

# Učebnicový korpus a jeho využití pro výuku češtiny jako cizího jazyka

*Pavčina Vališová  
Ústav českého jazyka FF MU  
pauline@mail.muni.cz*

**ABSTRACT:** The aim of this paper is to present a project of development a small specialised corpus. The corpus consists of 17 textbooks of Czech language as a foreign language and can be used to research textbook language in comparison with the authentic language in national corpora, the terminology and instructions used in the books, the choice of vocabulary for different levels etc.

## 1 Korpusy ve výuce

Cílem tohoto příspěvku je představit projekt výstavby malého specializovaného korpusu, čímž je korpus učebnic češtiny cizího jazyka. Před představením konkrétních rysů korpusu a jeho možností využití bychom chtěli uvést základní přehled aplikace korpusové lingvistiky do jazykové pedagogiky.

### 1.1 Přímé využití korpusů

Využití korpusů v jazykové výuce se tradičně dělí na dvě linie: přímou a nepřímou (U obou možností můžeme využívat, jak národní korpus, tak i specializované korpusy pro pedagogické účely). Přímé využití, někdy nazývané Data-driven learning, spočívá ve vyhledávání v korpusových nástrojích studenty cílového jazyka a odhalování gramatických pravidel či významů slov díky kontextu, popř. v jeho tzv. slabé variantě, používání výukových aktivit obsahujících vybrané konkordanční řádky učitelem (Johns, 1991; Bernardini, 2004; Gabrielatos, 2005; Römerová, 2008; Flowedew, 2009). Mezi výhody DDL je využití autentického reálného jazyka, induktivní přístup a samostatnost studentů. Mezi nevýhody patří nejen náročnost vyhledávání, ale i technické obtíže, stejně jako neznalost korpusu. Tato metoda bývá většinou omezena na univerzitní prostředí a na angličtinu, i když v současnosti přibývá studií ze středního nebo základního školství (Braunová, 2007; Thompson – Sealey, 2004) i experimenty u dalších jazyků včetně češtiny (Osolsobě, 2010; Vališová – Osolsobě, 2012), a to zvláště díky novým uživatelsky přátelským rozhraním korpusových vyhledávačů. DDL je však stále spíše výjimečně používaný doplňkový zdroj dat.

Kromě techniky zde bývá za podstatnou překážkou označováno také velké množství dat v korpusu, když zahrnuje i málo frekventované či substandardní tvary a rovněž i specifická témata textů, zvláště odborných či beletrie, jež se nehodí pro jazykovou výuku nerodilých mluvčích. Zde se nabízí dvě cesty: jak už bylo zmíněno, tzv. slabá varianta DDL, jež spočívá ve výběru či úpravě konkordancí učitelem a přípravě tištěných cvičení; druhou možností je pak vytvoření malého specializovaného korpusu pro pedagogické účely, jehož texty budou tematicky ovlivněny cílovou skupinou studentů-cizinců (Braunová, 2005). Takový korpus patří mezi žánrové korpusy a studentům umožňuje učit se typické výrazy nebo obraty pro jednotlivá témata nebo situace, jež jsou součástí komunikační kompetence cílového jazyka. Na rozdíl od velkého korpusu zde nedochází k dekontextualizaci, tj. text není vyjmut z kontextu, pro který byl určen. Texty nebo rozhovory jsou autentické, ale jejich témata jsou relevantní z hlediska výuky jazyka.

K příkladům takového korpusu uvedeme dva projekty univerzity v Tübingenu zaměřené na výuku angličtiny (Braunová, 2010). ELISA<sup>1</sup> zahrnuje 25 rozhovorů s rodilými mluvčími v délce od 5 do 15 minut. Korpus tedy obsahuje videa, jejich přepisy i vyhledávač konkordancí. Rodilí mluvčí mluví různými variantami angličtiny – pochází z Ameriky, Austrálie, Skotska nebo Irska, a mluví o svém životě. Škála témat je tím samozřejmě omezená. Celkem pokrývají 11 témat, jako například: náš kraj, výuka a vzdělávání, práce, plány do budoucna apod. Autoři vychází z předpokladu, že konkordanční řádky a seznamy slov mají smysl pouze, pokud jsou zapojeny do vhodných aktivit. Proto každé interview zahrnuje cvičení využívající v něm obsaženou slovní zásobu, a tak mohou studenti s rozhovorem pracovat detailněji. Výhodou je, že slyšeli celé interview, proto znají širší kontext vět obsažených ve cvičeních.

