesterreicher W., Raible W. (eds.) 2001. Language Typology and Language Universals. Berlin, Mouton de Gruyter.

Do you understand Clonish?

Grandon Goertz[1], Terese Anderson[2], Evan Ashworth[1]
[1]University of New Mexico, US
[2]University of Chicago, US

https://doi.org/10.36505/ExLing-2024/15/0014/000639

Abstract

Artificial intelligence (AI) voice cloning programs produce speech clones that are advertised as being very close to or even identical to naturally spoken language. We hypothesize the clones have measurable differences compared to natural speech, and these differences would be useful in a forensic analysis. This exploratory study compares the differences in the F1-F2 vowel spaces of naturally spoken speech to those of cloned speech produced by an AI cloning program. This investigation uses an innovative analysis program written specifically for this research.

Keywords: Clonish, centroid, cloned speech, vowel space

Introduction

Perceptually, cloned speech often compares to natural spoken speech and formant comparisons may not reveal differences between cloned and natural speech. We theorize that cloned speech, which by definition should be an exact numerical duplicate of naturally produced speech, is in fact different. This difference can be determined by comparing the centroids of the naturally spoken vowels to the centroids of the cloned vowels. The second part of this experiment compares the centroids of the total vowel spaces for both naturally spoken vowels and the cloned vowels.

This research employs new technology employing a modified Matlab® CentroidPolygon.mlx computer file which calculates vowel centroids and compares the centroids of the naturally spoken speech to the centroids of the cloned speech. The advantage of using centroids for comparison is that centroids consider the entire vowel portion of a w ord, centroids figure in the frequency wavelengths of each formant, centroids do not trim or discard data, and the centroid is the mathematical geometric center of each vowel or vowel space (Anderson et al., 2003).

Methods and materials

This research first used sentences that had been produced for a previous perception-production study that compared English and Greek vowel spaces (Botinis, et al., 2022). Native New Mexico English speakers each recorded carrier sentences containing the target words: *bit, beat, bet, bat*, *boot, butt, bought*,

and *bot*. Speech recordings were produced by both speakers using a Røde N microphone in a GretchKen™ Industries acoustic sound booth.

These eight carrier phrases were designed to produce the monothongs /iː, ɪ, e, æ, uː, ʌ, ɔ, ɑː/, which represent the corner vowels and edges of the English speakers' vowel spaces. The sentences were spoken clearly, and the key words were spoken with brief silence before and after each word. The vowel portions of the spoken target words were extracted with PRAAT, formants were computed, and the data was stored in a spreadsheet.

The Speechify ™ cloning program was trained on the 16 New Mexico English carrier sentences for both speakers. This program was then used to produce clones of the same sentences on which it was trained producing clone *bit*, clone *beat*, clone *bet*, clone *bat*, clone *boot*, clone *butt*, clone *bought*, and clone *bot*. for both male and female speakers. The vowel portions of the cloned words were extracted with PRAAT, clone formants were computed, and the formant clone data was stored in a spreadsheet.

The Matlab® CentroidPolygon.mlx program was used to calculate the centroids of the sixteen naturally spoken target vowels and the sixteen cloned speech vowels. The vowel spaces for both naturally spoken vowels and cloned vowels showed that the clone vowel spaces were modified and reshaped. This information is presented in Table 1 in the results section.

For the second part of this experiment, different sound recordings were necessary to test the viability of the computer algorithm. In this exercise, four participants each read a list of 82 monosyllabic words representing the vowels and diphthongs of American English with a variety of consonantal boundaries, producing 328 speech tokens. Speech recordings were produced by three female speakers and one male speaker using a Røde N microphone in a GretchKen™ Industries acoustic sound booth.

The words containing the /iː, ɪ, e, æ, uː, ʌ, ɔ, ɑː/ monothongs, which represent the corner vowels and edges of the English speakers' vowel spaces were identified in the lists and used in the study. The vowel portions of these target words were extracted with PRAAT, formants were computed, and the data was stored in a spreadsheet.

Speechify ™ was then trained on the 328 New Mexico English speech tokens, and this program was used to produce clones of the same words. Cloned vowel corner and edge vowels were again identified for the clone vowels space. The vowel portions of the cloned words were extracted with PRAAT, cloned formants were computed, and the clone data was stored in a spreadsheet.

The CentroidPolygon.mlx program was used to calculate the centroids of the individual vowels and the centroids of the vowel spaces for both natural speech and cloned speech. Chart 1 shows the locations of the natural spoken vowel

space centroids and the cloned vowels space centroids, and how much the centroids have moved.

Results

The spoken English natural vowels and the cloned English vowels were plotted on the same F1-F2 scale for comparison. It was expected that cloned vowel locations would be located close to the natural speech locations. This is based on the ideal that cloned words are perceived the same as original speech.

In contradiction, it was found that cloned vowel centroids were not located anywhere near to the corresponding centroids of natural speech. The change was especially evident in the cases of the New Mexico English vowels, /iː, æ, uː, ʌ, ɔ/ with F1 and F2 values. The change in vowel locations was also irregular and not predicted by the data.

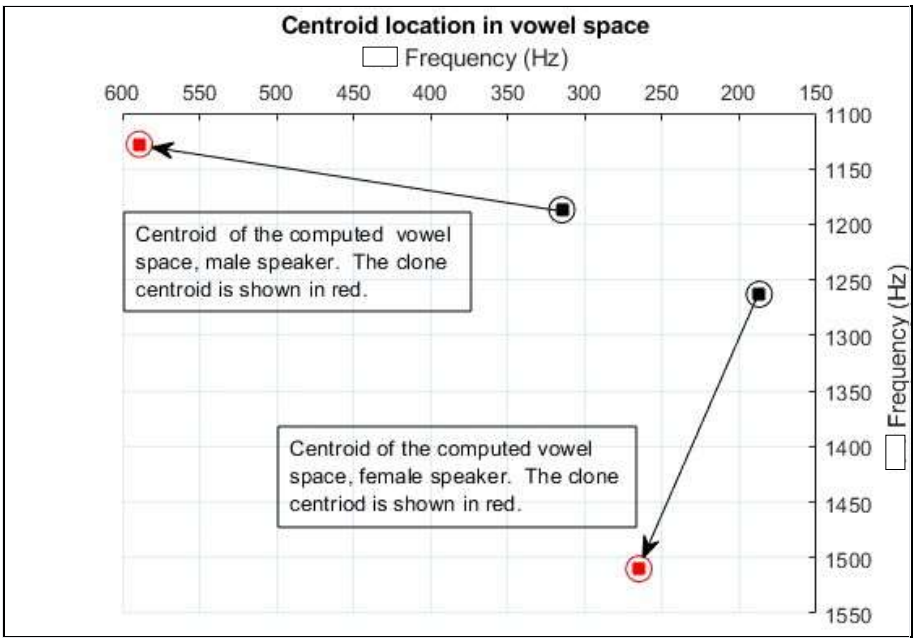
The centroids of both the natural speech and the cloned vowel spaces were computed, to show a net vowel space movement.

Table 1. The centroids values of naturally spoken and cloned English vowel spaces define different vowel spaces.

	F1 (Hz)			F2 (Hz)	
	Mean	SD		Mean	SD
Female speaker	187.	137.84		1263.3	245.72
Female cloned speech	265.1	89.73		1510.8	278.21
Male speaker	315.8	151.52		1187.2	246.50
Male cloned speech	589.2	230.80		1128.1	237.04

Significant formant changes were noted when the natural speech of each speakers' 82-word set was compared to their corresponding cloned speech word set, but the cloned vowels did not exhibit a pattern in the formant frequency changes. Therefore, centroids were computed for the total vowel space for both natural speech and the clones. Centroids represent the center of the total vowel space and show the differences between natural and cloned vowels in terms of frequency changes. The centroids for the male speaker's natural speech and clone speech, and the female speakers' natural speech and clone speech show, in summary form, the vowel shape differences.

Chart 1. A plot showing the centroid locations of the vowel spaces defined by the corner and edge vowels (/iː, ɪ, e, æ, uː, ʌ, ɔ, ɑː/) and the centroid locations of the cloned vowel spaces.



Discussion and conclusions

This research shows the vowel space centroids for the naturally spoken vowels are significantly different from the centroids of cloned vowels, and this change would be forensically useful to identify cloned speech. We understand that this is a limited study, but the mathematical algorithms that were developed are well-suited for additional study in speech analysis.

References

Anderson, T., Botinis, A., Goertz, G., Kontostavlaki, A. 2022. Vowel characterization by centroids. ExLing 2022 Paris: Proceedings 13th International Conference of Experimental Linguistics, October 13-16. Paris, France.

Botinis, A,, Goertz, G., Kontostavlaki, A., Anderson, T. 2023. Vowel discrimination of American English. ExLing 2023 Athens: Proceedings 14th International Conference of Experimental Linguistics, October 13-16. Athens, Greece.

The MathWorks Inc. 2023. MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com.

Enhancing Mandarin speakers' English pronunciation through AI-powered speech recognition technology

Wience Wing-sze Lai
The Hong Kong Polytechnic University, Hong Kong

https://doi.org/10.36505/ExLing-2024/15/0015/000640

Abstract

This study examines the effectiveness of "Reading Progress", an AI-powered speech recognition tool in Microsoft Teams, in enhancing the pronunciation and speaking skills of Mandarin speakers of English. The study aimed to determine how targeted reading practice could improve pronunciation, fluency, and confidence. Quantitative data from pre- and post-tests showed significant increases in pronunciation accuracy, with participants improving their articulation of challenging words and phrases common in job interviews. Errors with vowel sounds, consonant dropping, and simplifying consonant clusters were notably reduced. This study highlights the potential of AISRT in phonetic training for non-native speakers through personalized feedback and adaptive learning strategies.

Keywords: AI-powered speech recognition, pronunciation improvement, Mandarin speakers of English, Quantitative data analysis, phonetic training

Introduction

The integration of artificial intelligence (AI) in educational tools has opened new possibilities for enhancing language learning, particularly in the realm of phonetics. "Reading Progress", an AI-powered speech recognition tool embedded within Microsoft Teams, represents an innovative approach to improving English pronunciation, fluency, and intonation among Mandarin and Cantonese speakers. This study focuses on utilizing this technology to facilitate self-guided reading practice and provide real-time feedback to learners.

Objective

The primary objective of this study is to determine the effectiveness of "Reading Progress" in facilitating the training of English pronunciation, fluency, and intonation. By integrating this tool into the language learning curriculum, the study aims to evaluate its impact on learners' spoken English skills.

Literature Review

Recent studies have highlighted common English pronunciation difficulties faced by Mandarin and Cantonese speakers, such as confusion of vowel

distinction, dropping of final consonants, and simplification of consonant clusters or syllable structures (Zhang 2021; Liu 2019; Chen & Wang 2020). Additionally, issues with the misplacement and realization of lexical stress have been noted (Li 2020; Guo 2021). Previous research has examined various phonetic training tools, but there remains a gap in specifically addressing the integration of real-time feedback mechanisms to improve pronunciation among Chinese speakers (Derwing & Munro 2015; Thompson & Gabel 2018). By incorporating tools like "Reading Progress" into language learning, educators can provide tailored support to address specific pronunciation challenges (Wong, Tong & Lai 2022; Wang et al. 2021).

While existing studies have explored the effectiveness of different phonetic training tools, there is limited research on the impact of real-time feedback mechanisms integrated within AI-powered tools for pronunciation improvement among Chinese speakers. This study aims to fill this gap by investigating the potential benefits of "Reading Progress" in enhancing phonetic skills through personalized feedback and adaptive learning strategies.

Research questions

The study addresses the following research questions:

1. What are the differences between pre- and post-test measurements of participants' English pronunciation accuracy, fluency rates, and monotone scores?
2. How does participants' performance in these areas compare with their actual graded task performance in a mock job interview?
3. What specific phonetic aspects require further improvement?

Methodology

Participants

The research focuses on a first-year cohort (N=27) of two-year top-up degree programmes, taking a Professional English Communication course. A majority of these students came from various regions of mainland China (17 students) and had no prior EMI experience, with limited English proficiency.

Methods and procedures

Fifteen participants (Hong Kong: 7, mainland China: 8) engaged in a structured programme where they read aloud passages specifically selected for their relevance to job interview scenarios. The tool provided immediate, AI-generated scores on pronunciation accuracy, pacing, fluency, and intonation.

Besides, markups on the reading text allowed students to quickly identify mispronunciations, omissions, insertions, repetitions, self-corrections, monotone, extended pause, and rush reading. Quantitative data were collected

through the pre-test (with the "Coach function" disabled by the lecturer) and the post-test (with the "Coach function" available to students, allowing unlimited attempts before the deadline) measuring pronunciation accuracy and fluency rates, and monotone scores. The above findings (controlled practice) were also compared with the participants' performance in the actual graded task of a mock job interview (spontaneous speech).

Results

Pre-test vs post-test results

In the comparison between pre-test and post-test results, there was a general improvement in fluency, accuracy, and intonation in the post-test. This was evidenced by an increase in the number of correct words per minute for both local students (+4%) and mainland students (+5%), as well as an improvement in accuracy rate (local: +4%, mainland: +3%). Additionally, there was a drop in monotone score for both groups (local: -1%, mainland: -5%). Notably, mainland students showed greater improvement in fluency (+1%) and intonation (+4%) compared to local students, while local students exhibited a greater improvement in accuracy (+1%) than their mainland counterparts.

Mock job interview performance of students having completed RP practice vs those not

When examining the mock job interview performance of students who completed the Reading Progress (RP) controlled practice versus those who did not, it was found that students who participated in the RP practice scored higher overall in the mock job interviews. Specifically, local students scored +10% higher and mainland students scored +9% higher than those who did not engage in the practice. Among the four assessment criteria of the actual graded task of a mock job interview—spontaneous speech—students who completed the RP practice scored relatively higher in pronunciation (local: +13%, mainland: +8%), accuracy (local: +15%, mainland: +6%), content and structure (local: +6%, mainland: +12%), and fluency, intonation, and confidence (local: +5%, mainland: +6%).

Phonetic aspects for improvement

Regarding phonetic aspects for improvement, local students need to work on replacing /n/ with /l/ and altering word endings, such as missing 'ed' or adding 's'. Mainland students face challenges such as replacing /v/ with /w/, /θ/ with /s/, simplifying diphthongs like replacing /aɪ/ and /ɪə/ with /e/, /əʊ/ with /u/, adding /y/ to word-final /ʧ/ and /ʤ/, combining syllables in polysyllabic words, and misplacing word stress.

Conclusion

This study contributes to experimental phonetics by highlighting the potential of AISRT in phonetic training, particularly for non-native speakers. While previous studies have examined the effectiveness of various phonetic training tools, there remains a gap in research specifically addressing the integration of real-time feedback mechanisms in improving pronunciation among Chinese speakers with different exposures to EMI. Studies on AISRT tools like "Reading Progress" have shown promising results in enhancing learners' speaking skills through personalized feedback and adaptive learning strategies.

By integrating tools like "Reading Progress" into language learning curricula, educators can offer tailored support that addresses specific pronunciation issues, thereby equipping students with the skills necessary for successful professional communication. As the demand for proficient English speakers grows, implementing these tools can bridge communication gaps for non-native speakers, ultimately enhancing their employability and professional success.

References

Chen, H., Wang, J. 2020. Consonant cluster simplification in Mandarin ESL learners. Phonetics and Phonology Studies 12(3), 98-110.

Derwing, T.M., Munro, M. J. 2015. Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research. Amsterdam, John Benjamins Publishing Company.

Guo, Y. 2021. Realization of English stress by Mandarin speakers. Linguistic Insights 33(1), 45-58.

Li, X. 2020. Lexical stress placement in Mandarin speakers of English. International Journal of Language and Linguistics 5(4), 299-310.

Liu, Y. 2019. Final consonant deletion in Cantonese learners of English. Language Learning Journal 47(1), 67-80.

Thompson, R., Gabel, M. 2018. Real-time feedback in pronunciation training: A review. Second Language Research 34(2), 225-240.

Wang, T., Chen, X., Li, H. 2021. Adaptive learning strategies in AI-powered pronunciation tools. Computer Assisted Language Learning 34(6), 589-607.

Wong, R. F., Tong, X., Lai, E.M. 2022. Integrating AI tools in ESL education: A case study. Educational Technology Research and Development 70(1), 77-95.

Zhang, X. 2021. Phonetic challenges of Mandarin speakers in English pronunciation. Journal of Linguistic Research 45(2), 123-135.

Language learning and cognitive restructuring in bilingual French-German event processing

Pernelle Lorette[1], Theresa Gorski[1], Helen Engemann[1], Eva Soroli[2,3]

[1]University of Mannheim, Germany
[2]University of Lille, France
[3]STL CNRS Lab UMR8163, France

https://doi.org/10.36505/ExLing-2024/15/0016/000641

Abstract

This study examines the influence of linguistic properties on cognitive processing of motion events in French-German sequential bilinguals (L2-speakers). Employing verbal production and categorization tasks paired with eye-tracking, the results indicate that L2-speakers overall align their Manner lexicalization with German L1-speakers. However, low-proficiency L2-speakers exhibit reduced Manner encoding for low-Manner-salient events (e.g., walking). While preparing to verbalize, L2-speakers' visual attention develops throughout the learning process: dwell time on Manner-salient areas increases with longer L2-immersion – even exceeding L1 patterns before stabilizing. In the categorization task, a trend towards more Manner-congruent choices with increasing proficiency did not reach significance, pointing to important individual variability. These findings suggest that cognitive restructuring requires extensive experience with the target language, is task-dependent, and develops progressively.

Keywords: motion event processing, French-German bilinguals, linguistic typology, eye-tracking, linguistic and cognitive restructuring.

Introduction

Languages exhibit substantial cross-linguistic variability in how motion event semantic components are encoded (Talmy 2000). Speakers of verb-framed (VF) languages, such as French, typically encode Path (P) within the verb, relegating Manner (M) to peripheral expressions or omitting it entirely—e.g., *Il traverse la rue (en courant)* ('he crosses the street (running)'). In contrast, speakers of satellite-framed (SF) languages like English or German encode M in the verb while using peripheral elements to represent P—e.g., *Er rennt über die Straße* ('he is running across the street').

Recent research has started to explore whether and to what extent such linguistic constraints influence cognitive processing and attentional mechanisms. Lupyan et al. (2020) suggest that language shapes cognitive representations, leading speakers of typologically different languages to perceive reality differently. For example, Soroli (2024) showed that typological differences between English and French influence spatial thinking, motion event categorization, and gaze behavior. French speakers prioritize P more than

English speakers, with differences not only in verbalizations, but also in similarity judgments and some measures of visual attention, additionally guided by the nature and salience of the target events, and by the degree of language involvement in the tasks. In the domain of second language (L2) acquisition, Flecken et al. (2015) showed that while typological differences between French and German extend beyond the verbal domain —e.g., French speakers focus on moving entities and P-endpoints, and German speakers prioritize M— advanced French learners of German adopt German-like M-verb encodings, but retain French-like pre-speech attention patterns, reflecting persistent L1 influences in L2 cognitive processing. Wang and Li (2021) found that multilingual speakers undergo some cognitive restructuring toward the target language in both linguistic encoding and non-linguistic conceptualization, but the degree of restructuring depends on the level of exposure, suggesting that language learning entails internalizing progressively a new way of thinking.

Despite increasing interest, little is known about how cognitive shifts unfold across language combinations, experimental paradigms, and bilingual profiles. This study examines whether acquiring a typologically different L2 affects only verbalization or also influences visual exploration and categorization. L2 acquisition may involve either a transfer of entrenched L1 patterns (Flecken et al. 2015), or gradual cognitive reorganization (Wang & Li 2021). By integrating linguistic and cognitive paradigms, this research investigates whether L1-French/L2-German bilinguals adopt L1-German patterns in verbalization, visual attention and categorization of motion events.

Methods

28 L1 German speakers (26 females) and 21 L1-French/L2-German bilinguals (14 females) participated (*M* age = 28, *range* = 19-68) in Production and Categorization tasks (cf. Soroli, 2024). Among the L2 participants, 10 were high-proficient (*M* Lextale score = 85, *SD* = 3.5) and 11 low-proficient (*M* Lextale = 69, *SD* = 5.1). Participants first completed a socio-demographic background questionnaire and the LexTALE German proficiency test (Lemhöfer & Broersma 2012), and then were tested in the verbalization task, where they had to describe 30 videos featuring an individual performing a motion event. These events involved different Ms (with and without instruments) and Ps (with zero, one or two boundary crossings). Eye movements were recorded at 1000Hz via an Eyelink Portable Duo and Dwell times on M-salient areas (e.g., legs of the individual and, when applicable, instrument used) were measured via dynamic areas of interest (AOIs). The second task was a categorization task, where participants had to judge similarity between a target video and two options: one M-congruent (same M, different P) and one P-congruent (same P, different M) variant. Participants' choices were recorded via button press.

Results

Data from the verbalization task were analyzed using generalized linear mixed models on M-encoding in verbs. Fixed factors included language group (L1, L2-high, L2-low), M-type (with, without instrument), and their interaction, with a random intercept for item and a random slope for M-type per subject. The interaction of language group and M-type significantly improved model fit ($\chi^2(2) = 6.26$, $p = .044$). Specifically, the L2-low group showed lower M-verb encodings compared to the L1 group ($b = -1.49$, $z = -2.74$, $p = .006$) for events involving an instrument, but not for motions without an instrument ($p > .05$). No differences emerged between the L2-high and L1 groups ($p > .05$).

Linear mixed models were conducted on dwell time spent in M-salient AOIs during preparation for verbalization, with language group, M-type and P-type (zero-, one-, two-boundary crossings) as fixed factors, a random intercept for item, and random slopes for M-type and P-type per subject. Results indicated significant effects of M-type ($b = 224.78$, $t = 2.32$, $p = .027$) and P-type ($b_{one\text{-}boundary} = -621.74$, $t_{one\text{-}boundary} = -5.94$, $p_{one\text{-}boundary} < .001$; $b_{two\text{-}boundary} = 362.82$, $t_{two\text{-}boundary} = 2.79$, $p_{two\text{-}boundary} < .001$), but no effect of language: the L2-high ($M = 1345$, $SD = 809$, $b = 7.47$ $t = .07$, $p = .945$) and the L2-low groups ($M = 1263$, $SD = 794$, $b = 39.49$, $t = .38$, $p = .707$) did not differ significantly from L1 speakers ($M = 1281$, $SD = 819$) with regard to M-AOIs. However, further analyses on bilinguals' data showed a significant effect of L2-immersion length (i.e., months spent in Germany) on dwell time in M-AOIs. Longer immersion was associated with increased visual attention to M-AOIs ($b = 419.83$, $t = 3.99$, $p = .003$), even exceeding the visual attention of L1 speakers to M, before stabilizing at L1 levels, after several years in Germany (see Figure 1).

The categorization data was analyzed using generalized linear mixed models on M-congruent choices with language group, M-type and P-type as fixed factors, and random intercepts for items and for subjects. Although a descriptive trend indicated higher proportions of M-congruent choices in the L1 ($M = .51$, $SD = .5$) compared to the L2-high ($M = .41$, $SD = .49$) and L2-low groups ($M = .36$, $SD = .48$), the effect of language group was not statistically significant ($p > .05$) due to important intergroup variability. Additionally, significantly more M-congruent choices were found for one-boundary compared to zero-boundary crossings ($b = 1.31$, $z = 3.21$, $p = .001$). For L2 participants, L2 immersion descriptively increased the probability of making a M-congruent choice, but this trend did not reach statistical significance in the mixed models ($b = .73$, $z = 1.57$, $p = .116$).

