Abs-240
Antti Arppe
From modeling lexical synonyms to constructional alternations

Recently, large-scale corpus-based studies (Arppe 2008 and Divjak 2010) have explored the phenomenon of near-synonymy exploiting a battery of univariate and multivariate statistical techniques, based on a wide range of contextual linguistic features at the morphological, syntactic and semantic level. Noteworthy is that this research has demonstrated an extension of the application of statistical methods from dichotomous linguistic settings to polytomous ones, i.e., concerning more than two possible alternative outcomes with a similar meaning; however, these studies have focused on the lexical level of semantic similarity.

Although corpus-based work has also been conducted on the syntactic level concerning constructional alternations (alternatively synonymous structural variants, see Biber et al. 1998), e.g. concerning the English possessive constructions (Gries 2002 and Rosenbach 2003), the English verb-particle placement (Gries 2003a), and the English dative alternation (Gries 2003b and Bresnan et al. 2007), such work has predominantly been restricted to dichotomous alternatives. Nevertheless, it is clearly evident in general grammatical descriptions, e.g. Biber et al. (1999), that there are often more than two possible constructional alternatives, which clearly motivates a shift of interest in also constructional studies from pairs to sets with more than two alternative members.

This paper demonstrates the application of multivariate statistical analysis to the English (1) active vs. (2) be-passive vs. (3) get-passive alternation (see e.g. Biber et al. 1999), focusing on those verbs which occur in all three alternative syntactic constructions. In addition to explanatory variables based on current literature, i.e. register, stative vs. dynamic distinction, long vs. short form, length of subject/agent phrase, and preferences of the node verb itself, the underlying linguistic analysis also incorporates the scrutiny of the morphological and syntactic structure as well as semantic subcharacterizations of the context associated with the verbs in question.

In the statistical analysis, the application and results of both (1) polytomous logistic regression (see e.g. Arppe 2008) as well as its novel extension to corresponding (2) polytomous mixed-effects logistic regression modeling are demonstrated. The key benefit of these two statistical methods is that they allow us to (1) estimate the relative weights of the linguistic explanatory variables in natural terms as odds, as well as to (2) model the impact of their joint occurrence in various combinations as expected probability distributions for the alternative constructions; moreover, with the latter model we can (3) directly incorporate the effect of extralinguistic factors such as individual speaker/writer preferences.

The results support a probabilistic view of the relationship between linguistic usage and the underlying linguistic system, in which only a minority of linguistic choices are categorical - instead, most contexts exhibit degrees of variation as to their outcomes, resulting in proportionate choices over longer stretches of usage in texts or speech.

REFERENCES

Arppe, A. 2008. Univariate, bivariate and multivariate methods in corpus-based lexicography – a study of synonymy. Publications of the Department of General Linguistics, University of Helsinki, No. 44.

Biber, D., S. Conrad, and R. Reppen. 1998. Corpus linguistics: Investigating language structure and use. Cambridge: Cambridge University Press.

Biber, D., S. Johansson, G. Leech, S. Conrad, E. Finegan. 1999. The Longman grammar of spoken and written English. London: Longman.

Bresnan, J., A. Cueni, T. Nikitina, and R. H. Baayen 2007. Predicting the Dative Alternation. In: Cognitive Foundations of Interpretation. Boume, G., I. Kraemer, and J. Zwarts. Amsterdam: Royal Netherlands Academy of Science, pp. 69-94.

Divjak, D. 2010. Structuring the lexicon: a clustered model for near-synonymy. Berlin: Mouton de Gruyter. Cognitive Linguistics Research.

Gries, S. Th. 2002. Evidence in linguistics: Three approaches to genitives in English. In: Brend, R. M., W. J. Sullivan and A. R. Lommel (eds.). LACUS Forum XXVIII: What Constitutes Evidence in Linguistics? Fullerton: LACUS, pp. 17–31.

Gries, S. Th. 2003a. Multifactorial analysis in corpus linguistics: a study of particle placement. London: Continuum.

Gries, S. Th. 2003b. Towards a corpus-based identification of prototypical instances of constructions. Annual Review of Cognitive Linguistics, Vol. 1, pp. 1-27.

Rosenbach, A. 2003. Aspects of iconicity and economy in the choice between the s-genitive and the of-genitive in English. In: Rohdenburg, G. and B. Mondorf (eds.), Determinants of Grammatical Variation in English. Berlin/New York: Mouton de Gruyter, pp. 379-411.