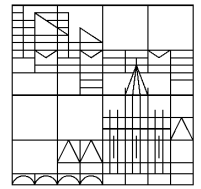
European Research Council
Established by the European Commission

Universität Konstanz



Sociolinguistic Typology Meets Historical Corpus Linguistics

Molly Rolf and Raquel Montero Estebaranz

3rd June 2024

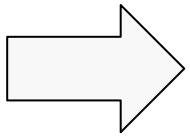
STARFISH

SOCIOLINGUISTIC TYPOLOGY AND RESPONSIVE FEATURES IN SYNTACTIC HISTORY

A question

German and English are sisters; both are part of the Germanic language family, yet English has lost its gender and case systems whilst German has not. Why?

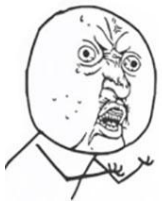
	gender and number					
	Masculine		Neuter		Feminine	
	Singular	Plural	Singular	Plural	Singular	Plural
case	**Strong Noun Declension**					
	engel 'angel'		scip 'ship'		sorg 'sorrow'	
Nominative	engel	englas	scip	scipu	sorg	sorga
Accusative	engel	englas	scip	scipu	sorge	sorga/sorge
Genitive	engles	engla	scipes	scipa	sorge	sorga
Dative	engle	englum	scipe	scipum	sorge	sorgum
case	**Weak Noun Declension**					
	nama 'name'		ēage 'eye'		tunge 'tongue'	
Nominative	nama	naman	ēage	ēagan	tunge	tungan
Accusative	naman	naman	ēage	ēagan	tungan	tungan
Genitive	naman	namena	ēagan	ēagena	tungan	tungena
Dative	naman	namum	ēagan	ēagum	tungan	tungum



Singular	Plural
angel	angels

German vs English

du	-	you
dich	-	you
dir	-	you
Sie	-	you
Ihnen	-	you
ihr	-	you
euch	-	you



Me during my french exam wondering what is the gender of a washing machine

Why has English lost its gender and case systems whilst German has not?

Project STARFISH

S- sociolinguistic
T- typology
A- and
R- responsive
F- features
I- in
S- syntactic
H-history

Investigating the role of so-called responsive, L2- difficult morphosyntactic features in language change.

Trudgill's (2011) Sociolinguistic Typology

Different types of contact situation -> different types of change:

- Adult L2 language acquisition leads to **simplification**.
- Child bilingualism leads to **complexification**.

Simplification **occurs when a society contains a high proportion of post-critical-period acquirers – "a result of the lousy language-learning abilities of the human adult".**

Walkden & Breitbarth (2019): uninterpretable features

Following Trudgill, Walkden & Breitbarth (2019) equate complexity with L2 difficulty.

Interpretability Hypothesis (e.g. Tsimpli & Dimitrakopoulou 2007): uninterpretable features are not accessible to adult L2 learners as part of the initial state.

In Minimalist syntactic theory, **uninterpretable** features:

- are present only within the syntax
- have no interpretation at the interfaces (i.e. no semantic content)

All else being equal, we predict that in sociohistorical situations in which adult L2 learners are dominant, uninterpretable features will typically be lost.

The WHAT, WHERE, WHEN, WHO and HOW?

The questions needing answers in our investigation into a case of language change:

WHAT the linguistic variants are
WHO is in contact, the people involved
WHERE the contact took place
WHEN the contact took place
HOW the contact took place- at the individual level and at the population level

The historical corpus approach

We make use of the historical record to examine the nature of morphosyntactic simplification.

Advantages of this method:

- using diachronic corpora provides the WHERE and WHEN answers:

-we expect more simplification in areas with higher numbers of adult L2 speakers
-when the change takes place, and if the change is short-term or long-term

- ability to track the speed and behaviour of a morphosyntactic change quantitatively.

Roadmap

Part I: Testing Trudgill's Sociolinguistic Typology Using Corpora

- Case study: Null subjects in Latin American Spanish
- Case study: Morphological case loss in Balkan Slavic

Question period- about 10 mins

Part II: Combining Methodologies

- Population dynamics: modelling language change and contact
- Case study: subjunctive morphology in Spanish

Question period- about 10 mins

Part I.
Testing Trudgill's Sociolinguistic Typology Using Corpora

Structure of Part I

1. Case study: Null subjects in Latin American Spanish

- Null subjects as an L2 difficult feature
- Contact and research question
- Corpus and results

2. Case study: Morphological case loss in Balkan Slavic

- Morphological case as an L2 difficult feature
- Contact in Balkan Slavic
- A nanosyntactic KP and case loss predictions
- Corpus and results

1. Case study: Null subjects in Latin American Spanish

Case study: Null subjects in Latin American Spanish

WHAT: Null subjects

Spanish is a null subject language (NSL)

Spanish [consistent NSL]: (**Nosotros**) queremos ir a la playa

English [non-NSL (NNSL)]: *(**We**) want to go to the beach

- **However, overt pronoun expression is higher in certain varieties of Latin American Spanish (LAS):**

Ellos me dijeron que ***yo*** tenía anemia . . . Si ***ellos*** me dicen que ***yo*** estoy en peligro cuando ***ellos*** me entren la aguja por el ombligo, ***yo*** me voy a ver en una situación de estrés.

> 'They told me that I had anemia . . . If they tell me that I am in danger when they put the needle in my belly-button, I am going to find myself in a stressful situation.' (Toribio, 2000:319, ex. 3e).

Null subjects are L2 difficult

- Null subjects seem to be harder to acquire, particularly for L2 speakers (Bini 1993, Pérez-Leroux & Glass 1999, Margaza & Bel 2006, Sorace 2011, Tsimpli & Lavidas 2019).
- Increasing the use of overt pronouns seems to be an act of simplification then.

The uninterpretable features on null subjects lead to misanalysis & overcompensation with features that are interpretable: overt subjects.

Contact

- The countries with grey have large Afro-Hispanic populations descended from the slaves taken to the Caribbean and coastal South America.
- Their languages reflect the same kind of change predicted- null subject loss.

Contact and research question

- Specifically, during the colonial period enslaved Africans were brought over to Latin America.
- These adult learners of L2 Spanish might have struggled acquiring the L2-difficult null subject system, preferring overt pronouns.
- Their children would then have nativized this system.
  - This is exactly the scenario Sessarego (2013) proposes for Latin American Spanish where Afro-Hispanic Language of the Americans (AHLAs) are these nativized varieties, showing many non-standard features such as SV order.

Main research question:

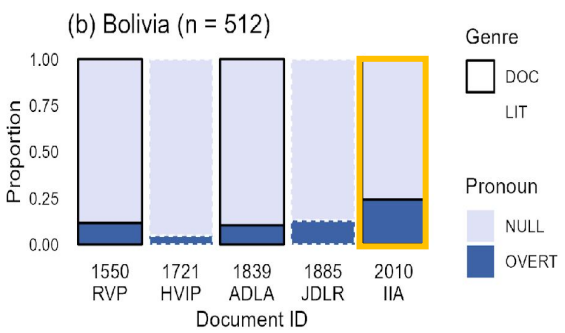
- **Does overtness increase diachronically?**

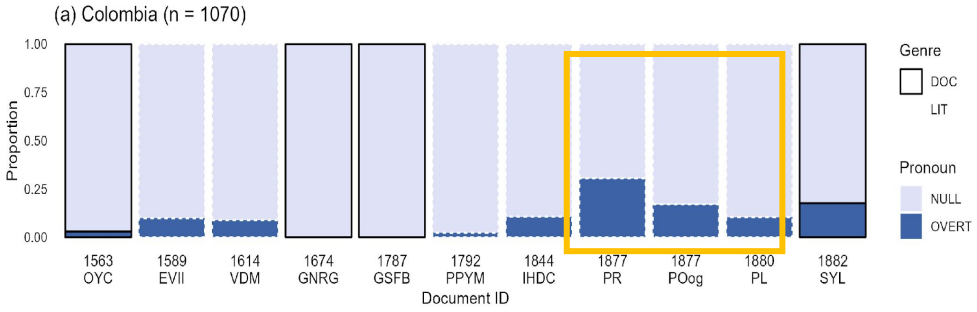
Corpus: CORDELES

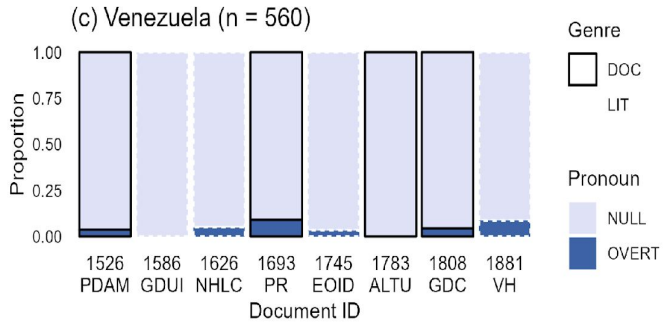
- **This is the main historical corpus**
  - 57 texts (~2-3k words each)
  - 8 countries (7 from the Caribbean and Central and South America, plus Spain as a control)
  - 16th-19th centuries
- **Each century + country combination ideally has 2 texts, one from each genre:**
  - Literature (e.g. novels, plays, poetry)
  - Documents (e.g. newspapers, legal documents, letters)
- **5 supplemental AHLA texts**

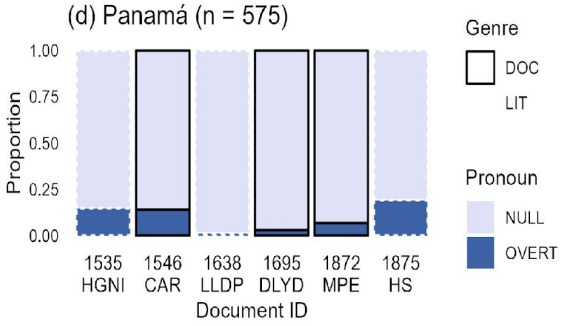
	CARIBBEAN/CENTRAL			SOUTH AMERICAN				SPAIN
	DR	PANAMÁ	CUBA	PERÚ	COLOMBIA	BOLIVIA	VENEZUELA	
16TH								
LIT	ENT	*HGNI*	*HDLI*	*HNMI*	*EVII**	--	*GDUI*	LAH
DOC	SDJ	*CAR*	*DRF*	*NDP*	OYC	*RVP*	PDAM	CAN
17TH								
LIT	*DPHJ*	LLDP*	*EDP**	*CEVP**	*VDM*	--	*NHLC*	DQ
DOC	--	DLYD	LCDH	CPVV	*GNRG*	--	PR	ACRA
18TH								
LIT	**LIVIE**	--	PJFC*	PAD	PPYM	HVIP	*EOID*	ARJD
DOC	ASD	--	SPPH	MC	GSFB	--	ALTU	EAU
19TH								
LIT	**GAL***	HS*	**ADUE**	**MYT**	**IHDC**	JDLR	VH	CPC
DOC	**ALD**	MPE	GDLH	**CRP**	**SYL**	ADLA	GDC	QDEV

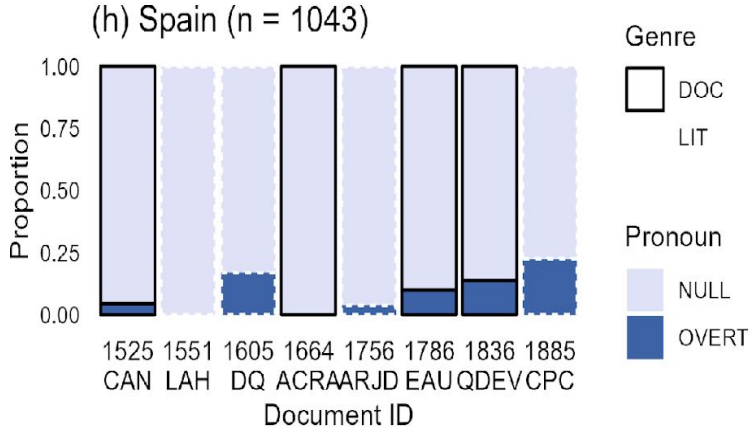
Results: Pronoun realisation

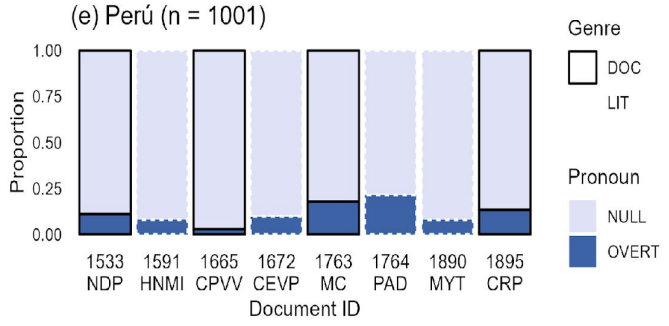


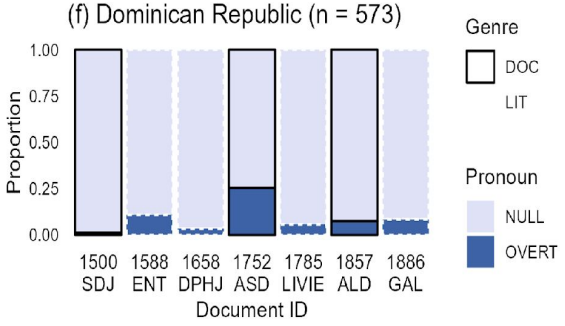


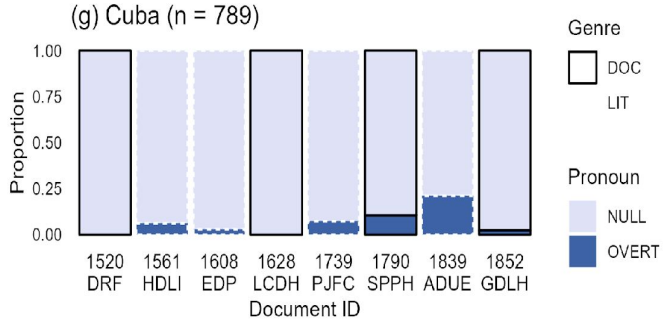








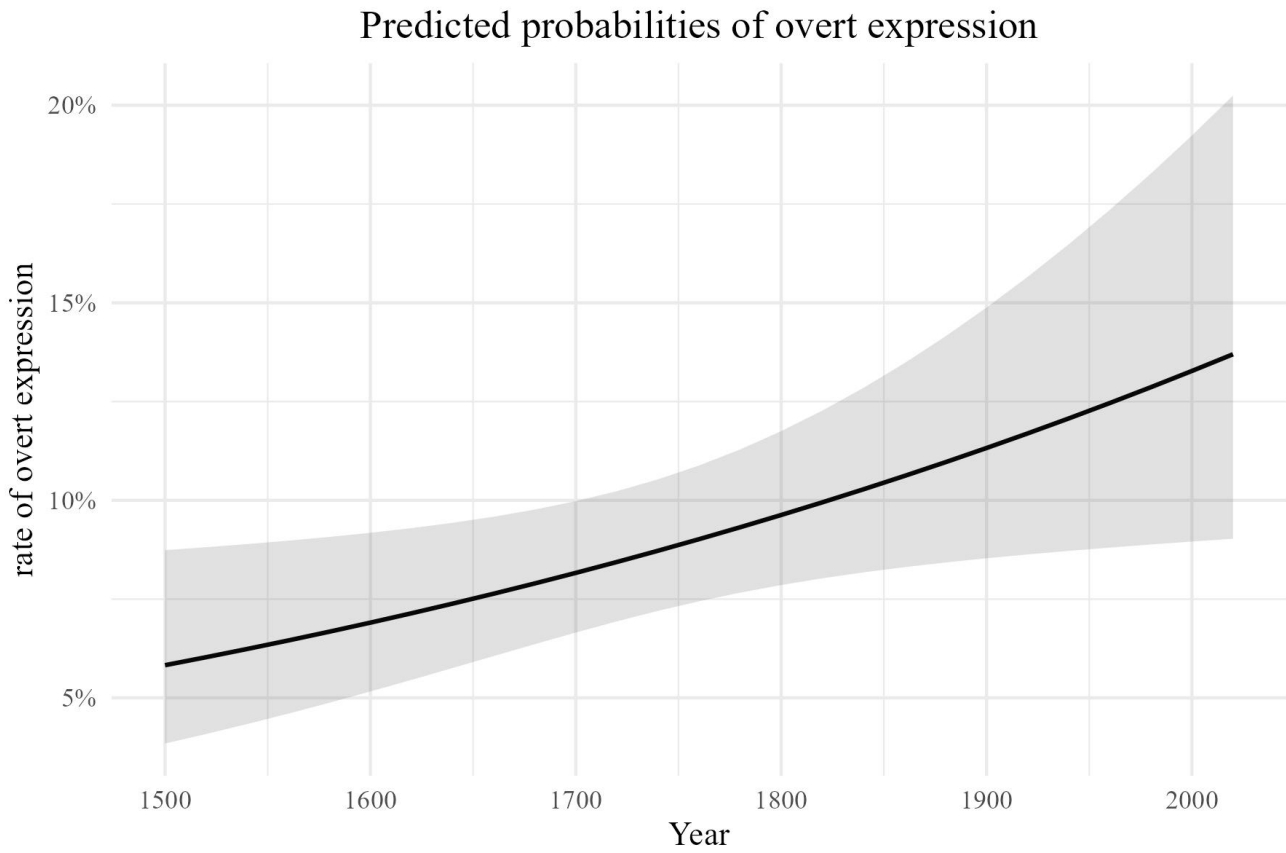




There also seems to be something else at play as the most overt texts are also very discourse heavy: orality

Model results

	Dependent variable:
	SPE
Year	0.247
	(0.108)
	p = 0.022*
ORSCORE	0.457
	(0.138)
	p = 0.001***
Constant	−2.360
	(0.107)
	p = 0.000***
Observations	5,707
Log Likelihood	−1,911.580
Akaike Inf. Crit.	3,831.160
Bayesian Inf. Crit.	3,857.758
Note:	*p<0.05; **p<0.01; ***p<0.001



2. Case study: Morphological case loss in Balkan Slavic

Case study: Morphological case loss in Balkan Slavic

WHAT: Morphological case

- Syntactic/structural Case is associated with uninterpretability in Minimalism
- Semantic case (inherent/lexical) perhaps with interpretability

- The function of uninterpretable Case features e.g. NOM and ACC are to determine positioning and behaviour of the subject and object, whereas other morphological cases seem more determined by semantics.

