



1. Introduction

Main aspects:

- Aimed at reporting the **phoneme frequency in Mapudungun**
- **Mapudungun** is a language spoken in the Southern Cone of South America.
- A lexicon of mono-morphemic units –considered as types– was used.
- A corpus was prepared using a Mapudungun-Spanish Dictionary published in 1916, –an exceptional lexicographical work (Salas 1985, Molineaux 2016).



Fig. 1: Mapuche territory

Phoneme frequency can:

- characterise a particular language (Tambovtsev, 2010),
- establish general trends within languages (Gordon, 2016),
- make comparisons among world languages (Everett, 2018),
- inform applied linguistics (language development and teaching, clinical studies of language-related diseases, diachronic studies, etc.).

The **phonemic inventory** of Mapudungun contains 6 vowels and 22 consonants, which is considered average (Maddieson, 2013).

	fr	cn	bk	stop	lab	den	alv	ret	pal	vel
close	i		u	p	t	t	ʈʂ	ʈʂ	ʈʂ	k
mid	e	ə	o	m	ɲ	n		ɲ	ɲ	ŋ
open	a			f	θ	s		ʃ		
				lat	l	l		ʎ		
				app	w			ɭ	j	ɰ

Fig. 2: Phoneme inventory

2. Methods

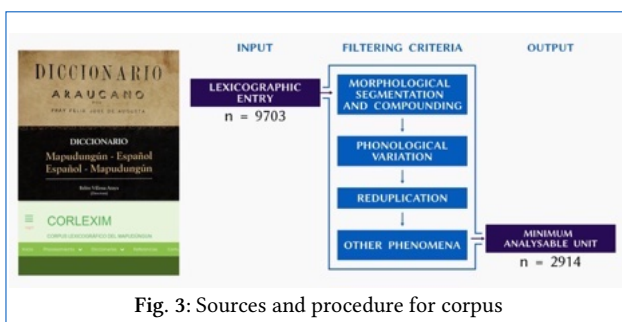


Fig. 3: Sources and procedure for corpus

- This is a quantitative descriptive research project of a non-experimental cross-sectional type.
- Morphosyntactical aspects were filtered out in order not to skew the results: inflectional and derivational suffixes, syntactic reduplication, compound words and nominal incorporation were considered.

- (1) kimeltuchefe **Derivation and flection**
kim -el -tu -tʃe -fe
know -CAUS -APL -person -AG
'teacher, lit. that who makes people know'
- (2) chafokutran **Compounds**
tʃafo + kufsan
cough + disease
'a cold with cough'
- (3) yofyofkūn **Reduplication and other phenomena**
jof -jof -ka -n
boil-RED -RES -INF
'to boil with sound'

Other aspects were considered, such as the exclusion of loanwords from Spanish and Quechua, as well as phonemes in interjections, due to their phonotactic anomaly. A list of 2914 mono-morphemic roots was gathered, from which the frequency rate of the 28 phonemes in Mapudungun was counted, using R.

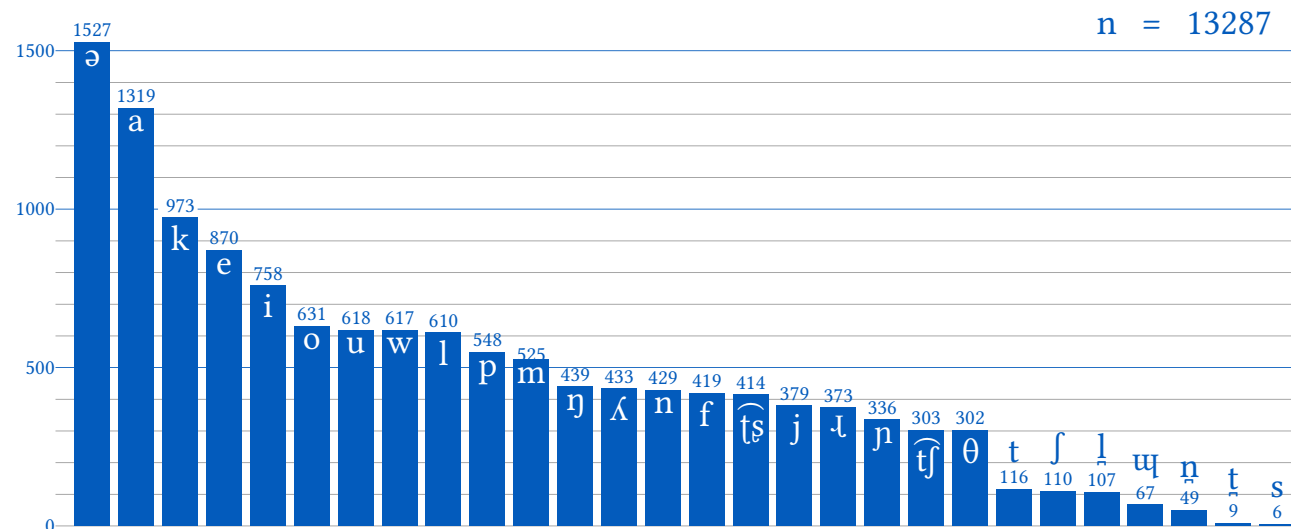


Fig. 4: Absolute frequencies.

3. Results and analysis

Phoneme frequency of Mapudungun:

- It shows a typically shaped distribution for this type of phonostatistical study (Martin, 2007).
- A subset of the phonemes accounts for a high percentage of the whole distribution.
- The first seven phonemes, for example, correspond to 50% of the total count, while the remaining 19 phonemes have gradually less frequencies, with some segments showing relatively imperceptible rates.

It is of note that:

- The highest count is for mid vowels /ə/ and /a/, /k/ and the remaining vowels.
- 75% of the phonemes with the highest frequency are sonorant segments (except for /k/ and /p/).
- The eight coronal obstruents have low frequency distribution (alveolars being in the lowest end).

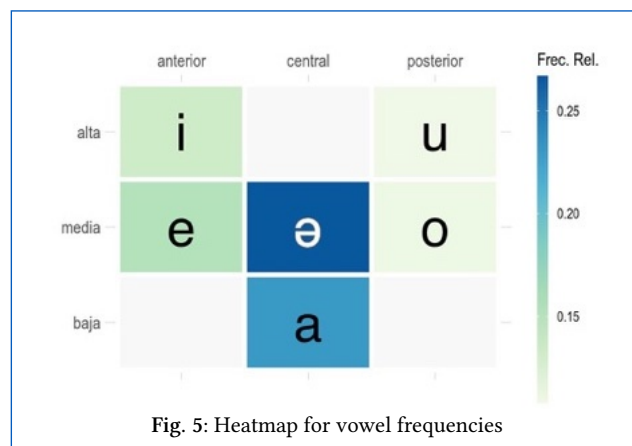


Fig. 5: Heatmap for vowel frequencies

Dental consonants have a low frequency in the corpus, which coincides with the correlation between intra-language frequencies and cross-language frequencies (Gordon, 2016). Given that the phonological contrast between dental and alveolar counterparts is typological marked (Molineaux, 2018), it was expected that the four dental phonemes in Mapudungun would have low frequencies. Similarly, in typological terms, it is noteworthy that /t/ and /s/ have low frequencies, since they tend to be an unmarked segments in world inventories (Gordon, 2016).



Fig. 6: Heatmap for consonant frequencies

4. Discussion

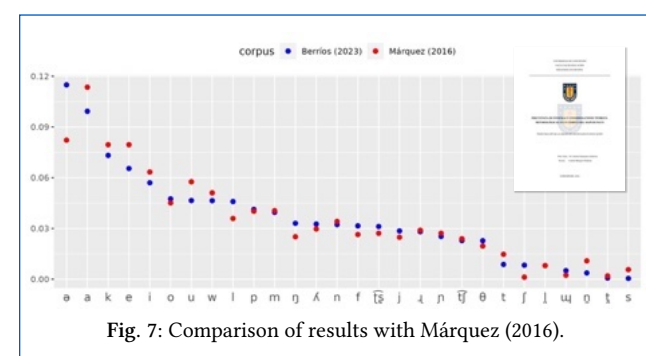


Fig. 7: Comparison of results with Márquez (2016).

A comparison was made with another study carried out with a smaller sample of dictionaries published in the 1990s, with a similar methodology (i.e. Márquez 2016). Similar results were obtained when comparing the relative frequencies in both studies.

Interestingly, the largest difference between both phonostatistical studies was the relatively higher frequency of /ə/ in our study, compared to the previous study, in which it occupied the second place. This may be related to sociolinguistic phenomena given that the materials used in Márquez (2016) and in our study cover different times periods. Thus, it is possible that language change may be altering the phonemic frequency distributions. An interesting notion to further explore is that of 'lexical impoverishment' (Choidi & Loncon 1999), in which own lexical items are lost due to contact with a majority language. It could be hypothesised that vocabulary with the /ə/ segment has becoming lost over time, while being replaced with loanwords from Spanish, which usually do not have that segment. Similarly, the productive derivation system of the language allows coinage of new words, which may amplify the effect if items without /ə/ are used to create new lexicon. All in all, a more in deep analysis is needed in order to identify this situation.

An interesting idea for further exploration is the apparent relationship between two articulatory axes and frequency distribution, depicted in Fig. 8.

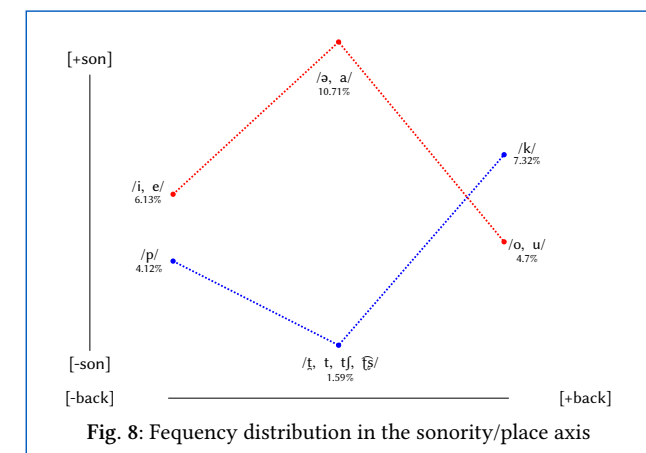


Fig. 8: Frequency distribution in the sonority/place axis

References

- Chiodi, F. & Loncon, E. (1999). Crear nuevas palabras. UFRO-CONADI.
- Everett, C. (2018). The similar rates of occurrence of consonants across the world's languages. *Language Sciences*, 69, 125–135.
- Gordon, M. (2016). *Phonological Typology*. Oxford University Press.
- Maddieson, I. (2013b). Consonant-Vowel Ratio. In M.S. Dryer, y M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. [2022-04-27 <http://wals.info/chapter/3>]
- Márquez, C. (2016). Frecuencia de fonemas y consideraciones teórico-metodológicas en un corpus del mapudungun. Thesis.
- Martin, A. T. (2007). *The evolving lexicon*. Dissertation.
- Molineaux, B. (2016). Vigencia del Dicionario Araucano de Félix de Augusta... *Boletín de Filología*, 51(1), 187–209.
- Molineaux, B. (2018). Pertinacity and change in Mapudungun stress assignment. *IJAL*, 84(4), 513–558.
- Salas, A. (1985). Fray Felíz José de Augusta. Su aporte... *CUHSO*, 2(2), 197–272.
- Tambovtsev, Y. (2010). Euphony in world languages. *Lingua Posnaniensis*, 52(1), 99–111.

