Preferred lexicalization and language contact

Eitan Grossman & Dmitry Nikolaev
Department of Linguistics
The Hebrew University of Jerusalem

Today's talk

Does language contact influence preferred lexicalization patterns in the domain of basic lexical valence orientation?

Background to today's talk

"In Russian, 'learn' is the reflexive of 'teach'; in Mongolian, 'teach' is the causative of 'learn'; in Mandarin, 'teach' and 'learn' are separate verbs. In each of these three languages, the kind of formal pairing found in 'teach' and 'learn' is repeated in many other pairs of verbs with analogous semantic relationships" (Nichols et al. 2004: 149).

Nichols, Johanna, David Peterson & Jonathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8: 149-211.

Background to today's talk

"These are not random facts. It will be shown here that languages can be typologized into a few broad groups: those that tend to treat intransitives as basic or simple and transitives as derived or complex, those that do the reverse, those that treat both as derived, and those that treat both as underived. This distinction is not a mechanical reflection of the presence of causative, middle, etc. morphology in a language, but a deep-seated principle governing lexicalization as well as grammar. We will speak of this large domain of facts and types as the LEXICAL VALENCE ORIENTATION of a language."

Background to today's talk

"...the lexical valence orientation of a language is the preferred or predominant or most common form of lexicalization or valence-related derivation, and usually it is lexically and/or grammatically basic or privileged… Types, that is, are statistical tendencies and not categorical or exclusive. Any verb can have its own particular and often idiosyncratic synchronic grammatical behavior no matter how strong or regular the typological propensities of a language. Lexicalization, in short, is a word-by-word matter that takes place against a background typological orientation."

Background to today's talk

Haspelmath (1993):

Some members of causal-noncausal pairs tend to be coded as causatives, while others tend to be coded as anticausatives.

a. Russian: inchoative derived from causative
   causative: *rasplavit'* 'melt (tr.)' →
   inchoative: *rasplavit'-sja* 'melt (intr.)'
b. (Khalkha) Mongolian: causative derived from inchoativ
   causative: *xajl-uul-ax* 'melt (tr.)' ←
   inchoative: *xajl-ax* 'melt (intr.)'

More on the typology of inchoative/causative verb alternations*

Martin Haspelmath
Free University of Berlin

Really crucial to say at the outset

We'll see a few different 'functional' explanations (in Bickel's sense), but this isn't what's going to interest us and we won't survey the whole range of 'functional' accounts of coding alternations involved in causal:noncausal pairs.

Rather, we'll focus on evaluating whether language contact ('event-based triggers') accounts for some of the observed cross-linguistic diversity.

Formal coding types 1: causatives

The transitive is derived from the intransitive:

Korean (isolate, Korea)

kkhul-ta	*kkhul-**i**-ta*
'boil' (intr.)	'boil' (trns.)

Formal coding types 2: anticausative

The intransitive is derived from the transitive

Maltese (Semitic, Malta;)

n-*ħaraq*	*ħaraq*
‘burn’ (intr.)	‘burn’ (trns.)

Formal coding types 3: equipollent

In equipollent alternations, both the intransitive and the transitive are derived.

Udmurt (Uralic, Russian Federation)

aźin-**sky**-ny	aźin-**ty**-ny
‘develop’ (intr.)	‘develop’ (trns.)

Formal coding types 4: labile

In labile alternations, the same form is used both intransitively and transitively.

English

‘freeze’ (intr.)

‘freeze’ (trns.)

Formal coding types 5: suppletion

Different roots are used for the intransitive and transitive.

Modern Hebrew (Semitic, Israel; pers. knowledge)

met	*harag*	(cf. *he-mit* ‘put to death’)
‘died’	‘killed’	

Three ways of clustering the alternations

1. On the one hand, causative and anticausative alternations are directed. Equipollent, labile, and suppletive alternations are undirected.

DIRECTED	UNDIRECTED
C,A	E,L,S

Three ways of clustering the alternations

2. On the other hand, causative, anticausative, and equipollent alternations all involve derivation, while labile and suppletive alternations arguably don't.

DERIVED	UNDERIVED
C, A, E	L, S

Three ways of clustering the alternations

3. On the third hand, all of the alternations involve the same root for the intransitive and the transitive, while suppletive alternations don't.

SAME ROOT	DIFFERENT ROOT
C,A,E,L	S

A 21 language sample

Table 1. The 21 languages of the sample

(Indo-European)	Russian, Lithuanian, German, English, French, Rumanian, Greek, Armenian, Hindi-Urdu
(Finno-Ugric)	Hungarian, Finnish, Udmurt
(Afro-Asiatic)	Arabic, Hebrew
(Turkic)	Turkish
(Mongolian)	Khalkha Mongolian
(Nakho-Daghestanian)	Lezgian
(Kartvelian)	Georgian
(Niger-Congo)	Swahili
(Austronesian)	Indonesian
(unclassified)	Japanese

Verb meanings examined

Table 2. The 31 inchoative/causative verb pairs

1. 'wake up (intr.)/(tr.)'	12. 'change (intr.)/(tr.)'	22. 'finish (intr.)/(tr.)'
2. 'break (intr./(tr.)'	13. 'melt (intr.)/(tr.)'	23. 'turn (intr.)/(tr.)'
3. 'burn (intr.)/(tr.)'	14. 'be destroyed/destroy'	24. 'roll (intr.)/(tr.)'
4. 'die/kill'	15. 'get lost/lose'	25. 'freeze (intr.)/(tr.)'
5. 'open (intr.)/(tr.)'	16. 'develop (intr.)/(tr.)'	26. 'dissolve (intr.)/(tr.)'
6. 'close (intr.)/(tr.)'	17. 'connect (intr.)/(tr.)'	27. 'fill (intr.)/(tr.)'
7. 'begin (intr.)/(tr.)'	18. 'boil (intr.)/(tr.)'	28. 'improve (intr.)/(tr.)'
8. 'learn/teach'	19. 'rock (intr.)/(tr.)'	29. 'dry (intr.)/(tr.)'
9. 'gather (intr.)/(tr.)'	20. 'go out/put out'	30. 'split (intr.)/(tr.)'
10. 'spread (intr.)/(tr.)'	21. 'rise/raise'	31. 'stop (intr.)/(tr.)'
11. 'sink (intr.)/(tr.)'		

Main findings (1)

Languages differ in their macro-preferences.

1. Some languages have an overwhelming preference for directed alternations, others for undirected alternations.

2. Some languages generally prefer one type of alternation, e.g., causative, anticausative, etc.

Table 3. Expression types by language

	total	A	C	E	L	S	A/C	% non-dir.
Russian	31	23	0	5	0	3	46.00	26
German	31	14.5	0	4	11.5	1	29.00	53
Greek	31	13.5	0	0	16.5	1	27.00	56
Rumanian	30	24	1	0	3	2	24.00	17
French	31	20.50	2	0	7.5	1	10.25	27
Lithuanian	31	17.5	6	6	0.5	1	2.92	24
Hebrew	31	20.5	7.5	2	1	0	2.73	10
Arabic	31	17	8.5	3	1	1.5	2.00	18
Georgian	31	9	4.5	15.5	0	2	2.00	56
Armenian	31	16	8.5	5.5	0	1	1.88	21
Swahili	31	11	11	8	0	1	1.00	29
Finnish	28	12	13.5	0.5	0.5	1.5	0.88	9
Udmurt	31	10.5	12.5	4.5	2.5	1	0.84	26
Hungarian	31	7	9	12	0	3	0.78	48
Lezgian	31	8	12	6	5	0	0.66	35
Hindi-Urdu	31	7.5	14	7.5	2	0	0.54	31
Turkish	30	9	17.5	2.5	0	1	0.51	12
Mongolian	31	6	22	2	0	1	0.27	10
Indonesian	31	0	14	17	0	0	0.04	55
English	31	2	0	1	25	3		94
Japanese	31	3.5	5.5	20.5	0.5	1		71
total	636	243	164.5	128.5	69	310		

Abbreviations:
A = anticausative alternation
C = causative alternation
E = equipollent alternation
L = labile alternation
S = suppletive alternation
A/C = ratio of anticausative to causative pairs
% non-dir. = percentage of non-directed pairs

Main findings (2)

Individual meanings differ in terms of their coding preferences.

Proposal: a scale of spontaneity of occurrence.

If an event is likely to occur spontaneously, causative coding is preferred.

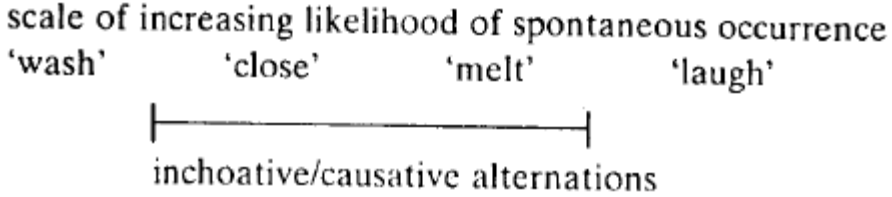
If an event is unlikely to occur spontaneously, anticausative coding is preferred.