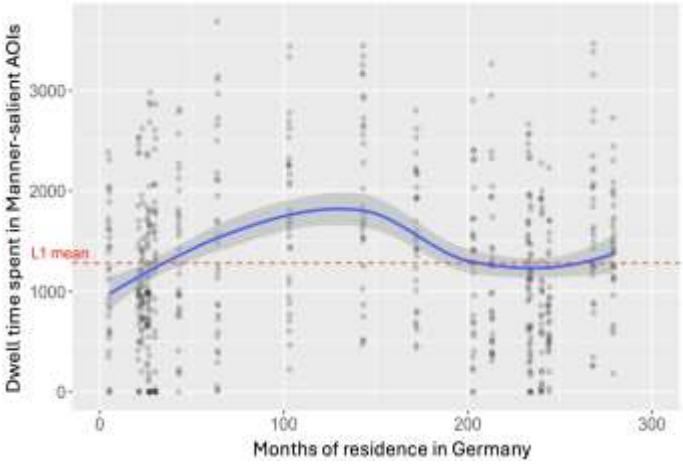


Figure 1. Dwell time on M-salient AOIs as a function of L2-immersion length.

Conclusion

The findings suggest that while L2-learners can approximate target verbalization patterns early in the learning process, aligning event cognition with target language patterns may require extensive L2 immersion. Additionally, event-specific characteristics, such as the type of M or P presented, may modulate both verbal and non-verbal behavior, independently of linguistic background. We conclude that reconceptualization is a dynamic process, impacting attention allocation and categorization, and that factors such as the salience and nature of events, as well as the extent of cumulative immersion in the L2, must be further considered when examining this process.

References

Flecken, M., Carroll, M., Weimar, K., Von Stutterheim, C. 2015. Driving Along the Road or Heading for the Village? Conceptual Differences Underlying Motion Event Encoding in French, German, and French-German L2 Users. The Modern Language Journal, 99, 100–122. doi: 10.1 111/modl. 12181

Lemhöfer, K., Broersma, M. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. Behavior research methods, 44, 325-343.

Lupyan, G., Rahman, R.A., Boroditsky, L., Clark, A. 2020. Effects of Language on visual perception. Trends in Cognitive Sciences, 24(11), 930 - 944.

Soroli, E. 2024. How language influences spatial cognition, categorization of dynamic motion events and gaze behavior: a crosslinguistic comparison. Language & Cognition, 16(4): 924-968.

Talmy, L. 2000. Toward a cognitive semantics. The MIT Press.

Wang, Y., Wei, L. 2021. Cognitive restructuring in the multilingual mind: language-specific effects on processing efficiency of caused motion events in Cantonese–English–Japanese speakers. Bilingualism: Language & Cognition, 24(4), 730–745.

Corpus linguistics for the acquisition of English lexicon

Fausto Luna, César Aguilar
Universidad Veracruzana, Mexico

https://doi.org/10.36505/ExLing-2024/15/0017/000642

Abstract

This work is a continuation of the research made by Luna & Aguilar (2022), and Aguilar & Luna (2023), focused on analysing what impact the use of linguistic corpora has on the teaching of English, specifically on students who have not had any previous approach to corpus linguistics. Our theoretical framework is in line with the proposals made by Granger (1998, 2004) regarding the use of corpora for learning English, as well as that of Boulton (2009, 2010) on language learning based on data recognition, via the use of digital corpora. Likewise, studies carried out by and Moon (2010), Szudarski (2018) or Friginal (2018) regarding the employ of corpus for vocabulary acquisition can also be mentioned here. What we present here is a quasi-experimental study, applied to a population of 36 participants, divided into two groups: one will be the intervention group and the other will be the control group. The results of this study might allow us to observe and understand the impact that teaching techniques and resources from corpus linguistics has on the acquisition of advanced vocabulary in English as well as the differences between the skills that English learners have to increase their lexicon, whether or not they are able to use computer tools such as electronic corpora. And, to recognise if the design of a course model for learning advanced English -which can be an important support for successfully passing a certification exam such as the CAE.

Keywords: CAE, corpus linguistics, English, learning, lexicon

Introduction

The use of Corpus linguistics (CL) has turned up as an increasingly relevant research topic in recent years, in particular, regarding its application in language teaching. Conspicuously, Sylviane Granger (2004) stands out as one of the key figures and pioneers in the field of corpus linguistics for foreign language learning. In this scope, corpus linguistics has proven to be a highly valuable tool for teaching foreign languages, particularly English.Consequently, this methodology, as defined by scholars such as McEnery (1996), Granger et al. (2002), and Bolaños (2015), can be employed not only for general English instruction but also to enhance specific language skills that students require to develop as part of their training.

Research problem

The research problem highlights that students at the BA Lengua Inglesa at the Languages School who are taking or have already concluded the Educational Experience (EE) Advanced English, as well as the graduates from the same educational program are not able to reach the language level desired on account of several circumstances even when they accomplished all the English levels offered by such program. As a result, several students do not feel confidence, and in turn, they find taking a C1 certification test quite complex.

Hypothesis and null hypothesis

The hypothesis presented in this paper states that the use of a CL made of sample tests of CAE might be employed as an essential tool which can allow students to acquire advanced English lexicon in order to pass such English certification successfully. Furthermore, the learners might increase their lexicon significantly. Notwithstanding, while it is true that the utilization of CL might lead to an increase in the range of vocabulary students and graduates possess, there is also the possibility of stating a null hypothesis. This is to say that, using CL could not produce significant improvements in lexicon acquisition or certification outcomes, suggesting that other factors could influence such results.

Research question and objective

This study aims to find out and asses the effect of corpus linguistic utilization on the development of advanced English lexical competence among English Language undergraduates and graduates. A pre- and post-test design is also employed, comparing an intervention group (receiving corpus-based instruction) with a control group, so as to examine the influence of the use of CL on lexical proficiency. In this regard, the research question in this work is stated as follows: What is the influence of using the CL on the lexical competence in English of the participants of the Bachelor's Program in English Language?

Background

Figure 1. The state of the art.

Methodology

This paper involves a multi-faceted research approach, combining exploratory, correlational, comparative, and quasi-experimental designs to achieve a comprehensive understanding of the research phenomenon.

Research design

It is imperative to highlight that, a mixed-method approach was adopted, integrating both quantitative and qualitative methods to collect and analyse data. The quantitative component involves the collection of numerical data through surveys and the units of analysis, while the qualitative component consists of the collection of textual data through semi-structured interviews.

Context and informants

This research paper has been developed with informants who are current students or graduates from the BA Lengua Inglesa from the Languages School of the Universidad Veracruzana. Such higher education institution is located in the capital city Xalapa, in Veracruz, Mexico.The intervention group consisted of 18 participants, comprising 12 undergraduate students and 6 graduates. All participants were enrolled in both the Bachelor of Arts in Lengua Inglesa (12 students) and the Bachelor of Arts in Enseñanza del Inglés (1 virtual student). A notable characteristic of this group was that all participants had completed an advanced English course, demonstrating a high level of proficiency in the language. Demographically, the participants' ages ranged from 21 to 47 years old. In terms of language proficiency, all participants had achieved a level of B2-B2+ according to the Common European Framework of Reference for Languages (CEFR). This homogeneity in language proficiency ensured that the intervention was tailored to the participants' needs and abilities.

Data collection methods

This research work employed a multi-method approach to collect data. A demographic questionnaire has been used to gather information about participants' background and characteristics. A satisfaction survey is administered to assess participants' perceptions and satisfaction with the intervention. Semi-structured interviews are conducted to gather in-depth information about participants' experiences and opinions. Furthermore, participant observation is definitely used so as to gather data on informants' behaviour and interactions during the intervention.

Units of analysis

Figure 2. Units of analysis.

Preliminary results, discussion and conclusion

First of all, the informants' language level at the beginning of the course ranged from B2 to B2+. In terms of certification background, the participants reported holding various certificates, including TOEFL ITP (14.3%), PET (9.5%), EXAVER 3 (9.5%), while 19% do not hold any certificate. Remarkably, all participants (100%) agreed on the relevance of Lexicon in a certification test. Aside from that, preliminary findings from this ongoing research indicate that participants initially struggled with the CAE diagnostic test, obtaining low grades and citing the test's complexity. Additionally, informants demonstrated limited familiarity with advanced lexicon, admitting to numerous unknown terms in an online test. Nonetheless, participants in the intervention group have shown keen interest in utilizing the Sketch Engine platform and Corpus Linguistics tools to improve their vocabulary register. With regular sessions and instructor guidance, participants are making progress in handling the platform and understanding new lexicon. While only preliminary conclusions can be drawn at this stage, the study aims to yield positive and significant results that support its main hypothesis. Further results will be reported upon completion of the research.

References

Aguilar, C., Luna, F. 2023. Today is another day to find you: Using Corpus Linguistics For Language Learning. In Gómez, L.; González, Ch., Lees, J. (eds.), ICERI2023 Proceedings 16th International Conference of Education, Research and Innovation, IATED Academy, Valencia, Spain: 8042-8050.

Boulton, A. 2009. Data-driven learning: reasonable fears and rational reassurance. Indian Journal of Applied Linguistics, 35(1): 81-106.

Boulton, A. 2010. Data-Driven Learning: taking the computer out if the equation. Language Learning, 60(3): 534-572. Friginal, E. (2018): Corpus Linguistics for English Teachers, London, Routledge. Granger, S. (1998): Learner English on Computer, London, UK, Longman.

Granger, S. 2004. Computer learner corpus research: Current status and future prospects. In Connor U., Upton, T. (eds.), Applied Corpus Linguistics: a Multidimensional Perspective. Brill, Amsterdam: 123-145.

Luna, F., Aguilar, C. 2022. Corpus linguistic tools and terminology translation tasks. In Botinis, A. (ed.), Proceedings of 13 th International Conference of Experimental Linguistics, International Society of Experimental Linguistics, Paris, France: 121-124.

What do we read aloud and how fast?

Philippe Martin
Université Paris Cité, France

https://doi.org/10.36505/ExLing-2024/15/0018/000643

Abstract

We don't read Word. By. Word, but instead we do read (and speak) accent phrases by accent phrases, i.e., by sequences of words containing only one (non-emphatic) stressed syllable. Only when we don't recognize globally a string of written text or when we learn another language, more precisely when we learn to read in a language we are learning, that we proceed or may proceed syllable by syllable, to progressively reach reading word by word sequences, and finally accent phrase by accent phrase to attain reading fluency aloud or silently.

Keywords: accent phrases, silent reading, theta brain oscillations, delta brain oscillations

What do we read?

We don't read Word. By. Word. and even less Syl. La. Ble. by Syl. La. Ble, but instead we do read (and speak) accent phrases by accent phrases, i.e., by sequences of words containing only one (non-emphatic) stressed syllable. Only when we don't recognize globally a string of written text or when learn to read in a language we are learning, that we proceed or may proceed syllable by syllable, to progressively reach reading word by word sequences, and finally accent phrase by accent phrase to attain reading fluency aloud or silently. In lexically stressed languages such as English or Italian, accent phrases, speaking and reading units, contain a single content word together with its associated grammatical words (pronoun, conjunction, prepositions…).

Reading French

In French however, accent phrases can contain more than one lexical item and are limited only by the maximum time it takes to pronounce them orally, 1250-1350 ms, or silently, a minimum time of 250 ms (Martin Ph. 2014). Accent phrases are also constrained by the minimum gap existing between two consecutive stressed syllables, about 250 ms. Accent phrases appear as groups of words already stored in the reader and speaker memory, with their lexical and depending grammatical words, ready to be used without any syntactic assembly effort from the speaker/reader. This is also supported by the fact that hesitations and reformulations are done with complete accent phrases in spontaneous speech, and not partially completed (Blanche-Benveniste, 2003).

Some facts about reading

Eye movements

Reading, either aloud or silently, involve eye saccades, which operate in a limited character span of some 5 to 10 characters before the focal point, and some 10 to 15 characters after the focal point, and take 70 to 80 ms (Gagl and al., 2022). These 15 to 25 characters encompass some 5 to 8 syllables, assuming an average of 3 characters per syllable.

Therefore, the eye saccades span encompasses the average number of syllables per accent phrase, varying from 1 to 8. Each eye saccade going to the next focal point is actually aiming to the next accent phrase, and not an isolated word inside the accent phrase.

Fast reading

While reading continuously, each saccade decodes an orthographic accent phrase (AP), from the AP reader lexicon entry. This conversion is synchronized by delta brain oscillations, operating in a 250 ms - 1350 ms duration range (Martin Ph., 2014).

For silent speech, there is no articulatory constrain that would slow down the reading process. Still, the recovering of the AP lexicon entry as synchronized by delta cannot go faster that 250 ms, i.e., 4 AP's per second.

This explains also the fact that we cannot have two successive stressed syllables closer than some 250 ms. Closer stressed syllables cannot be found in natural speech. When instantiated by a speech editor, the first syllable ceased then to be perceived as stressed while keeping the same acoustical characteristics.

Slow reading

If the reader is operating at a slower pace, for instance focusing on the text word by word, each word becomes an AP, and is instantiated orally or silently as an accent phrase with a stressed syllable. Going even slower, reading syllable by syllable or even character by character, each read unit actually becomes an accent phrase. Syllabic perception becomes then synchronized by delta brain oscillations.

Reading long words

Long orthographic words give an example of regulation by delta brain waves. In reading aloud, long words such as *paraskevidekatriaphobia* cannot be pronounced fast enough as one single AP. They require extra stressed syllables, usually located at some internal morphological boundary known or assumed by the reader. This applies as well to silent reading.

Speech short time memory

Breath groups in respiration cycles occur every 2-3 seconds. The expiration phase while phonation is realized should match with the listener short-time speech memory, also in the 2-3 second range (Martin R., 2014), in order to attain a successful dialog. The inspiration phase then occurs, with a minimum duration of 250 ms, giving the listener the opportunity to process the linguistic information perceived in the last 2 – 3 seconds.

Cognitive master processes

Speaking and reading aloud or silently are constrained along the time scale by the following processes:

- Theta brain oscillation: 4 Hz – 10 Hz, i.e.,100 ms - 250 ms period.
- Delta brain oscillation: 0.8 Hz – 4 Hz, i.e., 250 ms - 1250 ms period.
- Short-time speech memory (running speech): max 2000-3000 ms.
- Expiration phase in respiratory cycle: maximum about 3000 ms.

Speech synchronized processes

The expiration phase duration appears to be the master clock in the speech production system, conditioned by the short-time speech memory of the listener(s) (speaker included). Speech units are synchronized by theta and delta brain oscillations, giving a proper account for their respective range of duration:

- Syllabic duration: min 100 ms (Ghitza 2011), synced by theta oscillations.
- Consecutive stressed syllables minimal interval (stress clash): min about 250 ms (Martin, 2014) synced by delta oscillations.
- Accent phrase duration in silent reading: min about 250 ms (Martin 2014, Dunning 2010), synced by delta oscillations (Gagl, Martin, Rimmele 2022), whereas saccades in eye movement: take only 70-80 ms (Gagl et al., Rayner, Reichle et al., 2022).
- Accent phrase duration in oral reading: min 250 ms max 1350 ms (Gilbert, Boucher 2007, Martin 2014, Rimmele et al., 2021), synced by delta oscillations.
- Intonation Phrase and intermediate intonation phrase duration (CPS ended): 2000-3000 ms (Gilbert, Boucher 2007), constrained by short-term speech memory (Martin et al. 2014).
- Breath group -> Short-time speech memory (Martin et al.), IP and ip duration, also in silent reading.

Figure 1 gives an example of embedded timing of theta, delta oscillations, accent phrases and syllables.

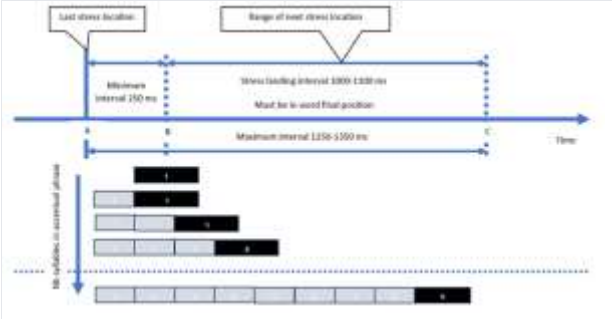


Figure 1. Reading aloud at 8 syllables / second.

References

Blanche-Benveniste C. 2003. La naissance des syntagmes dans les hésitations et répétitions du parler, in Araoui J.L. ed. Le sens et la mesure. Hommages à Benoît de Cornulier, Honoré Champion, Paris, 40-55.

Ciuffo, M., Myers, J., Ingrassia, M. et al. 2017. How fast can we read in the mind? Developmental trajectories of silent reading fluency. Read Writ 30, 1667–1686 (2017). https://doi.org/10.1007/s11145-017-9744-2.

Dunning, B. 2010. Speed Reading, Skeptoid Podcast. Skeptoid Media, Inc., 26 Oct 2010. Web. 12 Dec 2013. http://skeptoid.com/episodes/4229

Friederici A. D., Wartenburger I. 2010. Language and brain, Cognitive Science, vol. 10, 150-159, 2010.

Gagl B, Gregorova K, Golch J, et al. 2022. Eye movements during text reading align with the rate of speech production. Nat Hum Behav. 2022; 6(3):429-442. doi:10.1038/s41562-021-01215-4

Gilbert, A, Boucher, V. 2007. What do listeners attend to in hearing prosodic structures ? Investigating the human speech-parser using short-term recall, Proc. Interspeech 2007, 430-433.

Ghitza O. 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm, Frontiers in Psychol. 2011; 2: 130.

Martin Ph. 2014. Spontaneous speech corpus data validates prosodic constraints, Proc. 6th conference on speech prosody, Campbell, Gibbon, and Hirst (eds.), 525-529.

Martin R., Yan H., Schnur T. 2014. Working memory and planning during sentence production, Acta Psychologica, 152C,120-132.

Nowak, P. 2012. Speed reading tips: 5 ways to minimize subvocalization, http://www.irisreading.com/speedreading/speed-reading-tips-5-ways-to-minimize-subvocalization/

Reichle E. D., Pollatsek A. et al. 1998. Toward a model of eye movement control in reading », Psychological Review, vol. 105, n° 1, 125-157, 1998.

Rimmele J., Poeppel D., Ghitza O. 2021. Acoustically Driven Cortical δ Oscillations Underpin Prosodic Chunking. eNeuro. 2021; 8(4).

A dependency prosodic grammar for Italian

Philippe Martin
Université Paris Cité. France

https://doi.org/10.36505/ExLing-2024/15/0019/000644

Abstract

Research on Italian intonation is mainly based on the autosegmental-metrical framework, aiming to provide a prosodic grammar that adequately accounts for all well-formed sequences of phonological prosodic events. An alternative approach to account for sequences of tone patterns, or alternatively melodic contours, is presented here, using a set of dependency rules operating on accent phrases pitch events. The resulting prosodic dependency grammar was tested on a large set of more than 2700 recorded sentences read by up to 10 speakers, allowing to detect the prosodic invariants among the realizations observed between speakers when the dependency rules are applied.

Keywords: dependency grammar, Italian, prosodic structure, intonation, Siwis corpus

Introduction

In the past 30 years, a significant amount of research has been conducted on sentence intonation in Italian (e.g., Avezani, Gili Fivela 2020, De Dominicis 2010). A comprehensive list of references can be found in Gili Fivela & al. (2015)). Most studies are based on the autosegmental-metrical framework, although another approach can be found in Cresti, Moneglia (2018). Various aspects of the domain have been analyzed, including syntactic, semantic, and informational. Prosodic events are described as being aligned with syntactic boundaries, such as left and right dislocation, SN right edges, or related to informational categories, such as contrastive and narrow focus, yes/no questions, and exclamation. Although these efforts mostly pertain to short sentences in laboratory conditions, they aim to the elaboration of an intonation grammar capable of generating all well-formed sequences of prosodic events.

However, it is possible to approach prosodic structure as an autonomous linguistic object, a priori independent of the other structures of the sentence. As in the Autosegmental-Metrical (AM) model, the prosodic structure PS results from a hierarchical grouping of accent phrases (AP), the minimal prosodic units, into intermediate intonation phrases (ip), then into intonation phrases (IP), and finally into PS. Specific prosodic markers must then exist to indicate these successive groupings of accent phrases. The objective of prosodic phonology, as applied to Italian, is to describe the markers and their functions, rather than to describe well-formed sequences of tone patterns based on the sentence morphosyntactic structure.

Prosodic events

The SIWIS analysis corpus (Goldman et al. 2016) comprises over 2760 sentences read by up to 10 different native speakers of "standard" Italian, as spoken on public TV channels RAI. The corpus includes a balanced number of declarative and interrogative sentences, with word counts ranging from 3 to 48.

Prosodic events are typically associated with the most perceptually salient speech objects, i.e. vowels and possibly final voiced consonants of accent phrases stressed syllables.

Pitch accents are usually phonetically described in terms of pitch change, duration and intensity. Instead, the glissando parameter is used integrating all three acoustic parameters into one to characterize melodic variations in the sentence. Although an approximation of the listener's actual perception, the glissando value compared to a threshold allows to distinguish pitch changes perceived as such from those perceived as static tones (Rossi, 1971).

Rather than use high or low melodic targets located on accent phrase boundaries, melodic contours located of stressed syllable are retained as markers of the prosodic structure. The classes of prosodic events are then:

a. Sentence terminal déclarative reaching the lowest pitch value in the sentence **Cdec** ↓ (L*L%)
b. Sentence terminal interrogative reaching the highest pitch value in the sentence **Cint** ↑ (H*H%)
c. Rising above the glissando threshold **Cris** ↗ (L*H-)
d. Falling above the glissando threshold **Cfal** ↘ (H*L-)
e. Rising or falling below the glissando threshold **Cneu** → (H*)

Two additional contour categories are also taken into consideration: 'complex' which is instantiated on two syllables in the accent phrase, the stressed and the final (merged into one syllable if the final syllable is stressed), rising or generally falling below the glissando threshold on the accent phrase stressed syllable, and rising above the glissando threshold on the accent phrase last syllable vowel for interrogative, or below the glissando threshold but higher frequency level.

f. Sentence terminal interrogative complex (a variant of Cint) (LL*H%)
g. Sentence non-terminal complex **Ccom** √ (LL*H-), corresponding to the H* boundary tone in AM.

The crucial aspect here concerns the glissando value of each prosodic event. This value distinguishes between melodic changes above or below the threshold, which are perceived as static tone rather than a melodic change. In classical annotation, all such changes would be labeled as H*L- or L*H-, regardless of their glissando value relative to the threshold.

Prosodic dependency rules

Prosodic dependency rules govern the merging of accent phrases or groups of accent phrases based on the phonological contours described above.

Cneu → ⇨ {**Cfal** ↘, **Cris** ↗, **Ccom** √, **Cdec** ↓, Cint ↑} The neutral contour, rising or falling but below the glissando threshold, depends on any other class contour located later in the sentence (dependency "to the right").
Cris ↗ ⇦ **Cfal** ↘ (Dependency "to the left"). The falling contour above the glissando threshold depends on the first Cris found earlier in the sentence. No falling contour above the glissando threshold can be in initial position in the PS. However, falling contours below the GT may occur in initial position.
Cris ↗ ⇨ {**Ccom** √, **Cdec** ↓} The rising contour above the glissando threshold depends on the first Ccom √ or Cdec ↓ found later in the sentence (dependency "to the right").
Ccom √ ⇨ **Cdec** ↓. The complex continuity contour depends on the terminal Cdec occurring later in the sentence.

Examples

Examples in Figures 1 and 2 demonstrate how the merging of the first three accent phrases of the examples varies in congruence with syntax.

