

Associations between constructions and collexemes: Is there one measure of attraction?



Natalia Levshina
natalevs@gmail.com



Which frequency measure(s) represent(s) the associations between construction C and lexeme L the best?

(relative) frequency of L in C

- proponents:** Goldberg et al. 2004; Bybee 2010
- evidence:** facilitation of C learning, grammaticality judgements

contingency-based bidirectional measure, e.g. FET log-p

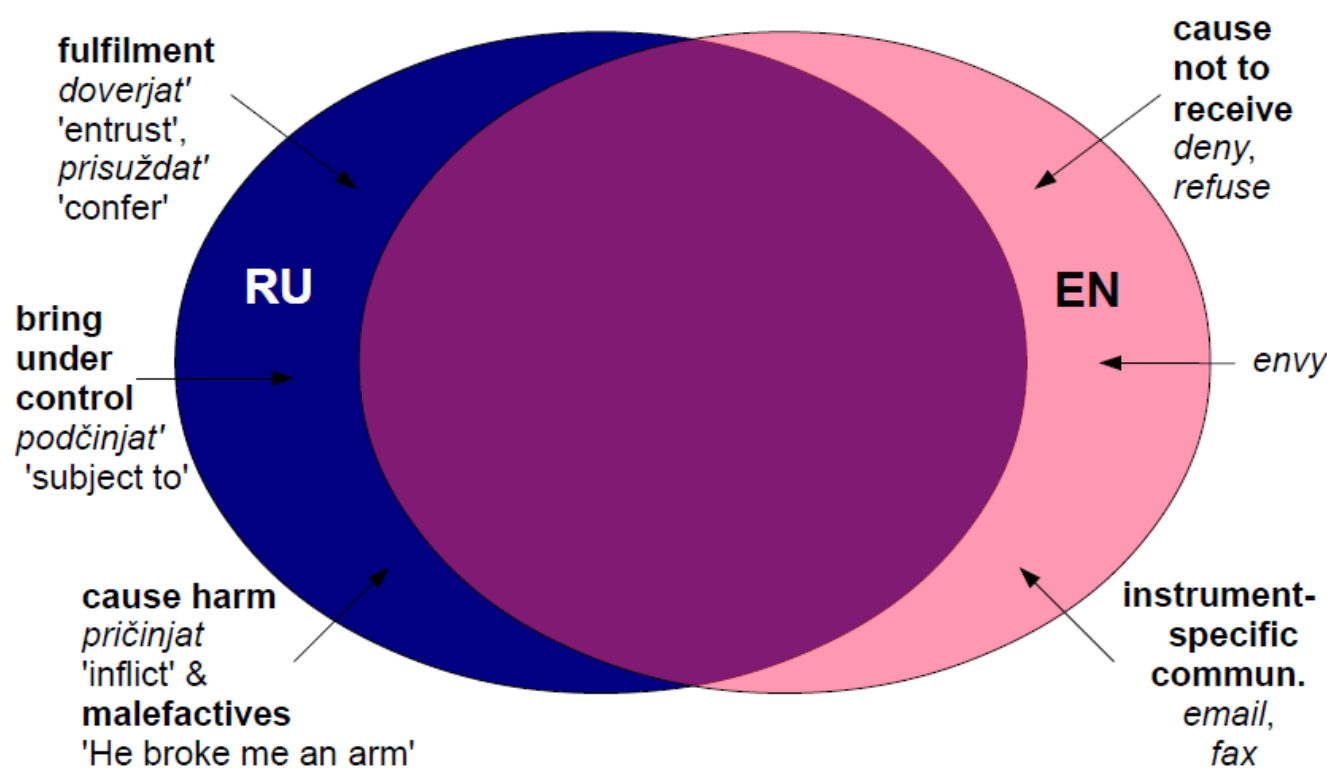
- proponents:** Gries et al. 2005, 2010; Ellis & Ferreira-Jr. 2009 (cf. also Wiechmann 2008)
- evidence:** sentence completion, self-paced comprehension task, ESLearners' uptake

frequency effects may differ

- proponents:** Divjak 2008 (to some extent); Schmid 2010 ('attraction' vs. 'reliance')
- evidence:** not much so far

