# Usage-based approach to discourse through keyword analysis

Masako Ueda Fidler & Václav Cvrček
*Brown University & Charles University*

This co-authored paper will be an attempt to demonstrate the relevance of the usage-based approach (Hopper 1987, Bybee 2010) to describing genre-specific discourse structure. In a way similar to how grammar and semantics of lexicon emerge from language usage, we anticipate that higher-level linguistic patterns such as genre-specific discourse properties can emerge through language use, especially through changing patterns of interaction between keywords (KWs) and context and through the interaction between KWs.

KWs are word forms obtained statistically (e.g. chi-square, log-likelihood tests), i.e. outputs by comparing the target text(s) against the background of a larger corpus that reflects the general usage of the language (the reference corpus) (Scott and Tribble 1996, Baker and Ellece 2011). They are word forms that occur in a text more frequently than expected by chance alone and are often closely connected to the overarching themes and genre of a text or set of texts. Shifts in KW links (links between one KW and other KWs) and shifts in their contextual use over time reflect the changing discourse role of the KW.

To demonstrate the relationship between emerging changes in genre-specific discourse practice and KW behavior, we will examine the New Year's Addresses from 1975 to 1989 by President of the socialist Czechoslovakia. The choice of texts is motivated by four factors: we can keep two properties of the texts nearly constant (author and genre); the texts are often assumed to be static, therefore posing a reasonable challenge to our hypothesis; large and balanced language corpora are available; and we have developed a beta-version of a keyword analysis application for Czech called KWords.

Our preliminary findings suggest that even a single KW can point to emerging changes in discourse structure. For example, while the word lidu 'of people' is linked to KWs representing more tangible KWs (economy and international politics) in mid-1970s, it is increasingly linked to abstract KWs (ideology) and less to action towards late 1970s. This may be seen as an indication of an emerging pattern of a schematic discourse (with decreasing substance) where people themselves are becoming increasingly peripheral to discussions of the country's future. Analysis of the texts will be based on the quantitative data obtained from our KWA application: frequencies of KWs, their ranking, links among KWs, contexts in which KWs occur in 15 texts. Each text contains approximately 1,000 words and will be contrasted with texts from large language corpora available from the Czech National Corpus.

## References

Bybee, J. 2010. Language, usage and cognition.

Baker, P. and Ellece, S. 2011. Key terms in discourse analysis.

Hopper, Paul. Emergent grammar. Berkeley Linguistics Society 13: 139-157. Berkeley Linguistics
        Society, vol. 13 (1987), 139-157

Czech National Corpus (www.korpus.cz).

Scott, M. and C. Tribble. 2006. Textual patterns: Keyword and corpus analysis in language education.