Loose and tight languages

A typology based on associations between constructions and lexemes

NATALIA LEVSHINA[1] and JOHN A. HAWKINS[2,3]

[1]Max Planck Institute for Psycholinguistics,

[2]University of California Davis, [3]Cambridge University

DOI: 10.5281/zenodo.5215108

A synopsis

- The processing of thematic roles depends on case marking, word order, but also crucially on the semantics of individual lexemes.
- This purely lexical-semantic processing could not be systematically studied until now cross-linguistically.
- With the help of large syntactically annotated corpora, we can now measure and compare the strength of filler-slot associations in different languages.
- The correlations between the strength of these associations and diverse morphosyntactic strategies in languages reveal a remarkable gradient typology.

DOI: 10.5281/zenodo.5215108

LOOSE

Basic grammatical relations have wide semantic range

TIGHT

Basic grammatical relations have narrow semantic range

DOI: 10.5281/zenodo.5215108

Hawkins 1986:121–127, 1995, 2019; Müller-Gotama 1994

Loose English vs. tight German

English has fewer semantic restrictions on the subject than German (e.g. locative, temporal, instrumental and other subjects)

- **1979** witnessed twenty big firms go bankrupt.
- ?1979 sah 20 grosse Firmen pleite gehen.

- **The roof** was leaking water.
- *Das Dach tropfte Wasser.

- **His second goal** ended the match.
- *Sein zweites Tor endete das Spiel.

Plank 1984

DOI: 10.5281/zenodo.5215108

LOOSE

TIGHT

English

Chinese,
Indonesian

German, Japanese,
Korean, Malayalam,
Russian, Turkish

DOI: 10.5281/zenodo.5215108

Hawkins 1986:121– 127, 1995, 2019; Müller-Gotama 1994

Other properties of tight languages

- more explicit grammatical coding, e.g.:
    - formal case marking
    - less optionality in the use of complementizers and relativizers

- verb-final languages are regularly tight rather than loose

- avoidance of raisings and long-distance WH-movements

- fewer cases of category ambiguity
    - e.g. German *Buch*$_{noun}$ – *buchen*$_{verb}$ *vs.* English *book*$_{noun}$ – *book*$_{verb}$

- a narrower set of subcategorization frames for verbs
    - e.g. German *öffnen* – *sich öffnen vs.* English *open*$_{tr}$ – *open*$_{\mathrm{intr}}$

DOI: 10.5281/zenodo.5215108

Hawkins 1986, 1995, 2019

A constructionist corpus-based perspective

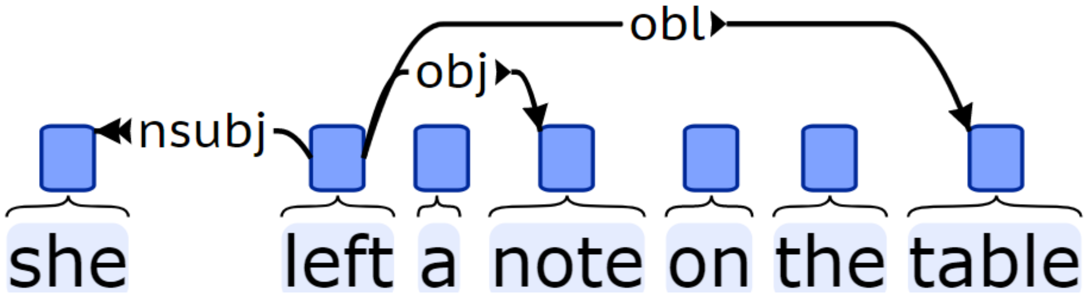
- Tight languages have tight associations between constructional slots and lexical fillers, while loose languages have loose associations.

- We can measure these associations using corpora by computing Mutual Information of lexemes and constructional slots:

$$I\ (Lex;\ Dep) = \sum_{i,j} p\left(lex_i, dep_j\right) log\ \frac{p\left(lex_i, dep_j\right)}{p\left(lex_i\right) p\ \left(dep_j\right)}$$

DOI: 10.5281/zenodo.5215108

Universal Dependencies



DOI: 10.5281/zenodo.5215108

Zeman et al. 2020

Fragment of a matrix

Lexeme	Intrans. subject	Trans. subject	Object	Oblique/ IO
hunter/NOUN	64	40	22	30
evening/NOUN	100	38	150	1145
street/NOUN	155	34	466	1331
t-shirt/NOUN	7	3	118	36

DOI: 10.5281/zenodo.5215108

Fragment of a matrix

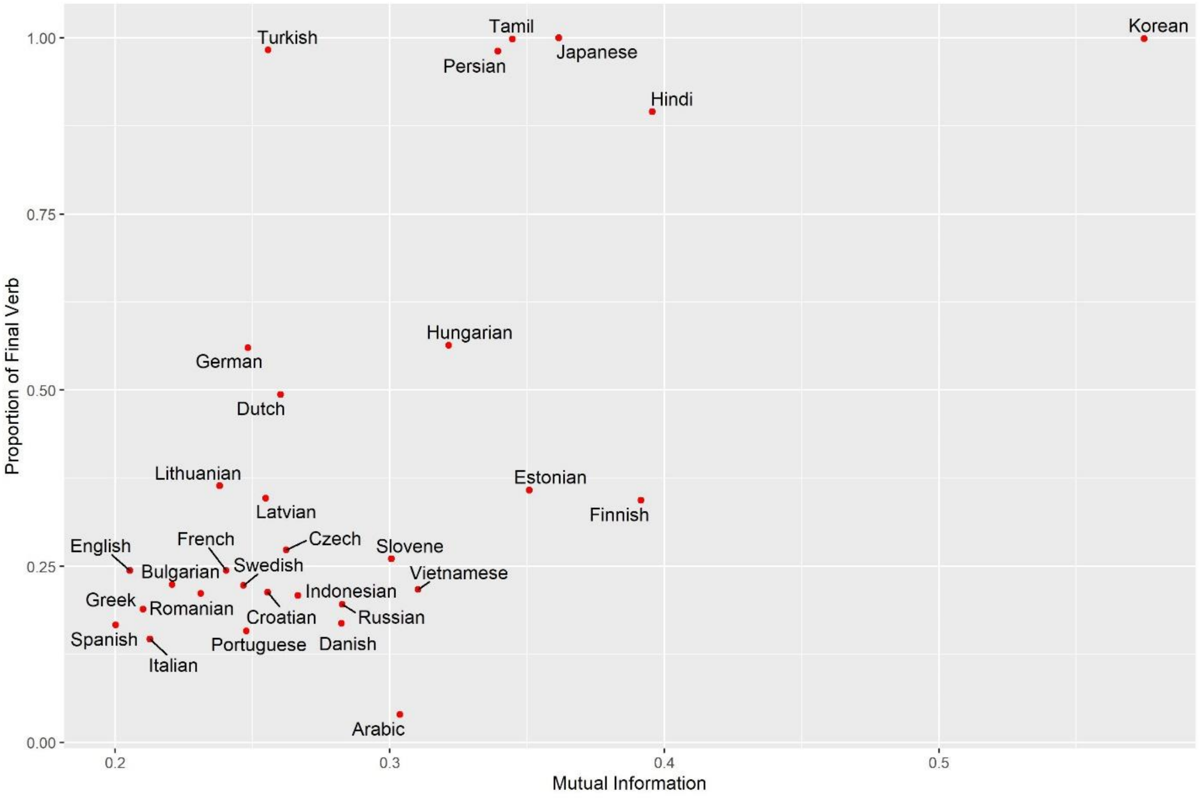
Lexeme	Intrans. subject	Trans. subject	Object	Oblique/ IO
hunter/NOUN	[illegible]	[illegible]	[illegible]	30
evening/NOUN	[illegible]	[illegible]	[illegible]	1145
street/NOUN	[illegible]	[illegible]	[illegible]	1331
t-shirt/NOUN	7	3	118	36

