

Grammaticalization paths and clines: Induction and visualization

Natalia Levshina

Leipzig University

Workshop “Parallel text analysis in diachronic research”

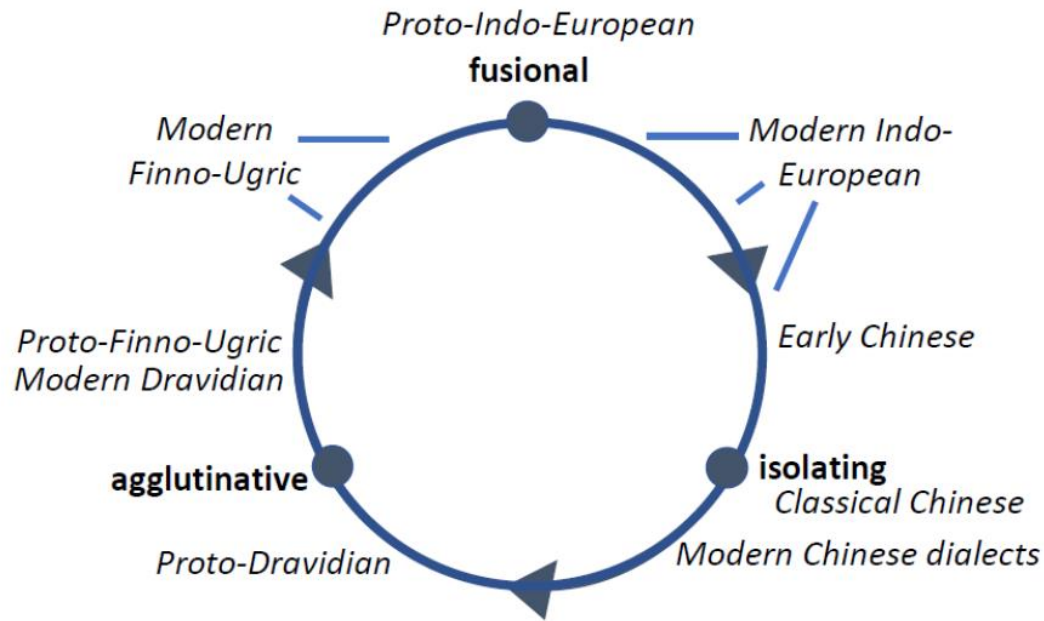
Marburg, 23 February 2018

Grammaticalization paths, scales, clines, etc.

- Global ones, e.g. grammaticalization scales

LEVEL	discourse		syntax		morphology		morpho- phonemics	
TECHNIQUE	isolating	>	analytic	>	synthetic- agglutinating	>	synthetic- flexional	> zero
PHASE		↑		↑		↑		↑
			syntacticization		morphologization		demorphemicization	loss
PROCESS			grammaticalization					

Cycle of language change



Local scales

- Germanic sandwich (English > Dutch > German)
- Romance sandwich (French > Italian > Spanish)

Outline of the talk

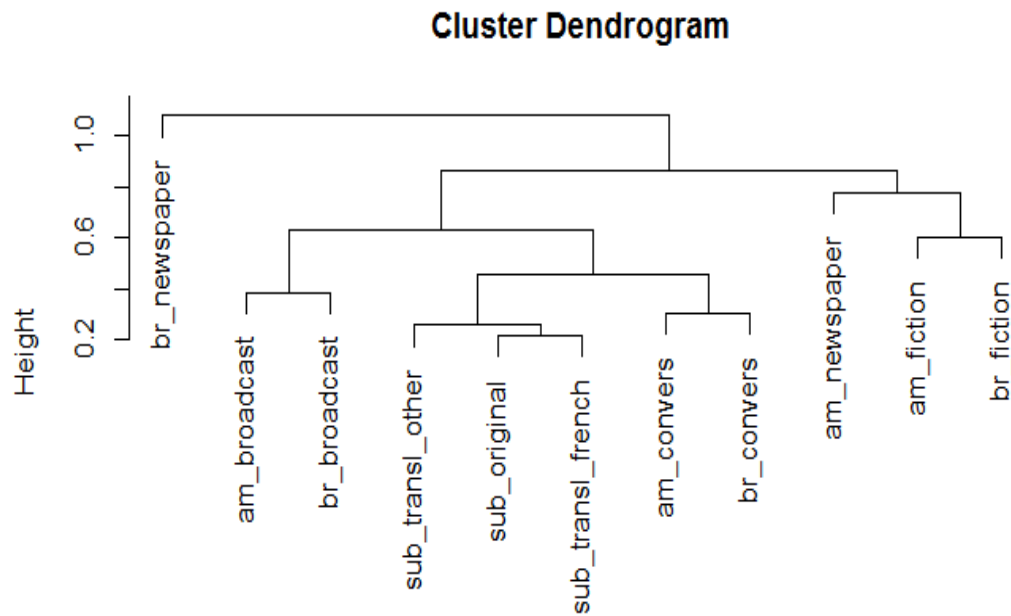
- Local cline:
 - Grammaticalization of **causative auxiliaries MAKE** in Romance languages

- Global cline:
 - Grammaticalization of **case markers** (based on English, Finnish, German, Russian and Turkish)

ParTy corpus

- Corpus of subtitles (films and TED talks)
- Minimum 10 languages per film/unit
- Sentences in English versions aligned with sentences in other languages, pairwise
- Partly available from <https://github.com/levshina/ParTy-1.0>

Why subtitles?



Based on the frequencies of 3-grams (Levshina 2017)





Short
8 oz.



Tall
12 oz.



Grande
16 oz.



Venti
20 oz.

A local cline:

Grammaticalization of causal auxiliaries **MAKE**
in Romance

Analytic causatives as comparative concepts

FUNCTION: An AC designates a causative event, which involves a causing event (or state) and a caused event (or state), and their participants, most importantly, the causer and the causee. The causer initiates or is responsible for the causing event, whereas the causee is the entity that brings about the caused event (state). There can be also other participants involved (such as the affectee, i.e. the final affected entity). The causing event is underspecified.

FORM: An AC is a construction that consists of two VERBS and their arguments. One VERB (V1) represents in an abstract way the causing event, whereas the other VERB (V2) represents the caused event. The order of the predicates may vary. The clauses should be closely integrated: at least some arguments of V2 should be grammatically dependent on V1.

Examples of ACs

- Don't **make** me cry.
- **Let** my people go.
- You're **forcing** me to be the voice of reason.
- 6 careers that **allow** to you to travel around the world.

Films



Dataset

- Translations in 18 European languages (15 Indo-European and 3 Finno-Ugrian languages)
- All ACs extracted manually from each doculect.
- 392 contexts with at least one language having an AC

Method

- Distance matrix based on weighted features: auxiliary (the greatest weight), form of effected predicate, transitivity, reflexivity, active/passive.
- Multi-Dimensional Scaling (cf. Wälchli & Cysouw 2012) with smacof
- An interactive plot with googleVis:
<http://www.natalialevshina.com/presentations.html>

Zooming in on Romance ACs

- ita: *fare* + Vinf
- fra: *faire* + Vinf
- spa: *hacer* + (NP) + Vinf
- por: *fazer* + (NP) + Vinf/Vinf_inflected
- rom: *a face* + *să* + Vsubj

Examples

- French, Amélie

*Amandine Poulain aime: (...) **Faire** briller le parquet avec des patins...*

Amandine Poulain likes: (...) polishing the parquet with slippers...

- Italian, Avatar

*Stronzate, **fammi** vedere!*

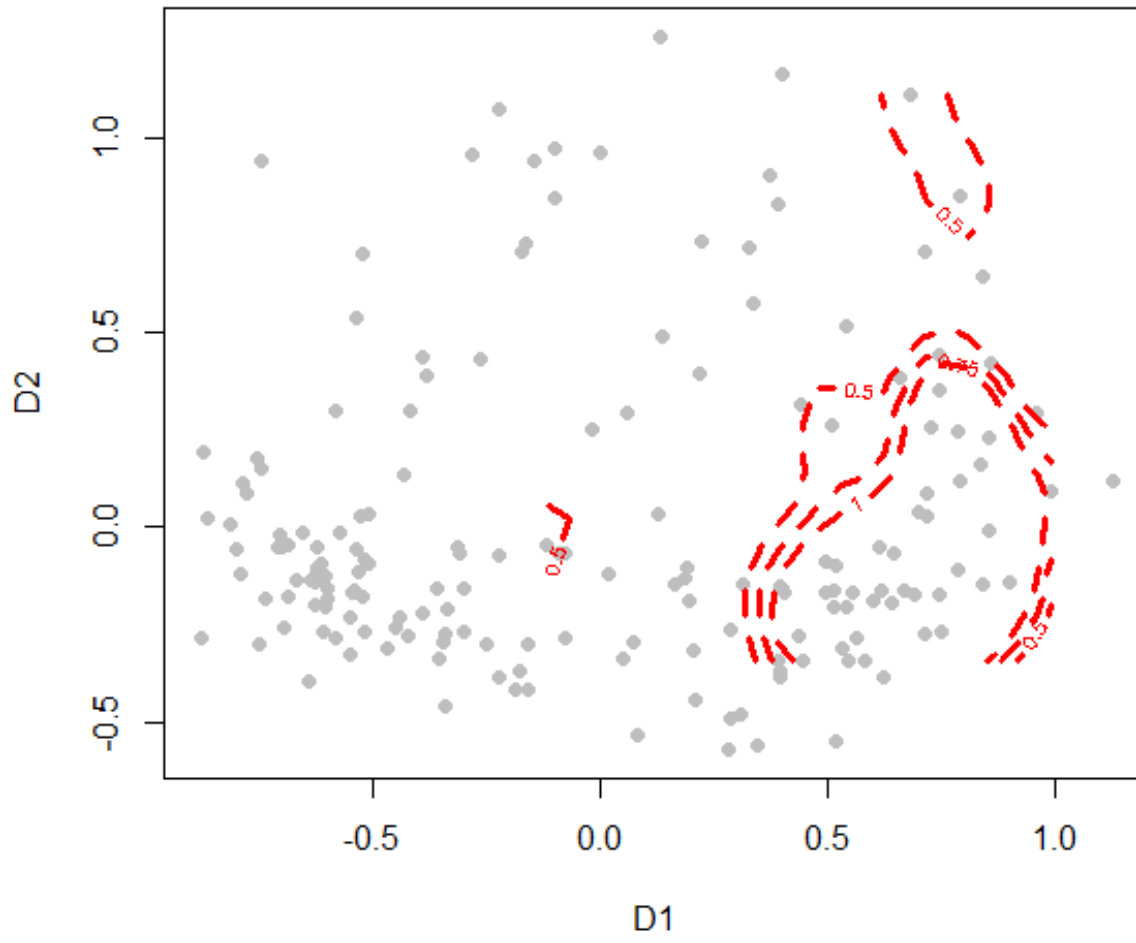
Bullshit make.me see

Bullshit, let me see that!