Druhý projekt zmíníme jen krátce, neboť jsme chtěli hlavně poukázat na tento typ pedagogického korpusu a jeho využití. Projekt Backbone<sup>2</sup> na téže univerzitě rovněž zahrnuje videorozhovory, tentokrát však pro celkem 6 jazyků: angličtinu (britskou a mezinárodní angličtinu jako lingua franca), němčinu, francouzštinu španělštinu, polštinu a turečtinu. Tento korpus je na rozdíl od projektu ELISA sponzorován evropskými fondy, a proto je přístupný zdarma. Jeho využití je momentálně testováno na středních i vysokých školách.

## 1.2 Nepřímé využití korpusů

Nepřímým využitím korpusu myslíme využívání korpusů lingvisty, učiteli nebo autory učebnic pro tvorbu výukových materiálů. Korpus je tedy pro ně zdrojem dat, a to jak kvantitativních, tak i kvalitativních. Frekvenční data jsou spolu tématem a budováním celkové komunikační kompetence kritérii pro rozhodování, v jaké fázi výuky jaký gramatický jev předkládat. Dále je možné z korpusu získávat typická slovní spojení, fráze či modelové věty (ať již autentické či upravené) pro tvorbu cvičení (Boulton, 2012; Frankenberg-Garcia, 2012).

K nepřímému využití korpusů ve výuce patří i specializovaný korpus vytvořený z učebnic. Ten sice neslouží jako zdroj dat pro tvorbu cvičení, ale jako soubor všech slov a frází, se kterými student přišel do styku. Samozřejmě, že tento korpus je limitovaný, neboť neobsahuje obrazový materiál učebnic. Hlavně však nemůže zahrnovat interakci ve třídě, ať již učitele a studenta nebo studentů mezi sebou. Přesto však může být přínosný, pokud data z pedagogického korpusu porovnáme s daty z korpusu národního (Gouvenurová – Meunierová, 2009).

## 2 Analýza učebnic a učebnicový korpus

Tento korpus nahrazuje a automatizuje klasickou analýzu učebnic, při které postupujeme stránku za stránkou. Během posledních dvaceti let, kdy se obor čeština jako cizí jazyk začal rychle rozvíjet a vzniklo mnoho nových učebnic, byla publikována také řada studií opírajících se o analýzu učebnic. Při bližším zkoumání zjistíme, že největší podíl tvoří bakalářské a diplomové práce různých bohemistických pracovišť v ČR, což dokazuje, že se čeština jako cizí jazyk řadí mezi nové disciplíny (viz tabulka č. 1). Za zmínku jistě stojí dvě monografie Milana Hrdličky, které se věnují předložkám (2000) a celé mluvnici (2009). Objevují se však i studie, jež nejsou zaměřené na gramatiku, ale na lexikum obsažené v učebnicích, metajazyk, tj. instrukce cvičení, nebo sociokulturní dovednosti, tedy tradice a společenské normy v ČR, jež se odráží v komunikačních situacích a taktéž i v jazyce.

<sup>1</sup> Viz [http://www.uni-tuebingen.de/elisa/html/elisa\\_info.html#demo](http://www.uni-tuebingen.de/elisa/html/elisa_info.html#demo)