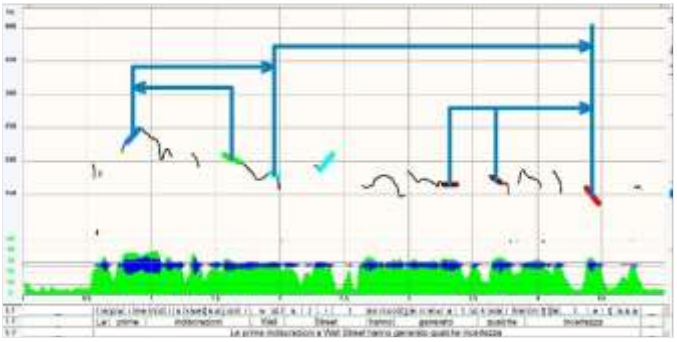


Figure 1. Color coded melodic contours aligned on stressed syllable vowels: **Cris** ↗(in blue) rising above the glissando threshold (GT), **Cfal** ↘ (in green) above GT, **Ccom** √ (in turquoise) complex non final, **Cneu** → (in brown) below GT, **Cneu** → below GT, and terminal **Cdec** ↓ (in red) [*Le pr**I**me*] ↗ [*indiscrezi**O**ni*] ↘ [*a W**A**ll StrEEt*] √ [*hanno gener**A**to*] → [*qu**A**lche*] → [*incert**E**zza*] ↓ "The first rumors on Wall Street have generated some uncertainty" (SIWIS it_a1_08_123).

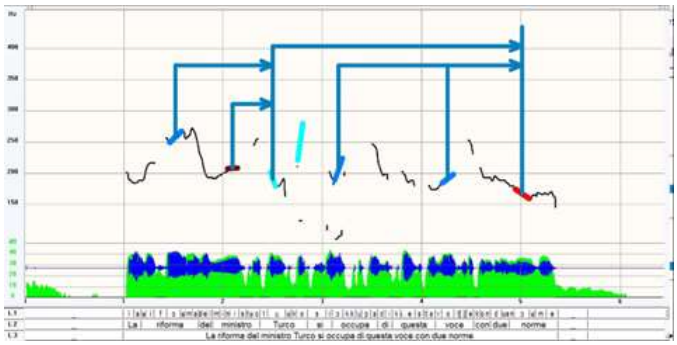


Figure 2. [*La rif**O**rma*] ↗ [*del min**I**stro*] → [*T**U**rc**O*] √ [*si **O**ccupa*] → [*di questa v**O**ce*] → [*con due n**O**rme*] ↓ "The reform of the Turkish minister deals with this item with two rules" (SIWIS it_c_22_164), is an example showing the local prosodic structure reflecting syntax ((*La riforma*) ((*del ministro*) (*Turco*))...

Table 1. Realizations of melodic contours by 10 speakers (Figure 1 example).

SIWIS ref	Le prime	indiscrezioni	A Wall Street	Hanno generato	qualche	incertezza
it_a1_08	↗	↘	√	→	→	↓
it_b_14	↗	↘	√	→	→	↓
it_b_17	→	→	√	→	→	↓
it_b_19	↗	↘	√	→	→	↓
it_b_21	↗	↘	√	↗	→	↓
it_b_18	↗	→	√	→	→	↓
it_b_27	↗	→	√	→	→	↓
it_b_40	↗	↘	√	→	→	↓
it_b_32	↗	↘	√	→	→	↓
it_b_38	↗	↘	√	↗	→	↓

References

Avesani, C., Gili Fivela, B. 2010. Curva melodica in Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/curva-melodica

De Dominicis, A. 2010. Intonazione in Enciclopedia dell'Italiano. https://www.treccani.it/enciclopedia/curva-melodica

Gili Fivela, B. et al. 2015. Intonational phonology of the regional varieties of Italian, Intonation in Romance, S. Frota, P. Prieto (ed.), Oxford U. Press, 140-197.

Cresti, E., Moneglia, M. 2018. The illocutionary basis of Information Structure. Language into Act Theory (L-AcT), in E. Adamou, K. Haude, M. Vanhove (eds.) Information structure in lesser-described languages, 359-401.

Goldman, J-Ph.et al. 2016. The SIWIS database. Proc. INTERSPEECH 2016 September 8–12, 2016, San Francisco, US.

Rossi, M. 1971. Le seuil de glissando ou seuil de perception des variations tonales pour la parole, Phonetica (23) 1-33.

An experimental study of ne-cliticization in Catalan

Clàudia Martínez-Hernández
Universitat Pompeu Fabra, Spain

https://doi.org/10.36505/ExLing-2024/15/0020/000645

Abstract

The Latin adverb inde, meaning 'from there,' has evolved into the partitive clitic ne/en in several Romance languages. While ne-cliticization has been extensively studied in Italian, research on its use in Catalan is both scarce and lacks empirical evidence. This study addresses this gap through two experiments involving acceptability tests and judgment tasks. The experiments challenge two established claims in the literature: (1) that ne-cliticization is dependent on the unaccusative status of the verb, and (2) that the presence of a locative element can cause unergative verbs to pattern like unaccusatives, thereby licensing ne as the main argument of the verb. The findings refute these claims, showing that ne can serve as the main argument of all intransitive verbs, irrespective of their unaccusative or unergative classification, and that locatives do not influence speakers' choices regarding ne-cliticization.

Keywords: partitive clitic, unaccusativity, locative, animacy

Introduction

This study investigates ne-cliticization as main argument of intransitive verbs in Catalan. The partitive clitic ne can serve different functions, including quantifying of a noun. In such cases, the clitic may represent the object of a transitive verb or the subject of an intransitive one, as shown in (1).

(1) Ja hi ha molts turistes. i a l'estiu **n**'arribaran més.
Already there are many tourists and in the summer **en**-arrive more
'There are already a lot of tourists and in the summer more of them will arrive'

This research focuses on the latter usage, where ne serves as the subject of an intransitive. The prevailing idea in the literature is that ne-cliticization is restricted to unaccusative verbs, based on Perlmutter's (1978) Unaccusativity Hypothesis, which posits that intransitive verbs can be divided into two categories, unaccusatives and unergatives, based on their semantic and syntactic behavior. Ne-cliticization has been widely defended as an unaccusativity diagnostic (Belleti and Rizzi, 1981; Burzio 1986)

and treated as a linguistic fact, despite studies that suggest that ne-cliticization may also be acceptable with unergative verbs.

In Martínez Hernandez (2024), I critically evaluate the limitations of using ne-cliticization as a diagnostic for unaccusativity and revisit the assumptions underlying the UH. In that study, I conducted an experiment to examine the acceptability and production of ne as the main argument of both classifications of verbs with other semantic variables controlled. The findings confirmed that the clitic is compatible with both verb types, contradicting the traditional hypothesis. However, the results also revealed variability in the judgements that could not be solely attributed to verb type.

Building in this initial findings, the present study seemed to test two additional variables while further challenging the dependency of ne-cliticization on the UH: (1) the animacy of the subject and (2) the influence of the presence of a locative. The latter has been proposed in the literature as a factor that might cause unergatives to "behave" like unaccusatives, thereby exceptionally licensing the partitive as the main argument (Torrego, 1989; Todolí, 2002).

Methodology

The methodology of the experiments is outlined as follows. Ninety participants took part in both experiment. The sample varied in terms of location within Catalan-speaking areas, as well as age and education level. Given that Catalan is spoken in a bilingual area alongside Spanish, and that the clitic does not exist in Spanish, the participants also completed a pretest to assess their knowledge of the general clitic system in Catalan. As mentioned previously, this experiment builds on an earlier study. The results showed that all verbs tested, both unaccusatives and unergatives, were acceptable with the clitic, however, certain verbs, while deemed acceptable, showed a lower rate of acceptability (just above 55% of participants). These verbs were selected for the new experiments to investigate whether the new variables influenced their acceptability. The goal of the first experiment was to test whether there was a difference regarding subject animacy. The experiment consisted of 10 sentences using 5 different verbs and 10 distractors. Each verb was tested twice with subjects of differing levels of animacy, allowing for direct comparisons. The verbs tested were: *volar, quedar, caure, tremolar* and *nedar* (*to fly, to be left, to fall, to tremble* and *to swim*). Subject animacy was controlled following the hierarchy proposed by Aissen (2003). Participants were asked to rate the sentences using a scale from 1 to 4. The other experiment aimed to test the effect of locatives in boosting the presence of the clitic. As in the first part, verbs were tested in pairs. However, this experiment involved a production task. The verbs *passejar, plorar, estudiar, aparèixer* and *treballar* (*to walk around, to cry, to study, to appear* and *to work*) were tested with and without a locative in otherwise identical sentences in order to evaluate the presence of the clitic.

Results

The results for both experiments showed a clear tendency that allow us to draw strong conclusions. Regarding the animacy experiment, an example of the results for one of the verb pairs is provided in Figure 1. Instead of conducting a binary acceptability task, using a graded scale allowed us to observe tendencies within the acceptable range. Considering that ratings of 3 and 4 indicate acceptability, we observed with the verb *caure* that a higher rate of 3 occurred when the subject was human, while a higher rate of 4 occurred when the subject was inanimate. This was general tendency was observed in all pairs of verbs, with lower animacy subjects yielding higher acceptability rates.

As for the production experiment, the results were similarly conclusive. An example of the findings of verb *treballar* is shown below. There was indeed a difference when the sentence presented a locative but not pertaining to the partitive clitic. Catalan also features a locative clitic, *hi*, which represents among other elements a locative expression. The results of the verb *treballar* in the graph, reflect the same trend that all the other verbs presented as well. When a locative was included, the number of participants choosing to use the partitive clitic decreased, with many opting instead for the locative clitic or a combination of the two. Therefore, we can conclude that the behavior of both clitics is opposite to that described in the literature. When a locative is added, it intuitively boosts the presence of the locative clitic, and as a consequence, decreases the use of the partitive.

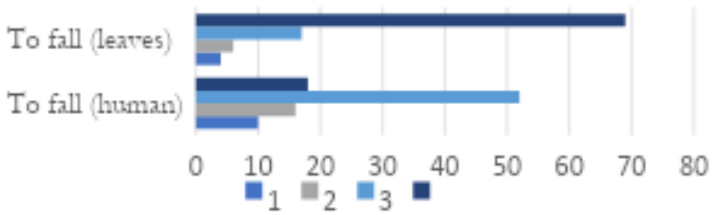


Figure 1. Acceptability results for verb *caure* (*to fall*) with different animacy in the subject.

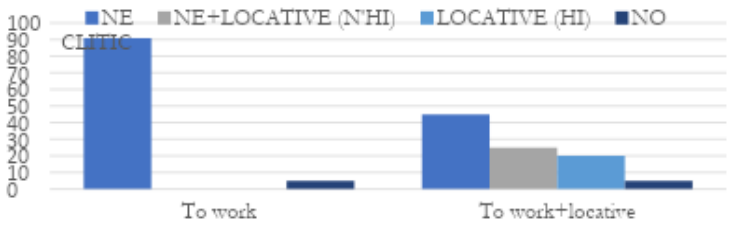


Figure 2. Production results of the clitic with verb *treballar* (*to work*), with and without locative.

Discussion

The results of this research confirm that, at the very least, *ne*-cliticization does not depend on the Unaccusativity Hypothesis. Moreover, these findings pose serious challenges to the proposal and its traditional diagnostics. This means partitive *ne* in Catalan is felicitious when cliticizing with any intransitive verb. However, this research has helped to refine our understanding of the clitic's behavior. The role of the locative, which has traditionally been linked to the clitic as a booster of its presence, has been shown not to boost the partitive. Instead, it increases the presence of locative clitic, as would have been intuitive in the first place. Additionally, the partitive has demonstrated a preference for arguments with relatively low animacy. All of the features identified in this research are associated with low prominence which aligns with the nature of the clitic, also semantically low prominent, by its indefinite and non-specific nature. In any case, this research supports the idea that a compositional interpretation of events account better for the facts than a lexicalist analysis of the verbal unit. The findings show the need to move away from rigid frameworks in favor of more dynamic approaches that account for the interplay of syntax and semantics in cliticization.

References

Aissen, J. 2003. Differential object marking: Iconicity vs. economy. Natural language & linguistic theory, 21(3), 435-483.

Belletti, A., Rizzi, L. 1981. The syntax of "ne": some theoretical implications. The Linguistic Review 1, 11–-154.

Burzio, L., 1986. Italian Syntax: A Government-Binding Approach. NY, Springer.

Martínez Hernandez, C. 2024. Ne-cliticization as unaccusativity diagnostic in Catalan. In Tănase-Dogaru, M (ed.)., RRL (LXIX 3-4., 447-460) Editura Academiei Romane.

Perlmutter, D. 1978. Impersonal passives and the unaccusative hypothesis. Proceedings of the Annual Meeting of the Berkeley Linguistics Society 38, 157–189.

Todolí, J. 2002. Els pronoms. In J. Solà, M. R. Lloret, J. Mascaró, M.Pérez Saldanya (eds.), Gramàtica del Català Contemporani, Barcelona, Empúries, 1337–1433.

Torrego, E. 1989. Unergative-unaccusative alternations in Spanish. MIT Working Papers in Linguistics 10, 253–269.

Morphological derivation as a tool for creating new words that speak to French speakers

Fanny Meunier, Raphaël Fargier, Katerina Palasis
Université Côte d'Azur, CNRS, BCL, France

https://doi.org/10.36505/ExLing-2024/15/0021/000646

Abstract

In this study, we investigate the role of morphological structure, semantic interpretability and conceptual synonymy in the processing of pseudowords in a lexical decision experiments in French. We used four types of stimuli: 1- non-morphological pseudowords (an English example would be °*sportume*); 2- semantically non-interpretable pseudowords (°*rapidation*); 3- semantically interpretable pseudowords (°*rapidify*); 4- synonym pseudowords (°*gardenist*). The results show significant effects, on lexical decision, of morphological structure and semantic interpretability, but not of conceptual synonymy, illustrating the time-course of a pseudoword's journey from representations of orthographic form to computation of meaning.

Keywords: morphology, derivation, lexical decision, pseudoword

Introduction

In natural languages, many words can be deconstructed into morphemic units, as exemplified by the French term *océanique* ("oceanic"), which consists of the root *océan-* ("ocean") and the adjectival suffix *-ique*. In languages such as English and French, the formation of complex words generally involves attaching an affix to a base morpheme, with the meaning of the resulting word often derivable from the meanings of its constituent parts. Morphologically complex words account for approximately 73% of the entries in a French dictionary (Le Robert Méthodique, 2000) but often with a very low frequency (New et al. 2004). Although listed in the dictionary, these words are unlikely to have been encountered by most speakers; nevertheless, they are often understood during conversations.

Everyday, speakers of a language encounter unfamiliar words with no difficulties to understand them and even more without noticing they are new. Often these are novel combinations of existing morphemes. Take, for example, the pseudoword °*futurness*. It combines the root *futur* with the suffix *-ness*, found in existing complex words like *happiness*. The root conveys the core meaning, while the suffix provides information about grammatical and semantic categories; *vintageness* thus refers to the "quality of something vintage". An English speaker intuitively understands these rules and categories, enabling

them to grasp the meaning of these words and define them even without context.

From a psycholinguistic point of view, the fact that language users can understand complex words that they have never encounter before suggests that some form of morphological information is available in the mental lexicon. This raise the question of how are people able to tell if a given letter string corresponds to a word of their language or not. When performing a lexical decision task, are participants consulting a stored whole-word representation of existing complex words or are they relying on the combination of the two morphemes? Existing research has shown that the lexical decision task is sensitive to morphological structure, grammaticality and semantic interpretability (Beyersmann et al. 2020).

In the experiment presented here, we look at lexical decision on pseudowords for which we manipulated their morphological structure, and explore the role of semantic features. In particular, we tested the role of synonymy: Is it more difficult to reject a pseudoword like *broadify*° because it has the same compositional meaning as *broaden*, an existing word built on the same root? Or would the existence of *broaden* in the mental lexicon make clearer that *broadify*° is not an existing word?

Experience

Materials

We tested four different types of pseudowords, identical to those used as primes in Longtin and Meunier (2005) and Meunier and Longtin (2007): a) non-morphological pseudowords, which were made with an existing French root and an existing orthographic, non-suffixal ending (e.g. *rapiduit*°, made with *rapide* "rapid" and *-uit*); b) non-interpretable pseudowords, made of the ungrammatical combination of an existing root and an existing suffix of French (e.g. *sportation*°, made of the French noun *sport* and the suffix *-ation*, which creates nouns out of verbs); c) interpretable pseudowords, made of the grammatical and semantically interpretable combination of a root and a suffix (e.g. *rapidifier*° "rapid+ify"); d) synonym pseudowords, which were constructed by swapping the suffix of an existing derived word for a semantically equivalent suffix (e.g. *jardineur*°, which was coined by changing the profession suffix *-ier* of the existing word *jardinier* "gardener" with the profession suffix *–eur*; an English equivalent would be *gardenist*. All the pseudowords respected the graphemic and phonological rules of French and could be easily pronounced.

List composition

Overall, we tested 29 non-morphological pseudowords, 28 non-interpretable pseudowords, 28 interpretable pseudowords and 29 synonym pseudowords.

These were distributed among 3 lists, to avoid root repetition. One hundred and ninety fillers (76 simple pseudowords and 114 existing words) were added to the 38 test pseudowords of each list.

Participants

Thirty native speakers of French participated to the experiment. They were randomly assigned to one of the three lists.

Procedure

Each stimulus was presented on the screen for 2000 ms or until the participant pushed a button. The inter-stimulus interval was of 750 ms. Stimuli were randomly presented, in the middle of the screen. Participants were told that they were going to see letter strings and that they had to say, as quickly and accurately as possible, if the letter string corresponded to a word of French or not. The session started with a practice run of 16 items.

Results

Overall, the error rate averaged 5.5 %, a standard rate for such type of task. Response times and error rates for the critical pseudowords were analysed with a GMM including Condition (the 4 types of pseudowords) as a fixed factor and participants and items as random ones.

Condition was significant for response times ($X2(4)=9373$, $p<.001$) and error rates ($X2(3)=63.2$, $p<.001$). Planned comparisons showed that non-morphological complex pseudowords are easier and faster to reject (1.4 % of errors; 844 ms, SD=266) than morphologically complex pseudowords even if non-interpretable (6.1 %; 950 ms, SD=300), and in turn non-interpretable pseudowords are easier and faster to reject than interpretable pseudowords (20.4 %; 1054 ms, SD= 355) and synonym ones (28.9 %; 1051 ms, SD=339); all $ps<.01$ with Bonferroni corrections. These two latter conditions are equivalent and no effect of synonymy is observed neither on response times nor on error rates.

Discussion and conclusion

Our results showed a significant effect of morphological structure and semantic interpretability, both in the response time and the error rates for the different types of pseudowords. Overall, this shows how lexical decision latencies and error rates, in addition to priming effects (e.g. Meunier, Longtin 2007), reflect the time-course of a pseudoword's journey from representations of orthographic form to computation of meaning. Non-morphological pseudowords are rejected at the very

beginning of the recognition process as they cannot be successfully mapped onto orthographic representations during an early, morpho-orthographic decomposition stage. In contrast, non-interpretable pseudowords are correctly processed at this morpho-orthographic decomposition stage but can be rejected since later semantic/syntactic integration fail to generate a coherent grammatical and semantic interpretation of their constituent morphemes. Both interpretable and synonym pseudowords are processed successfully at both processing stages, as they can be easily decomposed and interpreted. The existence of a synonymous word in the mental lexicon does not appear to modulate the decision time. Such model not only explains how language users process novel complex words with such ease, but also why unknown morphologically complex words encountered in real life and in an appropriate sentential context are so easily processed.

References

Beyersmann, E., Mousikou, P., Javourey-Drevet, L., Schroeder, S., Ziegler, J. C., Grainger, J. 2020. Morphological processing across modalities and languages. Scientific Studies of Reading, 24(6), 500-519.

Le Robert Methodique, 2000. Dictionnaire méthodique du français actuel. Le Robert, Paris, 1 vol. 1645 p. rédaction dirigée par J. Rey-Debove.

Longtin, C.-M., Meunier, F 2005. Morphological decomposition in early visual word processing, Journal of Memory and Language, 53, Issue 1, 26-41.

Meunier, F., Longtin, C. 2007. Morphological decomposition and semantic integration in word processing. Journal of Memory and Language, 56, 457-471.

New, B., Pallier, C., Brysbaert, M., Ferrand, L. 2004. Lexique 2 : A New French Lexical Database. Behavior Research Methods, Instruments, & Computers, 36 (3), 516-524.

Factors affecting resolution of anaphora with collective nouns in Russian

Anna Moskaleva
Lomonosov Moscow State University, Russia

https://doi.org/10.36505/ExLing-2024/15/0022/000647

Abstract

It is assumed that anaphoric agreement is more sensitive to semantic rather than grammatical information of an antecedent. Higher linear distance between a pronoun and its antecedent has proven to trigger semantic agreement, yet hierarchical distance is hardly examined in the research field. In the eye-tracking experiment, we investigated factors of linear and rhetorical (hierarchical) distance as well as the factor of working memory volume in their ability to facilitate interpretation of the number feature of a collective noun in Russian. The results indicate that linear distance is of great importance in interpreting pronouns, and the same is true for hierarchical distance.

Keywords: anaphoric agreement, rhetorical structure, collective nouns, eye-tracking, experimental linguistics

Introduction

The dualistic nature of collective nouns (holistic and distributive interpretations) is considered a source of variable patterns in noun-pronoun agreement in terms of number feature. One of the key factors influencing the preference for number is linear distance (Levin 2001). In a multi-factorial approach to referential choice (Kibrik 1996), one of the main discourse factors influencing referent activation in the listener's cognitive system is rhetorical distance, consisting of elementary discourse units that convey communicative information. The activation correlates with the probability of being pronominalized by the speaker.

In a normal case, morphosyntactic number on an agreement target corresponds to the number morphologically expressed on a controller, but there are cases where a violation of such alignment can be due to the influence of conceptual number on the controller. The extent to which number value is prone to inherit conceptual plurality correlates with the type of dependency according to Corbett's Agreement Hierarchy (Corbett 1979).

Hypotheses

The aim of this study is to understand the effects of linear and rhetorical distance and verbal working memory (VWM) in relation to the resolution of

anaphora with collective nominals. For this purpose, an eye-tracking experiment has been carried out to reveal the significance of these factors.

In the present study, the following hypotheses have been established:

1. An increase in rhetorical distance leads to a better processing of plural pronouns, and the opposite is true for singular pronouns.
2. An increase in linear distance leads to a better processing of plural pronouns, and the opposite is true for singular pronouns.
3. Rhetorical distance plays a greater role in pronoun processing than linear distance.
4. For participants with smaller VWM volume, it will take more time to process singular pronouns with an increase in rhetorical and/or linear distance.

Methodology

Thirty-two native speakers of Russian (students of MSU) participated in the experiment. The experiment was constructed in PsychoPy-2022.2.5., and the eye-tracking data have been collected with the help of eye-tracker Tobii Pro Spectrum 600. Participants were presented with composed texts using self-paced reading technique (non-cumulative). They had to read the texts silently and press the space button to proceed further. About 30% of texts were followed by a comprehension question alluding to one of the text constituents. The subjects were to respond out loud. Each of the 8 experimental lists contained 32 experimental items as well as 32 filler texts. All items were organized according to Latin Square Design.

Stimuli

The experiment has 2x2x2 design, and the following independant variables: rhetorical distance, linear distance, and number feature of the pronoun. The experimental texts have the same rhetorical structure (with an accuracy of relation direction) for all 8 conditions in one block. No plural noun phrases, nouns with variable gender, comitative phrases, animate referents, and nouns with the same gender as the collective noun are present in the texts (except the clause containing pronoun). The linear and rhetorical distance units represent finite clauses. The pronoun and its antecedent take the direct object position of the indefinite-personal clause.

An example of rhetorical structures for the experimental block is shown in Table 1, where four of eight conditions are given in one cell representing singular and plural pronoun cases.