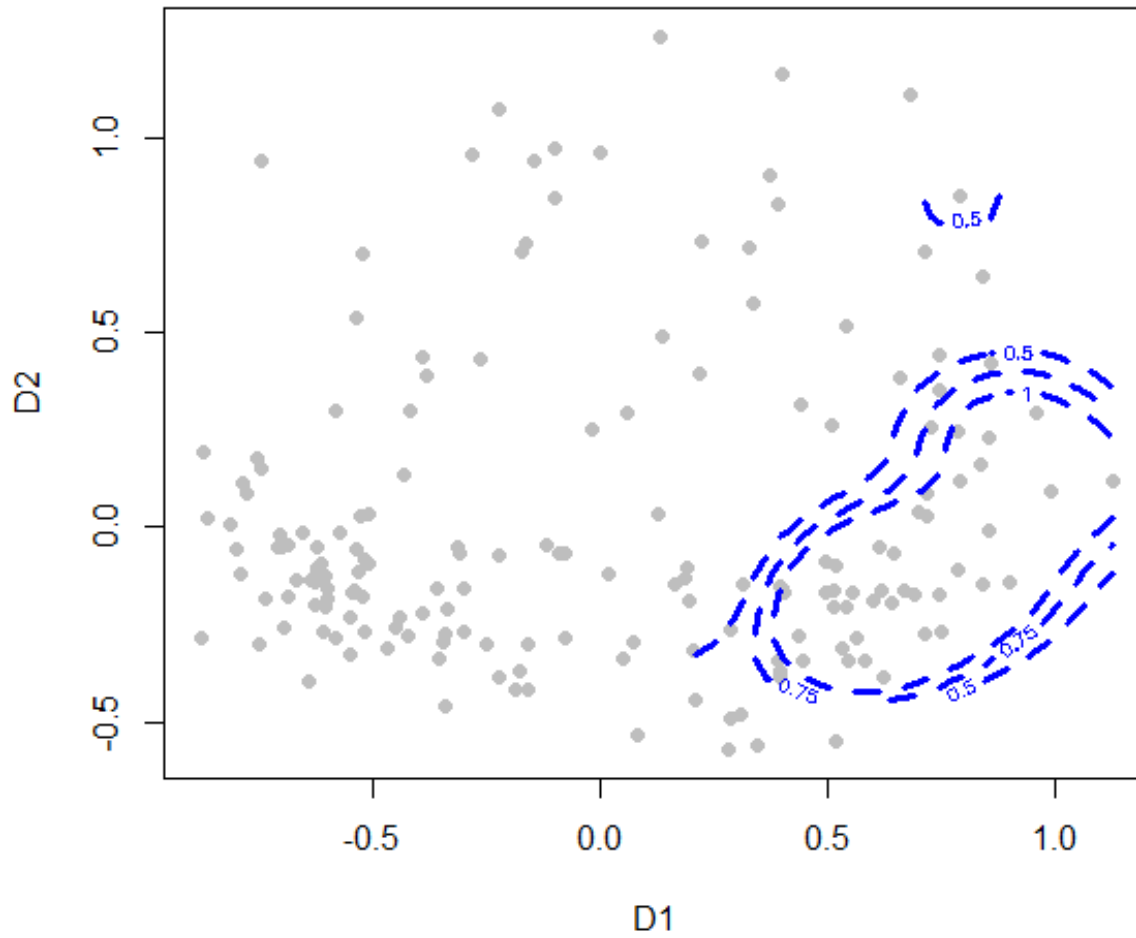
Romance sandwich(es)

- Lamiroy 2011:
 - French > Italian > Spanish
 - Based on a large number of diverse phenomena
- Soares da Silva 2012:
 - Italian > French > Spanish > Portuguese
 - Based on causative auxiliaries only

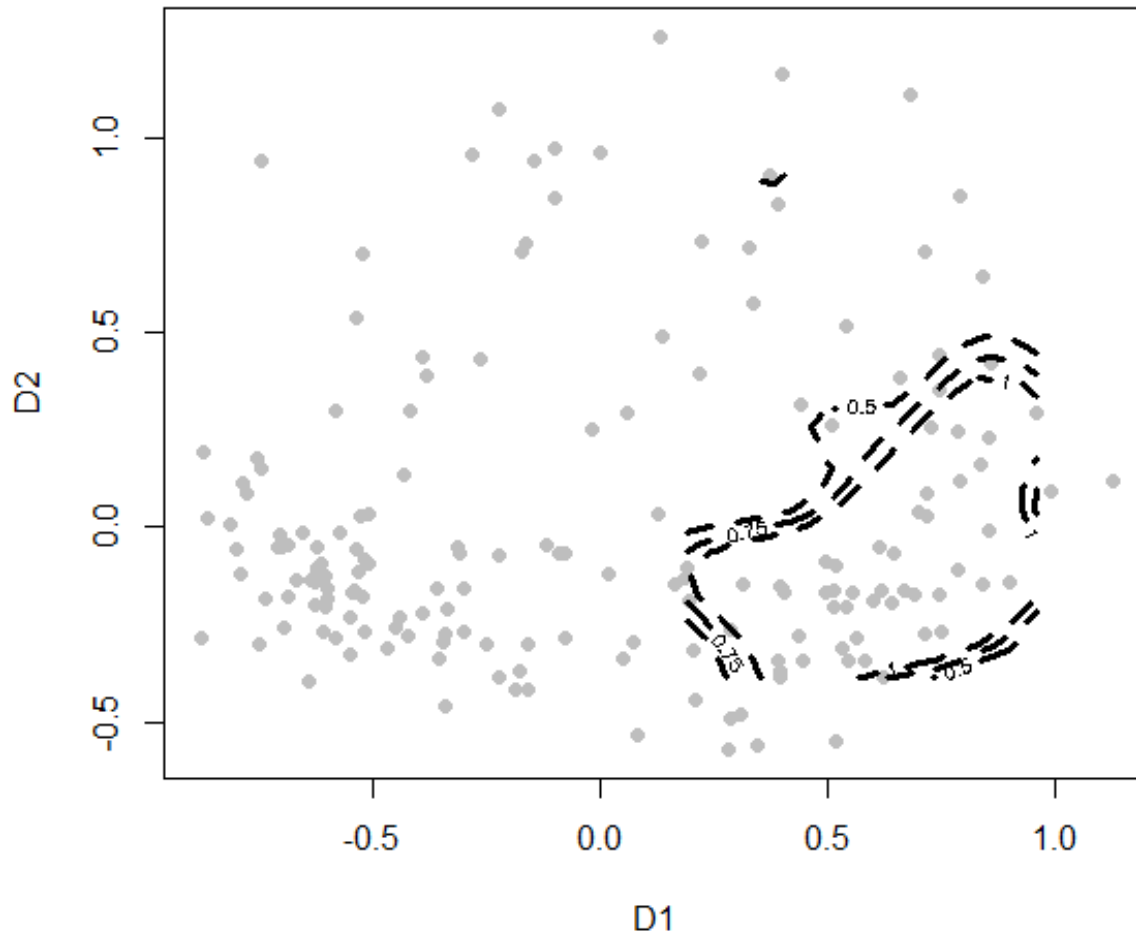
Romanian



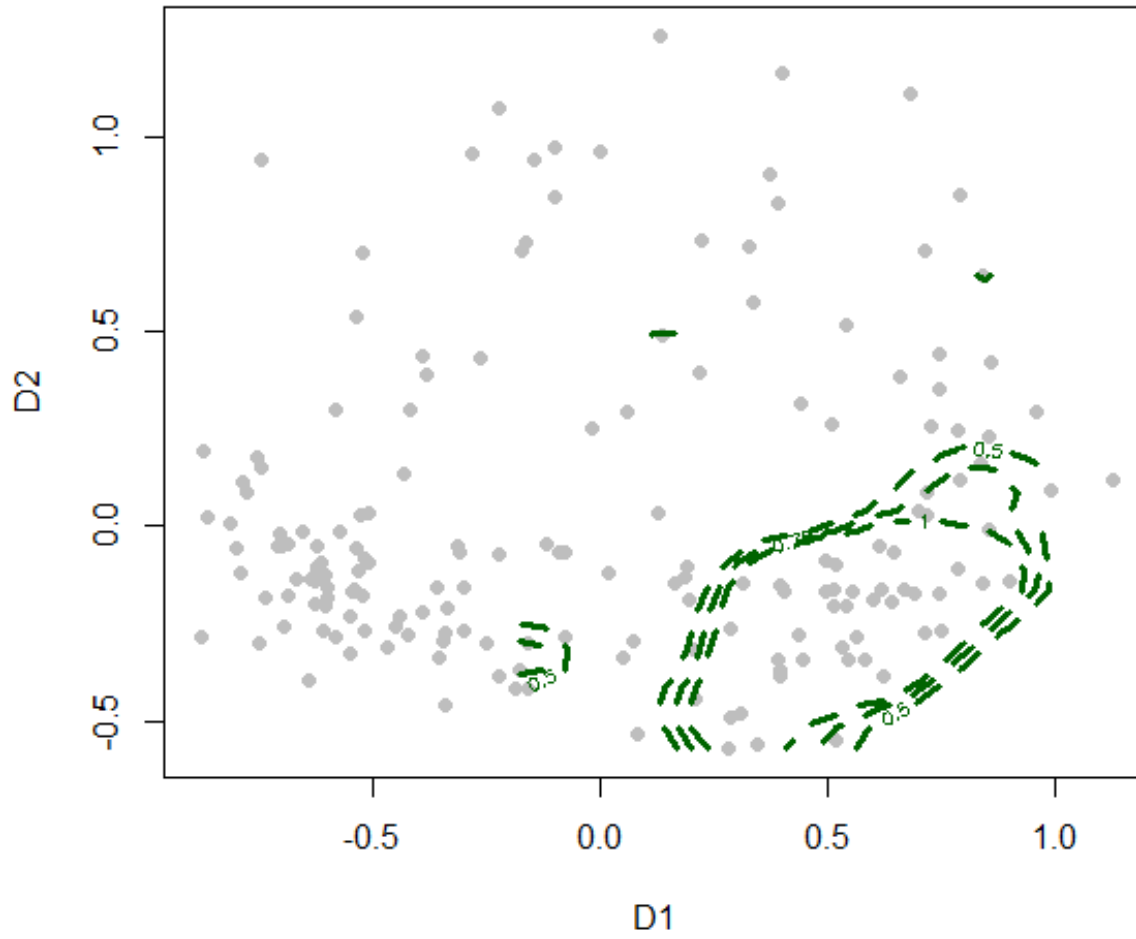
Portuguese



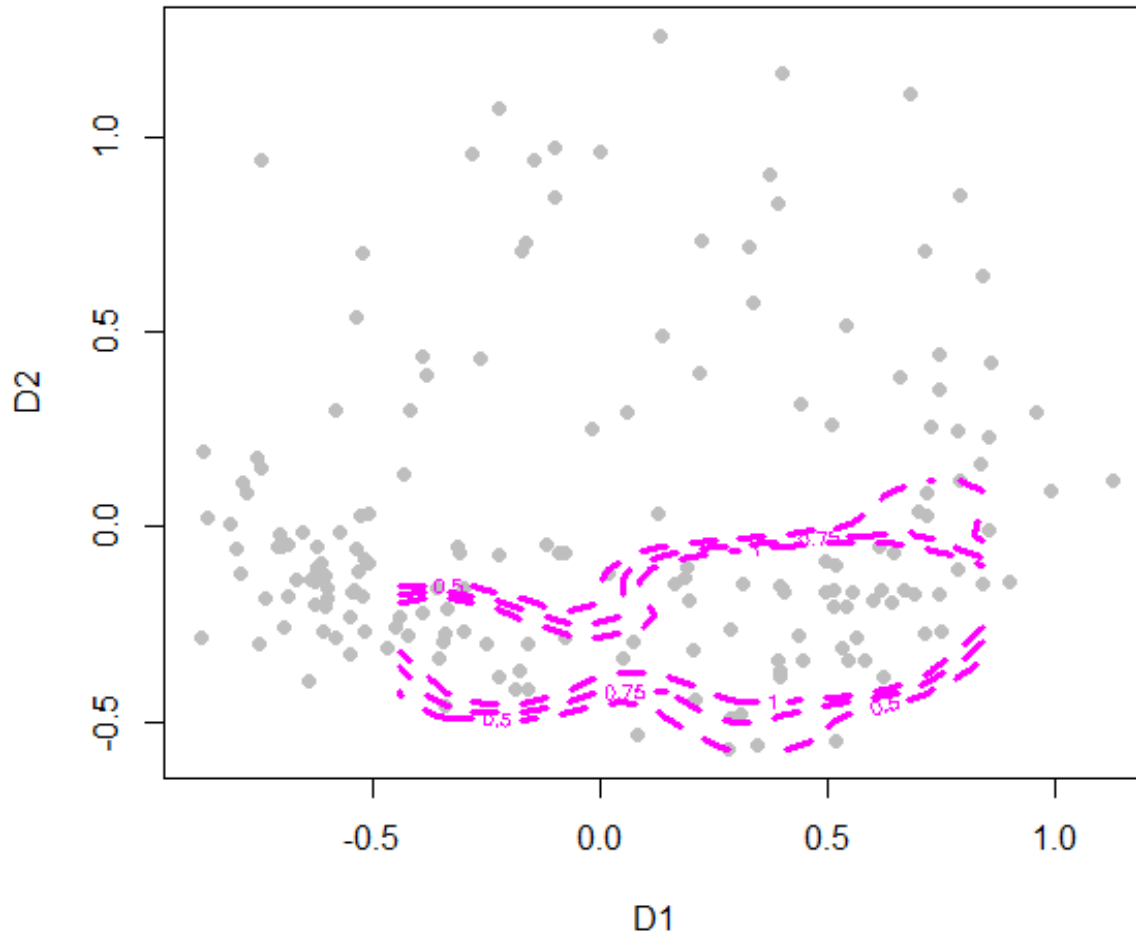
Spanish



French



Italian



Summary

Romanian < Portuguese < Spanish < French < Italian

- Semantics:



