Explaining the cross-linguistic distribution of argument-coding patterns
Universität Potsdam
March 21, 2023

Measuring predictability of argument realization patterns in bivalent verbs

Sergey Say
serjozhka@yahoo.com
University of Potsdam

Introductory examples

(1) *Karl wartet auf Marie* NOM; auf+ACC
'Karl is waiting for Marie.'

(2) *Mir fehlt ein Euro* DAT; NOM
'I am one Euro short.'

- both are conventional for native speakers
- both are (?) semantically motivated
- to the same degree?
- can this be measured and explained?

Structure of the talk

- Background and goals
- The database: BivalTyp
- Predictability: introducing *π*
- Results
  - verbs
  - languages
- Conclusions and implications

Background and goals

- All (?) languages have minor (a.k.a. non-canonical) valency patterns
- (Until recently) underrepresented in typological research
- Goal: to fill this gap for bivalent verbs

Background and goals

- Which factors determine valency class assignment in individual languages?
- To what extent are valency classes similar across languages? To what extent are they variable?

The database: BivalTyp

Sergey Say (ed.). 2020–... BivalTyp: Typological database of bivalent verbs and their encoding frames.
St. Petersburg: Institute for Linguistic Studies, RAS.
(Available online at https://www.bivaltyp.info)*

*All credit for building the web-page goes to Dmitry Nikolaev

The database: BivalTyp

- First-hand data provided by language experts
  - St. Petersburg-style typology

- Questionnaire with 130 verbs given in context
  - Wordlist-based approach [Nedjalkov 1969, Bossong 1998, Nichols et al. 2004, Nichols 2008, Malchukov & Comrie (eds.) 2015, etc.]

The database: BivalTyp

#21 (Peter was crossing the river in a boat)

‘Peter	**reached**	the bank’
X		Y

#22 (The wall was covered with fresh paint)

‘Peter	**touched**	the wall’ (and got dirty)
X		Y

=> Two pre-defined arguments (X, Y) for each predicate

The database: BivalTyp

- The sample: currently 92 languages, mainly spoken in Northern Eurasia

The database: BivalTyp

- A big **THANK YOU** to language experts

Anna Alexandrova, Daria Alfimova, Ekaterina Aplonova, Peter Arkadiev, David Avellan-Hultman, Aleksandra Azargaeva, Mislav Benić, Sandra Birzer, Alena Blinova, Nadezhda Bulatova, Denis Creissels, Michael Daniel, Varvara Diveeva, Sergey Dmitrenko, Vladimir Fedorov, Timothy Feist, Dmitry Gerasimov, Elena Gorbova, Olga Gorickaja, Ingunn Hreinberg Indriðadóttir, Ildar Ibragimov, Emil Ingelsten, Vasilisa Kagirova, Maxim Kloczenko, Maria Khazhomia, Maria Kholodilova, Mikhail Knyazev, Elena Kolpachkova, Daria (Suetina) Konior, Yukari Konuma, Elena Kordi, Richard Kowalik, Kirill Kozhanov, Irina Külmoja, Olga Kuznecova, Timur Maisak, Anastasia (Borisovna) Makarova, Anastasia (Leonidovna) Makarova, Ramazan Mamedshaxov, Solmaz Merdanova, Stepan Mikhajlov, Daria Mischenko, Zarina Molochieva, George Moroz, Rasul Mutalov, Galina Nekrasova, Johanna Nichols, Dmitry Nikolaev, Ajtalina Nogovitsyna, Sofia Oskolskaya, Maria Ovsjannikova, Anastasia Panova, Elena Perekhvalskaja, Natalia Perkova, Krasimira Petrova, Inna Popova, Maria Pupynina, Tatiana Repnina, Neige Rochant, Alexander Rostovtsev-Popiel, Daria Ryzhova, Sergey Say, Ekaterina Sergeeva, Ksenia Shagal, Mayya Shlyakhter, Natalia Stoynova, Ksenia Studenikina, Evgenija Teplukhina, Mladen Uhlik, Anastasia Vasilisina, Arseniy Vydrin, Natalia Zaika, Andreja Žele, Ekaterina Zheltova, Vasilisa Zhigulskaja, Anastasia Zhuk

The database: BivalTyp

- 10694 entries (130 predicates in 92 lgs – 1266 gaps):
  - language ID
  - predicate ID
  - verb
  - valency pattern
  - (for 30 languages: interlinearized examples)
- The database is searchable, sortable and mappable by predicates, languages, valency patterns, etc.

- **Further contributions are very welcome!**

Predictability: introducing *π*

How to measure “predictability”?

- Discrete semantic roles?
  - if defined on *a priori* grounds, they are not suitable for empirical-typological studies (Bickel et al. 2014, etc.)
- An alternative:
  - Use other languages as predictors, that is, as proxies for the meaning of arguments

Predictability: introducing π

- Assume there are only 4 verbs that belong to a certain class A in a given L1:

	L1
V_i	A
V_j	A
V_k	A
V_l	A
…	

Predictability: introducing *π*

- Explore the encoding of the corresponding verbs in L2:

	L1	L2
V_i	A	W
V_j	A	W
V_k	A	W
V_l	A	W

…

- The valency patterns of these 4 verbs in L2 seem to be predictable given the system of L1

Predictability: introducing *π*

- Explore the encoding of the corresponding verbs in L2:

	L1	L2
V_i	A	X
V_j	A	Y
V_k	A	Z
V_l	A	W
…		

- The valency patterns of these 4 verbs in L2 seem to be totally unpredictable given the system of L1

Predictability: introducing π

- Real-data example

	Russian	Kalmyk
‘be afraid’	NOM_GEN	NOM_ABL
‘reach’	NOM_GEN	NOM_DAT
‘avoid’	NOM_GEN	NOM_ACC
‘forfeit’	NOM_GEN	NOM_ABL
‘be ashamed’	NOM_GEN	NOM_ABL

- From the perspective of Russian, the encoding of the Kalmyk equivalents of ‘be afraid’, ‘forfeit’ and ‘be ashamed’ is more predictable than that of the other two verbs (‘reach’, ‘avoid’)

Predictability: introducing π

- Individual predicate, two languages:

$$\pi\,(V_i)(L_j \rightarrow L_k) = p\,(Class(V_i, L_k)|Class(V_i, L_j))$$

e.g.

$$\pi\,(\text{'reach'})(\text{Russian} \rightarrow \text{Kalmyk}) = 1/5 = 0.2$$

$$\pi\,(\text{'be\_afraid'})(\text{Russian} \rightarrow \text{Kalmyk}) = 3/5 = 0.6$$

Predictability: introducing π

- Individual predicate, one language: explore its behaviour from the perspective of as many other languages as there are available

$$\pi(V_i)(L_k) = \frac{\sum_{j=1}^{n} \pi(V_i)(L_j \rightarrow L_k)}{n}$$

e.g. π ('be_afraid') (Kalmyk) = 0.53

Predictability: introducing π

- Individual predicate, many languages: average predictability

$$\pi(V_i) = \frac{\sum_{j=1}^{n} \pi(V_i)(L_j)}{n}$$

e.g. π ('be_afraid') = 0.42

Results

(1) *Karl wartet auf Marie* [NOM_aufACC]

'Karl is waiting for Marie.'