Table 4. Expression types by verb pairs

	total	A	C	E	L	S	A/C
18. 'boil'	21	0.5	11.5	3	6	0	0.04
25. 'freeze'	21	2	12	3	4	0	0.17
29. 'dry'	20	3	10	4	3	0	0.30
1. 'wake up'	21	3	9	6	2	1	0.33
20. 'go out/put out'	21	3	7.5	5.5	3	2	0.41
11. 'sink'	21	4	9.5	5.5	1.5	0.5	0.42
8. 'learn/teach'	21	3.5	7.5	6	2	3	0.47
13. 'melt'	21	5	10.5	3	2.5	0	0.48
31. 'stop'	21	5.5	9	3.5	3	0	0.61
23. 'turn'	21	8	7.5	4	1.5	0	1.07
26. 'dissolve'	21	10.5	7.5	2	1	0	1.40
3. 'burn'	21	7	5	2	5	2	1.40
14. 'destroy'	20	8.5	5.5	5	1	0	1.55
27. 'fill'	21	8	5	5	3	0	1.60
22. 'finish'	21	7.5	4.5	5	4	0	1.67
7. 'begin'	19	5	3	3	8	0	1.67
10. 'spread'	21	11	6	3	1	0	1.83
24. 'roll'	21	8.5	4.5	5	3	0	1.89
16. 'develop'	21	10	5	5	1	0	2.00
15. 'get lost/lose'	21	11.5	4.5	4.5	0	0.5	2.56
21. 'rise/raise'	21	12	4.5	3.5	0	1	2.67
28. 'improve'	21	8.5	3	8	1.5	0	2.67
19. 'rock'	21	12	40	3.5	1.5	0	3.00
17. 'connect'	21	15	2.5	1.5	1	1	6.00
12. 'change'	21	11	1.5	4.5	4	0	7.33
9. 'gather'	21	15	2	3	1	0	7.50
5. 'open'	21	13	1.5	4	2.5	0	8.67
2. 'break'	21	12.5	1	4	3.5	0	12.50
6. 'close'	21	15.5	1	2.5	2	0	15.50
30. 'split'	20	11.5	0.5	5	3	0	23.00
4. 'die/kill'	21	0	3	1	1	16	—
total	636	243	164.5	128.5	69	31	

Main findings (2)

To the left are events that are most unlikely to occur spontaneously, to the right events that are most likely to occur spontaneously.



The probability of anticausative coding is highest towards the left, and for anticausative coding to the right.

Fast forward...

Haspelmath et al. (2014) claim that there is no **direct** link between semantics and coding.

Rather

Coding asymmetries are the result of a form-frequency correspondence.

In verb pairs in which the event tends to occur spontaneously (e.g., ‘dry,’ melt,’ ‘freeze’), the noncausal member is relatively more frequent, so that the causal member tends to be coded overtly (as **causative**).

In verb pairs in which the event tends to not occur spontaneously, i.e., to require external force, the causal member is relatively more frequent, so that the noncausal member tends to be coded overtly (as **anticausative**).

The experiment

A sample of 7 languages with extensive corpora.

	Language	Data type	Modality	Total number of words	Identification
1	English	Various	Spoken & written	100 million	Manual
2	Japanese	Various	Written	66 million	Exhaustive
3	Maltese	Various	Written	100 million	Manual
4	Romanian	Newspapers	Written	5 million	Manual
5	Russian	Various	Spoken & written	300 million	Exhaustive
6	Swahili	News texts	Written	12.5 million	Exhaustive
7	Turkish	Newspapers	Written	20 million	Manual

Main findings

Substantial support for the form-frequency hypothesis.

Nichols, Peterson & Barnes (2014)

Similar in approach to Haspelmath (1993), with a partially overlapping set of meanings, more distinctions of coding types and a larger and more balanced sample (80/150).

Table 2. *Verb pairs surveyed*

Pair	Plain	Induced
1	'laugh'	'make laugh, amuse, strike as funny'
2	'die'	'kill'
3	'sit'	'seat, have sit, make sit'
4	'eat'	'feed, give food'
5	'learn, know'	'teach'
6	'see'	'show'
7	'be/become angry'	'anger, make angry'
8	'fear, be afraid'	'frighten, scare'
9	'hid, go into hiding'	'hide, conceal, put into hiding'
10	'(come to) boil'	'(bring to) boil'
11	'burn, catch fire'	'burn, set fire'
12	'break'	'break'
13	'open'	'open'
14	'dry'	'make dry'
15	'be/become straight'	'straighten, make straight'
16	'hang'	'hang (up)'
17	'turn over'	'turn over'
18	'fall'	'drop, let fall'

Importantly

In both studies, ‘functional’ theories are privileged.

What about event-based factors?

The crux of the matter (in today's context)

Both Haspelmath (et al.) and Nichols et al. identified areal signals in their samples.

Haspelmath (1993) identifies a European preference for anticausatives.

Evidence is found for a west-east split within geographically diffused families, mainly Indo-European.

The crux of the matter (in today's context)

Nichols et al. (2004) identifies numerous areal signals:

1. Augmentation (roughly analogous to causativization) is strongly preferred in north Asia and in western and eastern North America.
2. It is dispreferred in Australia and in Europe and Africa.
3. Animate reduction (roughly analogous to anticausativization) is strongly dispreferred in north Asia and preferred in Central America.
4. Inanimate reduction is preferred in Europe and dispreferred in western and eastern North America, as well as in Central America.

The crux of the matter (in today's context)

1. Ambitransitivity (= lability) is common only in inanimates, and it is preferred in Europe and dispreferred in the Americas and in the Pacific Rim.

2. Equipollent alternations have a roughly converse distribution, preferred in Central America and western North America, and dispreferred in Africa.

3. Augmentation of inanimate verbs is generally preferred worldwide. It is dispreferred only in Europe.

The crux of the matter (in today's context)

Overall:

1. Transitivizing languages are common only in Eurasia and N. America, and are completely absent from the Pacific.

2. Undirected alternations are common in Africa and the Pacific.

3. Languages without a distinct preference are common in the Americas.

4. Northern languages tend to prefer directed correspondences, while southern languages tend to prefer undirected correspondences.

So….

1. In a sample biased towards Eurasia, languages show a west-east split, with Europe showing a preference for anticausatives/reduction. This shows up even when taking inheritance into account.
2. In a balanced worldwide sample, where genetic relationships were factored out, significant (and insignificant but possibly telling) areal signals showed up.

All in all, it looks like preferred lexicalization patterns in this domain are prone to contact-induced change.

Some further indications

1. Kulikov & Lavidas (2015) point to an areal split within Indo-European, such that verb lability increased in the western languages (e.g., Romance and Germanic) and decreased in the eastern languages (e.g., Indo-Aryan and Armenian).
2. Coptic and Koine Greek, which were in intensive contact in Late Antique Egypt, both developed an increased preference for verb lability (Grossman 2017, Lavidas 2009).
3. Russian Yiddish has moved away from the Germanic profile towards a strong detransitivizing preference as in Russian, while United States Yiddish has shifted towards a preference for labile verbs as in English (Luchina-Sadan, in prep.), as has Pennsylvania German (Goldblatt, in prep.).

A further experiment (Grossman & Nikolaev in prep.)

A statistical study of an extended sample of Old World languages, mostly from Eurasia, with a handful from Africa.

Based on the NINJAL World Atlas of Transitivity Pairs, which is in turn based on Haspelmath's (1993) original setup.

It includes his sample, and extends it with another 40+ languages.

WATP
THE WORLD ATLAS OF TRANSITIVITY PAIRS

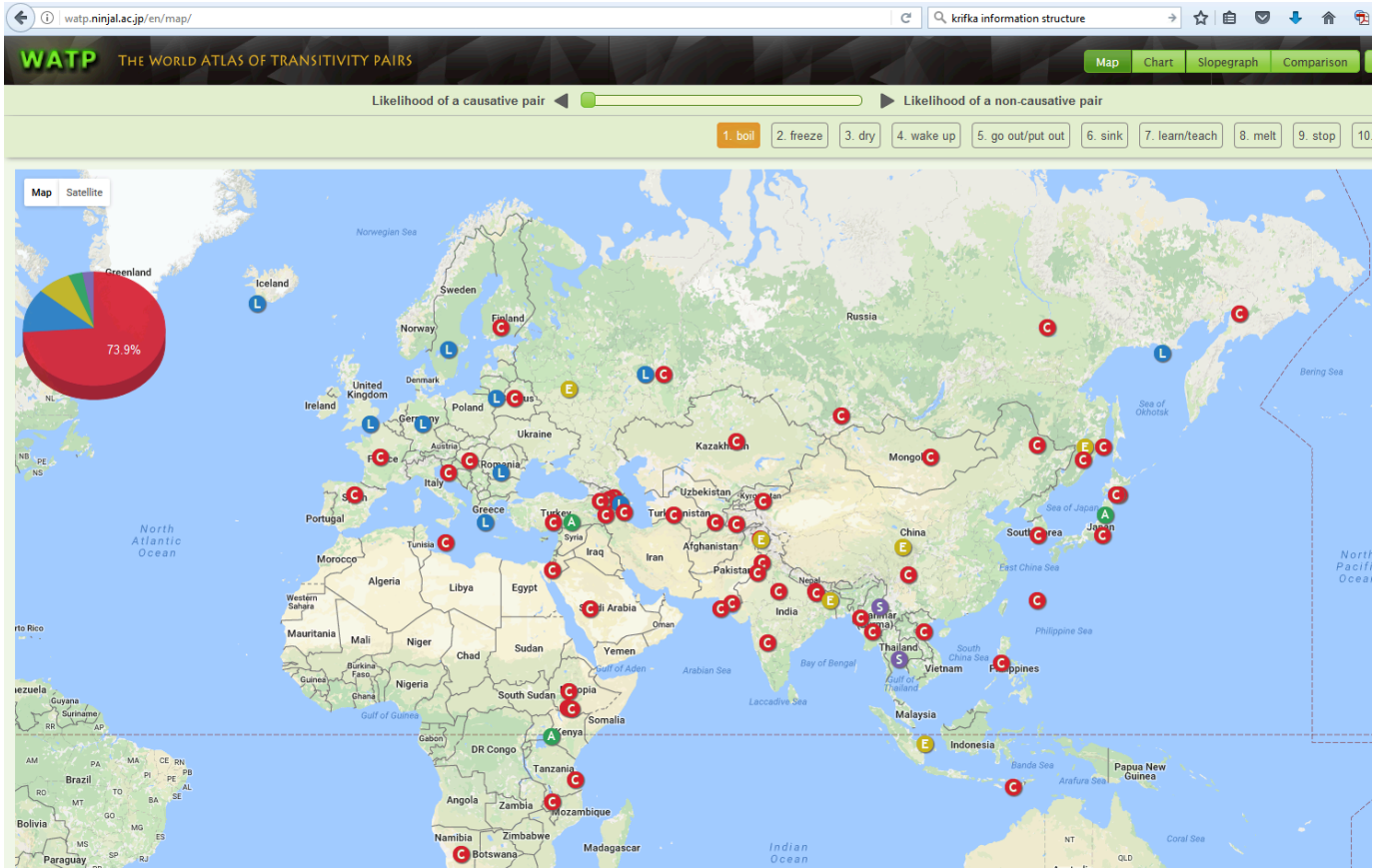
What is WATP?