Coercion
Lexical

Abstract causation
Grammatical

- Interesting: an iconic correlation in syntax:

V *să* V_{subj}

V NP V_{inf}/V_{infl}

V (NP) V_{inf}

VV_{inf}

VV_{inf}



Autonomy

Integration

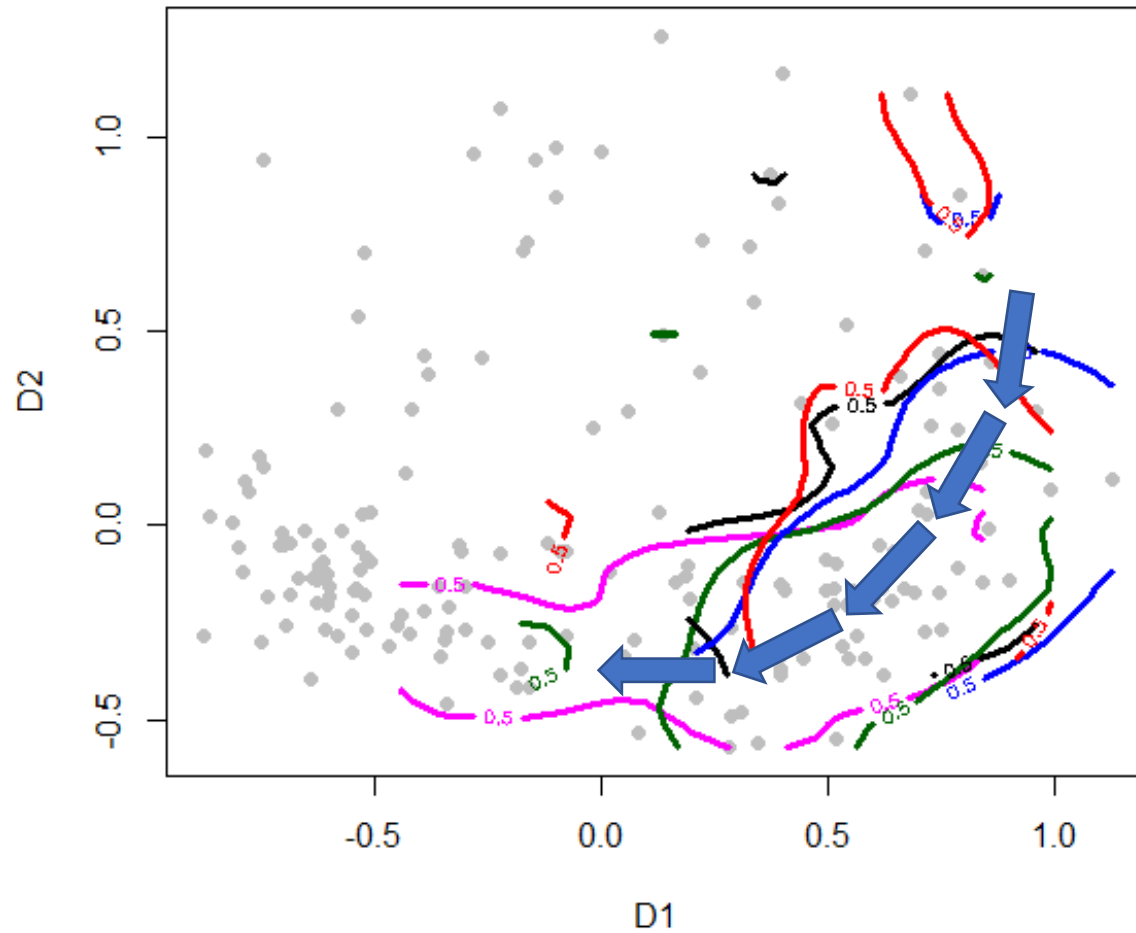
On semantic maps and diachrony

- “It should be stressed that the MDS method has been criticized because it cannot take into account diachronic information, if available (van der Auwera, 2008, 2013; Narrog, 2010). For example, there is no way to infer any directionality from Figure ...” (Georgakopoulos & Polis 2018)

On semantic maps and diachrony

- “It should be stressed that the MDS method has been criticized because it cannot take into account diachronic information, if available (van der Auwera, 2008, 2013; Narrog, 2010). For example, there is no way to infer any directionality from Figure ...” (Georgakopoulos & Polis 2018)
- Okay, but what about this?

Romance ACs



See also a different representation of diachronic information on MDS-based semantic maps in Cysouw & Forker 2009



A global cline:

Grammaticalization of case markers

Grammaticalization parameters

Parameter	Paradigmatic	Syntagmatic
Weight	Integrity	Structural scope
Cohesion	Paradigmaticity	Bondedness
Variability	Paradigmatic variability	Syntagmatic variability

Case hierarchy

- A hierarchy by Blake (2001):

nominative < accusative / ergative < genitive < dative < locative / prepositional < ablative / instrumental < *others*

- If a language has a case listed on the hierarchy, it will usually have at least one case from each position to the left. Thus if a language has a dative case it will have a genitive, an accusative or ergative or both, and a nominative.
 - E.g. German: nom < acc < gen < dat/obl

Case study

- Marking of different semantic roles of nouns
- Which roles are more grammaticalized than others?
- Onomasiological approach: from meaning to form
(we can't rely on language-specific categories!)

Data

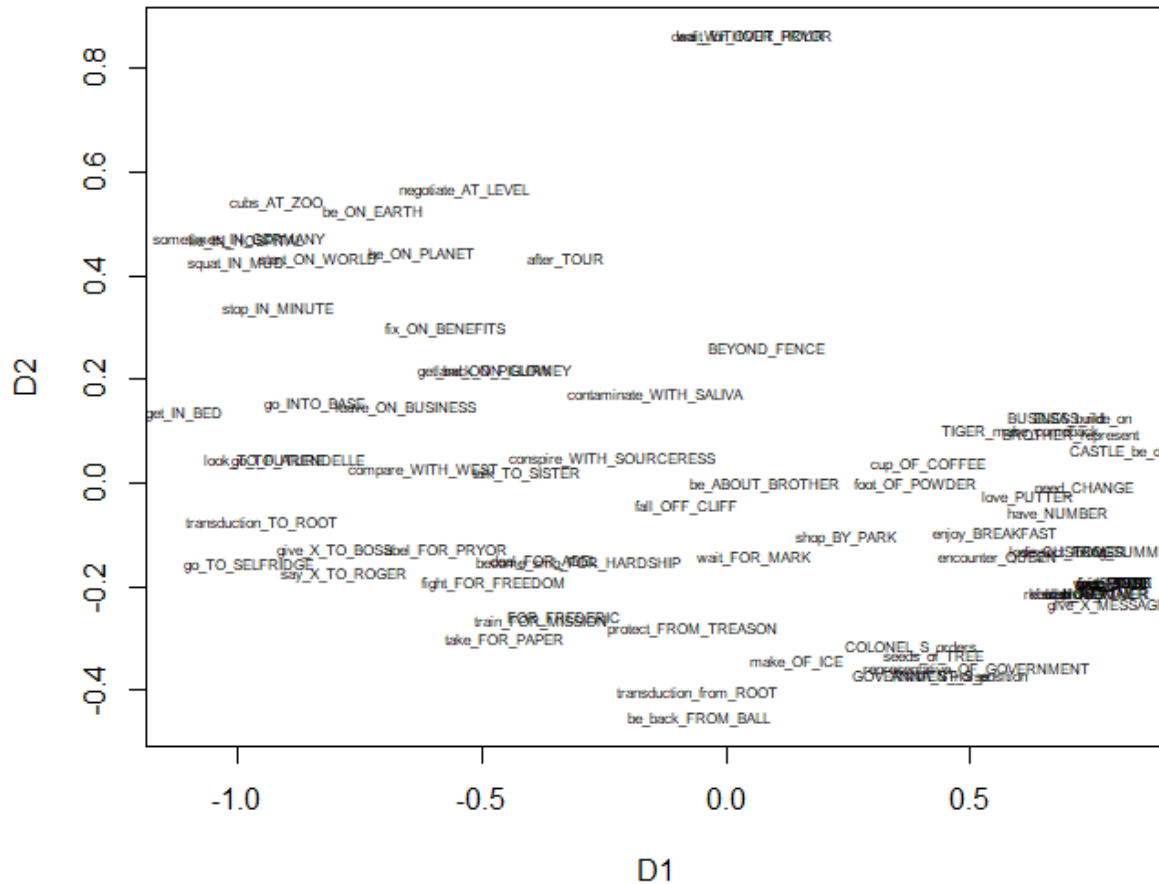
- eng, deu, fin, rus, tur
- ParTy corpus, 96 NPs, only SG



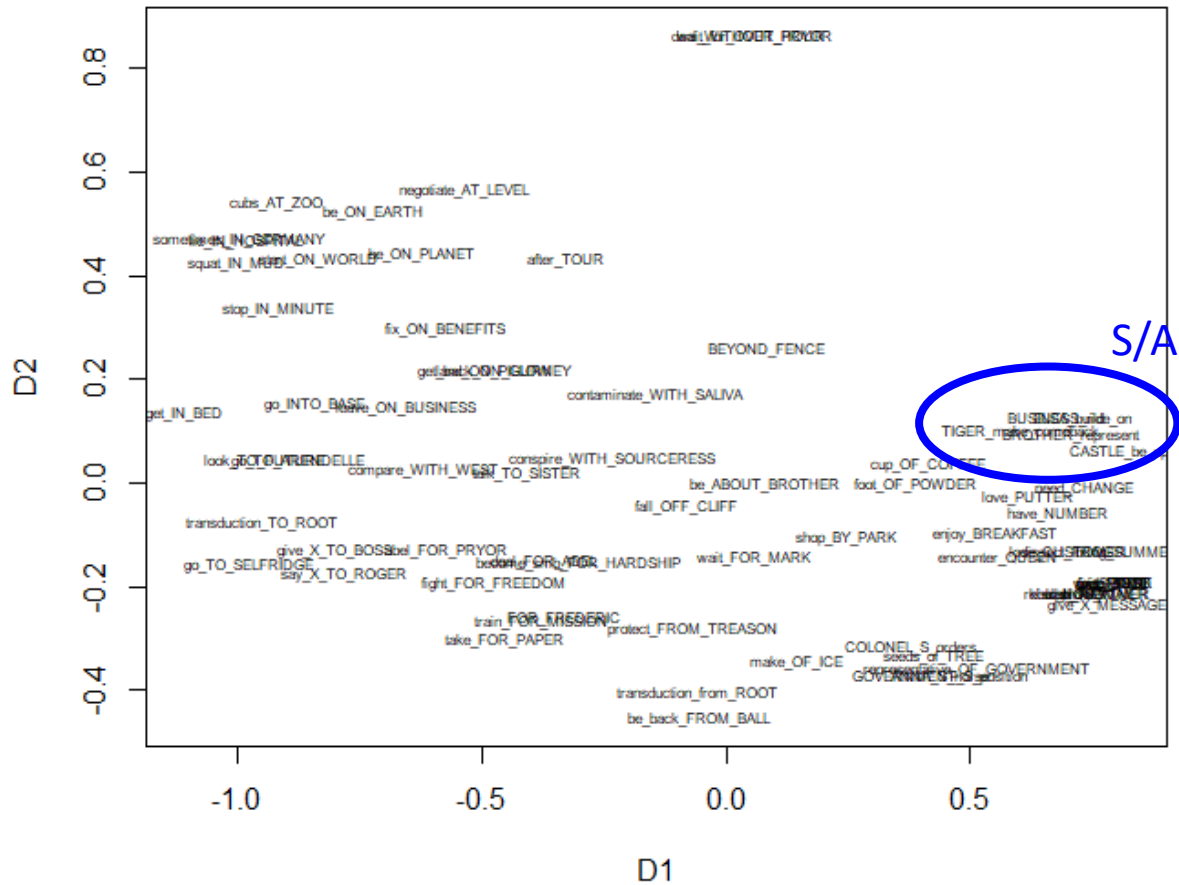
Examples of roles

- connect_TO_AVATAR
- come_out_FROM_LINK
- write_BOOK
- be_back_FROM_BALL
- train_FOR_MISSION
- dissect_FROG
- go_TO_SELFRIDGE
- love_PUTTER
- get_back_ON_GURNEY
- transduction_from_ROOT
- transduction_TO_ROOT
- contaminate_WITH_SALIVA
- COLONEL_S_orders
- hang_on_TILL_MORNING
- seeds_OF_TREE
- ELSA_build

