Each Baltic valency is Baltic in its own way? Evidence from parallel subtitles

Daria Alfimova, University of Potsdam

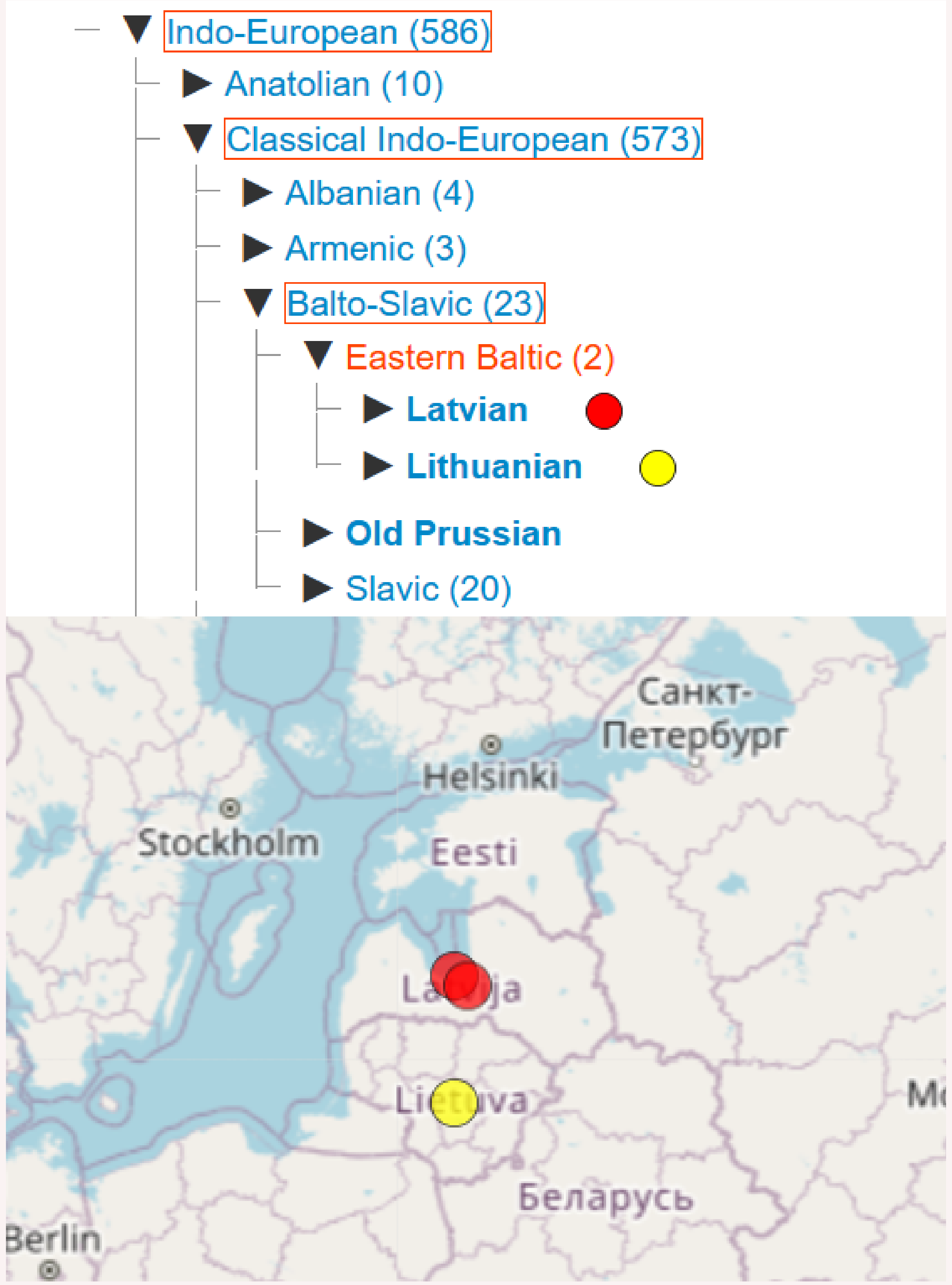
ALT 15, December 5, 2024

Workshop: Valency and valency alternations within and across language boundaries

Background

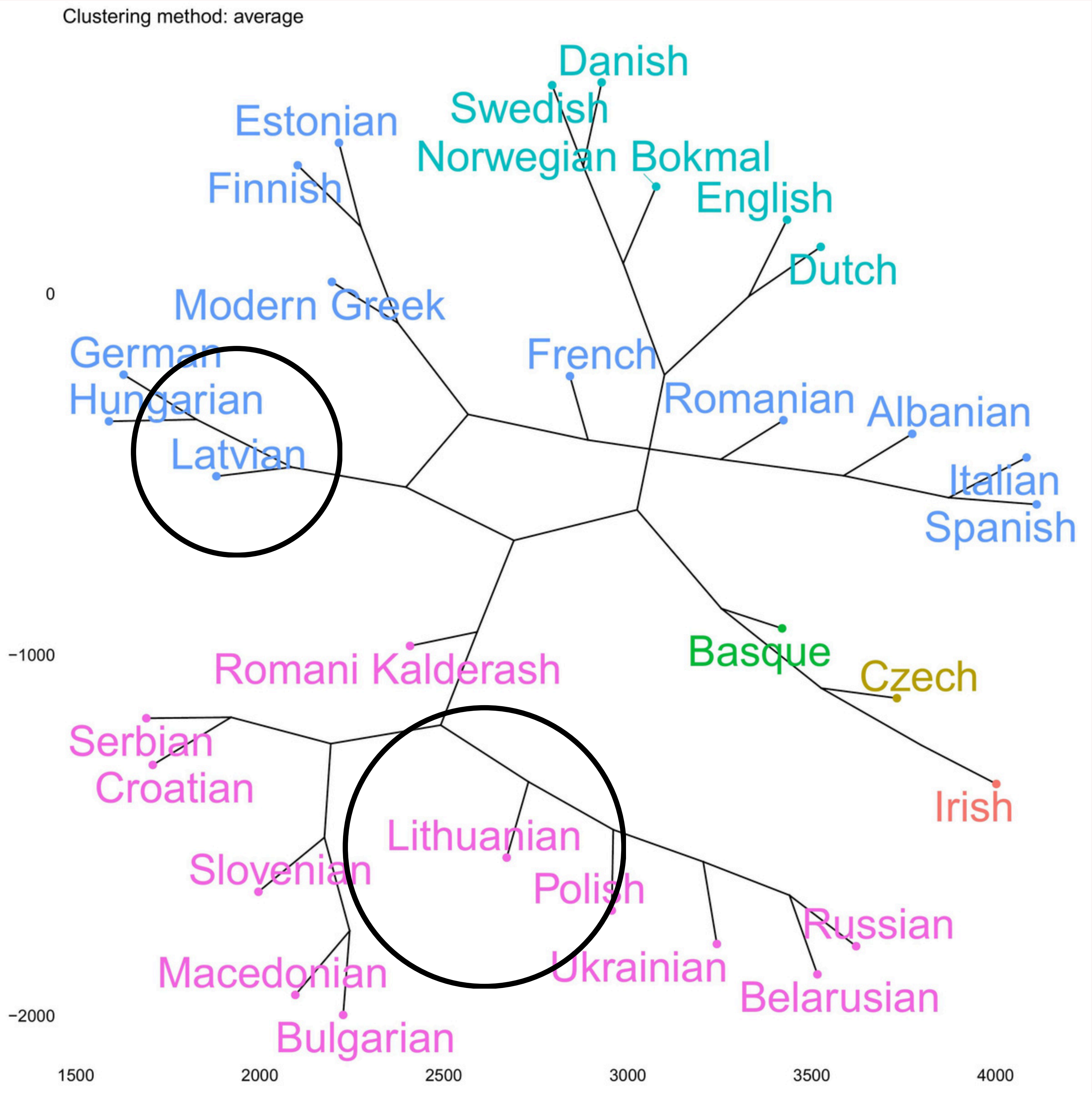
Latvian and Lithuanian:
neighboring territories, not mutually intelligible

Impressionistically, the use of verbs and valency patterns plays a significant role in the lack of mutual intelligibility



Hammarström et al. 2024

Previously on...



Alfimova 2024

Based on the valency patterns in the BivalTyp database (Say 2020-), Latvian and Lithuanian were shown to belong to different clusters (Alfimova 2024)

↓

usage-based approach?
(Bybee 2010; Diessel 2020)

Research questions

How many Baltic cognates are employed in parallel contexts, and does the parallel use of cognates trigger valency matches?

What is the role of prefixes, especially directional prefixes (such as Latvian *ie-* and Lithuanian *į-* ‘into’) in predicting valency matches?

Can valency mismatches in Baltic languages always be explained by language contact?

Data and descriptives

parallel corpus of film subtitles ParTy (Levshina 2016)

two films with both Latvian and Lithuanian

↓

“Inception” (Christopher Nolan, 2010): **2105 lines**

Annotation

Assumption: Overt coding of corresponding referents in both languages

(the term valency pattern is understood in its broad sense and
formally defined as the set of all referable participants of a single lexical verb)

↓

393 comparable verbs (lines)

Annotation: simplified example

source_sentence	source_negation	latvian_translation	ja_debitiv	latvian_comparable	lithuanian_translation	lithuanian_comparable	baltic_prefixes	baltic_dirprefixes	baltic_cognates	baltic_compared
Come back to reality, Dom.	0	Atgriezies īstenīb**ā**, Dom.	0	**LOC**	Sugrįžk **į** realybę, Domai.	**SD**	1	0	1	**FALSE**

(1) Atgriezies īstenīb-**ā**, Dom
come_back.IMP reality-**LOC** PN

(2) Sugrįžk **į** realybę, Domai.
come_back.IMP **to** reality.ACC PN

Annotation: simplified example

case → language-specific case annotation

adposition → comparative concepts:
spatial dynamic,
spatial static,
non-spatial temporal,
non-spatial non-temporal

