

Title: SyD – A Learners’ Tool for Exploring Variability in Language Corpora of Czech
Author: Václav Cvrček, Charles University, Prague

This paper describes the project and design of SyD, a new tool developed at the Institute of the Czech National Corpus, for exploring the variability of language in corpora of Czech, which can be useful for linguists as well as for learners of Czech. The simple Google-like opening form of this web-based application allows users to enter two or more variants. The application then looks for each variant in the corpora of Czech. SyD has two parts: synchronic and diachronic (hence the name SyD). The synchronic part, which is the more useful part for teaching Czech as a second language, searches for the variant strings in the corpora of contemporary spoken and written texts. It counts the overall absolute and relative frequencies of each variant, plus its partial frequencies for various subcorpora, e.g., how many times each variant occurs in specific text types or genres, or how it is pronounced by different groups of speakers (according to sex, age, education, region), etc. This can provide the user with straightforward and simple information about the stratification of Czech, which, due to its complexity, is quite confusing even for advanced learners of Czech.

vaclav.cvrcek@ff.cuni.cz