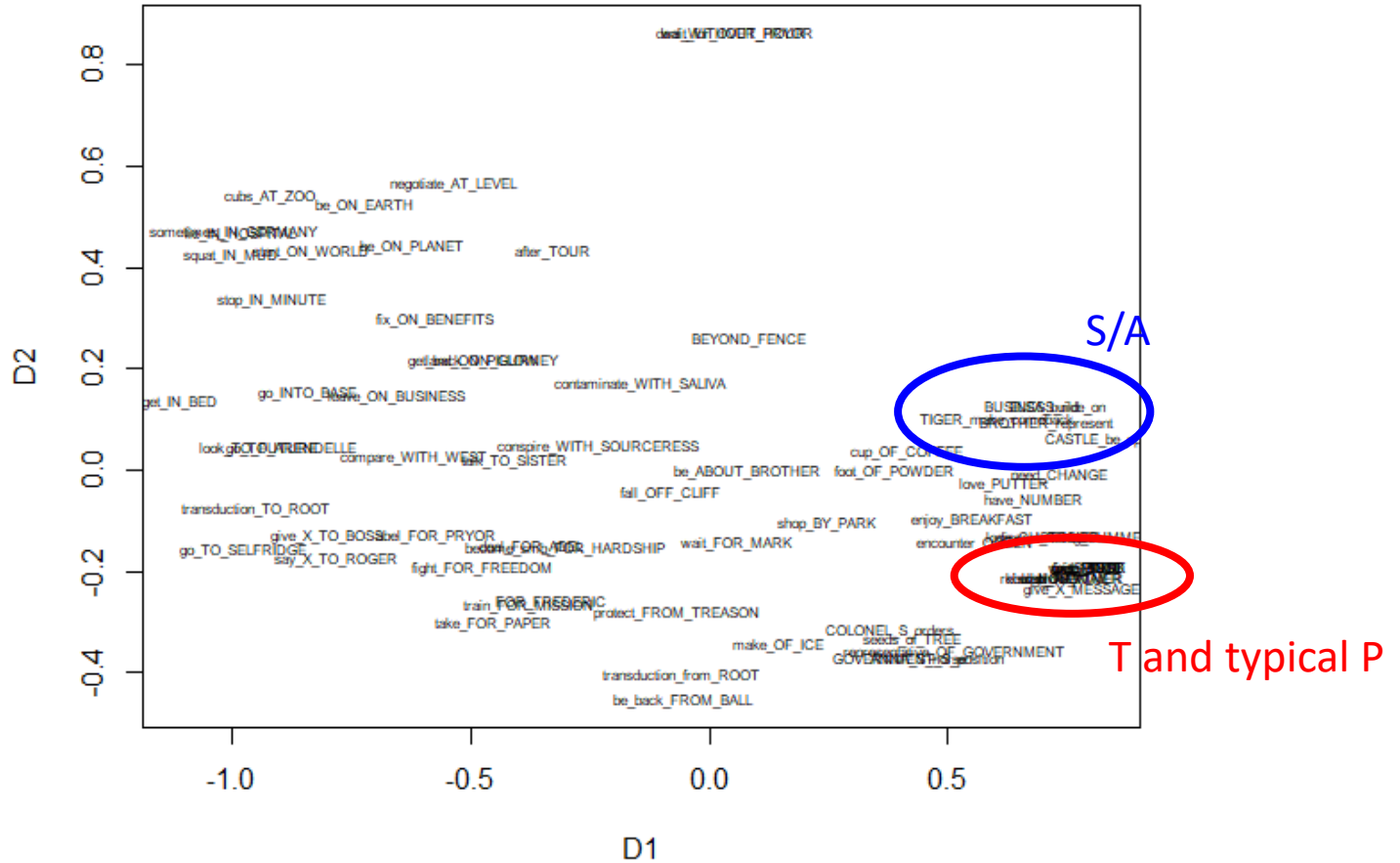
Ordinal MDS



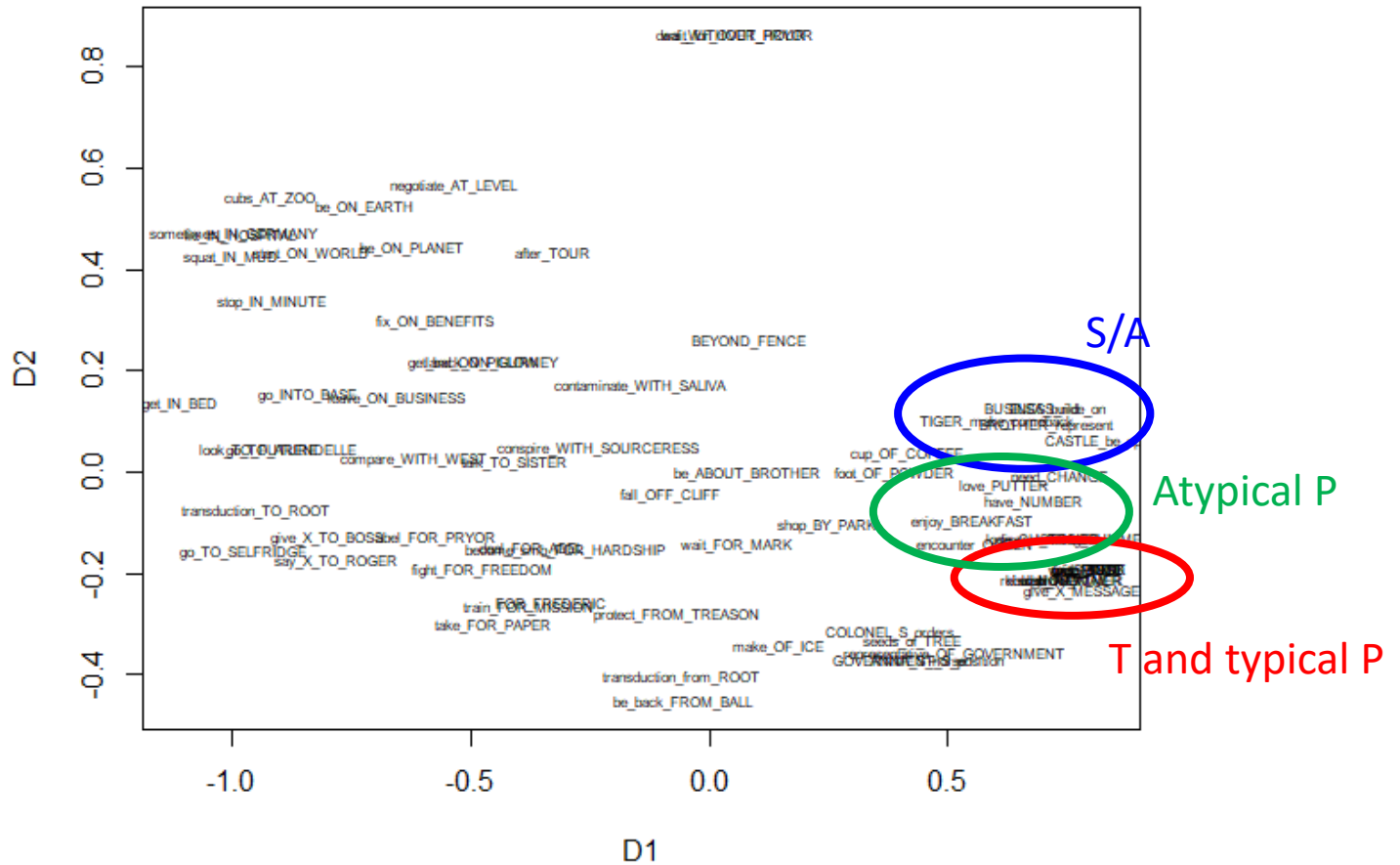
Roles



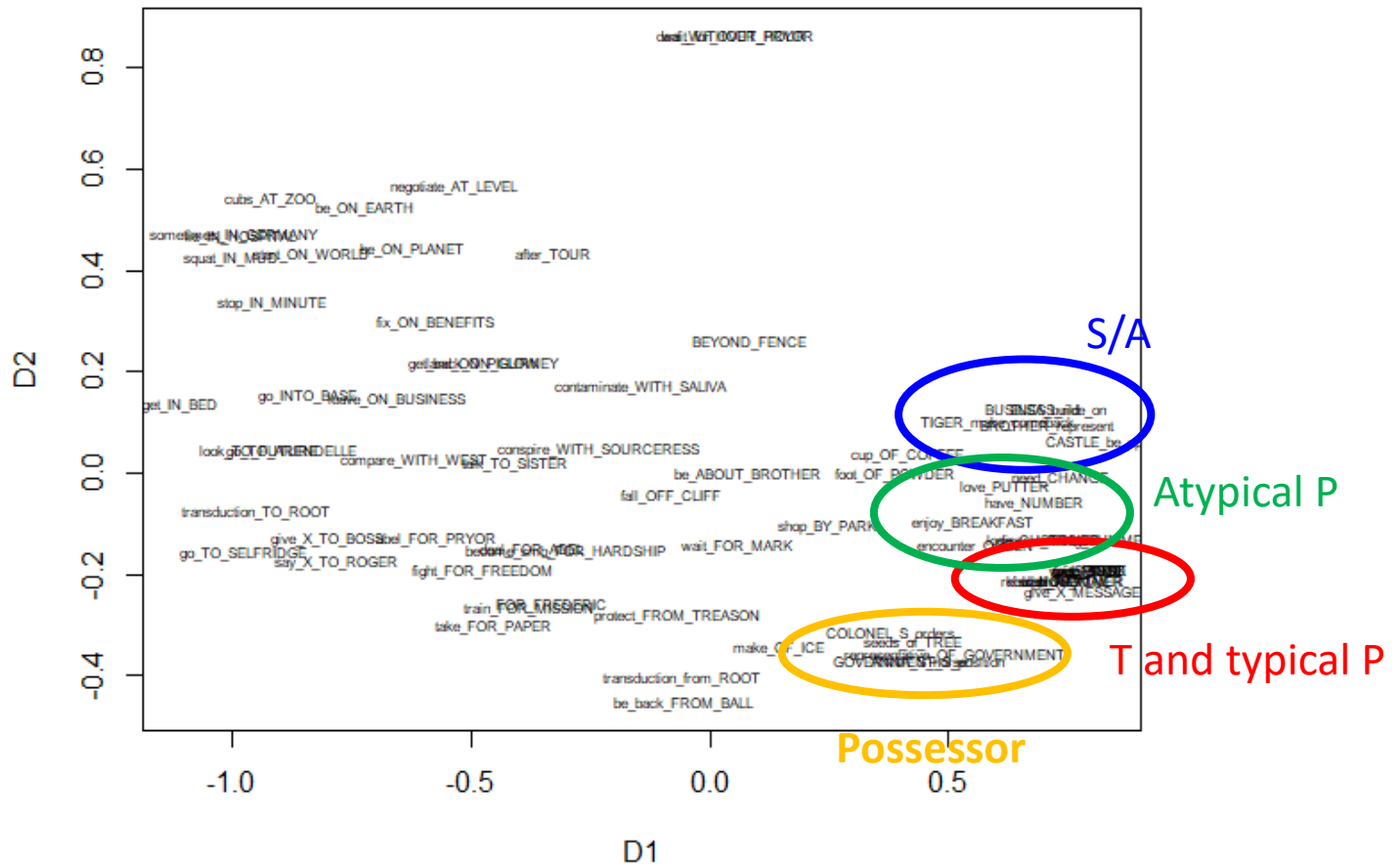
Roles



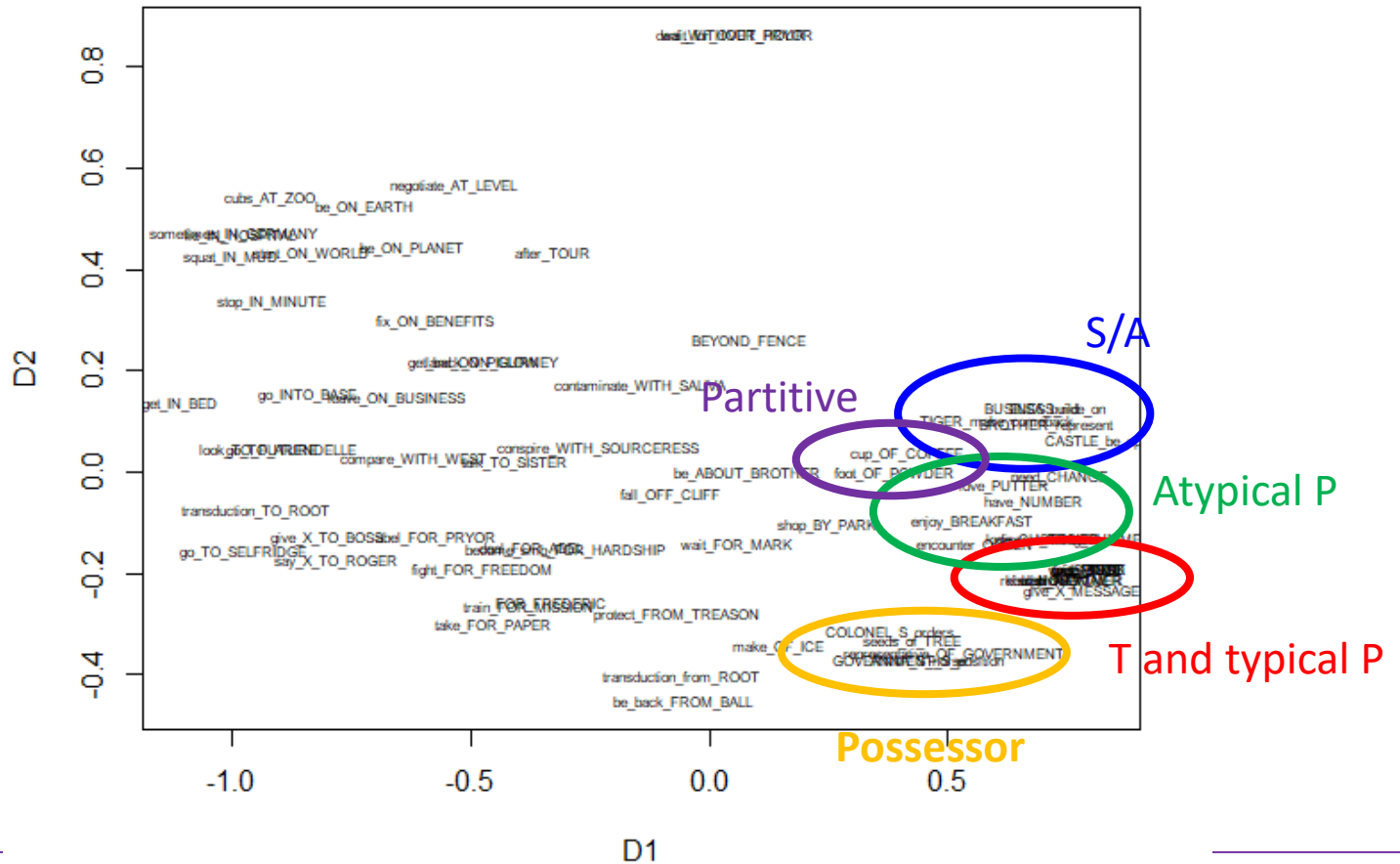
Roles



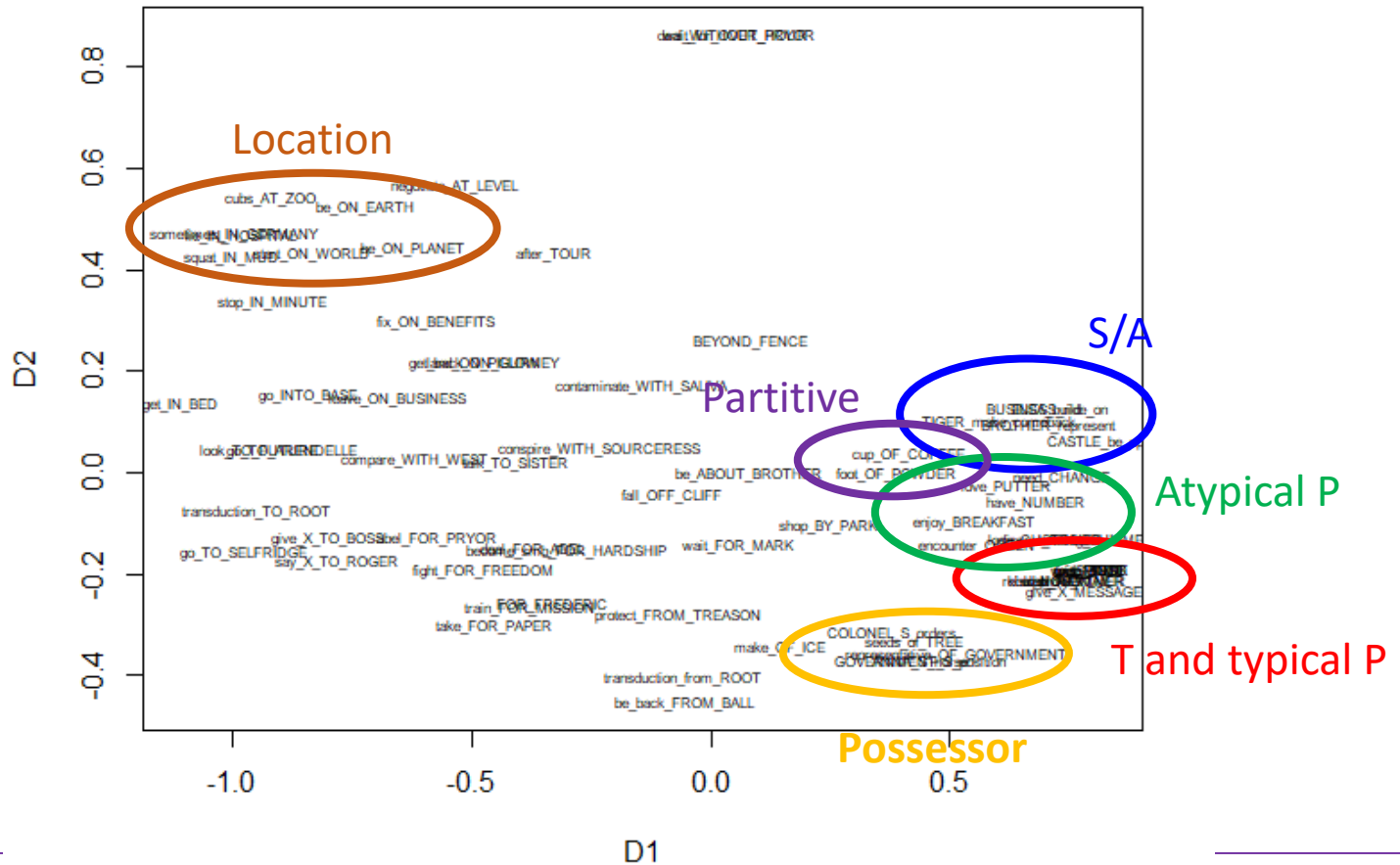
Roles



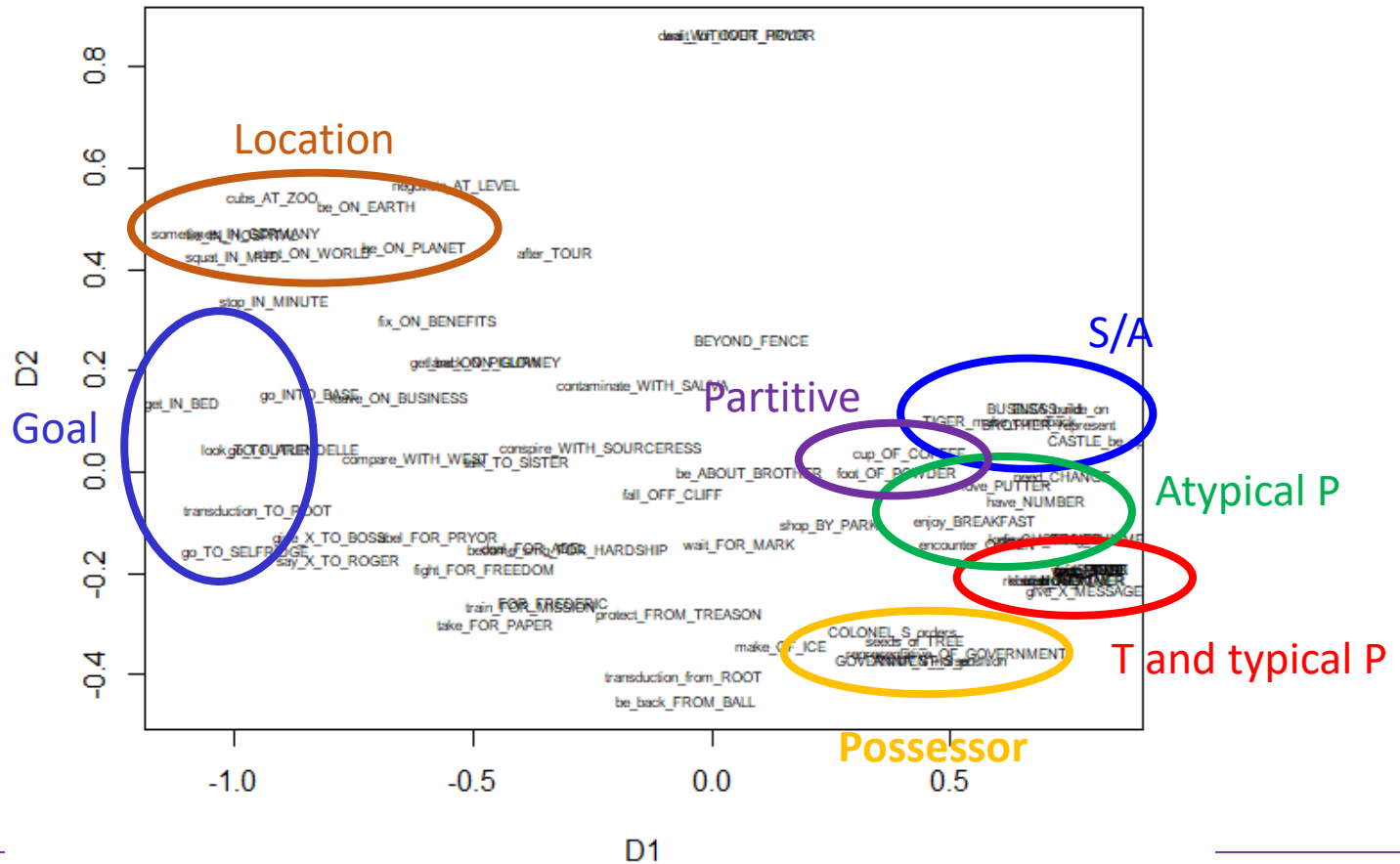
Roles



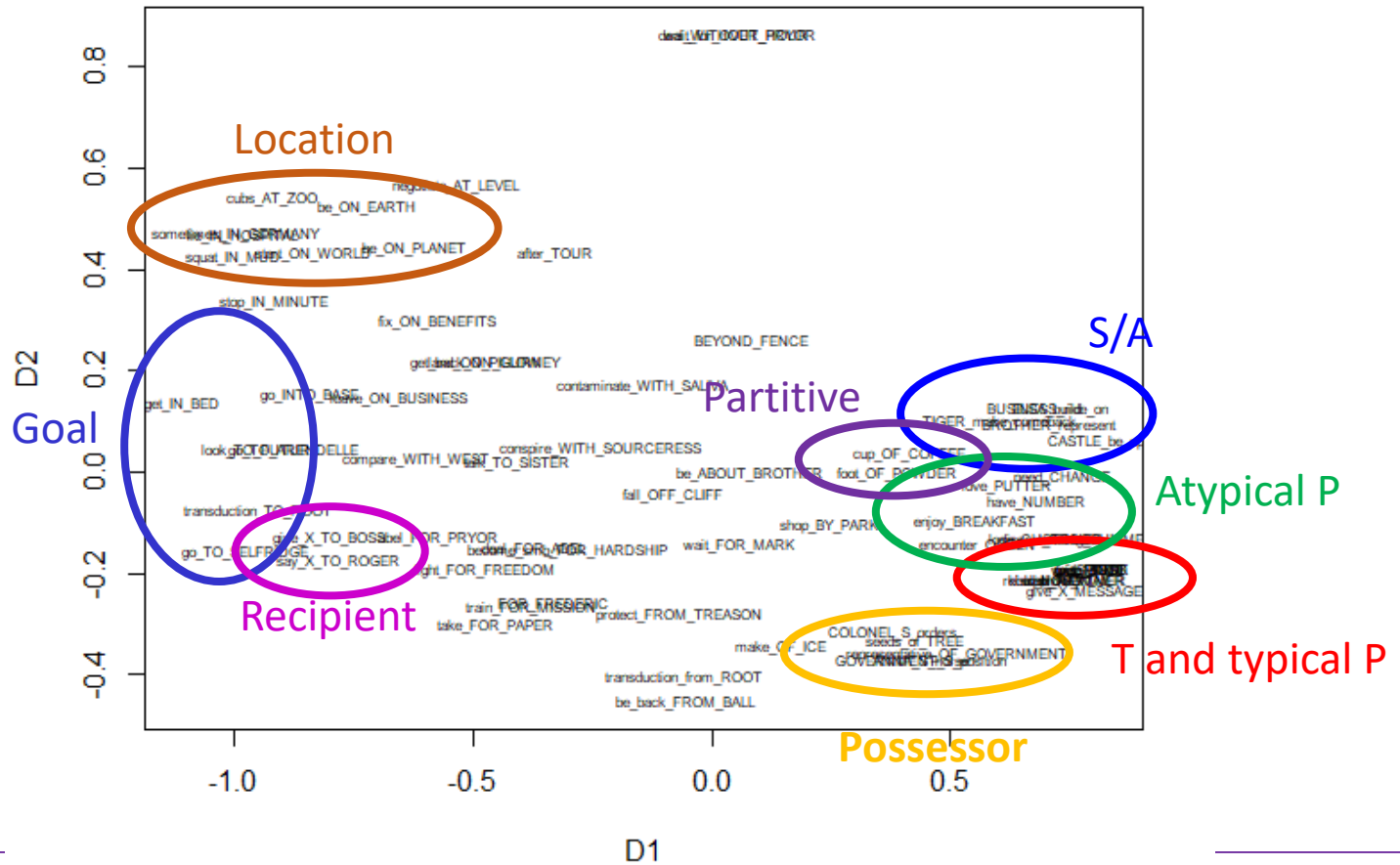
Roles



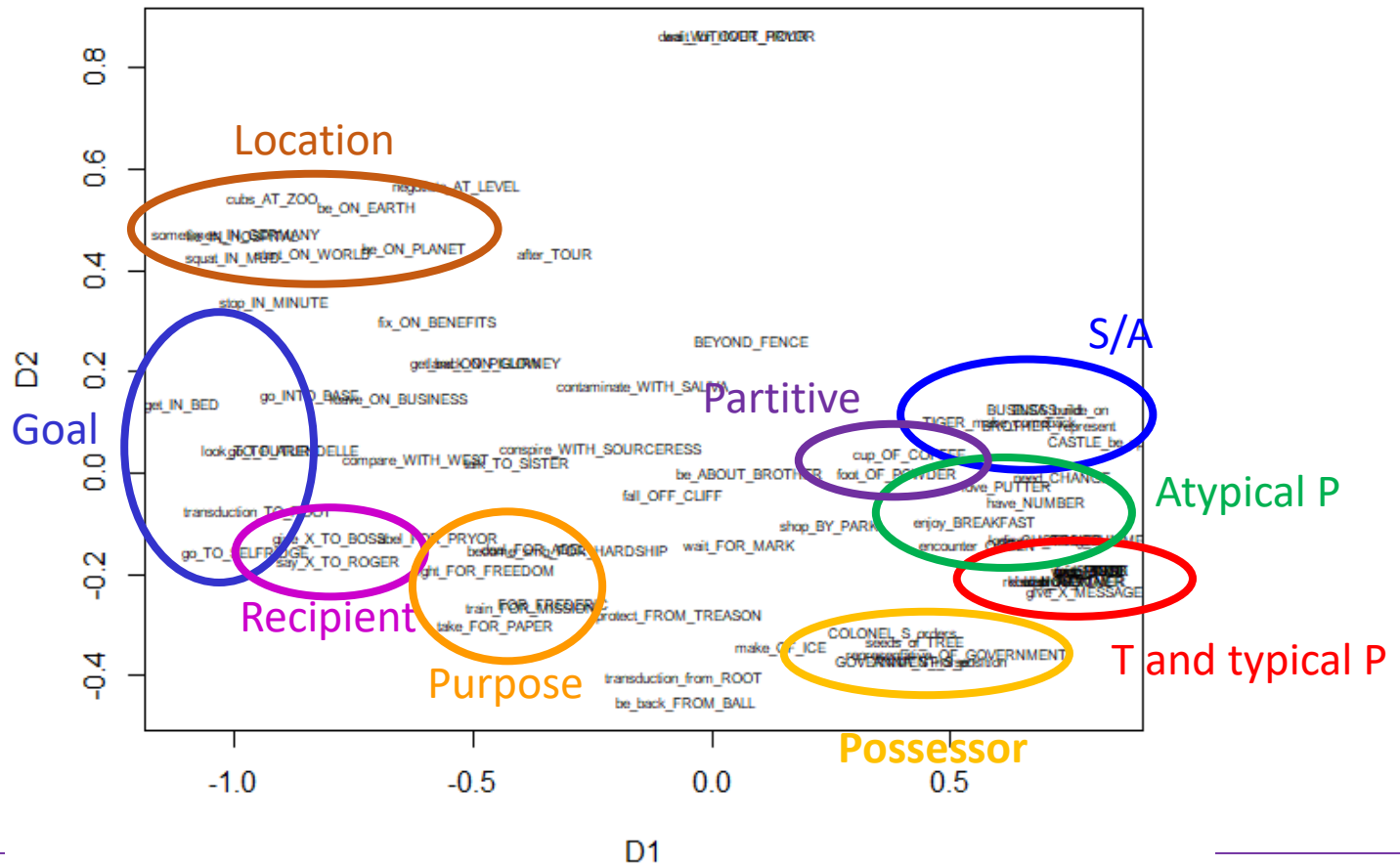
Roles