ABSTRACT



Aldo Berríos Castillo
a.p.berrios-castillo@sms.ed.ac.uk
<https://aldoberrios.cl>

Non-compositionality in sentiment: new data and analyses

Verna Dankers, Christopher G. Lucas

EMNLP-Findings 2023

1. Introduction

Sentiment computations largely adhere to the principle of compositionality:

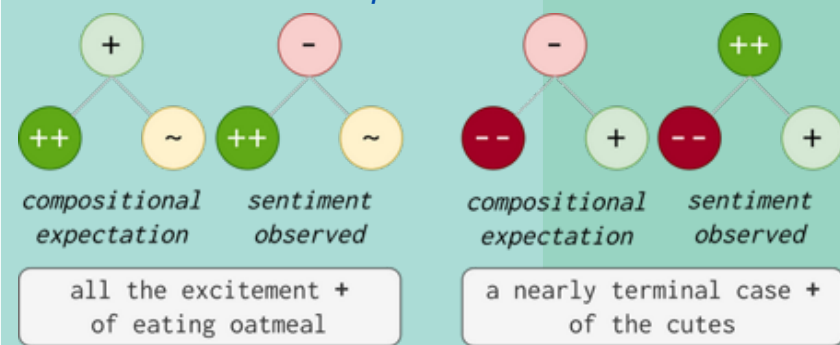
"If the meaning of a sentence is a function of the meanings of its parts then the global polarity of a sentence is a function of the polarities of its parts." (Moilanen & Pulman, 2007, p.1)

Nonetheless, many phrases seem non-compositional, e.g. due to sarcasm, formulaic language or the presence of contextual valence shifters.

How can we identify whether sentiment is non-compositional? We...

- design a protocol to obtain non-compositionality ratings
- create the **NonCompSST** resource containing ratings for 259 phrases
- illustrate how NonCompSST can complement existing sentiment analysis evaluation protocols in NLP

Illustrations of non-compositional sentiment we observe in our experiments

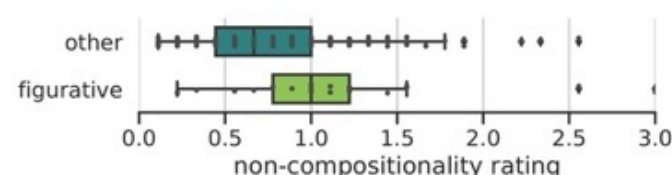


2. Methodology

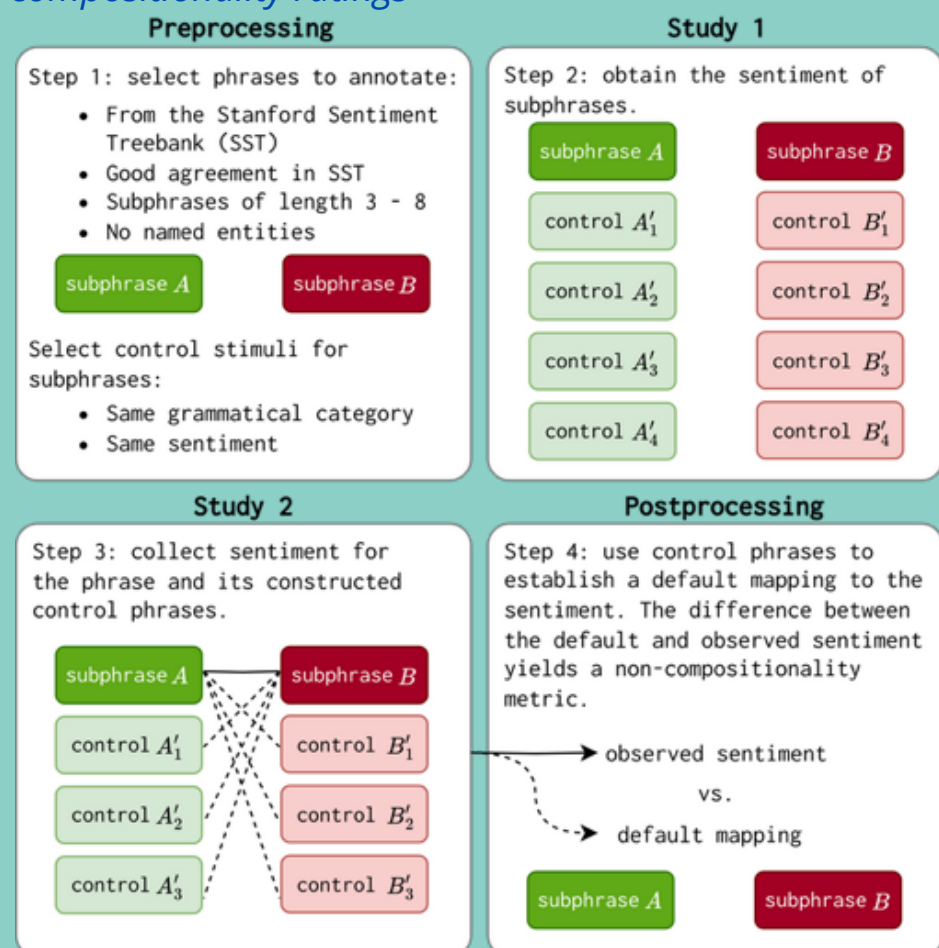
- Compositional processing applies a function to the polarity of subphrases to obtain the polarity of a phrase. Non-compositional phrases deviate from the expected output of that function (the default).
- We establish a **default mapping** by replacing the left and right subphrase with control stimuli and measure the difference in sentiment to the **observed sentiment**: this is our **non-compositionality rating**.
- This is applied to phrases from the Stanford Sentiment Treebank (SST, Socher et al., 2013), our resulting resource has non-compositionality ratings for 259 phrases.
- Sentiment annotations (on a 7-point scale) are collected via Prolific.

3. The non-compositionality ratings

- For most phrases, sentiment behaves compositionally; only 67 phrases have a rating that exceeds 1.
- Combining phrases with equal sentiments is more compositional than combining opposite sentiments.
- Among the non-compositional examples, we observe:
 - figurative language (e.g. "a pressure cooker" + "of horrified awe")
 - discourse relations (e.g. "fans of the animated wildlife adventure show" + "will be in warthog heaven")
 - common-sense reasoning (e.g. "the franchise's best years" + "are long past")

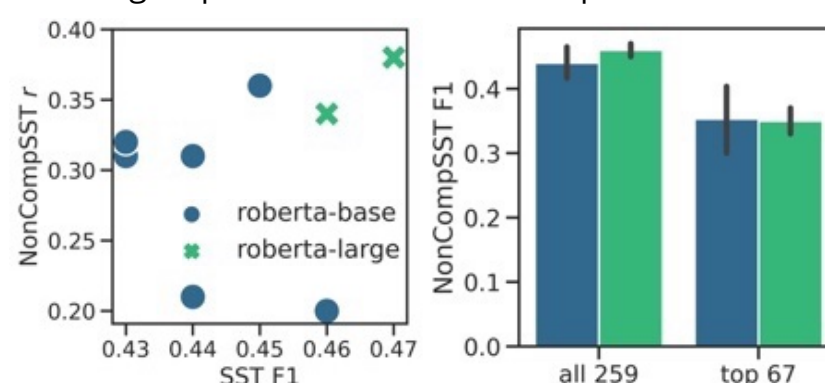


Overview of the methodology to obtain non-compositionality ratings



4. Model evaluation

- We evaluate RoBERTa-based (Liu et al., 2019) models:
 - a. Using F1-score on SST test data (using a 7-point scale);
 - b. Using Pearson's r comparing humans' and models' non-compositionality ratings.
- The results indicate NonCompSST can complement existing evaluation protocols.
- We also compute F1-scores over all 259 phrases, and over the 67 most non-compositional phrases: for the latter group sentiment is harder to predict.



Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv

Moilanen, K., & Pulman, S. (2007). Sentiment composition. RANLP

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. EMNLP

POTENTIAL OF BOOLEAN MONADIC RECURSIVE SCHEMES IN SYLLABIFICATION TYPOLOGY

Undergrad dissertation on computational phonology

INTRODUCTION

Boolean Monadic Recursive Schemes (BMRSs) were recently proposed as a framework for modelling phonological patterns. BMRS predicates are written in an *if... then... else...* structure, where the ellipses can be either a truth value or any other predicates in the same structure. They retain the typological strength of constraint-based theories by allowing structural rearrangement.

An example

- Hypothesised tone pattern:
 - Monosyllabic: /H/ → [H]
 - Disyllabic: /LH/ → [HH]
 - Trisyllabic: /HLH/ → [HHH]
- :BMRS analysis
 - $H(x)$ = if the underlying tone on x is high then \top else \perp
 - $H_o(x)$ = if $\bowtie(s(x))$ then $H(x)$ else $H_o(s(x))$

Note: Here, x stands for a syllable. \top for true, and \perp for false. $s(x)$ refers to the successor of x . \bowtie stands for the word's right boundary. $H_o(x)$ tells whether x should surface with a high tone.

- Evaluation of the trisyllabic pattern above

Input	\bowtie	H	L	H	\bowtie
$H(x)$	\perp	\top	\perp	\top	\perp
$H_o(x)$	\perp	\top	\top	\top	\perp
Output		H	H	H	

Syllabification: Why and How

WHY: Typology of feasible syllable shapes is well-documented. Existing findings can be used to build a BMRS account of syllabification. This domain is appropriate for testing BMRSs' potential in generating typology.

HOW: The BMRS analysis I proposed for syllabification consists of two groups of predicates. The first group deals with nuclei, onsets, and codas identification. The other decides how to fix unsyllabified segments: deletion, epenthesis, or just letting them surface faithfully.

- e.g., the default nucleus identification predicate is
 - $nucl(x)$ = if $V(x)$ then $C(p(x))$ LINE A: Onset obligatory
 else if $V(x)$ then \top LINE B: Onset optional
 else \perp

Note: $p(x)$ refers to x 's precedent. V stands for vowel and C for consonant.

- which implies an onset-obligatory language.
- **BMRSs' permutability:** Switching the order of LINE A and LINE B allows the vowel without a preceding consonant to become a nuclues, thus implying an onset-optional language.

This dissertation explores BMRS's typological potential through permuting the hierarchy of conditional statements inside BMRS predicates leads to distinct phonological patterns.

METHOD

The BMRS analysis I proposed includes five permutable predicates: $nucl(x)$, $onset(x)$, $coda(x)$, $repair(x)$, and $del(x)$. The first three determine the role of a segment in a sequence consists of 'C' and 'V'. The last two decide whether to fix an unsyllabified symbol and how to fix it (delete this segment when $del(x)$ returns true, otherwise do epenthesis). Among them, $coda(x)$ has six permutations, while each of the other four has two. I implemented the analysis with a Python script.

Implementation Component

Input test strings

- A list with strings consisting of one to six Cs or/and Vs in designed combinations
- e.g., "VC", "CCVCC", and "CVCCVC".

Grammars constructor

- Permutes all five permutable predicates in all possible ways
- Constructs grammars by combining every permutation of the five predicates with each other
- 96 ($2 \times 2 \times 6 \times 2 \times 2$) grammars in total

Ouput generator

- Inputs the test string list into each of the 96 grammars
- Collects the output patterns

Data organiser

- 36 distinct output patterns in total
- Maps each grammar to the generated output
- Among grammars producing the same output result, finds the shared constraint hierarchies of predicates within such grammars

RESULTS

The 36 output patterns can be grouped into 12 types according to the surfaced syllable shapes in each pattern. One specific shape corresponds to three possible output patterns. The 12 types of syllable shapes are: CV, CV(C), CV(C)(C), (C)CV, (C)CV(C), (C)CV(C)(C), (C)V(C), (C)V(C)(C), (C)(C)V, (C)(C)V(C), and (C)(C)V(C)(C). This matches well with empirical documentations on syllabification typology.