	RhetD = 1	RhetD = 2
LinD = 1	1-5 | 1-2 ← 3-5 | | 1← 2 **3** ← 4-5 | **4** ← 5	1-5 | 1-2 ← 3-5 | | 1← 2 **3** ← 4-5 | **4** → 5
LinD = 2	1-5 | 1-2 ← 3-5 | | 1← 2 **3** ← 4-5 | 4 → **5**	1-5 | 1-2 ← 3-5 | | 1← 2 **3** ← 4-5 | 4 ← **5**

Table 1. Example of experimental block. RhetD and LinD stand for rhetorical and linear distance accordingly.

Results

In this paper, only mean fixation durations of the second-pass pronoun reading are considered (Table 2). Confirmations of the hypotheses based on the metric of second-pass reading of the pronoun are demonstrated in Table 3.

A crucial disparity between singular and plural pronoun conditions was demonstrated with an increase in one distance and the other one being invariably high. The most challenging condition for a singular pronoun turned out to be the highest distance, both rhetorical and linear (178.7 ms), and the easiest one was when both distances were small (74.8 ms). The opposite is true in the case of a plural pronoun.

The verbal working memory factor did not exert any influence on the processing, thus statistics are not provided here. Irrelevant, not up-to-date assessment of VWM volume is deemed to be a potential reason for such a result.

Table 2. Mean duration of fixations (ms) for second-pass reading of a pronoun region.

Conditions	RhetD = 1, LinD = 1	RhetD = 1, LinD = 2	RhetD = 2, LinD = 1	RhetD = 2, LinD = 2
singular pronoun	74.8	83.9	117.6	178.7
plural pronoun	120.6	75.6	118.3	84.7

Table 3. Confirmation of the hypotheses. T, U stand for confirmation of a hypothesis by t-test or U-test respectively. The significance level for statistical tests is .05 (*) and .01 (**).Time() stands for reading time of singular or plural pronouns. >,< stand for more or less reading time accordingly.

LinD ↑ → time(sg) ↑	RhetD↑ → time(sg) ↑	LinD↑ → time(pl)↓	RhetD↑ → time(pl) ↓	RhetD>LinD → time(sg) > time(pl)	RhetD<LinD → time(sg) < time(pl)
No (RhetD=1), Yes (RhetD=2) (T*)	Yes (LinD=1) (T*, U*), Yes (LinD=2) (T**, U**)	Yes (RhetD=1) (T*), Yes (RhetD=2) (U**)	No (LinD=1), No (LinD=2)	No	No

Discussion

In the current study, an eye-tracking experiment has been conducted to investigate factors influencing resolution of anaphora with collective nominals, namely the factors of rhetorical and linear distance and the verbal working memory factor.

Linear distance facilitated less cognitive effort in the processing of plural pronouns, whereas rhetorical distance was found to cause overall cognitive load on the process of interpretation. The results have shown that a significant increase in reading time in the case of a singular pronoun was demonstrated when both distances were high, and no such effect was observed otherwise. The working memory effect has not been revealed in the experiment. The experimental findings also suggest that, apart from the type of dependency, the preference for conceptual or grammatical information correlates with the distance between a pronoun and its antecedent.

Acknowledgements

The research has been carried out as a part of a Bachelor's thesis at Lomonosov Moscow State University. I would like to express gratitude to my supervisor, professor Olga V. Fedorova, for the helpful guidance and provision of the eye-tracking equipment. I also thank all the participants who took part in the experiment for their valuable contribution.

References

Corbett, G. G. 1979. The Agreement Hierarchy. Journal of Linguistics 15, 203–224.

Kibrik, A.A. 1996. Anaphora in Russian narrative prose: A cognitive account. In B.Fox (ed.), Studies in Anaphora. Amsterdam and Philadelphia: John Benjamins.

Levin, M. 2001. Agreement with collective nouns in English. Department of English, University of Lund.

Speaker and prosodic peculiarity classification in emotional speech

Neda Mousavi[1], Sven Grawunder[1,2]

[1]Martin Luther University Halle-Wittenberg, Germany
[2]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

https://doi.org/10.36505/ExLing-2024/15/0023/000648

Abstract

In this study, the relationship between rhythmic metrics, emotion recognition, and speaker variability is investigated using the German emotional speech corpus (VMEmo). Using principal component analysis and linear discriminant, the results show accuracies close to 0.40 when rhythmic features from different acoustic domains of time, intensity, and frequency are merged to identify linguistic behavior. However, the fluctuating accuracies of 0.44 to 0.17 in classifying speakers based on specific rhythmic feature categories emphasize the significant differences within these feature subgroups. These variations suggest possible nuances or complexities that require deeper exploration and thorough investigation to better understand the differences between these features and their impact on speaker classification accuracy.

Keywords: rhythm, speaker classification, between-speaker variation, prosodic peculiarity, emotional speech

Introduction

In the literature, rhythmic patterns in speech have been introduced not only as potential markers for individual features but also as cues for emotion recognition (Lykartsis, 2020; Mefiah et al. 2015). In this study, we investigate the interaction between rhythmic measures, emotion recognition, and speaker variability, focusing on the German emotional speech corpus (VMEmo, Batliner et al., 2000). The emotional data in this corpus were elicited through an experimental method in which participants interact with machines in tasks such as making an appointment. The experiment was purposefully designed to elicit various emotional fluctuations in the speaker as a result of the machine's responses. However, emotional fluctuations were not labelled directly, rather the corpus tags mark the speaker's linguistic behavior during emotional expressions. With our focus on rhythmic patterns, we examined the prosodic dimensions of this linguistic behavior and considered the prosodic tags of each utterance, including peculiarities such as pauses between words, strong contrastive stress, pauses between syllables, syllable lengthening, etc. The research focused on answering these two questions: how accurately can we identify different prosodic strategies for expressing emotion based on rhythmic

indicators? and, how effectively can speakers be distinguished based on these rhythmic indicators?

Method

Data preparation involved extracting key information from the VMEmo corpus, including start and end times of human-generated segments, speaker phrases, and prosodic peculiarity tags. Machine-generated segments were excluded, and waveforms from 33 speakers were selected. Phrases under 4 seconds were omitted following Tilsen and Arvaniti (2013). Automatic segmentation was conducted using WebMAUS (Kisler et al., 2017), and annotations were enriched in Praat TextGrids with details on phrase numbers, peculiarity tags, and consonant/vowel intervals.

In feature selection, rhythmic indices such as %V, ΔC (Ramus et al., 1999), nPVIv, rPVIc (Grabe and Low, 2002), varcoC and CV rate (Dellwo, 2006) were examined. Following He and Dellwo (2016), intensity-based rhythmic indicators, including the SD and nPVI of peak and mean intensity values, were analyzed. Also referring to Mousavi and Grawunder (2023), metrics from the frequency domain reflecting intensity indicators were also included. In total, 14 rhythmic features were categorized into four subgroups: duration-based, intensity-based, frequency-based and all metrics.

Results

The study employed principal component analysis (PCA) and linear discriminant analysis (LDA) as the models of analysis. Figure (1) illustrates the distribution of the values of the first component across different speakers, sorted by the variance of values. The distribution of PC1 values within each violin plot shows the range and distribution of these values across the different phrases uttered by each speaker.

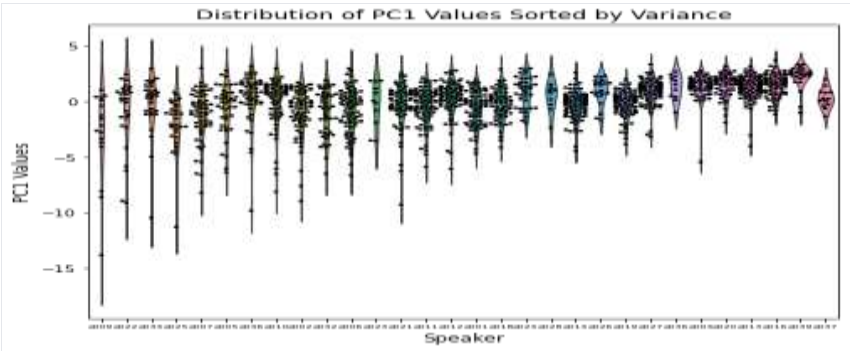


Figure 1. Visualizing within-speaker variability through PCA Analysis.

Figure (2) depicts the variance among speakers based on their PC1 and PC2 values. Closely clustered points indicate speakers with similar rhythmic patterns, while greater distances between points suggest greater variability in the rhythmic measures.

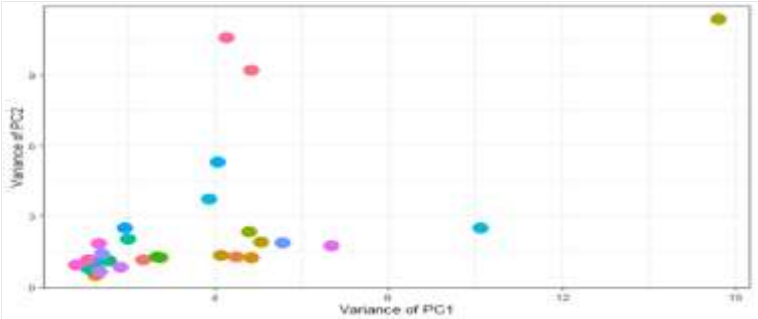


Figure 2. The variance among speakers based on their PC1 and PC2 values

Following this, we employed linear discriminant analysis for two main tasks: classification of phrases into prosodic peculiarities, regardless of speaker identity and evaluate the discriminative power of rhythmic features in distinguishing prosodic patterns, and speaker recognition, which focuses on speaker identification under the abstraction of emotional context. In addition, rhythmic groupings were compared across acoustic domains for their discrimination potential.

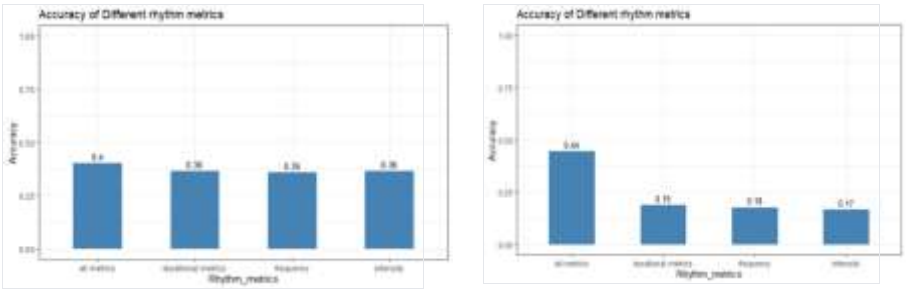


Figure 3. The accuracy of the recognition of prosodic peculiarities (left) and speaker (right)

While the differences in accuracy between these rhythmic metric categories suggest slight discrepancies in their effectiveness, the significance of these discrepancies requires further investigation. However, the notable differences in these accuracy values suggest possible differences in the discriminative abilities of these metrics for speaker discrimination.

Discussion

In this study, we investigated the effectiveness of rhythmic features extracted from vowel and consonant intervals in three different acoustic domains. The results show an approximate accuracy of 40% in distinguishing linguistic behaviors using rhythmic features in different acoustic domains. Furthermore, speaker identification achieved a similar level of accuracy when merging features from all acoustic domains. These results are consistent with previous research using rhythmic indicators extracted from different approaches to rhythm measurement, e.g. from Music Information Retrieval (Lykartsis, 2020). However, the interpretation of the accuracy levels achieved depends on their context of use and must be evaluated in accordance with the intended application and study objectives. Further research could also include a detailed investigation of the robustness of these rhythmic features in different emotional states, different linguistic contexts, or even in specific communication environments to allow a more nuanced understanding of their applicability.

References

Batliner, A., Huber, R., Niemann, H., Nöth, E., Spilker, J., Fischer, K. 2000. The Recognition of Emotion. In: Wahlster, W. (Ed.), Verbmobil: Foundations of Speech-to-Speech Translation. Artificial Intelligence. Springer, Berlin, Heidelberg.

Dellwo, V. 2006. Rhythm and speech rate: a variation coefficient for deltaC. In Karnowski P. & Szigeti, I. (ed.) Language and language processing. Frankfurt am Main: Peter Lang, 231-241.

Grabe, E., Low, E.L. 2002. Acoustic correlates of rhythm class. In: Gussenhoven, Warner (Eds.), Laboratory Phonology, vol. 7. Berlin: Mouton de Gruyter: 515–546.

He, L., Dellwo, V. 2016. The role of syllable intensity in between-speaker rhythmic variability. International Journal of Speech, Language & the Law, 23(2), 243-273.

Kisler, T., Reichel, U.D., Schiel, F. 2017. Multilingual processing of speech via web services. Computer Speech & Language, 45, 326–347.

Lykartsis, A. 2020. On the analysis of speech rhythm for language and speaker identification. PhD dissertation, Technische Universität Berlin.

Mefiah, A., Alotaibi Y.A., Selouani S.A. 2015. Arabic speaker emotion classification using rhythm metrics and neural networks. In 2015 23rd European Signal Processing Conference (EUSIPCO), 1426–1430. IEEE.

Mousavi, N., Grawunder, S. 2023. Persian speaker classification using rhythmic features. In Draxler, C., editor, Studientexte zur Sprachkommunikation: Elektronische Sprachsig-nalverarbeitung 2023, pages 194–201. TUDpress, Dresden.

Ramus, Fr., Nespor, M., Mehler, J. 1999. Correlates of Linguistic Rhythm in the Speech Signal. Cognition, 73, 265-292.

Tilsen, S., Arvaniti, A. 2013. Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. JASA, 134(1), 628–639.

Priming effect for noun phrase modification in Bengali?

Md-Wasim Odud, Samar Husain
Indian Institute of Technology Delhi, India

https://doi.org/10.36505/ExLing-2024/15/0024/000649

Abstract

This study examines the impact of syntactic complexity on structural priming effects in the context of noun phrase relativisation in Bengali. The result indicates that speakers of SOV languages, when presented with a binary choice of constructions with different levels of syntactic complexity, are predisposed towards the simpler structure, driven by underlying processing strategies, overriding any priming effect during online comprehension tasks.

Keywords: structural priming; syntactic complexity; noun phrase modification; Bengali

Introduction

Despite the extensive documentation of the robustness of structural priming effect (Mahowald, James, Futrell & Gibson, 2016), the role of syntactic complexity in influencing this phenomenon has not been comprehensively investigated in the priming literature. This suggests a need for deeper examination into how the degrees of syntactic complexity affect the strength and likelihood of priming effects.

Research question

The primary research question of this study probes the predictions and the potential limitations of the two dominant theories on structural priming; namely, the 'residual activation theory' (Pickering & Branigan, 1998) and the 'implicit learning hypothesis' (Bock & Griffin, 2000) by focusing on the effectiveness of syntactic priming across simple vs complex/rare sentence structures, particularly, in the context of noun phrase relativisation. These theories have been pivotal in understanding how priming operates, yet their scope in the context of diverse syntactic structures, particularly in less commonly studied languages, remains an area for further exploration. Thus, in alignment with the perspectives put forth by Jaeger & Norcliffe (2009), this study has been conducted in Bengali, an Indo-Aryan, South Asian, SOV language that offers a distinct syntactic typology.

Experiment overview

In order to test the predictions of the 'residual activation theory' (Pickering & Branigan, 1998) and the 'implicit learning hypothesis' (Bock & Griffin, 2000) regarding the impact of syntactic complexity on structural priming effects, we conducted a written sentence completion task (Pickering & Branigan, 1998) on noun phrase relativisation in Bengali which we report below.

Design and materials

The experiment had two conditions. For SRC primes and ORC primes, each item of this experiment contains a block of 3 sentences. The first sentence established the context and described a subject with two different modifiers. These modifiers were simple participial phrases that could be relativised either into an SRC or an ORC. However, the modifiers had been designed in a way such that one of the modifiers could be more easily translated into a subject relative clause (e.g. "The girl wearing a blue frock" to "The girl who is wearing a blue frock") while the other one could be more easily translated into an object relative clause (e.g. "The girl liked by Arijit" to "The girl whom Arijit likes") in Bengali. Following the context sentence, the participant was introduced to the prime sentence; a subject relative clause in the SRC prime condition and an object relative clause in the ORC prime condition. The last sentence then asked the participants to describe the subject. The target fragment (e.g. "That girl…") was designed to force a relative clause description of the subject. Participants were also told beforehand that they could only pick one of the two modifiers to describe the subject. Additionally, in this experiment, the entire block, including the context sentence, the prime sentence, the final question for description and the target fragment was displayed together on the screen to maintain the continuity of the participants' reading experience and to prevent the loss of any contextual details from the sentences. To ensure that the order of presentation had no influence on the priming behavior, the experiment balanced the sequence of the SRC-inducing modifiers and the ORC-inducing modifiers across conditions. This was implemented to rule out any potential biases that the order might impart on the participants' choices.

In all, 18 items were prepared each for the 2 conditions. In addition 72 fillers were interspersed randomly along with the critical items. A total of 2 lists were prepared for this experiment following a latin-square design. Each experiment session started with 5 practice trial. A total of 41 university students participated in this experiment.

To address concerns about the potential complexity of the context sentences, twenty-three participants separately evaluated these sentences on a 7-point Likert scale, with 1 as 'very unnatural' and 7 as 'very natural'. The mean

rating obtained was 4.47 (SD = 0.34), which suggests that participants generally perceived the sentences as moderately favorable.

Response coding procedure

The response data were coded as 1s and 0s for logistic regression analysis. For the SRC prime and ORC prime conditions, responses were coded as '1' if priming was observed, i.e., if the target fragment was completed using the prime structure (i.e. SRC in SRC prime condition and ORC in ORC prime condition) and '0' if priming was not observed, i.e., if the target fragment was not completed using the prime structure.

Results

The statistical analyses has been done using the generalized linear mixed-effects model with logit-link function using the lme4 package (Bateset al., 2015) in R. The result reveals no significant difference between the subject relative clause (SRC) primes in condition (a) and object relative clause (ORC) primes in condition (b) **(z = -0.13**). The result seems to indicate the prime conditions do not have any effect on the participant's default completion choices (63% SRC completions for SRC prime; 64% SRC completions for ORC prime. Similarly, 35% ORC completions for SRC prime; 36% ORC completions for ORC prime).

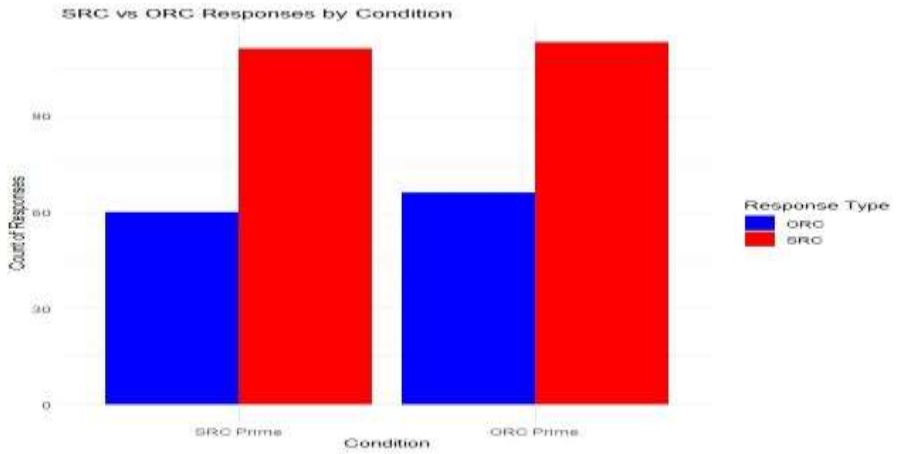


Figure 1: Response Distribution across SRC Prime and ORC Prime Conditions

Discussion

Our interpretation of the result shifts from a priming-centric view to one where processing dynamics come to the forefront. This is especially relevant considering that SRCs are generally easier to process than ORCs, both in

production and comprehension (Hawkins, 1999; Gibson, 1998). Our reinterpretation of the results is further supported by the 'production-distribution-comprehension' (PDC) account proposed by MacDonald (2013), which posits that language comprehension is influenced by the distribution of syntactic structures in production within a language. Additionally, the significance of preverbal material in making robust predictions about upcoming verbal heads in SOV languages (e.g., Konieczny, 2000; Vasishth and Lewis, 2006; Husain et al., 2014) further corroborates our re-interpretation of the results. We believe that, in our experiment, the target fragment acting as this preverbal material, in conjunction with the complexity gradient between SRC and ORC, likely steered participants towards the syntactically simpler option. In light of these observations, we propose the 'Processing over Priming' (PoP) hypothesis which postulates that speakers of SOV languages, when presented with a binary choice of constructions with different levels of syntactic complexity, are predisposed towards the simpler structure, driven by underlying processing strategies and the distributional patterns of their language. Critically, this hypothesis predicts that the propensity for speakers of SOV languages to form simpler structures will override any priming effect during online comprehension tasks, specifically, when the available choice of constructions has been reduced to a binary option between a simple and a complex counterpart.

References

Arai, M. 2012. What can Head-Final Languages Tell us about Syntactic Priming (and vice versa)?. Language and Linguistics Compass, 6(9), 545-559.

Bock, J.K. 1986. Syntactic persistence in language production. Cognitive psychology, 18(3), 355-387.

Cleland, A.A., Pickering, M.J. 2003. The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. Journal of Memory and Language, 49(2), 214-230.

Comrie, B., Keenan, E.L. 1979. Noun phrase accessibility revisited. Language, 649-664.

Analysing bare nouns in Marathi

Shravani Patil
Tilburg University, The Netherlands
Bielefeld University, Germany

https://doi.org/10.36505/ExLing-2024/15/0025/000650

Abstract

This study uses translation mining methodology to analyse bare nouns in Marathi. (In)definite expressions from English were compared to their corresponding referential expressions in Marathi to understand the interpretation of bare nouns. We found that the numeral 'one' behaved as an indefinite article in Marathi, while bare nouns got a definite reading. Bare nouns, in cases of pseudo-incorporation, got an indefinite reading.

Keywords: Bare nouns, indefiniteness, pseudo-incorporation, Marathi, Hindi.

Introduction

Previous theories on bare nouns (BNs) lacked empirical support (Carlson, 1977; Chierchia, 1998; Dayal, 2004). Liu et al. (2023) empirically tested these theories for Hindi, Russian, and Mandarin, which are typologically different languages, and found that these articleless languages do not function similarly. However, it was unclear if the differences between them were due to typology. This study investigated how BNs function in Marathi, a traditionally articleless language, and compares it with Hindi.

Method

We employed the Translation Mining approach proposed by Bremmers et al. (2021) for our corpus study. This approach compares fragments from the source language with the target language(s) and works with the assumption that the form and meaning of expressions are kept constant throughout the text.

We conducted two studies based on two different datasets: (i) the first chapter of Harry Potter and the Philosopher's Stone which was a test corpus to frame a hypothesis, (ii) chapters 2 to 17 of Harry Potter and the Philosopher's Stone which tested the hypothesis we framed. We used the Marathi and Hindi translations of the first chapter of Harry Potter and the Philosopher's Stone as our data set since those were our target languages, and the source language was English. The Hindi data is adopted from Liu et al. (2023), which used the same methodology to study the distribution of BNs in Hindi, Russian, and Mandarin. We manually extracted the referential expressions, 'the N' and 'a(n) N', from

"Harry Potter and the Philosopher's Stone". These referential expressions were compared to their equivalents in Hindi and Marathi.

Results

Quantitative results

'a(n) N' in English (n=88) aligned with the numeral 'one' as well as with BNs. The singular definite, 'the N' in English (n=138) aligned with BNs and demonstratives in Hindi and Marathi, while the singular indefinite 'a(n) N' in English (n=88) aligned with the numeral 'one' and with BNs in the two languages.

Fischer's exact test revealed no significant differences in the distribution of BNs vs. numeral 'one' (singular indefinite contexts) and the distribution of BNs vs. demonstratives (singular indefinite contexts) in Hindi and Marathi ($p>.05$). See figures 1 and 2.