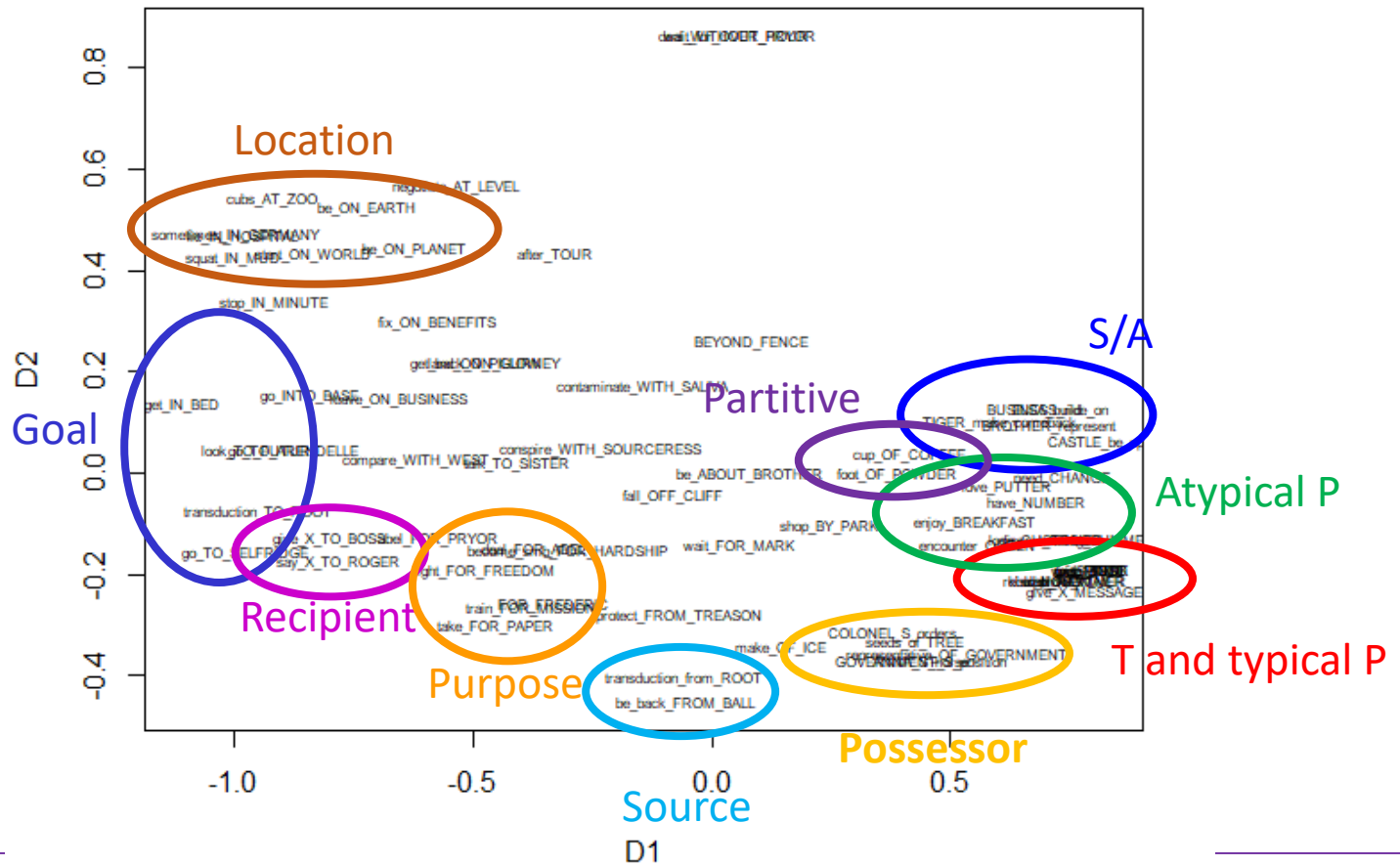
Roles



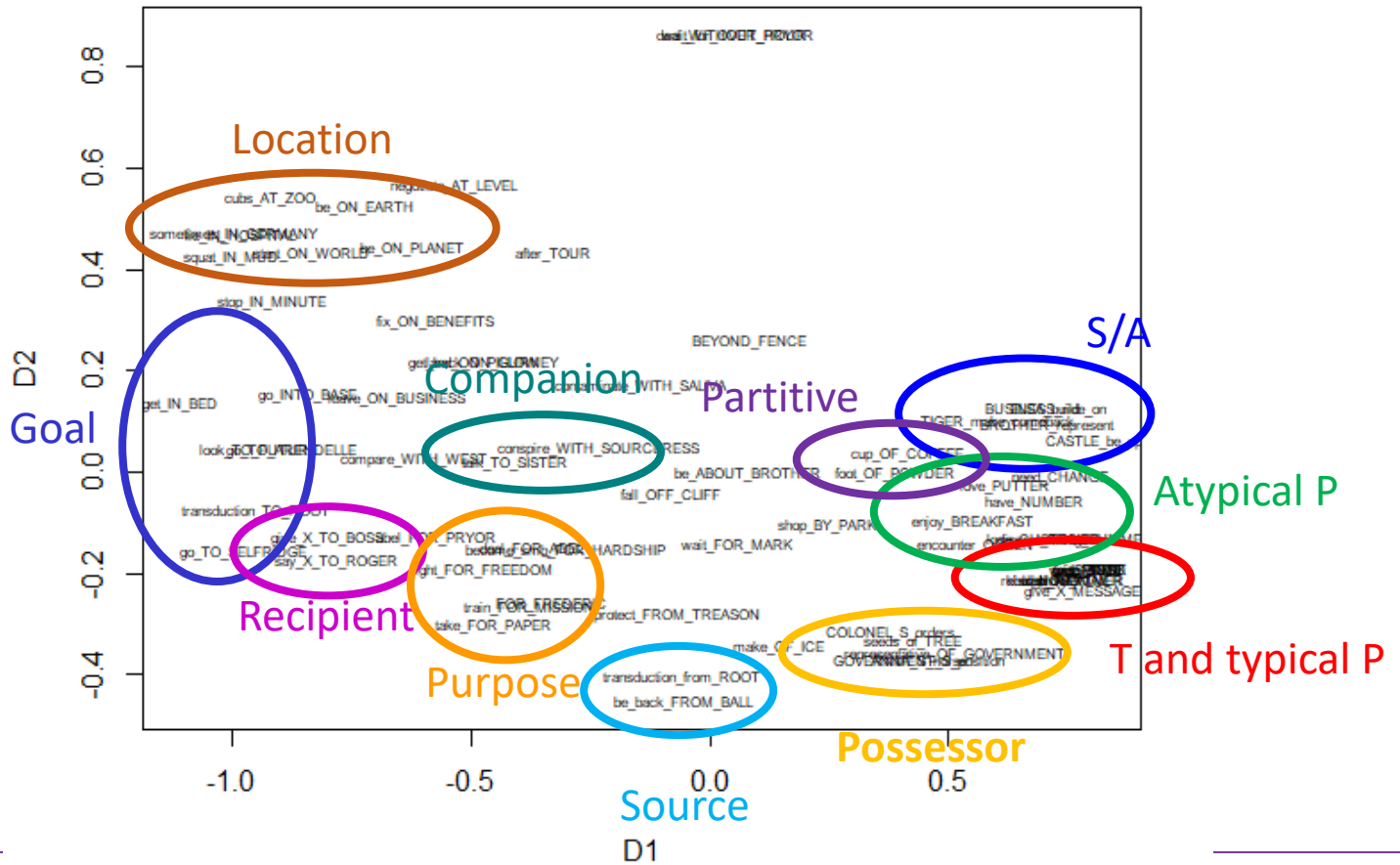
Roles



Roles



Roles



Operationalization of parameters

- Integrity: length of role marker in segments
- Syntactic scope: the entire NP with determiners, adjectives, etc. or only the noun. NP = 1, Noun = 0
- Bondedness:
 - Separability: Possible insertions between the nominal stem and the marker of the role. Yes = 1, No = 0
