Information-theoretic approaches to communicative efficiency: some controversial issues and open questions

Natalia Levshina, Leipzig University
natalevs@gmail.com

This paper provides a critical perspective on the recent information-theoretic (IT) approaches to linguistic reduction and enhancement of linguistic forms (see an overview in Gibson et al. 2019), which are based on the assumption that language users act rationally, trying to maximize the benefit-to-cost ratio of their linguistic behaviour. In this paper I will argue that this approach, albeit very promising and illuminating, has several caveats and problematic issues, which will be illustrated by cross-linguistic data, such as the following:

- the problem of measuring the costs and benefits of linguistic communication;
- the existing emphasis on the immediate linguistic co-text (e.g. n-grams) in measuring predictability, disregarding the larger context and encyclopaedic knowledge;
- the scarcity of studies that could help us understand which linguistic phenomena can be explained by the language users’ tendency to be communicatively efficient, and which should be attributed to other factors, e.g. automatization of articulation routines (Bybee 2006) or a mechanism that coordinates the speaker’s planning of articulation (e.g. Bell et al. 2009);
- insufficient recognition of the role of intersubjective rationality in communication, where the speaker and the hearer rely on the mutual assumption of each other’s efficient behaviour (Levshina 2018);
- a lack of consensus about the diachronic mechanisms that explain how efficient language structures emerge (cf. Cristofaro 2019).

In the second part, I address a more specific claim that highlights the importance of average contextual predictability (informativity) for explanation of different word lengths (Piantadosi et al. 2011; Gibson et al. 2019), while at the same time questioning the role of context-free frequency (Zipf 1935). Using large-scale English and Russian corpus data and a range of regression techniques, such as Poisson positive regression and ridge regression, I will show that a) both average predictability and frequency different types of probabilistic information play a role in predicting the length of a word; b) for particular word classes, the relative prominence of frequency is higher than that of contextual predictability.

References


