

THE CHOICE OF VERBS IN CZECH AS A FOREIGN LANGUAGE TEXTBOOKS

Abstract The aim of this paper is to present a project of compilation of a small specialized corpus, which consists of 7 textbooks of Czech language as a foreign language (CFL). The present analysis gives a preliminary account of a more extensive study of the choice of vocabulary in CFL textbooks. It focuses on the verb forms and the distribution of perfective and imperfective verbs in elementary level CFL textbooks.

1. Czech pedagogical corpus and its use

1.1. Traditional textbook analysis

During the last twenty years the discipline Czech as a foreign language (CFL) began to develop very quickly and many modern communicative textbooks were published as well as a lot of textbook analysis were conducted. If we take a closer look at these analyses, we can see that most of them are Bachelor or Master Theses. They focus on a large range of topics, e.g., grammar, word-formation, colloquial vocabulary or sociocultural skills in textbooks¹. Hrdlička wrote the only monographs and they deal with prepositions in textbooks² or the approach to teaching grammar in general³. He summarized his findings claiming there are two tendencies in CFL textbooks: 1. Overestimation of grammar and 2. Oversimplification of the language⁴. Therefore, a detailed research of textbook language in comparison with the Czech National Corpus (CNC) can be beneficial

¹ *Vališová P.* Korpus jako zdroj dat systémového popisu české konjugace při výuce češtiny jako cizího jazyka. Brno, 2009. Accessible at: <http://theses.cz/id/qw5wz5>

² *Hrdlička M.* Předložky ve výuce češtiny jako cizího jazyka. Praha. 2000.

³ *Hrdlička M.* Gramatika a výuka češtiny jako cizího jazyka. Praha. 2010.

⁴ *Ibidem.*

in producing teaching materials presenting contemporary usage and using authentic examples.

1.2. Pedagogical corpora

This type of pedagogical corpora is an example of indirect use of corpora in language teaching.⁵ It substitutes and automatizes traditional textbook analysis, which is carried out page by page⁶. The corpus of English as a foreign language (EFL) textbooks – TeMa, which has been collected at Louvain University⁷ and is pedagogically annotated, is considered a model for a pedagogical corpus. A similar project is being carried out by Römer⁸, who had collected a corpus of texts in EFL textbooks for Germans. A pedagogical corpus can be used not only to research the choice of vocabulary and compare it with authentic language, but also to focus on many other topics, such as language acquisition⁹, language of instructions, metalanguage, spoken and written variants, presentation of grammar or types of exercises (if the corpus includes a pedagogical annotation)¹⁰.

⁵ *Gabrielatos C.* Corpora and language teaching: Just a fling, or wedding bells? In: *TESL-EJ*, 2005, vol. 8/4, A1. P. 1–37; *Römer U.* Corpora and language teaching. In *A. Lüdeling – M. Kytö.* *Corpus Linguistics: an international handbook.* Berlin, 2008.

⁶ *Gouveneur C., Meunier F.* New types of corpora for new educational challenges. In *K. Aijmer* (ed.) *Corpora and Language Teaching.* Amsterdam, 2009.

⁷ <http://www.uclouvain.be/en-cecl-tema.html>

⁸ *Römer U.* A corpus-driven approach to modal auxiliaries and their didactics. In: *J. Sinclair* (ed.) *How to Use Corpora in Language Teaching.* Amsterdam, 2004. P. 185–199; *Römer U.* Looking at looking: Functions and contexts of progressives in spoken English and 'school' English. In *A. Renouf, A. Kehoe* (eds.). *The Changing Face of Corpus Linguistics.* Amsterdam: 2004.

⁹ *Huston S.* *Corpora in Applied linguistics.* Cambridge: 2004.

¹⁰ *Gouveneur C., Meunier F.* New types of corpora for new educational challenges. In *K. Aijmer* (ed.) *Corpora and Language Teaching.* Amsterdam, 2009.