$\pi = 0.12$

(2) *Mir fehlt ein Euro* [DAT_NOM]

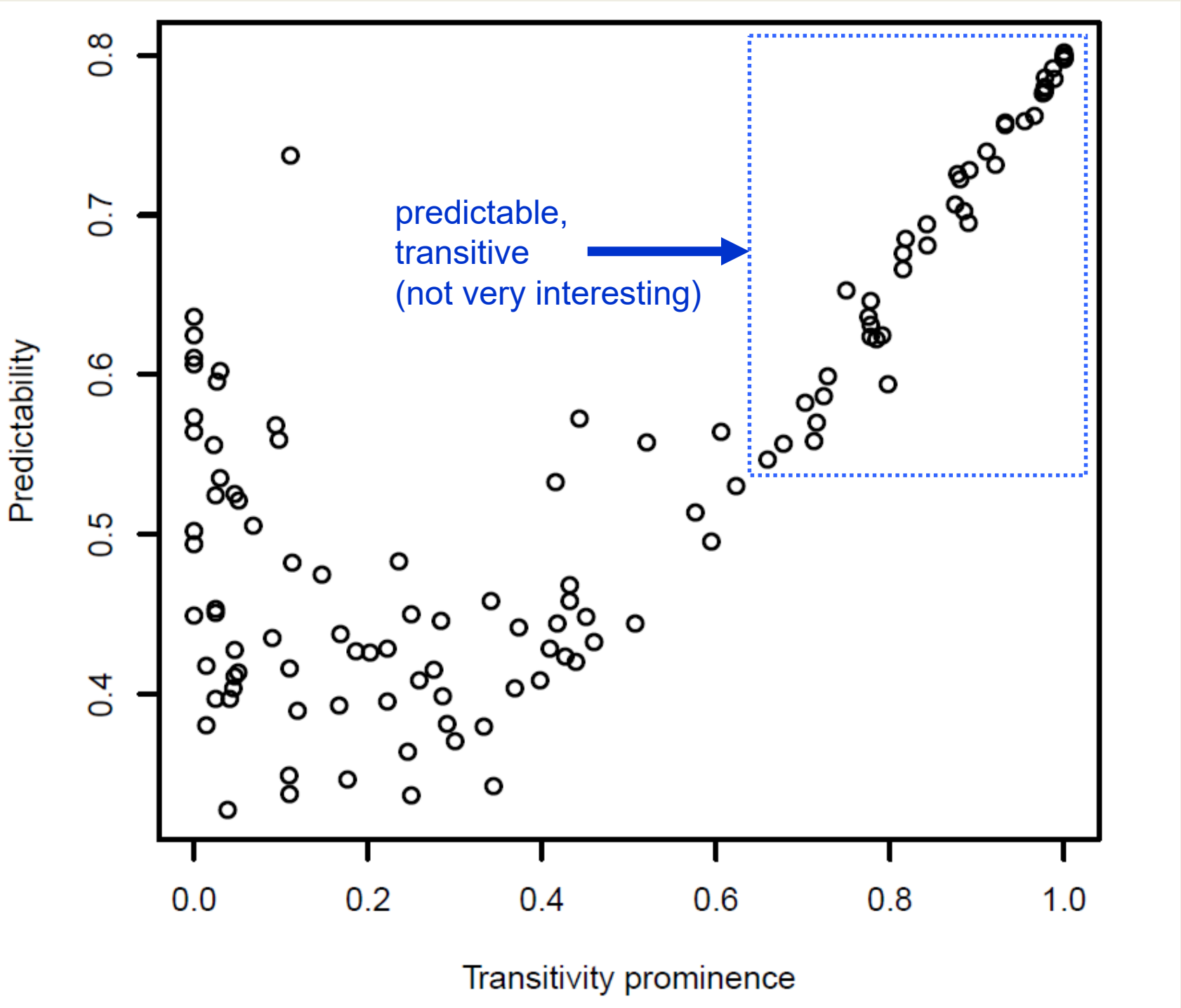
'I am one Euro short.'