CONCLUSION

This investigation confirms BMRSs' capability of predicting typology of syllable shapes by employing the framework's permutability and computability. Future examinations on BMRSs' typological potential can extend into other well-studied phonological or even morphological domains.

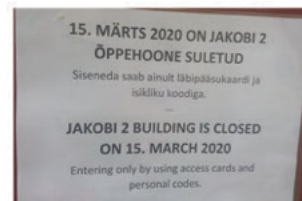
--- Presented by Yujia Sun Y.Sun-190@sms.ed.ac.uk ---

Linguistic autobiography as experience-based teaching in Linguistics

Liisa-Maria Komissarov, Linguistics MSc
L.Komissarov@sms.ed.ac.uk

INTRODUCTION

- The coursework of ‘Introduction to Linguistics’ at the University of Tartu is a **linguistic autobiography**: a 1800-2200 word paper where the students describe and analyse their own language-related experiences **according to a framework of materials and terminology covered in the lectures**.
- In order to walk the students through this process, they receive a task with
 - guiding **questions** (see below),
 - a list of up-to-date relevant **sources**,
 - a list of linguistic **terms** for each subtopic they have to write about.
- The tasks are written in an easy and accessible language, using the Socratic method inspired questions to prompt students to **contemplate, re-evaluate, and generalise their life experiences** and **associate and compare them with scientific theories** that they learned about in the lectures and the sources provided to them.



EXAMPLES

Example 1. The guidelines given to the students for the 1st draft's 1st subtopic

Can you ask your parents which were your first words? Can you explain, why these particular words? When did they appear?
What about longer utterances? Give concrete examples!

[...]

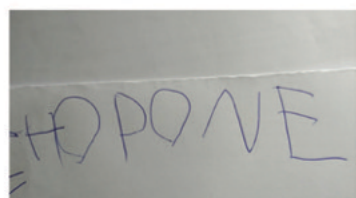
In which kind of a linguistic environment (i.e. monolingual, bilingual) have you acquired your first language(s)?

Do you consider yourself bilingual? Explain!

Do you remember or can you find out what was your first contact with a foreign language? Describe it!

Example 2. The respective section from a student's work

From the morphological traits that characterise child language, reduplication (e.g. auh-auh 'woof-woof', adaa-adaa 'bye-bye') was present in my speech, and from the phonological traits (Garmann et al. 2020) reduction (e.g. at 'auto' (Estonian for 'car')).



hobused									
@	p_v	u	s	@	t				
v	c	v	c	v	c				
hobune+d // _S_ pl n, //									

FEEDBACK

- We collect the **feedback** by students about how difficult and how interesting the course was for them.
- The majority of students admitted that the course was relatively difficult.
- However, a similar percentage of students claimed the course to be interesting for them – a sign that the course was challenging but **rewarding**.
- The written feedback also reveals that **students realise and value the experience they have gained**.

COURSE OUTLINE

- 3 ECTS
- 200-250 students
- **introductory course covering the main branches, concepts and functions of linguistics**
- weekly lectures and one coursework
- mandatory in the 1st semester for **undergraduate students** of Linguistics, Literature, Semiotics, Foreign Languages and Literature, Educational Sciences.

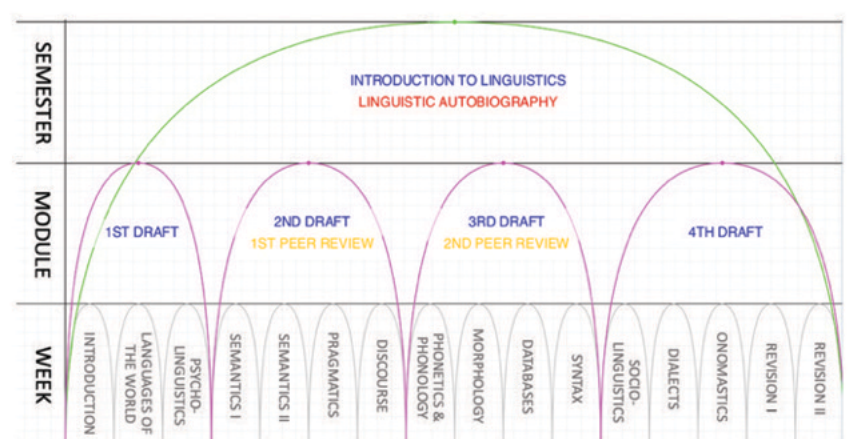


Figure 1. Course outline as a learning arch. Created in LearningArches by Kaospilot

TAKEAWAY

More emphasis should be put on tapping into the **personal experience of students** when teaching linguistics. This makes learning – and teaching – linguistics a lot more rewarding for everyone involved. Not every student will become a linguist, but they are all language users.

Excerpt from student feedback:

“Writing my own linguistic autobiography taught me to describe my own experiences and observances with scientific methods.”

FIND OUT MORE



Tragel, I., & Komissarov, L.-M. (2022). Combining subject-specific and subject-independent competencies in teaching Linguistics. 8th International Conference on Higher Education Advances (HEAd'22). Ed. Josep Domenech. Valencia: Editorial Universitat Politècnica de València, 1377-1384. <http://doi.org/10.4995/HEAd22.2022.14490>



'Falou alto': A DIACHRONIC ANALYSIS OF SOME MICROCONSTRUCTIONS OF THE [V + AA] SUBSCHEMA CONSTRUCTIONALIZED IN BRAZILIAN PORTUGUESE

THE FEDERAL UNIVERSITY OF RIO DE JANEIRO

Author: Raissa Romeiro Cumán

Email: raissacuman@letras.ufrj.br

Supervisor: Prof. Dr. Priscilla Mouta Marques

Prof. Dr. Graeme Trousdale

INTRODUCTION

This work consists of a diachronic analysis of some microconstructions of the [V AA] subschema in Brazilian Portuguese, namely: [falar alto], [bater forte], [fazer bonito], [fazer feio], [pagar caro], [pegar leve], [pegar pesado] and [dar certo] which are constructionalized in current Brazilian Portuguese (BP). To do this, we will use the theoretical-methodological contribution of Usage-Based Linguistics, which understands that constructions are the basic units of the language, related to each other in a constructional network. (GOLDBERG, 1995; 2006, CROFT, 2001).

GOALS

Our main objective is to carry out a diachronic analysis of some micro-constructions of the [Adverbial-Adjective Verb] subschema that are constructionalized in current Brazilian Portuguese, identifying the formal and functional characteristics presented by them, as well as the micro-steps involved in the change observed in these patterns and the contexts (Diewald, 2002, 2006) that might have motivated the neo-analyses that led to the emergence of such constructions in the language. Furthermore, the aim is to understand the links established between such micro-constructions and other nodes of the construction, such as the constructional network of light verbs.

Examples (1) to (4) present constructs (instantiations of use) of some of the constructionalized microconstructions under investigation (highlighted in bold) and constructs in which the same V + AA sequence is more compositional (highlighted in italics)

(1) "No teatro é até mais grave, porque os atores estão ali, despendendo uma energia diariamente, e, por isso, o foco e a atenção (do espectador) devem ser maiores. A pior atitude é *falar alto*, falar ao celular, ligar o telefone, porque a luz atrapalha demais. Interfere, incomoda e desrespeita quem está ali assistindo" (Corpus do Português ABA NOW 102 19-06-30 BR Estado de Minas)

'In theater it is even more serious, because the actors are there, spending energy every day, and, therefore, the focus and attention (of the spectator) must be greater. The worst attitude is to *speak loudly* [*speak loud*], talk on the cell phone, turn on the phone, because the light is too distracting. Interferes, bothers and disrespects those who are there watching'

(2) "Na novela O Sétimo Guardião, a curiosidade **fala alto** e ele se esforça para passar pelo local. Depois de muito esforço, Feijão tem uma grande surpresa: uma galeria com muitos artefatos antigos. Além dos objetos antigos, uma misteriosa figura de gato preto está desenhada na parede." (Corpus do Português ABA NOW 209 19-05-11 BR Gshow)

'In the soap opera *O Sétimo Guardião*, curiosity **gets the better of him** [**speaks loud**] and he makes an effort to pass by the place. After a lot of effort, Feijão has a big surprise: a gallery with many ancient artifacts. In addition to the antique objects, a mysterious black cat figure is drawn on the wall.'

(3) "O gol deixou a Suécia mais confiante. Aos 36 minutos, chegou com perigo mais uma vez. Rolfo deu um bom passe para Blackstenius, furando a defesa. A centroavante sueca *bateu forte* para o gol e a goleira alemã espalmou para fora." (Corpus do Português ABA NOW 203 19-06-29 BR EBC)

'The goal made Sweden more confident. In the 36th minute, he arrived dangerously once again. Rolfo made a good pass to Blackstenius, breaking through the defense. The Swedish center forward *hit the goal hard* and the German goalkeeper threw it wide.'

(4) "Meu neurologista Francisco Cardoso já nem me aconselha mais a fazer exercício – não quer perder tempo com conselhos que não serão certamente seguidos. Mas me aconselha muito tomar um avião para outros países, procurar novos conhecimentos, evitar pontos onde a saudade pode **bater forte**." (Corpus do Português ABA NOW 309 19-06-13 BR Estado de Minas)

'My neurologist Francisco Cardoso no longer advises me to exercise – he doesn't want to waste time with advice that will certainly not be followed. But I am highly advised to take a plane to other countries, seek out new knowledge, and avoid places where homesickness can **hit** [**hit hard**].'

Structural and pragmatic-discursive factors related to the previously mentioned microconstructions in different synchronies will also be analyzed, such as: the informational structure, the semantic nature of the subject and the elements in function of object, the degree of compositionality of the constructions, the presence of intervening elements and their nature and, finally, the textual genre/discursive domain in which the constructs occur

HYPOTHESIS

In this research, we hypothesize that constructions with an already constructionalized adverbial adjective would have been originated from a more compositional construction, in which the adjective is modifying the verb still in its original meaning. The construction [falar alto], for example, in which the verb *falar* indicates a verbal action and is being modified by the adjective *alto*, would have originated the constructionalized [falar alto] construction in which the verb *falar* is no longer a verb of saying and the meaning of the construct is not the sum of the meaning of its component parts.

Furthermore, we argue that these constructions, based on the constructionalization process (TRAUGOTT AND TROUSDALE, 2013), started to present functional and discursive-pragmatic characteristics different from those presented by the more compositional construction that would have originated them, such as, for example, a greater restriction for the presence of intervening elements, a change in the items that can occur in subject position (these being less agentive and less animated in the lexicalized construction) and a greater incidence of focus on the construction to the detriment of other elements present in the same sentence.

METHODOLOGY

For the development of this research, we will quantitatively and qualitatively analyze usage-based data collected in some corpora. To collect data from microconstructions in different synchronies, we will use the *Corpus do Português*, Genre/History tab, which contains 45 million words from different genres and discursive domains between the 13th and 20th centuries. To collect data from microconstructions in the current synchrony, we will collect data from the *Corpus do Português*, WEB/Dialects tab, which contains approximately one million web pages from four Portuguese-speaking countries (Brazil, Portugal, Angola and Mozambique), only data referring to Brazilian Portuguese will be analyzed.