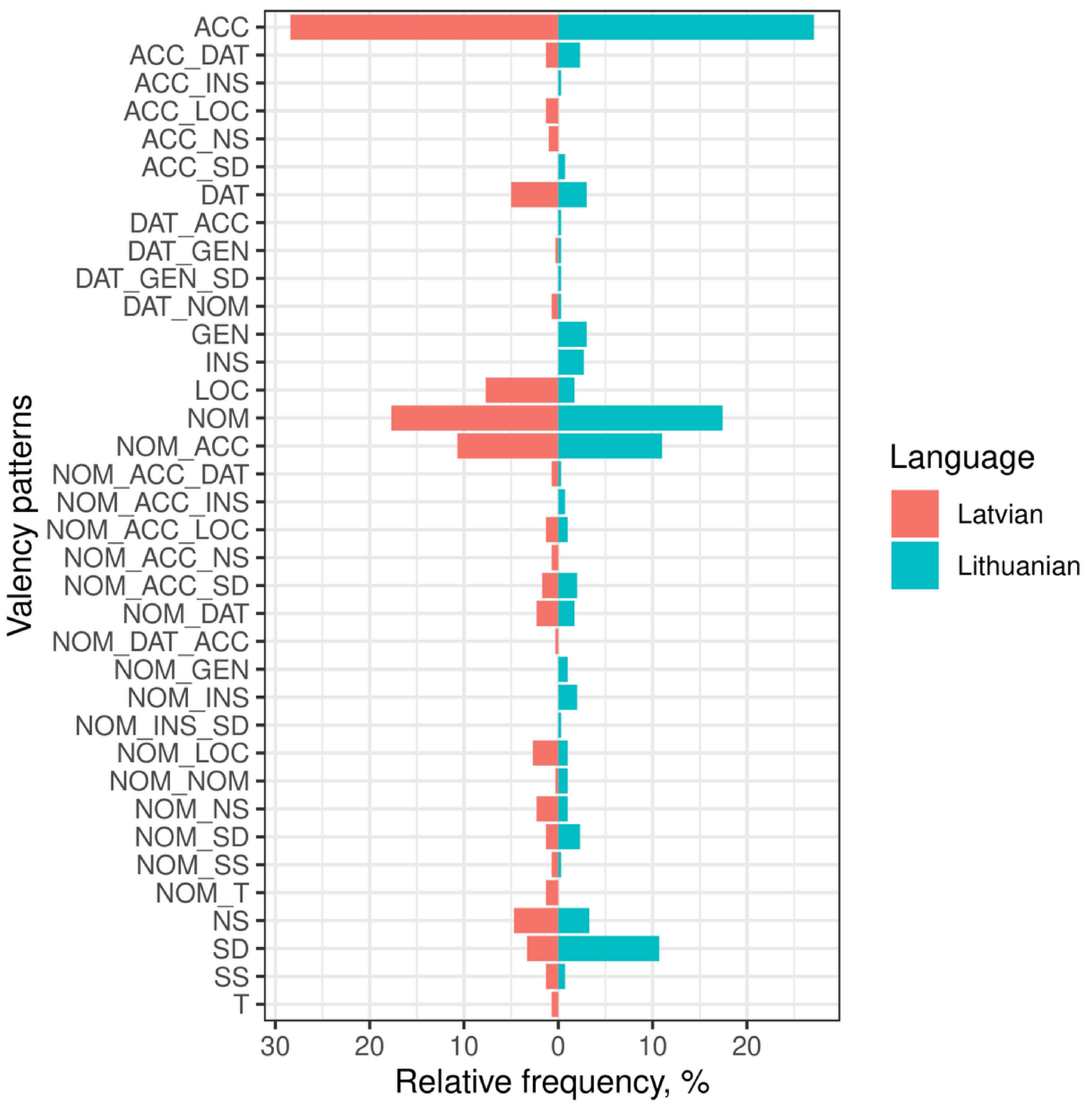
source_sentence	source_negation	latvian_translation	ja_debitiv	latvian_comparable	lithuanian_translation					
Come back to reality, Dom.	0	Atgriezies īstenībā, Dom.	0	**LOC**	Sugrįžk **į** realybę, Domai.	**SD**	1	0	1	**FALSE**

