

Abstract

Title: Corpus approach to teaching Czech as a foreign language

Author: Mgr. Bc. Adrian Jan Zasina

Department: Institute of the Czech National Corpus

Supervisor: Mgr. Lucie Lukešová, Ph.D.

Abstract:

Recently, we have experienced a rapid development in information technologies which is closely linked with the development of language corpora. Language corpora provide an indispensable amount of authentic data and information about how the real language works, therefore it is no wonder they deserve our attention in language teaching as well. We have seen a growing interest in the use of corpus resources in teaching Czech as a foreign language, however, the methodological issues dealing with specific applications have so far been rather neglected. Therefore, the present dissertation aims to systematically map the possibilities of incorporating corpora into current language teaching while considering the specific language needs of foreigners based on their typical errors. The emphasis is not only on a coherent methodological framework, but also on a practical demonstration of the work with corpus data, which is intended not only for corpus specialists but especially for teachers and their students.

The basis of the work lies in an extensive analysis mapping the most problematic phenomena in teaching Czech as a foreign language based on data from the CzeSL-SGT learner corpus. The research was carried out on the data containing texts written by students with a Slavic mother tongue. Automatic error annotation has shown that problems with writing long vowels are the most frequent ones. Furthermore, the most interesting phenomena, which were analysed in detail in the case studies, include: deverbal nouns ending on *-ání* vs. *-aní*, past participle, and competitive endings *-a* and *-u* in the genitive singular of inanimate masculine nouns. The corpus approach in teaching was presented in detail in these examples following a proposed three-step methodology: the identification of a problematic language phenomenon based on a learner corpus, its classification based on general corpus data, and the generalisation of results. Each case study also offered author's corpus exercises which could be used in resolving the presented issues with students.

This doctoral thesis also presents a didactical experiment conducted on students of Czech studies. The aim of the experiment was both to examine the possibility of introducing corpus methods into language teaching and to verify whether such an approach brings measurable advantages in the language development of students. The experiment confirmed that the proposed procedure, from the identification of a language problem, to its categorization based on general corpus data, and finally to the generalization of results, is functional in practice. In addition, the results pointed out that the group with the corpus approach in the long-term perspective tends to improve more in comparison to the traditional group.

Based on the present research, the most common issues of non-native Czech language from the perspective of Slavic speakers were discussed. Secondly, a coherent framework for examining language data with regard to teaching has been systematised and described. Finally, the functionality of this method has been verified in practice and original corpus exercises have been made available for further use.

Keywords: DDL, non-native Czech language, corpus linguistics, teaching Czech as a foreign language, language teaching, corpus methods