 - Phonological stability (e.g. vowel harmony in TUR, merge with article in GER, variants of 's in ENG). Yes = 1, No = 0

Example of coding: eng

- I just need you to give a message **to your boss**.
 - Length = 2
 - Scope = 1 (NP)
 - Separability = 1 (Yes)
 - Phonological stability = 1 (Yes)

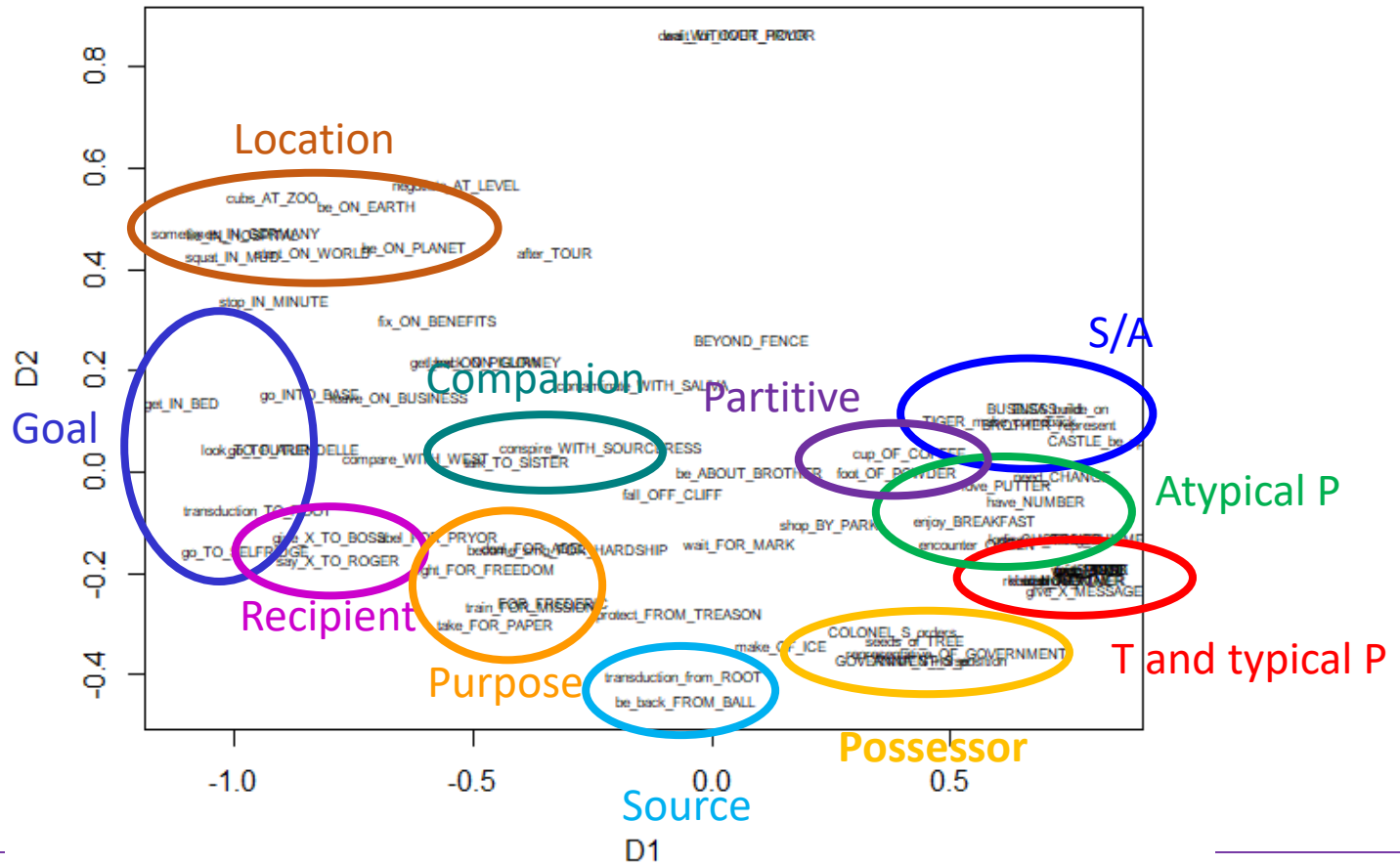
Example of coding: rus

- Ja ... xoču poprosit' vas peredat' poslanije