In each of the microconstructions that are the objects of this study, structural and pragmatic-discursive factors will be analyzed. They are:

- 1-the presence of intervening elements and the nature of them;
- 2- the informational structure of the microconstruction, verifying whether the focus is exclusive or shared with other items in the clause;
- 3- the elements functioning as subject in the sentences in which the construct is identified;
- 4- The transitivity of the verbs in these constructions and, where applicable, the items that occur as object;
- 5-the degree of compositionality of the microconstruction;
- 6- the different semantics of each microconstruction, depending on the context of use;
- 7- the genre/text type/discursive domain in which these constructs occur.

FINAL REMARKS

This work is at an early stage, and it is still too early to present preliminary results given the few occurrences collected at this time. We believe that this research can contribute to the mapping of the network of adverbial adjectives as well as the network of light verbs in Brazilian Portuguese.

REFERENCES

- BYBEE, J. *Language, usage and cognition*. Cambridge: Cambridge University Press, 2010.
- GOLBERG, A. E. *Constructions: a construction grammar approach to argument structure*. Chicago: The University of Chicago Press, 1995.
- TRAUGOTT, EC; TROUSDALE, G. *Constructionalization and constructional changes*. Oxford: Oxford University Press, 2013
- DIEWALD, G. A model for relevant types of contexts in grammaticalization. In *New Reflections on Grammaticalization*, 2002.
- LAMBRECHT, K. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. In *Cambridge Studies in Linguistics* 71, 1994.
- CROFT, W. *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press, 2001.



The impact of linguistic register variation in detecting morphosyntactic errors

Camilla Masullo¹, Alba Casado², Evelina Leivada³

¹Universitat Rovira i Virgili, ²Universidad de Granada, ³Universitat Autònoma de Barcelona - ICREA



UAB

Introduction

Linguistic register is defined as a variety of specialized language shaped by specific situational features, linguistic functions, social contexts, and communicative purposes (Biber & Conrad, 2009; Pescuma et al., 2023). Despite Register has been described by different linguistic and cognitive models (Hymes, 1974; van Dijk, 2005, 2008), few studies have focused on its effect on language processing (Bentum, 2022; Pescuma, 2022; Rotter & Liu, 2022; Wiese, 2022) revealing that listeners are sensitive to variation in linguistic register with consequences on language processing mechanisms. If we consider register knowledge as an inherent component of users' linguistic repertoires, it is reasonable to expect that individuals with different linguistic backgrounds exhibit some variation in the way they navigate the demands of various linguistic registers (Wiese & Rehbein, 2016). This difference could stem from the quantity and type of the received linguistic input: people with diverse linguistic trajectories may vary in the situational and communicative contexts in which they learn and use their languages (Backer & Bortfeld, 2021; Leivada et al., 2021 *inter alia*), and this could, in turn, affect their competence in handling register variation in their languages.

Our study

Considering the literature gap on the effect of register variation on language processing, we aim to examine whether a more (in)formal linguistic register strengthens or weakens the ability to detect morphosyntactic errors, specifically *Subject-Verb agreement mismatches*. Assuming that different linguistic backgrounds may bring to variation in the way people master linguistic registers (Wiese & Rehbein, 2016), we adopt a *comparative perspective*, including monolingual speakers and different groups of bilingual speakers in our sample.

Research questions

- I. Does register variation modulate the ability to detect morphosyntactic errors?
- II. Does register variation play the same role in detecting errors in monolingual and bilingual populations?

Methods

Participants (108)

- Italian “monolinguals” (27)
- Italian-Spanish bilinguals (27)
- Italian-Agrigentino bilinguals (28)
- Italian-Pavese bilinguals (26)

Task

- Auditory timed Acceptability Judgement task
- Background questionnaire based on LSBQ (Anderson et al., 2018)

Stimuli

- 40 test items with Subject-Verb agreement mismatches: 20 in **low register** (LR) and 20 in **high register** (HR), with the same semantic content.
- 60 grammatical fillers, and 20 ungrammatical fillers.



- *Il foglio degli sbirri comunali sono in un bordello assurdo.*
The sheet of the local cops are in a crazy mess.
- *Il documento dei poliziotti locali sono estremamente in disordine.*
The documents of the local policemen are very messy.

! We ran a **pilot test** on a sample of 6 neurotypical adult speakers of Italian, who were not subsequently included in the final sample, to ascertain a *clear distinction* in *linguistic register* between high- and low-register stimuli.

Analyses and Results

We ran LME and GLME models, setting *Accuracy* and *RTs* of test items as dependent variables.

FIRST SET OF ANALYSES (RQI)

Model1 = (g)lmer(Acc/RTs ~ Register * Group + Animacy + Gender + Age + Education + (1 + Register | Participant) + (1 | Item), data)

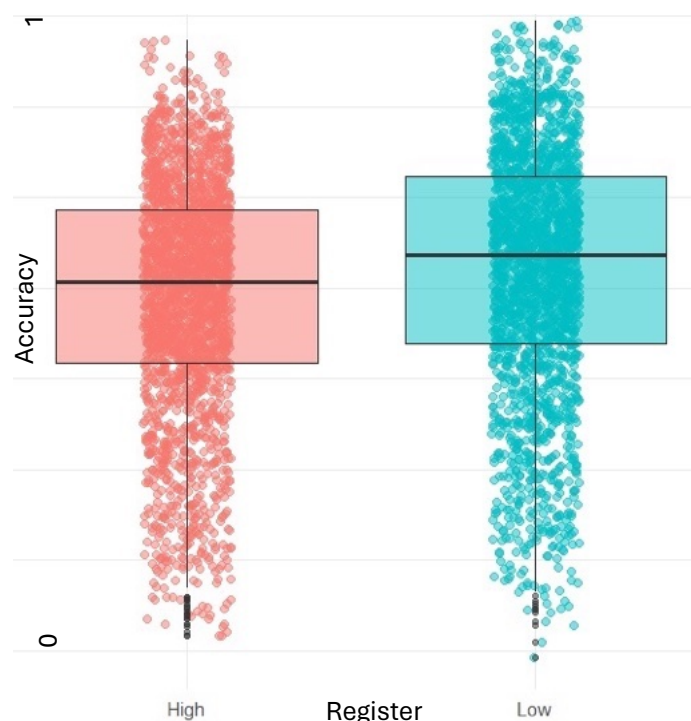


Figure 1. Register variation and accuracy rates.

- We found a *significant effect of Register* on the accuracy rates of the whole sample.

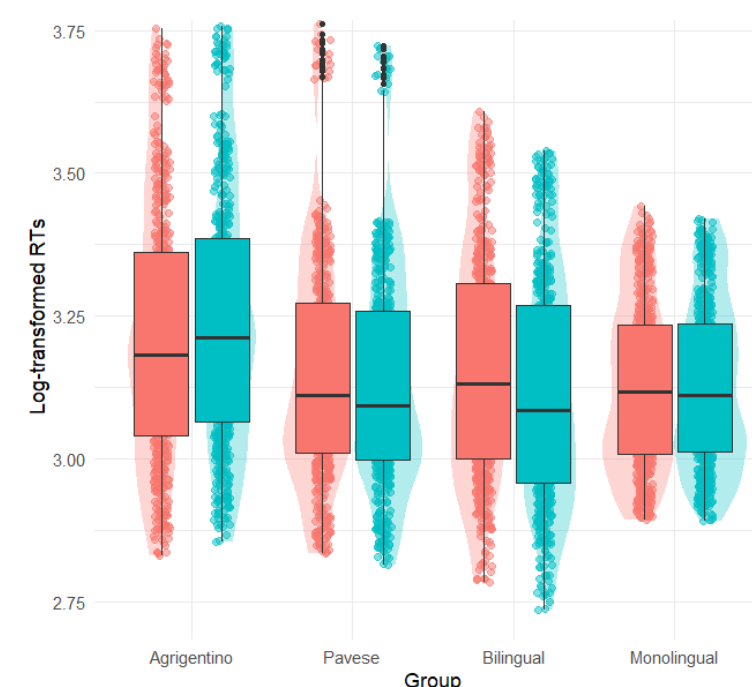


Figure 2. Register variation and RTs for each language group.

- We also found a significant effect of the *interaction* between *Register* and *Language Group* on RTs.

SECOND SET OF ANALYSES ON BILINGUALS (RQII)

Model2 = (g)lmer(Acc/RTs ~ Register * (Group + Italian Level + Dominance + Italian use + Dialect use + LanguageHighRegister + LanguageLowRegister + Switching) + Animacy + Gender + Age + Education + (1 + Register | Participant) + (1 | Item), data)

We investigated the effect of variables related to bilingual participants' language practices.

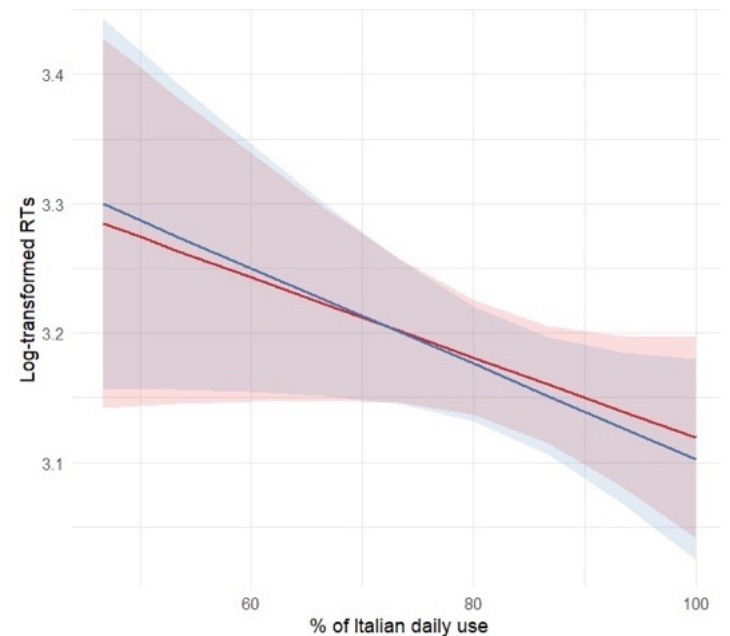


Figure 3. Italian use and RTs in low- and high-register stimuli for bilinguals.

- We found a main effect of the *percentage of Italian language use* on RTs, particularly pronounced for low-register stimuli.

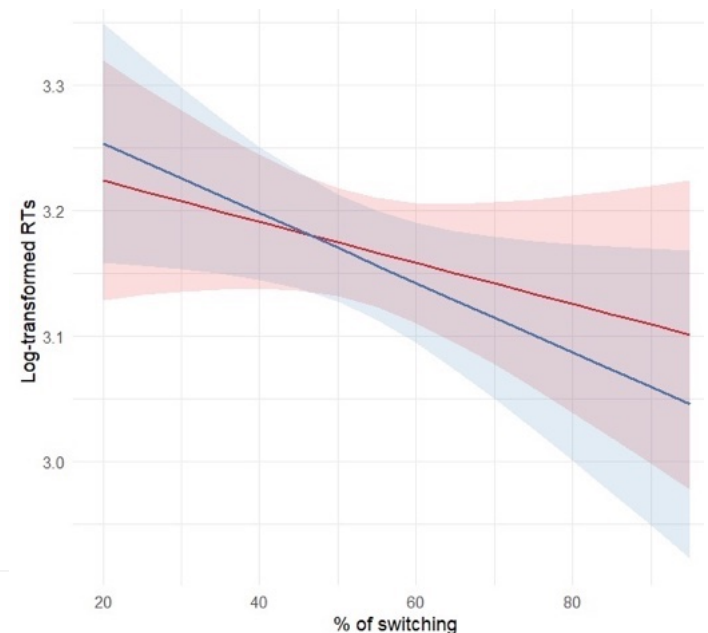


Figure 4. Switching and RTs in low- and high-register stimuli for bilinguals.