The World Atlas of Transitivity Pairs (WATP) is a geo-typological database of morphologically related transitivity pairs such as ak-u 'to open (intransitive)': ak-e-ru 'to open (transitive)' in Japanese, which participate in the causative alternation (doa-ga aita [The door opened] vs. Taro-ga doa-o aketa [Taro opened the door]). It consists of primary data from about 80 languages contributed by about 50 researchers, which can be downloaded for research purpose.

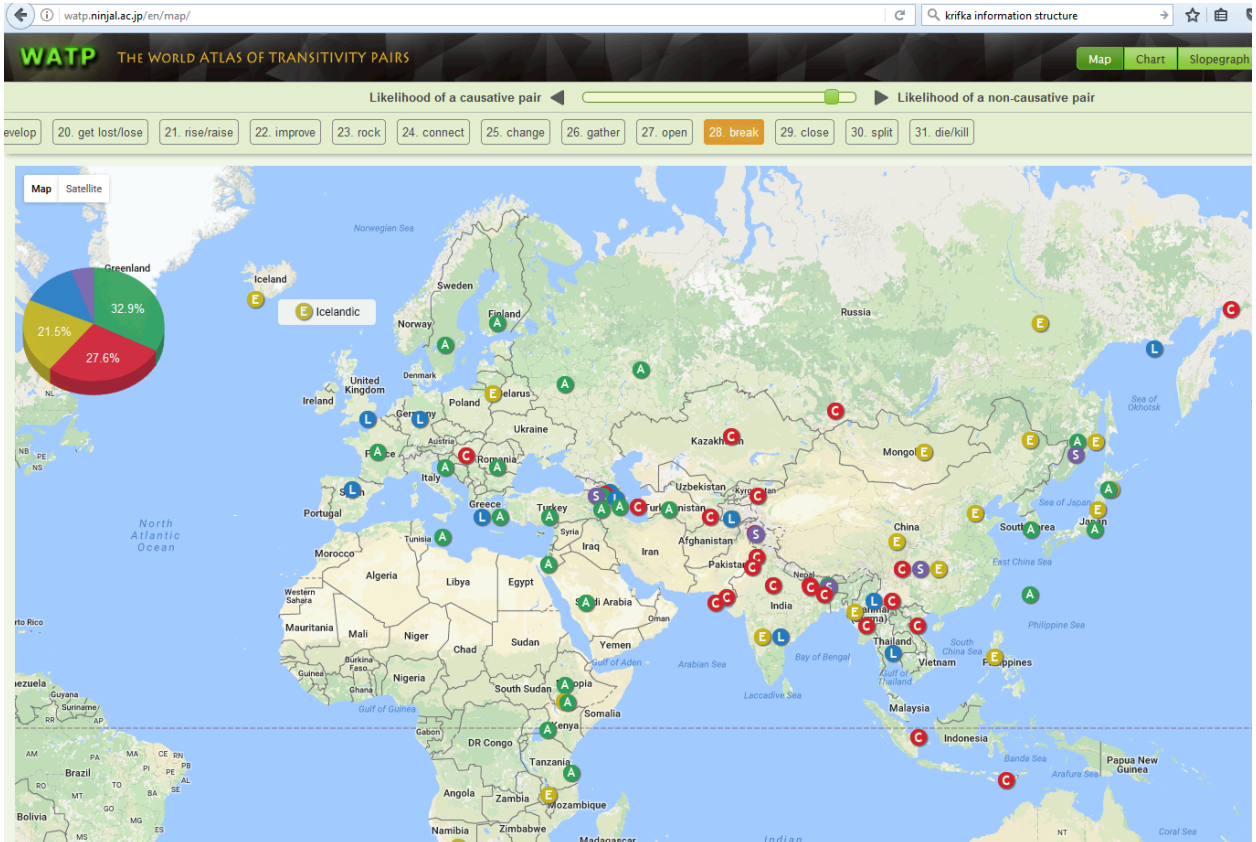
What does WATP do?

WATP offers visual representation of the geographical distribution of the formal relationship between the members of transitivity pairs from the

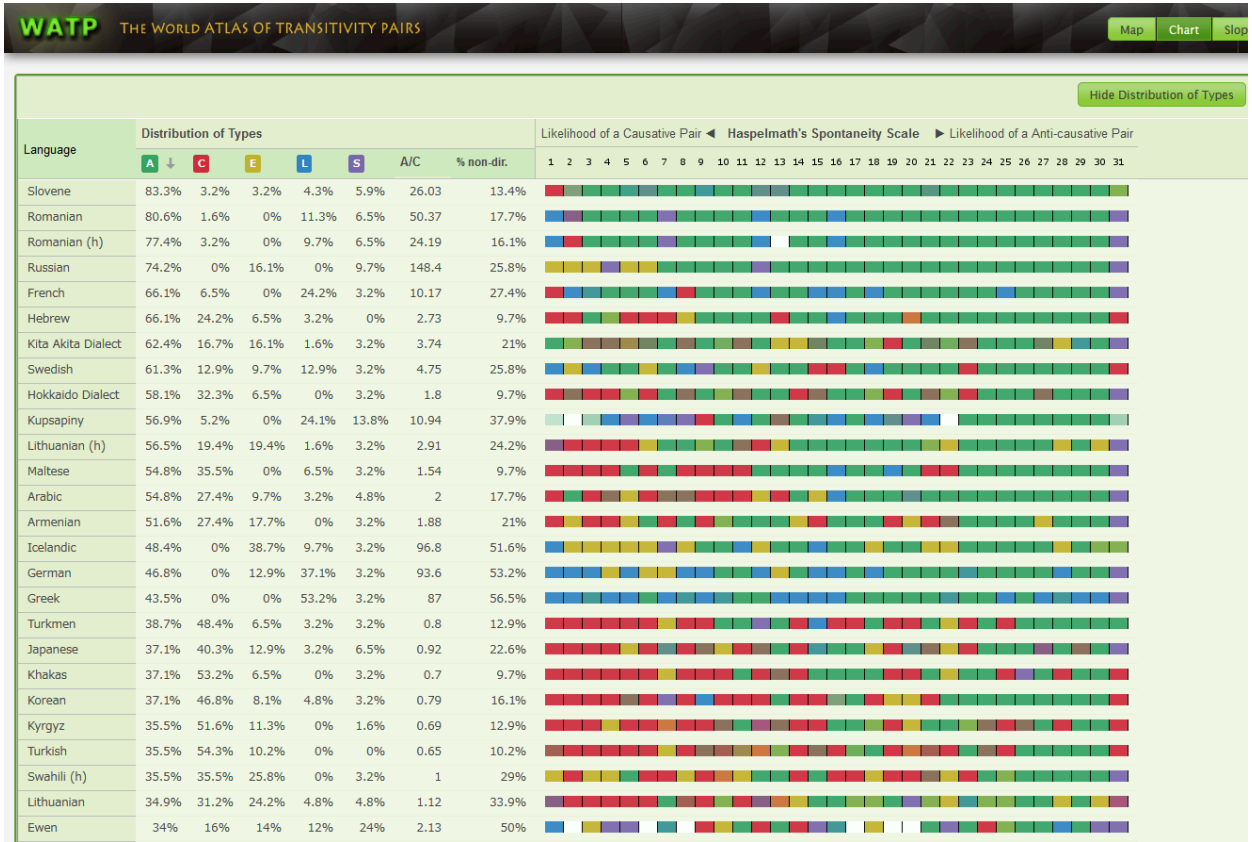
‘Boil’ across Eurasia (and its periphery)