The stronger the bias,
the tighter the language

DOI: 10.5281/zenodo.5215108

Mutual Information and verb-finalness

LOOSE



TIGHT

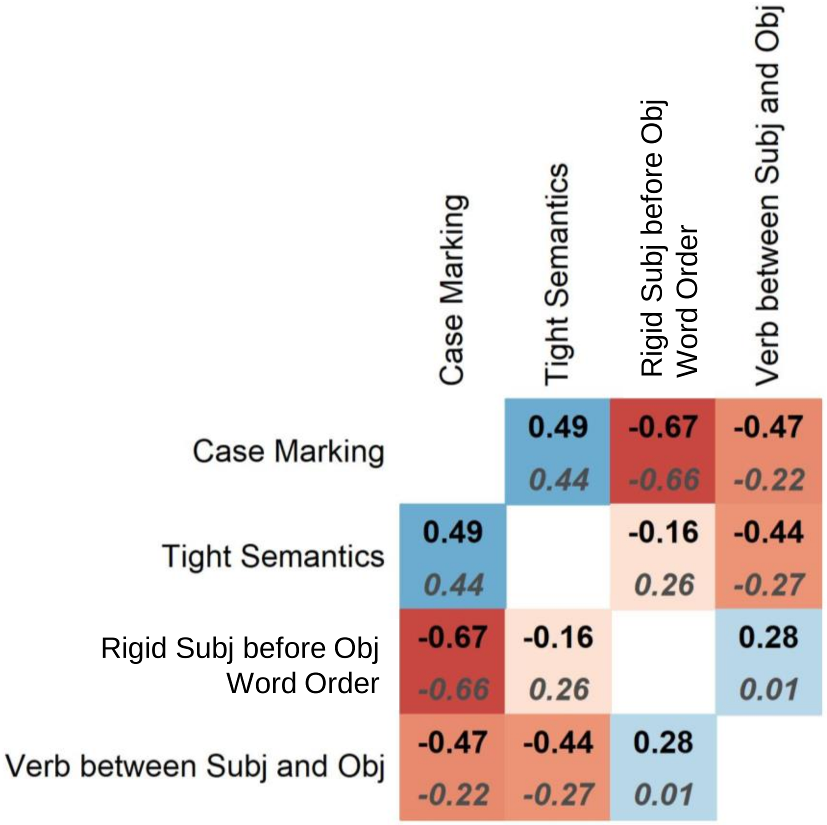
The positive correlation is supported by a Bayesian GLMM.
Bayesian R^2 = 0.85, 95%CI 0.66, 0.93.

DOI: 10.5281/zenodo.5215108

Levshina 2020 *TLT*

Tight semantics and other parameters of tr. Subject & Object

Correlations

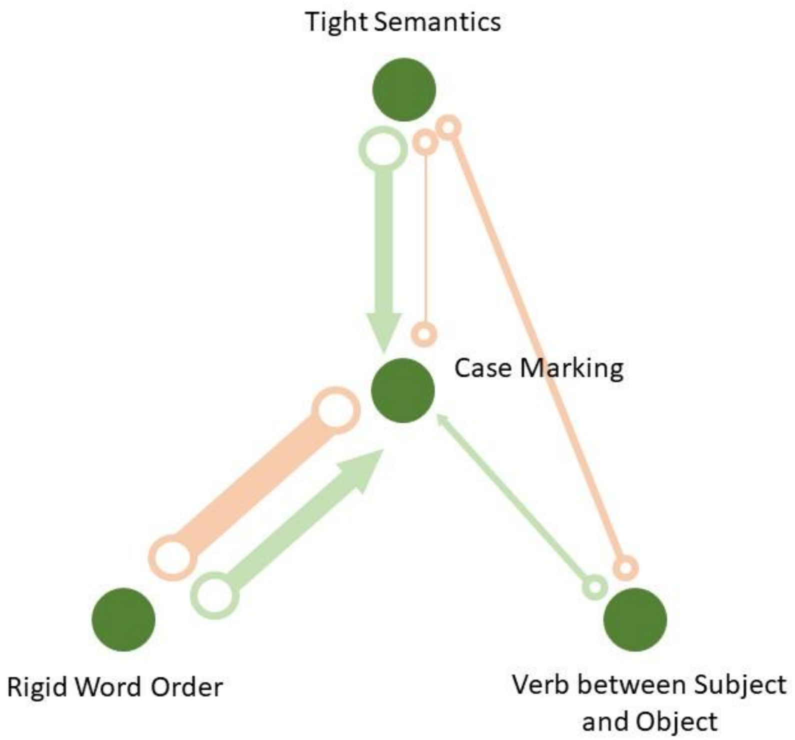


	Case Marking	Tight Semantics	Rigid Subj before Obj Word Order	Verb between Subj and Obj
Case Marking		0.49 *0.44*	-0.67 *-0.66*	-0.47 *-0.22*
Tight Semantics	0.49 *0.44*		-0.16 *0.26*	-0.44 *-0.27*
Rigid Subj before Obj Word Order	-0.67 *-0.66*	-0.16 *0.26*		0.28 *0.01*
Verb between Subj and Obj	-0.47 *-0.22*	-0.44 *-0.27*	0.28 *0.01*	

upper, black: absolute correlations

lower, grey: partial correlations

Causal network



Levshina 2021 *Front Psych*

DOI: 10.5281/zenodo.5215108

But what about verbs and subcategorization frames?

DOI: 10.5281/zenodo.5215108

P-lability

Alternation	Object of Transitive	Subject of Intransitive
Causative/inchoative alternation	Ann broke the cup.	The cup broke.
Middle alternation	The publisher sells the book.	The book sells well.
Induced action alternation	Sue jumped the horse over the fence.	The horse jumped over the fence.

DOI: 10.5281/zenodo.5215108

Levin 1993; Dixon 1994

A-lability

Alternation	Expressed object	Unexpressed object
Unspecified Object alternation	Jack ate the cake.	Jack ate.
Understood Body-Part alternation	The queen waved her hand at the crowd.	The queen waved at the crowd.
Characteristic Property alternation	Their dog bites people.	Their dog bites.

DOI: 10.5281/zenodo.5215108

Levin 1993; Dixon 1994

Outline

1. Loose and tight languages

2. A quantitative corpus-based study:
   - Corpora and annotation
   - Lability measures
   - Additional measures
   - Correlations

3. Discussion

DOI: 10.5281/zenodo.5215108

Corpora and annotation

- Leipzig Corpora Collection (Goldhahn et al. 2012)
  http://wortschatz.uni-leipzig.de/en/download/

- 30 online news corpora, 1M sentences in each:
  - Arabic, Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek (modern), Hindi, Hungarian, Indonesian, Italian, Japanese, Korean, Latvian, Lithuanian, Persian, Portuguese, Romanian, Russian, Slovenian, Spanish, Swedish, Tamil, Turkish, Vietnamese

- Annotated with the Universal Dependencies pipeline udpipe (Wijffels, Straka & Straková 2018).