I want ask you give message
vaš-emu **načalnik-u**.
your-DAT.SG boss-DAT.SG
- Length = 1
- Scope = 0 (Noun)
- Separability = 0 (No)
- Phonological stability = 0 (No)

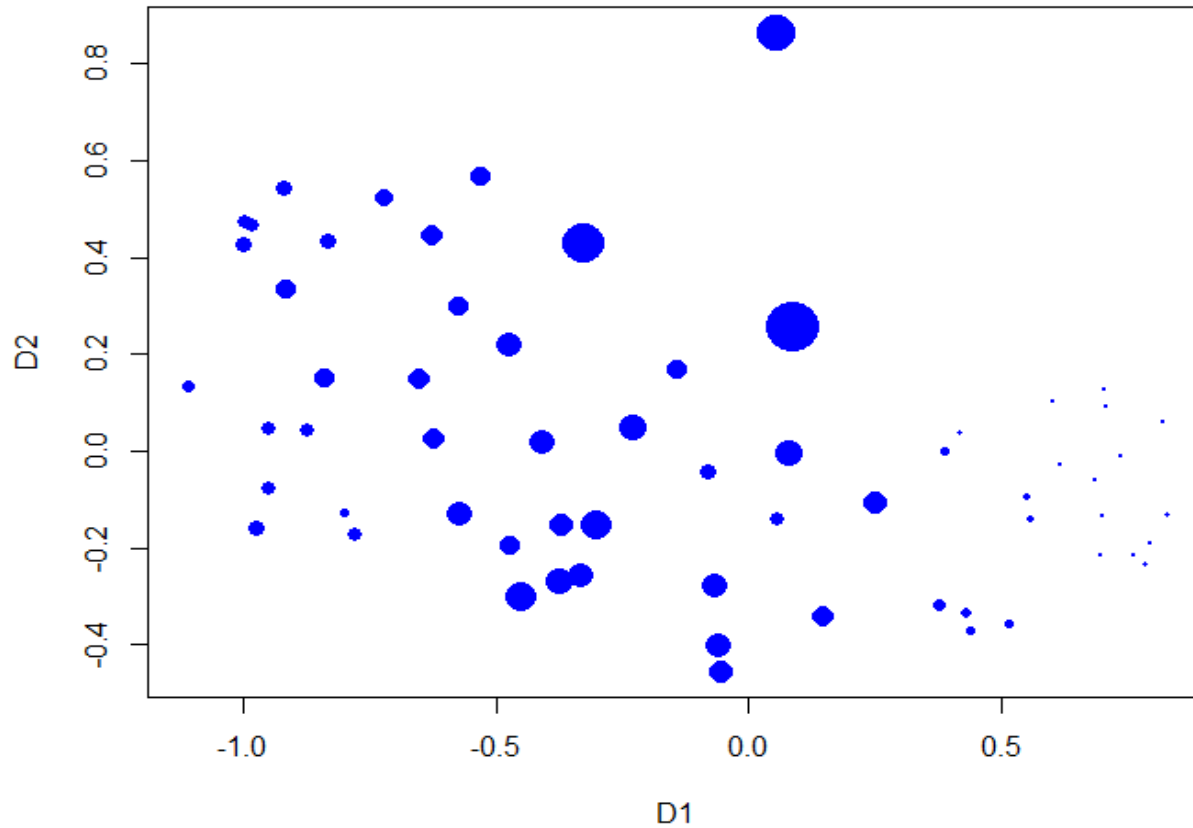
Example of coding: deu

- Ich will nur, dass Sie **Ihrem Chef-∅** etwas ausrichten.
 - Length = 0
 - Scope = 0 (Noun)
 - Separability = 0 (No)
 - Phonological stability = 0 (No)