- Also the *percentage of language switching* had a significant effect on RTs, which was more evident for low-register stimuli.

Discussion

- Linguistic register has an impact on *morphosyntactic processing* → it may function as a separate domain contributing to morphosyntactic processing, raising challenges for claims on the autonomy of morphology or syntax (Radford, 1988).

RQI

- High-register sentences may be associated with more formal contexts and thus perceived as correct.
- Familiarity with low register (Giménez Moreno, 2006) → less cognitive effort for processing low-register stimuli, and more cognitive resources for the processing of morphosyntactic traits.

RQII

- Monolinguals present comparable accuracy rates and RTs in low- and high-register stimuli → their linguistic repertoire is dominated by one language, which covers both low- and high-register contexts.
- For bilingual and bidialectal communities language use is *more diversified*.
- Lower accuracy rates in high-register stimuli and longer RTs in low-register stimuli for Italian-Agrigentino bilinguals → Prestige of Italian + Agrigentino dialect is more used in low-register contexts.

Corresponding author: camilla.masullo@urv.cat

Funding: CM has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945413 and from the Universitat Rovira i Virgili (URV). EL acknowledges funding from the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033) under the research project No. PID2021-124399NA-I00.

References





A Syntactic Study of Modern Chinese *Dui* DP *de* VP from the Approach of Distributed Morphology

Na Ning

China Scholarship Council

Guangdong University of Foreign Studies

The University of Edinburgh

1 Introduction

This paper discusses the syntactic structure of a widely used modern Chinese *dui*-construction, *dui* DP *de* VP (in some literature, DP is called NP), which has long been a hot issue among linguistic academia. There is growing body of literature that investigates the nature of *dui*-construction, primarily concentrating on:

- the position of the head
- the characteristics of *de*
- the reason behind the nominalization of VP
- the explanation to the contradiction to Endocentric Theory



2 Literature Review

2.1 Position of Head

2.1.1 *De* is the Head

Without syntactic class and semantic connotation, *de* is considered to head DeP, in all *de*-containing phrases (Si 2004), see figure-1.

2.1.2 VP is the Head

Xiang (1991) mentions that DP *de* VP can only be used as subject or object, headed by VP, which preserves its predicative property.

2.2 Identity of *de*

2.2.1 *De* as a Complementiser

Si (2002) holds that the phrase *Zhangsan de cengzai meiguo liuxue* 'Zhang San has studied abroad in America before' should be considered as a CP.

2.2.2 *De* as a Determiner

Simpson (1997;2001) claims that *de* in nominal structures, possessive constructions and relative clauses in Chinese Mandarin should be analysed as a D, see figure-2.

2.2.3 *De* as a Light n

Li (2015) holds that *de* in *de*-construction, either bare *de*-construction (X+*de*) or complex *de*-construction (X+*de*+Y), should be analyzed as a light n instead of C or D, see figure-3.

Doubt is
the key
to KNOWLEDGE

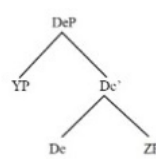


Figure-1 DeP

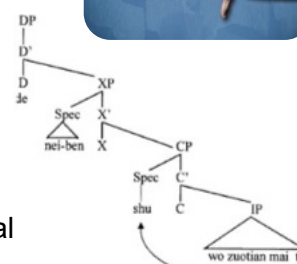


Figure-2 D as a Determiner

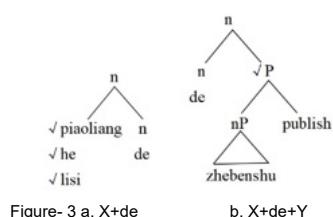


Figure-3 a. X+de

b. X+de+Y

2.3 Status of VP

2.3.1 V or VP

Xiong and Wang (2023) suggest that in *zhbenshu de chuban* 'the publication of this book', *chuban* 'publish' is a V rather than a VP, as is seen in (1):

(1) [[NP [DeP *zhe ben shu de*] [NP [N⁰]] [VP [V *chuban*]]]]

2.3.2 Nominalization or Relativisation

He and Wang (2007) have made a distinguish between nominalization and relativisation. He (2015) holds that Chinese *dui*-construction like *ta dui wo de piping* 'his criticism to me' is a relativised structure instead of a nominalized one, see figure-4.

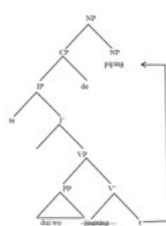


Figure-4 Relativisation of VP

3 Syntactic Analysis of *Dui*-Constructuon

3.1 Ambiguity Analysis

Three co-determinants impact the ambiguity of *dui* DP *de* VP

- the grammatical function of *dui*
- the argument structure of VP
- the animacy and semantic role of DP



3.2 Multiple Roles of *de*

De is possible to possess more than one identity, resting on the syntactic position where it locates, ranging from but not being limited to the realization of light n, Determiner, and Complementiser. And to realize its multi-function, *de* moves up forward stepwise, absorbed by the strong features of the head above. In this manner, *de* snowballs all the features and functions of the head position where it has the opportunity to stay.

3.3 A DM Approach to VP

- VP is a categoryless root with mere some general encyclopedic knowledge, which is first verbalized by light v, and then nominalized by light n
- In accordance with Decomposing Event Structure Model (Marantz 2008), initP, procP, resP introduce the correspondant arguments

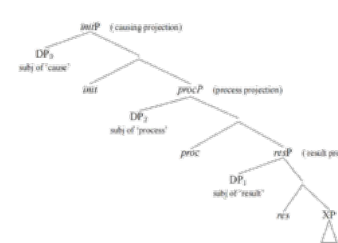


Figure-5 Decomposing Event Structure Model

3.4 A DP & PP Analysis

If *dui* DP *de* VP is analysed as a PP, *de* moves from light n to C, and DP is raised from [Spec, initP] to [Spec, TP]

If *dui* DP *de* VP is taken as a DP, *de* moves from light n to D, and DP is raised from [Spec, procP] to [Comp, PP]

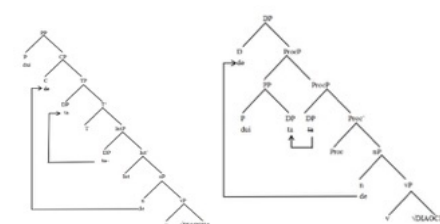


Figure-6 a. A PP Analysis b. A DP Analysis

4 Conclusion

Under the theoretical framework of Distributed Morphology (DM), this paper holds that *dui* DP *de* VP should be treated in two different ways:

- a *dui*-headed PP, with *dui* serving as the marker of topicalization in which *dui* is followed by a subject-predicate-structure-transformed CP--DP *de* VP
- a *de*-head DP, with *dui* functioning as a realisation of focalization which raises the object of VP into front of it

It also holds that VP, as a root, in *dui*- construction, has no specific category except for denoting a certain action state before entering the process of syntactic computation, until the light n, realized by *de*, assigns it the noun identity, which endows it a kind of referential nature.



Pragmatic Potentials of the Chinese Second-Person Pronoun 'Nin' (您) in Contemporary Sociolinguistic Contexts

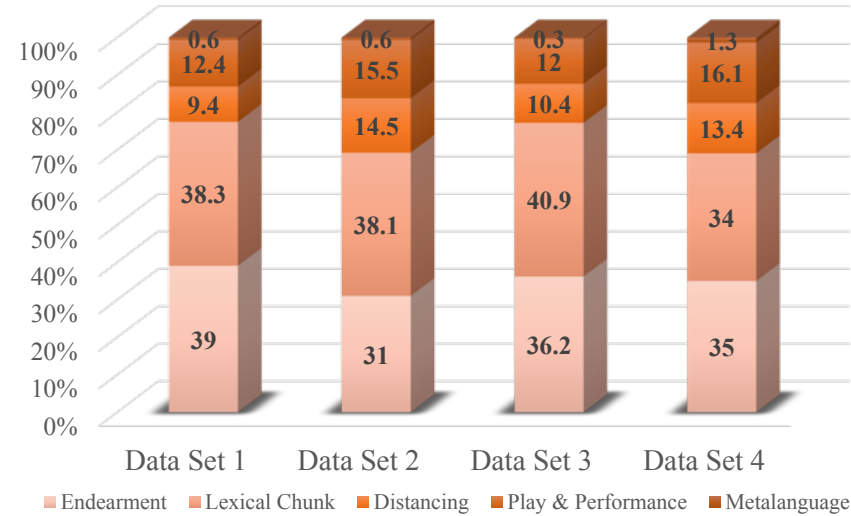
Chih-Cheng Wang (Roy). Applied Linguistics, School of Philosophy, Psychology, and Language Sciences

0. Questions

How do the second-person pronouns ni (你 & [你]) and nin (您 ; [您]) differ, and how is the latter used dynamically in everyday interaction? Where are interactional meanings (e.g. parody, sarcasm, etc.) from and how are they deployed? What can the alternation tell us about social life and stancetaking in Chinese?

3. Findings

Collocate	Rank	Freq (Scaled)	FreqLR	FreqL	FreqR	Likelihood	Effect
(Hello; ninhao)	1	7932	1450	1293	157	25.075	-0.165
(Thanks; xixie nin)	2	972	81	35	46	108.105	-1.298
(love you; ai nin)	3	732	59	31	28	85.621	-1.347
(Please)	4	96	42	7	35	26.330	1.094
What kind of (phone) plan do you (nin) use	7	24	16	6	10	23.835	1.702
(Are you; nin shi)	9	216	15	10	5	30.819	-1.562
Wish you happy birthday in advance;	10	18	13	9	4	22.232	1.817
(borrow your blessing; Jie nin ji yan)	12	156	6	3	3	36.979	-2.414
(Thank you-le)	13	132	5	2	3	31.577	-2.436



4. Discussion



The first-order indexical meaning of the second-person pronoun nin (**deference**) is bestowed with second-order indexical meanings (**endearment** and **distancing**) (Silverstein, 2003) that can be pragmatically mobilized to contradict the interlocutor or used in lexical chunks (Burdelski & Cook, 2012).

Speakers are able to agentively collapse the two central stances to engage in stylized performance or linguistic play (e.g. irony, joke, playful tone, etc.) in interaction.

1. Background

The semantics and syntactic structures of nin are detailed in previous studies (e.g. Xiang, 2019) but lack a coherent account, on which indexicality theory and stancetaking may shed some light. Stance is central to social life, and it even comes before interactional meanings (Ochs, 1990; Kiesling, 2009, 2022; Jaffe, 2016) and social identity or style (Coupland, 2007) in the indexical fields (Eckert, 2008)

2. Method BCC Corpus & AntConc



5. Takeaways & a Question

Pragmatic functions are mediated, and the social meanings of a lexical unit can go beyond one-to-one correspondence. Beyond structural and semantic correspondence, how is politeness, the traditionally assumed pragmatic function of nin, achieved and negotiated dynamically with body motion?

6. References

Anthony, L. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>

Burdelski, M., & Cook, H. (2012). Formulaic Language in Language Socialization. *Annual Review of Applied Linguistics*, 32, 173-188.

Cook, H.M. (2011). Language Socialization and Stance-Taking Practices. In A. Duranti, E. Ochs and B.B. Schieffelin (Eds.), *The Handbook of Language Socialization*. (pp.296-321). Wiley-Blackwell.

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge University Press.

Eckert, P. (2008). Variation and the indexical field. *Journal of sociolinguistics*, 12(4), 453-476.

Jaffe, A. (2016). Indexicality, stance and fields in sociolinguistics. In N. Coupland (Ed.), *Sociolinguistics: Theoretical Debates* (pp. 86-112). Cambridge University Press.