DOI: 10.5281/zenodo.5215108

Outline

1. Loose and tight languages

2. A quantitative corpus-based study:
   - Corpora and annotation
   - Lability measures
   - Additional measures
   - Correlations and causal network

3. Discussion

DOI: 10.5281/zenodo.5215108

P-lability in corpora

- Compute the frequencies of all verb lemmas (only predicates of main clauses) with the same noun as ‘obj’ and intr. ‘nsubj’

- Particle verbs and verbs with separable prefixes are treated as one unit (e.g. break + out, um+leiten)

Verb	Noun	Subject - Verb	Verb - Object
have	opportunity	0	375
die	people	64	0
open	door	36	149
begin	work	35	33

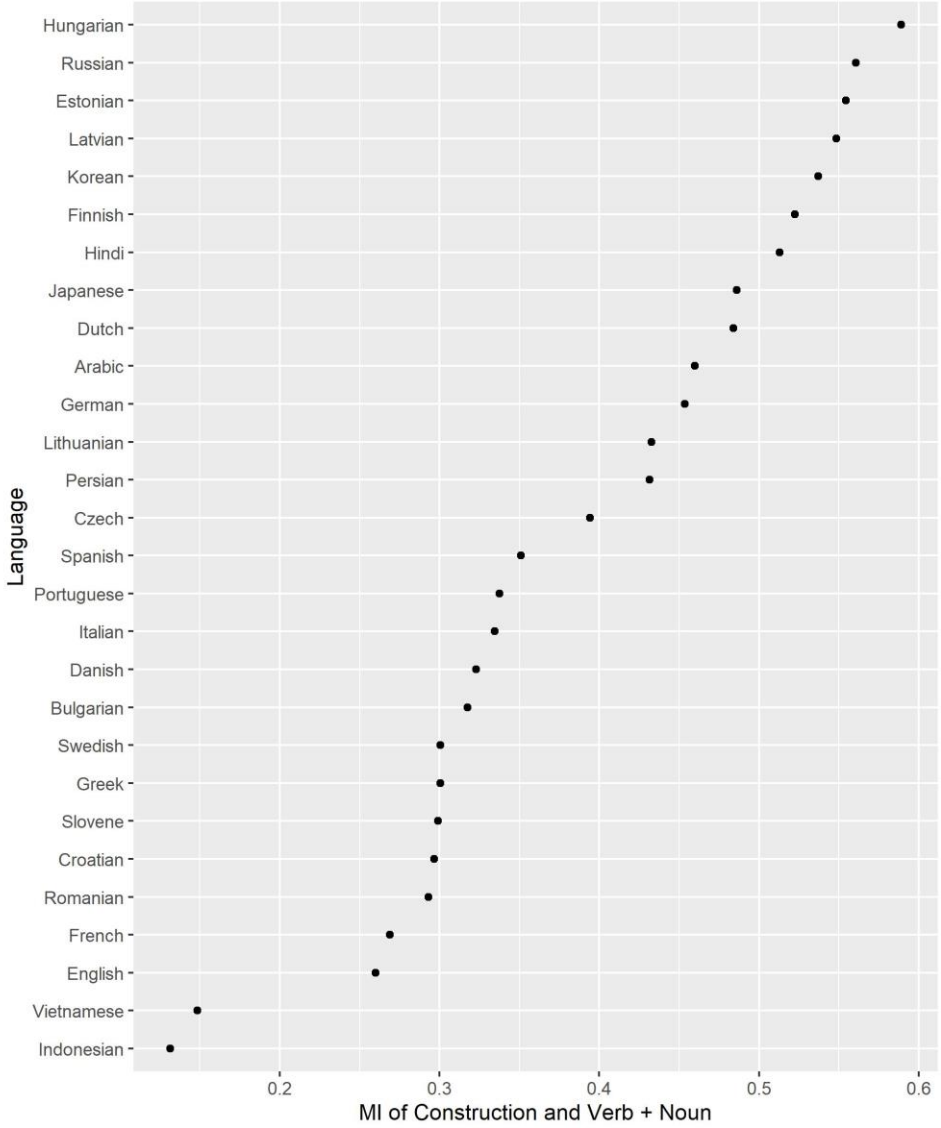
DOI: 10.5281/zenodo.5215108

Excluded cases

- Verbs with reflexive, passive, antipassive, middle morphology/ auxiliaries
    - Motivation: cross-linguistic differences in semantics and annotation
    - Consequence: we are primarily measuring looseness vs. non-looseness (the formal marking of which can be quite variable)
- Ditransitive clauses
- Data from Tamil and Turkish (strange issues with verb lemmas)

DOI: 10.5281/zenodo.5215108

P-lability MI scores



DOI: 10.5281/zenodo.5215108

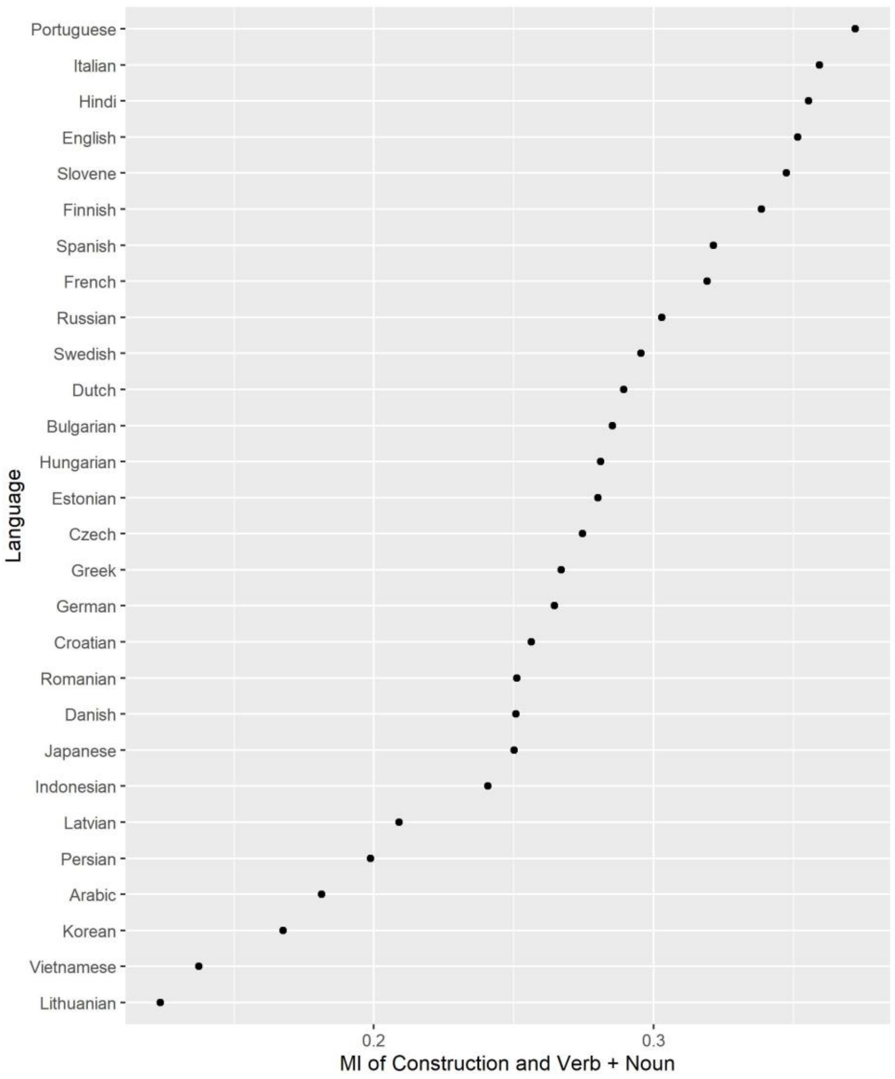
A-lability in corpora

- Find all verb lemmas with the same noun as ‘nsubj’ with and without ‘obj’ (nominal or pronominal object).

Verb	Noun (subject)	Transitive	Intransitive
be	idea	0	140
learn	student	21	35
play	team	55	47