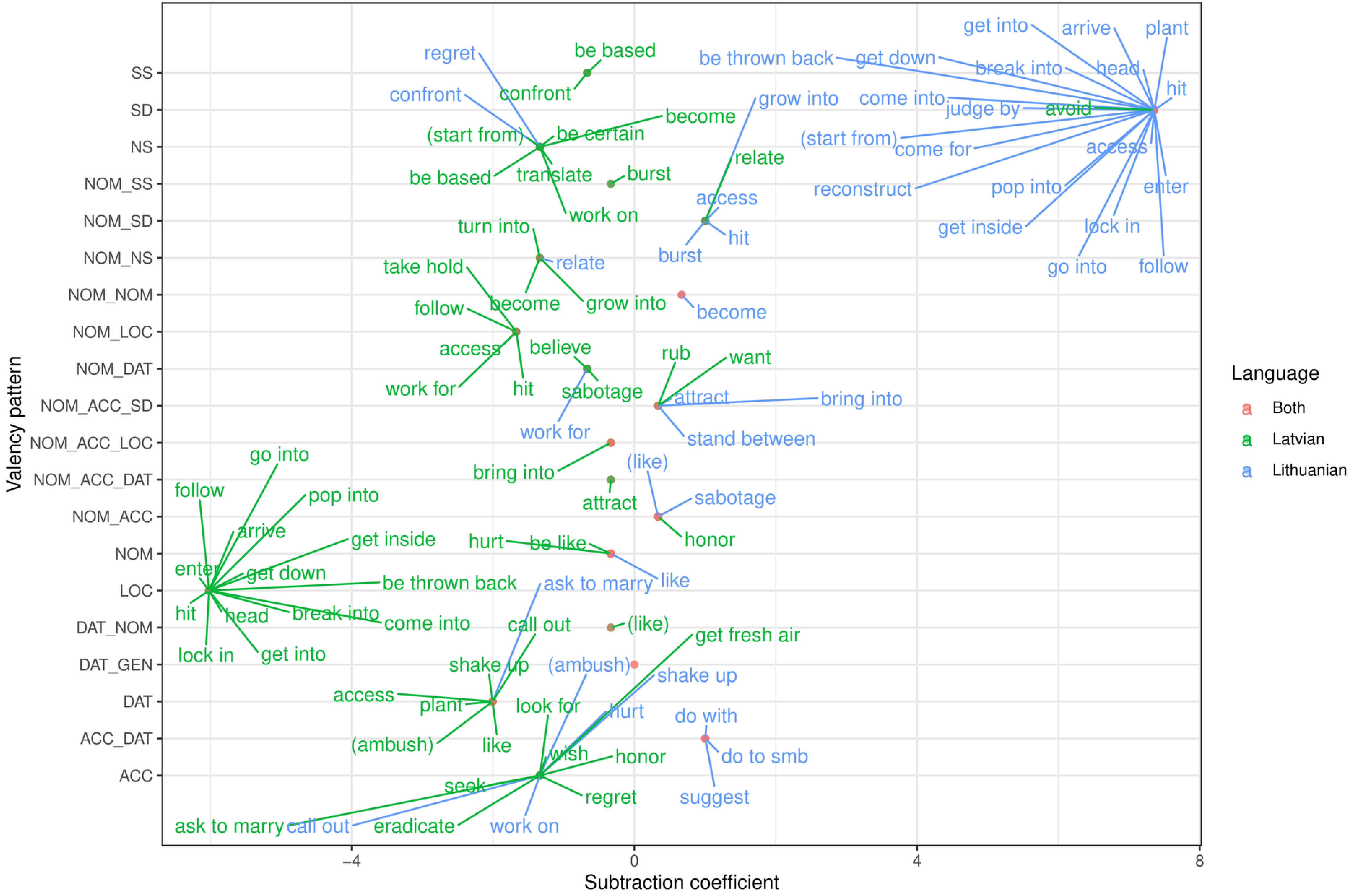
(1)Atgriezies īstenīb-**ā**, Dom
come_back.IMP reality-**LOC** PN

(2)Sugrįžk **į** realybę, Domai.
come_back.IMP **to** reality.ACC PN

Annotation: 3 contact languages

russian_translation	russian_comparable	polish_translation	polish_comparable	german_translation	german_comparable
Возвращай-ся **к** реальности.	**SD**	Wróć **do** rzeczywistości, Dom.	**SD**	Komm **in** die Realität zurück, Dom.	**SD**


ACC
ACC_DAT
ACC_INS
ACC_LOC
ACC_NS
ACC_SD
DAT
DAT_ACC
DAT_GEN
DAT_GEN_SD
DAT_NOM
GEN
INS
LOC
NOM
NOM_ACC
NOM_ACC_DAT
NOM_ACC_INS
NOM_ACC_LOC
NOM_ACC_NS
NOM_ACC_SD
NOM_DAT
NOM_DAT_ACC
NOM_GEN
NOM_INS
NOM_INS_SD
NOM_LOC
NOM_NOM
NOM_NS
NOM_SD
NOM_SS
NOM_T
NS
SD
SS
T
Valency patterns
30
20
10
0
10
20
Relative frequency, %
Language
Latvian
Lithuanian


Valency pattern
SS
SD
NS
NOM_SS
NOM_SD
NOM_NS
NOM_NOM
NOM_LOC
NOM_DAT
NOM_ACC_SD
NOM_ACC_LOC
NOM_ACC_DAT
NOM_ACC
NOM
LOC
DAT_NOM
DAT_GEN
DAT
ACC_DAT
ACC
–4
0
4
8
Subtraction coefficient
Language
Both
Latvian
Lithuanian
regret
be based
confront
confront
(start from)
be certain
become
be based
translate
burst
work on
turn into
relate
take hold
become
grow into
follow
access
believe
work for
hit
sabotage
work for
get into
arrive
plant
be thrown back
get down
break into
head
hit
grow into
come into
judge by
avoid
(start from)
come for
access
relate
reconstruct
pop into
enter
access
hit
burst
get inside
lock in
go into
follow
become
rub
want
attract
bring into
stand between
bring into
(like)
attract
sabotage
honor
go into
follow
pop into
arrive
get inside
hurt
be like
enter
get down
be thrown back
ask to marry
like
hit
head
break into
come into
call out
(like)
lock in
get into
get fresh air
shake up
(ambush)
shake up
access
plant
look for
hurt
do with
(ambush)
like
wish
honor
do to smb
seek
regret
suggest
ask to marry
call out
eradicate
work on

H1: Cognates exhibit fewer valency mismatches than etymologically unrelated verbs

(3) Cognates, match

Latvian (ACC)
Iznīcināt visu manu mantojumu?
destroy whole.ACC 1SG.POSS.ACC inheritance.ACC

Lithuanian (ACC)
Sunaikinti visą mano palikimą?
destroy whole.ACC 1SG.POSS inheritance.ACC

‘Destroying my whole inheritance?’