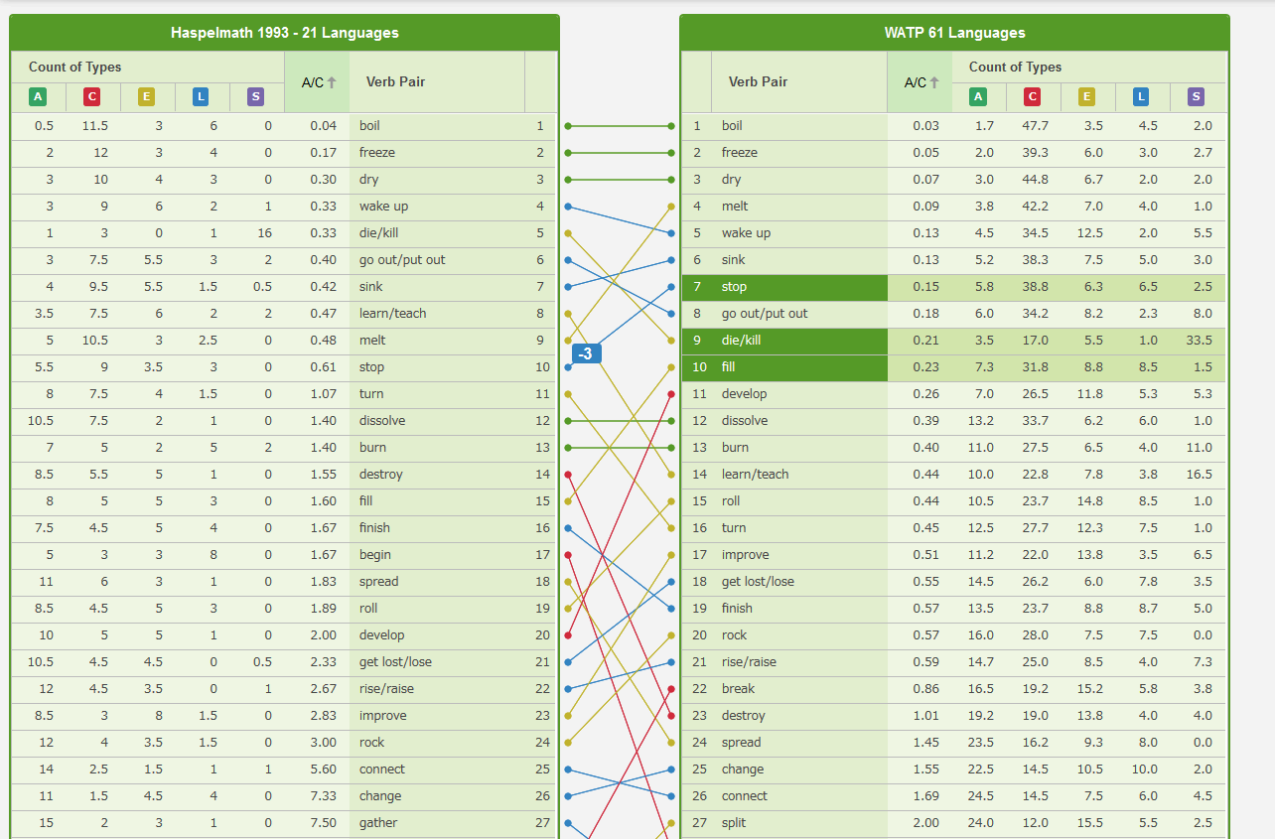
‘Break’ across Eurasia (and its periphery)



Macro-preference by language



Comparison of Haspelmath (1993) and WATP



Haspelmath 1993 - 21 Languages

Count of Types					A/C	Verb Pair	
A	C	E	L	S			
0.5	11.5	3	6	0	0.04	boil	1
2	12	3	4	0	0.17	freeze	2
3	10	4	3	0	0.30	dry	3
3	9	6	2	1	0.33	wake up	4
1	3	0	1	16	0.33	die/kill	5
3	7.5	5.5	3	2	0.40	go out/put out	6
4	9.5	5.5	1.5	0.5	0.42	sink	7
3.5	7.5	6	2	2	0.47	learn/teach	8
5	10.5	3	2.5	0	0.48	melt	9
5.5	9	3.5	3	0	0.61	stop	10
8	7.5	4	1.5	0	1.07	turn	11
10.5	7.5	2	1	0	1.40	dissolve	12
7	5	2	5	2	1.40	burn	13
8.5	5.5	5	1	0	1.55	destroy	14
8	5	5	3	0	1.60	fill	15
7.5	4.5	5	4	0	1.67	finish	16
5	3	3	8	0	1.67	begin	17
11	6	3	1	0	1.83	spread	18
8.5	4.5	5	3	0	1.89	roll	19
10	5	5	1	0	2.00	develop	20
10.5	4.5	4.5	0	0.5	2.33	get lost/lose	21
12	4.5	3.5	0	1	2.67	rise/raise	22
8.5	3	8	1.5	0	2.83	improve	23
12	4	3.5	1.5	0	3.00	rock	24
14	2.5	1.5	1	1	5.60	connect	25
11	1.5	4.5	4	0	7.33	change	26
15	2	3	1	0	7.50	gather	27

-3

WATP 61 Languages

	Verb Pair	A/C	Count of Types				
			A	C	E	L	S
1	boil	0.03	1.7	47.7	3.5	4.5	2.0
2	freeze	0.05	2.0	39.3	6.0	3.0	2.7
3	dry	0.07	3.0	44.8	6.7	2.0	2.0
4	melt	0.09	3.8	42.2	7.0	4.0	1.0
5	wake up	0.13	4.5	34.5	12.5	2.0	5.5
6	sink	0.13	5.2	38.3	7.5	5.0	3.0
7	stop	0.15	5.8	38.8	6.3	6.5	2.5
8	go out/put out	0.18	6.0	34.2	8.2	2.3	8.0
9	die/kill	0.21	3.5	17.0	5.5	1.0	33.5
10	fill	0.23	7.3	31.8	8.8	8.5	1.5
11	develop	0.26	7.0	26.5	11.8	5.3	5.3
12	dissolve	0.39	13.2	33.7	6.2	6.0	1.0
13	burn	0.40	11.0	27.5	6.5	4.0	11.0
14	learn/teach	0.44	10.0	22.8	7.8	3.8	16.5
15	roll	0.44	10.5	23.7	14.8	8.5	1.0
16	turn	0.45	12.5	27.7	12.3	7.5	1.0
17	improve	0.51	11.2	22.0	13.8	3.5	6.5
18	get lost/lose	0.55	14.5	26.2	6.0	7.8	3.5
19	finish	0.57	13.5	23.7	8.8	8.7	5.0
20	rock	0.57	16.0	28.0	7.5	7.5	0.0
21	rise/raise	0.59	14.7	25.0	8.5	4.0	7.3
22	break	0.86	16.5	19.2	15.2	5.8	3.8
23	destroy	1.01	19.2	19.0	13.8	4.0	4.0
24	spread	1.45	23.5	16.2	9.3	8.0	0.0
25	change	1.55	22.5	14.5	10.5	10.0	2.0
26	connect	1.69	24.5	14.5	7.5	6.0	4.5
27	split	2.00	24.0	12.0	15.5	5.5	2.5

Our experiment

The goal is to directly target the extent to which family and/or areality predict the distribution of coding types in a sample of Eurasian (+ a few African) languages.

1. Attempt to verify/correct/flesh out data with the help of language experts.
2. Settle discrepancies between Haspelmath (1993) and WATP contributors for particular languages (e.g., Georgian, Japanese, Swahili).
3. Unify methods of deciding on coding type, particularly within families.

1. Clustering/dimensionality reduction

The basic data item in the WATP and the database used for the talk is a set of codings for a particular pair of verbal meanings for a particular language, e.g.

Even, ‘learn/teach’, {L, A}

Using this data, we can compute differences between verbal pairs in different languages (‘learn/teach’ in Even vs. ‘learn/teach’ in English) and between different verbal pairs in the same language (‘learn/teach’ vs. ‘boil(in.)/boil(tr.)’ in English).

Measuring distances

The Jaccard distance metric was used to measure differences between verbal pairs:

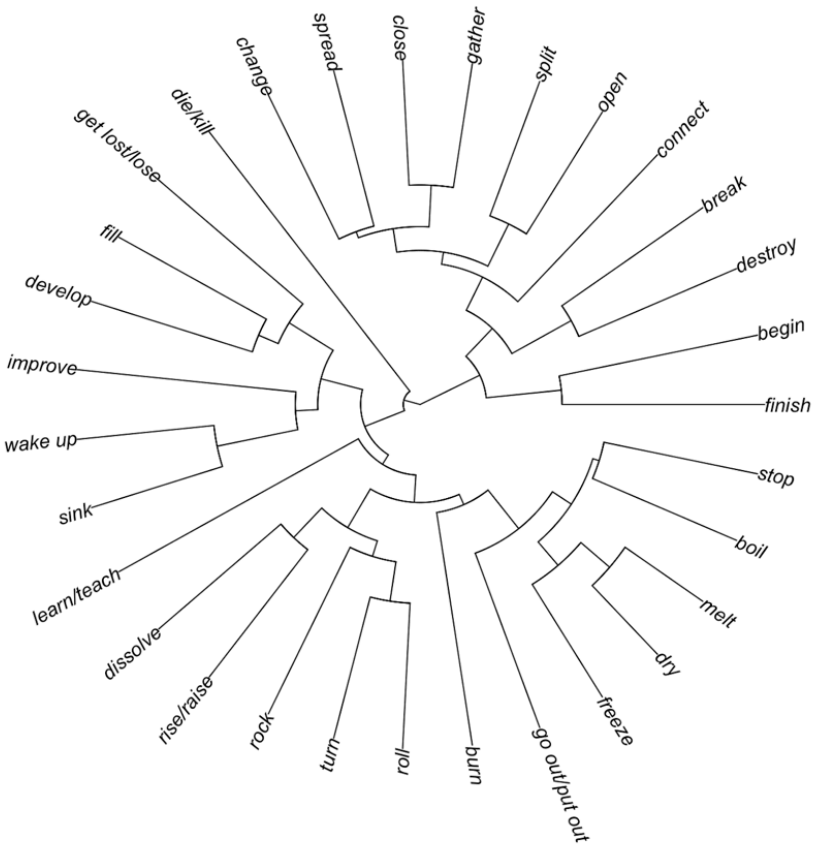
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