- I focus on discussion of morphological case in this talk: if we apply uninterpretability we would predict that the nominative and accusative cases are the most L2 difficult, which does not intuitively seem to be true...

Case study: Morphological case loss in Balkan Slavic

CASE	man, sg.	man, pl.
NOM	adam	adam-lar
ACC	adam-ı	adam-lar-ı
GEN	adam-ın	adam-lar-ın
DAT	adam-a	adam-lar-a
LOC	adam-da	adam-lar-da
ABL	adam-dan	adam-lar-dan
INS	adam-la	adam-lar-la

Case	Singular			Plural
	Feminine	Masculine	Neuter	All genders
Nominative	die	der	das	die
Accusative	die	den	das	die
Dative	der	dem	dem	den
Genitive	der	des	des	der

Case study: Morphological case loss in Balkan Slavic

	CASE	man, sg.	man, pl.
man (subject)	NOM	adam	adam-lar
man (object)	ACC	adam-ı	adam-lar-ı
man**'s** / **of** the man	GEN	adam-ın	adam-lar-ın
to the man	DAT	adam-a	adam-lar-a
in the man	LOC	adam-da	adam-lar-da
from the man	ABL	adam-dan	adam-lar-dan
with the man	INS	adam-la	adam-lar-la

- A shift in functional load from case morphology to adpositions.

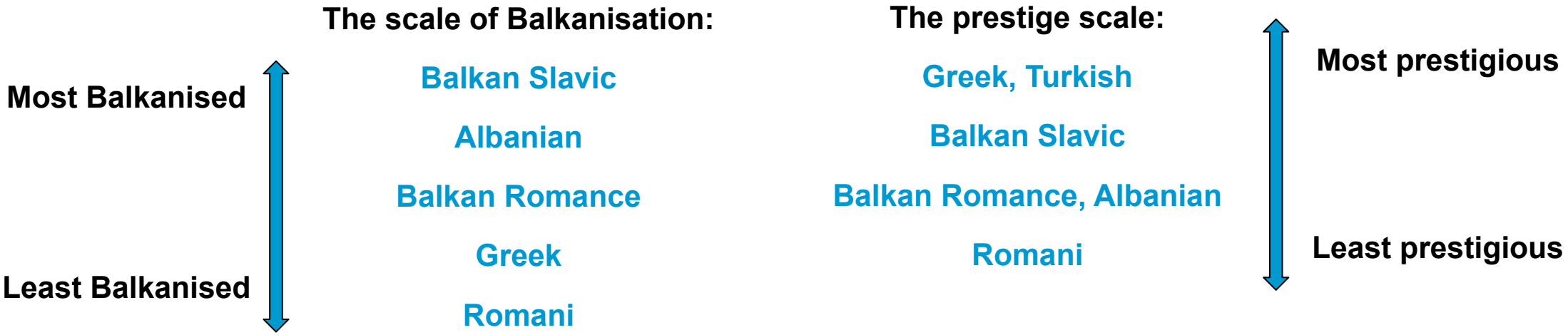
Case study: Morphological case loss in Balkan Slavic

Case systems of the Balkan Sprachbund languages:

- Greek- has retained some case: nom, acc and gen (from a previous 5-way case system)
- Albanian- has retained a four-way case distinction (nom, acc, gen-dat and abl)
- Romanian- the only Romance language to have retained case
- Romani- is the least impoverished case-wise with nom, acc/oblique, voc, gen, dat, loc, abl and instr
- **Bulgarian and Macedonian- nearly completely lost**

Contact in the Balkan Sprachbund

Lindstedt (2000, 2019) argues that an interplay of sociolinguistic factors in the Balkans could have led to the simplification of the case system in Balkan Slavic:

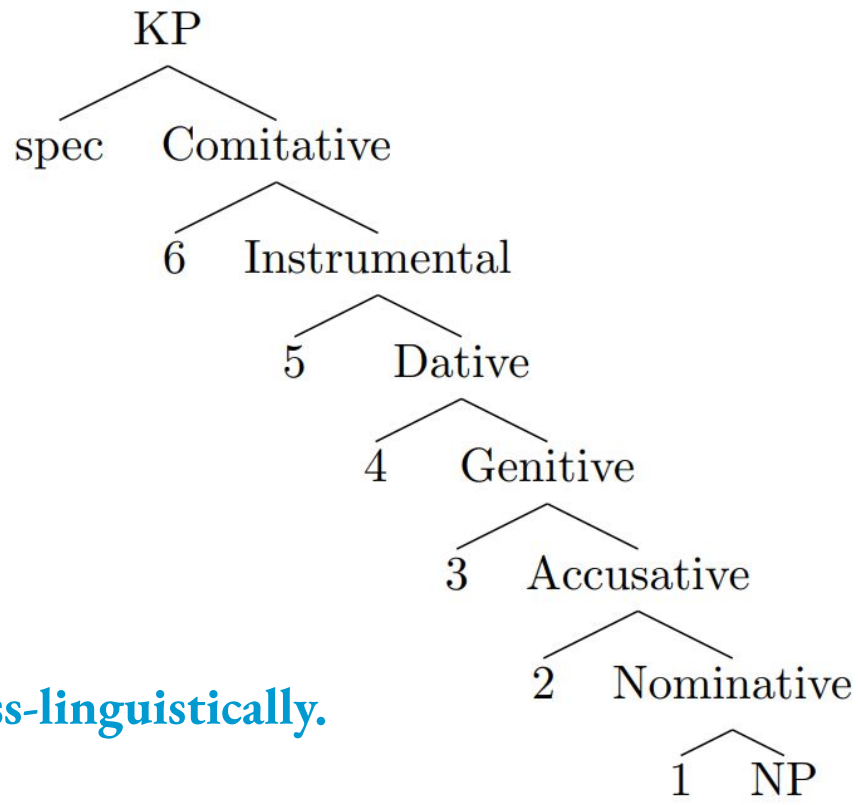


- Greek was at the top of the prestige scale and therefore had fewer multilingual L1 speakers but many L2 speakers.
- At the other end of the scale, Romani had the fewest L2 speakers but a very multilingual L1 population.
- Balkan Slavic fell somewhere in the middle of the prestige scale with many L2 speakers AND a high level of multilingualism.

Blake's Hierarchy and the KP (case phrase)

Blake's Hierarchy: NOM > ACC/ERG > GEN > DAT > LOC > ABL/INS > other

- A nanosyntactic approach to case: the cases in the KP (Caha, 2009) are in containment relationships: e.g. the dative contains the genitive and all the cases lower than it.
- If a language has a given case, it must also have all the cases to the left of the given case in the case hierarchy/sequence. In terms of syncretism, two or more cases can only be syncretic if adjacent on the case sequence.



The KP aims to capture case inventories and syncretisms cross-linguistically.

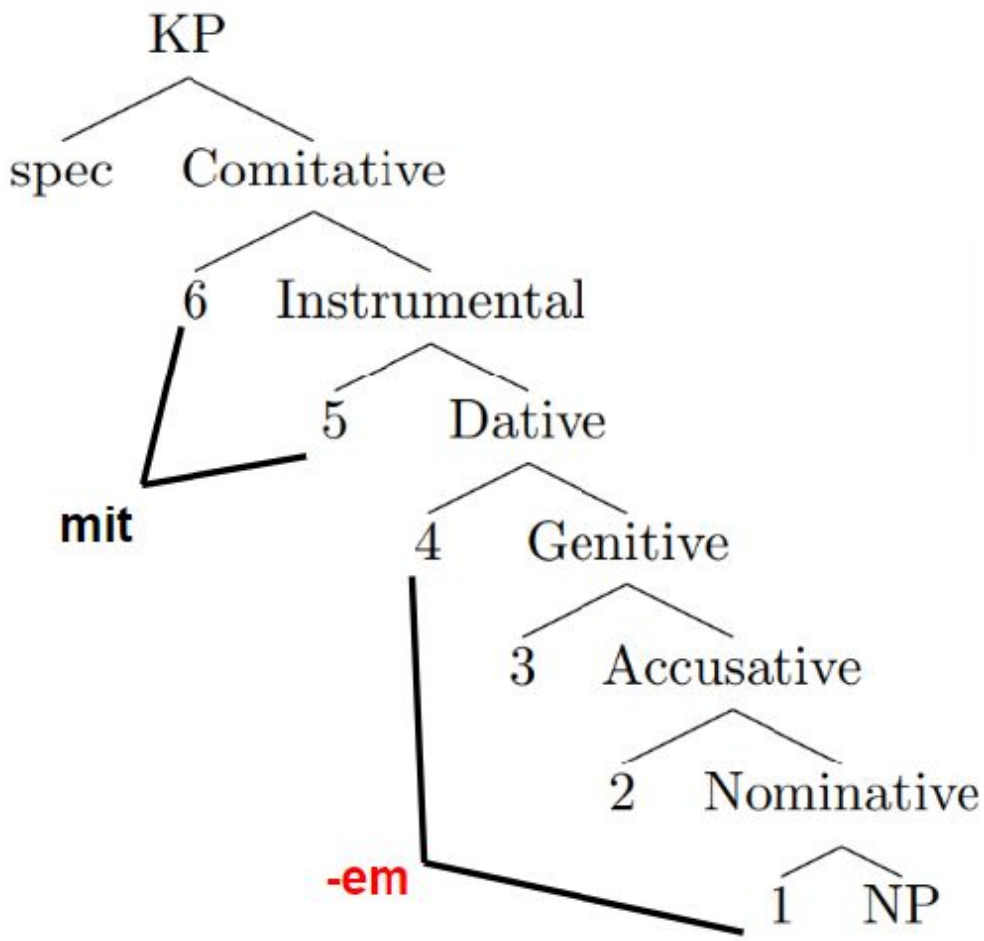
Case Phrase KP- adpositions too!

The same syntactic structure can be applied to adpositions.

'I am eating dinner with a friend.' Comitative

Adpositions taking certains cases on their nominals are also accounted for.

'Ich esse Abendessen mit einem Freund.' Comitative

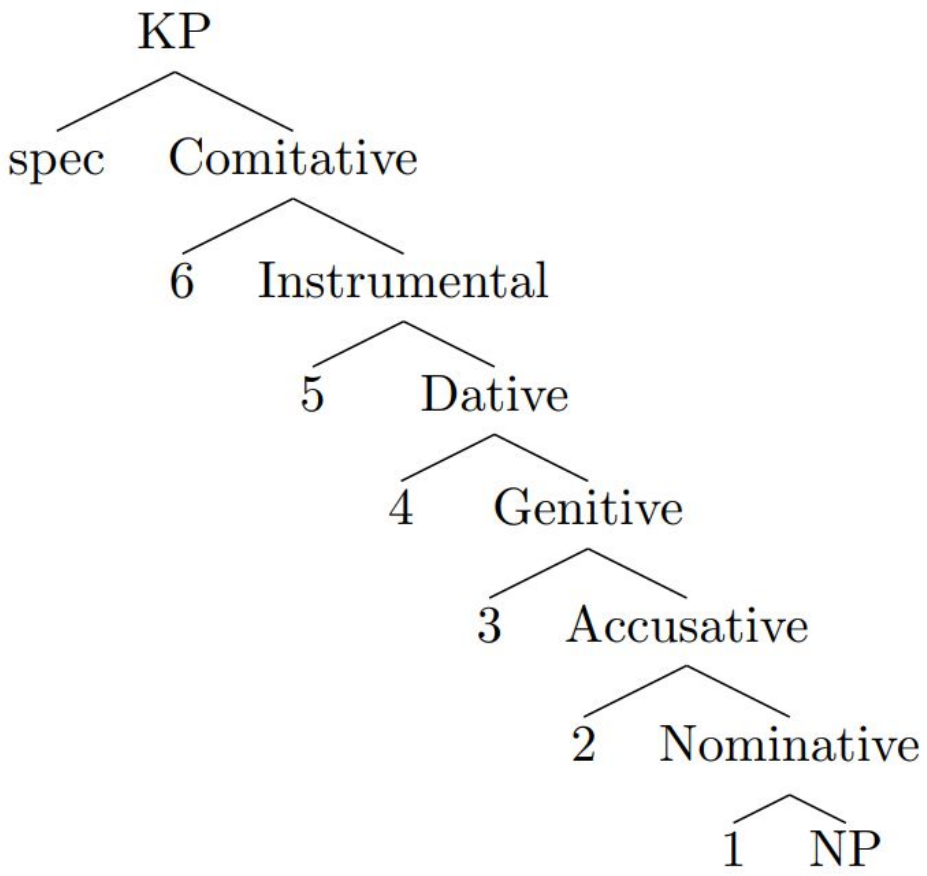


In this approach, the difference between adpositions and case markers is which morphological processes do or do not apply (in a DM kind of approach).

Adpositions and case markers have the same narrow syntactic representation.

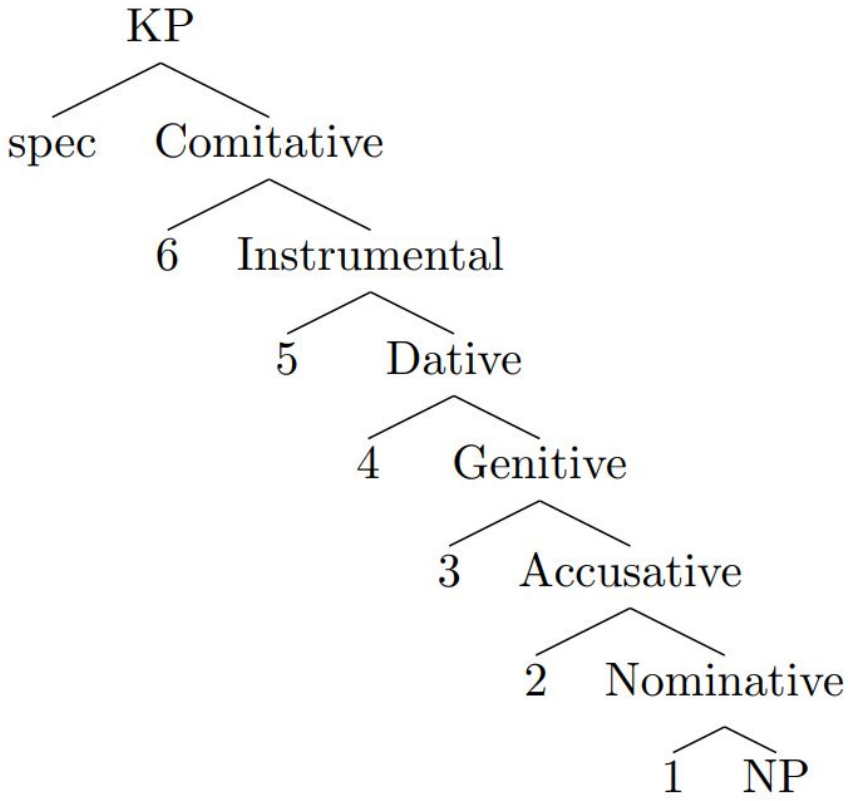
Diachronic predictions of FOFC- a functional load shift

The Final-Over-Final Condition (FOFC) when applied to the KP makes clear diachronic predictions. In a shift from morphological case realisation to prepositional, the shift must occur top-down.



Predictions

- The proportions of adpositional realisation should increase diachronically for each grammatical function, unless the proportion has already reached 100 percent, or in the instances of direct object and subject.
- The adpositional proportion of a given grammatical function should not overtake that of a higher function.

