

“Up close and personal vs. birds-eye view” of discourse:
a corpus study of perspective using Czech data

Václav Cvrček & Masako Fidler
Charles University & Brown University
vaclav.cvrcek@ff.cuni.cz, masako_fidler@brown.edu

Keywords: cognitive grammar, perspective-conceptualizer, discourse, corpus linguistics, Czech

Perspective, one of the essential building blocks in Cognitive grammar, affects how the perceived situation is represented; a perceived situation is rendered more prominent as perspective is positioned further away (Langacker 2002: 315-318). This paper provides corpus-based quantitative data to explore perspective on the level of discourse: i.e., how a text is perceived from differently-located points of perspective.

The study uses the corpus linguistic method of keyword analysis (KWA). It compares the target text (Ttxt) against the background of a reference corpus (RefC), and thereby identifies prominent words (keywords, KWs) that distinguish the Ttxt from the RefC (Scott 1999). KWA is used to probe various aspects of a Ttxt (cf. Culpeper 2002, Baker 2009), but it has been found that RefCs used in KWA can serve to represent different reader viewpoints (Fidler & Cvrček 2015).

In this paper KWA is applied to Ttxts of specialized genres and registers (e.g. medicine textbook). To our knowledge, this is the first study using KWA to explore perspectives of contemporary readers exposed to different genres. The Ttxts are contrasted against the background of several RefCs drawn from SYN2015, a general corpus of written contemporary Czech. These RefCs represent different reader viewpoints, e.g.: a corpus of fiction (prose) and corpora of non-fiction texts; the latter may be parceled out according to characteristics such as (a) basic research vs. application and (b) specialization (natural science, agriculture, and medicine). The fiction corpus is associated with the reader who is least exposed to the medical genre, while the corpora on medical research and its application are associated with the reader who is most familiar to the medical genre. The intuitive perception of register closeness can be objectivized by multi-dimensional model of register variation of Czech (Cvrček et al. 2018). The KWAs based on RefCs with closer proximity to the medical genre not only result in lower numbers of KWs but more importantly, they yield lower values of Difference Index to KWs. The results provide the discourse correlate of the role played by perspective observed in the grammatical construal: the reader familiar with the genre does not find the Ttxts to be unusual or striking from what s/he routinely encounters, while the reader inexperienced with the genre finds them unusual with a number of unfamiliar words that may become stumbling blocks in comprehension. The study uses several RefCs of different degrees of similarity to the Ttxt to illustrate the scalar nature of prominence in discourse as perspective moves further away from the genre in which the text is written.

References

- Baker, Paul. (2009). The question is, how cruel is it? Keywords in debates on fox hunting in the British House of Commons. D. Archer, ed. *What's in a Word-list?* Ashgate, 125–36.
- Culpeper, Jonathan. (2002). Computers, language and characterisation: An analysis of six characters in *Romeo and Juliet*. U. M. Marttala et al, eds. *Conversation in life and in literature: Papers from the ASLA Symposium*, 15:11–30.
- Cvrček, Václav, Zuzana Komrsková, David Lukeš, Petra Poukarová, Anna Řehořková & Adrian J. Zasina, A.J. (2019). From extra- to intratextual characteristics: Charting the space of variation in Czech through MDA. *Corpus Linguistics and Linguistic Theory*. [Ahead of print]
- Langacker, Ronald W. (2002). *Concept, Image, and Symbol. The Cognitive Basis of Grammar*. Berlin-New York: Mouton.
- Scott, Mike. (1999). *WordSmith Tools Help Manual*. Version 3.0.
- Křen, Michal et al. (2015). *SYN2015: A representative corpus of written Czech*. Institute of the Czech National Corpus, Charles University. (<http://www.korpus.cz>)
- Fidler, Masako & Václav Cvrček. (2015). A data-driven analysis of reader viewpoints: Reconstructing the historical reader using keyword analysis. *J. of Slavic Linguistics*. 23(2). 197-239.