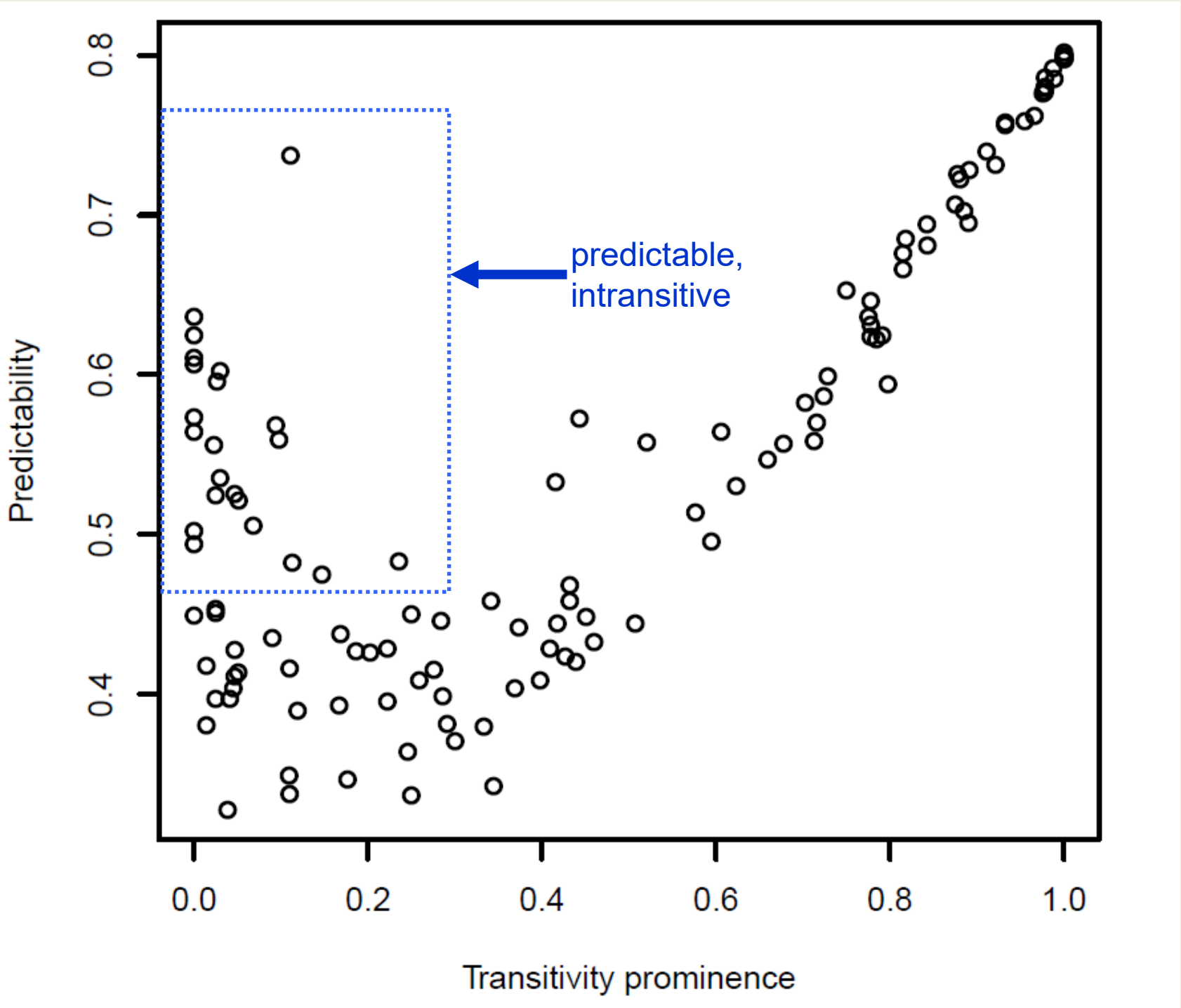
$\pi = 0.46$

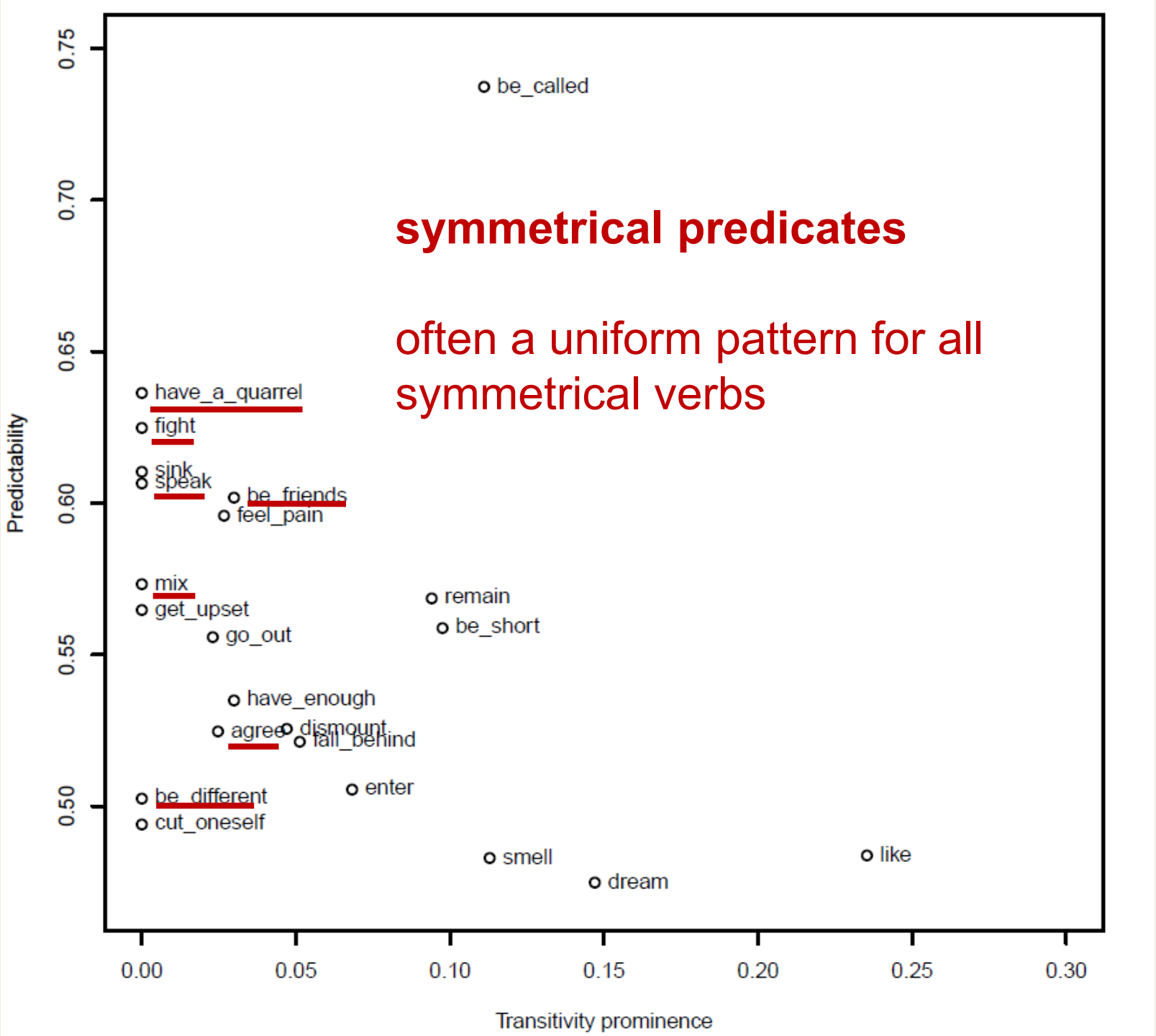
Results: verbs

- As expected, highly transitive predicates display high predictability scores
- E.g. the list of predicates with $\pi > 0.80$: all these verbs are invariably transitive in the data set

	predictability	transitivity ratio		predictability	transitivity ratio
'make'	0.80	1.00	'bend'	0.80	1.00
'eat'	0.80	1.00	'wash'	0.80	1.00
'drink'	0.80	1.00	'kill'	0.80	1.00
'take'	0.80	1.00	'sing'	0.80	1.00
'break'	0.80	1.00	'melt'	0.80	1.00
'write'	0.80	1.00	'fry'	0.80	1.00


predictable,
transitive
(not very interesting)
Predictability
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0
Transitivity prominence


predictable,
intransitive
Predictability
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0
Transitivity prominence


symmetrical predicates
often a uniform pattern for all symmetrical verbs
Predictability
Transitivity prominence
0.75
0.70
0.65
0.60
0.55
0.50
0.00
0.05
0.10
0.15
0.20
0.25
0.30
be_called
have_a_quarrel
fight
sink
speak
be_friends
feel_pain
mix
get_upset
go_out
remain
be_short
have_enough
agree
dismount
fall_behind
enter
be_different
cut_oneself
smell
dream
like

Results: verbs

- Shughni (> Iranian; Tajikistan)