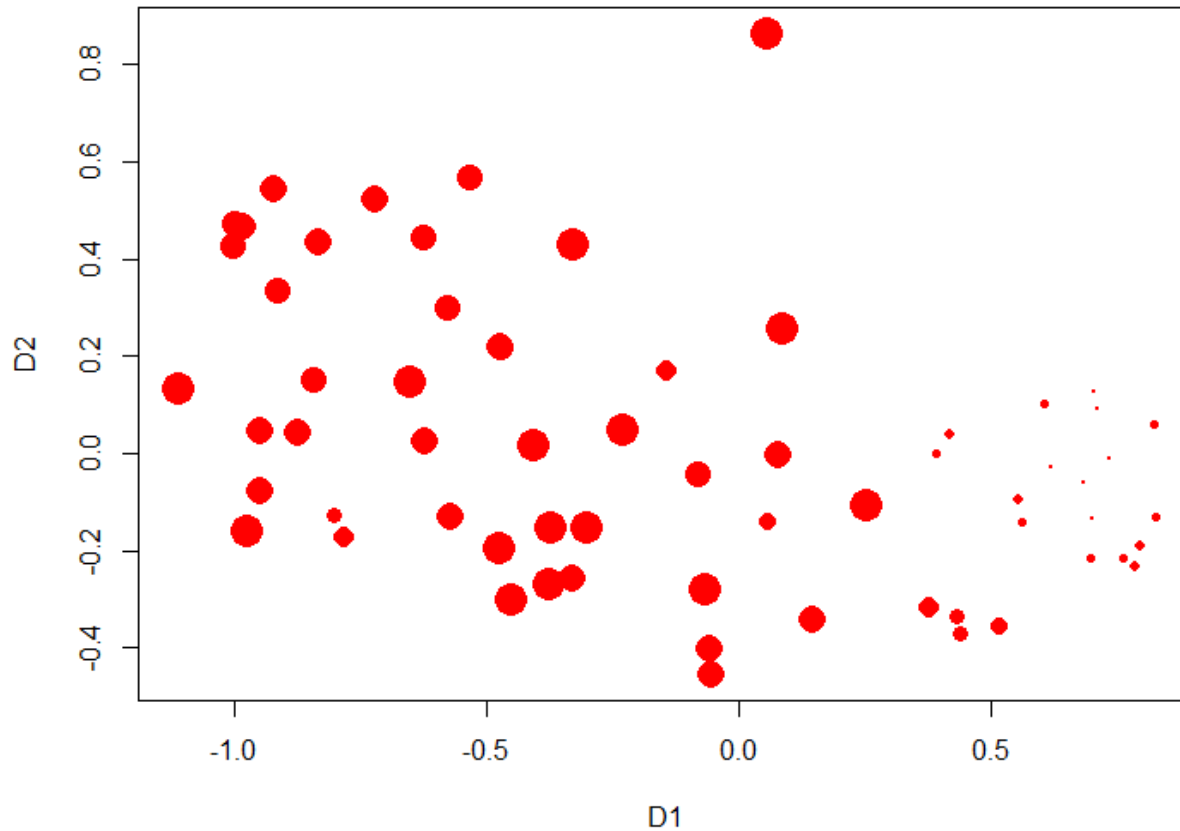
Roles



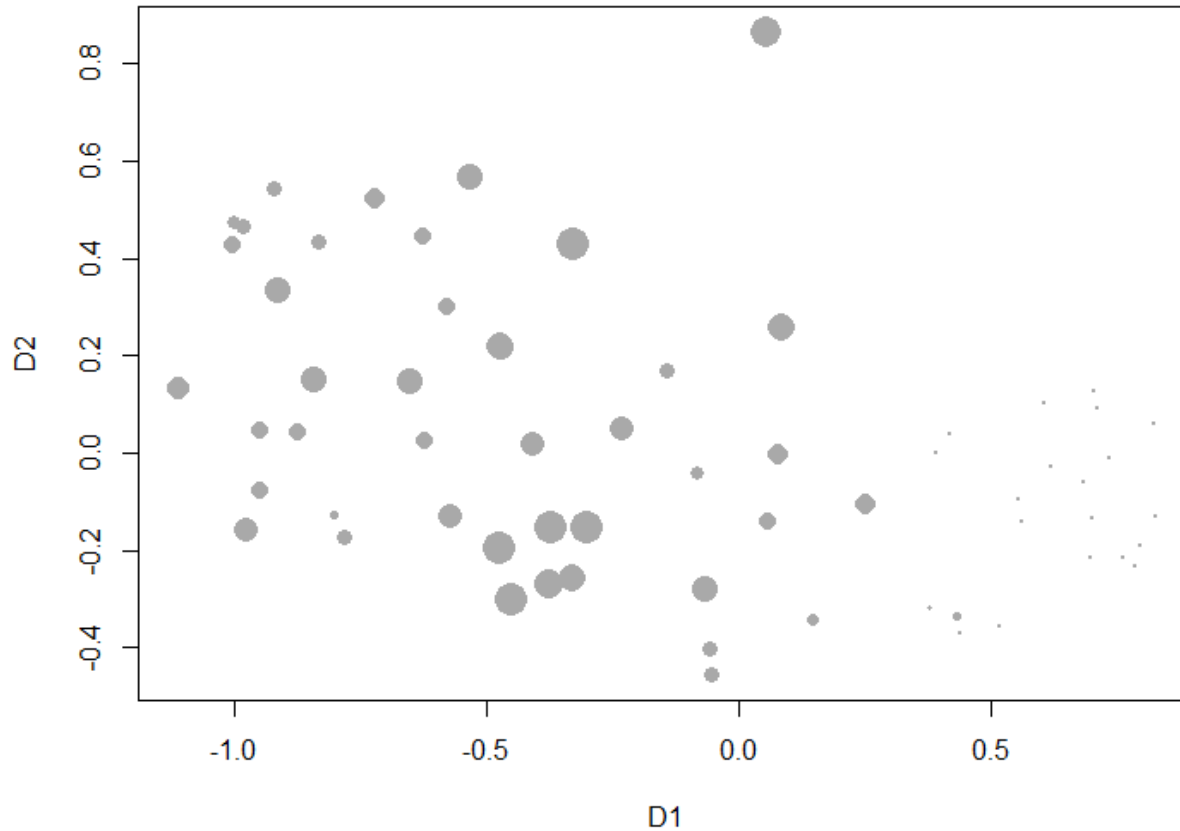
Average normalized length



Average scope/separability



Average phonological stability



Ranking of roles

Role	Length	Scope/ Separable	Phon. Stability	Mean for 3 parameters
S/A	0.037	0.08	0	0.034
Atypical Object	0.071	0.113	0	0.061
Patient	0.07	0.205	0	0.092
Theme (transfer)	0.083	0.225	0	0.103
Partitive	0.143	0.225	0	0.127
Possessor	0.182	0.46	0.06	0.234
Recipient	0.223	0.5	0.3	0.341
Location	0.326	0.775	0.491	0.531
Source	0.419	0.78	0.391	0.531
Goal	0.327	0.838	0.641	0.602
Topic (about)	0.512	0.8	0.6	0.637
Companion	0.502	1	0.7	0.734
Purpose	0.502	1	0.925	0.809

Zero, Suffix or Adposition?

Role	ENG	DEU	FIN	RUS	TUR
S/A	Zero	Zero	Zero (Suffix?)	Zero/Suffix	Zero/Suffix
Atypical Object	Zero	Zero/Suffix	Suffix	Zero/Suffix	Zero/Suffix
Patient	Zero	Zero/Suffix	Suffix	Zero/Suffix	Suffix
Theme	Zero	Zero/Suffix	Suffix	Zero/Suffix	Suffix
Partitive	Prep	Zero	Zero/Suffix	Suffix	Zero
Possessor	Suffix/Prep	Suffix/Prep	Suffix	Suffix	Suffix
Recipient	Zero/Prep	Zero/Suffix	Suffix (+ Post)	Suffix	Suffix
Location	Prep	Prep (+ Suffix)	Suffix	Prep + Suffix	Suffix
Source	Prep	Prep (+ Suffix)	Suffix	Prep + Suffix	Suffix
Goal	Prep	Prep (+ Suffix)	Suffix (+ Post)	Prep + Suffix	Suffix (+ Post)
Topic (about)	Prep	Prep (+ Suffix)	Suffix	Prep + Suffix	Suffix + Post
Companion	Prep	Prep (+ Suffix)	Suffix + Post	Prep + Suffix	Suffix + Suffix/Post
Purpose	Prep	Prep (+ Suffix)	Suffix + Post	Prep + Suffix	Suffix + Post

What we've learnt

- Blake's categorical hierarchy can also be interpreted as a continuum of grammaticalization of particular semantic roles.
- Why? A possible reason is frequency asymmetry. Who did what to whom? > Whose? > (To) Whom? > Where? > With whom? Why? About what? Etc. More frequent roles undergo grammaticalization faster.
- Partitives: probably due to high predictability given the 'classifiers', e.g. *a cup of tea* → *cuppa tea*.
- More research on the link between predictability and grammaticalization is needed.

So, what can parallel corpora do for (quantitative) diachronic research?

- Induction and visualization of local grammaticalization clines on probabilistic semantic maps
 - Methods: MDS, Multiple Correspondence Analysis (for construction-based maps) – this talk
- Induction and testing of global grammaticalization clines
 - Method: MDS (auxiliary) – this talk
- Detection and testing of changes in multiple probabilistic constraints
 - Methods: classification trees and random forests – good for small datasets! (cf. Levshina 2016, 2017)
- Investigation of diachronic development of language types, computation of transitional probabilities
 - Method: phylogenetic trees (Verkerk 2015)
- ...?

References

- Blake, Barry J. 2001. *Case*. Cambridge: Cambridge University Press.
- Dixon, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Cysouw, Michael & Diana Forker. 2009. Reconstruction of morphosyntactic function: Nonspatial usage of spatial case marking in Tsezic. *Language* 85 (3). 588-617.
- Georgakopoulos, Thanasis & Stéphane Polis. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass* (to appear)
- Hartmann, Irene, Martin Haspelmath & Michael Cysouw. 2014. Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language* 38(3). 463–484.
- Lamiroy, Béatrice. 2011. Degrés de grammaticalisation à travers les langues de même famille. In *Mémoires de la Société de linguistique de Paris : Nouvelle Série*, Vol. 19, 167–192. Leuven: Peeters.
- Lehmann, Christian. 2015. *Thoughts on Grammaticalization*. 3rd edn. Berlin: Language Science Press.
- Levshina, Natalia. 2015. European analytic causatives as a comparative concept. Evidence from a parallel corpus of film subtitles. *Folia Linguistica* 49(2). 487-520.
- Levshina, Natalia. 2016. Why we need a token-based typology: A case study of analytic and lexical causatives in fifteen European languages. *Folia Linguistica* 50(2). 507–542.
- Levshina, Natalia. 2017. A Multivariate Study of T/V Forms in European Languages Based on a Parallel Corpus of Film Subtitles. *Research in Language* 15(2). 153-172.
- Soares da Silva, Augusto. 2012. Stages of grammaticalization of causative verbs and constructions in Portuguese, Spanish, French and Italian. *Folia Linguistica* 46(2). 513–552.
- van Haeringen, C.B. 1956. *Nederlands tussen Duits en Engels*. The Hague: Servire.
- Verkerk, Annemare. 2015) Where do all the motion verbs come from? The speed of development of manner verbs and path verbs in Indo-European. *Diachronica* 32(1). 69-104.
- Wälchli, Bernhard & Michael Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3). 671–710.

Thanks!

The slides are available at

<http://www.natalialevshina.com/presentations.html>

The ParTy corpus is available (partly) at

<https://github.com/levshina/ParTy-1.0>