A case study of the Russian ditransitive construction

SEMANTICS



CORPUS DATA

- Russian National Corpus: syntactically annotated part (appr. 760 000 words)
- $C = V + NP_{DAT} + NP_{ACC}$ (also $NP_{DAT} + NP_{GEN}$ with mass nouns), in any order
- 667 occurrences of the ditransitive construction (automated search and manual cleaning)
- 119 verbs (aspectual pairs, e.g. *dat'* – *davat'* 'give', are represented by the imperfective form)

CORPUS ASSOCIATION MEASURES

	C1	-C1
L1	a	b
-L1	c	d

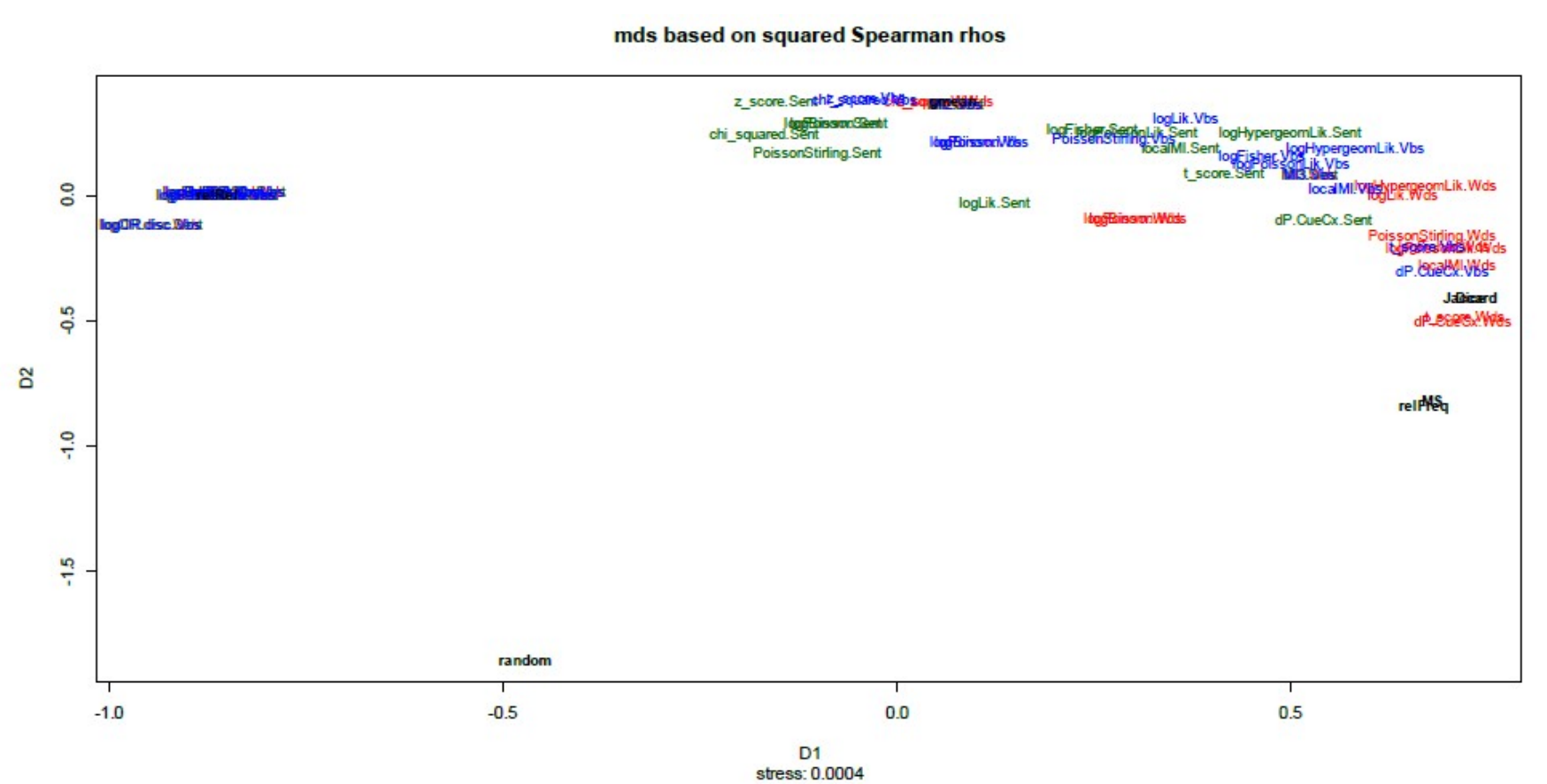
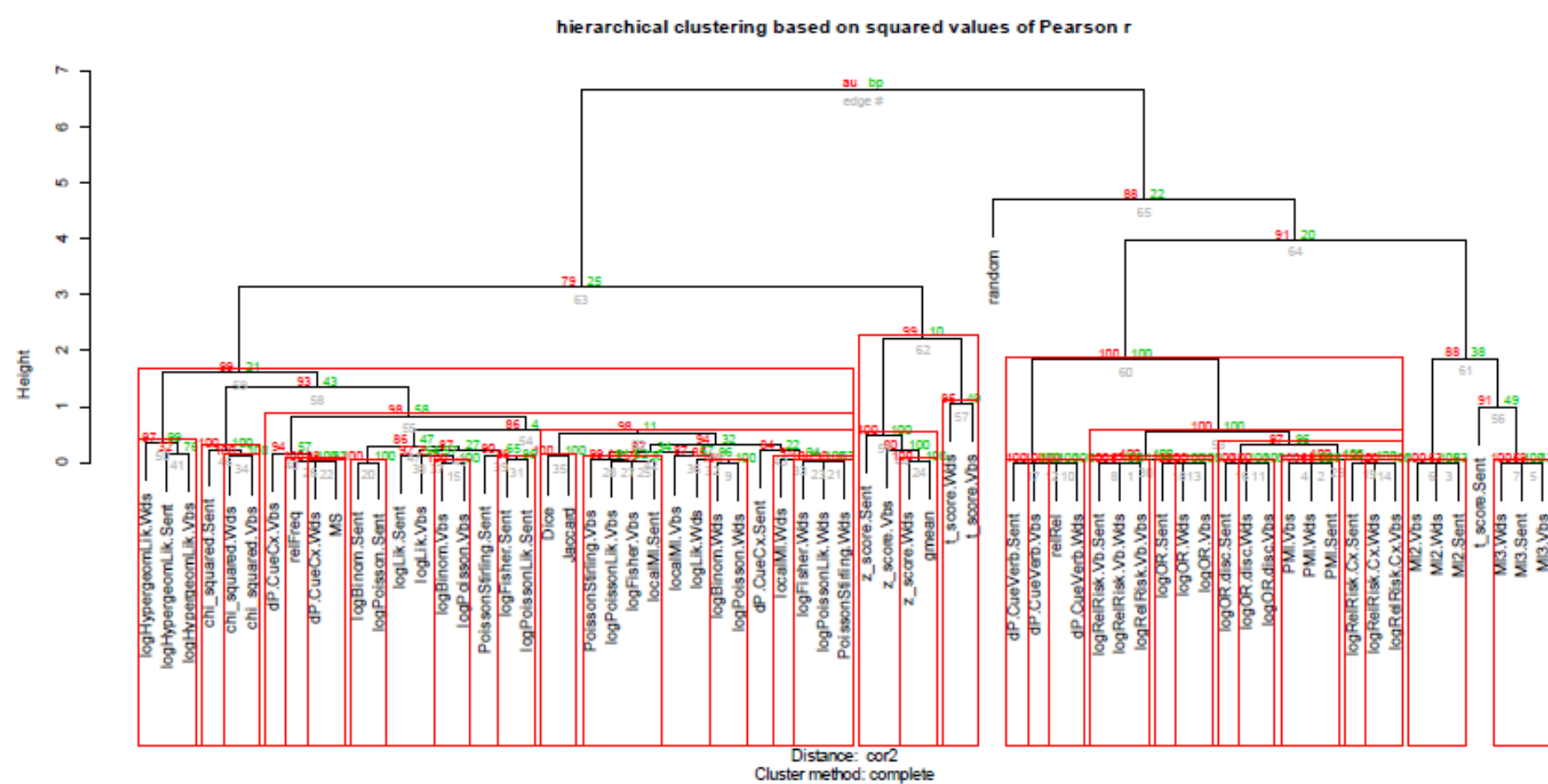
UNIDIRECTIONAL

- relative frequency
- reliance
- ΔP (construction or verb as cue)

