

Obsah / Content

Články / Articles

**Srovnání žánrů v korpusu na základě syntaktických funkcí substantiv
Comparison of genres in corpora on the basis of syntactic functions of
substantives /3**

Tomáš Jelínek

**Standarddeutsch messen? Frequenz und Varianz negativ-
-konditionaler Konnektoren**

**Jak měřit standardní němčinu? Frekvence a variantnost negativně podmiňovacích
konektorů /14**

Marek Konopka, Ulrich H. Waßner

Liší se mluvené a psané texty ve valenci?

Valency in spoken and written texts - is there a difference? /37

Marie Mikulová, Jan Štěpánek, Zdeňka Urešová

Jazyková potencialita: studium na bázi hapaxů legomenon

Language potentiality: A study on the basis of hapax legomena /48

Renata Novotná

**Konstruktionen mit formalem Objekt im Deutschen und ihre Entsprechungen im
Tschechischen**

Konstrukce s formálním objektem v němčině a jejich protějšky v češtině /59

Hana Peloušková

Recenze / Reviews

**Hladká, Z. – Martincová, O.: Slova v soukromých dopisech. Lexikografická sonda.
Brno: MU 2012.**

**Hladká, Z. a kol.: Soukromá korespondence jako lingvistický pramen. Brno: MU
2013. /71**

Ivana Bozděchová

Zprávy / News

Žánr epistolární v korpusu /76

František Čermák

Gramatika a korpus 2012: 4. mezinárodní konference, 28.-30. listopadu 2012 /78

Kateřina Najbrtová

CzeSL - Korpus češtiny nerodilých mluvčích

CzeSL - a Corpus of Non-Native Speakers' Czech /82

Svatava Škodová

Co je v ČNK nového IV /90

Lucie Benešová

Pokyny pro autory /93

Instructions for authors /95

Objednávka /97

Order form /98

Srovnání žánrů v korpusu na základě syntaktických funkcí substantiv

Tomáš Jelínek

Ústav teoretické a počítačové lingvistiky FF UK Praha

tomas.jelinek@ff.cuni.cz

Comparison of genres in corpora on the basis of syntactic functions of substantives

ABSTRACT: Large synchronic textual corpora of the Czech National Corpus are built as representative: they contain a balanced quantity of texts of various styles, divided into three genre subcorpora: fiction, technical/scientific literature and journalism. Comparisons of these genres have been performed on phonological and morphological level; in this paper, I deal with differences between genres on the surface-syntactic level. I use an automatic syntactic annotation of the SYN2005 corpus in the formalism of the analytical layer of the Prague Dependency Treebank. I compare the frequencies of syntactic functions of nouns in the three genres represented by the corresponding subcorpora of SYN2005. I also present a more detailed analysis of four syntactic phenomena: subtypes of the function of attribute in non-prepositional genitive; frequencies of groups of the type *pan Novák* (Mr. Novák); frequencies of the function of agent in passive constructions expressed by nouns in non-prepositional instrumental and the ratio of the expression of the nominal part of a verbal-nominal predicate by nominative and instrumental.

Significant differences found between genres in all the syntactic phenomena analyzed show that in comparing corpora one should carefully monitor their genre composition.

KLÍČOVÁ SLOVA: syntax, syntaktická funkce, korpus, žánr, reprezentativnost

KEY WORDS: syntax, syntactic function, corpus, genre, representativeness

Standarddeutsch messen? Frequenz und Varianz negativ-konditionaler Konnektoren

Marek Konopka, Ulrich H. Waßner

unter Mitwirkung von Sandra Hansen¹

Institut für Deutsche Sprache (IDS), Mannheim

konopka@ids-mannheim.de, wassner@ids-mannheim.de

Measuring Standard German? Frequency and Variance of Negative-Conditional Connectors

ABSTRACT: Among the German negative-conditional connectors in the range of *consequens* markers there are the prototypical cases *sonst* and *ansonsten*. Morphological alternatives (*sonsten* and *ansonst*) are rarely mentioned in contemporary grammars and dictionaries but they actually occur with considerable frequency. The four connectors are used in two functions: as a conjunctive adverb which can occupy various positions within the sentence or as a specific kind of subordinating conjunction (*Postponierer*). The large IDS corpora allow us to reveal specific distributions of the lexemes and of their different ways of use. Comparing the frequencies and the distributions can indicate to which extent the phenomena are part of the standard language. The paper will report on the results and demonstrate how the findings can be deduced from the corpora. It will draw conclusions for assessing the acceptability of the variants and the extent to which they can be considered standard language additionally testing statistical instruments to visualise and calculate the variance of phenomena as association plots and DP_{norm} .