2. The choice of vocabulary in textbooks

The previous manual textbook analysis showed that CFL textbooks usually tend to simplify the language as much as possible and try to avoid perfective verbs despite the fact that some of them belong to the most frequent vocabulary, e.g., *dokázat, najít, přijít, zapomenout, zůstat*, as the data from CNC shows¹¹. Therefore, the research questions are: Which verbs should we choose to present on elementary level? Should we present perfective verbs on this level already? The method is corpus-driven¹²: first, the verbs presented in 7 textbooks of Czech as a foreign language for elementary level are examined and then compared with the description of A1 level according to the Common European Framework for Languages (CEFR)¹³ and also with the data from the CNC¹⁴.

2.1. The description of A1 level for Czech

The A1 level according to the CEFR is considered as the lowest level, which does not cover almost any grammar¹⁵. Nevertheless, Czech, as a highly inflected language, has to include more grammatical features than analytical languages in its level description according to the CEFR. The students on A1 level can «use simple phrases and sentences to describe where I live and people I know»¹⁶

¹¹ Vališová P. Korpus jako zdroj dat systémového popisu české konjugace při výuce češtiny jako cizího jazyka. Brno, 2009.

¹² Huston S. Corpora in Applied Linguistics. Cambridge: 2010.

¹³ The A1 level in Czech is also a level of the Examination of the Czech Language for Permanent Residence in the CR: <http://check-your-czech.com/>

¹⁴ <http://korpus.cz/>

¹⁵ The Common European Framework in its political and educational context. Accessible at: http://www.coe.int/t/dg4/linguistic/source/framework_en.pdf

¹⁶ European Levels – Self Assessment Grid. Accessible at: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&ved=0CEwQFjAD&url=http%3A%2F%2Feuropass.cedefo>

and without the basic grammar minimum, it would be made impossible for the student to understand and speak¹⁷.

The grammatical minimum on A1 level for Czech does not mean that the student know e.g. how to create accusative of particular gender but they should be able to order in the restaurant, ask in the shop etc.¹⁸ The priority is the situation in which the student should participate by using simple sentences and phrases.

The topics recommended for this level are following¹⁹: everyday life situations (e.g. shopping), office, hospital, social intercourse, accommodation, public transport, orientation in the city, telephoning.

2.2. The choice of verbs in elementary textbooks

The aspect category lies on the borderland between grammatical category and lexical meaning of the verbs²⁰. The most of Czech verbs exist in pairs which differ in expressing finished action or emphasizing the result of the action (e.g. *udělat, koupit*) and imperfective verbs which express unfinished action or process (e.g. *dělat, koupit*), but the category is far more complex – the verbs can be created by prefixes or suffixes which can change the meaning of the verb as well. That is why the students on A1 usually learn present, past and future tense of imperfective verbs in Czech, the aspect category itself is a part of the syllabus later (levels A2 and B1). Sometimes, the textbooks tend to simplify the language as much as possible and use imperfective verbs even in sentences, where it sounds very unnaturally.

p.europa.eu%2Fen%2Fresources%2FEuropean-language-levels-cefr%2Fcef-ell-

¹⁷ *Cvejnová J. et al.* Metodika přípravy ke zkoušce z českého jazyka pro žadatele o trvalý pobyt (úroveň A1), 2010. Accessible at: http://cestina-pro-cizince.cz/uploads/Dokumenty/cest_pro_ciz_final.pdf

¹⁸ *Hádková M., Líněk J., Vlasáková K.* Čeština jako cizí jazyk. Úroveň A1, 2005. Accessible at: <http://www.msmt.cz/mezinarodni-vztahy/referencni-urovne-pro-cestinu-jako-cizi-jazyk>, P. 3.

¹⁹ *Ibidem*, P. 6.

²⁰ *Cvrček V. et al.* Mluvnice současné češtiny. Praha, 2010. P. 245.

In order to examine which verbs are used in the elementary textbooks, 7 contemporary textbooks of Czech for foreigners were used to create a small corpus²¹. They were scanned, converted from pdf to text files and then cleaned. The final corpus (UcKo-A1) consists of almost 62 thousand words. It consists of all the texts, dialogs (including gap-filling texts) and model sentences used for presenting grammatical features or lexical meaning (usually in a form of small dialogs). Afterwards, the corpus was compiled by using Sketch Engine²² and the list of lemmas (frequency more than 5 tokens) was extracted and then the verbs sorted manually.