<sup>2</sup> Viz <http://u-002-segsv001.uni-tuebingen.de/backbone/moodle/>

**Tab. č. 1      Přehled analýz učebnic češtiny jako cizího jazyka<sup>3</sup>**

| Oblast výzkum | Autor                     | Téma                              | Počet učebnic | Úroveň | Typ práce        |
|---------------|---------------------------|-----------------------------------|---------------|--------|------------------|
| Gramatika     | Drlíková (2008)           | vid                               | 4             | A1-B2  | bakalářská práce |
|               | Hrdlička (2000)           | předložky                         | 38            | A1-B2  | monografie       |
|               | Hrdlička (2009)           | česká mluvnice                    | 33            | A1-B2  | monografie       |
|               | Lukšija (2010)            | místní předložky                  |               |        | bakalářská práce |
|               | Vališová (2011)           | slovesné třídy                    | 23            | A1-A2  | diplomová práce  |
|               | Bičišťová (2011)          | číslovky                          | -             | -      | bakalářská práce |
|               | Lukšija (2012)            | deklinace substantiv              | 10            | A1-A2  | diplomová práce  |
|               | Kodajková (2012)          | gramatické kategorie slovesa      | -             | -      | diplomová práce  |
| Lexikum       | Kovaříková (2009)         | obecná čeština                    | -             | -      | bakalářská práce |
|               | Nezvalová (2010)          | slovotvorba                       | 3             | A1-A2  | bakalářská práce |
|               | Nováková (2010)           | obecná čeština                    | 4             | B1-B2  | diplomová práce  |
|               | Ocztošová (2012)          | slovní zásoba pro německy mluvící | -             | -      | bakalářská práce |
| Metoda        | Hrdlička (2009)           | komunikační metoda                | 33            | A1-B2  | monografie       |
|               | Hladíková (2011)          | signální gramatika                | 4             | A1     | diplomová práce  |
| Metajazyk     | Hradilová (2010)          | terminologie jmenné flexe         | 3             | A1-A2  | odborný článek   |
|               | Osolobě – Vališová (2010) | lingvistická terminologie         | 17            | A1-A2  | odborný článek   |
| Text          | Volnogradská (2012)       | typy textů                        | 3             | A1     | diplomová práce  |

Podívejme se nyní na to, jak tuto analýzu zautomatizovat. Chtěli bychom stručně představit již existující korpusy učebnic anglického jazyka. Úplně prvním korpusem byl The TOEFL 2000 First Academic Spoken and Written Corpus z roku 2002, což je kolekce americké angličtiny (celkem 2,7 mil), z čehož necelých 800 tis slov pochází z učebnic. Tento korpus však sloužil i jiným účelům než pouze pro analýzu učebnic.

Druhým počinem na tomto poli je německý German English as a Foreign Language Textbook Corpus (GEFL TC) z roku 2004. Je tvořen učebnicemi angličtiny určenými pro německé mluvčí a zahrnuje 100 tis slov. Jeho tvůrkyně Römerová již publikovala i

<sup>3</sup> Uvedené diplomové práce pochází z bohemistických pracovišť MU v Brně, UK v Praze, TUL v Liberci, ZCU v Plzni a UJEP v Ústí nad Labem. Tam, kde jsme neměli možnost nahlédnout, tzn. práce nejsou přístupné on-line, informace o počtu učebnic a jejich úrovni neuvádíme.

několik studií, ve kterých srovnává tento učebnicový jazyk s autentickým jazykem z britských národních korpusů (Römer, 2004, 2005).

Existuje i několik dalších malých korpusů, např. angličtiny pro čínské mluvčí, nyní bychom se však chtěli zaměřit na projekt univerzity v Lovani, neboť ten se liší rozsahem i pojetím. Tím je korpus TeMa<sup>4</sup> z roku 2007. Dříve uvedené korpusy zahrnovali převážně texty nebo dialogy, zatímco korpus TeMa již obsahuje celé učebnice, neboť jeho cílem je výzkum všech částí učebnic. Také jeho rozsah je daleko větší, a to asi 724 tis slov z celkem 32 učebnic obecné angličtiny. Navíc díky tzv. pedagogické anotaci nabízí mnohem větší pole využití (Gouveneurová – Meunierová, 2009).

