Title: Keyword Analysis with a Usage-Based Perspective: A Preliminary Study in Czech
Author: Václav Cvrček, Charles University, Prague and Masako Fidler, Brown University

The main idea of this paper is that interpretation of a text emerges in a way parallel to what is observed in usage-based grammar: just as language usage shapes grammar, language usage shapes the interpretation of the text. We demonstrate this hypothesis with the help of keyword analysis (KWA). KWA statistically identifies word forms (keywords, KW) by comparing the target text with a larger corpus that reflects the general linguistic patterns of the language (Scott and Tribble 2006, Baker and Ellece 2011).

KWs are connected to the overarching themes and genre of the target text. Shifts in KW links (co-occurrences of two KWs in a context window) over time reflect the changing discourse role of the KW, just as shifts in the syntactic and semantic contexts reflect the changing grammatical function of a form, cf. Bybee's exemplar model (2010). For example, the KW *lidu* '(of) people' occurs consistently in the Czechoslovak Presidential New Year's Addresses (NYAs) from 1975–77. A notable difference in KW links is observed between 1975 and 76: *lidu* is linked to KWs representing economy, the USSR, standard of living and results only in 1975. *Lidu* is linked increasingly to KWs concerning communist and socialist ideology and to political entities ('the National Front', 'the (political) system') in 1976 and 1977. Simultaneously, however, *lidu* is linked less to future intentions and abilities from 1975 to 77 ('we can', 'we will' in 1975, 'we want' in 1976, and no inflected verb form as a KW in 1977). 'People' are therefore seen as less active participants in events, and increasingly as part of discussions of the political system. The shift in KW links for *lidu* suggests a subtle shift in the usage of the word *lid*, which in turn signals a shift in its discourse role.

We compare the rankings of KWs in Gustáv Husák's NYAs, the relationship among KWs within each speech, and the collocational patterns of recurring KWs to stretch the notion of usage-based grammar to discourse. The results are drawn from the Czech National Corpus, including the language data used in Čermák, Cvrček and Schmiedtová 2010.

Bybee, J. 2010. Language, usage and cognition.
Baker, P. and Ellece, S. 2011. Key terms in discourse analysis.
Čermák, F., V. Cvrček, and V. Schmiedtová, eds. 2010. Slovník komunistické totality.
Czech National Corpus (www.korpus.cz).
Scott, M. and C. Tribble. 2006. Textual patterns: Keyword and corpus analysis in language education.