(4) Unrelated, mismatch

Latvian (NOM_T)
Es neizturami skumstu pēc tevis.
1SG.NOM unbearably miss after 2SG.GEN

Lithuanian (NOM_GEN)
Aš labai tavęs ilgiuosi.
1SG.NOM very 2SG.GEN miss

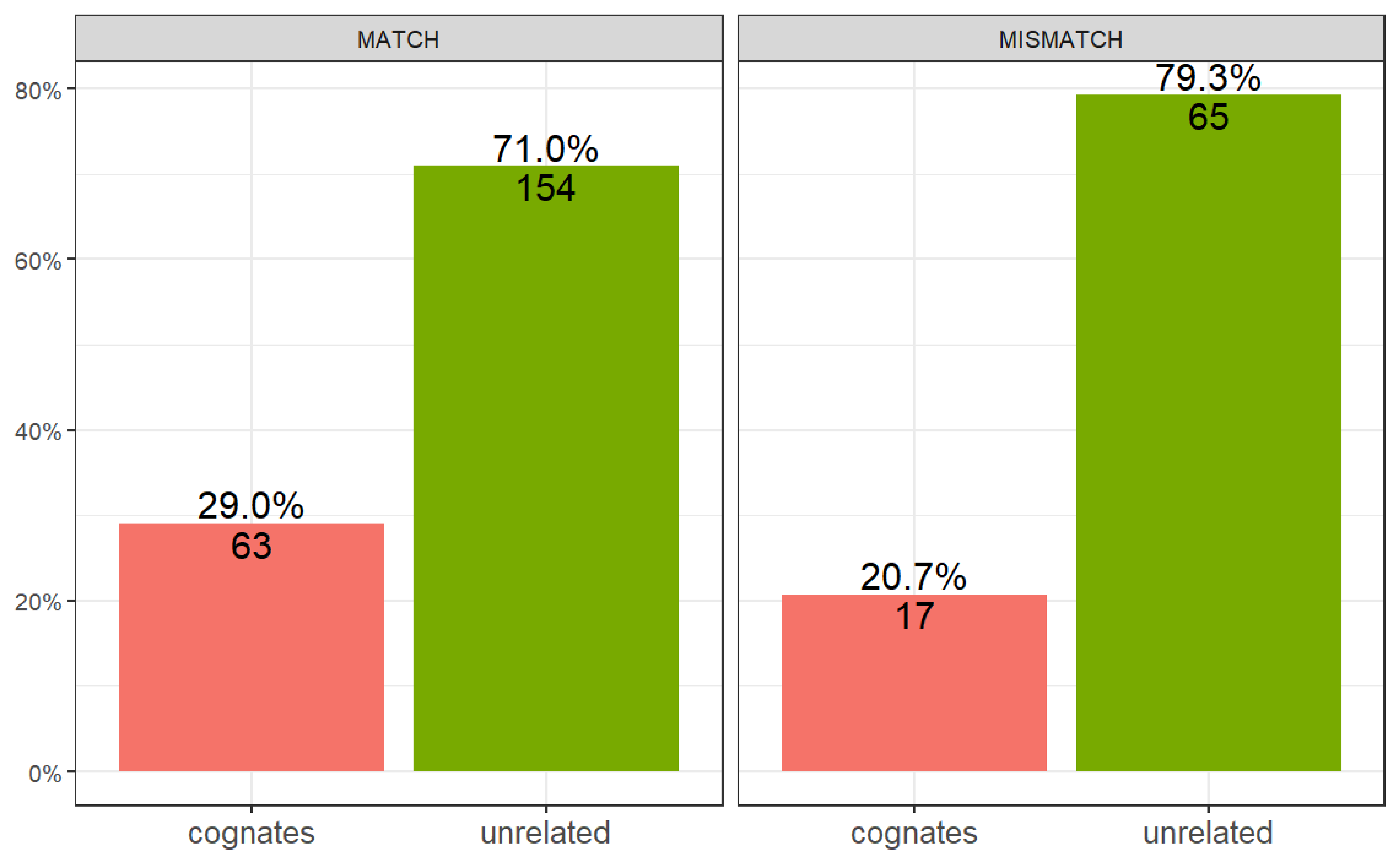
‘I miss you more than I can bear…’

H1: Cognates exhibit fewer valency mismatches than etymologically unrelated verbs

the original set of **393 comparable** parallel lines was reduced to **299** by excluding overly frequent lexical meanings (N ≥ 5), which overwhelmed the data with one and the same valency pattern

'know', 'do', 'build', 'tell', 'wait', 'need', 'have', 'create', 'talk', 'take somewhere', 'look', 'come back'

~~H1: Cognates exhibit fewer valency mismatches than etymologically unrelated verbs~~



$\chi 2 = 1.6902$, df = 1, $p \approx 0.19$

Main results

1

Baltic cognates do not show significantly more valency matches than unrelated verbs.

This suggests that the 'genealogical factor', often cited as influencing valency matches, is not necessarily rooted in shared verb origins.

Instead, it is likely driven by a shared approach in assigning specific marking devices to certain semantic roles, along with a common logic in word formation.

H2: Parallel prefixation increases the chances of a valency match

(5) Prefixed, match

Latvian (NOM_ACC_SD)
Mēs **iz**-vilkām no jums visu informāciju.
1PL.NOM **ex**-tract from 2PL.DAT whole.ACC information.ACC

Lithuanian (NOM_ACC_SD)
Mes **iš**-traukėm iš jūsų visą informaciją.
1PL.NOM **ex**-tract from 2PL.DAT whole.ACC information.ACC

‘We extracted every bit of information you had in there.’

