
Simulating the development of encoding asymmetries in argument marking

Sander Lestrade
Centre for Language Studies, Radboud University Nijmegen

s.lestrade@let.ru.nl

How are motivating factors (such as Jakobson's markedness, Haiman's iconicity of complexity, and Zipfian communicative efficiency) turned into linguistic conventions? In his ERC proposal, Haspelmath (2015) suggests that "higher-frequency items are more predictable than lower-frequency items, and predictable content need not be expressed overtly or can be expressed by shorter forms." Still, it is unclear what exactly are the mechanisms that create and maintain efficient structures language.

I have developed a computational model of event communication in which language evolution, or rather, at present, the development of argument marking, can be simulated. In such a model it is necessary to be maximally explicit about the mechanisms involved. I slightly deviate from Haspelmath's suggestion and tease apart the hypothesized effects of frequency and predictability, assuming that predictable parts of meaning need not be said explicitly, and that frequent words are more likely to be used again (because of a lower activation threshold) and pronounced less carefully (because of pronunciation automatization). In this talk I will introduce the model showing how asymmetrical argument marking may emerge in a protolanguage from these assumptions.

Agents communicate about automatically generated events in their virtual world. The speaker has to find an adequate wording for a target event that is sufficiently distinctive given the situational context in which other events are ongoing too (i.e., if there are similar distractor objects, referential expressions have to be more specific; cf. Levelt 1989). This also involves making clear the distribution of predicate roles over the event participants in the communicated event if necessary (contrast *man book read*, which is non-ambiguous in the absence of marking, with *man woman see*). If the hearer correctly identifies the event the speaker is talking about, the agents mark the successful usage of the words that constitute the utterance, remember the exact meaning for which the words were used, and next either switch turns to go on with their conversation or end it, after which two new agents are randomly selected for

a new conversation (cf. Steels 1997).

Over time, words may grammaticalize (Heine and Kuteva 2007). Two important mechanisms in this process are *erosion* (frequent forms being pronounced sloppily and eventually becoming represented accordingly; Nettle 1999) and *desemanticization* (frequent meanings becoming more general as a function of the different contexts in which they are used; Bybee 2010). If a meaning becomes more general, it can be used in even more contexts, and if a form becomes too short to stand on its own, it is suffixed to its host (Bybee, 1985).

As relative frequency plays a role in word activation, items that have previously been used for role disambiguation are more likely to be considered again. And as there are only two roles to be kept apart, a previously used role marker is often found good enough. As a result, its frequency of usage increases further, as well as the variation of its usage contexts. Because of the former, its form is likely to erode; because of the latter, its meaning is likely to bleach. Thus, it may end up as a (differentially used) case marker eventually.

Also the asymmetrical development of person marking follows relatively straightforwardly. Speech participants figure in many events. Because of the resulting frequent and varied usage, words referring to local persons are more prone to erosion and desemanticization. Differently from role markers, however, which do not have a referential function, once the form of a referential expression becomes too short to refer properly, a more expressive copy has to be recruited. The erstwhile local pronoun attaches to the verb indexing the person of its helper (Ariel, 1999). In contrast, third person pronouns often cannot be used, as they would lead to ambiguity because of the object distractors in the situational context.

References: • Ariel, M. (1999): "The development of person agreement markers: From pronouns to higher accessibility markers". In: *Usage based models of language*, 197-260. • Bybee, J.L. (1985): *Morphology. a study of the relation between meaning and form*. John Benjamins. • Bybee, J.L. (2010): *Language, Usage and Cognition*. Cambridge University Press. • Haspelmath, M. (2015): "Grammatical Universals". (<http://research.uni-leipzig.de/unicodas/grammatical-universals/>) • Heine, B. and T. Kuteva (2007): *The genesis of grammar: a reconstruction*. Oxford University Press. • Levelt, W. J. M. (1989): *Speaking. From intention to articulation*. MIT. • Nettle, D. (1999): *Linguistic diversity*. Oxford University Press. • Steels, L. (1997): "Constructing and sharing perceptual distinctions." *Machine Learning*. In: *ECML-97*, 4-13.