

Corpus-based resources for L1 and L2 teaching of Czech

Lucie Chlumská (Charles University, Czech Republic), Anna Čermáková (Charles University, Czech Republic) and Pavlína Vališová (Masaryk University, Czech Republic)

Introduction

Applications of corpus linguistics to language teaching began as early as the late eighties and nineties (e.g. Higgins 1988, Stevens 1991, J. Flowerdew 1993) and to this day, considerable amount of work has been done in this area, mostly in the second language acquisition and in teaching language for specific purposes. It is not surprising that it is English that has been in the centre of attention: as second language (ESL), as lingua franca (ELF), or as language for specific purposes (ESP). Currently, the learner corpora research is one of the most progressive areas of corpus linguistics (c.f. e.g. the ICLE¹ project launched by Sylviane Granger and her colleagues at the University of Louvain).

However, the research on corpora in language teaching seems to have rather neglected the idea of using corpora and corpus-based resources in L1 teaching, especially at primary and secondary school levels. To our knowledge, the CLLIP project (Sealey & Thompson 2004) employing a corpus-based approach to teach 8-10 year-olds and the web-based Englicious project² (developed by the team of Survey of English Usage, UCL) are one of the rare exceptions in the past few years.

This paper aims to introduce our joint effort to a) provide a basis for a theoretical framework to using corpus-based teaching materials for Czech, both as L1 and L2, incl. the identification of specific needs of L1 and L2 learners, and b) to present sample activities of real-life application of corpus-based exercises at the secondary school level (L1) and in the foreign learners' classroom (L2).

Teaching Czech as L1 and L2

Czech presents, in comparison to English, a rather specific language situation: it is a 'small' language, which lacks some of the resources taken for granted in teaching e.g. English; there is no contemporary general dictionary, no school dictionaries, and the traditional mother tongue teaching curriculum in schools is heavily relying on teaching grammar and orthography rules and favouring literature lessons at the expense of language. According to the PISA cross-national assessment, Czech pupils are consistently below the OECD average in reading skills and text comprehension.

This urgent need to improve and change the mother tongue teaching in schools has been now acknowledged and an extensive research team from The

¹ <https://www.uclouvain.be/en-cecl-icle.html>

² <http://www.englicious.org/>

Faculty of Education³, Charles University, is currently mapping the situation in primary schools in order to suggest changes. These changes should also involve the development of new, corpus-based, resources.

Teaching Czech as a foreign language has also been rather traditional and conservative. Due to the fact that Czech is a highly inflected language, textbook exercises mostly focus on grammar drills and use invented model sentences and texts (very few textbooks use authentic examples). Learners of Czech as L2 tend to be of different mother tongues and it is common to have a very heterogeneous classroom – that is why teaching resources are not usually developed for a specific target group of non-native speakers of Czech.

One of our long-term objectives therefore is to make use of extensive corpus resources that are available for Czech, which include several national corpora, large parallel corpus and various specialised corpora, including e. g. a corpus of student writing. We aim to create teaching resources both for L1 and L2. Our efforts are currently being supported by various outreach activities aimed at both students and teachers; where – while raising awareness of corpus resources – the resources are at the same time being tested providing thus important user feedback. We have now established cooperation with the publishing house Fraus, a major Czech publisher of educational resources and developing further cooperation with the Faculty of Education within the above mentioned project. Currently the most visible output are corpus-based exercises incorporated into the Fraus textbooks and our platform www.korpus.cz/proskoly aimed at teachers (see below).

Comparing the needs of L1 and L2 learners

Since learners of Czech as L1 and L2 have different needs, it is necessary to first identify their specific nature and learning requirements. The following summary is based on our own hands-on experience.

In the L1 learning context, Czech is taught from the age of 6 to the secondary school graduate level of 18-19 year-olds. Corpus-based language exercises can be used at most levels (possibly excl. the youngest pupils who learn to read and write), e.g. our pilot project with 10-12 year-old students demonstrated that they could successfully navigate through the corpus interface and enjoyed working with real data (e.g. comparing selected features of spoken and written language, discovering word meanings in their natural context, or engaging in activities with a parallel corpus including Harry Potter series). However, at the same time it pointed to a need for a less complex interface should corpora be used hands-on during classes.