BIDIRECTIONAL

- degree of association (e.g. log-odds)
- significance of association (e.g. FET log-p)
- information-theoretical (e.g. PMI)

RELATIONSHIPS BETWEEN CORPUS ASSOCIATION MEASURES

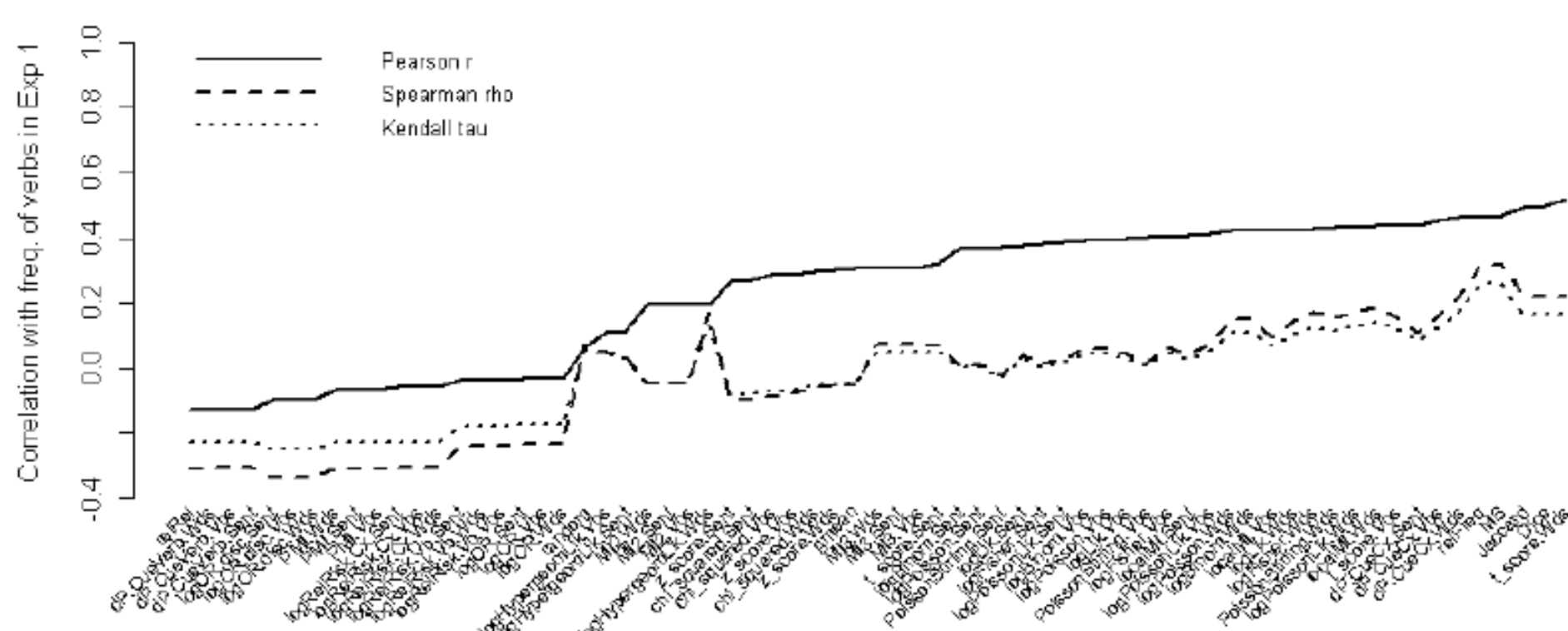


EXPERIMENT 1 (CONSTRUCTION > VERB)

- design:** constructions as stimuli, verbs as responses
- subjects:** 66 native speakers
- online form
- Stimuli:** 10x4 nonce word structures (16x4 fillers), e.g.

