Testing iconicity: A quantitative study of causative constructions based on a parallel corpus of film subtitles

Aims

Form-function isomorphism has been a prominent topic in functionally oriented typology. In this study we focus on iconicity of cohesion, i.e. correlation between the conceptual integration of events and their formal integration (e.g. Haiman 1983). The object of our study is causative constructions, such as *cause X to die*, *make X dead* and *kill* in English, which differ with regard to the degree of formal integration of cause and effect. To the best of our knowledge, the evidence in favour of such isomorphism has been based primarily on isolated, often self-constructed examples; quantitative empirical studies are still lacking. The present study aims to fill this gap. We use corpus data from a sample of ten languages that represent different language families (according to the Ethnologue classification): Finnish, French, Hebrew, Indonesian, Japanese, Korean, Mandarin Chinese, Thai, Turkish and Vietnamese, and employ cutting-edge statistical methods (namely, ordinal regression with mixed effects) in order to put the iconicity hypothesis to test.

Data

For this study we use a self-compiled parallel corpus of film subtitles in ten above-mentioned languages plus English. Subtitles are chosen because they represent informal language and contain highly diverse causative situations in comparison with other massively parallel corpora. First, we extract approximately 250 exemplars of different causative events (e.g. ‘X causes Y to die’ or ‘X causes Y to break’) from the English subtitles. Next, we check how these events are verbalized in each of the ten languages, and classify the language-specific causative expressions into several constructional types: analytic, resultative, morphological and lexical (cf. Comrie 1981), which are defined as comparative concepts (Haspelmath 2010). The English exemplars are also coded for more than a dozen semantic variables that have been mentioned in typological literature (intentionality of causation, control of the causee, etc.), among which Dixon’s (2000) parameters of semantic variation between more and less compact causatives.

Statistical analyses and preliminary results

We use a mixed-effect ordinal logistic regression with the constructional types as the response, the semantic variables as fixed effects and the multilingual exemplars and individual languages as random intercepts and slopes. Since the semantic parameters are highly intercorrelated, we also use Multiple Correspondence Analysis as a dimensionality-reduction technique, which enables us to simplify the model. The preliminary results suggest that the iconicity hypothesis in general holds: the less cohesive the form, the less direct causation and the more autonomous the Causee. However, there is some language-specific variation in the effect of the semantic predictors, as can be seen from the random slopes in the mixed-effect model. We will also discuss the observed form-function correlation from the
point of view of an alternative account based on the Principle of Economy and frequency effects, which has been recently developed by Haspelmath (e.g. 2008), and propose a unified model of form, frequency and function.

References