On the other hand, L2 learners of Czech are mostly adults and they learn Czech for very different reasons: most often it is their need for “survival” in a foreign country while studying/working in the Czech Republic. The main goal for everyone is to successfully communicate with native speakers. They especially need to learn how to combine words, how to use them in context and what is appropriate in different

³ Project PRIMUS/HUM/19 entitled *Didaktika českého jazyka v současném vzdělávacím kontextu* (Czech Language Didactics in Current Educational Context) led by Stanislav Štěpáník, PhD.

communicative situations, since colloquial, spoken Czech is very different from the formal, written language (on the morphological, syntactic as well as lexical level).

What both groups have in common is the need to master Czech orthography rules which may often seem counterintuitive and full of exceptions. The possibility to teach difficult spelling through concordances has already successfully been tested e.g. on French (Tyne 2016).

Types of corpus-based materials

Corpora in teaching can be used both directly (“hands-on”, using computers/tablets/mobile phones) and indirectly (“paper” exercises based on corpus data). The first option seems to be at the moment less favoured by teachers (due to too complex corpus query interface and computer lab accessibility); however, based on our experience, students themselves enjoy working with some of our online corpus-based tools, e.g. SyD – a corpus-based tool comparing variants in spoken and written Czech. To support the second, more favoured, option, i.e. ready-made exercises, we have launched an online repository of ready-made L1 and L2 corpus-based exercises for teachers⁴.

In general, we have been working with the following typology of corpus-based exercises (based on Vališová 2016) with the aim to identify the most suitable resources for both L1 and L2. In addition to the direct approach (searching corpora), the corpus-based materials can be based on e.g. 1) observation – a selection of concordance lines with one word or a phrase, two words / phrases for comparison, 2) gap-filling – one or more words missing in the concordance lines, 3) mixed concordances – matching the corresponding left and right context, 4) identification – matching a node word with its collocates or assigning the right text type, 5) mix – combination of the above-mentioned types.

Sample activities

In the L1 environment, we have provided corpus-based exercises (mostly focusing on lexicology, word formation, or orthography) for Czech language textbooks for 7th-9th grade (13-15-year old students) as well as for secondary schools (examples will be provided in the presentation). Currently, we are testing the corpus-based approach on the secondary school level in direct cooperation with students themselves – presenting findings about language to their classmates and assessing the difficulty of exercises.

Our L2 activities are based on the corpus of written contemporary Czech and consist of authentic or partly edited sentences. Some of the exercises have already been successfully tested in an English speaking classroom, A2 level. The exercises focus on problematic words and their use (e.g. *žít* vs. *bydlet* – both translated as ‘to live’ into English). The testing showed that the ideal model is to introduce several sentences in concordances to observe the behaviour of the word (its meaning and

⁴ www.korpus.cz/proskoly

typical context) and then to present more examples in gap-filling exercises. Gap-filling exercises are more demanding for the teacher to prepare, as it is necessary in Czech to put into the brackets additional morphological information such as person, singular / plural, or tense for verbs. However, the great benefit is in practicing meaning and grammar at the same time.

References

- Flowerdew, J. (1993). Concordancing as a tool in course design. *System* 21(2), 231–244.
- Higgins, J. (1988). *Language, learners and computers*. London: Longman.
- Sealey, A. & Thompson, P. (2004). "What do you call the dull words?" primary school children using corpus-based approaches to learn about language. *English in Education* 38(1), 80–91.
- Stevens, B. (1991). Concordance-based vocabulary exercises: A viable alternative to gap-fillers. *English Language Research Journal* 4, 47–61.
- Tyne, H. (2016). Developing spelling with data-driven learning in L1 French, presented at TALC12 (Teaching and Language Corpora Conference), Giessen 2014. Abstract available at <http://www.uni-giessen.de/faculties/f05/engl/ling/talc/home/programme/abstracts>.
- Vališová, P. (2016). Korpus ve výuce češtiny jako cizího jazyka – typy cvičení. In I. Starý Kořánová & T. Vučka (Eds.), *Čeština jako cizí jazyk VIII. Sborník příspěvků z VIII. mezinárodního symposia o češtině jako cizím jazyku*. Praha: Univerzita Karlova.