Najiba	*Safina=qati*	*δêd*	*kiǩ-t*
PN	PN=INS	quarrel	do-3SG

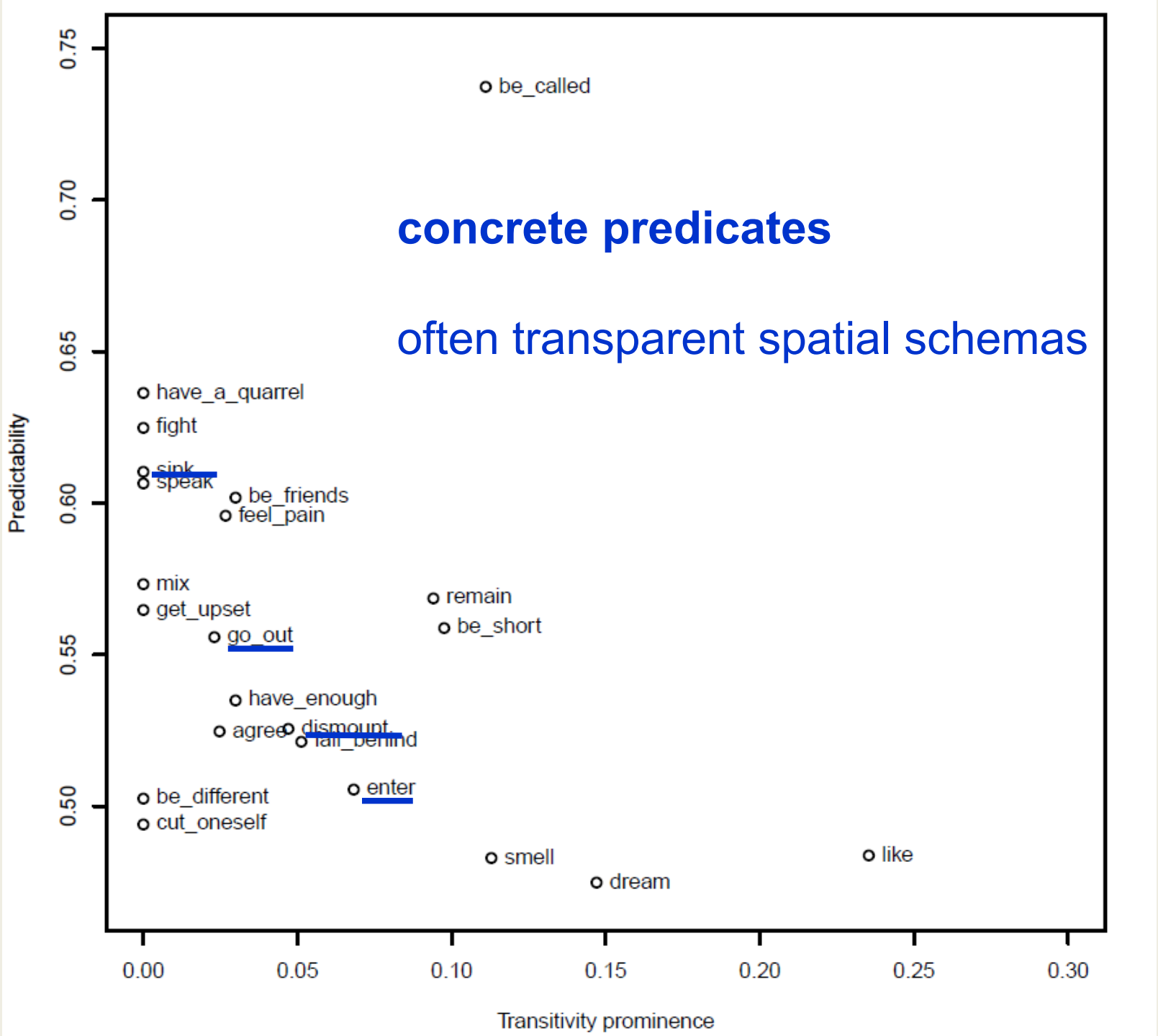
'Najiba is fighting with Safina.'

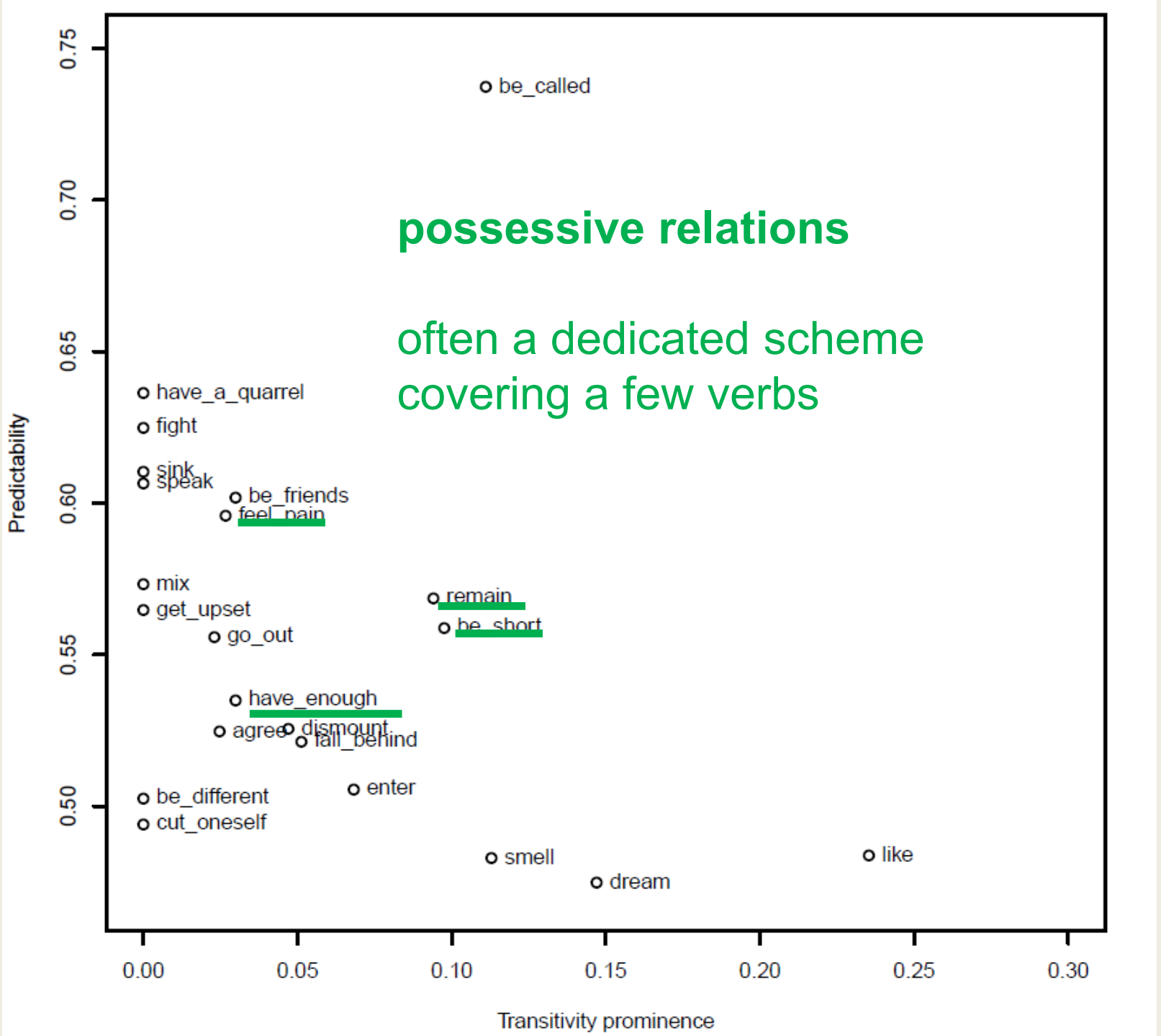
Asal	*ǩuvd=qati*	*alalaš*	*sut*
honey	milk=INS	mix	go.M.PST

'The honey got mixed with the milk.'