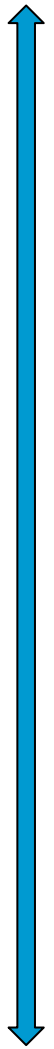


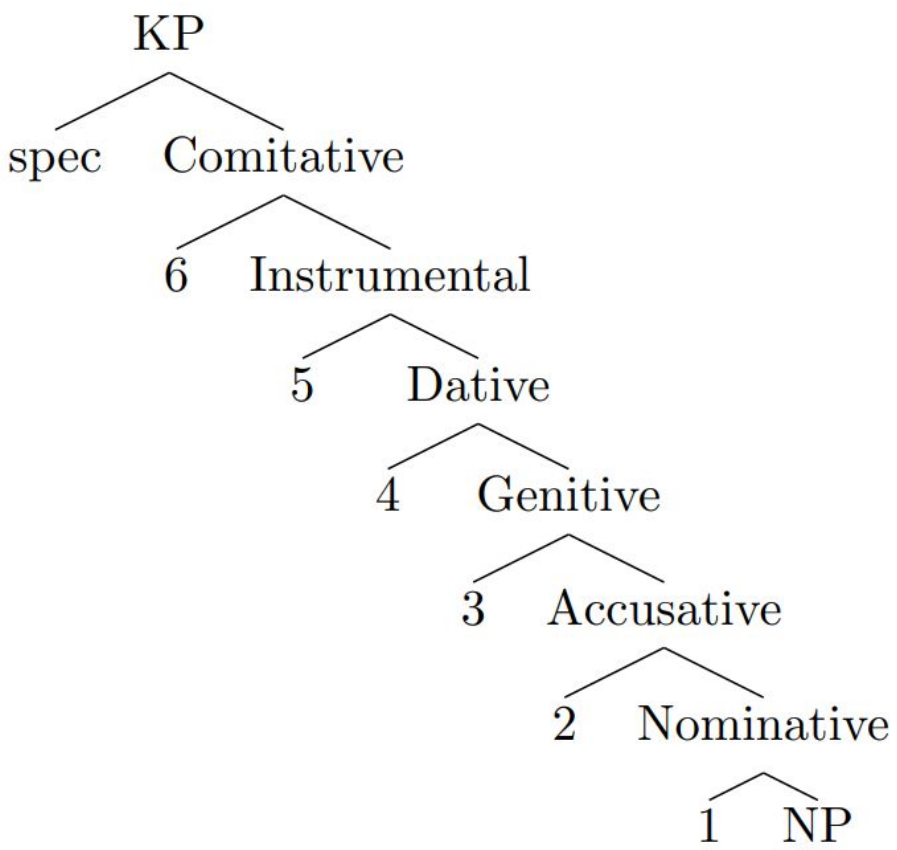
Predictions

Highest proportion of prepositional realisation

A change from m-case realisation to adpositional realisation, moving top-down in the KP.

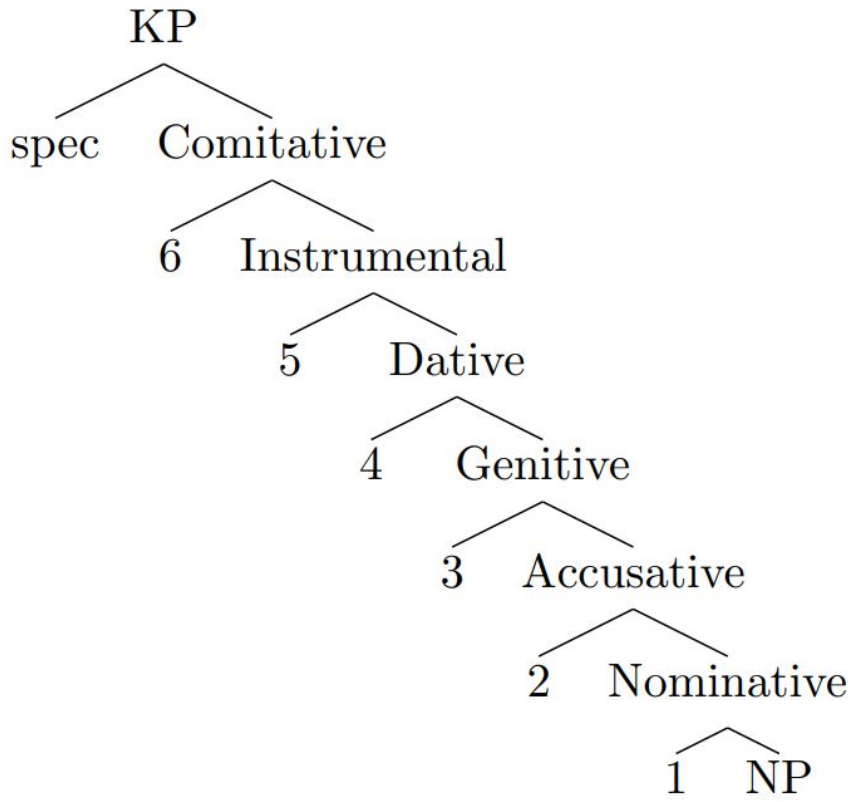
Lowest proportion of prepositional realisation





Corpus: Annotated corpus of Pre-standardised Balkan Slavic Literature

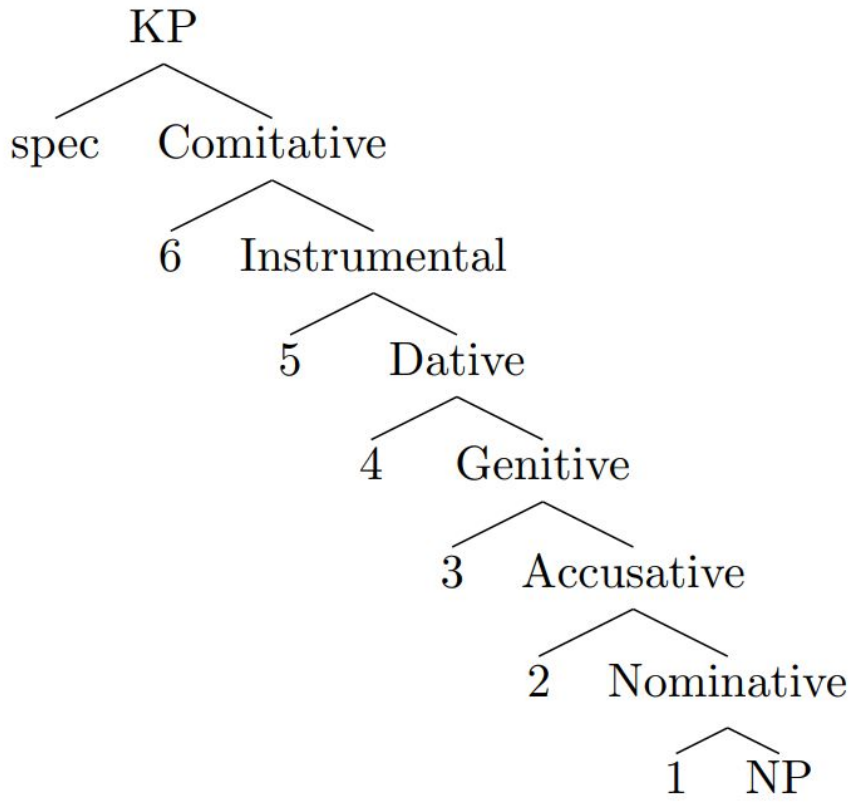
- The corpus texts I include are of various genres and are from the 1300s to 1860 from areas in which Balkan Slavic was spoken (mostly in the modern Bulgarian area).
- Wahlström (2015) dates the bulk of the change to be complete by the 16th century.
- For the later texts in the corpus therefore, the bulk of the change is predicted to have taken place.
- We can use corpus data to track if this is true i.e. how many of the grammatical functions have switched to prepositional realisation in the texts from the end of the 16th century?



Methodology- examining function realisation

- It was necessary to implement the existing corpus annotation to be able to directly compare the grammatical function realisation of adposition phrases with case on nominal phrases. The proportion of adpositional realisation vs non-adpositional realisation (i.e. case-marking or bare) was calculated for each grammatical function in each text.
- The grammatical functions focussed on here are the following:

Comitative
Associative
Instrumental
Goal
Recipient
Possessive
Direct object (theme direct object)
Subject (agentive subject)



Methodology- examining function realisation

Instrumental non-adpositional realisation:

togda	*napisovaaše*	*aleǯandrъ*	*farižь*	*črъveněm*	*vinomъ*	*na*	*bělomъ*	*oubrousě.*
then	wrote	Alexander	Paris	red.LOC*	wine.INS	on	white.LOC/INS	tablecloth.LOC

'Alexander Paris was then writing **with the red wine** on a white napkin.'

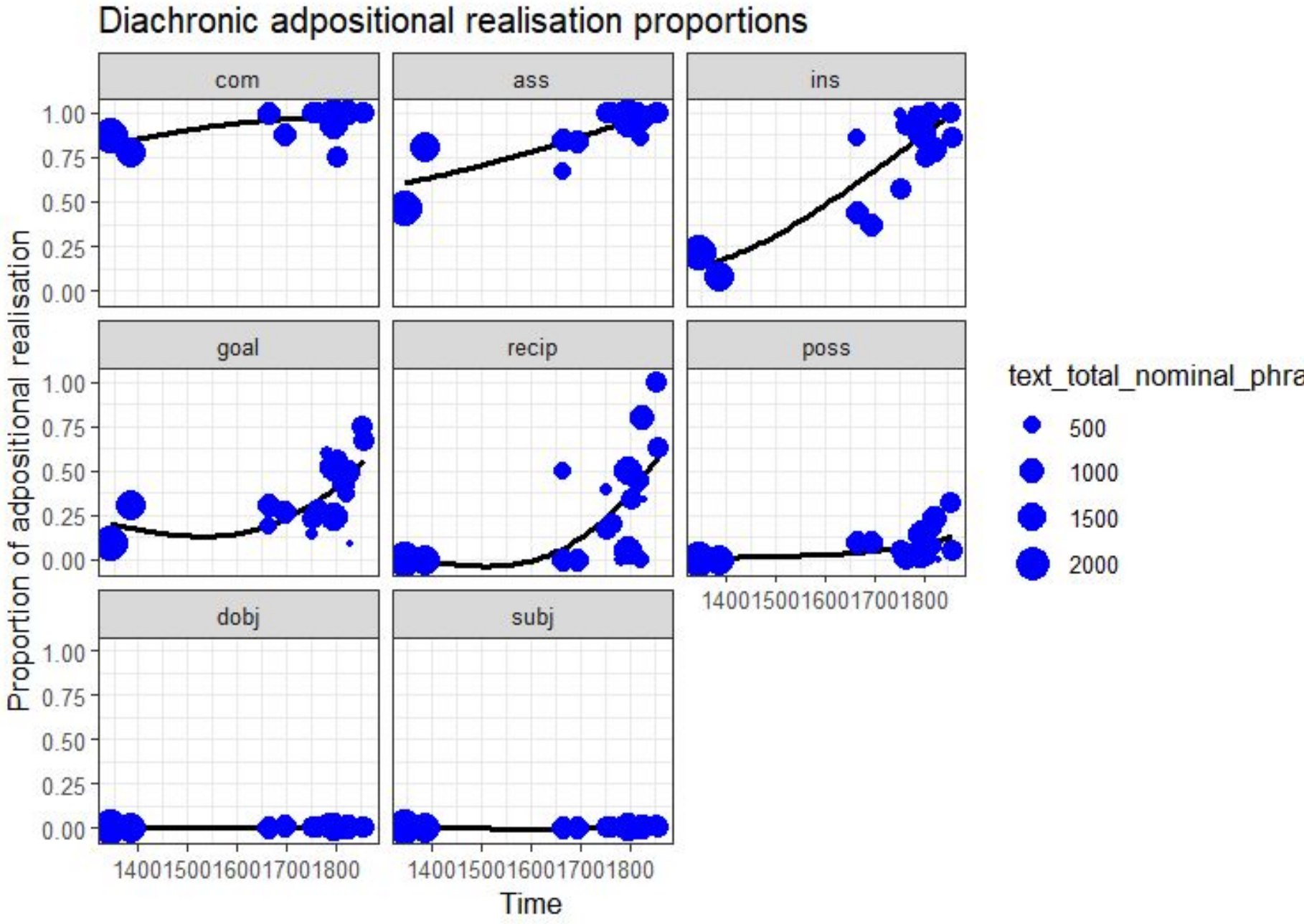
Instrumental adpositional realisation:

... *do*	êdinъ	gráťь	štáťь	da	sá	bię^ťь	sá-sь	kalíčь
until	one	city	want	to	self	beat	with	sword.NOM

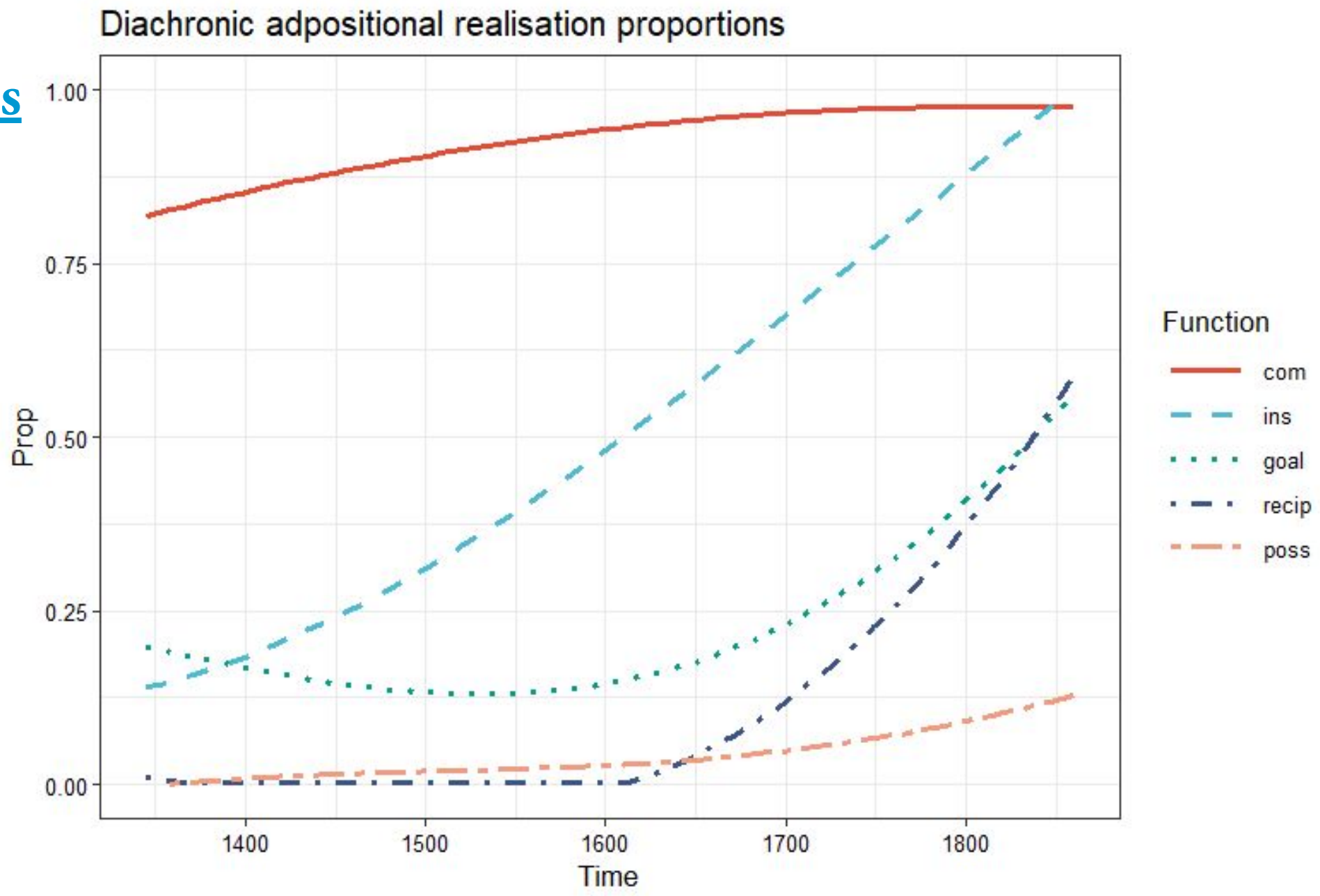
'... they will fight for every city **with swords**.'

*Annotator has noted that this suffix could alternatively be a 'generalised pronominal stem ending'.