$$d_J(A,B) = 1 - J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}.$$

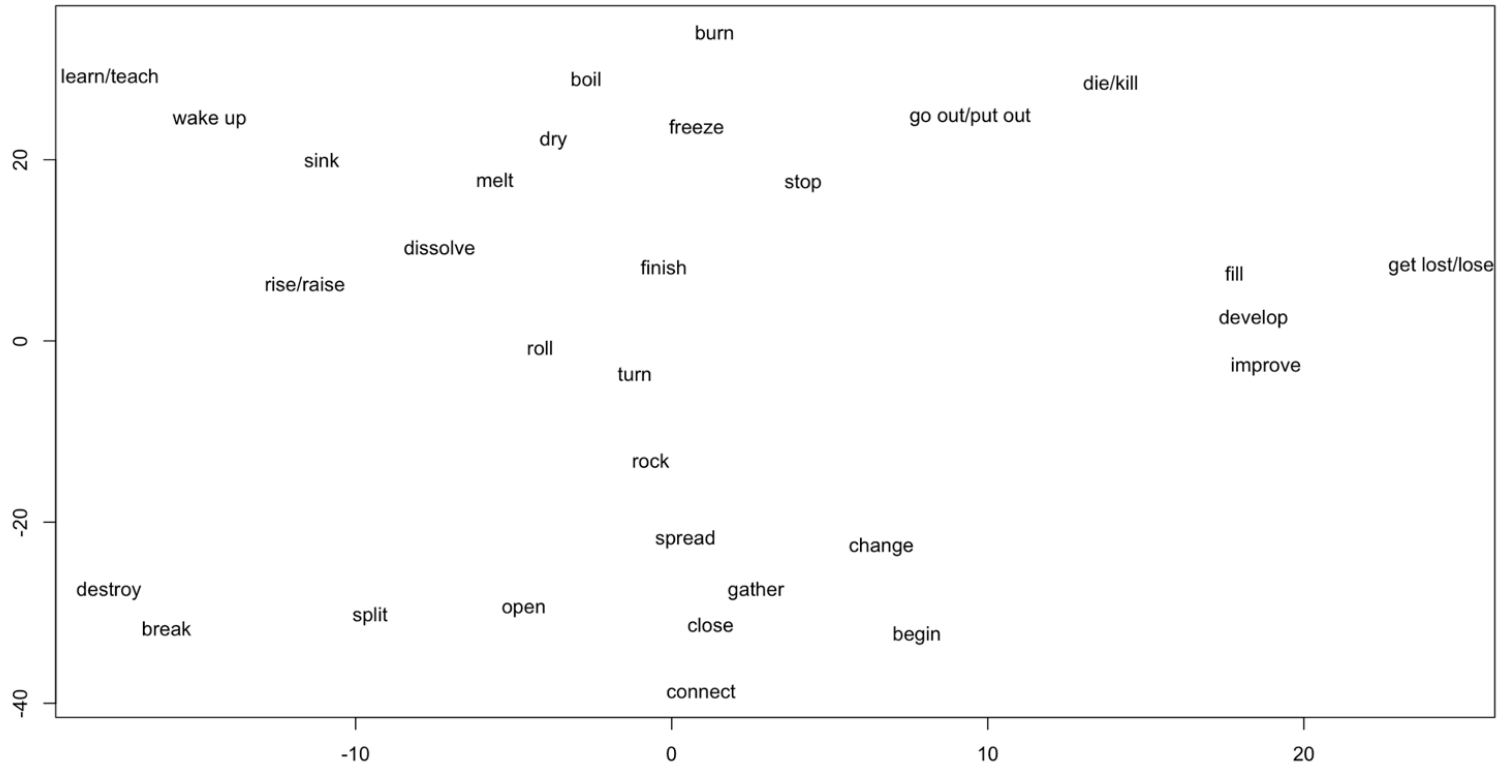
E.g. $D_{\text{'learn/teach'}}(\text{Even}_{\{L,A\}}, \text{Lezgian}_{\{A\}}) = 1 - ½ = ½$

Jaccard distances for two given verb pairs were summed for all languages to measure distances between verb pairs, and distances for all verb pairs for given two languages were summed to measure distances between these languages.

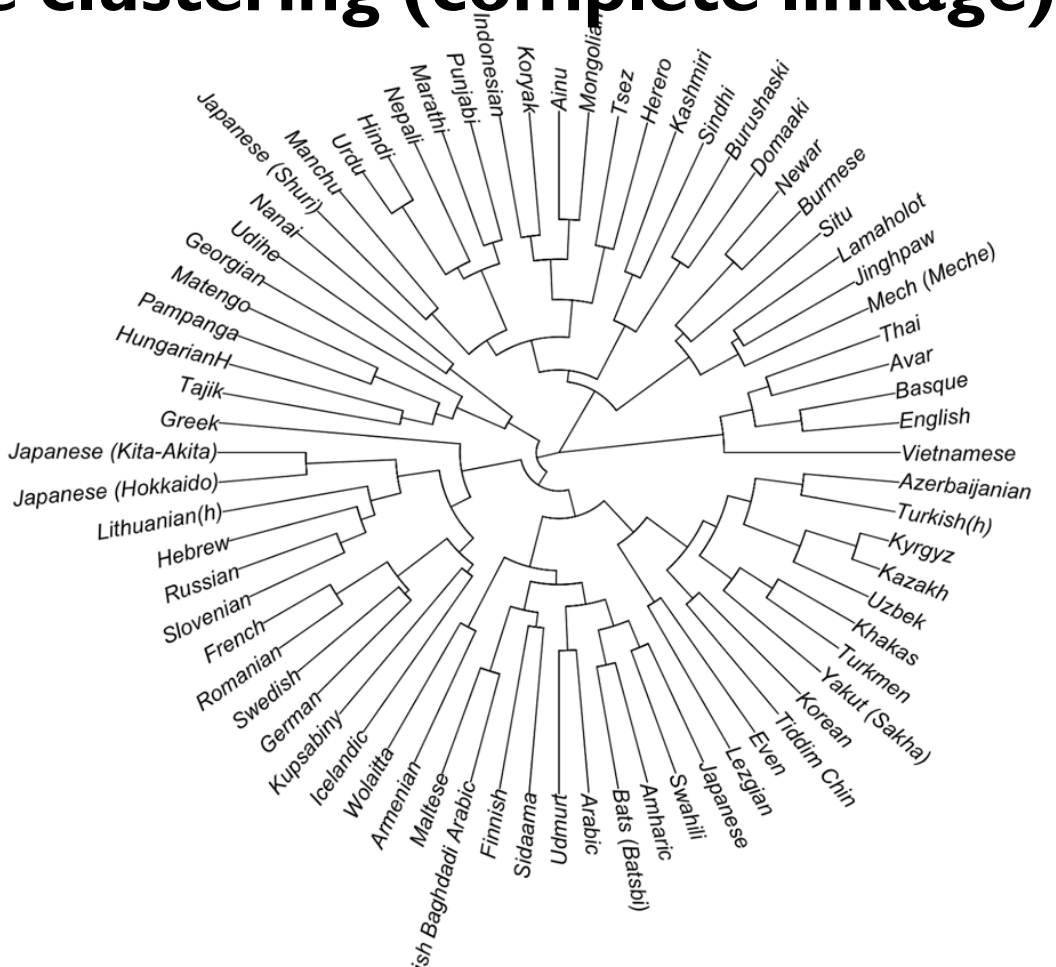
Verb clustering (complete linkage)



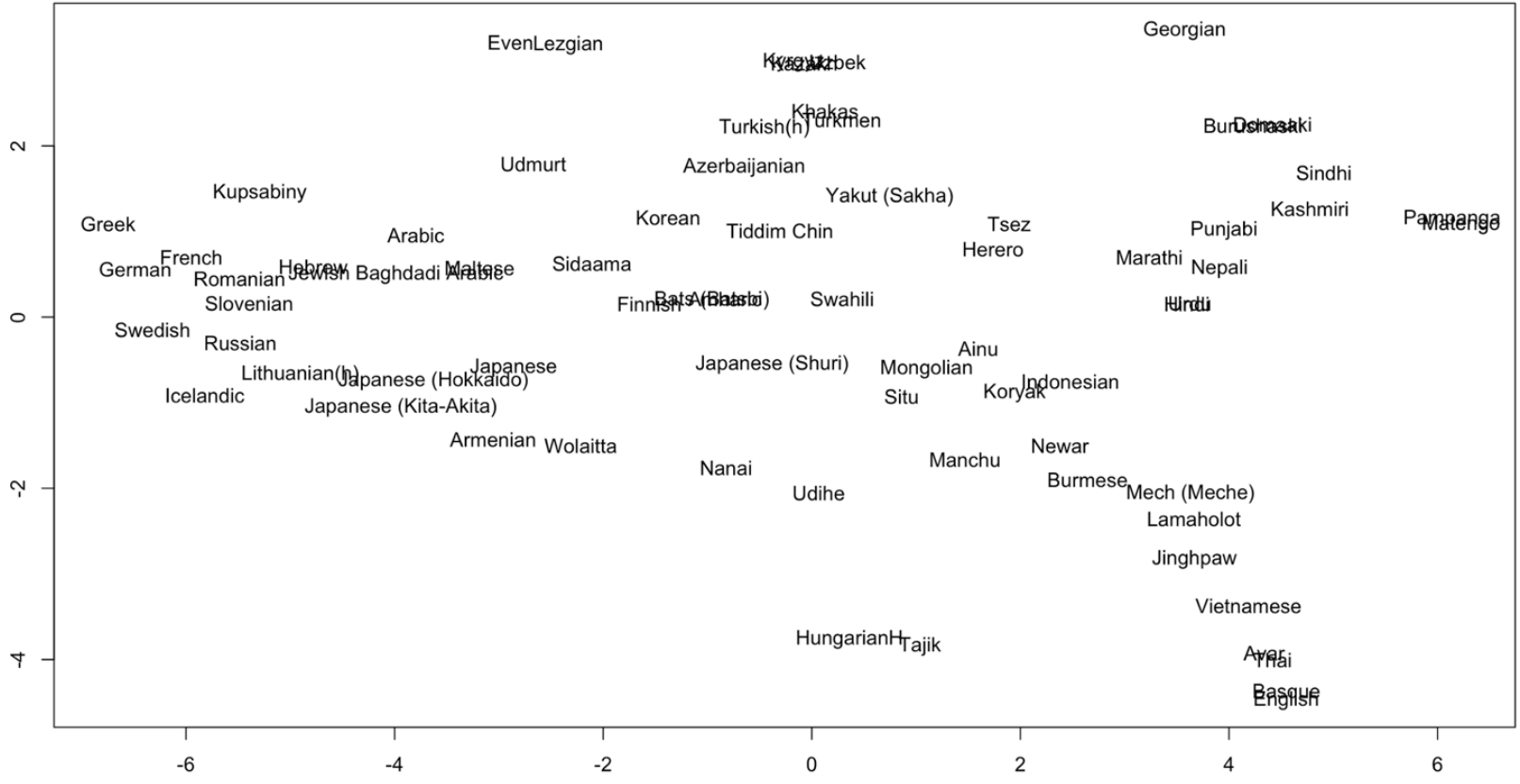
2-D projection (t-SNE)