Korpus TeMa je rozdělen na 4 subkorpusy: texty, přepisy poslechů, cvičení a instrukce. U každé učebnice je vždy uveden zdroj, a to konkrétně v tomto pořadí: 1. název učebnice, 2. jestli je to učebnice nebo cvičebnice, 3. lekce, strana, cvičení a nakonec 4. typy cvičení. Pro čtvrtou část korpusu autoři vytvořili speciální značky pro pedagogickou anotaci, tzn. značkují každé cvičení podle jeho typu, např. doplňování (slova z výběru, slova ve správném tvaru, libovolného slova), spojování apod. Při doplňování je samozřejmě označeno i správná forma slova, která se má doplnit.

## 2.1 Český učebnicový korpus

V rámci své doktorské práce plánujeme vytvořit malý specializovaný korpus z učebnic češtiny jako cizího jazyka a korpus TeMa považujeme za jeho vzor. Protože korpus ještě není dokončen, chtěli bychom Vás seznámit se složením tohoto korpusu a poté také s různými možnostmi využití tohoto typu korpusu.

Učebnicový korpus by měl být složen ze 17 současných učebnic češtiny pro cizince (viz tabulka č. 2) – všechny byly vydány po roce 2000, některé jsou však staršího data, neboť se jedná o další vydání, což ukazuje, že učebnice se stále využívá pro výuku. Úroveň učebnic jsme zvolili A1 až B1 podle Evropského referenčního rámce pro jazyky, a to z toho důvodu, že úroveň B1 je označována za tzv. prahovou, ale také z praktického hlediska – pro tyto úrovně máme nejvíce učebnic, pro úrovně vyšší již málo, proto by nebylo možné srovnání. Celkový počet učebnic je ve skutečnosti o něco málo vyšší, uvedla jsem 17, protože, pokud má učebnice více dílů, ale stejného autora či autory, označujeme ji jako jednu učebnici. Výběr učebnic jsme uskutečnili nejen podle data vydání a úrovně, ale i na základě toho, které učebnice učitelé často využívají (k tomu jsme využili studie Koláčkové (2010) i vlastních zkušeností z praxe), ale i faktu, aby byly zahrnuty učebnice pro různé cílové skupiny studentů, např. anglicky mluvících, Slovanů, azylantů apod.

**Tab. č. 2 Seznam učebnic pro plánovaný učebnicový korpusu**

|   |   |
|---|---|
| 1 | Adamovičová, A. – Ivanovová, D. (2007): <i>Basic Czech I, II</i> . Praha: Karolinum.        |
| 2 | Bischofová, J. et al. (2007): <i>Čeština pro cizince a azylanty B1</i> . Brno: SOZE.        |
| 3 | Bořilová, P. – Holá, L. (2011): <i>Česky krok za krokem 2</i> . Praha: Akropolis.           |
| 4 | Cvejnová, J. (2011): <i>Česky, prosím</i> . Praha.  |
| 5 | Čechová, E. – Remediosová, H. (2005): <i>Chcete mluvit česky? 1, 2</i> Liberec: Harry Putz. |
| 6 | Froulíková, L. (2008): <i>Adam a Eva v Českém ráji</i> . Praha: Academia.                   |
| 7 | Hádková, M. (2005): <i>Čeština pro cizince a azylanty A1, A2</i> . Brno: SOZE.              |
| 8 | Holá, L. (2006): <i>New Czech Step by Step</i> . Praha: Akropolis.                          |
| 9 | Holá, L. (2010): <i>Čeština Express 1, 2</i> . Praha: Akropolis.                            |

<sup>4</sup> Viz <http://www.uclouvain.be/en-cecl-tema.html>

|    |  |
|----|--|
| 10 | Kestřánková, M. et al. (2010): <i>Čeština pro cizince B1</i> . Brno: Cpress.   |
| 11 | Matula, O. (2007): <i>Český den. Kurz českého jazyka pro azylanty navazující na Manuál pro učitele českého jazyka pro cizince bez znalosti latinky</i> . Praha: Člověk v tísní o.p.s., Projekt Varianty. |
| 12 | Nekovářová, A. (2006): <i>Čeština pro život - 15 moderních konverzačních témat</i> . Praha: Akropolis.   |
| 13 | Parolková, O. (2004): <i>Czech for foreigner s 1, 2</i> . Praha: Bohemika.   |
| 14 | Pintarová, M. – Režková, I. (1995): <i>Communicative Czech. Elementary Czech. Intermediate Czech</i> . Praha: Univerzita Karlova.  |
| 15 | Rigerová, K. (2000): <i>Czech for Everyone</i> . Praha: K. Rigerová.   |
| 16 | Štindl, O. (2008): <i>Easy Czech. Elementary</i> . Praha: Akronym.   |
| 17 | Štindlová, B. (2008): <i>Česky v Česku 1, 2</i> . Praha: Ústav jazykové a odborné přípravy UK: Akropolis.  |