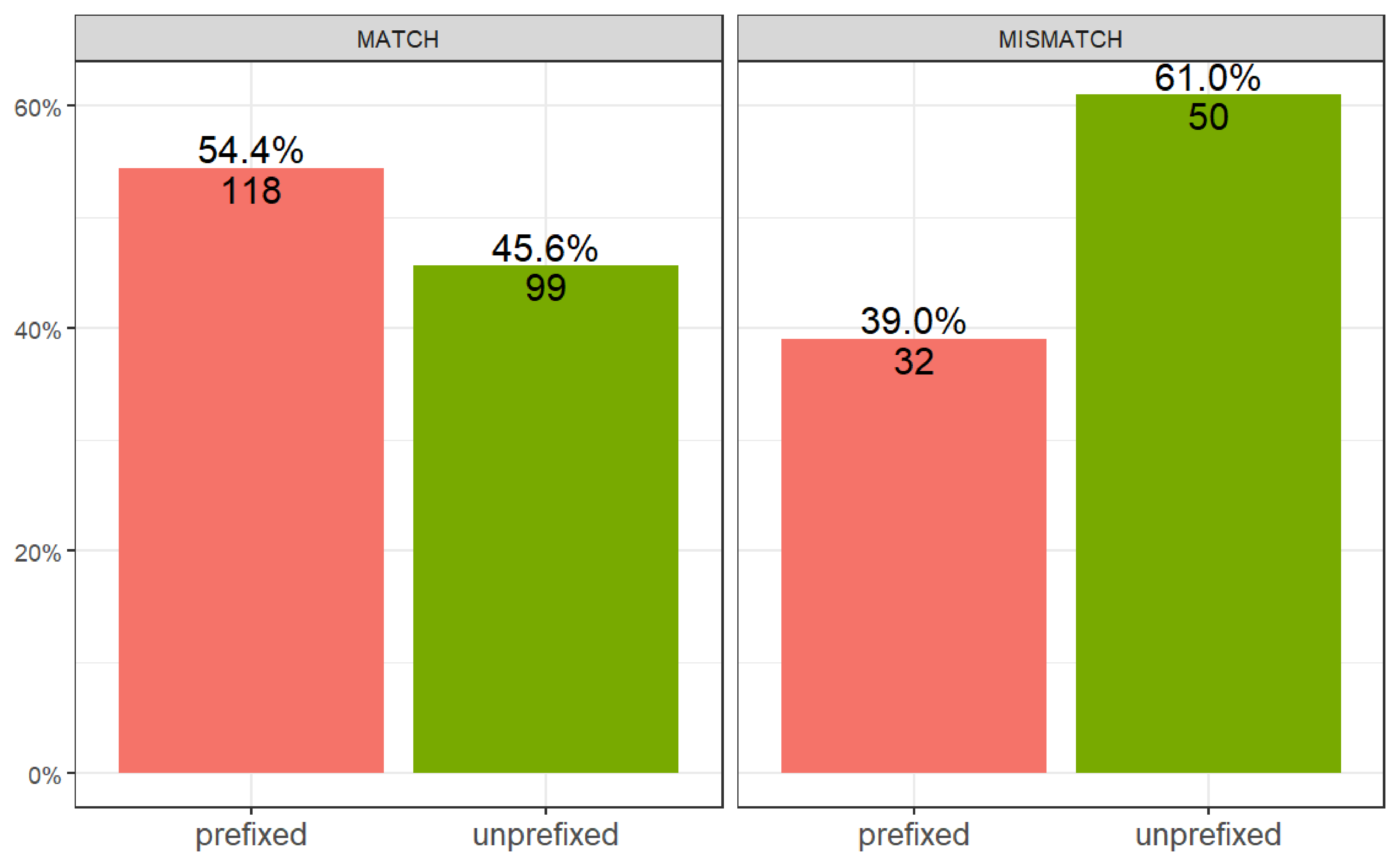
(6) Unprefixed, mismatch

Latvian (DAT_NOM)
Taču kam tu tici?
but who.DAT 2SG.NOM believe

Lithuanian (INS_NOM)
Bet kuo tu tiki?
but who.INS 2SG.NOM believe

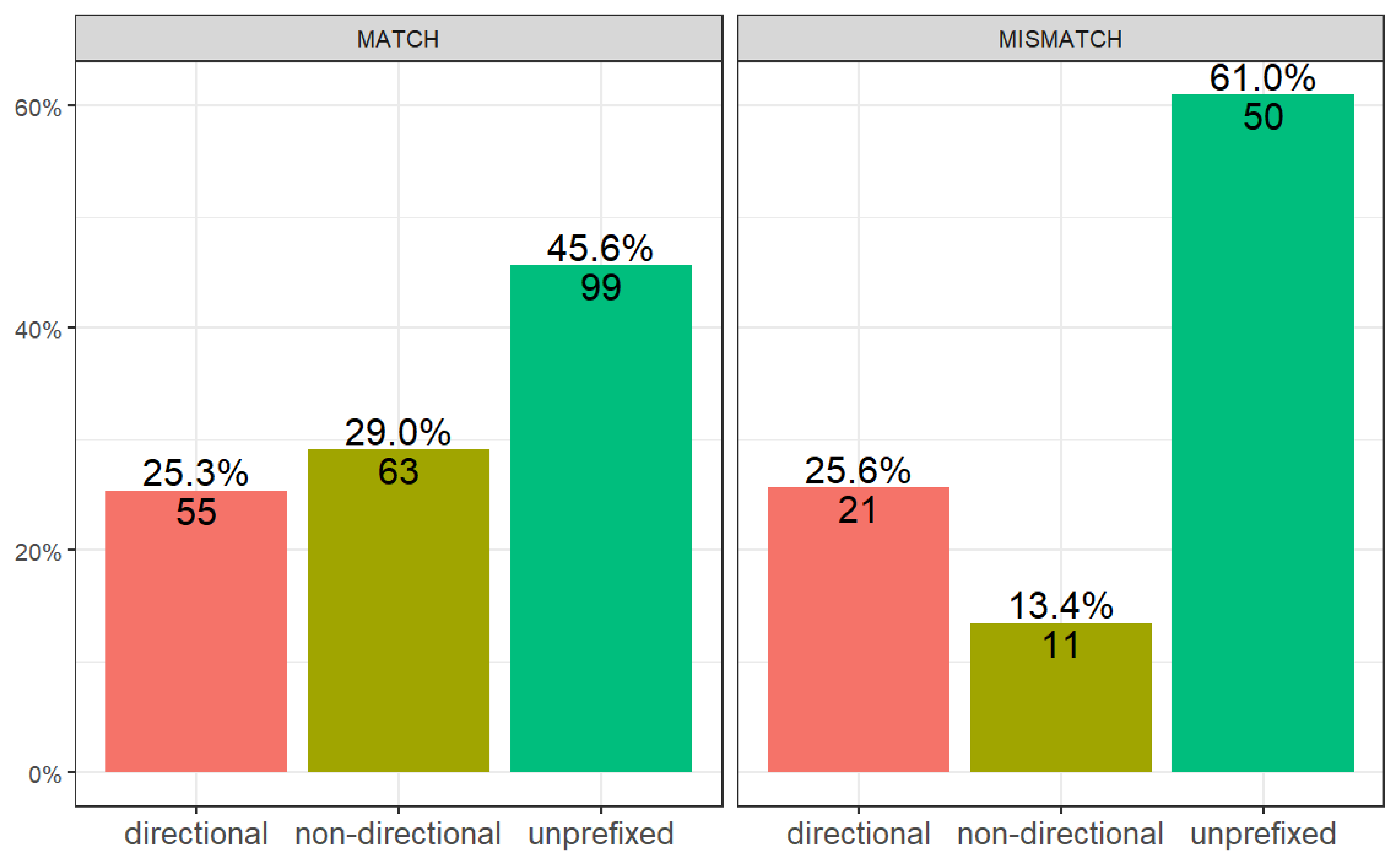
‘But what do you believe?

H2: Parallel prefixation increases the chances of a valency match



$\chi 2 = 5.0142$, df = 1, $p \approx 0.025$

H2: Parallel prefixation increases the chances of a valency match



Main results

2

Prefixed verbs match in valency significantly more frequently than unprefixed verbs.

This, again, highlights the importance of word formation for the valency choice.

Latvian appears to be more precise with prefixal direction encoding than Lithuanian, which might compensate for the multifunctionality of the Latvian Locative case.

H3: Baltic mismatch → greater-than-chance probability of a match with at least one contact language

H3: Baltic mismatch → greater-than-chance probability of a match with at least one contact language

there are 219 data points coming from three contact languages (German, Polish, Russian)

H3: Baltic mismatch → greater-than-chance probability of a match with at least one contact language

82 mismatches in Baltic valency patterns:

19 (23%) - no valency match with any contact language for either of the Baltic languages

10 (12%) both languages match with one of the contact languages while not matching in valency with each

11 (13%) - Latvian matches with one of the contact languages while not matching with Lithuanian

42 (51%) - Lithuanian matches with one of the contact languages while not matching with Latvian

H3: Baltic mismatch → greater-than-chance probability of a match with at least one contact language confirmed for Lithuanian

82 mismatches in Baltic valency patterns:

19 (23%) - no valency match with any contact language for either of the Baltic languages

10 (12%) both languages match with one of the contact languages while not matching in valency with each