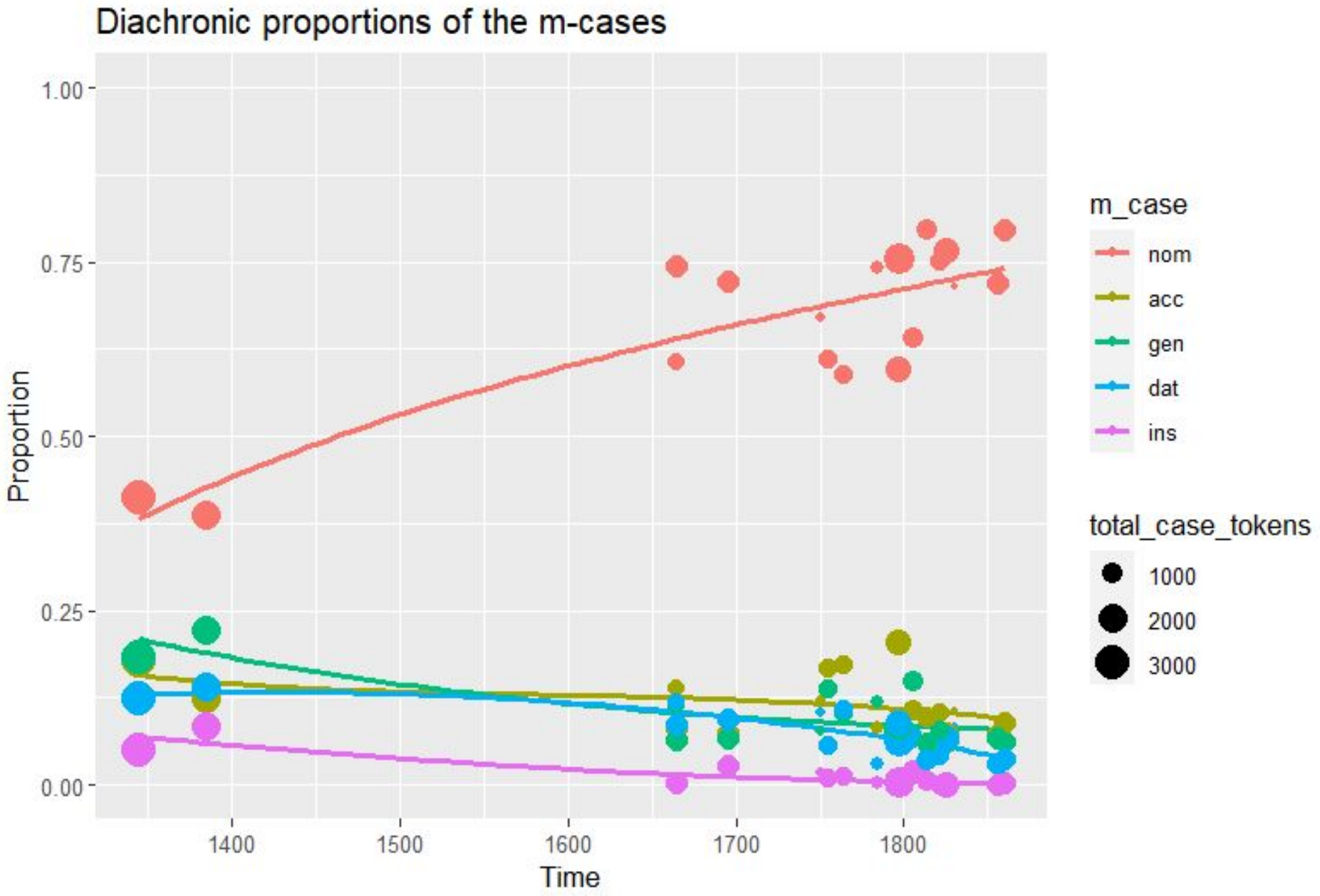
Results


Diachronic adpositional realisation proportions
com
ass
ins
goal
recip
poss
dobj
subj
Proportion of adpositional realisation
1.00
0.75
0.50
0.25
0.00
1400150016001700 1800
Time
text_total_nominal_phra
500
1000
1500
2000

Results


Diachronic adpositional realisation proportions
Prop
1.00
0.75
0.50
0.25
0.00
1400
1500
1600
1700
1800
Time
Function
com
ins
goal
recip
poss

Results


Diachronic proportions of the m-cases
Proportion
1.00
0.75
0.50
0.25
0.00
1400
1500
1600
1700
1800
Time
m_case
nom
acc
gen
dat
ins
total_case_tokens
1000
2000
3000

Interim summary

- Investigating morphosyntactic simplification with corpora, based on Trudgill's sociolinguistic typology and the Interpretability Hypothesis.
- Two case studies of morphosyntactic simplification.
- Increase in the simplex overt subjects in Latin American Spanish.
- Order of loss of morphological cases follows predictions of Caha's KP and FOFC.

Any questions?

Part I References (1)

Biberauer, Theresa, Anders Holmberg & Ian Roberts. (2014). *A syntactic universal and its consequences*. Linguistic Inquiry 45. 169–225.

Bini, M. (1993). "La adquisición del italiano: Más allá de las propiedades sintácticas delparámetro pro-drop." In J. M. Liceras (Ed.), *La lingüística y el análisis de los sistemasno nativos:* 126–139. Dovehouse Editions Canada.

Blake, Barry J. (1994). Case. Cambridge University Press, Cambridge, second edn.

Caha, P. (2009). *The Nanosyntax of Case*. Ph.D. thesis, University of Tromsø.

Holmberg, Anders. (2000). Deriving OV order in Finnish. In Peter Svenonius (ed.), *The derivation of VO and OV*, 123–152. Amsterdam: John Benjamins.

Lindstedt, J. (2000). *Linguistic Balkanization: Contact-induced change by mutual reinforcement*

Lindstedt, J. (2019). Diachronic Regularities Explaining the Tendency towards Explicit Analytic Marking in Balkan Syntax, in *Balkan Syntax and (Universal) Principles of Grammar*.

Pérez-Leroux, A. T., & Glass, W. R. (1999). "Null anaphora in Spanish second language acquisition: Probabilistic versus generative approaches." *Second Language Research*, 15 (2): 220–249.

Margaza, P., & Bel, A. (2006). "Null subjects at the syntax–pragmatics interface: Evidence from Spanish interlanguage of Greek speakers." In M. Grantham O'Brien, C. Shea, &J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006):* 88–97. Cascadilla Proceedings Project.

Sessarego, Sandro. (2013). "Afro-Hispanic Contact Varieties as Conventionalized Advanced Second Languages". *IBERIA* 5 (1): 99-125.

Šimko, Ivan. (2020). *Annotated Corpus of Pre-Standardized Balkan Slavic Literature*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1368.

Part I References (2)

Sorace, Antonella. 2011. “Pinning down the concept of “interface” in bilingualism”. *Linguistic Approaches to Bilingualism* 1(1): 1-33.

Toribio, Almeida J. (2000). “Setting parametric limits on dialectal variation in Spanish”. *Lingua: International Review of General Linguistics* 110 (5): 315–341.

Trudgill, P. (2011). On the sociolinguistic typology of linguistic complexity loss. *Language Documentation & Conservation, Special Publication* No. 3, 90-95.

Tsimpli, Ianthi Maria & Dimitrakopoulou, Maria. (2007). *The Interpretability Hypothesis: Evidence from wh-interrogatives in second language acquisition*. Second Language Research 23: 215–242.

Wahlström, Max. (2015). *The loss of case in Bulgarian and Macedonian*. Doctoral thesis, University of Helsinki, Helsinki, Finland.

Walkden, G., and A. Breitbarth. (2019). Complexity as L2-difficulty: Implications for syntactic change, in *Theoretical Linguistics 45(3– 4)* , 183–209

Measuring Orality

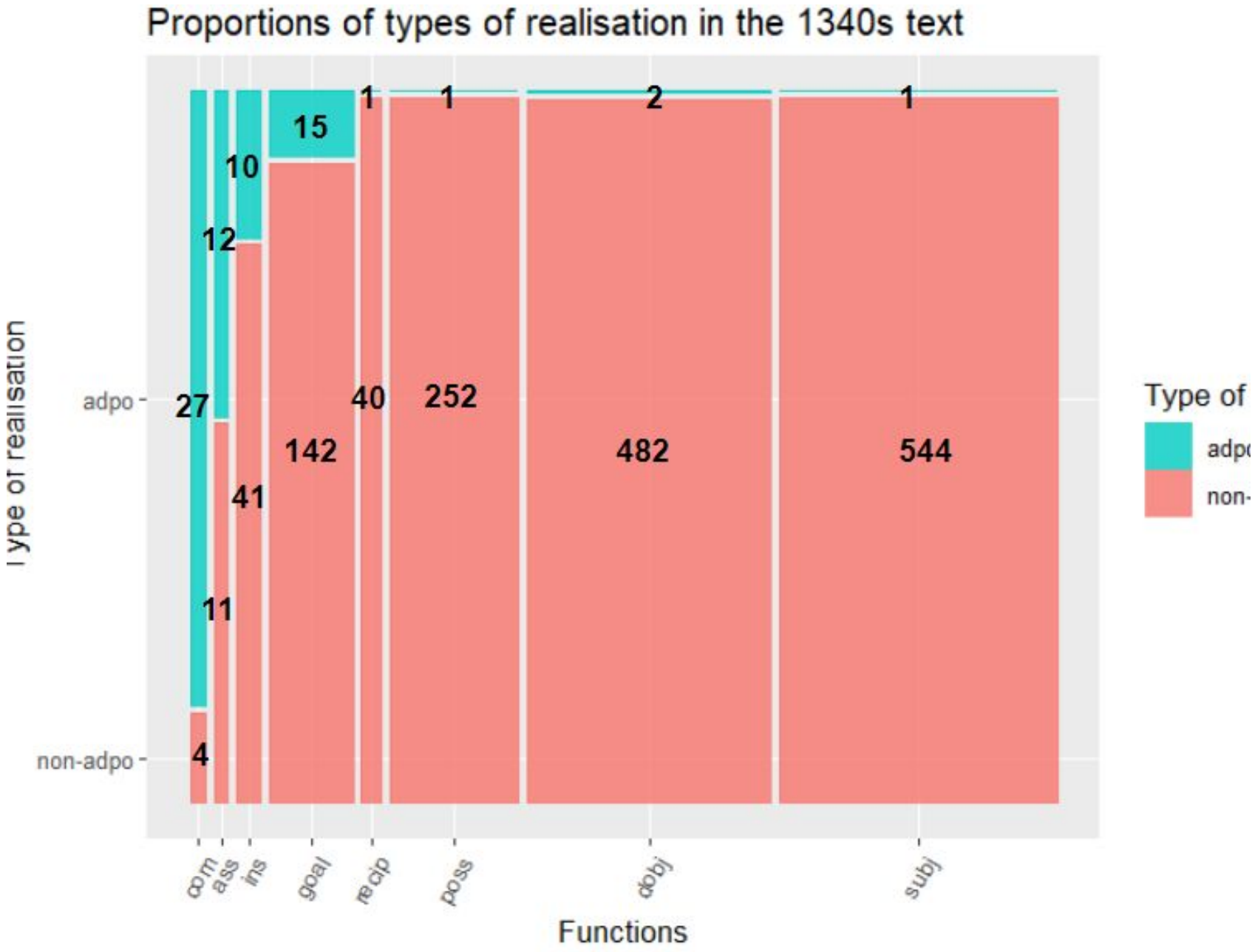
- **Rosemeyer (2019) measured orality levels in a diachronic corpus of Brazilian Portuguese plays**
- **The plays followed a shift toward reflecting spoken speech over the centuries**

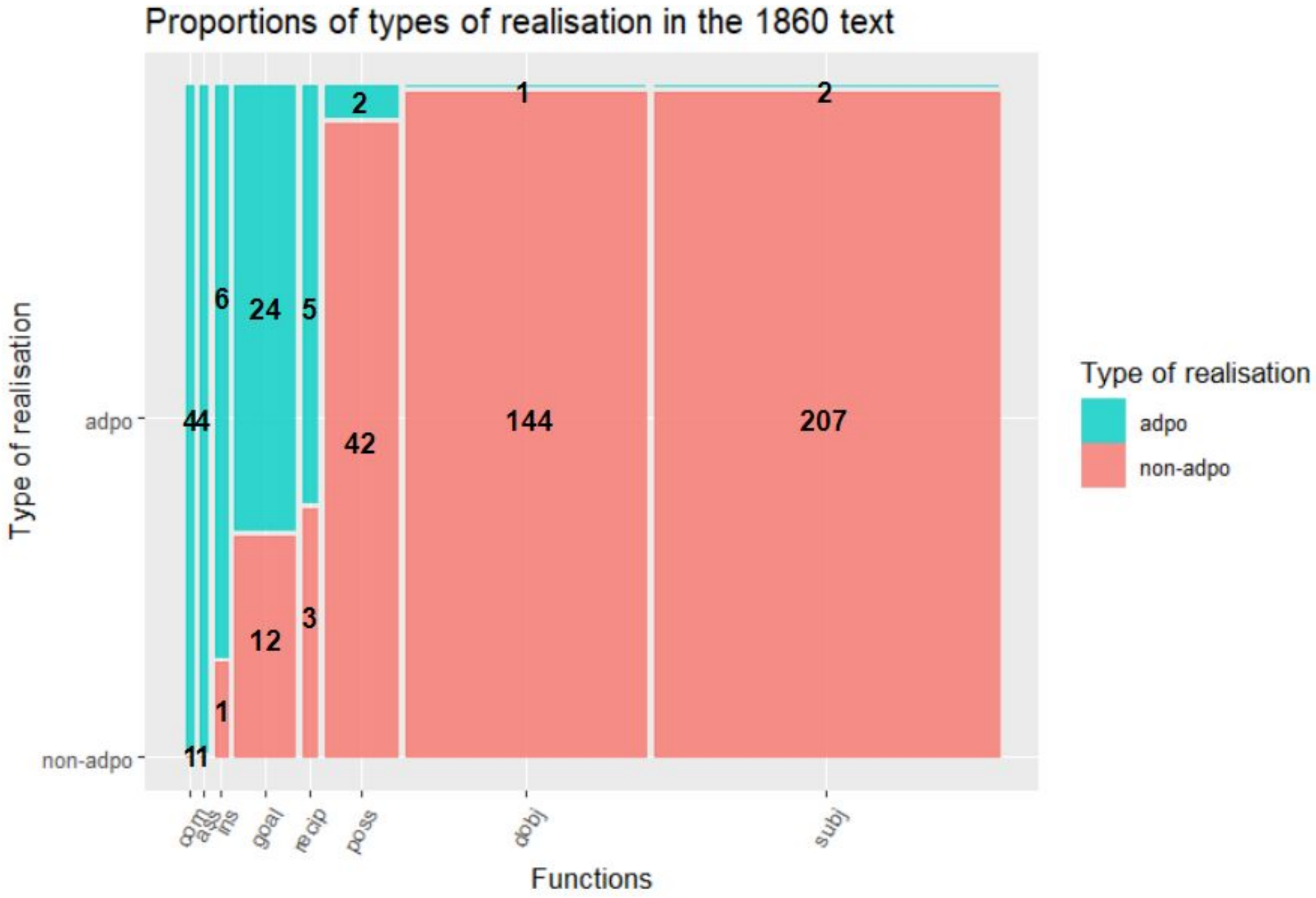
Rosemeyer (2019) variables:

- Present progressive
- Demonstrative neuter pronouns
- Time and place adverbs
- Discourse markers
- Private verbs

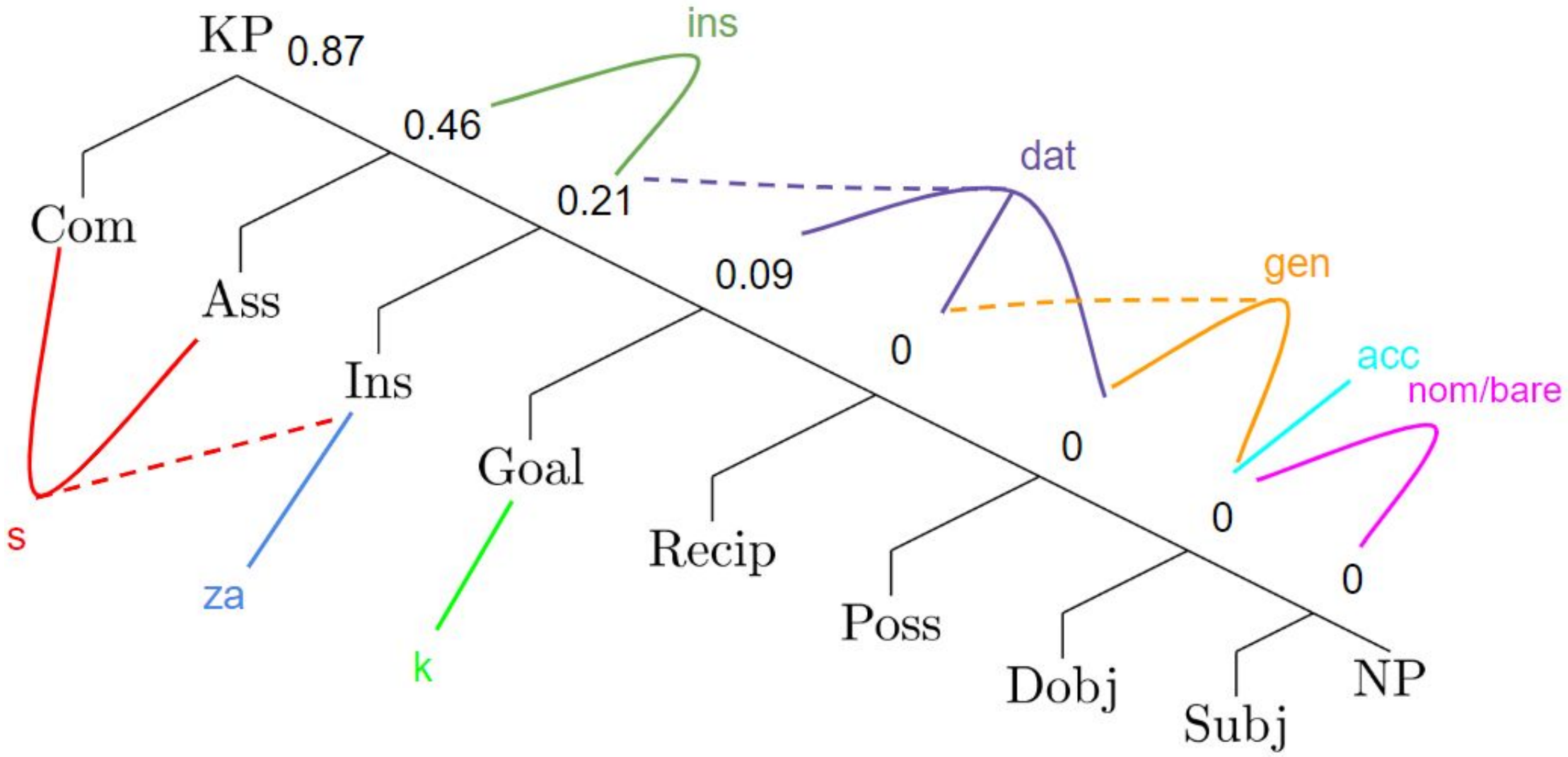
Our variables:

