

Keynote Speech

Using Universal Dependencies for testing hypotheses about communicative efficiency

Natalia Levshina
Centre for Language Studies
Radboud University, Nijmegen, The Netherlands

Abstract

There is abundant evidence that language structure and use are influenced by language users' tendency to be efficient, trying to minimize the cost-to-benefit ratio of communication (e.g., Hawkins, 2004; Gibson et al., 2019; Levshina, 2022). In my talk I will show how data from corpora annotated with Universal Dependencies can be used for testing hypotheses about the role of communicative efficiency in shaping up language structure and use. The hypotheses are as follows:

1. As discussed by typologists (Sapir, 1921; Sinnemäki, 2008), rigid word order can compensate for lack of formal marking of core arguments. The hypothesis is then that there are positive correlations between the entropy of subject and object in a transitive clause in a corpus and the relative frequency of disambiguating case forms or verb forms. These correlations are expected to minimize the articulation effort involved in the use of argument flags or indices.
2. There is a positive correlation between semantic tightness (Hawkins, 1986), operationalized as Mutual Information between lexemes and syntactic roles, and the relative frequency of verb-final clauses in a corpus. Strong associations between lexemes and roles help to avoid the costs of reanalysis in verb-final languages.
3. There is a negative correlation between the relative frequency of verb-final clauses in the clause and the average number of overt core arguments, which helps to save processing costs required for keeping longer dependencies in mind (cf. Ueno & Polinsky, 2009).

These hypotheses will be tested on corpus data annotated with Universal Dependencies, with the help of mixed-effects models with genealogical and geographic information as random effects.

References

- Gibson, Edward, Richard Futrell, Steven P. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen & Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Science* 23(5): 389-407. <https://doi.org/10.1016/j.tics.2019.02.003>
- Hawkins, John A. 1986. *A Comparative Typology of English and German: Unifying the Contrasts*. London: Croom-Helm.
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Levshina, Natalia. 2022. *Communicative Efficiency: Language Structure and Use*. Cambridge: Cambridge University Press.
- Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt.
- Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. In: Matti Miestamo, Kaius Sinnemäki and Fred Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, 67–88. Amsterdam: John Benjamins.

Ueno, Mieko & Maria Polinsky. 2009. Does headedness affect processing? A new look at the VO-OV contrast. *Journal of Linguistics* 45: 675–710.