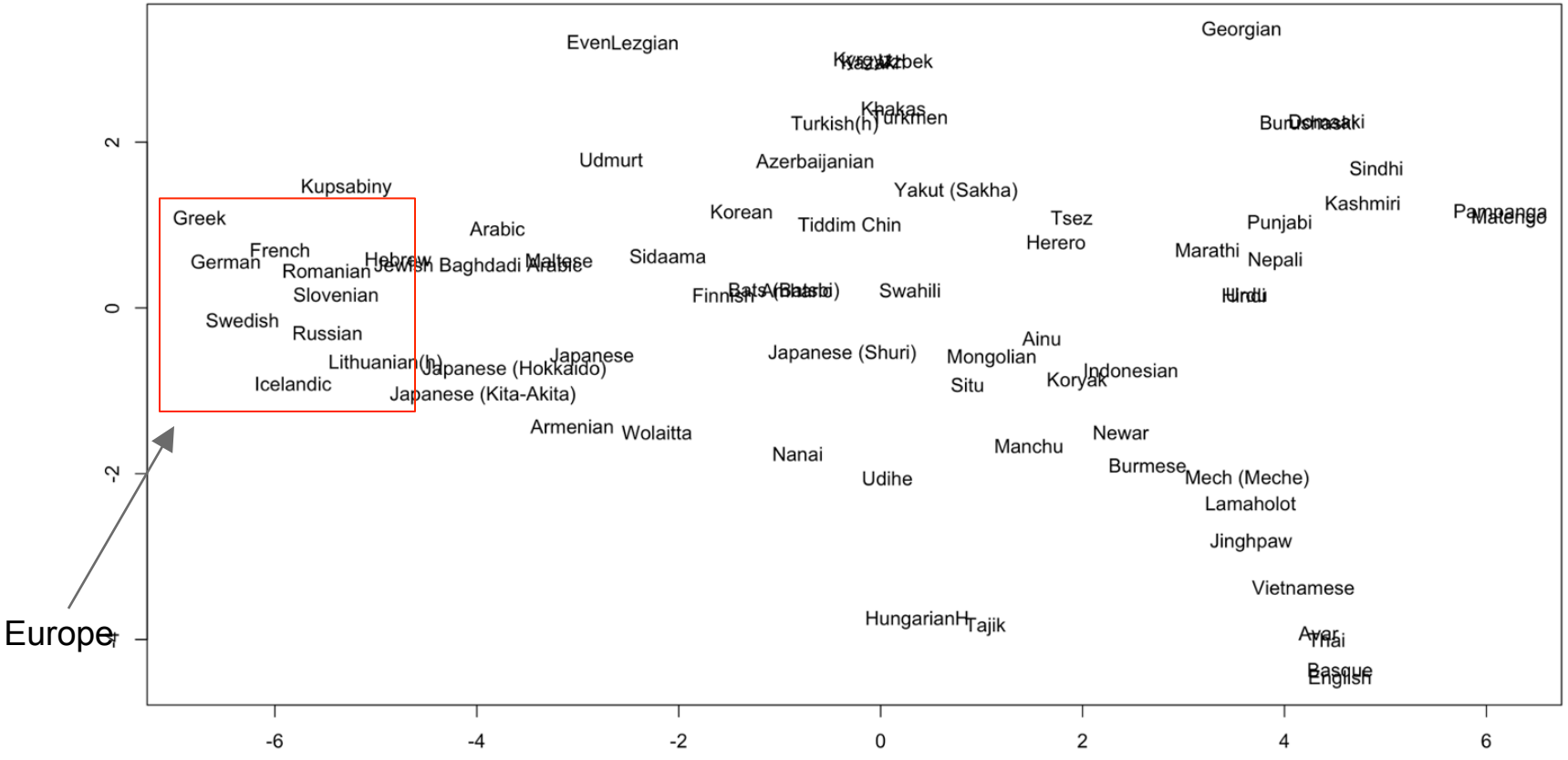
Language clustering (complete linkage)



2-D projection (t-SNE)



2-D projection (t-SNE)



The areal perspective

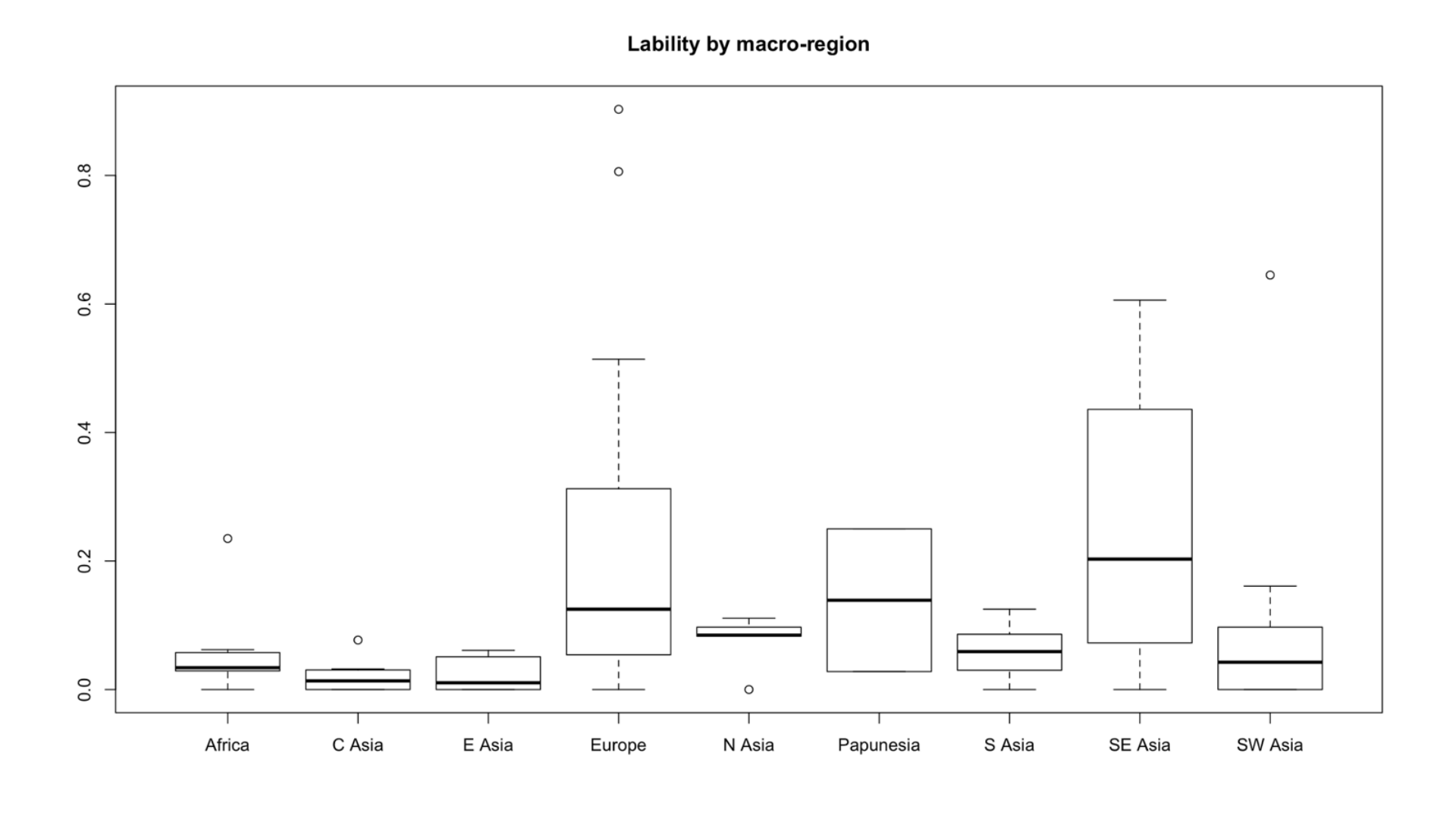
Predictive Areality Theory (Bickel and Nichols 2006 and subsequent):

- Use geographical areas determined on the basis of non-linguistic criteria (geography, settlement history, etc.) to predict linguistic properties.
- Combine with the Family Bias Method (the first lecture) to identify biases within genealogical groupings.