- Progressive
- Demonstrative neuter pronouns
  - *esto/eso/aquello*
- Time and place adverbs
  - *aquí/ahora*
- Private verbs
  - *pensar* 'to think' / *creer* 'to believe'

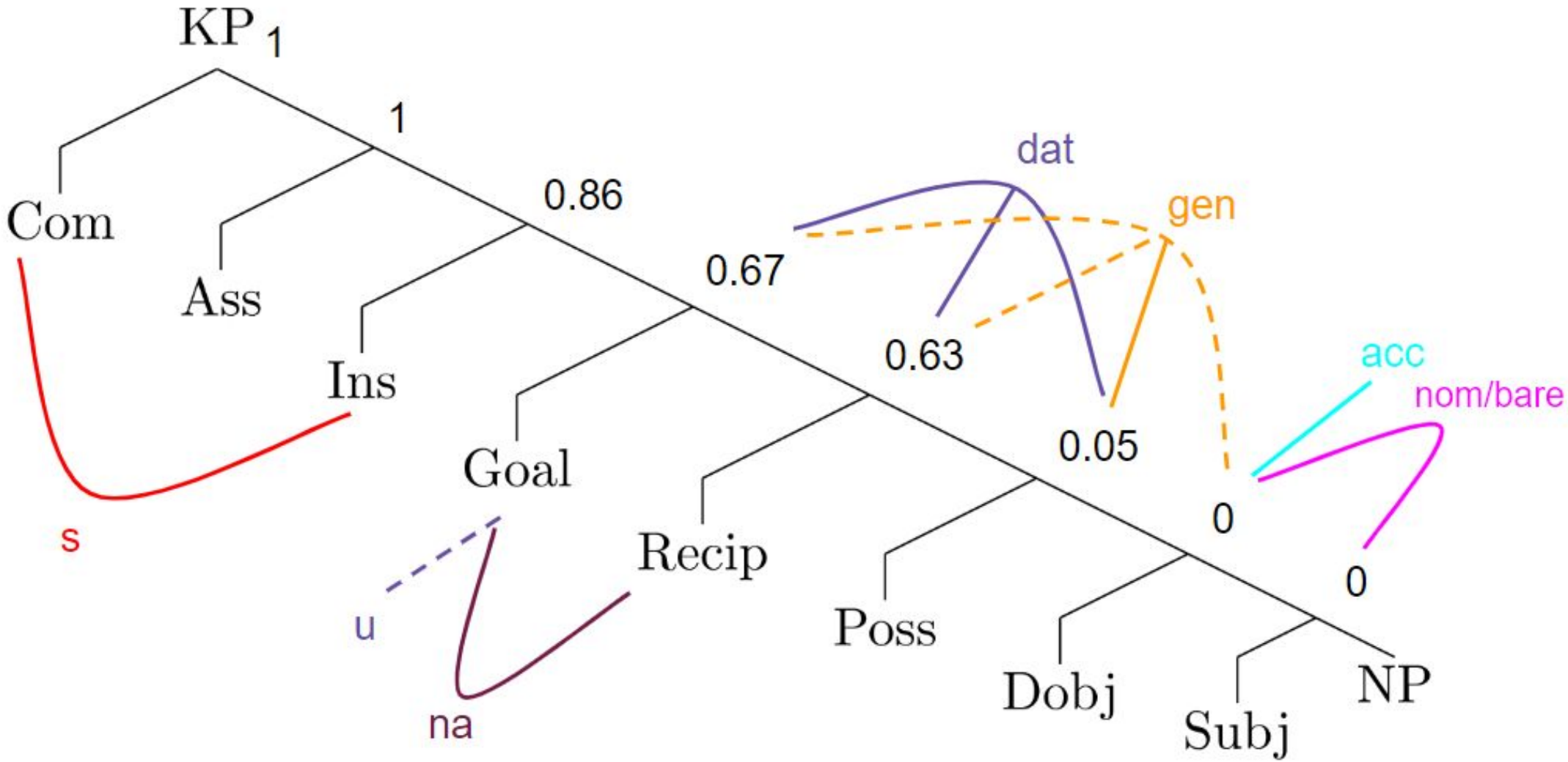

Proportions of types of realisation in the 1340s text
Type of realisation
adpo
non-adpo
27
4
12
11
10
41
15
142
1
40
1
252
2
482
1
544
com
ass
ins
goal
recip
poss
dobj
subj
Functions
Type of
adp
non-

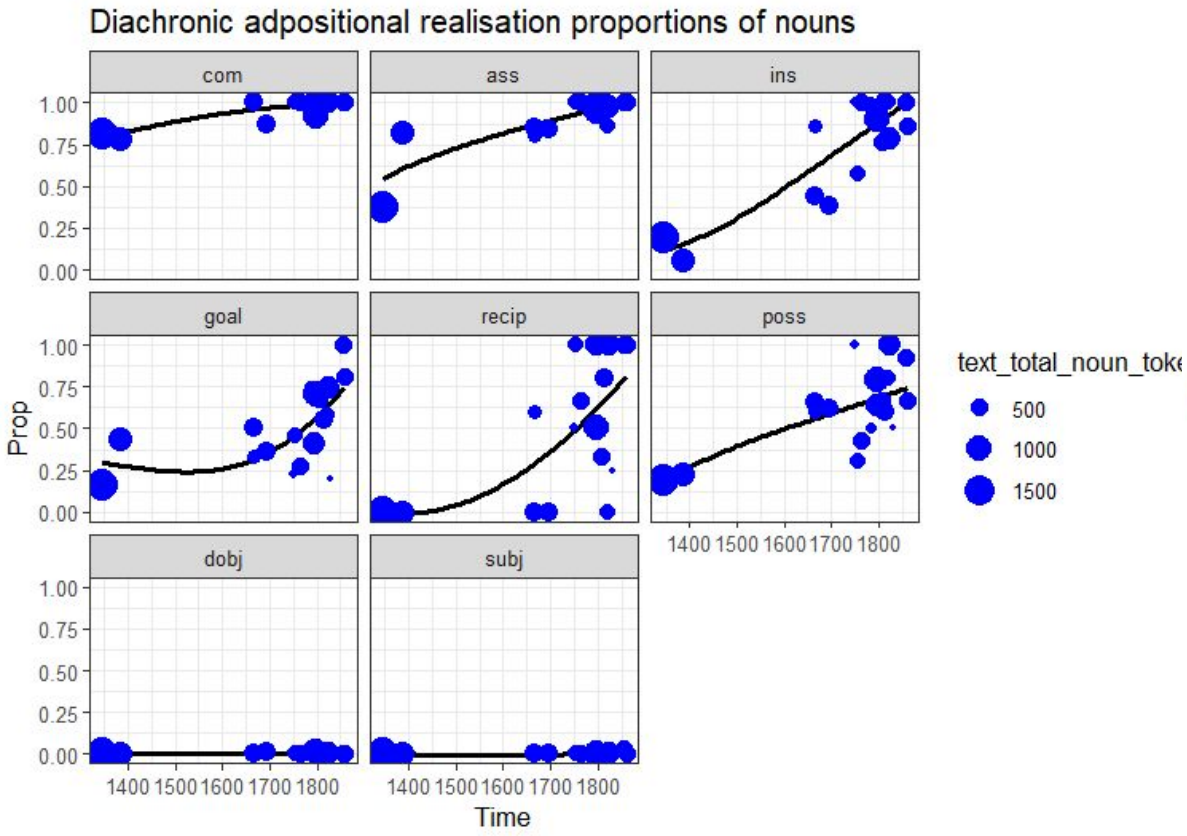

Proportions of types of realisation in the 1860 text
Type of realisation
adpo
non-adpo
44
11
6
1
24
12
5
3
2
42
1
144
2
207
com
ass
ins
goal
recip
poss
dobj
subj
Functions
Type of realisation
adpo
non-adpo

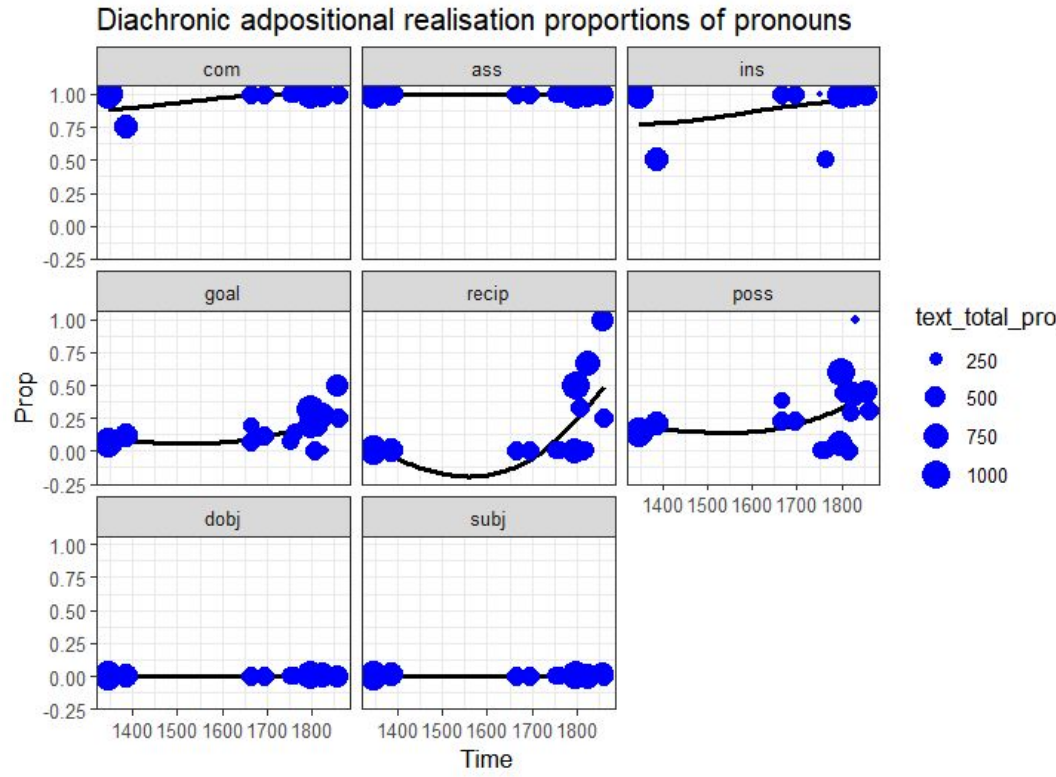
Oldest text- 1340s
Adpositional realisation proportions, adpo forms, m-cases and their syncretisms

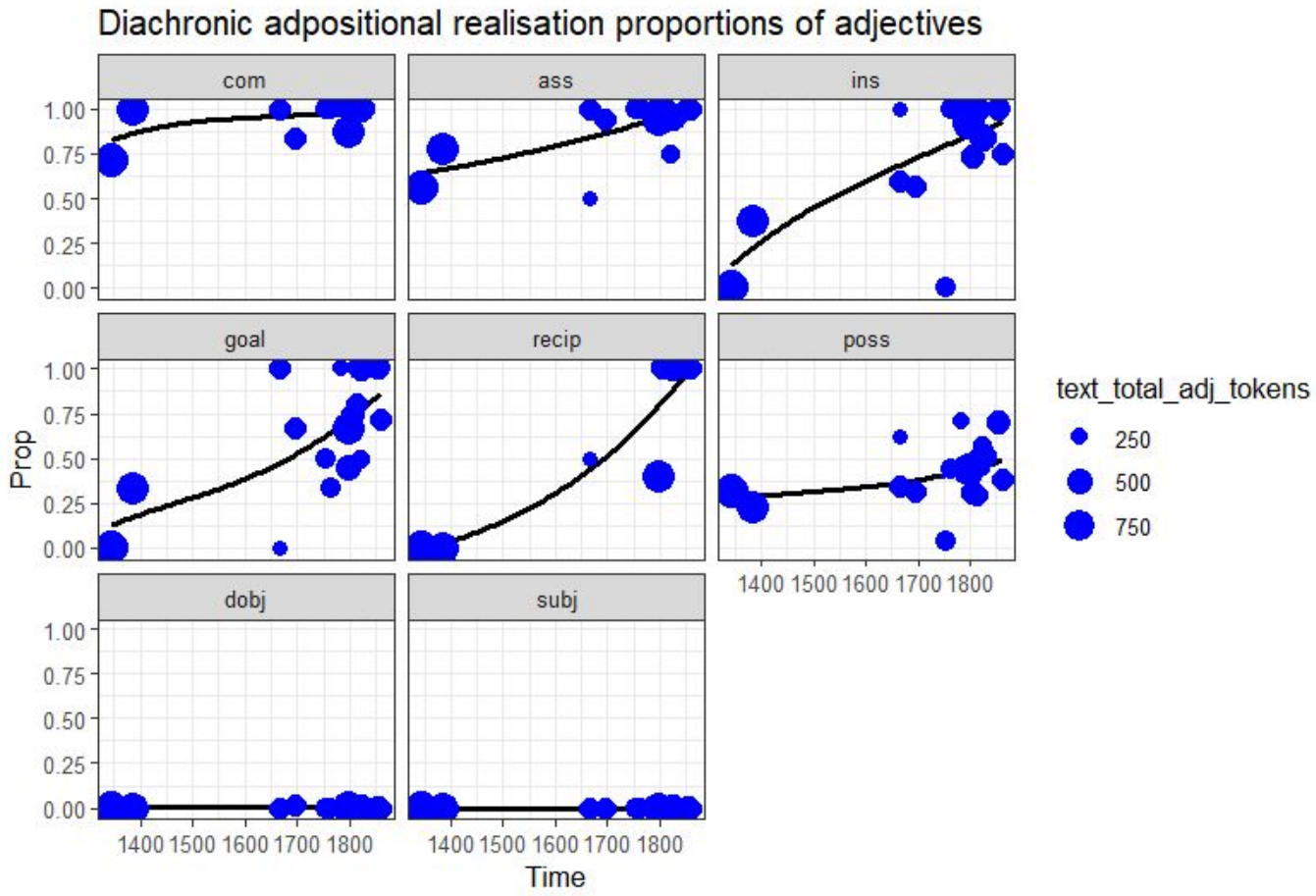


Newest text- 1860
Adpositional realisation proportions, adpo forms, m-cases and their syncretisms




Diachronic adpositional realisation proportions of nouns
com
ass
ins
goal
recip
poss
dobj
subj
Prop
Time
1.00
0.75
0.50
0.25
0.00
1400 1500 1600 1700 1800
text_total_noun_toke
500
1000
1500


Diachronic adpositional realisation proportions of pronouns
com
ass
ins
goal
recip
poss
dobj
subj
Prop
Time
1.00
0.75
0.50
0.25
0.00
-0.25
1400 1500 1600 1700 1800
text_total_pro
250
500
750
1000


Diachronic adpositional realisation proportions of adjectives
com
ass
ins
goal
recip
poss
dobj
subj
Prop
1.00
0.75
0.50
0.25
0.00
1400 1500 1600 1700 1800
Time
text_total_adj_tokens
250
500
750

Part II. Combining methodologies

STARFISH PROJECT

Presented by: Molly Rolf and Raquel Montero Estebaranz

Universität Konstanz

June 2024

STARFISH

Introduction

- So far we have seen the benefits of taking a Corpus Linguistics approach to explore contact induced language change.
- However, there are other equally interesting approaches that have been used:

Typology	**Experimental**
Lupyan and Dale (2010)	Berdicevskis and Semenuks (2022)
Bentz and Winter (2013)	Atkinson, Smith, and Kirby (2018)
Sinnemäki and Di Garbo (2018)	Raviv, Meyer, and Lev-Ari (2019)
...	...