There were 124 verbs and only 21 (17%) from them were perfective. If we look closer on the concordances, we can see that most of the perfective verbs are presented only in one or two forms, mostly as a part of a phrase. Some of the verbs are presented only in imperative:

- 1) prominout (36 tokens, 4 books) – promiňte (to excuse)
Aha. < Promiňte >, jak se to píše?
- 2) posadit se (7 tokens, 4 books) – posad'te se (to sit down)
Pojd'te dál a < posad'te > se.

The most of the verbs are presented only in infinitive, usually after a modal verb:

- 3) opravit (9 tokens, 3 books) (to repair)
Dobrý den, potřebuju < opravit > boty.
- 4) pomoci (6 tokens, 4 books) (to help)
Můžete mi < pomoci > prosím vás?

²¹ The criteria for choosing the textbooks were: 1. Published after 2000, 2. A1 level only. The textbooks used for creating this corpus are: *Adamovičová A., Ivanovová D.* Basic Czech I. Praha, 2007; *Cvejnová J.* Česky, prosím. Praha, 2011; *Hádková M.* Čeština pro cizince a azylanty A1. Brno, 2005; *Holá L.* Čeština Express 1, 2. Praha: 2010; *Matula O.* Český den. Praha, 2007; *Štindl O.* Easy Czech Elementary. Praha: 2008; *Štindlová B.* Česky v Česku 1, 2. Praha, 2008.

²² <http://sketchengine.co.uk/>

- 5) přijít (43 tokens, 2 books) (to come)
Nemůžete < přijít > zítra?

Some of the verbs occur only in the past tense:

- 6) narodit se (17 tokens, 6 books) (to be born) – narodil/a se
Jan se < narodil > 13. 5. 1980.
- 7) umřít (5 tokens, 2 books) – umřel/a (to die)
Němcová < umřela > v roce 1862.

Some of the verbs are presented as a part of a particular phrase which the students memorize:

- 8) vzít (8 tokens, 5 books) – vezmu si (to take)
< Vezmu > si čtyři johlíky.
- 9) sejít se (10 tokens, 5 books) – sejdeme se (to meet)
Dobře, < sejdeme > se tam v 7.10.

The results show that the students on A1 level are usually not familiar with aspect category yet and do not know how to create perfective verbs but they already use some of them in particular forms in phrases. What is surprising, that some of the verbs (e.g. *přijít – to come*) were included in 2 or 3 books only, which means the authors tried to avoid using perfective forms in order to simplify the language. Our question is: Why to avoid the perfective verbs when it is enough to know one verb form only?

2.3. The choice of verbs in the A1 level CEFR description

The description of A1 level was written in 2005 while all the books from this corpus later (2005–2010) and thus should be informed of this document. The descriptor includes a list of vocabulary with a reference to a particular topic – the compulsory vocabulary is marked, the rest is considered as recommended for passive knowledge. 162 verbs were sorted from the compulsory list and 51 (31%) of them were perfective, which is much more than in the textbooks (see Fig. 1). Did the authors of the textbooks use the description? Since this corpus includes the texts and dialogs from the textbooks, it was

expected there would be more perfective verbs – the passive knowledge or guessing the meaning of the word from the context belong to reading activities. The reason could be the fact that some of the verbs in the descriptor are not suitable for this level, e.g. *proclít* (clear customs).

2.4. The frequency of selected verbs in Czech National Corpus

Two verbs from the A1 level description (*proclít, zamluvit*) and two verbs from the textbook corpus (*přijít, říct*) were searched in the SYN2010 – part of CNC, representative written corpus of contemporary Czech language (100 mil words).²³

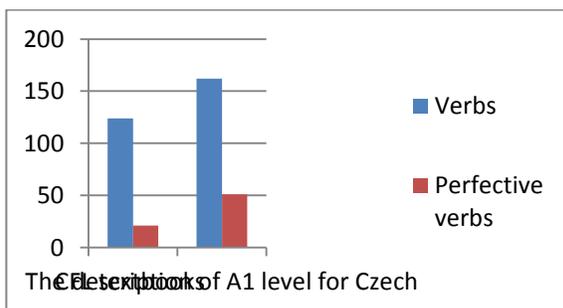


Fig. 1. The verbs in CFL textbooks and in the description of A1 level for Czech

Proclít has only 16 tokens and the dialog in the customs is not a part of the A1 syllabus – that is why it is considered as unsuitable for this level.