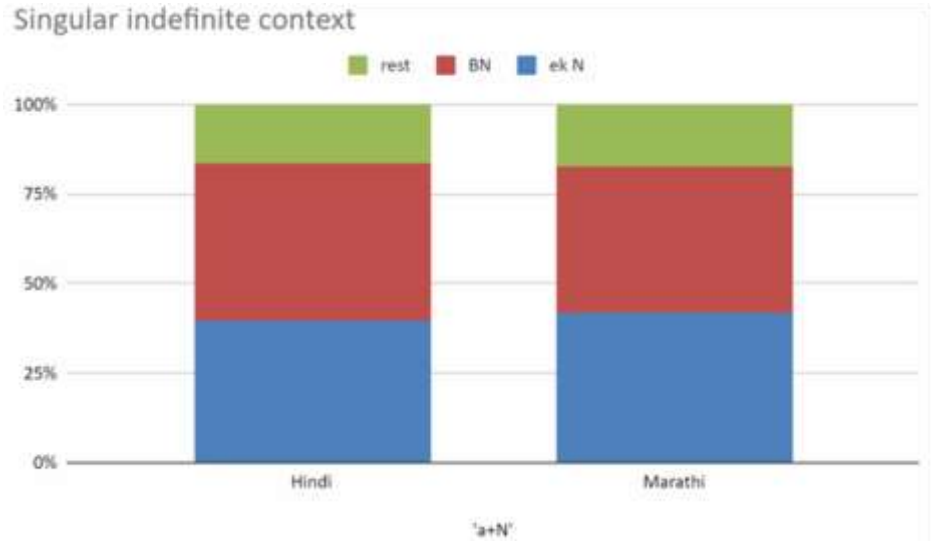


Figure 1: Relative frequencies of BN, numeral `one' + N and rest translations of English indefinite singulars (a(n) + Nsg) in Hindi and Marathi.

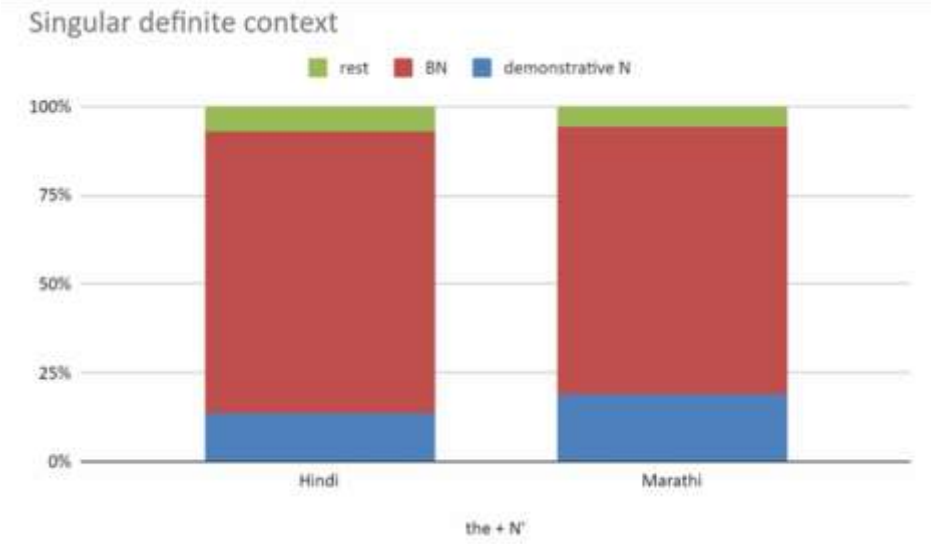


Figure 2: Relative frequencies of BN, demonstrative + N and rest translations of English definite singulars (the + Nsg) in Hindi and Marathi.

Qualitative results

Caseless BNs in object position in Hindi and Marathi have an indefinite reading as they correspond to 'a(n) N' in English. Accusative case-marked nominals have a definite reading as they correspond to 'the N' in English. However, there were also some contexts where caseless BNs received an indefinite reading. These were contexts of Pseudo Incorporation (PI); caseless BNs in object position are PIed in Hindi and Marathi and have an indefinite reading.

Discussion

Example (1) illustrates that the Marathi bare noun 'ɾisiʋəɾ' (receiver) aligns with the English definite NP 'the receiver' and gets a definite reading. However, in example (2), the Marathi bare noun 'nəkaʃa' (map) gets an indefinite reading as it gets incorporated with the verb 'ʋatʃ-əɳ' (reading).

(1) ɻa-ni ɾisiʋər punha kʰali ʈʰeʋ-la.
He-ERG receiver.M again down keep-PFV.M.
'He put the receiver back down.'

(2) ɾəsʈɻa-ʧʃa ʋələn̪-a-ʋər misʈər ɖuɾsli-na pəhili ʧʃəməʈkaɾik
Street-GEN corner-OBL-on mister Dursley-ERG.HON first strange
goʃʈə ɖɻs-li - nəkaʃa ʋatʃ-əʈ əsle-lə mandʒər
thing.F see-PFV.F - Map.M read-PROG be.PROG-N cat.N.
'It was on the corner of the street that he noticed the first sign of something peculiar- a cat reading a map.'

We conclude that, as observed for Hindi (Liu et al., 2023), Marathi also has an indefinite article: Marathi uses 'numeral one' as an indefinite article. BNs in Marathi generally have a definite interpretation unless they occur without an accusative case marking, in which cases they get PIed and have an indefinite interpretation.

This study contributes to the literature on articleless languages by showing how grammatical as well as lexical factors like typology and PI influence (in)definite interpretation of BNs.

Acknowledgements

This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – CRC-1646, project number 512393437, project A3.

References

Bremmers, D., Liu, J., van der Klis, M., Le Bruyn, B. 2021. Translation Mining: Definiteness across Languages—A Reply to Jenks (2018). Linguistic Inquiry. 53. 1-30.

Carlson, G. 1977. A unified analysis of the English bare plural. Linguistics and Philosophy, 1, 413–457.

Chierchia, G.1998. Reference to Kinds across Languages. Natural Language Semantics, 6(4), 339–405.

Dayal, V. 2004. Number Marking and (in)Definiteness in Kind Terms. Linguistics and Philosophy, 27(4), 393–450.

Liu, J., Patil, S., Seres, D., Borik, O., Le Bruyn, B. 2023. 'Articleless' languages are not created equal. Proceedings of Sinn Und Bedeutung, 27, 381–398.

Toward a description of digor lexical prosody

Varvara Petrova
Lomonosov Moscow State University, The Russian Federation

https://doi.org/10.36505/ExLing-2024/15/0026/000651

Abstract

In this paper, I report on the initial results of instrumental research on the lexical prosody of Digor Ossetic (Iranian ‹ Indo-European). Despite speakers of Digor having some intuition regarding the placement of stress, no consensus on its acoustic cues exists in the literature. In this study, I analyzed disyllabic and trisyllabic words with vowels of the same quality within the word form. The results show a tendency for the ultimate vowel of the word to be the longest and for the initial vowel to possess the highest intensity value, albeit trisyllabic words with weak vowels were shown to serve as an exception to the latter trend. This and other relevant acoustic patterns may support the existence of quantitative stress in Digor Ossetic.

Keywords: stress correlates, lexical prosody, vowel length, vowel intensity, Ossetic language.

Introduction

Digor is one of the two main dialects of Ossetic, alongside Iron. Both Iron and Digor have their own standard variety; in this paper, I discuss Standard Digor. The total number of speakers of Ossetic is around 450 000, of which around 100 000 speak Digor.

While the suprasegmentals of Iron Ossetian have already been described in detail, no conclusion on the nature of Digor stress has been reached thus far. M. Isaev [1966] described Digor lexical accent as dynamic but claimed that the intensity of vowels correlated with their phonological length. E. Henderson [1949], on the other hand, believed that “what is felt to be the accent is in most cases closely linked with vowel quantity”. She noted that stress was not recognized by native speakers of Digor and the placement of stress was subject to a high degree of variability.

For my study, stress is defined as increased prominence of one vowel per phonological word. I do not discuss secondary stress. Since vowel quantity and intensity were mentioned by previous researchers as likely correlates of Digor stress and are two of the three most widespread phonetic features to correspond with stress cross-linguistically [Gordon & Roettger 2017], they were considered as potential acoustic cues for lexical stress in Digor. Given that both of these characteristics are affected by vowel quality, the word list for this research consists only of multisyllabic words with vowels of the same quality

within a word form. Two features, length and intensity, were compared separately, with the expectation to see either a) consistent dominance of vowels in the same position, indicating the feature to be a correlate for fixed stress, b) consistent dominance of a certain vowel within a set of utterances of the same word as opposed to dominance of another vowel for another word, implying variable stress, or c) no consistent dominance of one vowel, meaning that stress in Digor, if it exists at all, does not correlate with a given feature.

The speakers reported stress to be not easily distinguishable and to rarely fall on the leftmost syllable. Speakers familiar with Iron Ossetic noted that, unlike the latter, Digor stress can fall on the third and further syllables from the left. One of the speakers suggested that stress can move to the leftmost syllable when the word is under emphatic stress (see Perception of length).

Methodology and data

Data for this study were collected in Vladikavkaz in August 2023. They were obtained from four consultants aged 20 to 67, who were all either university graduates or undergraduate students and spoke Russian fluently. The sound was recorded on Zoom H5 recorder using a WH20XLR microphone. The consultants were asked to translate the words selected from the dictionary from Digor to Russian. After the subset of words that exist in their personal vocabulary has been determined, the consultants were asked to pronounce these words in a natural manner (elicitation type 1) and then repeat it twice with a pause after each utterance. Thereafter the utterances of the same words were elicited in phrases like *ʑɜʁaj […]* “Would you say [word]” (type 2) and *t͡ʃi ʑɜχtaj? – […]* “What did you say? – [word]” (type 3). In total, 267 utterances of disyllabic and trisyllabic were analyzed in this research.

Results

Data presented below consist of values normalized as follows: $(x-x_{mean})/\sigma$, where x_{mean} is a mean value for all of the vowels in all of the utterances of a single lemma by a single speaker. Alpha level of .01 is used in hypothesis tests.

Vowel length

MWU tests were run on each pair of groups presented in Figure 1a. The difference in duration of non-ultimate and ultimate vowels was shown to be statistically significant: $p < 0.001$ for the two vowels of disyllabic words, $p < 0.001$ for the last vowel of trisyllabic words in comparison to the first and the second; $p < 0.001$ for all ultimate vowels compared to all non-final vowels. There is no difference between the ultimate vowels of disyllabic and trisyllabic words in terms of length ($p = 0.225$); differences between the first vowels of disyllabic and trisyllabic words are considered insignificant given that $\alpha = 0.001$

($p = 0.022$). The second vowels of trisyllabic words are both significantly longer than the first ($p = 0.008$) and shorter than the third ones ($p < 0.001$). The penultimate vowels were shown to be significantly shorter in disyllabic than in trisyllabic words ($p < 0.001$).

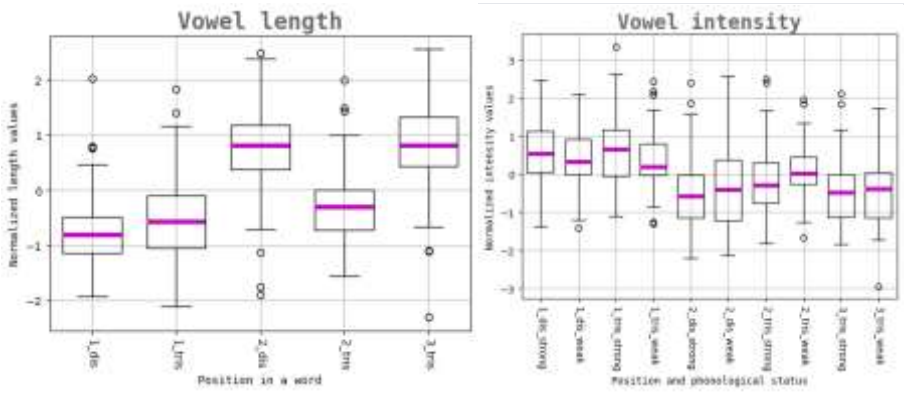


Figure 1 (a-b). Normalized vowel length by position in a word (1, 2, 3), number of syllables in a word (dis for disyllabic, tris for trisyllabic) and, for intensity, vowel type (weak [i, ɜ, u] or strong [a, o, e]).

Vowel intensity

MWU tests have shown intensity to be significantly greater in the first syllable of disyllabic words in comparison to the second regardless of phonological strength of vowels ($p < 0.001$ for both strong and weak vowels; see Figure 1b); for phonological status of vowels see Table 1. However, in trisyllabic forms first vs non-first vowel contrast is only confirmed for words with strong vowels ($p < 0.001$ when compared to the second and the third vowels separately; no substantial difference is found between the groups of the second and the third vowel in terms of intensity: $p = 0.012 > \alpha$). The difference between the intensities of the first and third vowels is also relevant for words with weak vowels ($p < 0.001$), but differences between the first and the second were found to be statistically insignificant ($p = 0.03$), which indicates that there exists at least one type of words with no clearly dominant vowel in terms of intensity. This finding weakens the claim for dynamic stress in Digor.

Possible arguments for quantitative stress

The study did not discover any minimal pairs which would unambiguously indicate a certain type of stress in terms of 1) its main acoustic correlate and 2) contrastiveness or lack thereof. The more indirect findings that could speak in favor of quantitative contrastive stress are presented below.

Words with no quantitative dominance of the final vowel

One lemma (*dʒaman* 'time') was found to counter the tendency for the longest last vowel, which agrees with the speakers' reports of Digor words to rarely be stressed on the first syllable. MWU test run on 18 utterances of *dʒaman* and *36* utterances of *arʁaw* 'tale' and *sajtan* 'devil' (both with the dominant last vowel) showed that the "paroxytonic" group lacks contrast in length between the syllables ($p = 0.021 > \alpha$), unlike the "oxytonic" group (for which $p < 0.001$). The lack of contrast instead of the expected significantly greater length of the first vowel could be explained by the partial compensation, which may occur due to frequent decrease in speech rate toward the end of the utterance.

Perception of length

One speaker reported a stress shift to the first vowel in emphatic contexts (elicitation type 3; see Methodology and data). This introspective assumption was proven wrong within the full dataset. However, an elimination of quantitative vowel contrast in the emphatic context caused by the lengthening of the first vowel ($p = 0.016 > \alpha$; compare to $p < 0.001$ for non-emphatic contexts type 1 and 2) characteristic of this speaker's idiolect could be a proof of an existing connection between vowel length and the perception of stress.

Conclusion

The results show a tendency for Digor vowels of the last syllable to be the longest within the word and vowels of the first syllable to have the greatest intensity, given that all of the vowels in the word represent the same phoneme. Trisyllabic words with weak vowels, however, counter the latter generalization. This fact, along with multiple observations on the rarity of word-initial stress made by our consultants, indicates that the presence of dynamic stress in Digor is unlikely. It could be argued that evidence from perception and word forms with initial vowel lengthening point toward a contrastive quantitative stress. However, given the lack of prototypical minimal pairs or formal perception experiments, we find this hypothesis to be in need of further testing.

References

Boersma, P. 2001. Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341-345.

Gordon, M., Roettger, T. 2017. Acoustic correlates of word stress: A cross-linguistic survey. Linguistics Vanguard. 3. 10.1515/lingvan-2017-0007.

Henderson, E. 1949. A Phonetic Study of Western Ossetic (Digoron). Bulletin of the School of Oriental and African Studies, University of London, vol. 13, no. 1, 36-79.

Isaev, M. 1966. [Digor dialect of Ossetic: phonetics, morphology]. Moscow: Nauka

On the syntactic symmetry of coordinate structures

Adam Przepiórkowski[1,2,3], Katarzyna Kuś[2], Agnieszka Patejuk[1], Berke Şenşekerci[2]

[1]Polish Academy of Sciences, Poland
[2]University of Warsaw, Poland
[3]Massachusetts Institute of Technology, US

https://doi.org/10.36505/ExLing-2024/15/0027/000652

Abstract

This paper examines the syntax of coordinate structures, focusing on the debate between asymmetric and symmetric approaches. Asymmetric theories argue that the first conjunct heads the coordination and determines its distribution. They seem to be supported by examples where predicates like *annoy* and *depend* govern mixed-category coordinations (e.g., NP and CP). Through acceptability judgement experiments, we challenge this argument on the basis of an alternative analysis where such structures involve a coordination of a PP and a CP. Using the Thermometer Method, we find that 2 out of 8 tested predicates (e.g., *annoyed*) accept CP complements as readily as PP complements, while others show reduced acceptability but not outright ungrammaticality. Additionally, the difference in acceptability between purportedly grammatical PP&CP coordinations and ungrammatical CP alone is marginal. These results debunk a key argument for asymmetric theories and provide empirical support for symmetric approaches to coordination.

Keywords: coordinate structures, symmetry of coordination, Thermometer Method, acceptability judgement task

Syntax of coordinate structures

Some of the most prominent approaches to the syntax of coordinate structures in generative linguistics (e.g., Munn 1993 and Zhang 2009) and in corpus linguistics (Universal Dependencies; de Marneffe et al. 2021) are asymmetric: the first conjunct is assumed to head the coordinate structure and determine its distribution. The main argument for such approaches is based on a few examples from Sag et al. (1985), including (1)–(2).

(1) Pat was annoyed by [[NP the children's noise] and [CP that their parents did nothing to stop it]].

(2) You can depend on [[NP my assistant] and [CP that he will be on time]].

In (1), '(annoyed) by' governs a coordination of an NP and a CP, even though it is only compatible with an NP ("…annoyed by the children's noise" is fine), and not with a CP ("…annoyed by that their parents…" is bad). In (2), a similar pattern is observed for 'depend (on)'. This is assumed to provide an argument for asymmetric approaches to coordination: apparently only the category of the first conjunct determines the distribution of the coordinate structure.

Acceptability judgement experiments

The aim of this paper is to debunk the above argument. We report the results of acceptability judgement experiments which show that sentences such as (1)–(2) may have a different syntactic structure than indicated by bracketing in (1)–(2), one that does not support that argument. For example, (1) may be analysed as in (3), i.e., as involving a coordination of a PP and a CP.

(3) Pat was annoyed [[PP by the children's noise] and [CP that their parents did nothing to stop it]].

On this analysis, the purported asymmetry vanishes: the coordinate structure occurs in a position in which both conjuncts are fine alone: both (3') and (3") are acceptable.

(3') Pat was annoyed [PP by the children's noise].
(3") Pat was annoyed [CP that their parents did nothing to stop it].

A series of experiments – to be described in detail in a full-length publication – followed the Thermometer Method (Featherston 2008, 2009), where an open-ended scale is presented with two anchor items visible throughout the experiment: one that is relatively unacceptable (assigned a score of 20) and one that is relatively acceptable (assigned a score of 30). To better assess the absolute acceptability of test items, experiments also included standard items proposed in Gerbrich et al. 2019 as a subset of fillers. The standard items are a set of 15 sentences grouped into 5 classes, A—E, based on their level of acceptability. Each class has 3 members: sentences in class A and B are grammatical, ones in C are marginal, and ones in D and E are ungrammatical. Additionally, we applied a set of exclusion criteria based on reaction time (Juzek 2015: 249–251), comprehension questions, linguistics background, and responses to extremely grammatical and ungrammatical fillers.

In the first experiment, we checked whether predicates such as 'annoyed' and 'depend' take CP complements; only if they do not, does the argument go through. We tested 8 predicates discussed in the relevant literature: 'annoyed (by)', 'ashamed (of)', 'familiar (with)', 'account (for)', 'depend (on)', 'speak (about)', 'suffer (from)', 'talk (about)'.

Out of these 8 predicates, 'annoyed (by)' and 'ashamed (of)' turned out to be as acceptable with CP complements as they are with PP complements (both $p > .05$), so they do not provide any support for asymmetry. The remaining 6 predicates exhibited varying degrees of acceptability with CP complements, and all were statistically significantly less acceptable with CP complements than with PP complements. However, the CP variants of these predicates were statistically significantly higher not only than that of standard items E, but also than that of standard items D. Therefore, it cannot be concluded that these six predicates categorically reject CP complements.

In the second experiment, we compared directly the acceptability of the purportedly fully grammatical PP&CP coordinations ("…depend on… and that…", as in (2) above) with the acceptability of the purportedly ungrammatical CP alone ("…depend that…", as in (4) below). We focused on the 6 predicates that survived the first experiment, excluding 'annoyed (by)' and 'ashamed (of)'.

(4) You can depend [CP that he will be on time].

For 'account (for)', the difference between PP&CP and CP items was not statistically significant, contrary to the prediction of the asymmetric account. For the other 5 predicates, CP items were significantly less acceptable than those involving the PP&CP coordination. However, in all cases, the acceptability difference between PP&CP and CP was relatively small and the average acceptability of CP items was closer to that of grammatical B items or marginal C items, rather than to the ungrammatical D or E items. Therefore, the difference between PP&CP and CP cannot be interpreted as a clear-cut distinction between grammaticality and ungrammaticality.

Conclusion

In summary, our experimental results reveal that CP complements are not categorically ungrammatical for the tested predicates, with some showing relatively high acceptability. Moreover, the observed difference between PP&CP coordination and CP alone does not support a strict binary distinction between grammatical and ungrammatical structures. These findings weaken the empirical foundation of asymmetric theories of coordination, indirectly adding support to symmetric approaches to coordination, recently advocated in Przepiórkowski 2022 and Neeleman et al. 2023.

Acknowledgements

The project has been financed by the OPUS grant 2023/49/B/HS2/02546 "Coordination: Symmetric or Asymmetric?" from the National Science Centre, Poland.

References

de Marneffe, M.-C., Manning, C.D., Nivre, J., Zeman, D. 2021. Universal Dependencies. Computational Linguistics 47(2), 255–308.

Featherston, S. 2008. Thermometer judgements as linguistic evidence. In Riehl, C. M., Rothe, A. (eds.), Was ist linguistische Evidenz?, 69–90. Aachen, Shaker Verlag.

Featherston, S. 2009. A scale for measuring well-formedness: Why syntax needs boiling and freezing points. In Featherston, S., Winkler, S. (eds.), The Fruits of Empirical Linguistics. Volume 1: Process, 47–73. Berlin, Mouton de Gruyter.

Gerbrich, H., Schreier, V., Featherston, S. 2019. Standard items for English judgment studies: Syntax and semantics. In Featherston, S., Hörnig, R., von Wietersheim, S., Winkler, S. (eds.), Experiments in Focus: Information Structure and Semantic Processing, 305–327. Berlin, Mouton de Gruyter.

Juzek, Tom. S. 2015. Acceptability judgement tasks and grammatical theory. Ph.D. dissertation, University of Oxford.

Munn, A.B. 1993. Topics in the Syntax and Semantics of Coordinate Structures. Ph.D. dissertation, University of Maryland.

Neeleman, A., Philip, J., Tanaka, M., van de Koot, H. 2023. Subordination and binary branching. Syntax 26, 41–84.

Przepiórkowski, A. 2022. Coordination of unlike grammatical cases (and unlike categories). Language 98, 592–634.

Sag, I.A., Gazdar, G., Wasow, T., Weisler, S. 1985. Coordination and how to distinguish categories. Natural Language and Linguistic Theory 3(2), 117–171.

Zhang, N.N. 2009. Coordination in Syntax. Cambridge University Press.

