Title: A Tool for Keyword Analysis in Czech
Author: Václav Cvrček, Charles University, Prague

Corpus-based keyword analysis is a rapidly-evolving and widely used method for interpreting all sorts of texts (including literary) in English (Baker and Ellece 2011, Scott and Tribble 2006). Despite the fact that its use is widely appreciated and keyword analysis has been used even in some studies of Czech (Čermák et al. 2010) there has been no computational tool freely available to scholars who want a quantitative basis for their interpretation of texts.

This paper presents a web-based keyword analyzer for Czech developed at the Institute of the Czech National Corpus, which is capable of handling large texts and provides the user with simple and straightforward information about keywords in target texts. The text entered by the user is tokenized by the application and the frequency for each word type is counted. The user has a choice of corpus that he/she would like to use as a general background against which the target text is compared. Once the analysis is done, the user may look at the text with highlighted keywords, examine the keyword list that contains frequency and statistical measures developed in the process of identification, look at the dispersion plot (depicting the spread of keywords in the text), and explore the collocations of keywords, which sheds light on the use of these words and the role they play in a text.