Kiesling, S. F. (2009). Style as Stance: Stance as the Explanation for Patterns of Sociolinguistic Variation. In A. Jaffe (Ed.), *Stance: Sociolinguistic Perspectives* (pp. 171-194). Oxford University Press.

_____. (2022). Stance and stancetaking. *Annual Review of Linguistics*, 8(1), 409-426.

Ochs, E. (1990). Indexicality and socialization. In J. Stigler, R. Schweder, & G. Herdt (Eds.), *Cultural Psychology: Essays on Comparative Human Development* (pp. 287-308). Cambridge University Press.

Silverstein, M. (2003). Indexical Order and the dialectics of sociolinguistic life. *Language and Communication*, 23(3-4), 193-229.

Xiang, X. (2019). Personal pronouns in Chinese discourse. In She, C. (Ed.) *The Routledge Handbook of Chinese Discourse Analysis*. Routledge.

Xun E.D., Rao, G.Q., Xiao, X.Y. & Zang, J.J. (2016). The construction of the BCC corpus in the age of big data. *Corpus Linguistics* (01), 93-109+118.





and the Fixation of Essentially Contested Concepts

Semantic Originalism

By **Tristan B. Borresen**
School of Philosophy, Psychology and Language Science
University of Edinburgh

Acknowledgements
I want to thank my two supervisors Sacha Bourgeois-Gironde and Alda Mari, and Institut Jean Nicod, PSL.

Purpose

To show that it is not possible to fix the meaning of essentially contested legal concepts in contrast to what semantic originalism maintains.

A Linguistic Fact

"Domestic violence" (in the 4th Amendment) meant originally "violence within the borders of a country", **not** conjugal abuse. We need to **fix our interpretation to its original meaning**.



I. Introduction

What are the requirements for a theory of constitutional interpretation?

- The source and grounds of legal decisions come from the **legal text**.
 - The ideal law is where the judge always **"translates"** the written text into its legal **"effect"**.
 - Therefore, first step to constrain and justify the course of legal action is to decide what the **law actually means**.
 - That is, a **hermeneutic framework** to understand the law.
 - The most important type of legal text is the **constitution**.
 - The constitution must have a **consistent meaning**; a meaning that can resist the **evolution** of time.
 - However, like all types of language, **the meaning of the constitutional text changes over time**.
- ” A theory of legal interpretation must provide an **optimally consistent interpretation** of the law despite the **drift of its meaning over time**.



II. Semantic Originalism

One theory obtain these requirements is Lawrence Solum's *semantic originalism*

- Semantic originalism posits that the original semantic meaning of the US Constitution is the **canonical** meaning.
- Lawrence Solum has formulated a powerful account on semantic originalism.
- In his account, the **original semantic meaning** of the Constitution amounts to



” [the] function of the **conventional** semantic meanings and patterns of usage that **prevailed** among **competent speakers** of the natural language at the time the texts were written.

The Fixation Thesis: the original meaning of the constitutional text is fixed at the time each provision is framed and ratified.

The Linguistic Fact Thesis: The original meaning of the Constitution is a linguistic fact; it is discovered.



III. Essentially Contested Concepts

However, some words in the US Constitution are vague, they are *essentially contested concepts*.

- Consider “freedom of speech,” “cruel and unusual punishment,” or “due process”.
- Solum explains that these vague terms from the Constitution are (i) abstract and (ii) essentially contested concepts (ECC).
- It is essentially impossible to settle an agreement of what they mean.
- However, in disputing, people do not talk past each other, « they agree to disagree. »
- They **agree** on the **concept** but have **different conceptions** of « freedom of speech ».
- Solum’s solution (2008) is this: we fix the original meaning of the Constitution to the **concept**, not to a particular **conception**.
- When a legal term is essentially contested, there is a **gap** between its **linguistic interpretation** and its **legal application**.

Interpretation	Restricted to the semantic meaning of the legal term.
Construction	Involves the normative aspects of the legal term

” ... **interpretation** leads to the **concept**, and then **construction** selects a particular **conception**.
Solum 2008 92



IV. Unfixable Contested Concepts?

Solum's Semantic Originalism does not obtain the requirements of legal hermeneutics

1. The Fixation Thesis offers a consistent meaning of the Constitution by fixing it to the original, public meaning.
2. If the meaning of the text is fixed, then the interpretation submitting to that meaning is constant and hence, consistent over time.
3. What enables consistency is that there is a fixed meaning that can be distinguished from other alternatives.
4. However, because the semantic meaning of essentially contested concept (ECC) is unsettled, it is not possible to distinguish one interpretation of an ECC from another.
5. Accordingly, any interpretation of an ECC is not consistent.
6. **As a result, The Fixation Thesis does not accommodate ECCs; they are not fixable.**

Conclusion

The price of **abstraction** in language is a loss of specifics. However fixing the meaning of a term T is to **specify** T given some **contextual restrictions**.
Therefore, Solum’s claim that the fixed interpretation of ECCs is abstract is **contradictory: No meaning can be fixed and abstract at the same time.**

Dialectal Variation in Voice Onset Time: Welsh-English and Midlands-English Dialects.

Rebekah Elliot

1. INTRODUCTION

Question: Do dialects with different language family origins have different Voice Onset Times?

Investigating the VOT's of voiceless plosives in East-Midlands-English dialects and Northern Welsh-English dialects:

/p/ /t/ /k/

2. BACKGROUND

Sociological features impacting VOT:

- Age of acquisition
- Language profile
- Gender

Docherty et. al. (2011):

Found differences in VOT for speakers along the Scottish-English border. Scottish dialects' VOT longer by ~10ms

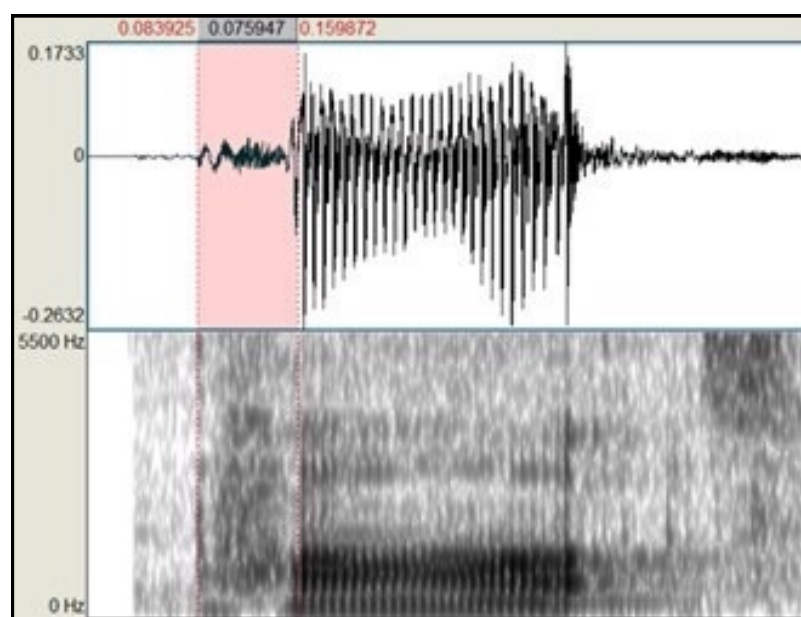
3. APPROACH

- 3 Welsh participants (North Wales)
- 4 English participants (East-Midlands)

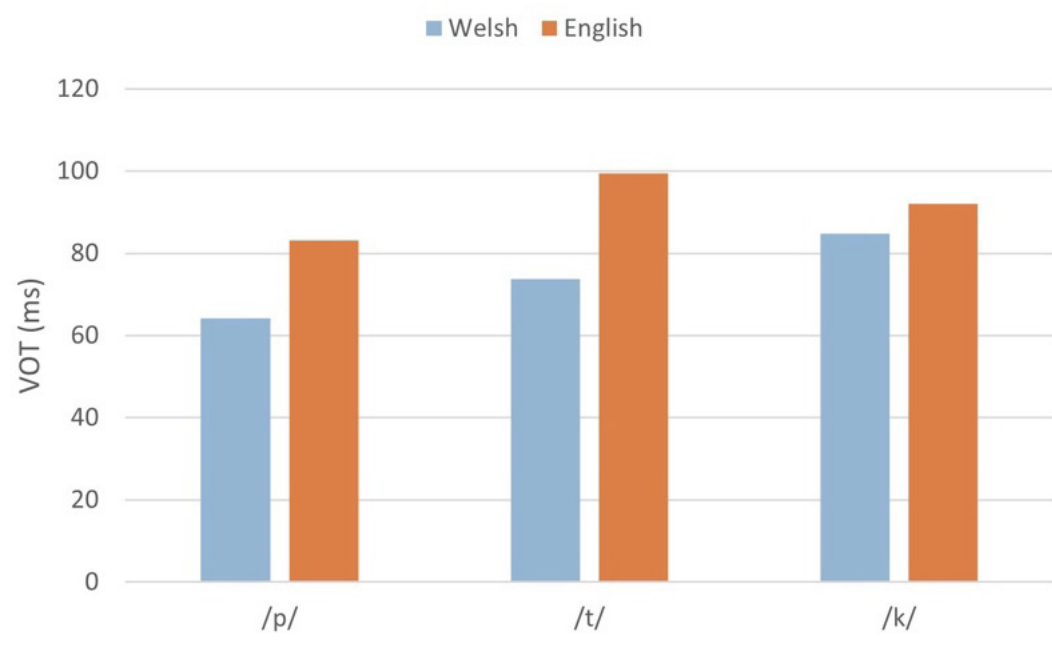
Recorded reading of word list

6 instances of each token = total of 18 tokens collected for each speaker

Audio analysed using Praat



Spectrogram of male English participant saying 'part' - highlighted /p/



4. RESULTS

/p/ - 19ms difference

/t/ - 26ms difference

/k/ - 8ms difference

Female Midlands-English speakers had longer VOT than male Midlands-English speakers

Male Welsh speakers had longer VOT than female Welsh speaker

5. DISCUSSION

- Preservation of phonological features from the Welsh language on the Welsh dialect?
- Presence of social influences on VOT (gender, age)
- In what contexts do aspects of dialectal transfer occur? (i.e. Docherty et. al. 2011)
- VOT as a regional variable?
- Is this difference perceptible? Does this phenomena occur in other dialects of English?

References:

- Docherty, G. J., Watt, D., Llamas, C. Hall, D. (2011). Variation In Voice Onset Time Along The Scottish-English Border. 17th International Congress of Phonetic Sciences, 591-594.
- Harada, T. (2007). The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. IRAL - International Review Of Applied Linguistics In Language Teaching, 45(4).
<https://doi.org/10.1515/iral.2007.015>
- Whiteside, S., P., Henry, L., Dobbin, R. (2004). Sex Differences In Voice Onset Time: A Developmental Study Of Phonetic Context Effects In British English. The Journal of the Acoustical Society of America, 116(2), 1179-83. doi: 10.1121/1.1768256.

When Speech Becomes Writing: The Case of Disfluencies

Aida Tarighat, Martin Corley

s2184935@ed.ac.uk | martin.corley@ed.ac.uk



For human readers, **written disfluencies** (e.g., **um**) render the text more **speech-like** which in turn opens it up to **less literal interpretation**.



literal fluent	I'm sure negative peer pressure leads to mostly idiotic decisions.
literal disfluent	I'm sure negative peer pressure leads to mostly um idiotic decisions.
non-literal fluent	I'm sure negative peer pressure leads to mostly clever decisions.
non-literal disfluent	I'm sure negative peer pressure leads to mostly um clever decisions.