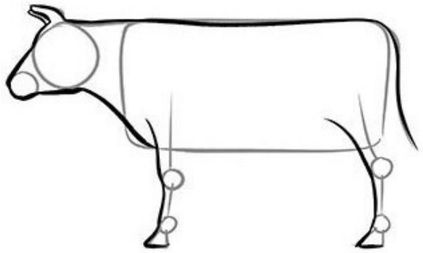
- Each of these approaches investigates the contact question from different levels of abstraction.

Abstraction

- Abstraction helps in the study of a limited set of properties in isolation.

realistic cow

Simplified cow



Spherical cow

- But it can sometimes lead to what has been termed as “negligibility assumptions”: the approach might neglect a causally relevant factor (Frigg and Hartmann 2006).
- “An issue as nuanced as the Trudgill conjecture ought to benefit from as many different lines of attack as possible” (Walkden and Breitbarth 2019: 33)

Finding the right balance of abstraction

Research Question: What are the possible outcomes of language contact?

Corpus Linguistics

- Real Data
- Noisy

Computational Modeling

- Predictions
- Idealized

STARFISH: Combining different methodologies

Structure of Part II

1.Modeling Language Change and Contact

1.1 VL Model

1.2. Long-term contact

1.3. Short-term contact

1.4. Fine-tuning Trudgill's Sociolinguistic Typology

2.Case study: subjunctive morphology in Spanish

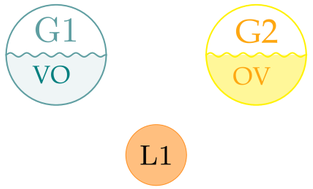
2.1. Polarity Subjunctive in Spanish

2.2. Contact in Medieval Spain

3.Conclusion

L1 acquisition: Variational Learner (Yang 2000)

Variational Learner

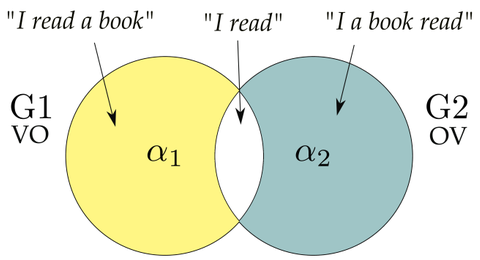


L1 leaner

- w_1 = Probability G1
- w_2 = Probability G2
- γ = Learning rate

(Yang, 2000)

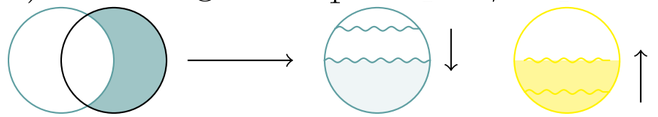
Linguistic Input



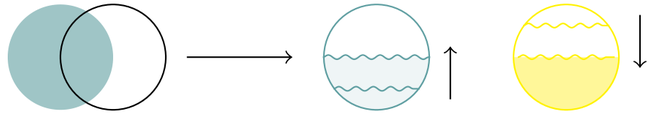
- α_1 = Probability of a G1 speaker uttering a sentence G2 cannot parse.
- α_2 = Probability of a G2 speaker uttering a sentence G1 cannot parse.
- If $\alpha_1 > \alpha_2$, G1 wins

Learning Algorithm

a) Punishing G1: $w_1' = w_1 - \gamma \cdot w_1$

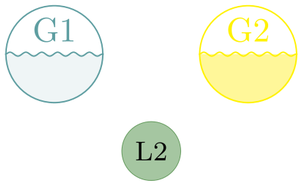


b) Rewarding G1: $w_1' = w_1 + \gamma(1 - w_1)$



L2 acquisition: Variational Learning (Kauhanen 2022)

Variational Learner

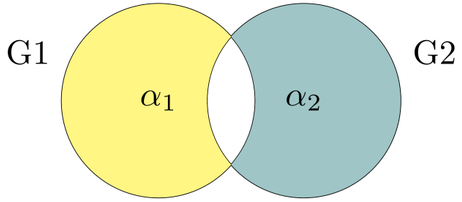


L2 leaner

- w_1 = Probability G1
- w_2 = Probability G2
- γ = Learning rate
- $\boldsymbol{\delta}$ = **L2 difficulty**

(Kauhanen, 2022)

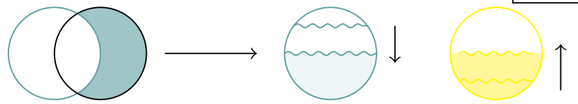
Linguistic Input



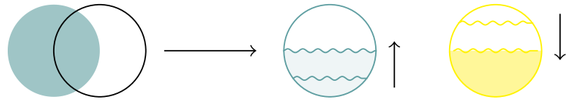
- α_1 = Probability of a G1 speaker uttering a sentence G2 cannot parse.
- α_2 = Probability of a G2 speaker uttering a sentence G1 cannot parse.

Learning Algorithm

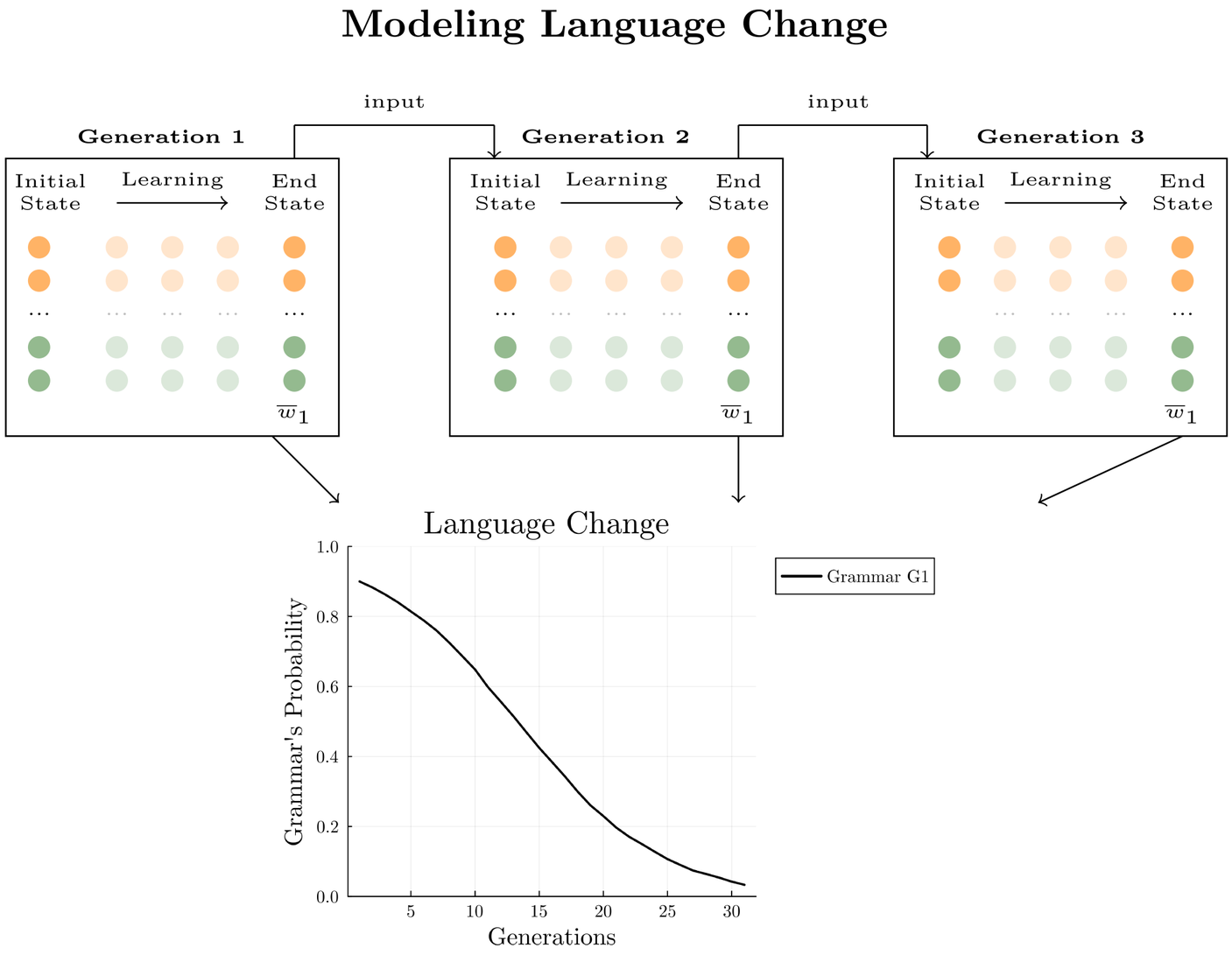
a) Punishing G1: $w_1' = w_1 - \gamma \cdot w_1$ $\boxed{-\boldsymbol{\delta} \cdot \boldsymbol{w_1}}$



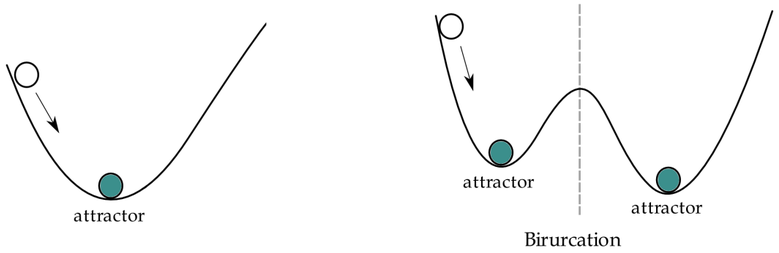
b) Rewarding G1: $w_1' = w_1 + \gamma(1 - w_1)$ $\boxed{-\boldsymbol{\delta} \cdot \boldsymbol{w_1}}$

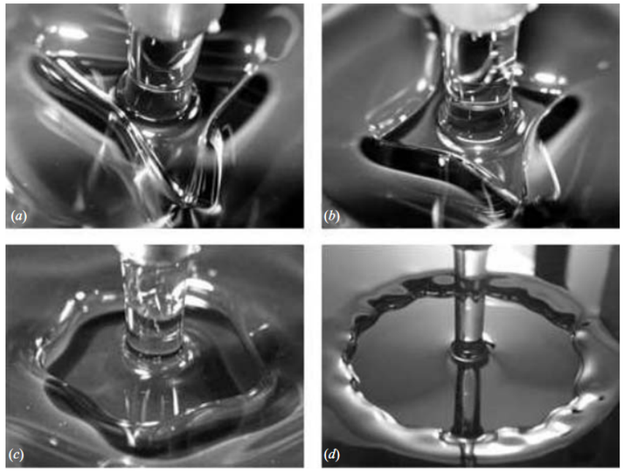


Language change: generations of VL



The Bifurcation Threshold (Kauhanen 2022)





- **Attractor**: a set of values towards which the system tends to over time.
- **Bifurcations**: the qualitative behavior of the system changes at certain values (one attractor changes to multiple ones).
- **Example from hydraulic jumps** : depending on the water flow the system stabilizes in a given behavior (Bush, Aristoff, and Hosoi 2006).
- **L2 Contact**:

a. If the number of speakers is above a certain threshold, the system will have one behavior.

b. If the number of speakers is below a certain threshold the system will have another behavior.

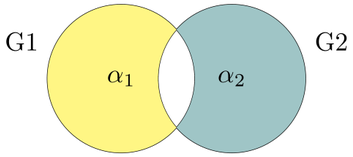
Bifurcation Threshold (Kauhanen 2022)

If the number of L2 speakers is below a threshold (σ_{crit}), the L2-difficult feature will be preserved, and if above that threshold the feature will be lost.

$$\sigma_{crit} = \frac{(\alpha - 1)(D + 1)}{\alpha D}$$

The α parameter

$$\alpha = \frac{\alpha_1}{\alpha_2} = \frac{\text{G1 advantage}}{\text{G2 advantage}}$$



The D parameter

$$D = \frac{d}{\alpha_2} = \frac{\text{G1 L2-difficulty}}{\text{G2 advantage}}$$

How the values affect σ_{crit}

α_1: 0.2

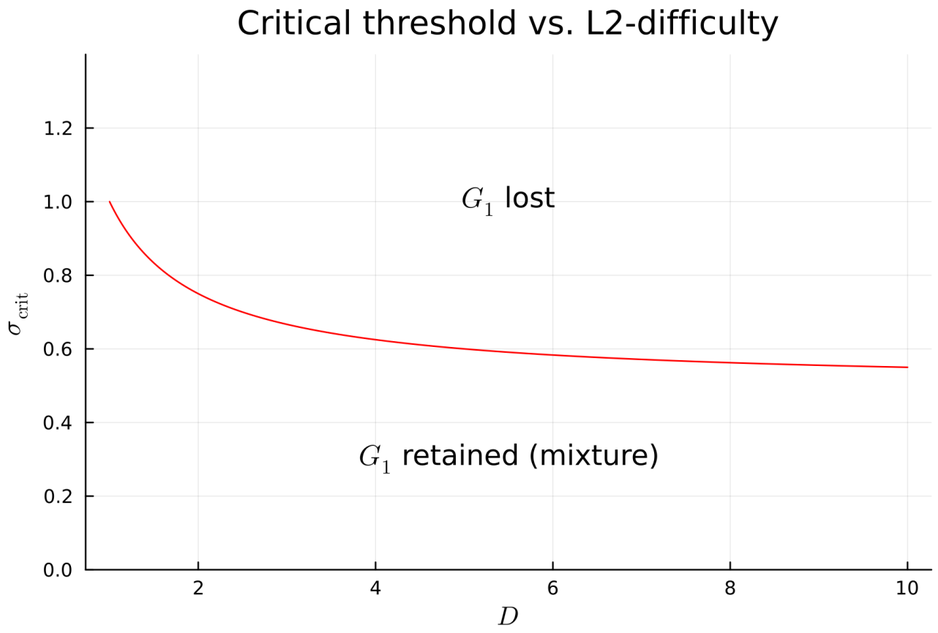
α_2: 0.1

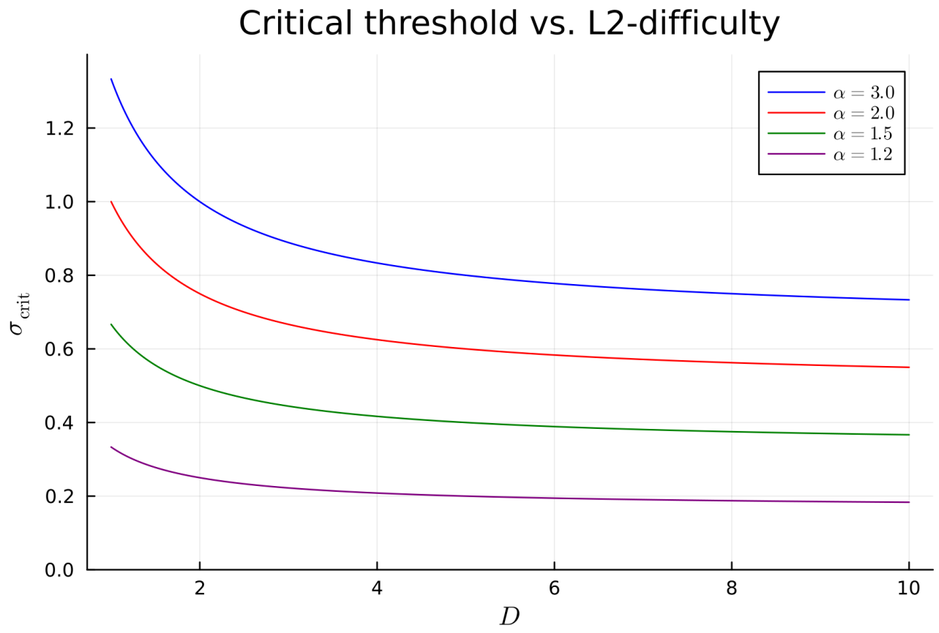
d: 0.3

Threshold = 0.67

Bifurcation Threshold (Kauhanen 2022)

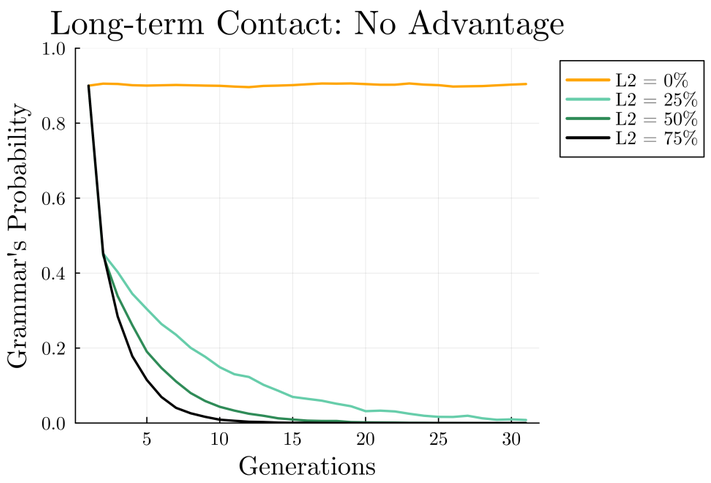
$$\sigma_{crit} = \frac{(\alpha - 1)(D + 1)}{\alpha D}$$