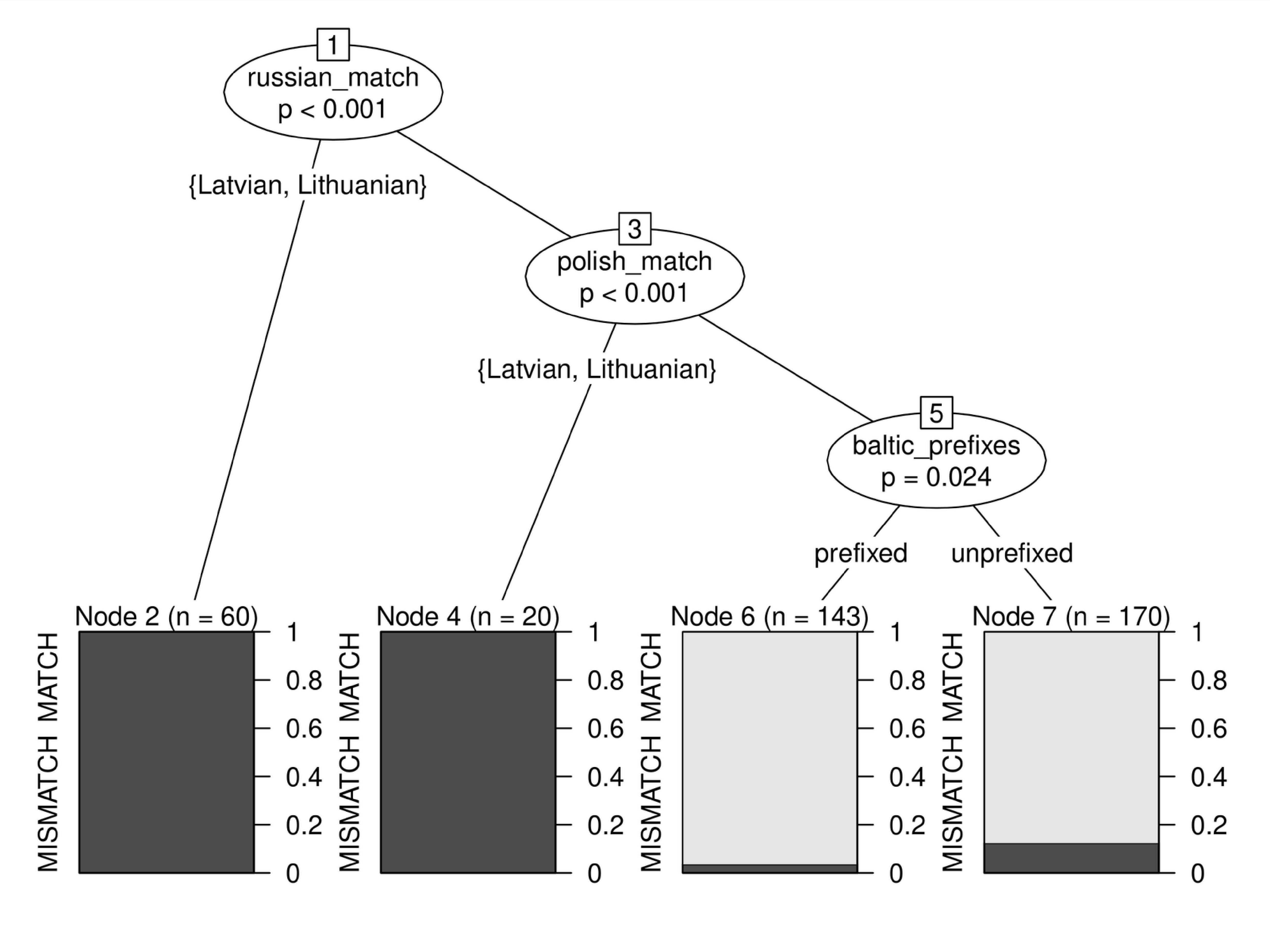
11 (13%) - Latvian matches with one of the contact languages while not matching with Lithuanian

42 (51%) - Lithuanian matches with one of the contact languages while not matching with Latvian

H3: Baltic mismatch → greater-than-chance probability of a match with at least one contact language **confirmed for Lithuanian**

	German	Polish	Russian	
Total number of matches with Latvian	13	10	8	31 (23%)
Total number of matches with **Lithuanian**	28	39	37	**104 (77%)**
	41 (30%)	49 (36%)	45 (33%)	**135**

H3



1
russian_match
p < 0.001
{Latvian, Lithuanian}
3
polish_match
p < 0.001
{Latvian, Lithuanian}
5
baltic_prefixes
p = 0.024
prefixed
unprefixed
Node 2 (n = 60)
Node 4 (n = 20)
Node 6 (n = 143)
Node 7 (n = 170)
MISMATCH MATCH
1
0.8
0.6
0.4
0.2
0

Main results

3

In case of valency mismatches in Baltic, Lithuanian alignes with the major contact languages significantly more frequently than Latvian.

Most likely, this follows from the fact that the Latvian language had/has more genealogically unrelated contact languages (Livonian, Estonian), which undermined the retention power of the closely related contact languages (Lithuanian, Russian).

Thank you! Key takeaways

The ‘genealogical factor’, often cited as influencing valency matches, is not necessarily rooted in shared verb origins. Instead, it is likely driven by a shared approach in assigning specific marking devices to certain semantic roles, along with a common logic in word formation.

The more (genealogically) diverse the contact languages are, the less retention power do closely related contact languages have.

References

Alfimova, Daria. 2024. Object Marking in Western Eurasia: The Circum-Baltic Area Dissolves into the Broader Areal Background. Journal of Language Contact, 17(2), 281-314. https://doi.org/10.1163/19552629-bja10065.

Bybee, Joan. 2010. Language, Usage and Cognition. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511750526.

Diessel, Holger. 2020. A dynamic network approach to the study of syntax. Frontiers in Psychology 11. https://doi.org/10.3389/fpsyg.2020.604853.

Hammarström, Harald & Forkel, Robert & Haspelmath, Martin & Bank, Sebastian. 2024. Glottolog 5.1. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.14006617 (Available online at http://glottolog.org, accessed on 4 December 2024.)

Levshina, Natalia. 2016. Verbs of letting in Germanic and Romance languages: A quantitative investigation based on a parallel corpus of film subtitles. Languages in Contrast 16(1), 84–117.

ParTy. Parallel Corpus for Typology. (Available online at https://github.com/levshina/ParTy-1.0, accessed on 4 November 2024).

Say, Sergey. 2014. Bivalent verb classes in the languages of Europe: A quantitative typological study. Language Dynamics and Change 4(1): 116–166. https://doi.org/10.1163/22105832-00401003.Say, Sergey (ed.). 2020-. BivalTyp: Typological database of bivalent verbs and their encoding frames. (Available online at https://www.bivaltyp.info, accessed on 4 December 2024.)