The verb *zamluvit* can be substituted with *rezervovat*, which is not only much more frequent according to the corpus data²⁴ (Fig. 2) but also international and easy to understand.

²³ Bonito interface was used for searching in CNC: <http://korpus.cz/bonito/>

²⁴ SyD (an interface for searching variants) was used: <http://syd.korpus.cz/>

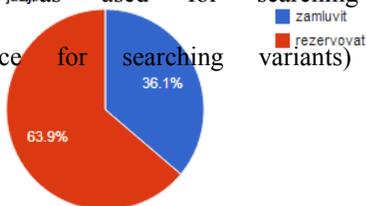


Fig. 2. The verbs *zamluvit* and *rezervovat* (to reserve) in corpus SYN2010 (SyD interface)

The verb *přijít* belongs to the most common verbs (almost 79.580 tokens – 16th most frequent verb in SYN2010) but was included only in two books from our textbook corpus. The most common verb forms are past tense: *přišel* (17.754 tokens) and perfective future tense: *přijde* (11.366). There is no reason in avoiding this verb when the students on this level already know how to conjugate the verb *jít* (to go), so it could be easy to just add the prefix *při-*.

The verb *říct* is included in every textbook in the corpus but the past form only in one of them (*Česky v Česku II*). If we look at the frequency of particular forms in SYN2010, it is visible that the past form *řekl/a* is the most frequent form (Fig. 3). The reason why it is missing is that the textbook usually present short dialogs, not a continuous text. Other reason could be the fact that this verb is irregular and all the forms should be memorized. The corpus helps us to choose the most frequent form (*řekl/a*) to present on elementary level.

Slovní tvary

[1] frekvence	psaný	mluvený
celkem	211753	0
řekl	98066	0
řekla	35114	0
říct	18857	0
říci	9924	0
řečeno	7444	0
Řekl	5876	0
řekne	5186	0
řekli	3845	0
neřekl	3220	0
řeknu	2841	0
Řekla	2617	0

Fig. 3. The frequency of word forms of the verb *říct* in corpus SYN2010 (SyD interface)

3. Conclusion

The corpus-driven analysis of the choice of verbs in CFL textbook corpus demonstrated that the textbooks on A1 level tend to avoid perfective verbs as much as possible. It was surprising that there were discrepancies in the description of A1 level for Czech according to CEFR and the vocabulary in CFL textbooks, even though all the authors of the textbooks recognize their books as A1 level according to the CEFR and one of the authors is even the author of the descriptor. Many perfective verbs are presented in the textbooks merely in one particular form to memorize (infinitive, imperative or past tense). Therefore, national corpus could be an excellent assistant in choosing the most frequent form and the suitable context.

It is argued that corpus evidence should be taken into account in order to achieve higher degree of naturalness in textbook language²⁵. The data from CFL textbook corpus prove that even elementary level can benefit from the national corpus in the choice of verbs. For further research would be useful to examine also the other parts of the textbooks (exercises and listening), create a list of verbs for each topic included in A1 level description using frequency data from national corpus, analyse the modal verbs in the textbooks and in the corpus or use the corpus as a source of data for authentic examples. This type of specialised corpus thus helps to improve the teaching materials to present contemporary and authentic language usage.

²⁵ *Bernardini S.* Corpora in the classroom. An overview and some reflections on future developments. In *J. Sinclair.* How to Use Corpora in Language Teaching. Amsterdam: 2004. P. 17–36; *Römer U.* Corpora and language teaching. In *A. Lüdeling, M. Kytö.* Corpus Linguistics: an international handbook. Berlin, 2008.