(The inverse procedure of traditional areal linguistics, which identifies linguistic areas only if the properties in question are shared by unrelated languages.)

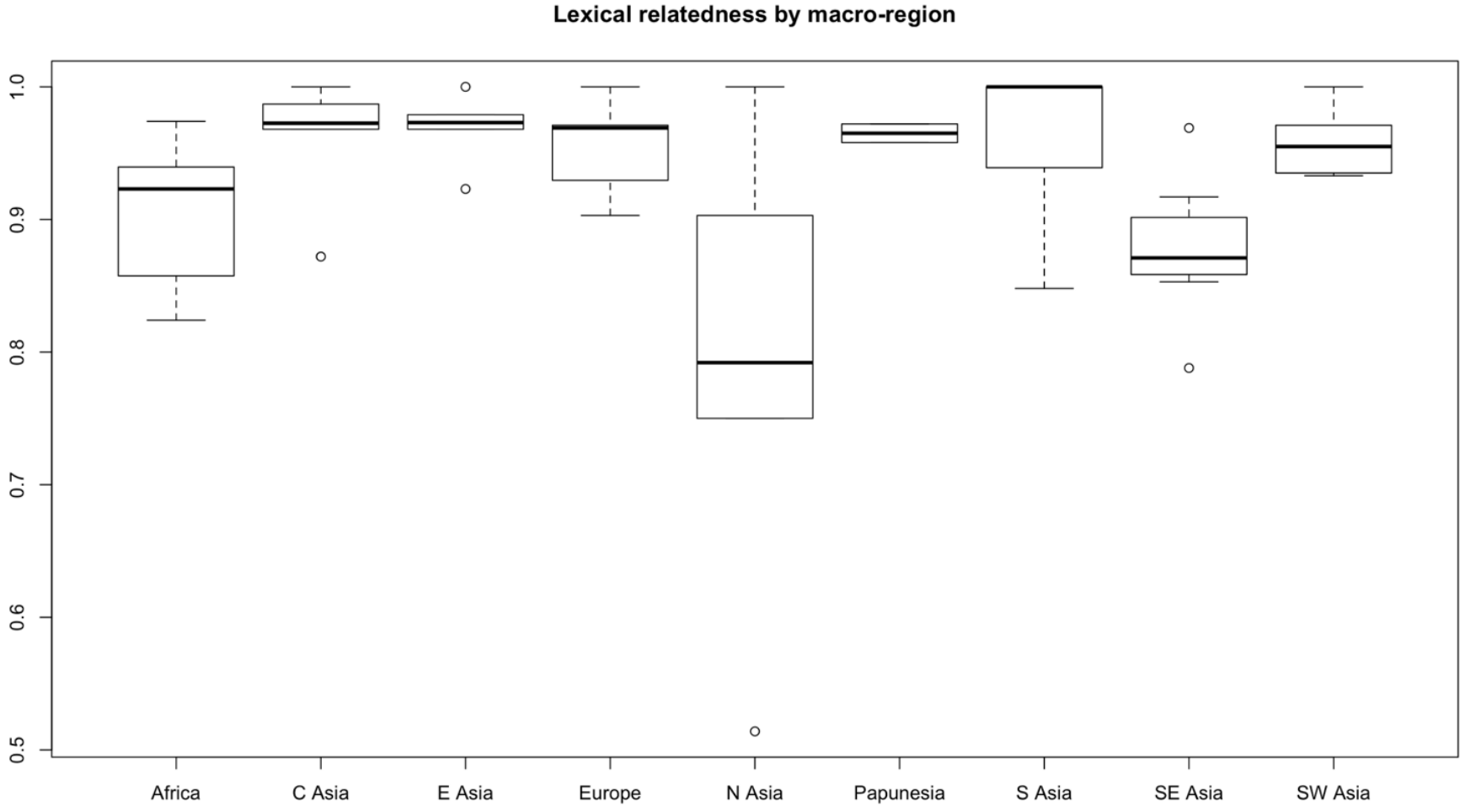
Areas used in this study

- Africa – very few data points in the sample, certainly wrt the area's ca. 2000 languages.
- Central Asia
- East Asia
- Europe
- North Asia
- Papunesia
- South Asia
- Southeast Asia


Lability by macro-region
0.0
0.2
0.4
0.6
0.8
Africa
C Asia
E Asia
Europe
N Asia
Papunesia
S Asia
SE Asia
SW Asia

Most areas don’t have much lability.

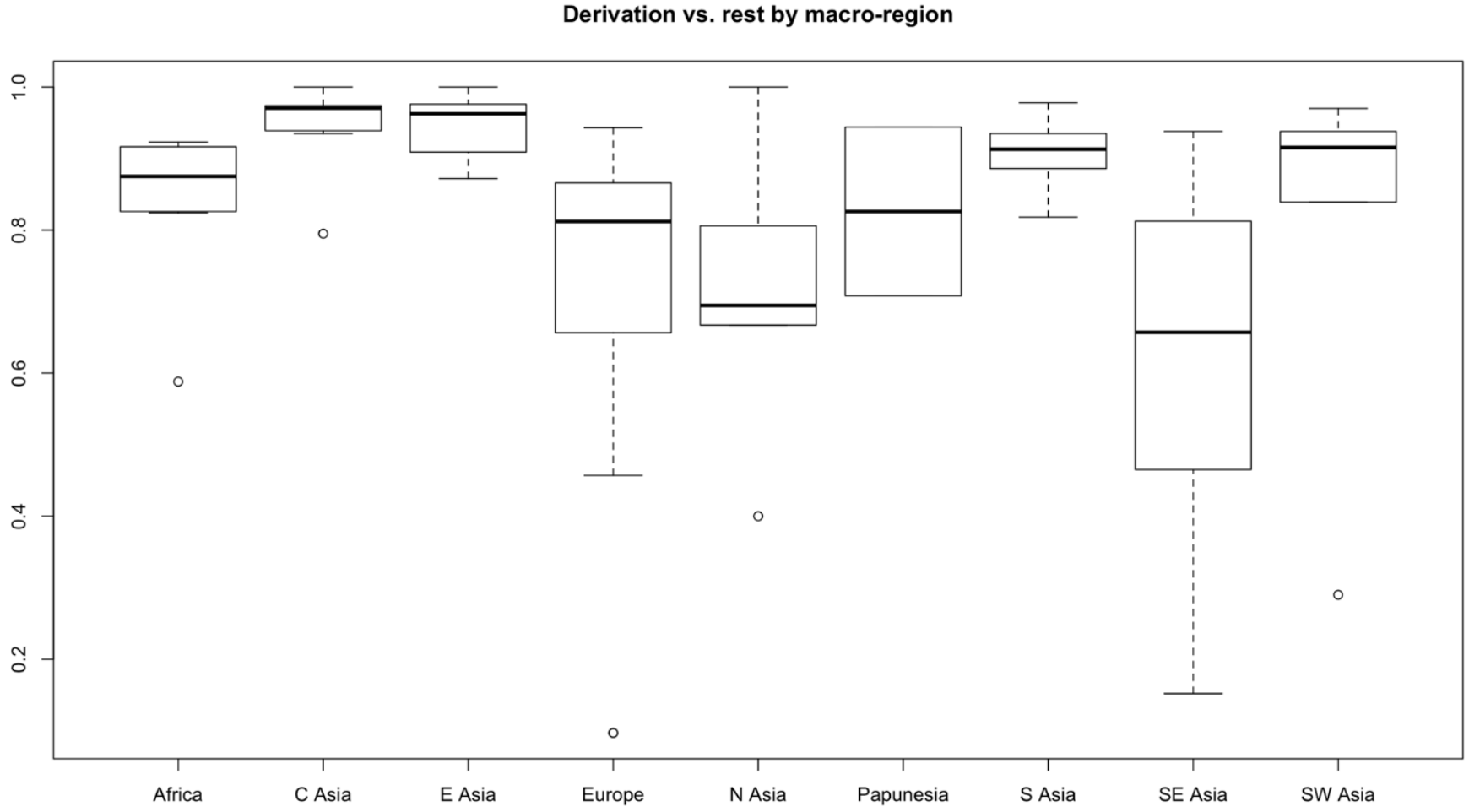
Southeast Asia and Europe are the most diverse, followed by Papunesia.