Ahmed	*Saida=qati*	*rozi*	*sut*
PN	PN=INS	contented	go.M.PST

'Ahmed agreed with Saida.'


concrete predicates
often transparent spatial schemas
Predictability
Transitivity prominence
0.75
0.70
0.65
0.60
0.55
0.50
0.00
0.05
0.10
0.15
0.20
0.25
0.30
be_called
have_a_quarrel
fight
sink
speak
be_friends
feel_pain
mix
get_upset
go_out
remain
be_short
have_enough
agree
dismount
fall_behind
enter
be_different
cut_oneself
smell
dream
like


possessive relations
often a dedicated scheme covering a few verbs
Predictability
Transitivity prominence
0.75
0.70
0.65
0.60
0.55
0.50
0.00
0.05
0.10
0.15
0.20
0.25
0.30
be_called
have_a_quarrel
fight
sink
speak
be_friends
feel_pain
mix
get_upset
go_out
remain
be_short
have_enough
agree
dismount
fall_behind
enter
be_different
cut_oneself
smell
dream
like

Results: verbs

- Turkish: the GEN_NOM pattern – 5 verbs in the data set

Mehmed-in	*şimdi*	*on*	*lira-sı*	*kal-dı*
PN-GEN	now	ten	lira-P.3	remain-PST

'Now Mehmet has 10 liras left.'

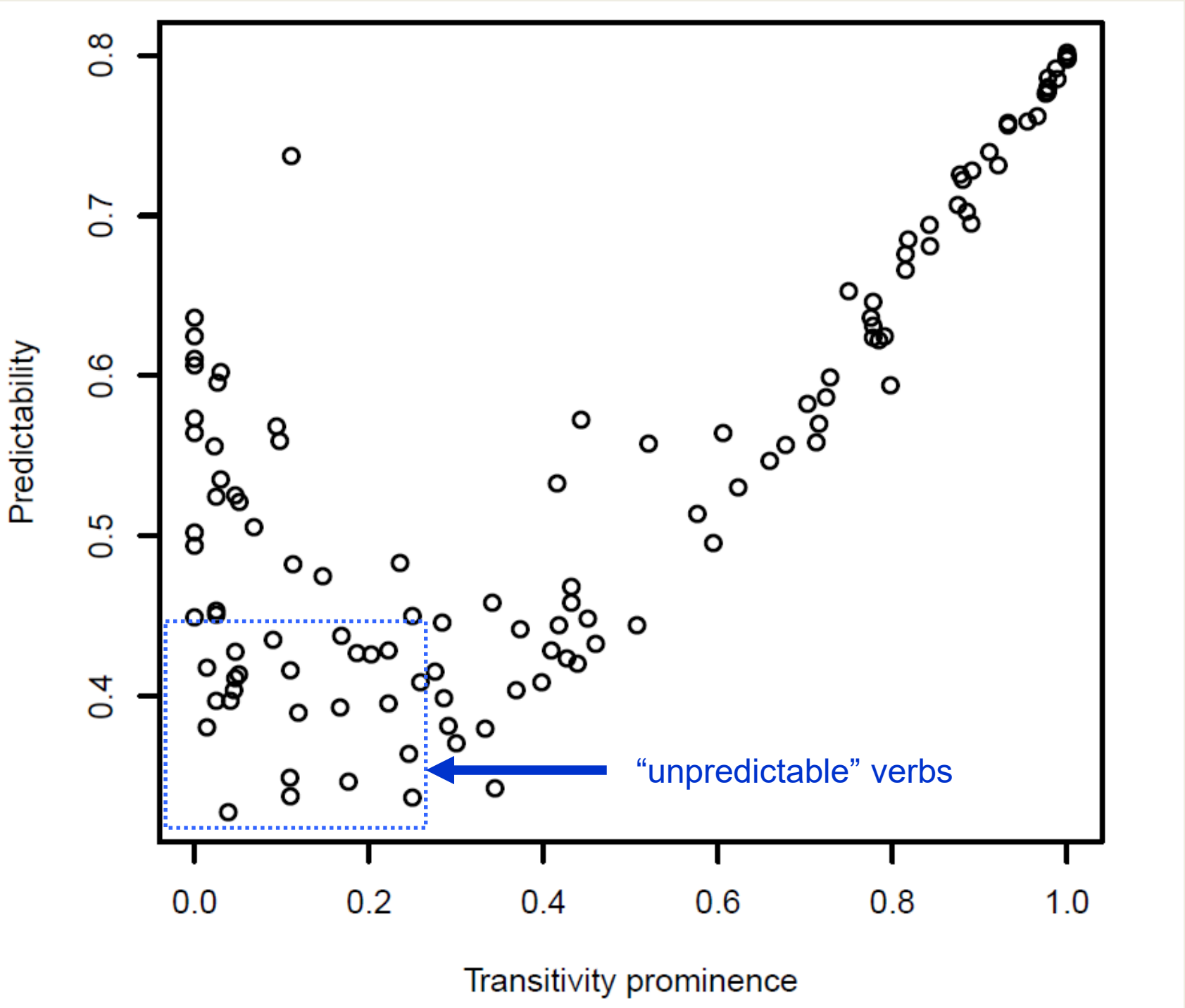
Mehmed-in	*araba-sı*	*var*
PN-GEN	car-P.3	there_is

'Mehmet has a car.'

Mehmed-in	*baş-ı*	*ağrı-yor*
PN-GEN	head-P.3	ache-PRS

'Mehmet has a headache.'

+ also 'have (illness)', 'be short'


Predictability
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0
Transitivity prominence
“unpredictable” verbs

Results: verbs

- The least predictable verbs are mainly **psychological** verbs
- Top 13 verbs with the lowest π-value (< 0.395):

	predictability	transitivity		predictability	transitivity
be_content	0.33	0.04	encounter	0.37	0.3
fall_in_love	0.34	0.25	influence	0.38	0.33
be_surprised	0.34	0.11	take_offence	0.38	0.01
be_fond	0.34	0.34	envy	0.38	0.29
be_squeamish	0.35	0.18	resemble	0.39	0.12
marvel	0.35	0.11	trust	0.39	0.17
enjoy	0.36	0.25			