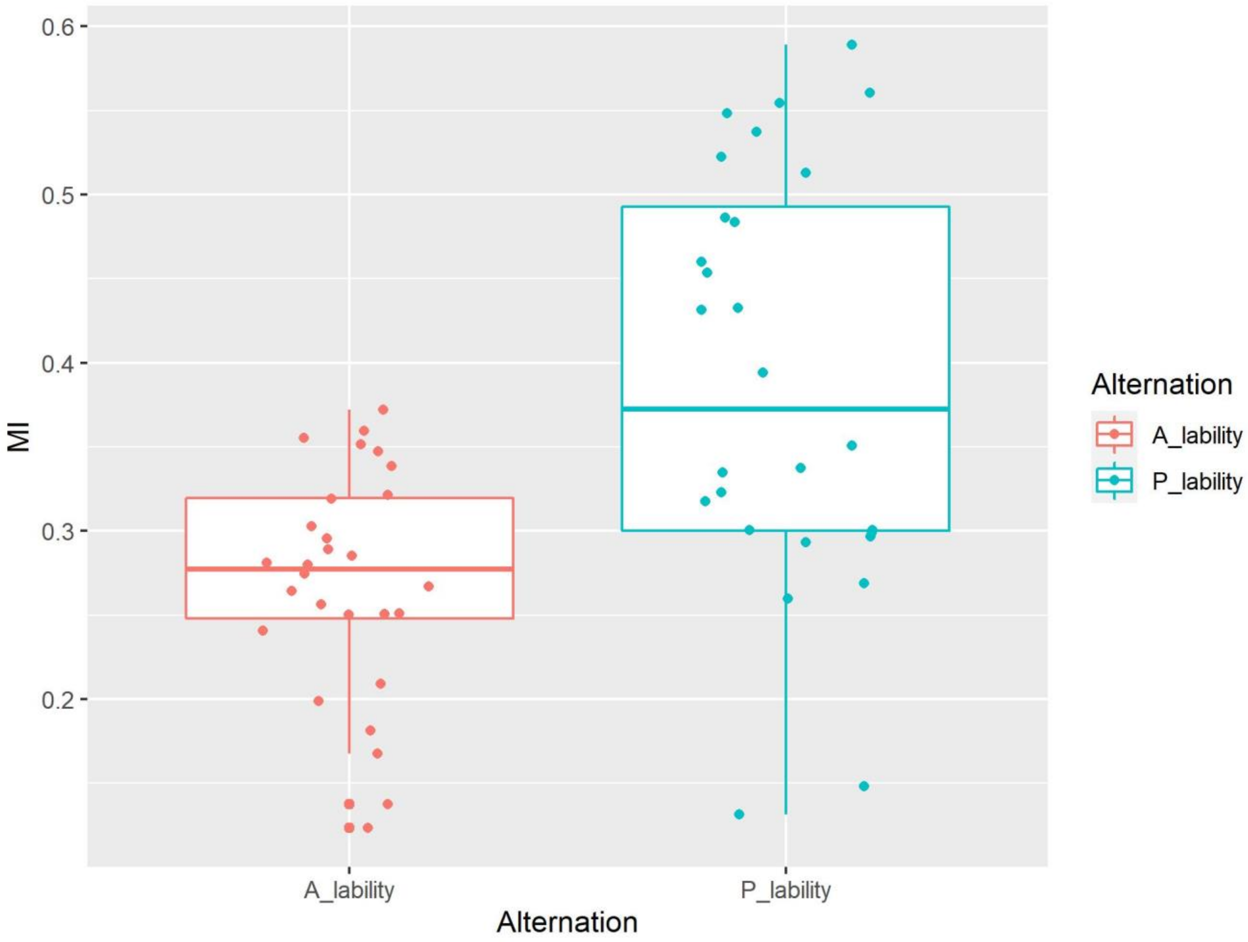
DOI: 10.5281/zenodo.5215108

A-lability MI scores



DOI: 10.5281/zenodo.5215108

Distributions of MI scores



DOI: 10.5281/zenodo.5215108

Outline

1. Loose and tight languages

2. A quantitative corpus-based study:
    - Corpora and annotation
    - Lability measures
    - Additional measures
    - Correlations

3. Discussion

DOI: 10.5281/zenodo.5215108

Additional measures

- Rigidity of Subject and Object order
- Proportion of lexical verb in the middle, between Subject and Object
- Case marking: MI of cases and roles (Subject and Objects)

Note: we examine only transitive subjects!

DOI: 10.5281/zenodo.5215108

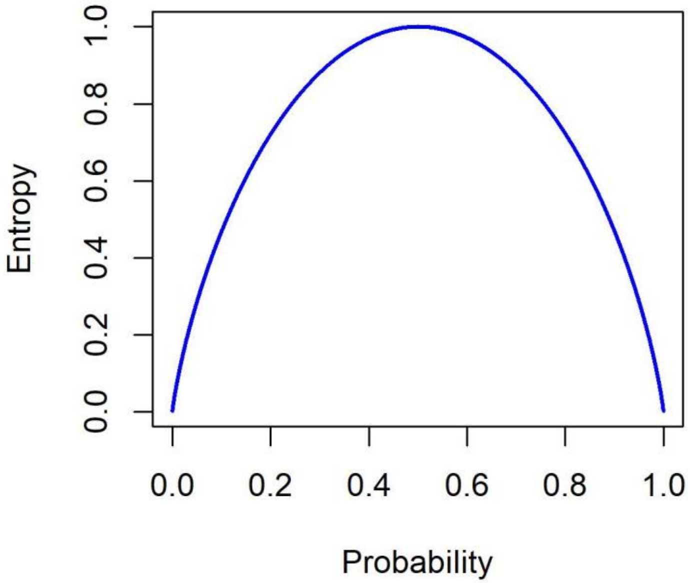
Subject – Object order: Shannon entropy

- Proportions of nsubj + obj and obj + nsubj (only common nouns) in a transitive clause

- The higher H, the greater the variability

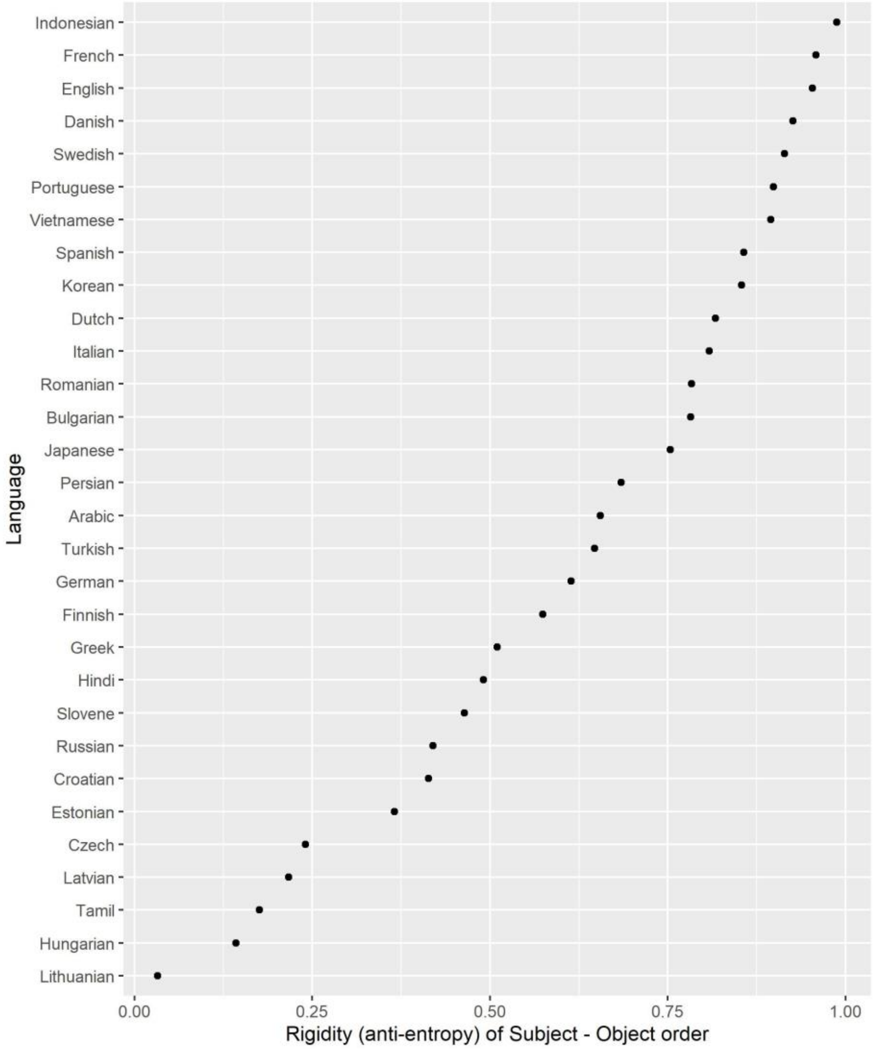
$$H(X) = -\sum_{i=1}^{2} P(x_i)\, log_2\, P(x_i)$$

- We use rigidity, which is 1 minus entropy.