Long-term contact

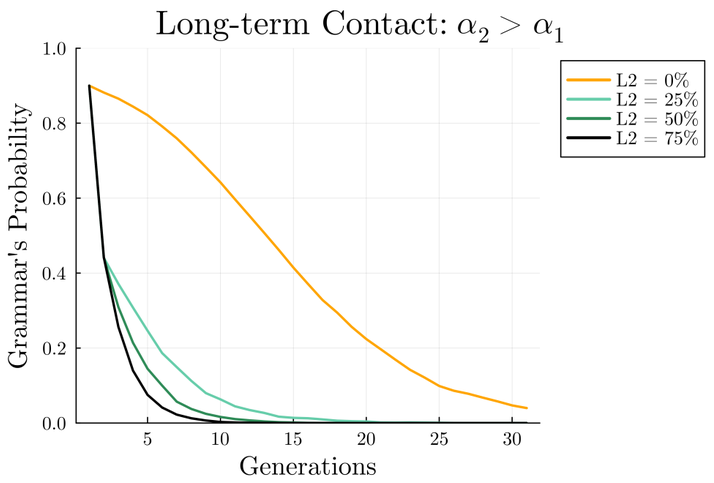
Values | *Simulation*



L2 difficult feature:

- Lost (speed depends on nº L2)

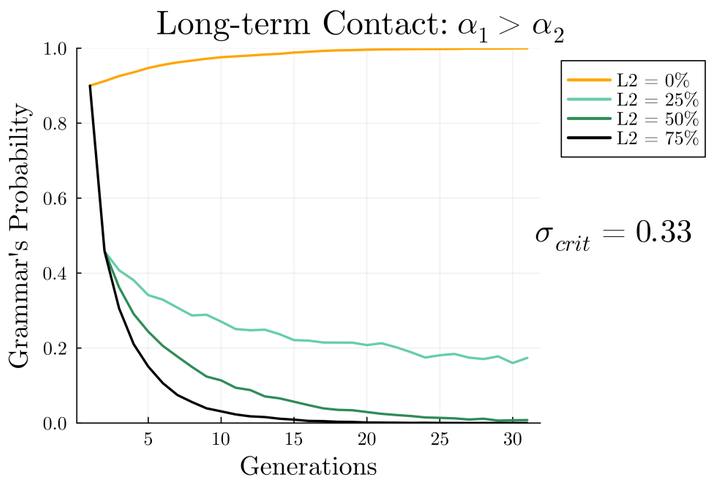
Values | *Simulation*



L2 difficult feature:

- Lost (speed depends on nº L2)

Values | *Simulation* | *Bifurcation*



L2 difficult feature:

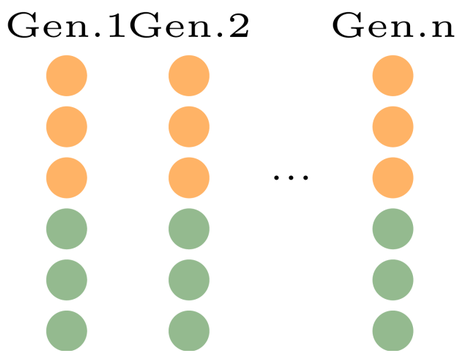
- Lost if nº L2 is above bifurcation threshold (σ_{crit})
- Retained if nº L2 is below bifurcation threshold (σ_{crit})

Types of Language Contact

Long-term contact

Prolonged influx of L2 speakers

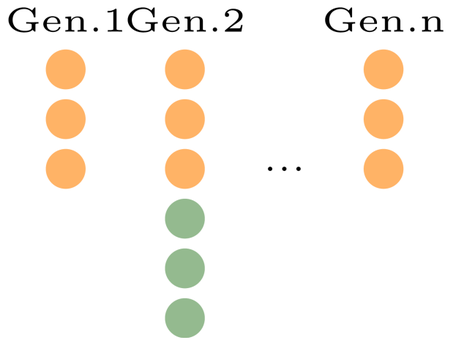
- Lingua Franca
- Trading areas (e.g., *Hanseatic League*)



Short-term Contact

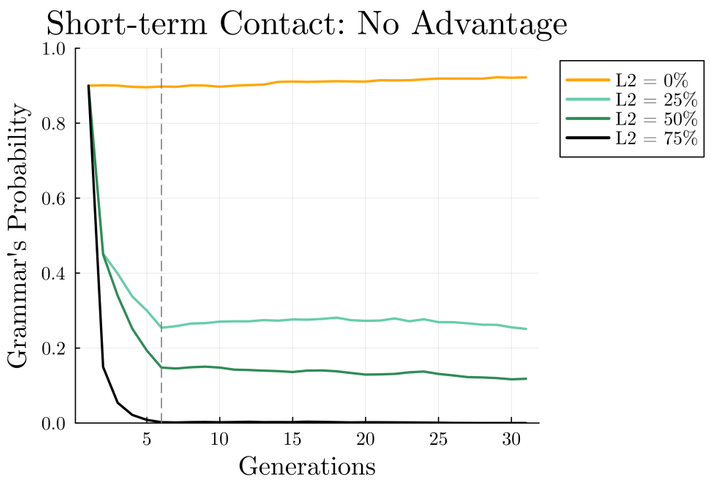
Short influx of L2 speakers

- Population movements (e.g., *Gastarbeiter*, *expulsions* etc.)



Short-term contact

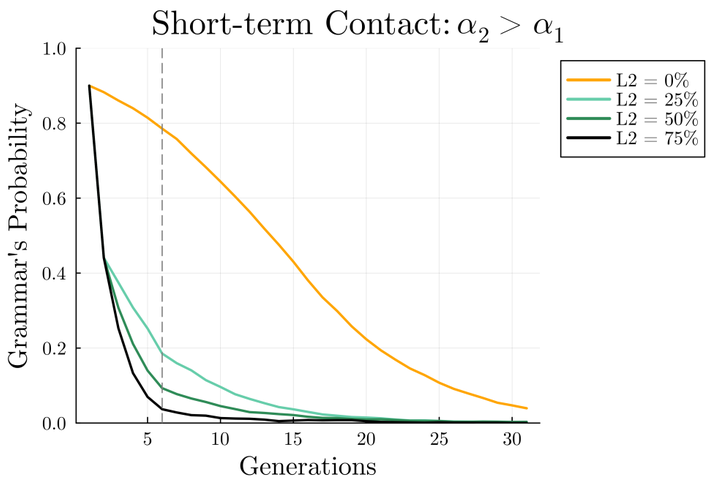
Values | *Simulation*



L2 difficult feature:

- Lost if G1's probability reaches 0 during contact
- Retained: if G1's probability doesn't reach 0 during contact

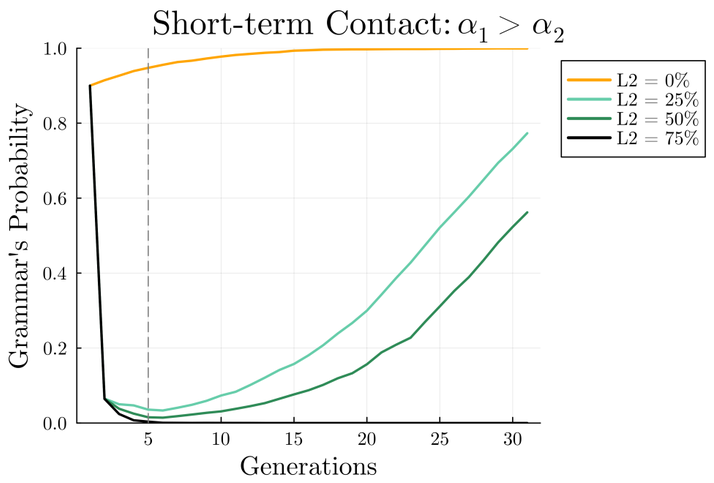
Values | *Simulation*



L2 difficult feature:

- Lost (speed depends on nº L2)

Values | *Simulation*



L2 difficult feature:

- Lost if G1's probability reaches 0 during contact.
- Retained if G1's probability doesn't reach 0 during contact.

Fine-tuning Trudgill's typology

- Trudgill (2010) : adult L2 difficult features will be lost in cases of contact characterized by a high number of adult L2 speakers.
- Proposal: L2 difficult features may be lost or retained, the outcome is determined by several factors:

Type of Contact	**Pred.**	**Advantages**	**Further Conditions**
Long-term	Lost	$\alpha_1 = \alpha_2$	–
	Lost	$\alpha_2 > \alpha_1$	–
	Lost	$\alpha_1 > \alpha_2$	$\sigma > \sigma_{crit}$ & $1 < \alpha <$ D+2
	Retained	$\alpha_1 > \alpha_2$	$\sigma < \sigma_{crit}$ & $1 < \alpha <$ D+2
Short-term	Lost	$\alpha_1 = \alpha_2$	w_1: 0 during contact
	Retained	$\alpha_1 = \alpha_2$	w_1: not 0 during contact
	Lost	$\alpha_2 > \alpha_1$	–
	Lost	$\alpha_1 > \alpha_2$	w_1: 0 during contact
	Retained	$\alpha_1 = \alpha_2$	w_1: not 0 during contact

Case Study: Polarity Subjunctive in Spanish

The Phenomenon: Polarity Subjunctive (PS) in Spanish

- Polarity Subjunctive: certain verbs, when negated, allow the embedded verb to be in indicative or subjunctive mood in Spanish (Quer 1998):

(1) Ariel sabía que hacía /*hiciese calor
Ariel knew that was.IND was.SUBJ warm
'Ariel knew it was warm.'

(2) Ariel **no** sabía que hacía /hiciese calor
Ariel not knew that was.IND was.SUBJ warm
'Ariel didn't know it was warm.'

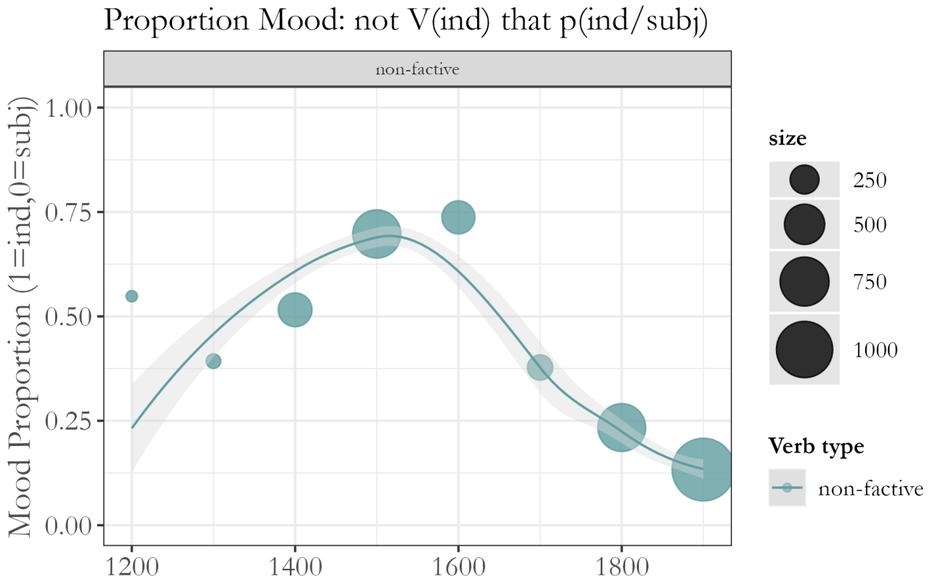
- Polarity Subjunctive has been shown to be particularly hard to master for L2 speakers of different languages (Borgonovo and Prévost 2003; Iverson, Kempchinsky, and Rothman 2008)

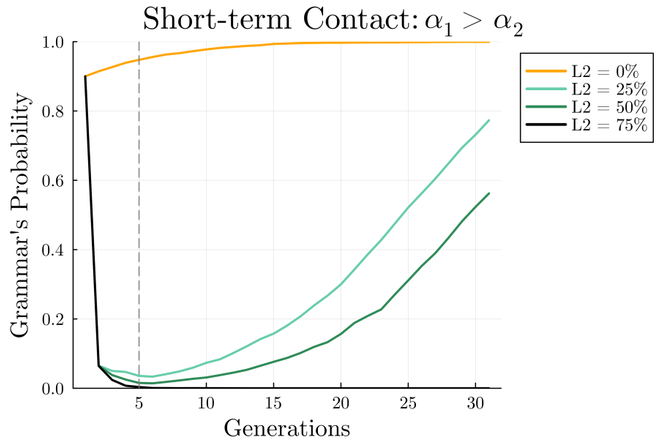
Corpus Study: mood in the history of Spanish

- **Corpus**: *Corpus del diccionario histórico de la lengua española* (CDHLE)
- **Data**: published in Spain; 12th-20th centuries
- **Searches**: neg + V(non-factive) + that … Vind/subj[1]
- **Verbs**: believe and say (most frequent in the corpus)

1. The diachronic evolution of factive verbs is a bit different. See Supplementary Slides

Corpus Study: results non-factive verbs

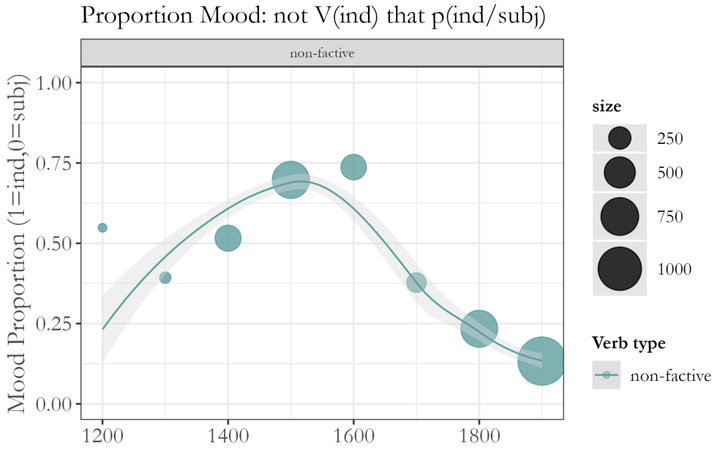




- Up-and-then-down behavior (non-monotonic)
- Could this be explained via language contact?
- There is one specific case from the models we saw before that resembles the change we see here.
- Was there any type of contact with L2 speakers between 1200-1600?

Contact with (Andalusi) Arabic

Video by Alexandre Vigo Wikipedia



Ruhstaller and Gordón Peral (2017), Vicente (2020)

- Contact: 711-1600
  - "Andalusi Arabic continued to be spoken in the Iberian Peninsula after the end of al-Andalus as a Muslim–Arab state in 1492 CE, as some of the Arabic-speaking population remained in certain regions up until the 17th century "
  - "In those areas which retained their Mudejar populations, contact between the two languages continued, and it was principally through these speech communities that Arabic terms were incorporated into Romance."
  - Effect of contact: mostly of lexical loanwords, but included some function words like the preposition *hasta*
- 1501: strict policy of intolerance & ban on the use of Arabic.

Contact with Portuguese

The Iberian Union: between 1580 and 1640. Dynastic Union of the Monarchy of Spain and the Kingdom of Portugal.

