
Explaining coding asymmetries: Frequency or informativity?

Donnerstag
09.03.2017
09:30 – 10:00
B4 1, 0.26

Natalia Levshina
Leipzig University

natalia.levshina@uni-leipzig.de

It is well known that more frequent grammatical categories tend to be less formally marked than less frequent ones Greenberg (1966). For example, the number markers of singular nouns languages are either as long as or shorter than those of plural nouns. However, more recent computational linguistic studies (e.g. Piantadosi et al. 2011) demonstrate that formal length is in fact better predicted by average informativity of a word (i.e. the inverse of the word's average conditional probability given the preceding context) than by its context-independent frequency. These findings support the theory of uniform informational density as a means of optimization of human communication (Jaeger 2010). The present paper aims to answer the question whether linguistic coding asymmetries can be better explained by differences in relative frequency (e.g. Greenberg 1966; Haspelmath 2008) or by those in informativity.

This study focuses on three well-known cases of coding asymmetries: singular/plural nouns, absolute/comparative forms of adjectives and cardinal/ordinal numerals. On the basis of the Google books n-grams in English, French, German, Italian, Russian and Spanish, I compute the normalized and relative frequencies, as well as the average contextual predictability scores based on n-grams with different n for samples of word forms representing the grammatical categories. Goodness-of-fit measures of mixed-effect logistic regression models are used to estimate how helpful frequency and informativity are in predicting whether a given word form belongs to one or the other category.

References: • Greenberg, J. (1966): *Language universals, with special reference to feature hierarchies*. The Hague: Mouton. • Haspelmath, M. (2008): Frequencies vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19 (1), 1–33. • Jaeger T.F. (2010): Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61, 23–62. • Piantadosi, S., H. Tily & E. Gibson. (2011): Word lengths are optimized for efficient communication. *PNAS* 108(9).