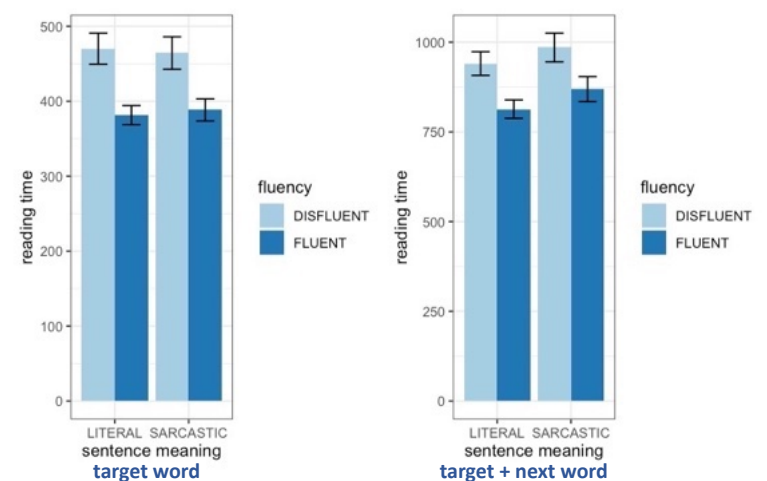


☹️ When preceded by **um**, sentences compatible with **non-literal** meaning are **not faster** to read.

😊 **Literal** sentences are **faster** to read than non-literal sentences.

😊 **Disfluent** sentences are **slower** to read than fluent sentences.

⌚ self-paced reading



99 participants, word-by-word self-paced reading of 24 items

⚙️ eye-tracking experiment

natural reading prosody not disrupted

48 counterbalanced pairs, same character count:

merry(literal)/feral (non-literal)
feral(literal)/merry (non-literal)

commas before and after **um**:

... be, um, merry when ...



The last few breaths of a dying language

A study of the Soo language relationship with its neighbours.

Hercules Singh Munda
MSc Linguistics (S2125642)

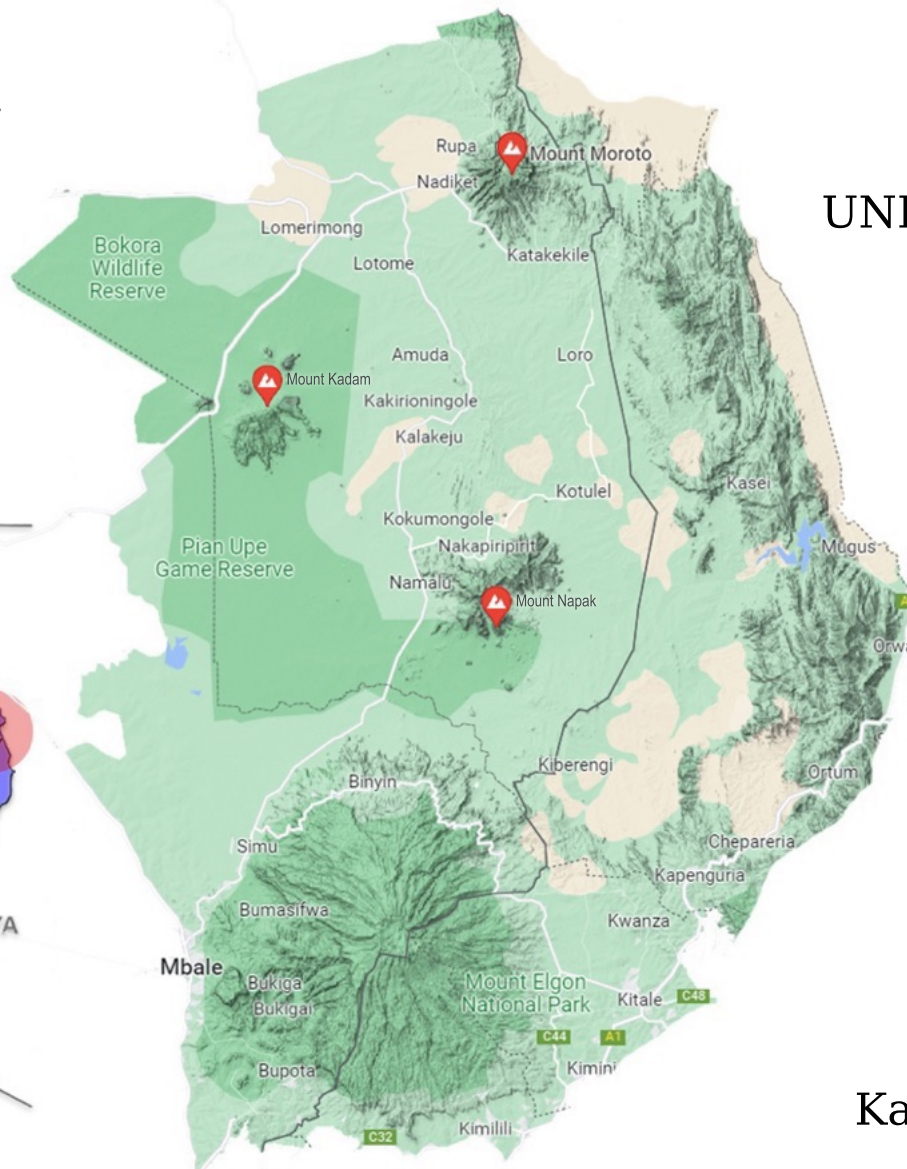
Soo Language

[ISO 3-639: teu]

//Soo (So/Tepes) is a
'Nilo-Saharan?' language.

//Soo along with Ik &
Nyangi, composes
the Kuliak language(s).

//Soo word order is VSO
(verb-subject-object).



Mount Moroto, Kadam & Napak

BACKGROUND

With less than 50 speakrs,
UNESCO has declared Soo as
Severely Endangered.

The Soo community
is spead across three
extinct volcanic mounition
of North-east Uganda.

Three major dialects:
Tepes (Mt. Moroto),
Katam (Mt. Kadam), and
Yog Tonji (Mt. Napak).

Soo speakers are
sorrounded by
Pokot (Southern Nilotic)
Turkana (Eastern Nilotic)
Karimojong (Eastern Nilotic)
communities.

THE PHONOLOGICAL INFLUENCE ON MORPHOLOGY

METHOD

1. Elicitation Session
2. Group Interview

DATA

9 Hour Audio (Raw)
Swadesh List & clan names
Historical narrative

PARTICIPANTS

12 elderly member with an average
age of 65 years (L:42,H:85)

//Soo community comprise of 36
clans based on animals, naturally
occurring phenomenons and
governance system.

*The data presented is an extension
of my first MA linguistics dissertation
and it is aimed to take on a detailed
phonological analysis of the Soo
language at a PhD level.

ANALYSIS*

1.Free variation - [c] & [k]
Speakers at Mount Kadam replaced
the -c- sound with -k- as follows:

Token: katecock (clan name)
Moroto: k-a-tɛ-cok
Kadam: k-a-tɛ-kok
Napak: ɲ-k-a-tɛ-cok

2 Affixation - [ɲ]

It was observed that speakers at
Mount Napak added -ɲ- sound to
every clan name, such as follows:

Token: Niparoto (clan name)
Moroto: ni-pa-ro-to
Kadam: ni-pa-ro-to
Napak: ɲi-pa-ro-to

SCOPE*

The properties of free variation and
affixation are so far only found in clan
names (i.e noun word class).

CONCLUSION

1. The data so far suggest that Soo
is influenced by it's neighbouring
language communities.

2. The speakers from Mt. Moroto
claim themselves as native.

3. Speakers at Mount Kadam &
Napak exhibit language variation
in terms of affixation*

4. With limited fluent speakers,
the Soo language is on part to
extinction & revitalization efforts.

REFERENCE

1. The Kuliak Languages of Eastern
Uganda. Heine, Bernd. 1976.
2. The So Language: A Grammar
Sketch. Beer, McKinney, Lokiru 2009.



A Preliminary Study of Tone Sandhi Pattern of Personal Names in Shanghai Dialect

BACKGROUND

In tonal languages like Shanghai dialect, tone sandhi is a well-known phenomenon that refers to the way in which the tone of its bearing unit changes in speech. Personal names, being a fundamental aspect of human identity, occur with high frequency in daily conversations. However, their sandhi pattern remains under-investigated, and thus poses a challenge for text-to-speech systems.

There are two tone sandhi patterns in the Shanghai dialect:

Left-Dominant pattern

The first syllable's tone influences the subsequent syllables in a sandhi chunk.

/34/ /23/ → [33 44]
ts^ho mi → ts^ho mi
'fry' 'noodle' 'fried noodles'
(nominal compound)

Right-Dominant pattern

The last syllable in a chunk remains unchanged, while the non-final syllable(s) of the sandhi domain neutralizing to some level tone.

/34/ /23/ → [33] [23]
ts^ho mi → ts^ho mi
'fry' 'noodle' '(to) fry noodles'
(verbal phrase)

Personal names exhibit both types of variations above. From pilot investigation of this thesis, native Shanghainese speakers would consider the names to be unnatural if pronounced in a way that deviates from the received pattern. This indicates that there should be some underlying constraints in personal name pronunciation. While previous studies have primarily focused on the sandhi patterns of actual words, the underlying constraints determining whether a name follows LD or RD remain unclear.

RESEARCH QUESTION

The thesis tries to build up a model to predict whether a name follows LD sandhi or RD sandhi. Based on the literature review, possible constraints come from:

- (A) Tone Categories and their Combinations
- (B) Presence / Absence of Zero-Onset
- (C) Frequency of the name's mention

The effect of (A) is examined in both di- and tri-syllabic compounds, while (B) is examined in di-syllabic names, and (C) in tri-syllabic names.

SUBJECT

This study recruited 30 subjects, born between 1965 and 1977.

MATERIAL

To eliminate the influence of phoneme level and experimental material sequence, all names, except for the familiar ones, were randomly generated. All reading materials were generated, shuffled, and managed with Python scripts. Every subject read 380 names one by one, including 190 randomly generated di-syllabic names, 125 random tri-syllabic names, and 65 tri-syllabic names of well-known people. Di- and tri-syllabic names appeared alternately to avoid the possible priming effect.

ANALYSIS

All independent variables here are categorical variables. The response variable is binary (sandhi pattern being either LD or RD). Hence, we assume a logistic regression model and random forest classifier for data analysis. We one-hot encode all categorical variables, remove missing values. We train the Random Forest classifier on a 70 percent split of our dataset, and test the model on the rest 30 percent.

RESULTS

Di-syllabic Names

The majority (86.87%) of di-syllabic names follows the Left-Dominant sandhi pattern. RD occur in specific tone combinations. Combinations with the presence of the falling tone (T1) usually signifies a deviation from the prevailing Left-Dominant pattern in disyllabic names.