Change in Portuguese: DOM

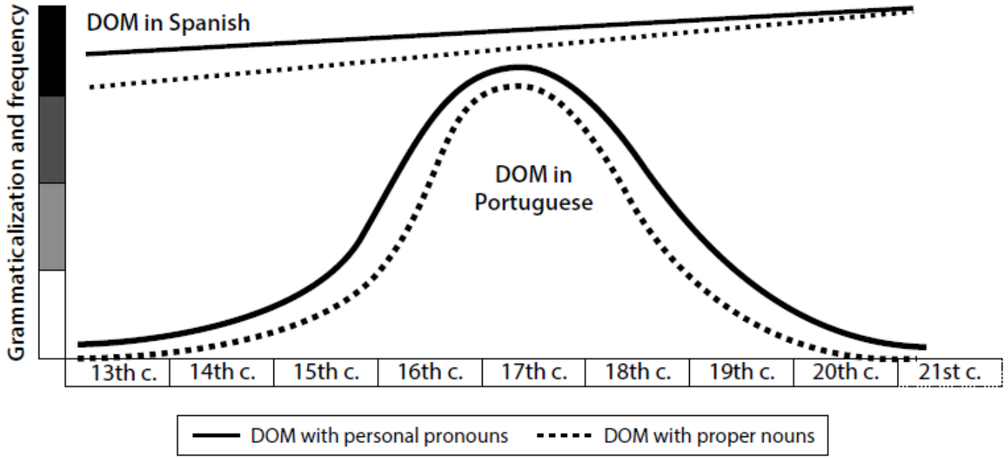
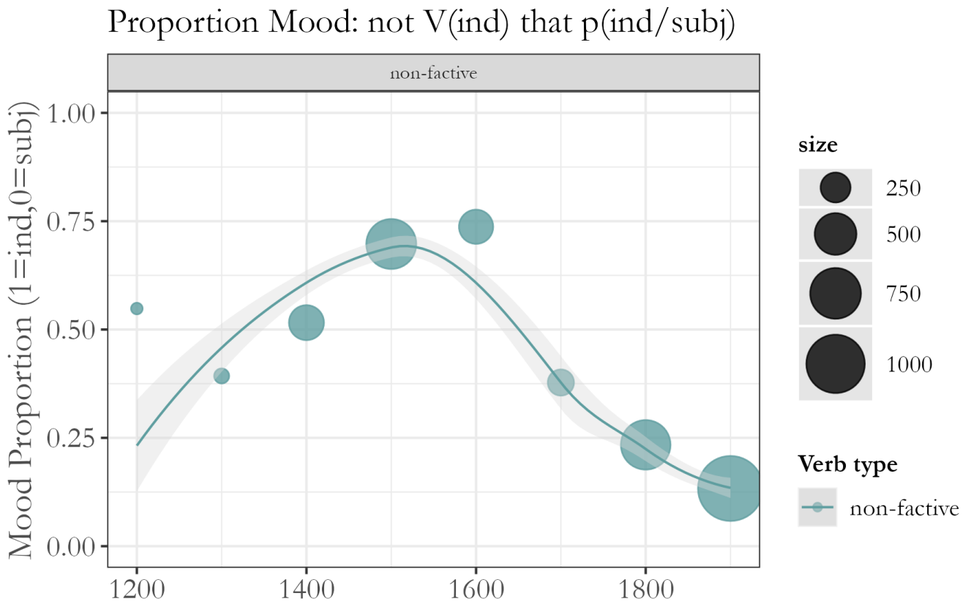


Figure 1. Grammaticalization and frequency of DOM in Spanish and Portuguese concerning personal pronouns and proper nouns.

It has been argued that the Union affected DOM in Portuguese (Döhla 2014)

Change in Spanish: PS

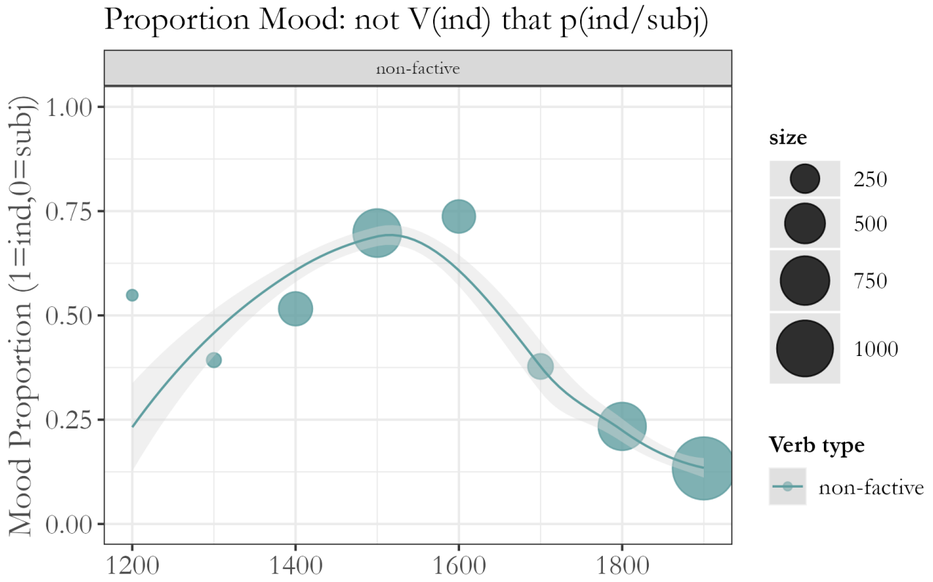


It may have had an impact on the Spanish language

Contact Explanation

The increase in the use of indicative during the period before 1700 may have been due to the presence of L2 speakers: andalusi arabic, andalusi romance, portuguese, etc.

Then, the number of L2 speakers in the population decreased and as the system went back to its original state.



Conclusion

- Part I. The advantages of using corpus data to test Trudgill's conjecture
  a. Advantage 1. Better understand WHERE and WHEN language change occurs.
  b. Advantage 2. Measure quantitatively how language change happens.
- Part II. The advantages of combining multiple approaches to tackle contact induced language change
  a. Advantage 1. More fine grained picture of language contact outcomes.
  b. Advantage 2. Explain cases in which despite contact the L2 difficult feature is preserved.

Thank you for listening!

molly.rolf@uni-konstanz.de

raquel.montero-estebaranz@uni-konstanz.de

gemma-hunter.mccarley@uni-konstanz.de

henri.kauhanen@uni-konstanz.de

george.walkden@uni-konstanz.de

https://www.ling.uni-konstanz.de/en/walkden/starfish/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 851423

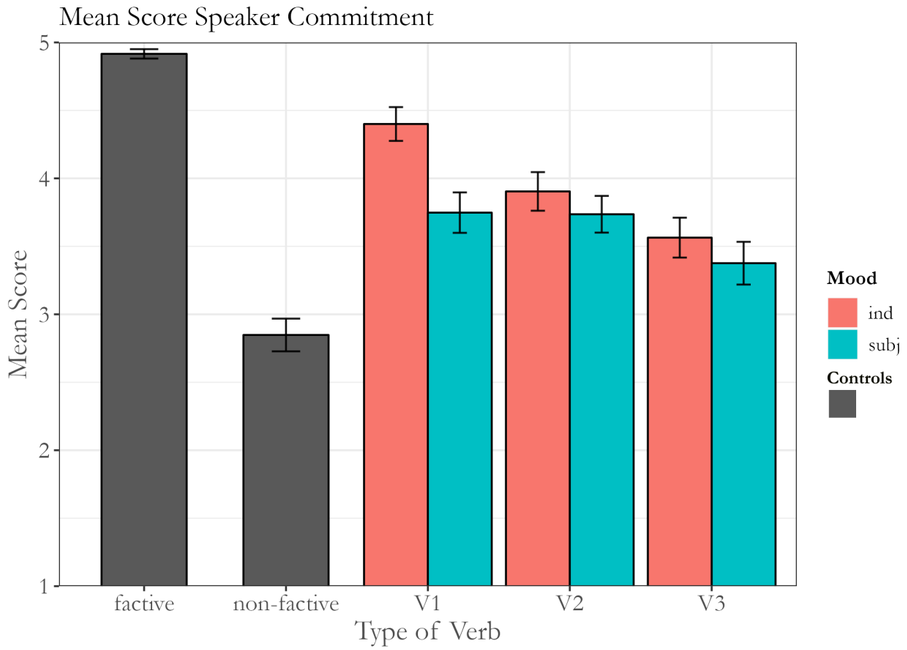
Universität Konstanz

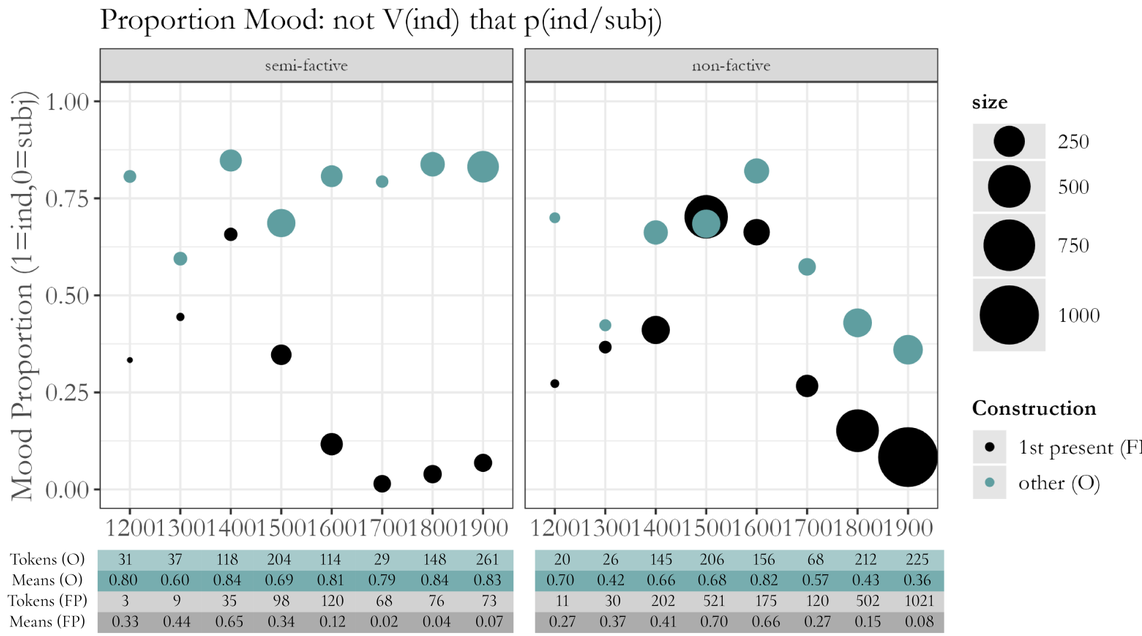
Funded by the European Union

European Research Council
Established by the European Commission

STARFISH

Appendix: verb classes

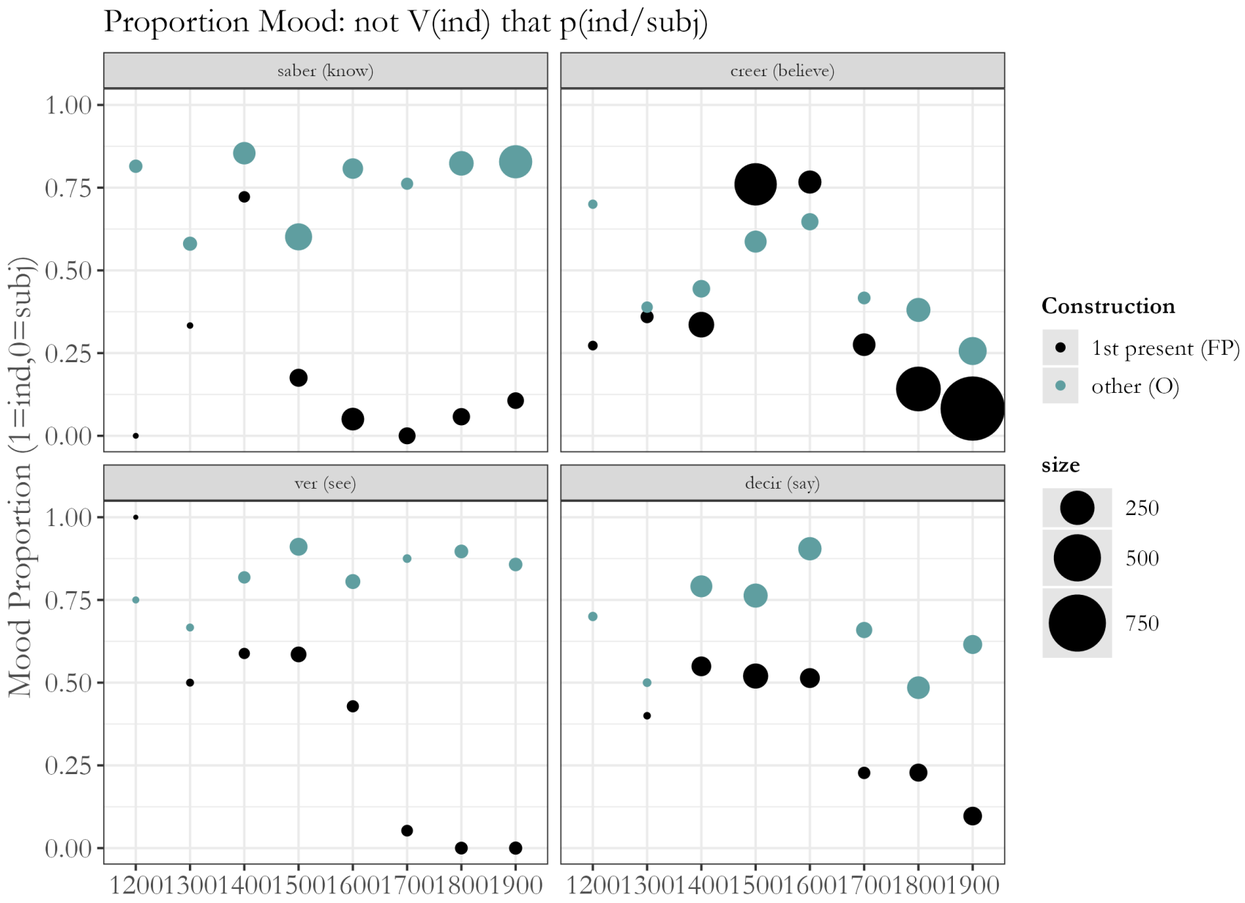




	semi-factive							
Tokens (O)	31	37	118	204	114	29	148	261
Means (O)	0.80	0.60	0.84	0.69	0.81	0.79	0.84	0.83
Tokens (FP)	3	9	35	98	120	68	76	73
Means (FP)	0.33	0.44	0.65	0.34	0.12	0.02	0.04	0.07

	non-factive							
Tokens (O)	20	26	145	206	156	68	212	225
Means (O)	0.70	0.42	0.66	0.68	0.82	0.57	0.43	0.36
Tokens (FP)	11	30	202	521	175	120	502	1021
Means (FP)	0.27	0.37	0.41	0.70	0.66	0.27	0.15	0.08

Appendix: individual predicates



Appendix: PS in Portuguese

They don’t believe that it is raining

a. Eles não acreditam que está.IND a chover!

b. Eles não acreditam que esteja.SUBJ a chover!

Part II References

Atkinson, Mark, Kenny Smith, and Simon Kirby. 2018. "Adult Learning and Language Simplification." *Cognitive Science* 42 (8): 2818–54.

Bentz, Christian, and Bodo Winter. 2013. "Languages with More Second Language Learners Tend to Lose Nominal Case." *Language Dynamics and Change* 3 (1): 1–27.

Berdicevskis, Aleksandrs, and Arturs Semenuks. 2022. "Imperfect Language Learning Reduces Morphological Overspecification: Experimental Evidence." *PLoS One* 17 (1): e0262876.

Borgonovo, Claudia, and Philippe Prévost. 2003. "Knowledge of Polarity Subjunctive in L2 Spanish." In *Proceedings of the 27th Boston University Conference on Language Development*, 1:150–61. Cascadilla Press Somerville, MA.

Bush, John WM, Jeffrey M Aristoff, and AE Hosoi. 2006. "An Experimental Investigation of the Stability of the Circular Hydraulic Jump." *Journal of Fluid Mechanics* 558: 33–52.

Döhla, Hans-Jörg. 2014. "Diachronic Convergence and Divergence in Differential Object Marking Between Spanish and Portuguese." *Stability and Divergence in Language Contact. Factors and Mechanisms. Amsterdam/Philadelphia: John Benjamins*, 265–89.

Frigg, Roman, and Stephan Hartmann. 2006. "Models in Science." https://plato.stanford.edu/entries/models-science/.

Iverson, Michael, Paula Kempchinsky, and Jason Rothman. 2008. "Interface Vulnerability and Knowledge of the Subjunctive/Indicative Distinction with Negated Epistemic Predicates in L2 Spanish." *Eurosla Yearbook* 8 (1): 135–63.

Kauhanen, Henri. 2022. "A Bifurcation Threshold for Contact-Induced Language Change." *Lossa: A Journal of General Linguistics* 7 (1).

Lupyan, Gary, and Rick Dale. 2010. "Language Structure Is Partly Determined by Social Structure." *PloS One* 5 (1): e8559.

Quer, Josep. 1998. "Mood at the Interface." PhD thesis, Universiteit Utrecht.

Raviv, Limor, Antje Meyer, and Shiri Lev-Ari. 2019. "Larger Communities Create More Systematic Languages." *Proceedings of the Royal Society B* 286 (1907): 20191262.

Ruhstaller, Stefan, and María Dolores Gordón Peral. 2017. "The Lexical Impact of Language Contact with Arabic on Spanish and Catalan." *Lexicographica* 33 (2017): 277–96.

Sinnemäki, Kaius, and Francesca Di Garbo. 2018. "Language Structures May Adapt to the Sociolinguistic Environment, but It Matters What and How You Count: A Typological Study of Verbal and Nominal Complexity." *Frontiers in Psychology* 9: 342569.

Trudgill, Peter. 2010. “Contact and Sociolinguistic Typology.” *The Handbook of Language Contact*, 299–319.
Vicente, Ángeles. 2020. “Andalusi Arabic.” *Arabic and Contact-Induced Change*, 225.
Walkden, George, and Anne Breitbarth. 2019. “Complexity as L2-Difficulty: Implications for Syntactic Change.” *Theoretical Linguistics* 45 (3-4): 183–209.
Yang, Charles D. 2000. “Internal and External Forces in Language Change.” *Language Variation and Change* 12 (3): 231–50.