Lexical relatedness by macro-region
1.0
0.9
0.8
0.7
0.6
0.5
Africa
C Asia
E Asia
Europe
N Asia
Papunesia
S Asia
SE Asia
SW Asia

Contrariwise

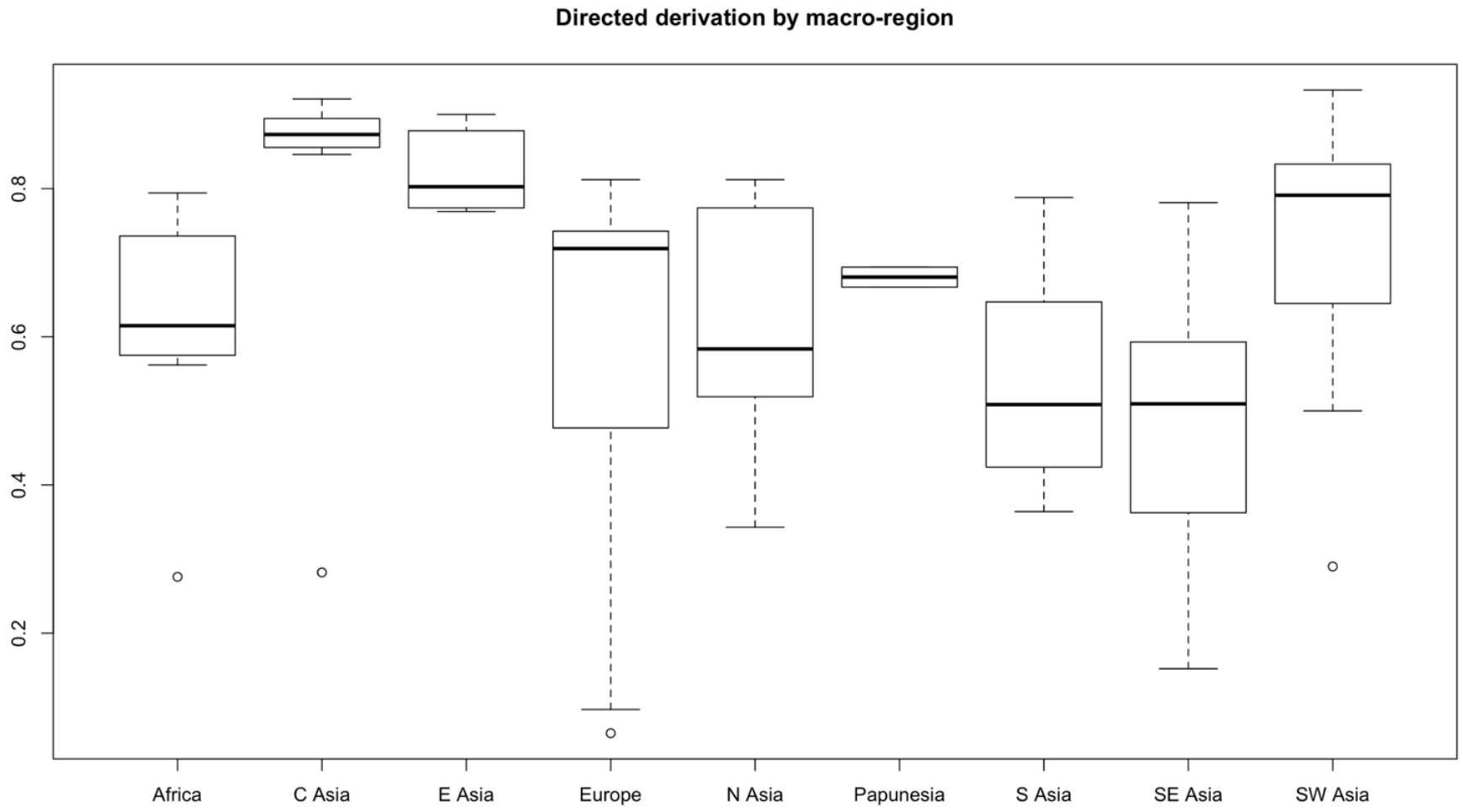
Most areas show a bias for lexical relatedness.

Some areas – North Asia, followed by Europe and South Asia – are more diverse in this respect.


Derivation vs. rest by macro-region
1.0
0.8
0.6
0.4
0.2
Africa
C Asia
E Asia
Europe
N Asia
Papunesia
S Asia
SE Asia
SW Asia

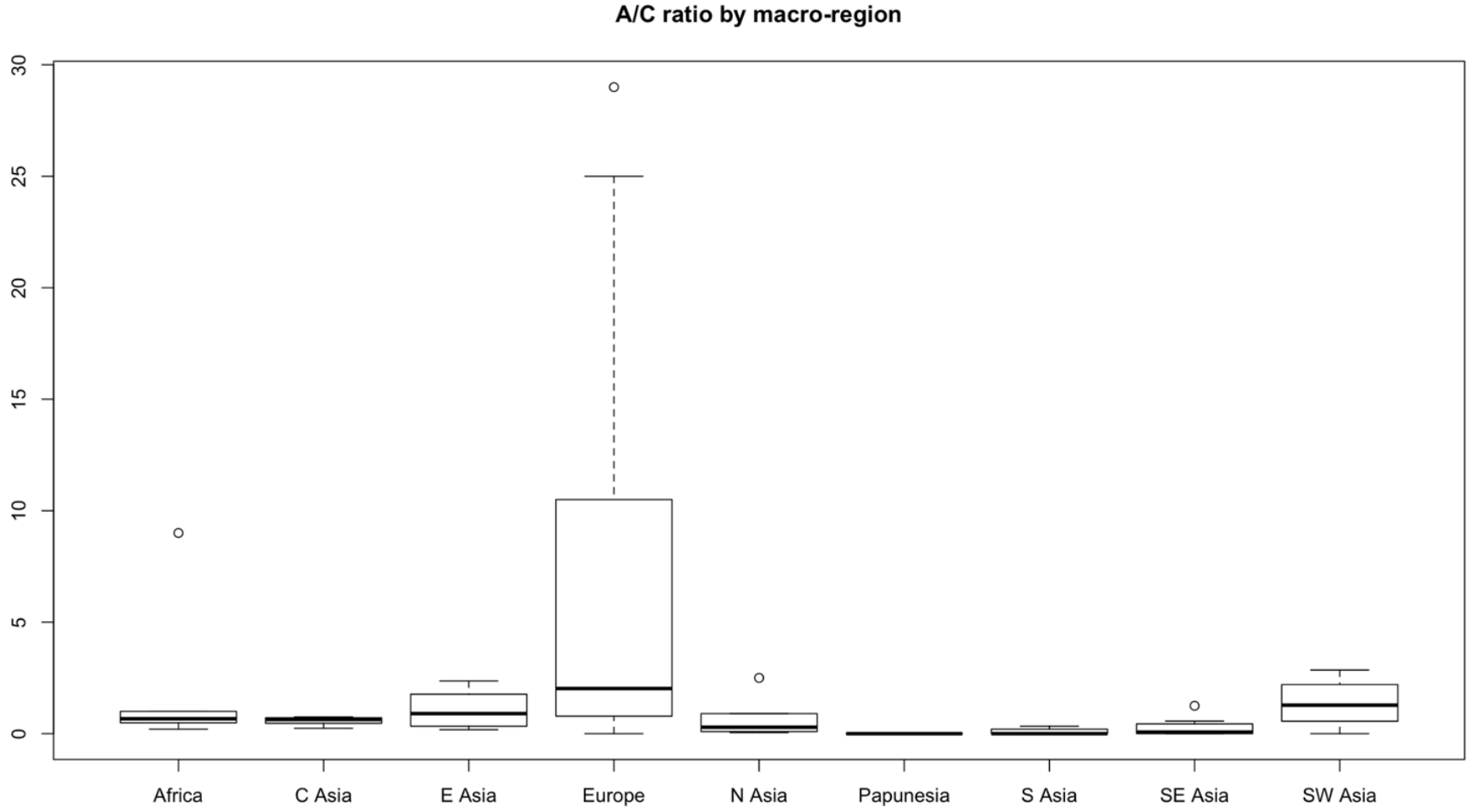
Most areas prefer derivation (as opposed to lability and suppletion).

But Southeast Asia seems to tolerate lability and suppletion, as do Europe, North Asia, and Papunesia to an extent.


Directed derivation by macro-region
0.2
0.4
0.6
0.8
Africa
C Asia
E Asia
Europe
N Asia
Papunesia
S Asia
SE Asia
SW Asia

Here we find more variation between areas, but most areas show a mild bias in favor of directed derivation (A/C, as opposed to S,L,E).

On the whole, Asia (except for SW Asia), seems to be more tolerant of undirected derivation.


A/C ratio by macro-region
30
25
20
15
10
5
0
Africa
C Asia
E Asia
Europe
N Asia
Papunesia
S Asia
SE Asia
SW Asia

Finally, all areas show a very low A/C ratio, except for Europe, which is more diverse and has more languages with a preference for anticausatives.

Findings

By eyeballing the plots, we see lots of apparent areal clustering, which suggests contact-induced change.

Findings

Significance testing (likelihood-ratio tests of mixed models with and without region as a fixed effect with random intercepts for phyla as a random effect) shows that **A/C ratio** and **lexical relatedness** show a significant dependence on the region when the phylum is controlled for.

Findings

Inverse tests show that phylum has a significant effect over and above that of region in the case of A/C ratio and a borderline significant effect in the case of lexical relatedness ($p = 0.054$). I.e., we see interaction of these factors.

Lability, **directed derivation**, and **derivation vs. the rest** are poorly explained by either predictor.

Data for most Eurasian regions outside Europe are clearly insufficient.

Also

We haven't characterized Eurasia in comparison to any other area of the world, so we don't know if there is a particular preferred lexical valence orientation characteristic of Eurasia.

In conclusion

Previous research on lexical valence orientation (and related notions) has pointed to several functional explanations for the distribution of associated coding types.

They have also pointed to areal symptoms of an 'event-based' causal factors.

Our study strongly corroborates Haspelmath's (1993) east-west split in Eurasia wrt A/C ratio, as well as some of Nichols et al.'s (2004) findings.

However

Most observed areal effects don't reach statistical significance.

Genealogical affiliation (phylum) has a significant effect for A/C ratio (and to a lesser extent, lexical relatedness).

Nonetheless

These non-significant areal preferences might provide hypotheses worth pursuing.

The bottom line

‘Event-based’ factors show a signal in Eurasian preferences for the ways in which causal:noncausal pairs are coded.

Next – to try to evaluate the contribution of ‘functional’ factors to the distribution of coding strategies within Eurasia (and beyond), and to see whether they interact with areality.

Thank you!