Elliptical appendices of relative clauses (EARs)

Tabea Reiner
Bamberg University, Germany

https://doi.org/10.36505/ExLing-2024/15/0028/000653

Abstract

This investigation is about the observation that certain relative clauses in German show peculiar appendices, involving an obligatory subject gap and lacking syntactic integration. The question raised here is whether these strings should be conceived of as an output of mental grammar or rather as systematic mistakes. This second notion would be plausible if a range of examples is poorly accepted in a controlled setting. Therefore, I conducted an experiment, more precisely: a pseudo response time experiment, based on real instances of the structure. In fact, the examples achieved quite high acceptance rates so that they cannot easily be dismissed as mere mistakes. They might, however, still represent grammatical illusions. It will be argued that they are not, leaving an inner-grammatical status of the structure as the only possibility.

Keywords: ellipsis, relativization, acceptability, response time experiment, illusion

Introduction

This paper is about sentences from German edited texts like (1).

1. Nachrichten, die man schaudernd liest
   news that one with.a.shudder reads

   und (*man) dabei dieses unendliche
   and (*one) at.the.same.time this infinite

   Grauendoch nicht begreifen kann.
   horror yet not comprehend can

   (KulturSPIEGEL, no. 2/2011, p. 38)

   'news that one reads with a shudder and yet cannot comprehend this infinite horror'

The pronoun *die* (line 1) introduces a relative clause. This relative clause appears to contain a coordinate structure, as suggested by the presence of the conjunction *und* (line 2). However, the immediately following string involves two peculiarities. First, it does not seem possible to realize the subject, which is odd for a language like German (Seiler 2015:244-245). Second, verb last order, more precisely: finite verb last order (*begreifen kann*), usually implies the presence

of an overt subjunction or relativizer – but there is no subjunction, nor can *die* relate to this string as it would act as a second, unselected direct object.

Thus, we observe an obligatory subject gap as well as unlicenced verb last order. These two properties are the defining characteristics of what are called here ELLIPTICAL APPENDICES OF RELATIVE CLAUSES (EARS). Their peculiarity raises the question whether they should be conceived of as an output of mental grammar or rather as systematic mistakes. The following sections try to answer this question by means of an experimental approach.

Testing EARs: inner- or outer-grammatical?

Conceptual issues around acceptability: It is well known that acceptability alone is neither a necessary nor a sufficient criterion for the inner-grammatical status of a given structure. It is not a necessary criterion because there are garden-path structures (– acceptable, + considered grammatical) and it is not a sufficient criterion because there are grammatical illusions (+ acceptable, – considered grammatical; cf. Haider 2011). However, EARs are too peculiar to qualify as garden-paths. This leaves us with two possibilities. Either EARs prove not to be acceptable in a controlled setting, which means that they can be plausibly dismissed as outer-grammatical – or they prove to be acceptable in a controlled setting, which means that their status has to be further discussed: do they derive from grammar or are they illusions?

Another conceptual issue with respect to acceptability is how to access it experimentally. Especially when working on standard varieties, as is the case in the present study, one is faced with the danger that participants will want to showcase what they have learnt at school rather than follow their intuition. The design presented in the following section tries to overcome this problem.

Design: I conducted a pseudo response time experiment. In more detail, the participants were asked to tell apart real German sentences from mock German sentences as quickly as possible. This means that, form the perspective of the participants, accuracy was presupposed while speed was emphasized. In actual fact, however, I was interested in the judgments as such and used time pressure only to enforce spontaneity. Technically, the task was implemented as a matching task with real measurement of response times at SoSci Survey.

As for the stimuli, I used four structural types:

- EARs from a convenience sample (edited genres throughout)
- Instances of asymmetric coordination (AC, cf. Reich 2009)
- Surface structures that can be analysed as deep coordination below the subject (DC1)
- Surface structures that can be analysed as deep coordination below non-subjects (DC 2)

AC is similar to EARs in two respects. First, it involves the (surface) coordination of unlikes, typically V2 + V1. Second, it involves an obligatory subject gap. However, AC is generally considered grammatical and acceptable in the literature. Likewise, DC1 and DC2 represent structures beyond doubt. Thus, the three structures serve as baselines for the acceptability of EARs.

The four structures were tested in four conditions, cf. Table 1.

Table 1: Conditions

	+ conjunction	– conjunction
+ subject		
– subject		

The presence of the subject was varied in order to double-check whether the subject gap is indeed obligatory. The presence of an overt conjunction was varied in order to explore syndesis as a factor for coordination acceptability.

As to the number of stimuli, EARs were represented by five examples in four conditions, the other structures were represented by one example each, again in four conditions. So, in total, there were 32 target items, mixed with filler and control items and presented in rotated order. 121 persons participated in the experiment, of which 103 native speakers remained after the data had been cleaned.

Results

The accumulated results are shown in Figure 1 – Figure 4.

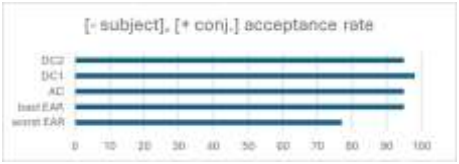


Figure 1: Results for condition [– subject, + conjunction]

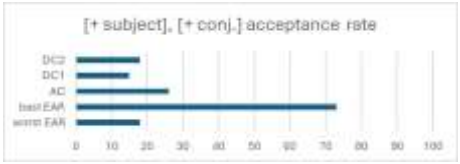


Figure 3: Results for condition [+ subject, + conjunction]

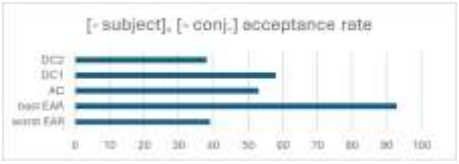


Figure 2: Results for condition [– subject, – conjunction]

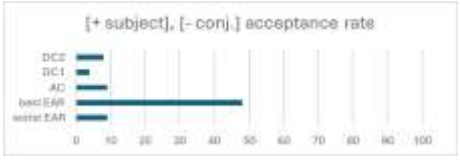


Figure 4: Results for condition [+ subject, – conjunction]

The figures suggest that EARs are not essentially worse than the baseline structures. Even more so, acceptability seems to depend more on the absence of the subject and the presence of a conjunction than on the type of structure.

Discussion

Of the two possibilities outlined on p. , the second one has been confirmed: EARs are sufficiently acceptable to count either as the products of mental grammar *or* as grammatical illusions. One way to decide between these two remaining options is provided by Meinunger's (2023) criterion for illusion status, as I understand it: any phenomenon is an illusion iff capturing it requires fundamental changes in the theory with little gain in empirical adequacy. In particular, an illusion is restricted to a small set of syntactic or semantic contexts.

EARs do not fulfil this criterion. They can be captured by a very general analysis of ellipsis (Reiner in prep.) and even in my small convenience sample they occur in a broad range of contexts: among the relative clauses are restrictive and non-restrictive ones, head-adjacent and dislocated ones; their introductions are simple or involve pied piping; the appendix starts with different conjunctions; the overall readings are symmetric or asymmetric (in a wide sense of the word) and, in addition, I even found the same structure in free relative clauses. Therefore, EARs do not qualify as grammatical illusions.

Conclusion

It was argued that EARs are acceptable without being grammatical illusions. Thus, they must be regarded as products of mental grammar.

References

Haider, H. 2011. Grammatische Illusionen. Lokal wohlgeformt – global deviant. Zeitschrift für Sprachwissenschaft 30(2), 223–257.

Meinunger, A. 2023. Unexpected finite verb forms in German: Cases of grammatical illusion? Linguistische Berichte (Special issue 34), 205–230.

Reich, I. 2009. "Asymmetrische Koordination" im Deutschen (Studien zur Deutschen Grammatik 75). Tübingen, Stauffenburg.

Reiner, T. in prep. An old analysis of ellipsis applied to new observations [working title].

Seiler, G. 2015. Syntactization, analogy and the distinction between proximate and evolutionary causations. In Adli, A., García García, M.,Kaufmann, G. (eds.) 2015, Variation in language: System- and usage-based approaches (Linguae & Litterae 50), 239–264. Berlin, De Gruyter.

Processing flexible argument structure in Dutch, German, and English

Anne Renzel
University of Münster, Germany

https://doi.org/10.36505/ExLing-2024/15/0029/000654

Abstract

Permissive subjects, which are non-agentive subjects combined with action verbs in the active form (e.g., 'A few years ago a penny would buy you two or three pins'; 'The tent sleeps four people'), are hardly found in German compared to English. For this contrast, previous research offers an explanation related to processing constraints, proposing that distinct processing strategies account for varying efficiency of processing permissive subjects. The differences in processing strategies are said to correlate with basic typological properties, specifically basic word order. In order to demonstrate cross-linguistic differences in the processing of permissive subjects, we conducted an experiment where native speakers of English, Dutch, and German performed a self-paced reading task. Based on the results, we argue that fundamental differences in processing strategies are present in English and Dutch on the one hand (look ahead parsing) and German on the other (look back parsing). However, our findings challenge the claim that different processing strategies are directly related to contrasts in the basic position of the verb. It is likely that also other contrasts in basic language properties could influence processing permissive subjects, most likely case.

Keywords: permissive subjects, sentence parsing, West Germanic languages, SVO/SOV word order, self-paced reading

Introduction

Present-day English allows a wide range of semantic roles to function as the subject. Permissive subjects, which are non-agentive subjects combined with action verbs in the active form (e.g., 'A few years ago a penny would buy you two or three pins'; 'The tent sleeps four people'), are hardly found in German compared to English (Rohdenburg 1974, Hawkins 1986, Müller-Gotama 1994, Dreschler 2020, Levshina 2020). For this contrast, previous research offers an explanation related to processing constraints, proposing that distinct processing strategies account for varying efficiency of processing permissive subjects. The differences in processing strategies are said to be linked to basic typological properties, specifically basic word order. If a language has SVO order (like English), permissive subjects can be better processed due to more routinized look ahead parsing strategies. In contrast, if a language is SOV (like German), parsers are more used to look back parsing strategies, leading to difficulties in processing permissive subjects (Hawkins 2014; Engelhardt et al., 2023). The

present study addresses the question if fundamental differences in the processing of permissive subjects can be revealed indicating look ahead or look back parsing, and if parsing strategies depend on typological features like SVO versus SOV word order. Additionally to English and German, we investigate Dutch, as it is clearly SOV but seems to allow slightly more diverse roles in the grammatical subject than German (Müller-Gotama 1994).

Method and materials

To investigate cross-linguistic processing of permissive, a self-paced reading experiment was conducted with native speakers of English (n=40), Dutch (n=45), and German (n=45). The experimental design manipulated AGENTIVITY of the subject phrase, distinguishing between non-agentive conditions (the permissive subjects) and agentive control conditions. Agentive test items contained the same critical verb as non-agentive stimuli but began with an animate noun phrase, which is a plausible agent for the critical verb. In addition, we also manipulated DOCUMENTEDNESS by including both in English documented permissive subjects (Doc) as well as extended ones (Ext) (e.g., 'The house lives four families'). The latter are not accepted in English (Rohdenburg 1974) and considered ungrammatical across all three languages. The inclusion of extended constructions provides an opportunity to investigate whether speakers apply similar processing strategies to both existing and non-existent constructions, independent of the frequencies of the constructions, thereby shedding light on how constraints on permissive subjects are shaped not only by linguistic norms but in particular processing routines.

Results

The RTs (Reading Times) of the verbs in non-agentive sentences and agentive sentences were analyzed using linear mixed-effects models to account for both fixed effects and random effects for participants and items. The model reveals a significant interaction between LANGUAGE and AGENTIVITY (Figure 1). All interactions in the model involving the factor DOCUMENTEDNESS are non-significant, which suggests that documentedness does not substantially modulate RTs in our dataset. The results indicate that German speakers experience difficulties in processing permissive subjects. English speakers, in contrast, demonstrate consistently fast processing across all constructions. Dutch speakers were expected to show comparable difficulties in processing permissive subjects as in German due to basic SOV word order in Dutch similar to that of German (based on Hawkins 2014, Engelhardt et al. 2023). However, RTs in Dutch reveal a faster processing of permissive subjects compared to German, and fall between those of English and German speakers.

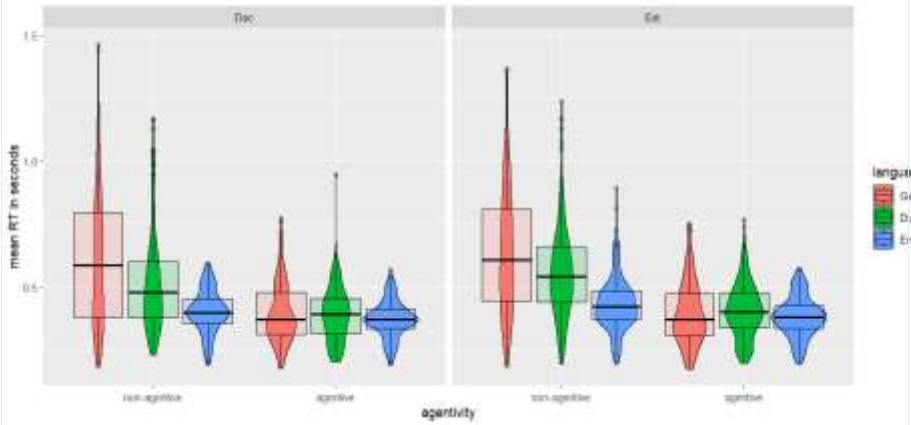


Figure 1. Effects of Language, Agentivity and Documentedness on RTs.

Importantly, this pattern is not limited to permissive subjects documented in English (Doc); our experiment suggests that these contrasts extend even beyond the known constructions. Extended (Ext), non-existent permissive subjects, too, are read faster in English and Dutch than in German. These effects cannot be attributed to usage-based frequency data and the prevalence of these constructions in the three languages. Therefore, our findings go beyond the established understanding that high-frequency constructions are cognitively easier to activate, and therefore easier to produce and process. Since we also observe a tendency among Dutch speakers to process non-documented permissive subjects significantly faster than speakers of German, our results not only indicate fundamental differences between English and German (SVO vs. SOV) in how vague semantics in permissive subjects are processed, but, importantly, also between Dutch and German (both SOV).

Discussion

In this study, we conducted a self-paced reading experiment to investigate differences in the processing of non-agentive subject-verb combinations in permissive subjects in English, Dutch, and German. The aim was to determine whether fundamental differences in processing strategies for this type of loose form-to-meaning mapping can be observed across the three languages, and to what extent these differences are influenced by basic word order properties (SVO vs. SOV). In English, all constructions in our experiment are processed fast, while German speakers show significant difficulty processing loose semantics in subject-verb combinations. In Dutch, too, permissive subjects are processed significantly faster than in German. We argue that speakers of the three languages employ different strategies to process non-agentive subject-verb combinations in permissive subjects. We interpret our findings to suggest

that in particular English speakers, as speakers of SVO languages, routinely employ look ahead strategies, which likely facilitate the processing of permissive subjects (Hawkins 2014; Levshina 2020). In contrast, speakers of SOV languages like German encounter greater challenges in processing permissive subjects, which we interpret as evidence of a reliance on look back strategies to earlier sentence material that must be integrated into the parsing string. According to Hawkins (2014) and Levshina (2020), this process is hindered by the vague semantics inherent to permissive subjects.

The observation that speakers of Dutch as an SOV language like German show faster processing speeds than German indicating look ahead processing mechanisms, too, appears to conflict with the hypothesis that look ahead parsing routines are directly related to SVO word order. In Dutch, other mechanisms than word order appear to facilitate the processing of vague semantics in permissive subjects. We argue that also other contrasts in language systems contribute to facilitating the processing of permissive subjects in Dutch compared to German. Verb-dependent processing triggered by the loss of case (Hawkins 2014:142) may be already activating a certain degree of looking ahead toward later sentence material and the integration of late frame selection, making these strategies more routine in both English and Dutch. Therefore, the absence of case marking in Dutch might explain the differences in processing efficiency of permissive subjects in Dutch and German.

References

Dreschler, G.E. 2020. 'Fifty pounds will buy me a pair of horses for my carriage': the history of permissive subjects in English. English Language and Linguistics, 24(4), 719–744.

Engelhardt, P.E., Filipović, L., Hawkins, J.A. 2024. Prediction in SVO and SOV languages: processing and typological considerations. Linguistics, 62(2), 349–383.

Hawkins, J.A. 1986. A comparative typology of English and German. Unifying the contrasts. London: Croom Helm.

Hawkins, J.A. 2014. Cross-linguistic variation and efficiency (First edition). Oxford linguistics. Oxford UK: Oxford University Press.

Levshina, N. 2020. How tight is your language? A semantic typology based on Mutual Information. In K. Evang, L. Kallmeyer, R. Ehren, S. Petitjean, E. Seyffarth, D. Seddah (Eds.), Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories (pp. 70–78). Stroudsburg, PA, USA: Association for Computational Linguistics.

Müller-Gotama, F. 1994. Grammatical Relations: A Cross-Linguistic Perspective on their Syntax and Semantics. Empirical Approaches to Language Typology. Berlin: De Gruyter.

Rohdenburg, G. 1974. Sekundäre Subjektivierungen im Englischen und Deutschen: Vergleichende Untersuchungen zur Verb- und Adjektivsyntax. Bielefeld: Cornelsen-Velhagen und Klasing.

A data-based sketch of Phonological Grammar for L2 prosody-syntax

Yuji Shuhama
Keiwa College, Japan

https://doi.org/10.36505/ExLing-2024/15/0030/000655

Abstract

This study investigates L2 learners' prosody to outline Phonological Grammar for pedagogical purposes. The prosody data from Japanese learners of English was collected by the following procedures. Fifty-nine learners participated in the production tasks consisted of arranging cards into sentences and reading them aloud appropriately for a given context. Seventeen students in the surveyed learners' group continued three-month lessons and took an interview test for the follow-up fluency assessment. The results indicated a significant correlation of prosodic naturalness with grammatical knowledge, showing that phrasal prosody could improve through the additional learning exposure while discourse/sentence prosody could not. These results lead to Phonological Grammar, a grammar that focuses on prosody tied to phrase structures.

Keywords: prosodic/syntactic phrasing, intonation, information structure, L2 English

Introduction

This study deals with L2 learners' knowledge for prosody-syntax mapping, in other words, how grammar and pronunciation are linked. It is known that structurally ambiguous utterances (e.g., "feel the frog with the feather") are resolved by means of prosodic cues such as pauses and vowel lengthening (Snedeker et al 2008). Similar use of prosody seems to be partly available by L2 learners for ambiguity resolution, but the underlying L2 phonology has not been extensively pursued.

Although prosody-syntax mapping is expected to be harder to acquire than individual sound/grammar knowledge (Sorace 2011), a recent theoretical study suggests that natural pitch patterns can be acquired implicitly by non-native speakers (Archibald and Croteau 2021). This research direction leads to a theoretical hybrid of L2 phonology and grammar (Phonological Grammar in my term), which promotes the learning of natural rhythm and intonation.

Based on the above background, this study addresses two research questions: 1) Can L2 learners make use of prosodic cues to disambiguate structurally complex utterances? 2) How can Phonological Grammar for L2 learners be outlined/visualized? To answer the questions, learners in different proficiency levels are examined because more proficient learners are expected to acquire prosody-syntax mapping better.

Prosodic disambiguation

Methods: A speech production task was conducted with 28 Japanese learners of English with CEFR B2, 31 learners with CEFR A2 as well as 8 native speakers. After a tutorial, they were asked to read a given context on a PC display. Here is an example of a context: "John likes to play with a ball. It's in a basket in his room. Now he tries to throw the ball far. He finds a little gap above the curtain."

Then they arranged word cards (*curtain*, *behind*, *throw*…) into a correct utterance (a method following Yang et al (2022)). The correct order was shown, and the participants were asked to read it aloud using appropriate prosody. The card arrangement and speech were video/audio-recorded to compare the participants' prosody-syntax mapping with native speakers' data.

Results: The result of the production task indicates a significant correlation between proficiency and prosody-syntax mapping. For example, ambiguous "throw the ball in the basket behind the curtain" could be prosodically disambiguated by more than half of the B2 learners, while most of the A2 learners could hardly prosodically disambiguate them.

Figure 1 shows how the above ambiguous utterance was prosodically disambiguated to mean taking out a ball in the basket and then throw it over the curtain. Overall, the native speakers used pauses, vowel lengthening, and pitch changes to effectively signal prosodic boundaries. The B2 learners tended to depend on pauses. The A2 learners' pauses were placed too randomly to signal the boundaries. In addition, the range of pitch by A2 and B2 learners was by far narrower than the native speakers' pitch.

These facts indicate that although using pauses seems relatively easy to acquire, prosodic cues uncommon in the learners' first language (Japanese, in this study) such as pitch changes are difficult to learn and use for disambiguation.

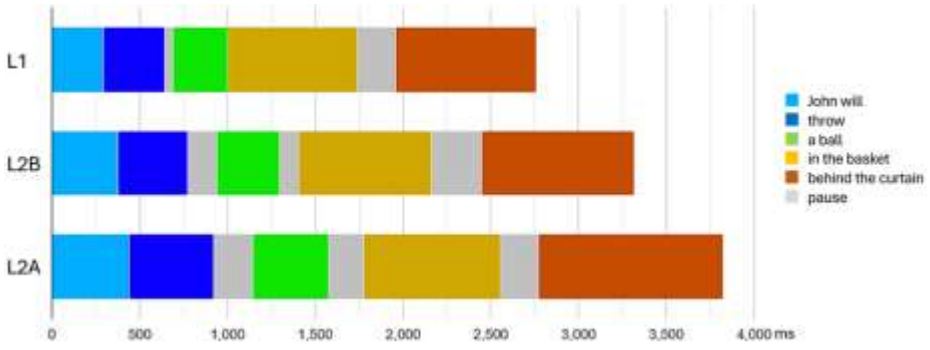


Figure 1. Time course for the target utterance

Pitch control in focus structures

Methods: Out of the 31 A2 learners in Part 2, 17 students joined a weekly English course for three months and then took writing and speaking tests. The speaking test was an interview-style exam based on one of the units in the coursebook. The students were asked to read aloud a passage, which begins with a focus-sensitive sentence, and orally answer the questions about the topic discussed in the passage.

The interview was audio-recorded, and the recorded sound files in MP3 format were analyzed for prosodic features using *Praat* (version 6.1.50). The students' pronunciation was rated on a scale from 1 to 5 according to clarity and fluency, and for a comparison purpose, the data were divided into better speech samples (Group A, n=9) and others (Group B, n=8), respectively.

Results: Figure 2 illustrates the waveform and pitch of the sentence pronounced by a female speaker of American English. The pitch curve shows a sharp rise and fall on a focused word *plants*. More importantly, there is a clear V-shaped rise-fall-rise on *isn't just*. The pitch ranges from 142.8 to 497.5 Hz, which is remarkably wider than the pitch by other two learner's groups.

Figure 3 shows the speech data of a female learner in Group A. There is a tendency in this group that a contrast of *plants* is slightly emphasized within the limited pitch range from 87.5 to 216.6 Hz. The whole sentence was read as fast as the native speaker's sample, which means that the learners in Group A could acquire sufficient fluency to read out in a smooth flow without unnecessary pauses.

To summarize, the comparison of pitch data (B Group omitted) shows that the pitch range of upper A2 learners of English is narrower than a native speaker's and the ability of pitch use varies among them. The learners with relatively high prosodic skills are able to control pitch to express focused information, while the ones with low prosodic skills are unable to control pitch nor read aloud smoothly.

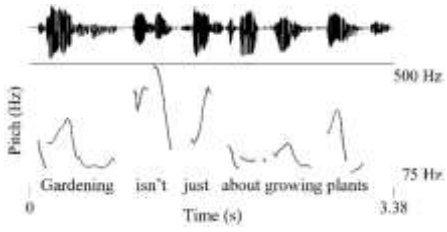


Figure 2. The speech data of a female speaker of American English.