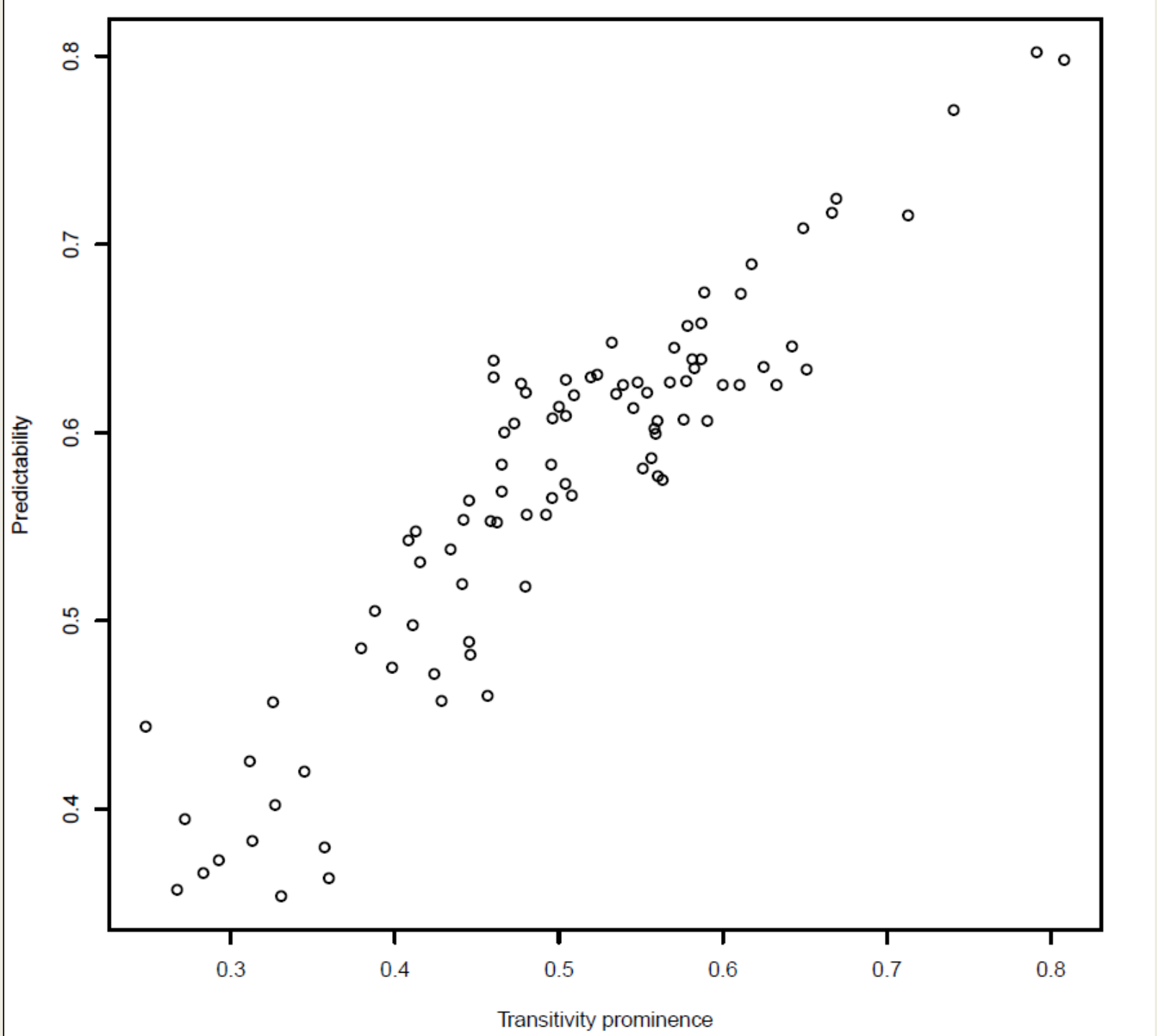
Results: verbs

- Some psychological verbs in Aghul:
  - various patterns
  - no obvious motivation

	verb	**pattern**
'be afraid'	*guč'a-*	DAT_POST.ELAT
'be glad'	*šad-x.u-*	ABS_DAT
'be squeamish'	*karih-t:i* + 'be'	DAT_ABS
'be content'	*rezi-di* + 'be'	ABS_SUPER
'fall in love'	*k:an-x.u-*	DAT_ABS
'trust'	*quχ.a-*	ABS_POST
'be angry'	*qel ke-*	DAT_SUBCONT.ELAT
'be surprised'	*ʔalamat x.u-*	ABS_SUPER
'take offence'	*qel aq'.u-*	ERG_SUBCONT.ELAT
'despise'	*alčaq-t:i fac.u-*	TR
'be shy'	*neč-t:i* + 'be'	DAT_SUBCONT.ELAT

Results: languages

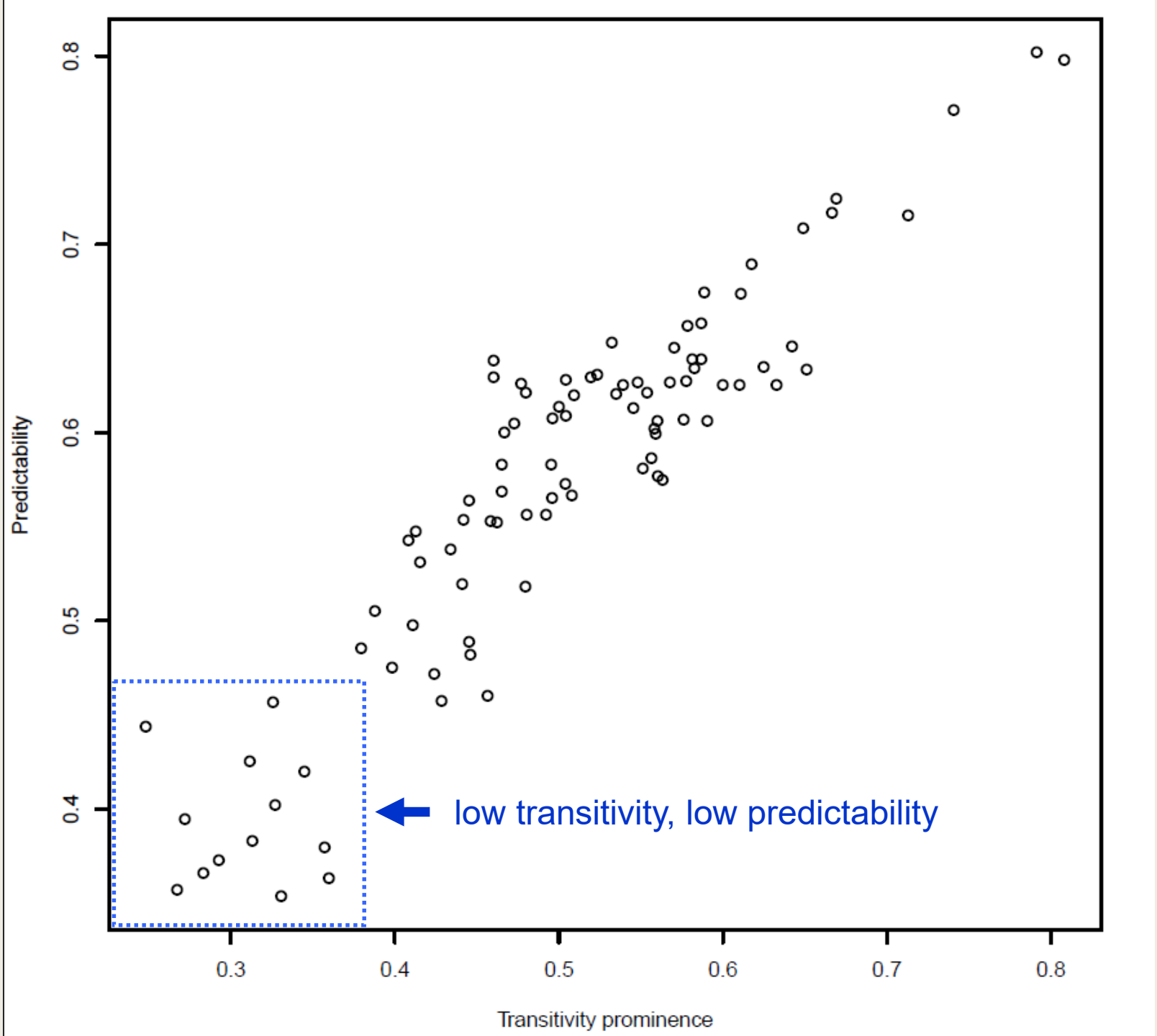
- Languages can differ in the degree of semantic motivation behind their valency classes
- Theoretically, this can be captured through calculating mean π-values across their lexica
- In reality, the main predictor of this mean is the language's transitivity prominence: the ratio of transitive verbs

