

English grammatical alternations and communicative efficiency: An information-theoretic approach based on Generalized Additive Models

Natalia Levshina (Leipzig University, Germany)

One of the central themes in corpus-based Construction Grammar has been the relationships between constructions and their collexemes (e.g. Stefanowitsch & Gries 2003 and later works). The present study focuses on one aspect of these relationships that has not received much attention, namely, maximization of communicative efficiency. If some collexemes are frequently used with a particular construction, e.g. the verb *give* with the ditransitive construction, one can say that this use has high predictability, or, using the terms from information theory (Shannon 1948), low information content/surprisal. In human communication, more predictable information tends to be expressed by shorter forms, and less predictable information is usually conveyed by longer forms. This idea is known as the principles of economy, least effort, etc. (e.g. Zipf 1949, Haiman 1980, Hawkins 2014). Some empirical support for predictability effects in constructional variation has already been found (e.g. Levy & Jaeger 2007).

The hypothesis is as follows: if there are two constructional variants that differ in formal length, the shorter variant is preferred when the collostructional combinations (e.g. the ditransitive construction + *give*) are more predictable, and the longer variant is chosen in less predictable contexts. This hypothesis is tested on the following functionally similar constructions:

- a) *help* + (to) Infinitive, e.g. *Mary helps John (to) cook the dinner*;
- b) *want to/wanna* + Infinitive;
- c) *go* (and) Verb, e.g. *Go (and) bring them in*;
- d) stative verbs + *home* vs. *at home*, e.g. *You should stay (at) home*.

In order to test this hypothesis, I use data from different English corpora and datasets (BNC, COHA, GloWbE and Google Book Ngrams). I fit several Generalized Additive Models with the short/long form as the response variable, and information content of a collexeme given the construction in question, and information content of the construction given the collexeme (cf. Attraction and Reliance in Schmid [2000]) as the predictors. The relevant variables known from the literature are controlled for. The statistical analyses support the research hypothesis.

References

- Haiman, John. 1983. Iconic and economic motivation. *Language* 59. 781-819.
- Hawkins, John A. 2014. *Cross-linguistic Variation and Efficiency*. Oxford: Oxford University Press.
- Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt & Thomas Hoffman (eds.), *Advances in neural information processing systems (NIPS)*, vol. 19, 849–856. Cambridge, MA: MIT Press.
- Schmid, Hans-Jörg. 2000. *English abstract nouns as conceptual shells. From corpus to cognition*. Berlin: Mouton de Gruyter.
- Shannon, Claude E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27. 379–423 & 623–656.
- Stefanowitsch, Anatol & Stefan Th. Gries. Collostructions: investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Zipf, George. 1949. *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press.