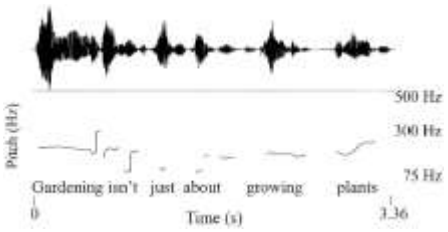


Figure 3. The speech data of a female participant in the A-samples

Discussion and implications

The production task in Part 1 shows that L2 learners can make limited use of prosodic cues such as pauses. The follow-up lessons and interview assessment in Part 2 suggests that pitch is learnable to mark focus structures even though its range is not as wide as native speakers' pitch. If the pitch change in a focused construction "not just A (but) B" is visualized for a publication purpose, I will illustrate it in a simple, easy-to-read fashion as follows:

	is	just	pla	peo
Gardening	n't	about growing	nts, it's about growing	ple

This study has two implications for learning and linguistic theory. A pedagogical implication is that since prosody-syntax mapping is more challenging to acquire than syntax itself, one of keys towards more proficient second language users is to learn how to put phonological chunks to their analyzable structures. Another implication from a theoretical perspective is that the findings in Part 1 supports Grillo and Turco's (2016) proposal by providing empirical evidence of prosodic disambiguation of prepositional phrases at different structural height.

Acknowledgements

This research was supported by JSPS KAKENHI Grant No. 23K00687 "Structuring second language prosody from the syntax-phonology interface perspectives". I am grateful to my students for data collection and to my colleagues for their help as native speaker informants.

References

Archibald, J., Croteau, N. 2021. Acquisition of L2 Japanese WH questions: Evidence of phonological contiguity and non-shallow structures. Second Language Research 37(4): 649-679.

Boersma, P., Weenink, D. 2021. Praat: doing phonetics by computer. An open computer program for speech analysis available at: https://www.fon.hum.uva.nl/praat/

Grillo, N., and Turco, G. 2016. Prosodic disambiguation and attachment height. A conference paper presented in Speech Prosody at Boston University. DOI: 10.21437/SpeechProsody.2016-242

Snedeker, J., Yuan, S. 2008. Effects of prosodic and lexical constraints on parsing in young children (and adults). Journal of Memory and Language 58(2): 574-608.

Sorace, A. 2011. Pinning down the concept 'interface' in bilingualism. Linguistic approaches to bilingualism 1(1), 1-33.

Yang, W., Yiting, G., Ying, F., Ying, S. 2022. Mental representations of time in English monolinguals, Mandarin monolinguals, and Mandarin-English bilinguals. Frontiers in Psychology 13, 1-15. DOI: 10.3389/fpsyg.2022.791197

Longitudinal acoustic analysis of /ɹ/ among Japanese returnee children

Yuting Sun, Tim Laméris
Leiden University, The Netherlands

https://doi.org/10.36505/ExLing-2024/15/0031/000656

Abstract

Japanese returnee children exhibit variations in their pronunciation of the English consonant /ɹ/ over a five-year period. Changes in their language environment are likely to influence early language acquisition. This study involves eight Japanese returnee children who were asked to describe the picture book "Frog, Where Are You?" The research compares the third fundamental frequency (F3) of the consonant /ɹ/ at three different times: immediately after the return, one year later, and five years later. The findings indicate that there is no positive correlation between changes in F3 and the duration of time since their return.

Keywords: Japanese returnees, acoustic analysis, longitudinal study

Introduction

Returnees are speakers who move (back) to their first language (L1) environment after having lived abroad for a certain period in an L2 majority language (Flore, 2010). This return migration typically constitutes a drastic reduction in exposure to the L2 and may lead to L2 attrition, as dominance in the second language diminishes in favor of the first language (Flores & Snape, 2021).

Native Japanese speakers often struggle to distinguish between the /ɹ/ and /l/ sounds in English. According to the Perception Assimilation Model (PAM) proposed by Flege in 1995, this difficulty can be attributed to the fact that both /ɹ/ and /l/ are assimilated into the Japanese /ɾ/ category, as their F3 frequency overlaps with those of the English /ɹ/ and /l/. Previous studies have reported a positive correlation between F3 frequency and /ɹ-l/ phoneme contrast. (Aoyama et al., 2004; Flege et al., 1995; Iverson et al., 2005; Shinohara & Iverson, 2021). In 1995, Fledge found that native Japanese speakers could not produce the distinction between /ɹ/ and /l/ as clearly as native English speakers, who had an average F3 frequency of 1750 Hz for /ɹ/ and 2854 Hz for /l/. In contrast, Japanese speakers did not show clearer F3 distinction, with an average of 2261 Hz for /ɹ/ and 2944 Hz for /l/. Furthermore, some studies explored the factors leading to the loss of contrast. Aoyama (2004) emphasized the age effect, discovering that children typically perform better than adults in

distinguishing the /ɹ-l/ contrast. In addition to age, prior research has also identified the influence of language environment. Kuhi et al. (2016) suggested that prolonged exposure to the Japanese language environment diminishes sensitivity to the critical F3 distinction for English /ɹ/ and /l/. More recently, Shinohara et al. (2021) employed a training method that demonstrated a positive effect, helping Japanese speakers improve their F3 distinction between /ɹ/ and /l/. Looking back on previous research it remains unclear how heritage children and returnees experience changes in their F3 distinction of the /ɹ-l/ contrast. This leads us to our research question:

How does F3 frequency change in the production of the English /ɹ/ in the *frog* among Japanese-English bilingual returnee children over five years? We hypothesize that there would be a correlation between F3 frequency and duration of return: the longer returnees have been back, the more their production of /ɹ/ may deviate from the standard F3 frequency, because of the L2 attrition. F3 values of /ɹ/ for the returnees would increase over time as they moved back to the Japanese environment and gradually lost phoneme contrast compared with the native speakers.

Methodology

For this study, we analyzed audio recordings from eight returnee children reported by Laméris et al. (2024). The audio recordings for these eight participants, measured at three-time points, consistently included the word "frog" across all recordings. The speech materials for analysis comprised 24 tokens (8 returnees × 3-time points). All their recordings were automatically saved as .mp3 files. We imported the relevant audio files and TextGrid files into Praat (Boersma and Weenink, 2019), and then extracted the F3 frequency of/ɹ/ in word *frog*.

All analyses were performed in R 4.2.2 (R Core Team, 2022). The independent variable in this study was the duration since the participants' return to Japan, categorized into 3-time points: immediately after return, one year after return, and five years after return. The dependent variable was changes in F3 frequency. A Pearson's correlation coefficient was computed to assess the relationship between time of return and F3 frequency across participants. The analysis revealed a Pearson correlation coefficient (r) of 0.035, indicating a very weak positive correlation. This correlation was not statistically significant, with a p-value of 0.868. The 95% confidence interval for the correlation coefficient ranged from-0.373 to 0.433. Figure 1 displays the individual changes in F3 values over time

Conclusion

The overall analysis showed no significant correlation between the duration of return to Japan and the F3 frequency of the /ɹ/ sound among Japanese returnee children. In future studies, we could increase the sample size and extend the duration of the experiment.

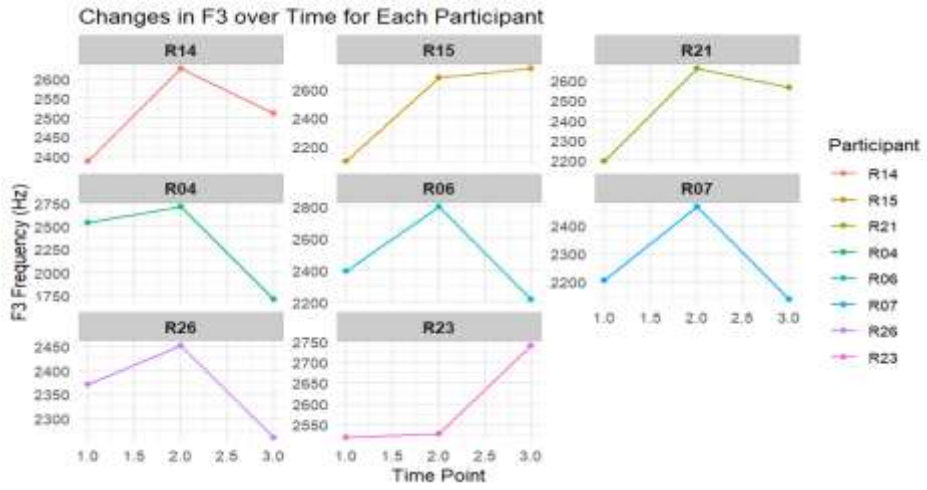


Figure 1. Changes in F3 Over Time Each Participant.

References

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /ɾ/ and English /l/ and /ɹ/. *Journal of Phonetics*, *32*(2), 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Boersma P and Weenink D (2019) Praat: doing phonetics by computer: 6.0.48 [computer program]. Available at: http://www.praat.org.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese Adults can Learn to Produce English /I/ and /l/ Accurately. Language and Speech, 38(1), 25–55. https://doi.org/10.1177/002383099503800102

Flores, C. (2010). The effect of age on language attrition: Evidence from bilingual returnees. Bilingualism: Language and Cognition, 13, 533–546. https://doi.org/10.1017/S136672890999054X

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic Training with Acoustic Cue Manipulation: A Comparison of Methods for Teaching English /ɹ/-/l/ to Japanese Adults. The Journal of the Acoustical Society of America, 118, 3267–3278. https://doi.org/10.1121/1.2062307

Laméris, T. J., Kubota, M., Kupisch, T., Cabrelli, J., Snape, N., & Rothman, J. (2024). Language change in Japanese–English bilingual returnee children over the course of five years: Evidence from accent-rating. Second Language Research, 02676583241230854. https://doi.org/10.1177/02676583241230854

Shinohara, Y., & Iverson, P. (2021). The effect of age on English /r/-/l/ perceptual training outcomes for Japanese speakers. Journal of Phonetics, 89, 101108. https://doi.org/10.1016/j.wocn.2021.101108

The acquisition of Hungarian demonstratives

Enikő Tóth, Péter Csatár, György Rákosi
University of Debrecen, Hungary

https://doi.org/10.36505/ExLing-2024/15/0032/000657

Abstract

This paper presents the results of a small corpus query and a production study exploring demonstrative use in child Hungarian. 13 children between the ages of 3;6 and 4;6 and 11 adults participated in a jigsaw puzzle completion task. We found a significant association between the relative distance of the referent from the speaker and demonstrative choice in both the target group and the adult control group. There is, nevertheless, a strong overall difference between the two groups: children use more distals than proximals, unlike adult controls, who reverse this pattern. We argue that Hungarian children may overuse the distal demonstrative because it serves as the default option in non-contrastive contexts, and is the unmarked form with grammaticalised, non-spatial functions.

Keywords: demonstratives, child language, exophoric, Hungarian.

Introduction

Hungarian has a two-term demonstrative system; *ez, ezek* 'this, these' are nominal proximal demonstratives, *itt* 'here' is an adverbial demonstrative, while *az, azok* 'that, those', and *ott* 'there' are distal. Nominal terms can be used as pronouns or as adnominal demonstratives, as illustrated by (1a) and (1b).

(1)	(a)	Ez	(a	kulcs)	nyitja	az	ajtó-t.
		this.NOM	the	key.NOM	open.3SG	the	door-ACC
		'This (key) opens the door.'					
	(b)	Ez-t	(a	kulcs-ot)	kérem.		
		this-ACC	the	key-ACC	want.1SG		
		'I want this (key).'					

Descriptive grammars of Hungarian assume that in adult language use, the choice of spatial demonstratives is primarily determined by the relative distance of the referent from the speaker (along with interactional factors, such as joint attention, contrastivity). Given that acquisition of demonstratives in Hungarian is an underexplored area, our study aims at assessing the basic distributional trends of demonstrative use in child language and at investigating the role of the distance factor in particular. We first report preliminary findings from a corpus study, and then describe the findings of a controlled elicitation task.

Corpus findings

We conducted a corpus analysis in the MONYEK corpus (Morphologically Disambiguated Corpus of Spoken Child Hungarian) (Mátyus and Orosz 2014), which contains spoken language data, specifically transcripts of interviews with children aged 4;5–5;6, where the children were asked to tell or recount stories based on visual prompts. Altogether, we examined 10,000 utterances (approximately 39,000 words), analyzing nominal demonstratives, specifically their nominative and superessive forms, and the two demonstrative adverbs.

As Table 1 shows, the relative frequency of proximal and distal terms is different in the two populations: children use fewer proximal demonstratives and more distal ones. In contrast, adult participants show the reverse pattern, characterized by a higher frequency of proximal terms. The observed difference is quite pronounced.

Table 1. Relative frequency of demonstrative use in the subcorpus.

	Child-directed speech	**Children's speech**
Proximal terms	357	156
Distal terms	220	318

Our results contradict the observation made by Diessel and Monakhov (2023: 936) based on data from the CHILDES database on a group of languages that does not include Hungarian. Diessel and Monakhov (2023) point out that the proportions of children's spatial terms correspond closely to those in child-directed speech. However, it is important to note that Diessel and Monakhov do not discuss the same age group, their results describe the language use of children aged 1;0-2;1 years. As previously noted, our corpus data include all occurrences and functions of demonstratives. To investigate whether a similar distributional pattern would emerge in a controlled experimental setting, we conducted a production study.

Controlled elicitation task

We adapted the experimental method described by Shin and Morford (2020) and conducted a puzzle completion task with children and adults as a control group. In a scripted conversational setting, the experimenter asked *which piece* questions: *Melyik darabon vannak a mozdony kerekei? 'Which piece has the wheels of the engine?'* to elicit spontaneous use of demonstratives in table-top space. The task was to put the puzzle together. Each session was video-recorded and concluded upon the completion of the puzzle. One session lasted approximately 4 minutes.

Participant and instructor were sitting face to face across a table. The participants were not allowed to touch the puzzle pieces, which were arranged

in three different regions: within easy arm's reach, at forced reach, and beyond reach. There were 13 child participants, 7 female and 6 male, their average age was 3 years and 9 months. We selected this age group based on findings from cross-linguistic research (Küntay and Özyürek 2006; Shin and Morford 2020), which indicate that while four-year-old children are capable of using a variety of demonstratives, their patterns of use differ significantly from those of adults. The children in our study were all monolingual Hungarian speakers.

The adult group included 11 participants, all students from the University of Debrecen (Hungary), with an average age of 23 years. Demonstratives were analysed according to two factors: (i) demonstrative type: proximal vs. distal, and (ii) placement of the puzzle piece (within reach, forced reach, or beyond reach). The children produced 72 tokens, while the adults produced 47 tokens. Detailed results are presented in Table 2.

Table 2. Relative frequency of demonstrative use in the production task.

	Children		**Adults**	
	Proximal	**Distal**	**Proximal**	**Distal**
within reach	14	8	16	0
forced reach	11	14	10	6
beyond reach	4	21	2	13
frequency (%)	40	60	60	40

Children generally produced either pronominal (*ez/az*) or adverbial (*itt/ott*) demonstratives, typically accompanied by a pointing gesture, to respond to the *which piece* questions. A significant association was observed between the referent's relative distance from the speaker and the demonstrative patterns in both the target group ($\chi^2(2) = 11.26$, $p<.01$) and the adult control group ($\chi^2(2) = 24.23$, $p<.01$).

A comparison of the choice of demonstratives between children and adults revealed a more consistent pattern among adults. Adults predominantly used proximal demonstratives when referring to puzzle pieces within easy arm's reach, while distal forms were favoured for objects located beyond reach. Although children also leaned toward distal terms with beyond-reach objects, their choices were less predictable when referring to items within easy arm's reach. Interestingly, the forced reach zone displayed the opposite trend, showing the greatest variability in demonstrative use across both groups.

In general, the children used proportionally more distals than proximals (as opposed to the adult control group). This confirms the findings of our preliminary corpus queries, since the children's utterances in the MONYEK corpus also exhibited a higher proportion of distal terms.

From a cross-linguistic perspective, Diessel and Monakhov (2023) also remark that English children aged 1;0-2;1 make extensive use of distal

demonstratives, and this is strengthened by the findings of González-Peña et al. (2020: 11), who also observe that in English, children predominantly use *that/there*, i.e., distal terms were present in a higher number of transcripts within the CHILDES database in the case of children aged 18-24 months. Unfortunately, we have not found any data on the same age group.

Regarding the results mentioned on English, Diessel and Monakhov (2023) assume that a possible explanation for the overuse of distals in English is that *that* is the default in non-contrastive situations in adult language use (Levinson 2018). Accordingly, we assume that Hungarian children may overuse the distal demonstrative as it is the default choice in non-contrastive situations, and it is also the unmarked demonstrative that has adopted grammaticalised, non-spatial functions. The distal form *az* 'that' serves as the diachronic source of the definite article *a/az* 'the' in Hungarian, and it also functions as the matrix pronominal associate of certain finite subordinate clauses and left-dislocated noun phrases. The distal default slowly fades out as children grow, but it is still present in the age group we investigated.

Acknowledgements

This research has been supported by the NKFIH K22_143417 research grant.

References

Diessel, H., Monakhov, S. 2023. Acquisition of demonstratives in cross-linguistic perspective. Journal of Child Language 50, 922-953.

González-Peña, P., Coventry, K.R., Guijarro-Fuentes, P. 2020. Acquisition of demonstratives in English and Spanish. Frontiers in Psychology 11: 1778.

Levinson, S.C. 2018. Introduction: Demonstratives: Patterns in Diversity. In Levinson, S. C., Cutfield, S., Dunn, M. J., Enfield, N. J., Meira, S. (eds.) 2018, Demonstratives in Cross-Linguistic Perspective, 1-42. Cambridge University Press.

Küntay, A.C., Özyürek, A. 2006. Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? Journal of Child Language 33: 303-320.

Mátyus, K., Orosz, Gy. 2014. MONYEK - Morfológiailag egyértelműsített óvodai beszélt nyelvi korpusz. [MONYEK - A morphologically disambiguated corpus of spoken child Hungarian.] Beszédkutatás 22, 237-245.

Shin, N.L., Morford, J.P. 2020. Demonstratives in Spanish: Children's developing conceptualization of interactive space. In Colomina-Almiñana, J. J., Sessarego, S. (eds.) 2020, Language Patterns in Spanish and Beyond, 285-301. New York, Routledge.

Form and meaning in L1 and L2 lexical selection

Andromachi Tsoukala[1], Margreet Vogelzang[1,2], Ianthi Tsimpli[1]
[1]University of Cambridge, UK
[2]Newcastle University, UK

https://doi.org/10.36505/ExLing-2024/15/0033/000658

Abstract

The present study examines how competing phonological, pragmatic and morphosyntactic cues influence L1 and L2 lexical selection. We used a novel text completion task which involved rhyme priming; the aim was to activate a target rhyming completion which, if produced, would violate either pragmatic or morphosyntactic constraints. In a pilot study, we tested 41 L1 English speakers and 45 L1 Greek L2 English speakers. Results showed that morphosyntactic constraints pre-empted erroneous rhyming completions, whereas pragmatic constraints did not, especially in L1 speakers. L2 speakers were not as susceptible to errors caused by phonological interference, consistent with previous literature suggesting reduced reliance on form and preferential reliance on meaning in the L2.

Keywords: L1-L2 lexical selection, phonological interference, rhyme, pragmatics

Introduction

The question of how we choose our words when we speak (a process known as lexical selection) has been debated in the language production literature, and different theoretical accounts have been proposed (see Kerr et al., 2023 for a recent review). For instance, serial models (e.g., Levelt et al., 1991; 1999) view lexical selection as a top-down, stage-based process. Firstly, semantic-conceptual and syntactic considerations in a given context lead to the activation of a lemma in the mental lexicon. This is followed by form encoding, which involves morphophonological specification, then phonetic encoding and, ultimately, articulation. Hence, the serial view holds that lemma selection is constrained by semantic and syntactic factors, and it is only after this selection has occurred that phonological information becomes available. In the words of Levelt et al. (1991), "a (lexical) item's semantic-syntactic makeup is accessed and used before its phonological makeup becomes available" (p. 122).

By contrast, interactive models (e.g., Dell, 1986; Dell & O'Seaghdha, 1991) argue that the flow of activation can be bidirectional, such that information at lower levels (e.g., activated phonological representations) can feed back up to affect lemma selection. Supportive evidence for the interactive view has come from analyses of speech errors which have shown mixed-error effects, i.e., that lexical errors share both semantic and phonological properties with target

words (e.g., saying "rat" instead of "cat") (e.g, Dell et al., 1997). Similarly, in a sentence completion study by Rapp & Samuel (2002), lexical selection was influenced by phonological features of preceding words in context. Participants were more likely to produce rhyming completions when rhyme prime words were present in the sentence fragments as opposed to when they were absent.

In the present study, we empirically test these models. Specifically, we explored whether a phonological competitor cue (rhyme) can override semantic and syntactic constraints and bias lexical selection in a novel text completion task. We tested both L1 and L2 speakers to examine if competing cues are weighted differently, leading to divergent production outcomes. Given that L2ers have been argued to show reduced sensitivity to form and overreliance on meaning when processing the L2 (e.g., Clahsen & Felser, 2006; Talamas et al., 1999), we hypothesised that a formal phonological cue would influence lexical selection to a lesser extent in L2 than in L1 speakers.

Methods

Forty-one L1 English university students based in the UK (52% female; Mean age = 21.4; SD = 2.14), and forty-five L1 Greek L2 English university students based in Greece (80% female; Mean age = 22.4; SD = 2.57) participated in this pilot study. The L2 participants' knowledge of English was at an intermediate or higher level, as was established through the British Council's English Level Placement Test (Mean score = 21.2; SD = 1.71, out of 25 max).

During the study, participants silently read two types of texts (N = 7 per type), as shown in Table 1. They produced a completion to fill in a blank space at the end of the texts. Both text types contained end rhyme to prime a target rhyming completion which would violate morphosyntactic (Text Type A) or pragmatic, world knowledge constraints (Text Type B). Participants produced a completion with no time limit, and responses were audio recorded for analysis.

Table 1. Examples of the two text types.

Text types	
A (morphosyntactic violation)	**B** (pragmatic violation)
Having pests at home is never nice. If you do, dispose of cheese and rice. That's of course if you have seen a ...	I didn't look above my head. I tried to cross the street instead. Thank god my mother thought ahead In fact, without her I'd be dead. So "Check the traffic lights", she said, "And wait until the man turns ..."
Completion targets	
Incorrect (rhyming) target: mice Correct (nonrhyming) target: mouse	Incorrect (rhyming) target: red Correct (nonrhyming) target: green

Responses were grouped into different categories (e.g., one-word or multiword, rhyming or nonrhyming, related or irrelevant completions). Out of all the completions produced, more than 50% contained one of the targets in both the L1 and L2 group, suggesting that the targets were quite predictable in the textual contexts we designed. For analysis, we focus only on one-word completions consisting of one of the target words (i.e., the correct and nonrhyming target or the incorrect and rhyming target), which represent 57% of the collected data. Data were analysed in R through generalised linear mixed effects models using the formula: glmer(Completion ~ Group*Text_Type + (Text_Type | participant) + (Text_Type | item).

Results

Results are plotted in Figure 1. Analyses revealed an effect of Text Type (beta = -4.19; p = 0.007; OR = 0.02); the odds of producing the correct (nonrhyming) target were lower when the texts involved a violation of pragmatics compared to morphosyntax. The effect of Group was significant (beta = 1.19; p = 0.002; OR = 3.27); L2 participants produced the correct target to a greater extent than L1 participants. The Group*Text Type interaction was nonsignificant, although the L1 participants produced numerically more incorrect target completions in the pragmatic violation texts (68.07%), whereas L2ers produced the correct target more often (55.25%).