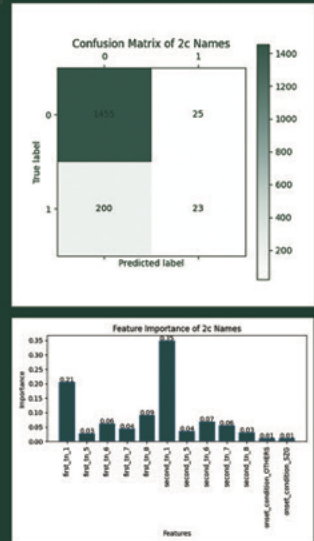
Tone Combination		Logistic Regression Coefficient				Random Forest Performance			
	RD%		Estimate	Std. Error	z value	p value			
71	48.60%	(Intercept)	0.92904	0.13792	6.736	1.63e-11 ***	0	0.88	0.98
61	44.67%	tone type (σ1)	-0.221348	0.01635	-13.54	< 2e-16 ***	1	0.48	0.1
51	41.36%	tone type (σ2)	-0.271065	0.0157	-17.26	< 2e-16 ***		accuracy	0.68
16	40.07%	onset condition (SZG)	-0.2276491	0.13636	-2.028	0.0426 *		macro avg	0.54
18	26.71%	onset condition (ZSG)	0.029078	0.12971	0.224	0.8226		weighed avg	0.83
17	22.37%	onset condition (ZSZG)	-0.2201	0.13591	-1.619	0.1053			
15	16.43%	subject	-0.021563	0.00323	-6.687	2.28e-11 ***			
55	12.95%								
65	10.00%								
11	8.33%								
75	7.87%								
76	6.21%								
56	4.19%								
81	3.93%								
78	2.78%								
66	1.67%								
85	1.67%								
58	1.06%								
77	0.56%								
86	0.56%								
87	0.56%								
57	0.53%								
68	0.33%								
67	0.00%								
88	0.00%								

Logistic Regression

The logistic regression model is a good fit for the di-syllabic data. The predictor variables "Tone Type(σ1)", "Tone Type(σ2)", "Onset Condition (SZG)", and "Subject" are statistically significant predictors.

Random Forest

The results of LR and RF support the feature combination Tone Type(σ1), Tone Type(σ2), Onset Condition (SZG/OTHERS). T1 exhibits a higher significance across all features.





- Spanish (adverb/preposition + subordinator marker *que*)

hasta que “until”, *sin que* “without”, *aunque* “although”

- German *als* “as, when, than”

[XP al [CP [C° so [...]]]] > [CP [C° also [...

Old High German:

- a. *bi namen uuéiz ich thih ál só man sinan drút scal*
by name know I you.ACC all so one his master shall
“I know you well by name, as one should know his master”

(Otfrid V, 8, 38)

- b. *Niman en was alse gut alse iob*
No-one NEG was as good as Job
“Nobody was as good as Job”

(Lil 4,3)



得天时者矣然而不胜者，
de tian - shi zhe yi ran er bu sheng zhe,
get good-time person PAR pronoun CONJ not win person,

是天时不如地利也。 (Mencius)
shi tian - shi bu ru di li ye
is good-time not equal(to) geographic benefits PAR

“They are given by Heaven the opportunity of time, and in such case their failure is because opportunities of time vouchsafed by Heaven are not equal to advantages of situation afforded by the Earth.”

然而

ran (pronoun) + *er* (conjunction) —> *ran-er* “however” (conjunction)



- 莫不， 莫非， 无不， 无非
- 以期， 以便， 以致， 以至



-
-

-

-

-
-

-

-

- a. [DP (al) die [N wile [CP daz/und/so]]] → [DP die [N wile [CP ∅ ...]]]
b. [DP die [N wile [CP ∅ ...]]] → [DP d' [N wile [CP ∅ ...]]]
c. [DP d' [N wile [CP ∅ ...]]] → [CP wile ...]

-

-



- 君 仁， 莫 不 仁。 Mencius 7 [Warring States]
Jun ren, mo bu ren.
ruler benevolent none not benevolent
“If the ruler is benevolent, then all will be benevolent.”

-

- 民 莫 不 知。 Zuo Zhuan [Warring States] 468 BC-300 BC
Min mo bu zhi.
people none not know
“All people know this.”

- 天下莫不代王怀怒。 Shiji [Western Han]
tian xia mo bu dai wang huai nu
world none not replace king carry anger
“All the people feel the king’s anger.”

-

- 四方正 远近莫不皆正也。 Qian Han Ji [Eastern Han]
si fang zheng yuan jin mo bu jie zheng ye
world righteous far near none-not all righteous PAR
“If the world is righteous, nations near and far all would be righteous.”

- 凡物莫不以适为得，以足为至。 Yi Wen Lei Ju [Tang] 624
fan wu mobu yi shi wei de yi zu wei zhi
all thing all use suitability as gain use sufficiency as perfection
“Things all regard suitability as gain, sufficiency as perfection.”

- 喏早晚不来，莫不又是谎么？ Xi Xiang Ji [Yuan] 1300
nuo zao wan bu lai mobu you shi huang me
ModP early late not come maybe again is lie SFP
“He doesn’t come sooner or later, maybe isn’t it another lie?”

-
-



Arbëreshë NPs at Syntax-Pragmatics Interface

(Jessica Rita Messina – University of Padova)

jessicarita.messina@studenti.unipd.it

1. Introduction

Piana degli Albanesi (Palermo, Sicily) is the main **italo-albanian community** in Sicily and one of the biggest and most active arbëreshe enclave in Italy. For over five centuries this Arbëreshe variety has been living in **close contact** with the Italo-Romance local dialects and, only recently, with Italian. Nowadays, Piana is a **multilingual setting**: speakers are bi-/trilingual and they can easily switch between the three codes of their own linguistic repertoire (Arbëreshe/Italian/Sicilian) in their daily social interactions. The intimate contact with the Romance varieties led to phenomena of **code-switching**, loan, **convergence** and **hybridization**, involving all the linguistic domains. Like the other arbëreshe varieties, Piana's variety is characterized by an extensive **Romance relexification** in conditions of arbëreshe-sicilian dialect bilingualism and code-switching between dialect, regional Italian and standard Italian (Savoia 2008).

2. Morphosyntax of albanian DPs

- Distinction Indefinite/Definite
Një djalë / djal-i 'a boy/the boy'
Art boy / boy-Def.Nom.ms
- Case+Definiteness
Një djalë / djal-i-n
Art boy-Acc/ boy-Def.Acc.ms
- Complementary Distribution between **Definite Inflection and Determiners**
Një vajzë/Ajo vajzë/vajz-a 'a girl/this girl/the girl'
Art girl / Dem girl / girl-Def.Nom.fs
- Definiteness vs Indefiniteness in Nominal modification (pre-adjectival/ pre-genitival Linkers).

3. Morphosyntax of arbëreshë DPs

Parameters a, b and c:

- (1) Një skolinë** 'a tie'
Një monopatinë 'a scooter'
Një felpa / 'na felpa 'a sweatshirt'
 - (2) Skolin-a** 'the tie'
Monopatinu /Ai monopatinë /
'u monopatinu / stu monopatinu 'the scooter'
felp-a / 'a felpa 'the sweatshirt'
 - (3) Marku ble një telefonë / 'u telefonu / stu telefonu**
'Mark buys a mobile phone/ the mobile phone'
 - (4) Marieja i bun 'na/una torta Lukës**
'Maria bakes a cake for Luca'
- **Indefinite Inflection**: -Indefinite Article **një**; -zero-marked N (arbëreshë and morphologically adapted Romance loans); - Non-adapted Romance N preceded by **një** and (occasionally) by Romance indefinite article ('**na** 'a').
- **Definite Inflection**: alongside arbëreshe finite morphology, also used for regularly morphologically integrated borrowed Ns, there is an alternative structure: Romance D (article and demonstrative) + Romance N (**një makina / 'a makina** 'a car' / 'the car')
- Within sentences: -Indefinite NPs in object position present arbëreshë structure Indef. Art.+N(uninflected); - Definite NPs in object position present Romance structure Art/Dem+N(Romance).

4. Convergence between Arbëresh and Romance

At the sentence level, it is observable an increasing use of Romance D, already used as an alternative in NPs → Speakers seem to prefer this structure in those contexts in which N is a Romance borrowing → **Romance D** are increasingly selected as a tool of **Definiteness** lexicalization.

Të bij-a-t /bij-a-t 'Le figlie/alcune figlie'

Art daughter-Nom.pl/daughter-Nom.pl

Ai ë i burrë çë më thërret sempri

This is Lnk man Comp Pron-Dat call always

'This is the man who always calls me'

→ Speakers start to associate **Definiteness** to a **proclitic morpheme** in pre-nominal position → this new definiteness properties distribution can be taken as an indicator of **language variation**, probably due to language contact.

5. DPs at the Syntax-Pragmatics Interface

I- The **Syntax-Discourse/Pragmatics Interface** is a *locus* of **Optionality** and Instability in bilingual grammar (Tsimpli & Sorace 2006; Sorace & Serratrice 2009). Although Articles are at the crossroads between all linguistic domains, their use depends on **Pragmatics factors** → a **Definite/ Specific DP** refers to a familiar NP, already implicitly or explicitly introduced in speech (**Common ground**); an Indefinite/unspecific DP can introduce a new NP (Heim 1982).

→ **Finite morphology** and **determiners** are 'anchor' points of the sentence into discourse (referentiality) and their optionality in early speech originates from the possibility of underspecifying functional heads as I and D in early grammar (cfr. Schaeffer 1994).

→ **Underspecified D** needs to be semantically interpreted via a Pragmatical assignment of a default "Familiar" value (deictically assigned): deictic interpretation is unavailable in the adult grammar.

II- **Concept of Non-Shared Assumptions** (pragmatic)(CNSA): Speaker and hearer assumptions are always independent.

→ As long as the CNSA fails to apply, children acquiring any language will over-generate utterances containing items which involve the Common Ground (Articles, Pronouns, Cleft...). (Schaeffer & Matthewson 2005): **D omission** and **D over-extension** in Early Language as a consequence of Underspecification.

III- **The Underspecification account** (Belletti et al. 2007) applies in bilingual contexts when one of the languages has a complex settings due to syntax-pragmatics interface conditions (Arbëreshe) and the other does not (Italo-Romance) → **Underspecification of the interpretable feature** [+Def].

Monolingual arbëreshe (one-to-one mapping):

POST-N ART ⇔ [+Def]

PRE-N ART ⇔ [-Def]

Attrited L1 arbëreshe grammar

POST-N ART ⇔ [+Def]

PRE-N ART ⇔ [-Def]

PRE-N ART ⇔ [+Def]

• The shift to the adult grammar involves a restructuring of the mapping between grammar and pragmatics: in the resulting attrited arbëreshë grammar the feature [+def] is mapped into two different D-structures (makin-a / 'a makina 'the car').

6. Conclusions

The **Syntax-discourse interface complexity** in terms of cognitive load to process in real-time communication, the **Underspecification** affecting Early Grammar leading to an ambiguous D usage (D Over-extension and D omission), related to an immature pragmatic system (lack of CNSA), and the **Language influence** exercised by the language with the less restrictive option (Italo-Romance varieties) should lead to a **morpho-syntactic reorganization** within **arbëreshë DPs**, concerning a new lexicalization of Definiteness.

Selected References:

- Belletti, A., Bennati, E. & Sorace, A. (2007) *Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian*, *Nat Language Linguistic Theory* 25, 657–689.
- Matranga, V. (2019) *Arbëreshë*, Versione 2, In Thomas Krefeld & Roland Bauer (a cura di) "Lo spazio comunicativo dell'Italia e delle varietà italiane", Versione 67 In *Korpus im Text*.
- Schaeffer, J., & Matthewson, L. (2005). *Grammar and Pragmatics in the Acquisition of Article Systems*, *Natural Language & Linguistic Theory*, 23, 53-101.
- Sorace, A. (2011) *Pinning down the concept of "interface" in bilingualism*, *Linguistic Approaches to Bilingualism*, 1, 1-33.
- Tsimpli, I.M. & Sorace, A. (2006) *Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena*, In *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 653-664). Somerville, MA: Cascadia Press.
- Savoia, L. M. (2008), *Studi sulle varietà Arbëreshe*, Università della Calabria, Dipartimento di Linguistica, Sezione di Albanologia.