DOI: 10.5281/zenodo.5215108

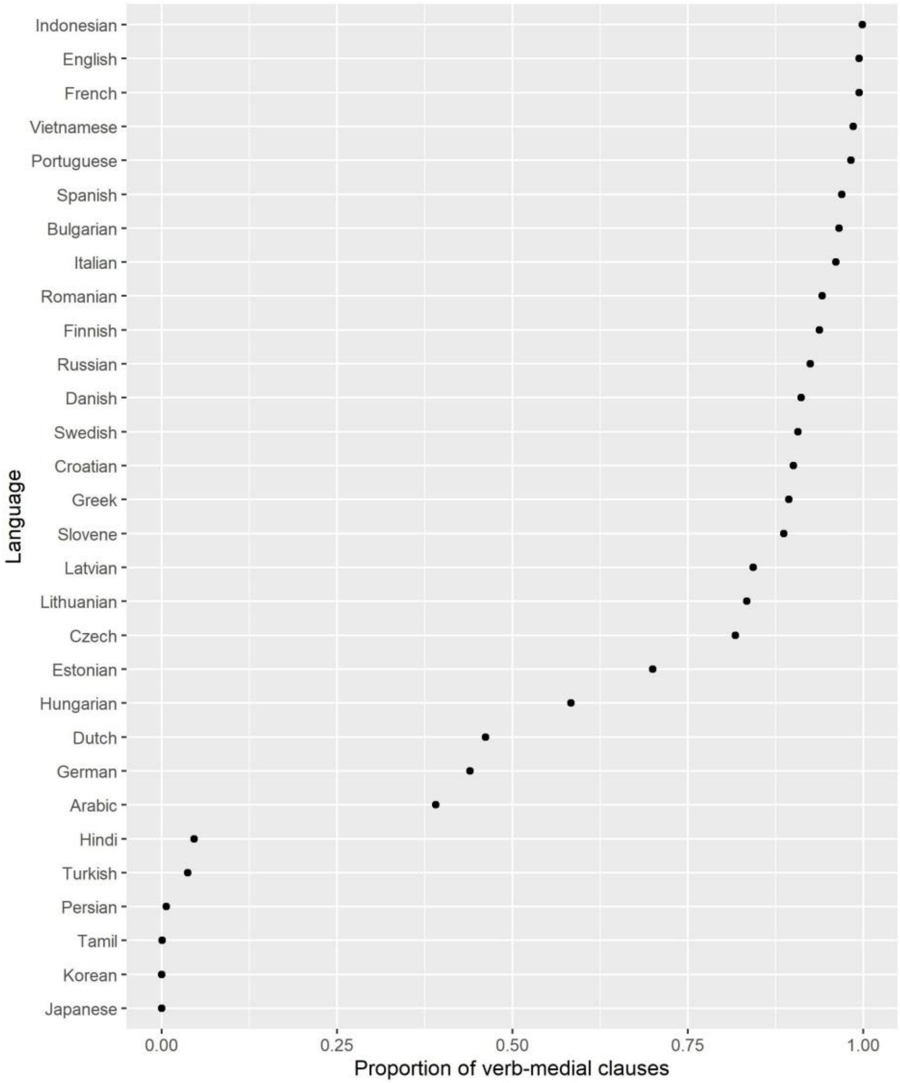
Rigidity of Subject – Object order



DOI: 10.5281/zenodo.5215108

Levshina 2021 *Front Psych*

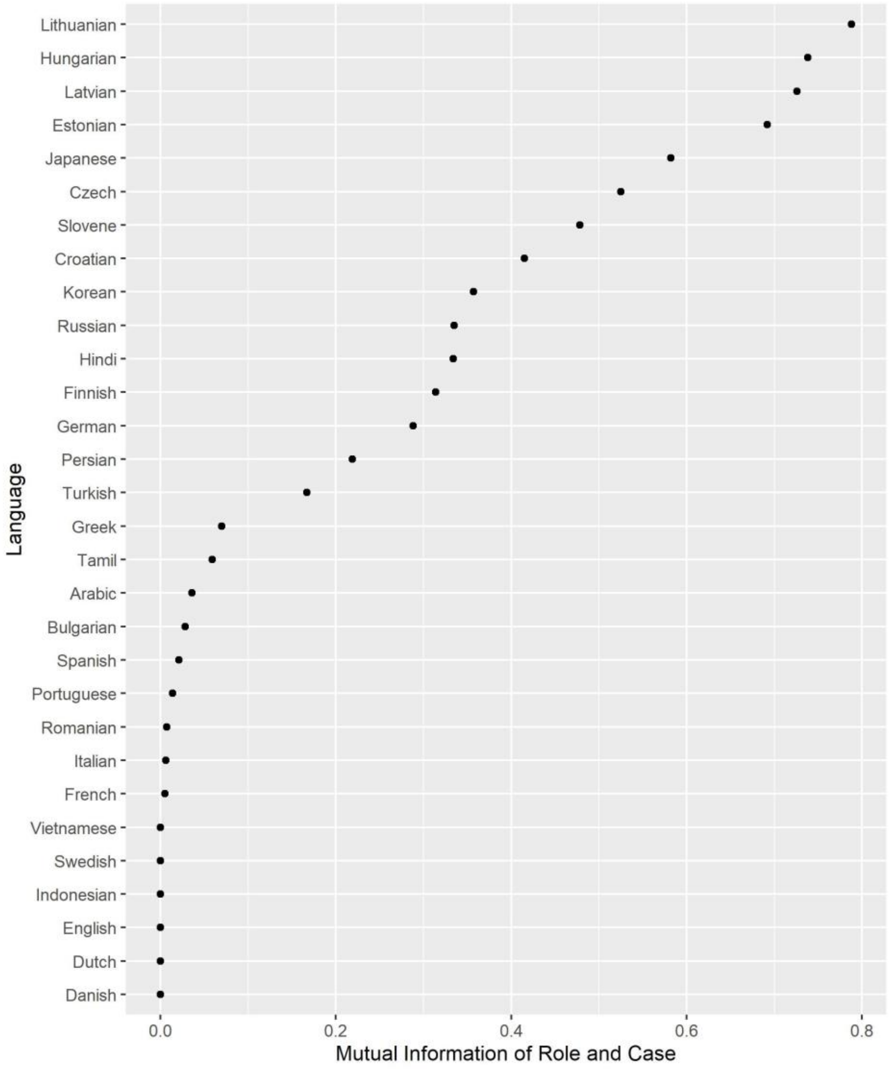
Proportions of Verb between Subject and Object