Při samotné tvorbě korpusu je tedy nejdříve nutné učebnice oskenovat, následné soubory pdf zkonvertovat do textového souboru (použili jsme program zabudovaný do Informačního systému Masarykovy univerzity) a poté vyčistit, tj. zbavit se obrazového materiálu a ručně opravit chyby, které při konverzi vznikly, např. když bylo některé písmeno nebo celé slovo rozeznáno špatně. V této fázi se momentálně tvorba korpusu nachází. Poté plánujeme využít nástroj WebBootCat vyhledávače Sketch Engine<sup>5</sup>, který umožňuje vložení vlastního korpusu ve formátu txt. Následně je možné vytvořit seznam nejfrekventovanějších (wordlist), ve srovnání s ČNK vygenerovat klíčová slova nebo vyhledávat slova a jejich kontexty jako v klasickém korpusu. Za použití jednoho z volně přístupných anotátorů bychom chtěli rozdělit korpus na 4 subkorpusy dle korpusu TeMa: texty, poslechy, instrukce a cvičení. Také bychom chtěli vytvořit subkorpusy podle úrovně SERRJ (A1, A2, B1)<sup>6</sup>. Záměrem je využití tohoto korpusu pro výzkum slovní zásoby s ohledem na jednotlivé úrovně SERRJ a s porovnáním s autentickým jazykem v ČNK.

Jaké možné problémy mohou nastat? Již při myšlence rozdělení knih podle SERRJ vzniká u některých učebnic překážka. Často jsou totiž orientovány na výuku více úrovní, např. A1 až A2, nebo dokonce A1 až B1. Je tedy obtížné někdy rozhodnout, ve které části učebnice určit tuto hranici mezi úrovněmi. Nejlépe na tom jsou v tomto ohledu učebnice nejnižší úrovně A1, zřejmě i kvůli tomu, že na této úrovni cizinci skládají Zkoušku z češtiny pro trvalý pobyt. Učebnic úrovně A1 je proto největší množství a často jsou autory cíleně omezeny na tuto nejnižší úroveň.

Další obtíže nastávají při tzv. čištění textového souboru, neboť v učebnicích nejsou čisté texty, ale často tabulky, věty nebo pouze slova či hlásky. Dále se tam vyskytují instrukce v různých mediačních jazycích, nejčastěji v angličtině. A nakonec učebnice jsou často barevné a obsahují obrázky, fotografiemi a grafikou, proto některé věty, např. ty, jež jsou otištěny světlým písmem na tmavém pozadí, nejsou programem vůbec rozeznány. Z tohoto důvodu je tvorba korpusu, přestože malého formátu, velmi časově náročná.

## 2.2 Využití učebnicového korpusu

Protože ještě není možné předložit konkrétní data na základě učebnicového korpusu, chtěli bychom uvést různé směry, jakými je možné se ubírat při využívání tohoto typu korpusu.

<sup>5</sup> <http://sketchengine.co.uk/>

<sup>6</sup> SERRJ – Společný evropský referenční rámec pro jazyky, viz <http://www.msmt.cz/mezinarodni-vztahy/spolecny-evropsky-referencni-ramec-pro-jazyky>

Dělíme je na dvě hlavní oblasti, a to: 1. Deskriptivní – popisují učebnicový jazyk jako specifický žánr a 2. Didaktickou oblast – zkoumá typy cvičení.

