Prototype identification methods in cognitive corpus linguistics – an overview

Keywords: prototype model of categorization, prototype identification, corpus data, methodology

The prototype model of categorization (Rosch, 1973 & 1975; Rosch & Mervis, 1975), which posits that natural categories are organized around highly salient category members, or prototypes, and assumes the existence of graded category structure, has proved to be a powerful descriptive and explanatory tool also for the organization of linguistic categories, especially when one considers the well-established tradition of cognitive semantic research in polysemy and synonymy (for overviews of the relevant studies, spanning nearly four decades, see for example Glynn, 2010 & 2014). At the same time, although the theoretical constructs such as prototype structure have remained stable in their impact and appeal to cognitive linguists, the field has undergone a major methodological transformation in the meantime (e.g., Gries & Stefanowitsch, 2006; Glynn & Fischer, 2010; Janda, 2013; Glynn & Robinson, 2014; Divjak, Levshina & Klavan, 2016). This shift away from self-constructed examples and intuition/introspection towards corpus data and quantitative methods has also profoundly influenced the way linguistic categories are analyzed, and the present paper aims to provide an overview of recent methodological approaches and techniques that have been employed to address the issue of prototype identification in polysemous and near-synonymous categories situated at various points along the lexis-grammar continuum.

In this presentation, we will refer to a number of recent publications in cognitive linguistics, all of which use corpus data (sometimes coupled with subsequent experimental validation) and follow the principle of total accountability, but employ techniques at different levels of quantitative sophistication, ranging from frequency of use (e.g., Gilquin, 2006) through collostructional analysis (e.g., Stefanowitsch & Gries, 2003) to hierarchical clustering (e.g., Divjak & Gries, 2006) and multiple correspondence analysis combined with loglinear modelling (e.g., Glynn, 2016). The papers in focus also represent a varied cross-section of research objectives: while some are case studies aiming to investigate the organization of a given category, be it a lexeme (e.g., Jansegers, Vanderschueren & Enghels, 2015) or a syntactic construction (e.g., Perek, 2014), others additionally propose and test a procedure to determine prototype structure (e.g., Gries, 2003), empirically evaluate criteria for prototypicality (e.g., Gilquin & McMichael, 2018) or suggest theory-driven improvements to current practices in the field (e.g., Glynn, 2016). All of them, however, epitomize the efforts of the discipline to take the usage-based commitment seriously (cf. Tummers, Heylen & Geeraerts, 2005), and the

purpose of this paper is to discuss the main tendencies and depict the full methodological arsenal we currently have at our disposal when it comes to identifying categorical relationships within and between linguistic categories on the basis of corpus data.

References

Divjak, D., & Gries, St. Th. (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, 2(1), 23–60.

Divjak, D., Levshina, N., & Klavan, J. (Eds). (2016). *Cognitive Linguistics: Looking Back, Looking Forward*. Special issue of *Cognitive Linguistics*, 27(4).

Gilquin, G. (2006). The place of prototypicality in corpus linguistics: Causation in the hot seat. In St. Th. Gries, & A. Stefanowitsch (Eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis* (pp. 159–191). Berlin and New York: Mouton De Gruyter.

Gilquin, G., & McMichael, A. (2018). Through the prototypes of *through*: A corpus-based cognitive analysis. *GCLA*, 6, 43–69.

Glynn, D. (2010). Corpus-driven Cognitive Semantics: Introduction to the field. In D. Glynn, & K. Fischer (Eds.), *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches* (pp. 1–41). Berlin and New York: Mouton De Gruyter.

Glynn, D. (2014). Polysemy and synonymy: Cognitive theory and corpus method. In D. Glynn, & J. A. Robinson (Eds.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy* (pp. 7–38). Amsterdam and Philadelphia: John Benjamins.

Glynn, D. (2016). Quantifying polysemy: Corpus methodology for prototype theory. *Folia Linguistica*, 50(2), 413–447.

Glynn, D., & Fischer, K. (Eds.). (2010). *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches*. Berlin and New York: Mouton De Gruyter.

Glynn, D., & Robinson, J. A. (Eds.). (2014). *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Amsterdam and Philadelphia: John Benjamins.

Gries, St. Th. (2003). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics*, 1, 1–27.

Gries, St. Th., & Stefanowitsch, A. (Eds.). (2006). *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. Berlin and New York: Mouton De Gruyter.

Janda. L. A. (Ed.). (2013). *Cognitive Linguistics: The Quantitative Turn (The Essential Reader)*. Berlin and Boston: Mouton De Gruyter.

Jansegers, M., Vanderschueren, C., & Enghels, R. (2015). The polysemy of the Spanish verb *sentir*: A behavioral profile analysis. *Cognitive Linguistics*, 26(3), 381–421.

Perek, F. (2014). Rethinking constructional polysemy: The case of the English conative construction. In D. Glynn, & J. A. Robinson (Eds.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy* (pp. 61–85). Amsterdam and Philadelphia: John Benjamins.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 111–144). New York: Academic Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192–233.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.

Stefanowitsch A., & Gries, St. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.

Tummers, J., Heylen, K., & Geeraerts, D. (2005). Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory*, 1(2), 225–261.