DOI: 10.5281/zenodo.5215108

Levshina 2021 *Front Psych*

Case marking: MI of case and Subject/Object roles



DOI: 10.5281/zenodo.5215108

Levshina 2021 *Front Psych*

Outline

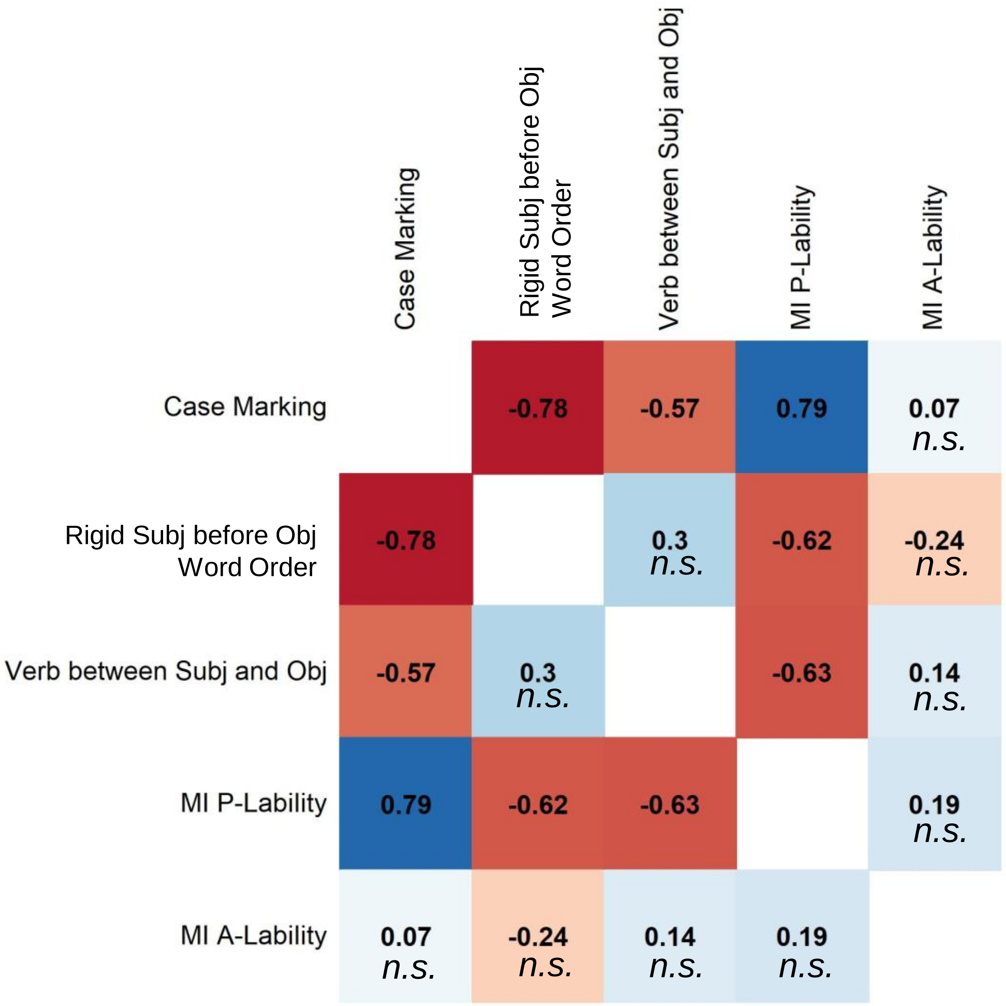
1. Loose and tight languages

2. A quantitative corpus-based study:
   - Corpora and annotation
   - Lability measures
   - Additional measures
   - Correlations

3. Discussion

DOI: 10.5281/zenodo.5215108

Spearman’s correlations



	Case Marking	Rigid Subj before Obj Word Order	Verb between Subj and Obj	MI P-Lability	MI A-Lability
Case Marking		-0.78	-0.57	0.79	0.07 *n.s.*
Rigid Subj before Obj Word Order	-0.78		0.3 *n.s.*	-0.62	-0.24 *n.s.*
Verb between Subj and Obj	-0.57	0.3 *n.s.*		-0.63	0.14 *n.s.*
MI P-Lability	0.79	-0.62	-0.63		0.19 *n.s.*
MI A-Lability	0.07 *n.s.*	-0.24 *n.s.*	0.14 *n.s.*	0.19 *n.s.*	

Note:
Genealogical relationships were controlled by sampling 1 language per genus (1000 samples).

DOI: 10.5281/zenodo.5215108

Outline

1. Loose and tight languages

2. A quantitative corpus-based study:
   - Corpora and annotation
   - Lability measures
   - Additional measures
   - Correlations

3. Discussion

DOI: 10.5281/zenodo.5215108

Conclusions: P-lability

- P-lability scores are strongly correlated with other properties of loose and tight languages:

LOOSE **TIGHT**

High P-lability (low MI-scores)	Low P-lability (high MI-scores)
No case marking	Rich case marking
Rigid order of Subject & Object	Variable order of Subject & Object
Verb-medial order	Verb-final order

DOI: 10.5281/zenodo.5215108

M A X
P L A
N C K

Conclusions: P-lability

- P-lability scores are strongly correlated with other properties of loose and tight languages:

LOOSE **High P-lability (low MI-scores)**	**TIGHT** **Low P-lability (high MI-scores)**
No case marking	Rich case marking
Rigid order of Subject & Object	Variable order of Subject & Object
Verb-medial order	Verb-final order

Strong associations between constructional slots and lexemes help in early and more reliable identification of thematic roles, alongside case marking.

DOI: 10.5281/zenodo.5215108

Conclusions: A-lability

- A-lability scores are not correlated with any of those properties. It is also more frequently found than P-lability.

- A possible explanation is that A-lability is driven by diverse communicative and cultural factors:
    - high accessibility, e.g. *Italy wins [the final]!*
    - conventionalized inferences, *e.g. He drinks again [liquor].*
    - Focus on action with low discourse prominence of object, e.g. *She chopped and chopped [e.g. meat].*
    - taboo, e.g. *Pat sneezed [mucus] onto the computer screen*.
    - tact, e.g. *I contributed [$1,000] to UNICEF*.

DOI: 10.5281/zenodo.5215108

Fillmore 1986; Goldberg 2005; see also Levshina 2018

References

- Dixon, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Fillmore, Ch.J. 1986. Pragmatically Controlled Zero Anaphora. In: *Proceedings of the Berkeley Linguistics Society* 12: 95–107.
- Goldberg, A.E. 2005. Argument realization: the role of constructions, lexical semantics and discourse factors. In: J.-O. Östman & M. Fried (eds.), *Construction Grammars: Cognitive grounding and theoretical extensions*, 17–44. Amsterdam: John Benjamins.
- Goldhahn, D., Th. Eckart & U. Quasthoff. 2012. Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In: N. Calzolari, Kh. Choukri, Th. Declerck et al. (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, 759-765. Istanbul: ELRA. URL http://www.lrec-conf.org/proceedings/lrec2012/pdf/327_Paper.pdf
- Hawkins, J.A. 1986. *A Comparative Typology of English and German: Unifying the Contrasts*. London: Croom-Helm.
- Hawkins, J.A. 1995. Argument-predicate structure in grammar and performance: A comparison of English and German. In: I. Rauch & G.F. Carr (eds.), *Insights in Germanic Linguistics*, *Vol. 1 Methodology in Transition*, 127–44. Berlin: Mouton de Gruyter.
- Hawkins, J.A. 2019. Word-external properties in a typology of Modern English: A comparison with German. *English Language and Linguistics* 23(3): 701–723.
- Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Levshina, N. 2018. *Towards a Theory of Communicative Efficiency in Human Languages*. Habilitation thesis, Leipzig University. DOI 10.5281/zenodo.1542857.
- Levshina, N. 2020. How tight is your language? A semantic typology based on Mutual Information. In: *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, 70–78. Düsseldorf: ACL. URL https://www.aclweb.org/anthology/2020.tlt-1.7.pdf
- Levshina, N. 2021. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. Frontiers in Psychology 12: 648200. DOI 10.3389/fpsyg.2021.648200.
- Müller-Gotama, F. 1994. *Grammatical Relations: A Cross-Linguistic Perspective on Their Syntax and Semantics.* Berlin: Mouton de Gruyter.
- Plank, F. 1984. Verbs and objects in semantic agreement: Minor differences between English and German might that might suggest a major one. *Journal of Semantics* 3(4): 305–360.
- Wijffels, J., M. Straka & J. Straková. 2018. udpipe: Tokenization, parts of speech tagging, lemmatization and dependency parsing with the UDPipe NLP Toolkit. R package version 0.7. URL https://CRAN.R-project.org/package=udpipe