Mezi deskriptivní způsoby využití patří:

1. Jak jsme již zmínili, výzkum slovní zásoby a komparace s autentickým územ v národních korpusech (Römerová, 2004, 2005; Gouveneurová – Meunierová, 2009).
2. Učebnicový korpus může být též využit pro studium osvojování jazyka, neboť můžeme vyhledat všechny kontexty, ve kterých se slovo či fráze vyskytují v učebnicích různých úrovní (Hustonová, 2010).
3. Pedagogické značkování však umožňuje i zkoumání učebnic z dalších perspektiv. Při rozdělení na subkorpusey podle úrovní lze zkoumat výběr slovní zásoby pro jednotlivé úrovně. Slovní zásobu specifickou pro konkrétní úroveň lze porovnat s popisem Rámce.
4. Lze porovnat input v podobě textů, poslechů i cvičení, tzn. jaká slovní zásoba se vyskytuje v té které části, a zda se od sebe liší.
5. Porovnáním kolokací slov ve cvičeních lze objevit rozdíly mezi učebnicemi. (Gouvernerová, 2008 v pilotní studii dle korpusu TeMa např. objevila v tomto směru mezi učebnicemi velké rozdíly).
6. Také lze zkoumat metajazyk, jenž učebnice využívají, a to nejen pro instrukce ke cvičením (Používají obecné výrazy nebo konkrétní? Imperativ či jiný tvar?), ale i při výkladu gramatiky. Zde se však často místo češtiny objevuje zprostředkovací jazyk. Rovněž důslednost ve využívání lingvistické terminologie je důležité téma pro výzkum. V naší studii z roku 2010 jsem společně s doc. Osolsobě zkoumala lingvistickou terminologii v 17 učebnicích češtiny pro cizince a objevili jsme velkou nekonzistentnost mezi českými termíny a mezinárodními výrazy, a to i v rámci jedné učebnice.
7. Další oblastí výzkumu je též typ textu, tzn. podíl psané a mluvené komunikace v učebnicích a různých jazykových variant, např. zastoupení spisovné/hovorové/obecné češtiny v učebnicích či v jednotlivých částech jedné učebnice.
8. Při možnosti automatického morfologického značkování, lze zkoumat i další jevy spojené s výukou gramatiky.

Z didaktického hlediska je možno zkoumat typy cvičení podle tématu, jevu nebo referenční úrovně. (Např. Gouvernerová ve výše zmíněné pilotní studii objevila, že typy cvičení se velmi liší dle úrovní, a také poukázala na nedostatek pozorovacích aktivit, které podle výzkumu v oblasti osvojování druhého jazyka hrají velmi důležitou roli). Nakonec je možné navrhnout zlepšení díky výsledkům současného výzkumu osvojování druhého jazyka i za pomoci současných korpusů rodilých i nerodilých mluvčích, tj. národních korpusů nebo korpusů žakovských, které se v současné době budují i pro češtinu jako cizí jazyk.

### **3 Závěr**

Na závěr bychom chtěli shrnout, že tento typ korpusu nám otevírá nové oblasti výzkumu učebnicového jazyka i didaktických metod, je však nutné mít na paměti, že učebnicový korpus je i přes svoje výhody velmi limitovaný tím, že učebnice redukuje pouze na text a nelze zahrnout interakci ve třídě, která je nedílnou součástí výuky jazyka. Učebnicový korpus také nabízí komparaci s autentickým jazykem národních korpusů, je však nutné vzít v potaz, že přestože studenti mohou těžit z autentického jazykového inputu, adaptované či vymyšlené texty jsou také nedílnou součástí učebnic, zvláště pro nižší pokročilosti. I přesto se však domnívám, že jeho výhody nad těmito limity převažují.