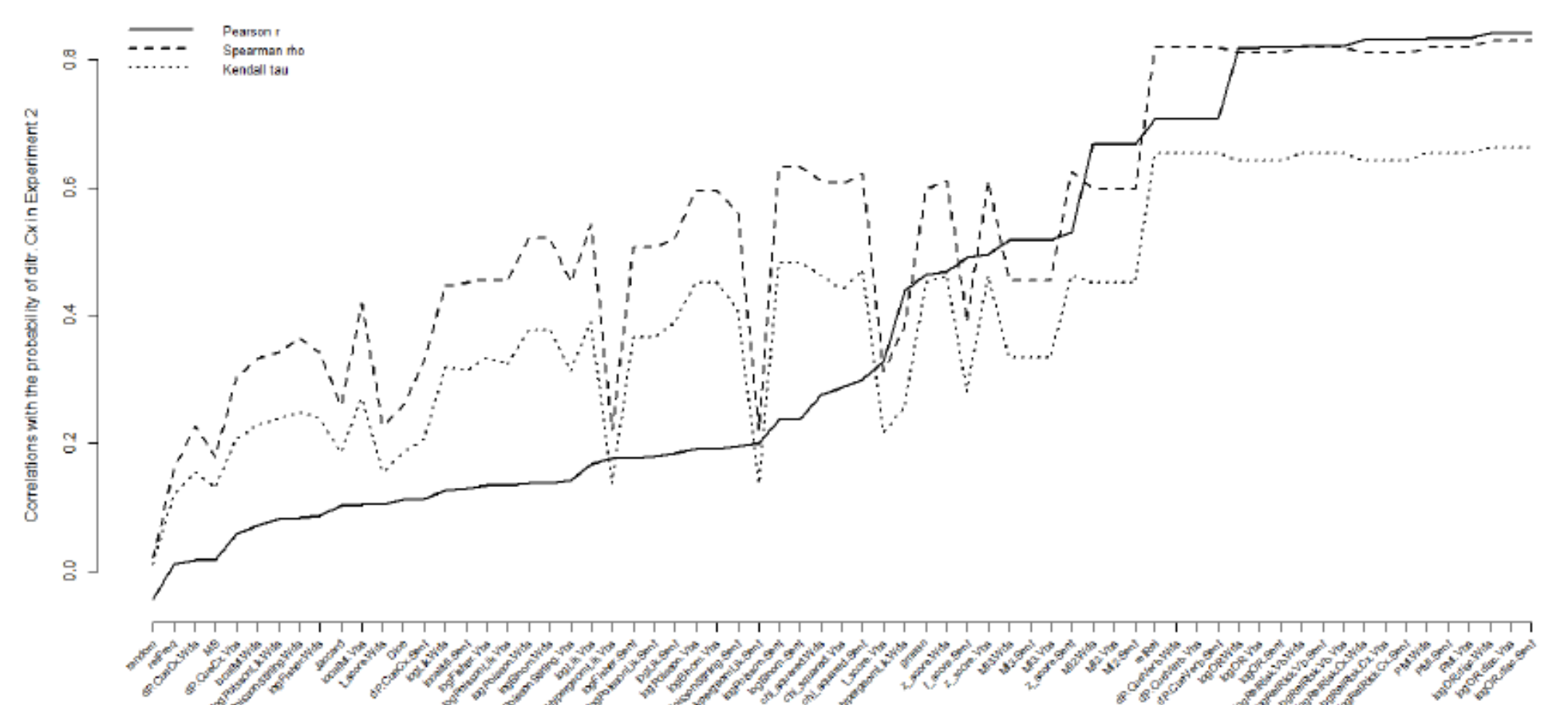
Секаб _____ (что делает?) его тому леваву.

- task:** write 3-5 verbs that fit the context
- measure:** overall frequency of each verb as response



EXPERIMENT 2 (VERB > CONSTRUCTION)

- design:** verbs as stimuli, constructions as responses
- subjects:** 41 students of Pskov State University
- stimuli:** 10x2 collexemes (16x2 fillers)
- task:** make a sentence with each verb
- measure:** proportion of sentences with DitrC in the responses



CONCLUSIONS

1 The corpus measures form two large stable clusters: one with the relative frequency of L in C and the other with reliance of L on C. Most corpus measures are very strongly correlated. Therefore, competitions between individual measures in prediction of experimental results (cf. previous research) should be taken with a grain of salt.

2 The measures from the relative frequency (attraction) cluster correlate positively with the experimental frequencies Construction > Verbs. The reliance cluster measures correlate positively with the experimental frequencies Verb > Construction. These facts can be interpreted as evidence of the bidirectionality of C-L associations.

3 The contingency-based measures (e.g. FET log-p) do not have any additional predictive power. The simple attraction and reliance seem to be the optimal measures of C-L association.