DOI: 10.5281/zenodo.5215108

MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS

WWW.MPI.NL

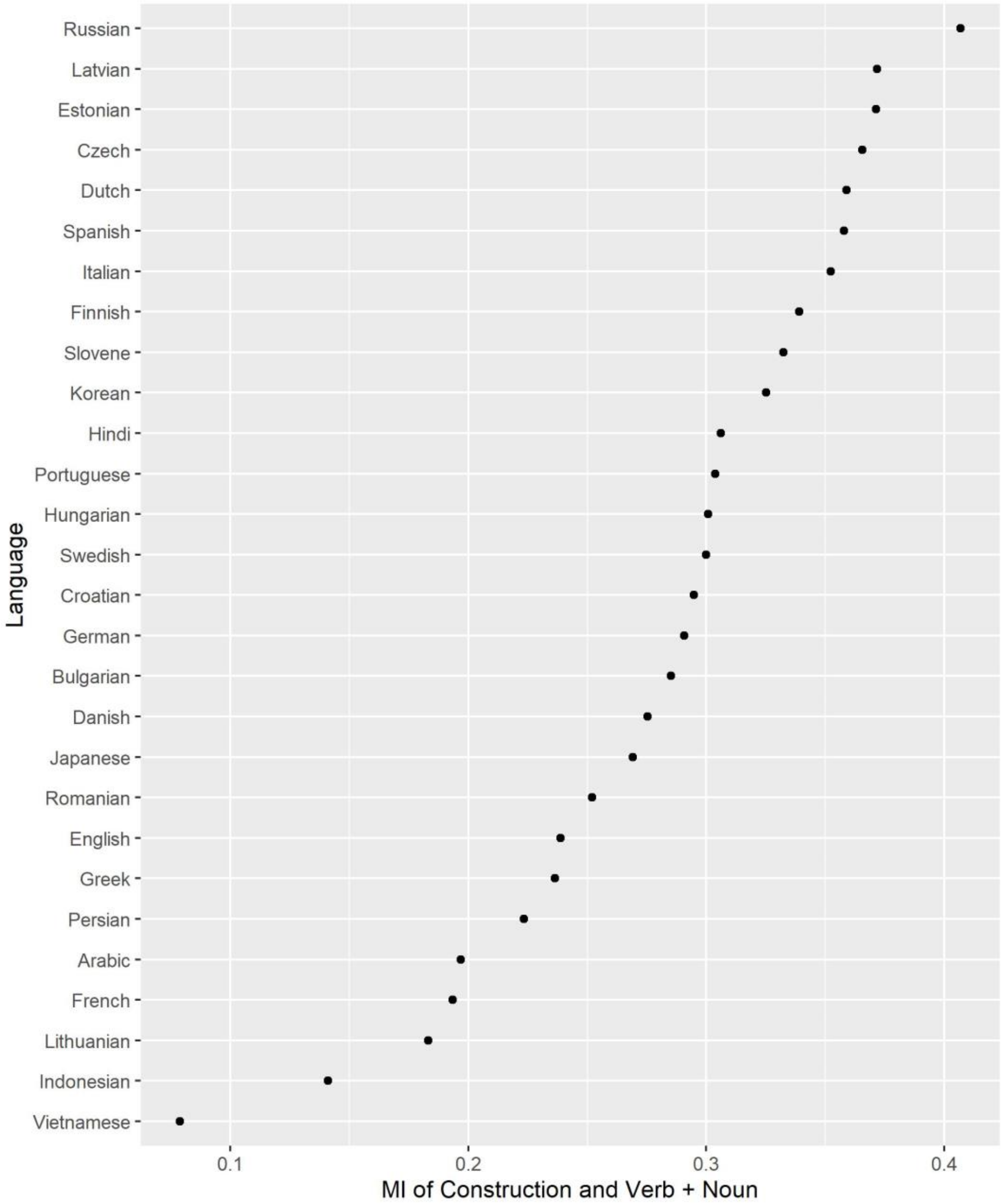
Additional stuff

- We also computed A-lability scores including not only nominal objects, but also objects expressed by ccomp (finite complement clauses) or xcomp (non-finite complement clauses).

- The slides that follow show the scores and correlations with this (broader) operationalization of A-lability. The P-lability scores are the same as above.

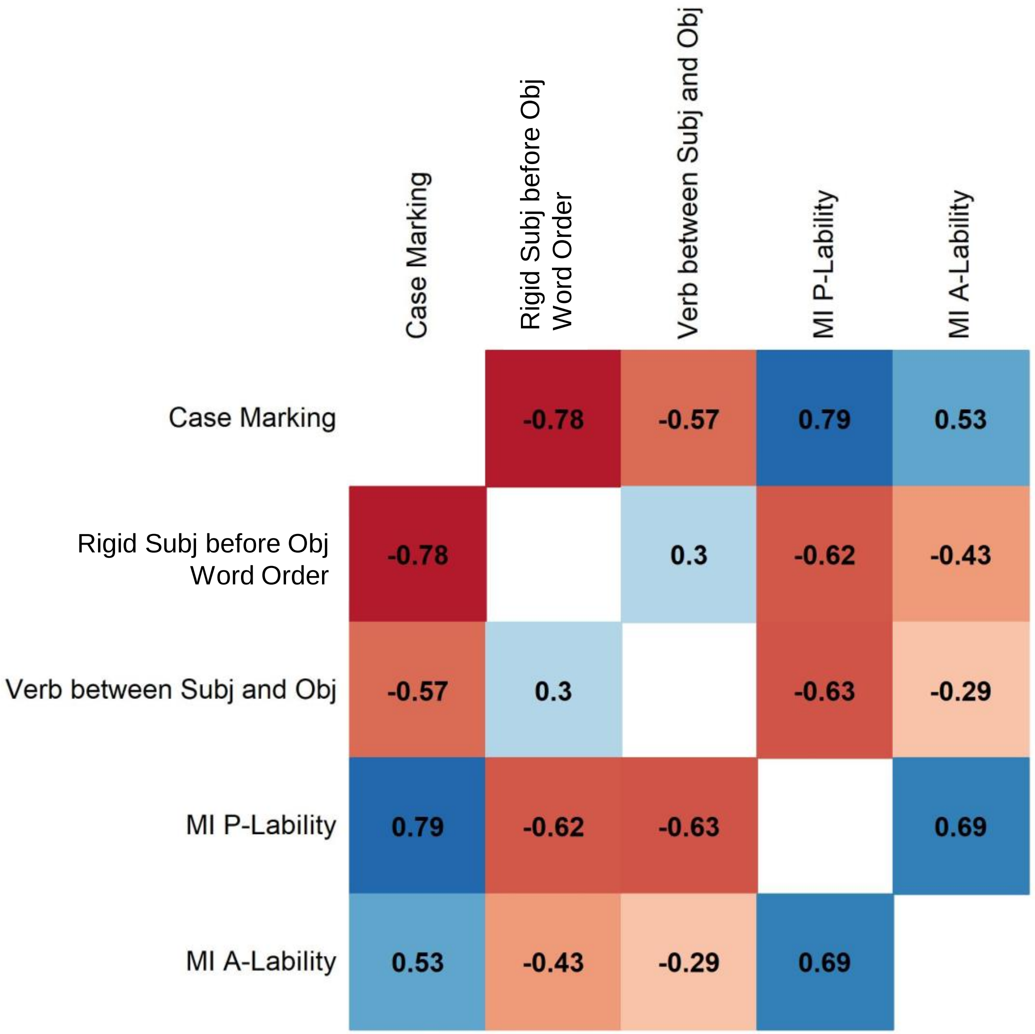
- The A-lability scores including clausal objects might overlap partly with raising (e.g. the verb *happen* is then sometimes intransitive, but also sometimes transitive, due to non-finite complements).