## REFERENCE

- BERNARDINI, S. (2004): Corpora in the classroom. An overview and some reflections on future developments. In: J. Sinclair. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, s. 17–36.
- BOULTON, A. (2012): Hands-on / hands-off: Alternative approaches to data driven learning. In: J. Thomas – A. Boulton (eds.). *Input, Process and Product. Development in Teaching and Language Corpora*. Brno: Masaryk University Press, s. 152–168.
- BRAUN, S. (2005): From pedagogically relevant corpora to authentic language learning contents. In: *ReCALL*, roč. 17, č. 1, s. 47–64.
- BRAUN, S. (2007): Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. In: *ReCALL*, roč. 19, č. 3, s. 307–328.
- BRAUN, S. (2010): ELISA – a pedagogically enriched corpus for language learning purposes. In *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Lang, Frankfurt/M, 25–47. Dostupné z WWW: <<http://epubs.surrey.ac.uk/1292/>>
- FLOWERDEW, L. (2009): Applying corpus linguistics into pedagogy. A critical evaluation. In: *International Journal of Corpus Linguistics*, roč. 14, č. 3, s. 393–417.
- FRANKENBERG-GARCIA, A. (2012): Integrating corpora with everyday language teaching. In: J. Thomas – A. Boulton (eds.). *Input, Process and Product. Development in Teaching and Language Corpora*. Brno: Masaryk University Press, s. 36–53.
- GABRIELATOS, C. (2005): Corpora and language teaching: Just a fling, or wedding bells? In: *TESL-EJ*, roč. 8, č. 4, A1, s. 1–37.
- GOUVERNEUR C. (2008): The phraseological patterns of high-frequency verbs in advanced English for General Purposes: a corpus-driven approach to EFL textbook analysis, In: F. Meunier – S. Granger (eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam & Philadelphia: Benjamins, s. 223–246.
- GOUVERNEUR, C. – MEUNIER, F. (2009): New types of corpora for new educational challenges. Collecting, annotating and exploiting a corpus of textbook material. In: K. Aijmer (ed.) *Corpora and Language Teaching*. Amsterdam: John Benjamins.
- HRDLIČKA, M. (2000): *Předložky ve výuce češtiny jako cizího jazyka*. Praha. Karolinum.
- HRDLIČKA, M. (2010): *Gramatika a výuka češtiny jako cizího jazyka*. Praha. Karolinum.
- HUSTON, S. (2010): *Corpora in Applied linguistics*. Cambridge: Cambridge University Press.
- JOHNS, T. (1991): Should you be persuaded. Two samples of data-driven learning materials. In: T. Johns – P. King (eds). *Classroom Concordancing: ELR Journal*, roč. 4, s. 1–16.
- KOLÁČKOVÁ, L. (2010): Hlediska, podle nichž učitelé ČCJ vybírají učebnice češtiny pro cizince. In: K. Hlínová (ed.). *Sborník Asociace učitelů češtiny jako cizího jazyka (AUCČJ) 2010*. Praha: Akropolis, s. 95–100.
- OSOLSOBĚ, K. (2010): Jak se učit česky s korpusem. In: *Přednášky a besedy ze XLIII. běhu Letní školy slovanských (bohemistických) studií*. Brno: Masarykova univerzita, s. 112–119.
- OSOLSOBĚ, K. – VALIŠOVÁ, P. (2011): Tagset korpusů ČNK z hlediska předpokládané znalosti gramatické terminologie u nerodilých mluvčích (Možnosti a meze využívání korpusů češtiny pro nerodilé mluvčí). In: I. Dominiková – m M. Lachout (eds.). *Lingua terminologica*. Praha: MUP, s. 141–156.

OSOLSOBĚ, K. – VALIŠOVÁ, P. (2012): Using data-driven method in teaching Czech as a foreign language. In J. Thomas – A. Boulton (eds.). *Input, Process and Product. Development in Teaching and Language Corpora*. Brno: Masaryk University Press, s. 183–194.

RÖMER, U. (2004): A corpus-driven approach to modal auxiliaries and their didactics. In: J. Sinclair (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 2004, s. 185–199.

RÖMER, U. (2005): 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: Rodopi.

RÖMER, U. (2008): Corpora and language teaching. In: A. Lüdeling – M. Kytö. *Corpus Linguistics: an international handbook*. Berlin: Walter de Gruyter GmbH & Co.

THOMPSON, P. – SEALEY, A. (2004): „What do you call the dull words?“ Primary school children using corpus-based approaches to learn about language. In: *English in Education*, roč. 38, č. 1, s. 80–91.