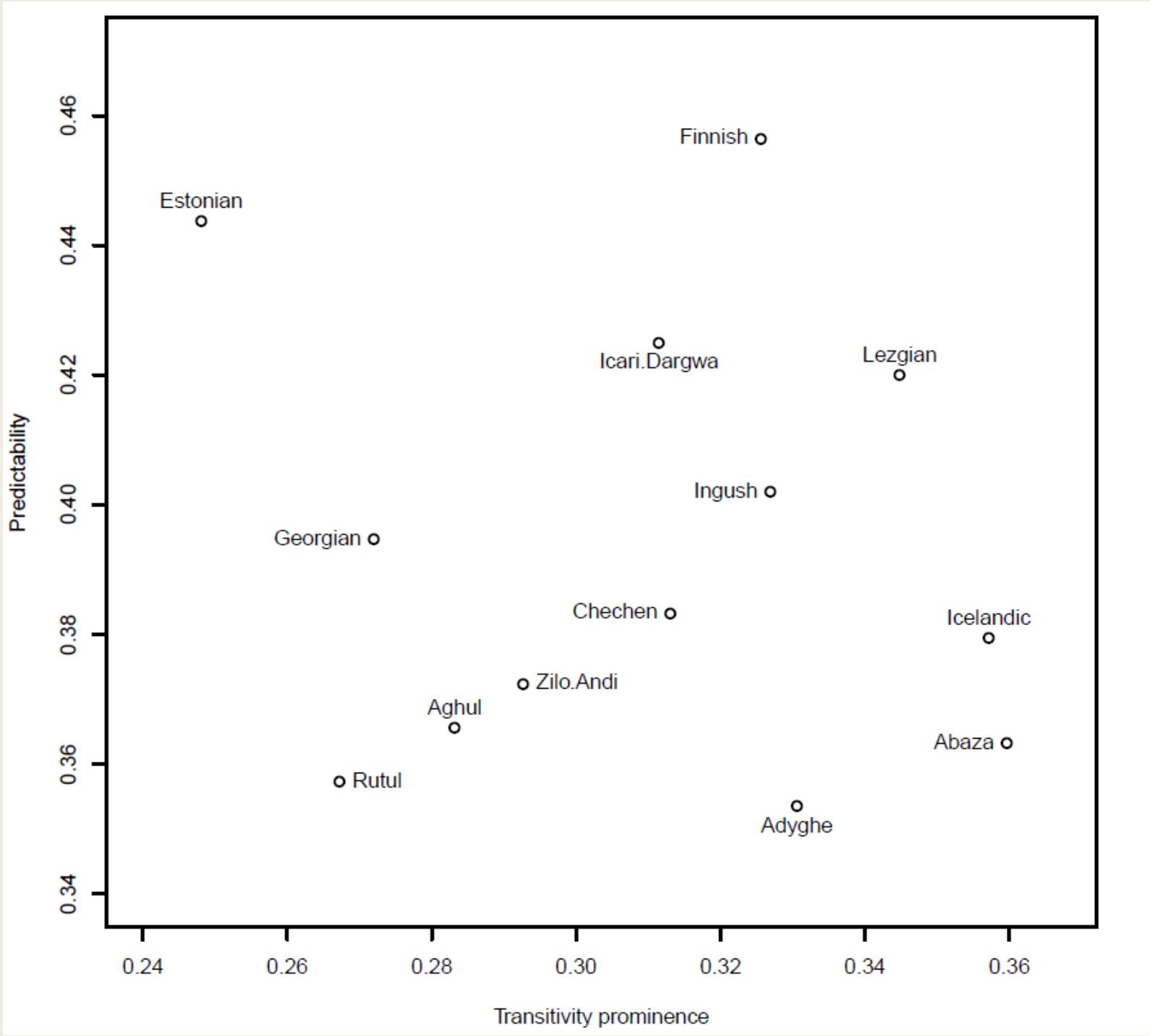

Predictability
0.4
0.5
0.6
0.7
0.8
0.3
0.4
0.5
0.6
0.7
0.8
Transitivity prominence

Results: languages

- The higher the ratio of transitive verbs, the higher the mean predictability

- However, there are promising fluctuations between languages with comparable transitivity prominence


Predictability
0.4
0.5
0.6
0.7
0.8
0.3
0.4
0.5
0.6
0.7
0.8
Transitivity prominence
low transitivity, low predictability