DOI: 10.5281/zenodo.5215108



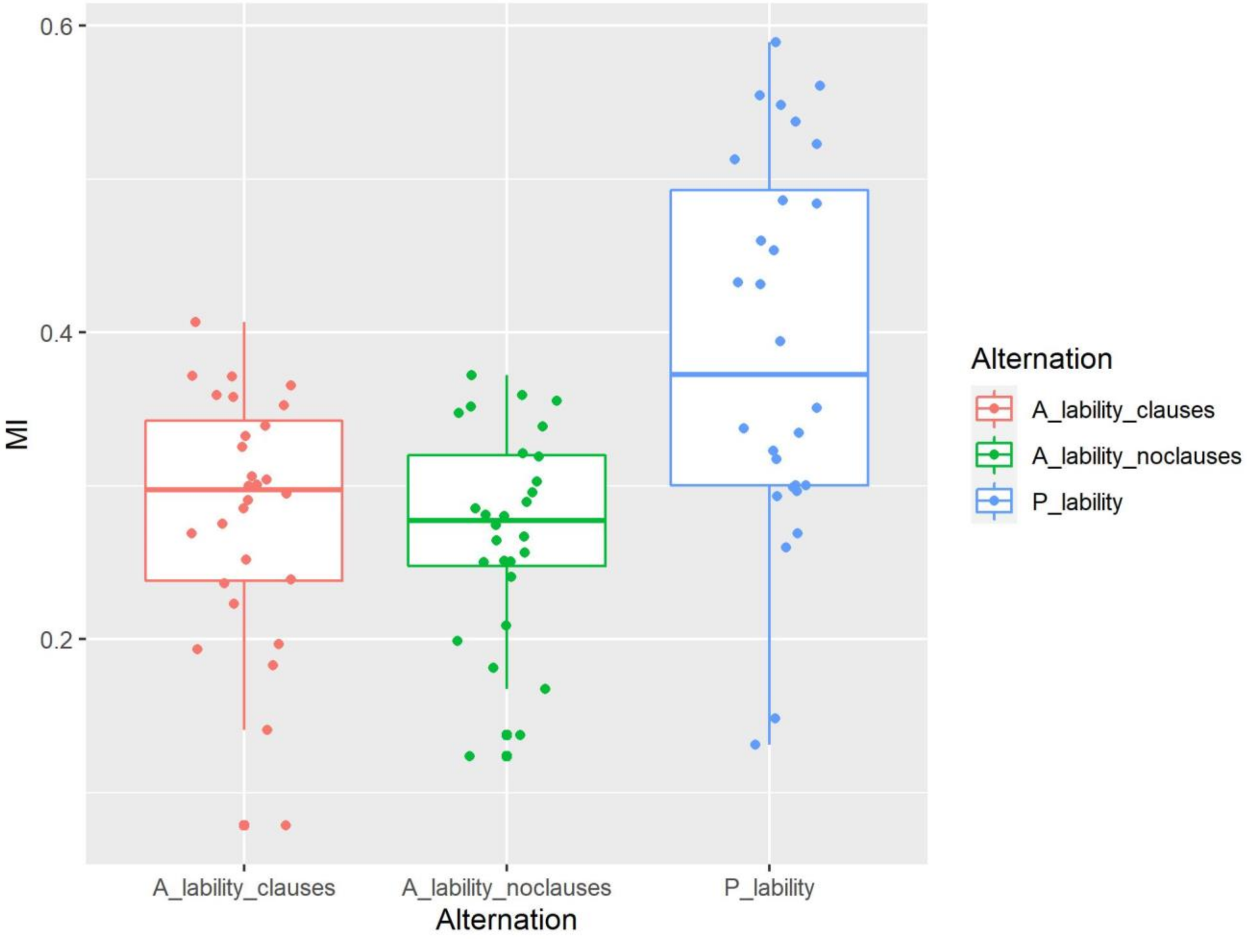
DOI: 10.5281/zenodo.5215108

Correlations with both types of A-lability



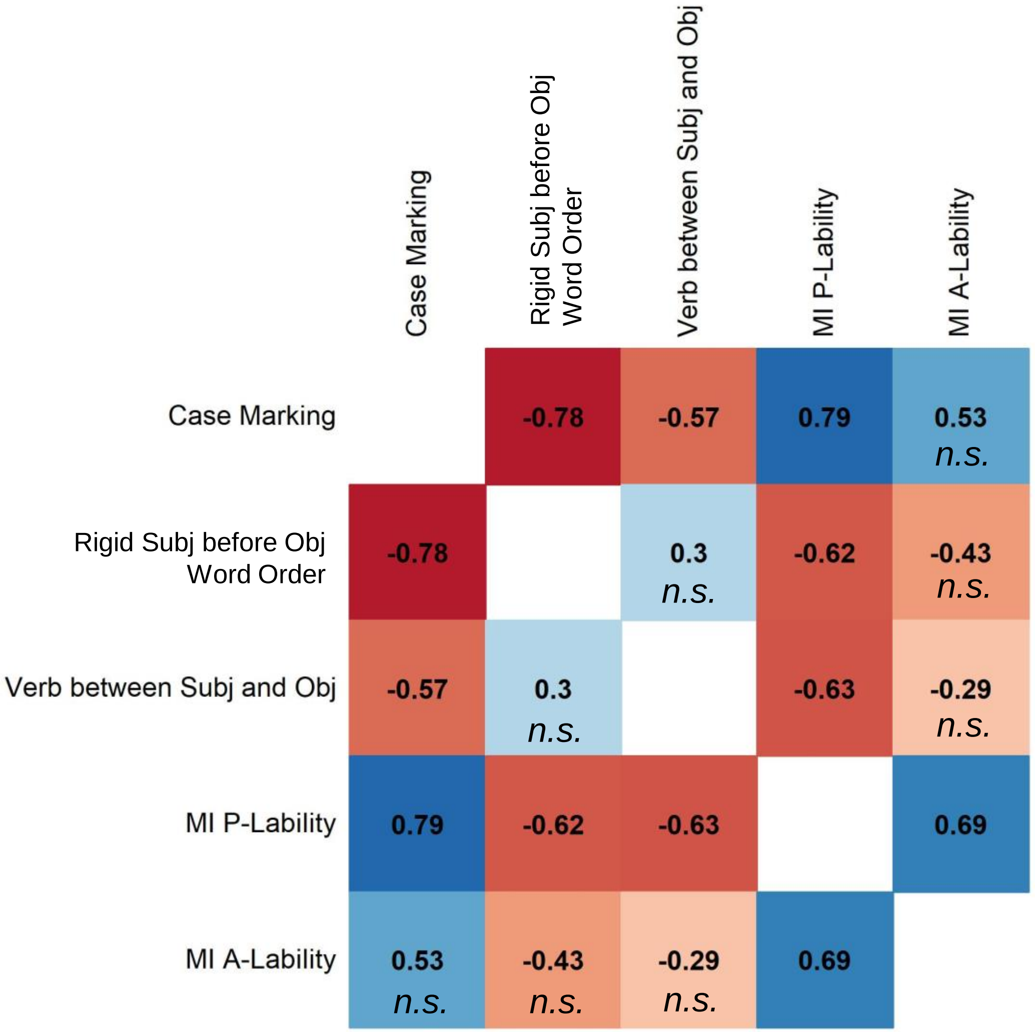
	Case Marking	Rigid Subj before Obj Word Order	Verb between Subj and Obj	MI P-Lability	MI A-Lability
Case Marking		-0.78	-0.57	0.79	0.53
Rigid Subj before Obj Word Order	-0.78		0.3	-0.62	-0.43
Verb between Subj and Obj	-0.57	0.3		-0.63	-0.29
MI P-Lability	0.79	-0.62	-0.63		0.69
MI A-Lability	0.53	-0.43	-0.29	0.69	

DOI: 10.5281/zenodo.5215108



DOI: 10.5281/zenodo.5215108

A-liability with clauses



	Case Marking	Rigid Subj before Obj Word Order	Verb between Subj and Obj	MI P-Lability	MI A-Lability
Case Marking		-0.78	-0.57	0.79	0.53 *n.s.*
Rigid Subj before Obj Word Order	-0.78		0.3 *n.s.*	-0.62	-0.43 *n.s.*
Verb between Subj and Obj	-0.57	0.3 *n.s.*		-0.63	-0.29 *n.s.*
MI P-Lability	0.79	-0.62	-0.63		0.69
MI A-Lability	0.53 *n.s.*	-0.43 *n.s.*	-0.29 *n.s.*	0.69	

Note:
Genealogical relationships were controlled by sampling 1 language per genus (1000 samples).

DOI: 10.5281/zenodo.5215108

A-lability with clauses: examples

Verb	Noun (subject)	Transitive	Intransitive
receive	family	97	0
focus	program	0	20
learn	student	37	19
say	office	65	10

DOI: 10.5281/zenodo.5215108