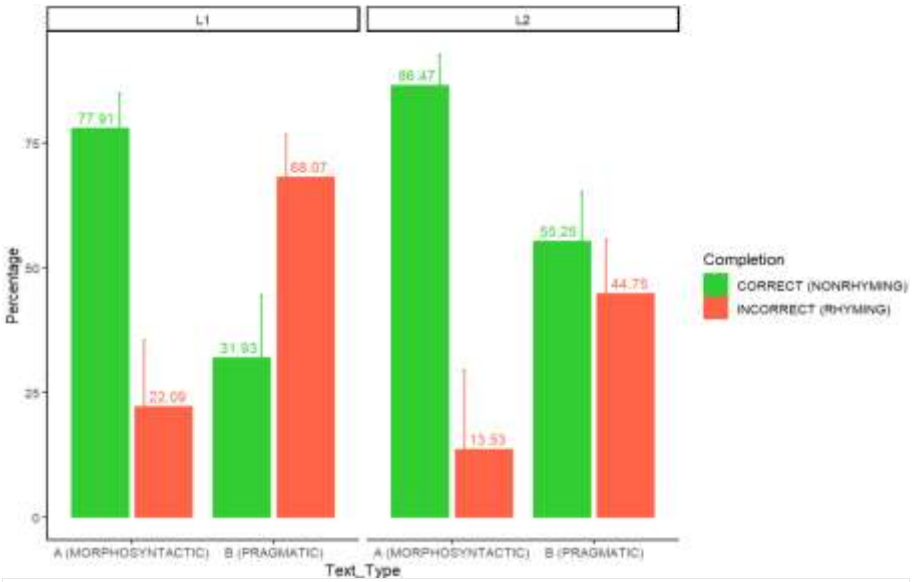


Figure 1. Mean % of correct and incorrect completions by Group and Text Type.

Discussion

Results partly support serial models, since phonological cues did not affect lexical selection in all contexts. Participants produced morphosyntactically well-formed completions more often than infelicitous rhyming responses, indicating that grammatical knowledge is more entrenched and thus accessible to block inappropriate phonological forms. Yet, we also found some support for interactive activation models, since participants, and particularly L1 rather than L2 speakers, were more likely to be influenced by rhyme and produce pragmatically infelicitous responses. These results align with previous work showing that phonological cues can affect lexical selection (e.g., Rapp & Samuel, 2002); they also extend it by providing preliminary, yet novel evidence that phonological cues can override top-down pragmatic constraints. Moreover, results indicated that, compared to L1 speakers, L2ers were overall less susceptible to errors caused by phonological interference, which may be due to unequal weighting or reduced sensitivity to formal cues, consistent with previous literature (e.g., Clahsen & Felser, 2006; Talamas et al., 1999). Finally, the development of this novel task may prove useful for future studies studying phonological influences on lexical selection in other contexts and populations.

References

Clahsen, H., Felser, C. 2006. Continuity and shallow structures in language processing. Applied Psycholinguistics 27(1), 107-126.

Dell, G.S. 1986. A spreading-activation model of retrieval in sentence production. Psychological Review 93, 283–321.

Dell, G.S., O'Seaghdha, P.G. 1991. Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). Psychological Review 98, 604 – 614.

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., Gagnon, D.A. 1997. Lexical access in aphasic and nonaphasic speakers. Psychological review, 104(4), 801.

Kerr, E., Ivanova, B., Strijkers, K. 2023. Lexical access in speech production: Psycho- and neurolinguistic perspectives on the spatiotemporal dynamics. In Hartsuiker, R. J., Strijkers, K. (eds.) 2023, Language production 32–65. Abingdon, UK: Routledge.

Levelt, W.J., Roelofs, A., Meyer, A.S. 1999. A theory of lexical access in speech production. The Behavioral and Brain Sciences 22(1), 1–75.

Levelt, W.J.M., Schriefers, H., Vorberg, D., Meyer, A.S., Pechmann, T., Havinga, J. 1991. The time course of lexical access in speech production: A study of picture naming. Psychological Review 98(1), 122–142.

Rapp, D.N., Samuel, A.G. 2002. A reason to rhyme: phonological and semantic influences on lexical access. Journal of Experimental Psychology: Learning, Memory, and Cognition 28(3), 564.

Talamas, A., Kroll, J.F., Dufour, R. 1999. From form to meaning: Stages in the acquisition of second-language vocabulary. Bilingualism: Language and Cognition 2(1), 45–58.

Disambiguation of French motion/location expressions

Carolin Ulmer
Freie Universität Berlin, Germany

https://doi.org/10.36505/ExLing-2024/15/0034/000659

Abstract

The present study investigates reading preferences for French sentences containing a Manner of motion verb and a prepositional phrase which are ambiguous between a location and a locomotion reading. We tested how different prepositions and different Manner of motion verbs influence reading preferences. 61 native French speakers judged different ambiguous sentences by selecting one of two suggested contexts, each clearly suggesting one of the two readings. Results show that overall, locomotion readings are possible with all tested prepositions, but more likely with *à travers* 'through' than with ambiguous prepositions. Manner of motion verbs fall into two classes, one allowing the locomotion reading and the other one not. We provide experimental evidence confirming patterns observed in previous corpus studies.

Keywords: motion events, boundary-crossing, disambiguation task

Introduction

French (like other Romance languages) has sentences with a Manner of motion verb and a prepositional phrase that are ambiguous between a location and a locomotion reading, see (1).

1. *Il court dans le jardin.* (a) 'He runs in the garden'. (b) 'He runs into the garden'.

Some authors have claimed that such sentences can only have reading 1a (Slobin, Hoiting 1994) whereas other authors have described that they are in fact ambiguous with the two readings in 1a and 1b (Aurnague 2008, Kopecka 2009, Sarda 2019). Whether this structure can have only the unbounded interpretation, as in 1a, or whether a bounded reading, as in 1b, is indeed possible has implications for the question whether there is a constraint on the expression of boundary-crossing events in certain languages (Slobin, Hoiting 1994, i.a.). So far, not much research has been conducted on the interpretation of these sentences. Elicitation studies on motion events report that native speakers of French utter sentences similar to (1) to describe bounded motion events (Pourcel, Kopecka 2005; Hendricks, Hickmann 2011). A corpus study by Kopecka (2009) found that both readings are possible for most of the verbs under investigation, however with a different distribution for all verbs.

Methods

Materials, participants and procedure

This study investigated the reading preferences of ambiguous motion/location sentences in an experimental setting. Stimuli consisted of sentences containing:

- one of six Manner of motion verbs: *courir* 'to run', *sauter* 'to jump', *ramper* 'to crawl' (which Fortis 2010 claims to be possible with both readings) and *nager* 'to swim', *marcher* 'to walk', *danser* 'to danse' (which Fortis 2010 claims to be possible with a location reading only)
- one of three prepositions: *dans* 'in(to)', *en dehors* 'outside/out of' (both ambiguous) and *à travers* 'through' (with a clear locomotion meaning)

Both ambiguous prepositions, in their locomotion reading, express not only the direction of the locomotion, but also the crossing of a boundary. Every sentence was presented with two contexts, each clearly suggesting one of the two readings, see Fig. 1. The order of the two contexts was counter-balanced. Participants were asked to decide which context the sentence fits in best.

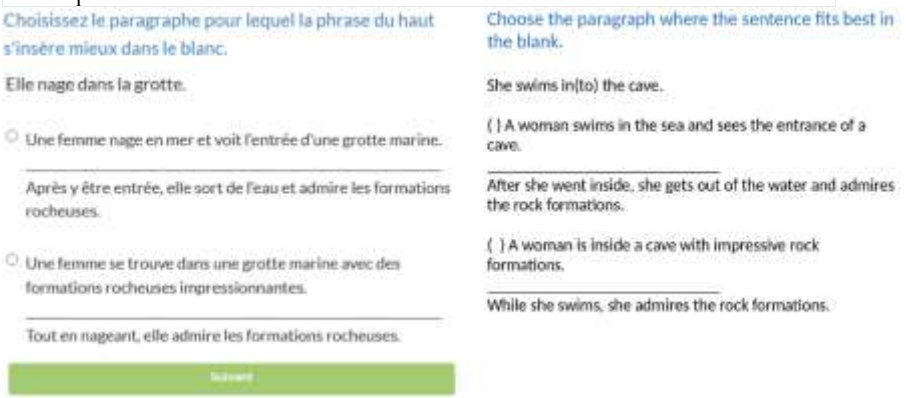


Figure 1. Presentation of French test items, here with English translations.

The experiment was carried out with Labvanced. It was run with 61 monolingual speakers of French recruited through Prolific in August 2023. Each participant judged three test sentences along with nine filler sentences. Ten (to eleven) judgements were elicited for each test sentence.

Hypotheses

1. The ambiguous prepositions (*dans* 'in/into', *en dehors de* 'outside/out of') were expected to behave differently to the directional preposition (*à travers* 'through'): Less boundary-crossing readings are expected with the ambiguous prepositions.
2. Manner verbs were expected to differ regarding the extent to which they allow the boundary-crossing reading: They fall into two classes, one allowing the BC reading, the other one not (cf Fortis 2010).

Results

The results both for different prepositions and for different Manner of motions verbs are shown below in Fig. 2 and 3, respectively.

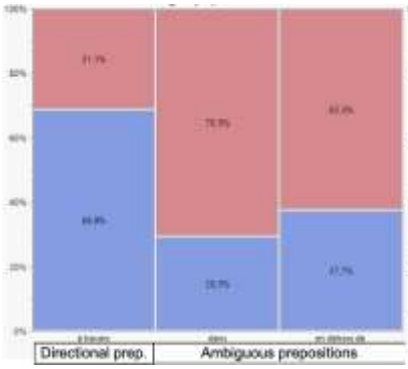


Figure 2. BC readings with different prepositions

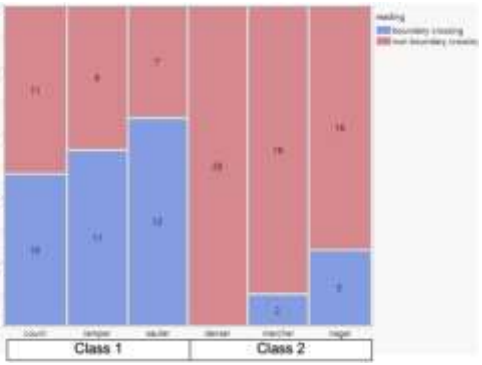


Fig. 3: BC readings with different Manner verbs.

Statistical analyses were carried out with R (R Core Team 2021). Results were examined with two mixed-effects logistic regression models using the glmer function contained in the package lme4 (Bates et al. 2015). The first model examined the likelihood of the interpretation as a boundary-crossing event depending on the type of preposition and the verb class. The directional preposition *à travers* 'through' significantly increased the likelihood of a boundary-crossing interpretation in comparison to the ambiguous prepositions (Estimate = 1.7545, $p < 0.001$). Verb class 1 significantly increased the likelihood of a boundary-crossing interpretation compared to verb class 2 (Estimate = 1.6054, $p < 0.001$). The second model examined the interaction of individual verbs and prepositions with regard to boundary-crossing interpretations. No significant difference between the two ambiguous prepositions *dans* and *en dehors de* was observed ($p = 0.31$). Regarding the different Manner of motion verbs, no significant difference among the verbs of class 1 was observed. In class 2 however, there was a significant difference in the interpretation of sentences containing *danser* and *nager* ($p = 0.002$).

Discussion

We have presented experimental evidence that a boundary-crossing reading is available for at least some sentences with a Manner of motion verbs and a prepositional phrase. The advantage of our approach lies in the comparability of results for different lexical items under the same conditions. Both our hypotheses were confirmed:

1. Less boundary crossing-readings occurred with the ambiguous prepositions *dans* 'in(to)' and *en dehors* 'outside/out of' than with the directional preposition *à travers* 'through'.
2. Different Manner of motion verbs show different preferences for the boundary-crossing reading. They seem to fall into two classes, only one of which clearly allows this reading. The status of *nager* 'to swim' however remains somewhat unclear; further research is required in this regard.

As our results come fairly close to those of the corpus study in Kopecka (2009), the experimental method designed here seems to be appropriate to investigate the ambiguous sentence type under discussion. As a next step, the experiment will be extended to include more prepositions.

References

Aurnage, M. 2008. Qu'est-ce qu'un verbe de déplacement?: critères spatiaux pour une classification des verbes de déplacement intransitifs du français. Congres Mondial De Linguistique Française - CMLF'08. 1905-1917.

Bates, D., Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67(1). 1–48.

Fortis, Jean-Michel. 2010. The typology of motion events. Space in Language - Leipzig Summer School 2010 –PART III.

Hendriks, Henriette & Hickmann, Maya. 2011. Expressing voluntary motion in a second language: English learners of French. In Cook, Vivian & Bassetti, Benedetta (eds.), Language and bilingual cognition, 315-340. New York: Psychology Press.

Kopecka, Anetta. 2009. L'expression du déplacement en français: l'interaction des facteurs sémantiques, aspectuels et pragmatiques dans la construction du sens spatial. Langages 173(1). 54-75.

Pourcel, Stéphanie & Kopecka, Anetta. 2005. Motion expression in French: Typological diversity. Durham and Newcastle Working Papers in Linguistics 11. 139-153.

R Core Team. 2021. R: a language and environment for statistical computing. Manual. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sarda, Laure. 2019. French motion verbs. Insights into the status of locative PPs. In: Aurnague, Michel & Stosic, Dejan (Hgg.), The Semantics of Dynamic Space in French: Descriptive, Experimental and Formal Studies on Motion Expression, 67-107. Amsterdam/Philadelphia: John Benjamins.

Hendriks, Henriette & Hickmann, Maya. 2011. Expressing voluntary motion in a second language: English learners of French. In Cook, Vivian & Bassetti, Benedetta (eds.), Language and bilingual cognition, 315-340. New York: Psychology Press.

Talmy, Leonard. 1991. Path to realization: a typology of event conflation. In: Sutton, Laurel, Johnson, Christopher & Shields, Ruth (Hgg.), Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society, 480-519. Berkeley, CA: Berkeley Linguistics Society.

Pronoun interpretation in German speech reports

Sebastian Walter[1], Yvonne Portele[1,2]
[1]Goethe-Universität Frankfurt, Germany
[2]Freie Universität Berlin, Germany

https://doi.org/10.36505/ExLing-2024/15/0035/000660

Abstract

We investigated the interpretation of German personal pronouns using a referent selection task. Participants read contexts introducing three characters followed by pictures depicting an utterance made by one referent. We manipulated the embedding of this utterance in either direct or indirect discourse and the inclusion of either a 1st, 2nd or 3rd person personal pronoun, leading to differences in the necessity to perform perspective shifts for correct interpretation. Referent selection data for pronouns show more errors in direct compared to indirect discourse, suggesting participants avoid perspective shifts. The 1st person pronoun shows a deviating pattern with more errors in indirect compared to direct discourse, in line with proposals highlighting the prominent perspective introduced by 1st person pronouns.

Keywords: pronoun interpretation, speech reports, direct discourse, indirect discourse

Introduction

Speech reports in direct discourse (DD; (1)) are verbatim quotes of what someone else said or thought (e.g., Davidson 1979). To correctly interpret the pronouns in (1), comprehenders must perform a perspective shift (shifting from the reporting speaker's to the reported speaker's perspective): The 1st person pronoun *I* refers to the reported speaker (*Peter*), the 2nd person pronoun *you* refers to the reporting speaker (the addressee of the utterance), and the 3rd person pronoun *he* refers to some other person. Indirect discourse (ID; (2)), by contrast, paraphrases what someone else has said or thought, without directly quoting them. For correct pronoun interpretation in (2), no perspective shift is necessary (i.e., interpretation from the reporting speaker's perspective): The 1st person pronoun *I* refers to the reporting speaker, the 2nd person pronoun *you* refers to the addressee in the current utterance context, and the 3rd person pronoun *he* refers to the reported speaker (*Peter*).

(1) Peter sagte: "Ich/Du/Er bringe/bringst/bringt Wein."
Peter said, "I/You/He will bring wine."

(2) Peter sagte, (dass) ich/du/er den Wein bringe/bringst/bringt.
Peter said (that) I/you/he will bring wine.

Köder et al. (2015) showed that Dutch speakers made more errors in identifying 2nd and 3rd person pronoun referents in DD compared to ID, whereas adequacy for 1st person pronouns in DD approximates that of ID.

This has been attributed to perspective shifts, which are necessary to correctly interpret pronouns in DD but not in ID. Based on the findings of Köder et al. (2015), we hypothesize to find higher error rates in DD compared to ID in our study on German. Furthermore, the error pattern for the 1st person pronoun might diverge from the 2nd and 3rd person pronouns.

Methodology

Participants

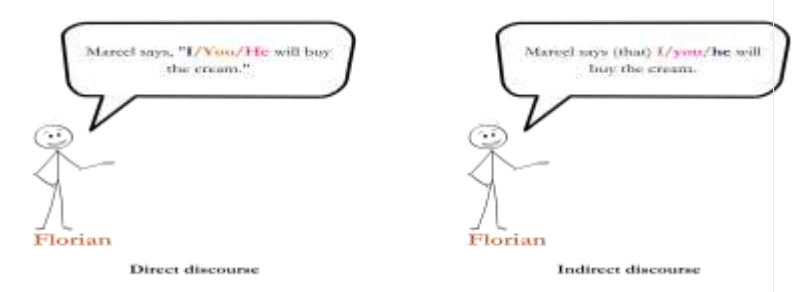
Ninety adult native speakers of German were recruited online via Prolific.

Materials and procedure

We created 24 experimental sets of written stimuli. An example is shown in (3). Each trial consisted of a context and a target picture. The context introduced two male referents in the first sentence who did something together (*Florian* and *Marcel* in (3a)). The second context sentence (3b) introduced a third male referent who joined the other two characters later (*Tim*). This referent then talked to the first referent. The context was followed by a depiction of the first character (presented as a stick figure with the respective proper name shown below) reporting what the second character said within a speech bubble. The speech bubble contained an utterance in DD or ID and contained a 1st, 2nd, or 3rd person pronoun (see (3c)). Following the context and the picture, participants continued to the next screen, where they were presented the prompt for the referent selection task. We asked participants who will perform the action mentioned in the speech bubble, giving them all three proper names as possible answers (three-alternative forced choice task). Materials were presented in German, we use translations in the pictures for reasons of space.

(3) a. Am Montag haben sich Florian und Marcel im Drogeriemarkt getroffen, um eine neue Sonnencreme zu begutachten.
On Monday, Florian and Marcel met at the drugstore to check out a new sunscreen.

b. Später kam Tim dazu und sprach mit Florian.
Later, Tim joined them and spoke with Florian.



Results

The proportion of (in)correct choices for each condition are shown in Figure 1. Overall, participants made more errors in DD (53.43%) than in ID (17.04%). The highest error rate was found for the 2nd person pronoun in DD (80.56%), whereas the lowest error rate was found for 3rd person pronouns in ID (3.6%). A generalized linear mixed effects logistic regression model (with a binary coding of wrong choice as dependent variable) revealed a significant interaction. There was a difference between speech reports for all three pronouns: Whereas more errors were made in DD compared to ID in case of the 2nd and 3rd person pronoun *you* and *he* (*you*: z ratio: 17.333, $p < .001$, *he*: z ratio: 14.446, $p < .001$), more errors in ID compared to DD were made for the 1st person pronoun *I* (z ratio: -8.310, $p < .001$).

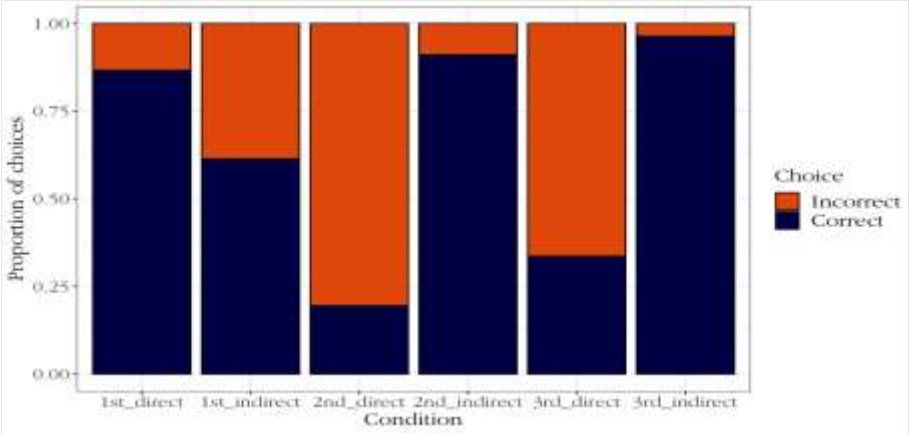


Figure 1. Proportion of incorrect (orange) and correct (blue) choices.

Discussion

The results confirm the overall pattern found by Köder et al. (2015) for Dutch. The main difference between their and our results is that in our study, participants performed significantly worse in the referent selection task for 1st person pronouns in ID compared to DD. In their study, by contrast, participants performed equally well in selecting the correct referent of the 1st person pronoun in DD and ID alike. Comprehenders seem to avoid perspective shifts, explaining the worse performance in DD for 2nd and 3rd person pronouns.

This interpretation is strengthened when looking at the type of wrong choices: for 2nd person pronouns, participants overwhelmingly chose the third referent, i.e., the addressee in the reporting context. For 3rd person pronouns, participants chose the second referent in most trials, i.e., neither speaker nor

addressee in the reporting context. 1st person pronouns deviate from this pattern. However, introducing a perspective by means of a 1st person pronoun makes this perspective more salient (Saure et al. 2023). Thus, the reported speaker's perspective is salient, pressuring participants to perform a perspective shift toward that speaker (= referent 2 in contexts), resulting in correct referent choices in most trials. This line of reasoning can also account for the ID condition: in the 2nd and 3rd person pronoun conditions, participants chose correctly because no perspective shift was needed. For 1st person pronouns, participants made more mistakes by wrongly performing a perspective shift toward the reported speaker. This wrong shift might be due to syntactic factors as the reported speech type conditions (= embedded sentence in ID vs. quote in DD) were syntactically similar, as the embedded clause in ID was introduced without a complementizer and in verb-second word order. The effect should thus be weaker when introducing the embedded clause in the ID condition by means of the complementizer *dass* ('that'), resulting in a verb-final structure. This might account for differences between this study and Köder et al. (2015) as Dutch only allows for ID to occur when introduced by a complementizer + verb-final word order (leaving aside methodological differences for now).

To conclude, direct discourse seems to be more demanding than indirect discourse in language comprehension. Future research should take into account the role of different syntactic structures (ID) as well as the processing of speech reports (e.g., on-line measurements, memory representations). Moreover, in a next step, it is worth investigating the effect speech-accompanying gestures have on the interpretation of pronouns in speech reports, as they have been argued to be cues of perspective shift (cf. Harris & Potts 2011).

References

Clark, H. H., Gerrig, R. J. 1990. Quotations as demonstrations. Language 66(4). 764–805. doi:10.2307/414729

Davidson, D. 1979. Quotation. Theory and Decision 11(1). 27. doi:10.1007/bf00126690.

Harris, J. A., Potts, C. 2011. Predicting perspectival orientation for appositives. In Bochnak, R. (ed.), Proc. 45th Annual Meeting of the Chicago Linguistic Society, 207-221, Chicago, IL, USA.

Köder, F., Maier, E., Hendriks, P. 2015. Perspective shift increases processing effort of pronouns: A comparison between direct and indirect speech. Language, Cognition and Neuroscience 30(8) 940-946. doi:10.1080/23273798. 2015.1047460.

Saure, C., Hinterwimmer, S., Jordan-Bertinelli, A. P. 2023. An experimental investigation of the interaction of narrators' and protagonists' perspectival prominence in narrative texts. Zeitschrift für Sprachwissenschaft 42(2), 341-372. doi:10.1515/zfs-2023-2009.

Index of names

ExLing 2024 Paris
Ebook ISSN: 2529-1092
Ebook ISBN: 978-618-84585-5-0