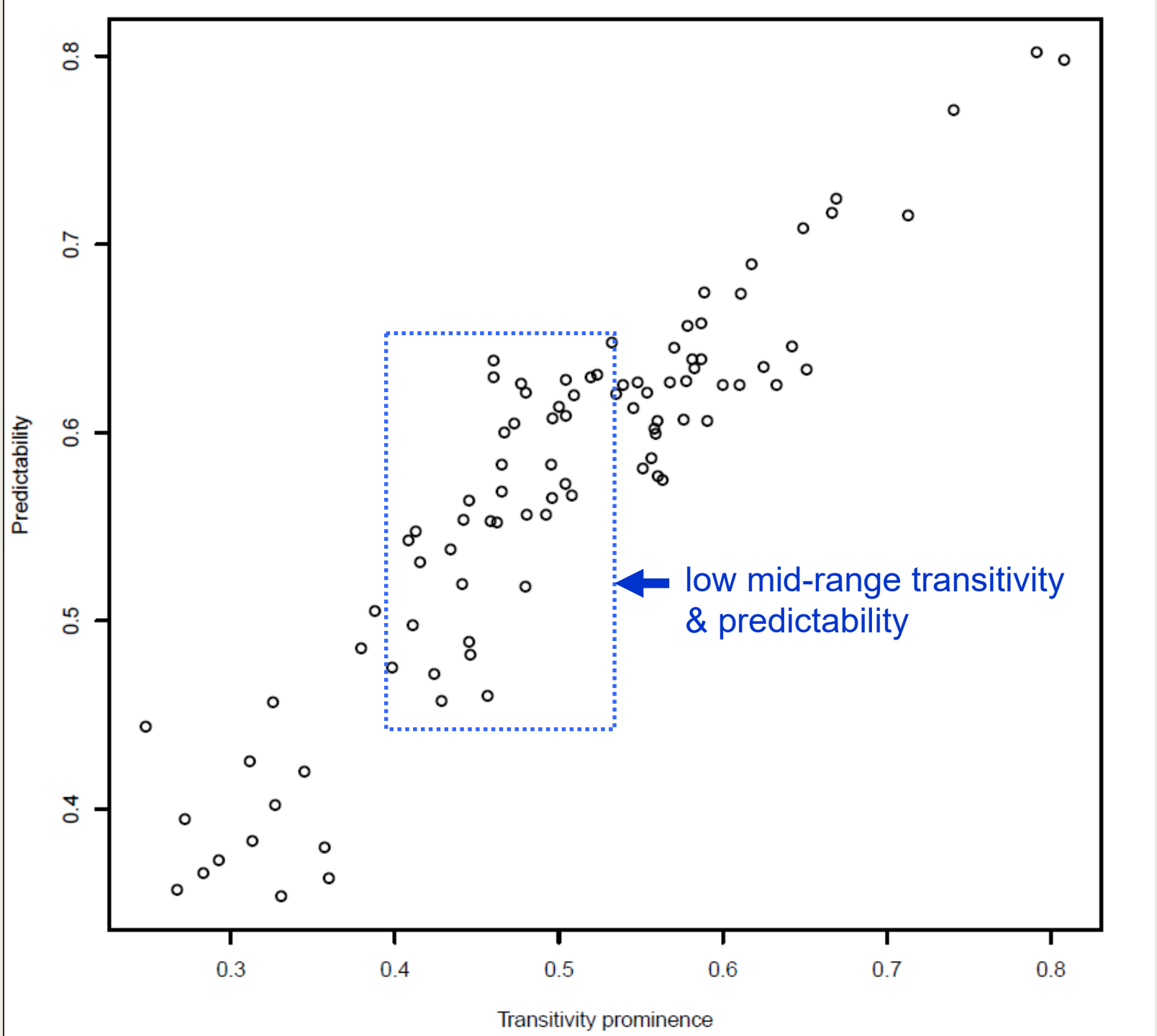

Predictability
0.46
0.44
0.42
0.40
0.38
0.36
0.34
Finnish
Estonian
Icari.Dargwa
Lezgian
Ingush
Georgian
Chechen
Icelandic
Zilo.Andi
Aghul
Abaza
Rutul
Adyghe
0.24
0.26
0.28
0.30
0.32
0.34
0.36
Transitivity prominence

Results: languages

- Indigenous languages of the Caucasus and Baltic Finnic are among languages with
  - the lowest transitivity prominence
  - the lowest predictability

- Northwest Caucasian (Abkhaz-Adyghe) languages have especially low predictability values
  - related to their non-transparent lexicalized preverbs?

- All things being equal, lower population size correlates with lower predictability (higher complexity)?


Predictability
0.4
0.5
0.6
0.7
0.8
0.3
0.4
0.5
0.6
0.7
0.8
Transitivity prominence
low mid-range transitivity
& predictability

Results: languages

- Predictability in the low mid-range area:
  - Turkic, Mongolic >
  - most Finno-Ugric, Baltic & Slavic >
  - Mande, Irish, Udi
- Simple case systems in Turkic & Mongolic? Few grammaticalized adpositions?
  - to be explored

Conclusions and implications

- Valency patterns are neither fully predictable, nor fully idiosyncratic.
  - $\Rightarrow$ Not to say it's sensational
  - $\Rightarrow$ But this can be measured!

- Verb meanings differ in terms of predictability of the respective verbs' valency behaviour

- Highly transitive verbs are cross-linguistically stable (and invariably predictable)

Conclusions and implications

- Low transitivity, high predictability:
  - symmetric predicates: ‘fight’, ‘speak’, ‘get mixed’, ‘agree’
  - some motion-related verbs: ‘go out’, ‘dismount’, ‘enter’, drown’
  - some verbs related to possession: ‘be short’, ‘have enough’, ‘remain’ + ‘feel pain’

Conclusions and implications

- Low transitivity, low predictability:
  - most verbs of emotions and other psychological verbs: ‘rejoice’, ‘be surprised’, ‘trust’, ‘fall in love’, ‘enjoy’…

  ⇒ No empirical justification for Experiencers and Stimuli as unified roles

  ⇒ Multiple models of construal in terms of more concrete (spatial) schemas

  ⇒ These models are largely idiosyncratic

Conclusions and implications

- Languages also differ in the degree of their verbs' predictability
- Highly transitive languages ignore the distinctions made by other languages and display higher predictability
- Given a certain level of transitivity prominence, languages fluctuate in terms of their predictability

  => genus-level genealogical signal?

- Intuitively, more "predictable" languages employ less variegated metaphors for abstract meanings

THANK YOU!

Selected references

- Bickel, Balthasar, Taras Zakharko, Lennart Bierkandt & Alena Witzlack-Makarevich, 2014. Semantic role clustering: An empirical assessment of semantic role types in non-default case assignment. Studies in language, 38 (3). Advances in research in semantic roles. 485–511.
- Bossong, Georg. 1998. Le marquage de l'expérient dans les langues d'Europe. In: Feuillet (ed.). Actance et valence dans les langues de l'Europe. Berlin: Mouton de Gruyter. 259–94.
- Dowty, David. 1991. Thematic proto-roles and argument selection. Language 67, 547–619.
- Fillmore, Charles. 1968. The case for case. In: E. Bach & R. Harms (eds), Universals in Linguistic Theory. Holt, Rinehard and Winston, New York. 1–88.
- Haspelmath, Martin. 1993. More on the typology of inchoative / causative verb alternations. In: Comrie, Bernard & Maria Polinsky (eds.) *Causatives and Transitivity.* Amsterdam: Benjamins. 87–120.
- Hopper, P.J., Thompson, S.A. 1980. Transitivity in grammar and discourse. Language, 56. 2. 251–350.
- Lehmann, Ch. 1991, Predicate classes and PARTICIPATION. H. Seiler & W. Premper, Partizipation. Das sprachliche Erfassen von Sachverhalten. Tübingen: G. Narr (LUS, 6). 183–239.

Selected references

- Lazard, Gilbert. 1994. L'actance. Paris: Presses Universitaire de France.
- Malchukov, A. L. 2005. Case pattern splits, verb types and construction competition. In M.Amberber and H. de Hoop, H. (eds), Competition and Variation in Natural Languages: the Case for Case. Oxford: Elsevier, 73-117.
- Malchukov, A. & B. Comrie (eds.). 2015. Valency classes in the world's languages. Berlin: Mouton de Gruyter.
- Nedjalkov, V.P. 1969. Nekotorye verojatnostnye universalii v glagolnom slovoobrazovanii. In: F. Vardul' (ed.). Jazykovye universalii i lingvisticheskaja tipologija. Moscow: Nauka. 106-114.
- Nichols, Johanna. 2008. Why are stative-active languages rare in Eurasia? Typological perspective on split subject marking. In Mark Donohue and Søren Wichmann (eds). The Typology of Semantic Alignment Systems, 121-139. Oxford: Oxford University Press.
- Nichols, Johanna, David A. Peterson & Jonathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8: 149–211.
- Tsunoda, T. 1981. Split case-marking patterns in verb-types and tense / aspect / mood // Linguistics. Vol. 19. P. 389–438.
- Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21. 385–396.
- Van Valin, Robert D., Jr. 1999. Generalized Semantic Roles and the Syntax-Semantics Interface. In: Corblin, Dobrovie-Sorin, and Marandin (Eds.) Empirical Issues in Formal Syntax and Semantics 2. The Hague: Thesus. 373–389.