KEY WORDS: corpus linguistics, grammar, standard German, connectors, conjunction, adverbials, variation

Jak měřit standardní němčinu? Frekvence a variantnost negativně podmiňovacích konektorů

ABSTRAKT: Prototypickými případy těchto konektorů jsou v němčině slova *sonst* (*jinak*) a *ansonsten* (*jinak*). Jejich morfologické alternativy *sonsten* a *ansonst* se v součas-

¹ Sandra Hansen hat computerlinguistische Aufgaben wie die Extraktion der Daten übernommen und zahlreiche statistische Analysen (vor allem in R) durchgeführt. Wir danken außerdem Saskia Schmadel für die Unterstützung bei mühsamen Korpusrecherchen und -auswertungen, bei der Erstellung von Abbildungen und der Formatierung. *Last not least* sind wir Anja Konopka für die wertvollen Korrekturen verbunden.

ných gramatikách a slovnících téměř nezmiňují, avšak jejich frekvence v textech není zanedbatelná. Všechny tyto čtyři konektory se užívají ve dvou funkcích: jako spojkové adverbium, jež může ve větě zaujímat různé pozice, nebo jako specifický druh subordinální konjunkce (postponent). Velké korpusy IDS umožňují odhalovat specifickou distribuci těchto lexémů a různých způsobů jejich užívání. Srovnávání frekvence a distribuce může indikovat, do jaké míry jsou tyto jevy součástí standardního jazyka. Tato studie přináší výsledky tohoto srovnávání a ukazuje, jak lze korpusové nálezy interpretovat. Prostřednictvím testování statistických nástrojů se vyslovuje k akceptabilitě variant a míře, v níž mohou náležet standardnímu spisovnému jazyku.

KLÍČOVÁ SLOVA: korpusová lingvistika, gramatika, standardní němčina, konektory, spojky, adverbiale, varianta

Liší se mluvené a psané texty ve valenci?

Marie Mikulová, Jan Štěpánek, Zdeňka Uřešová
Ústav formální a aplikované lingvistiky MFF UK Praha
{mikulova, stepanek, uresova}@ufal.mff.cuni.cz

Valency in spoken and written texts – is there a difference?

ABSTRACT: In this article, we introduce the Prague Dependency Treebank of Spoken Czech. The syntactic and semantic annotation of this corpus has led to the expansion of PDT-Vallex, a valency lexicon of Czech verbs, which has previously been linked only to the annotation of written texts. The expansion of the lexicon consisted of several steps: (i) new verbs were added to the lexicon; (ii) new meanings and new valency frames were added to verbs that had already been included in the lexicon; valency frames that had already been part of the lexicon were enriched with (iii) new participants (actants) and (iv) new formal realizations of participants (actants). All the above mentioned enrichments are (a) unmarked and based only on the addition of a new verb, a new meaning, a new participant (actant) or a new form, however, (b) some of them are influenced by the typical characteristics of spoken language. It would be almost impossible to find some of the verbs, some of the meanings, participants or forms in a written text. We believe that verbs in spoken language tend to exhibit different valency behavior than verbs in written texts. In this article we attempt to draw more general conclusions on the valency behavior of verbs in spoken language.

KEY WORDS: Prague Dependency Treebank of Spoken Czech, spoken language, valency lexicon

KLÍČOVÁ SLOVA: Pražský závislostní korpus mluvené češtiny, mluvená čeština, valenční slovník

Jazyková potencialita: studium na bázi hapaxů legomenon

Renata Novotná

Ústav Českého národního korpusu FF UK, Praha

renata.novotna@ff.cuni.cz

Language potentiality: A study on the basis of hapax legomena

ABSTRACT: The paper deals with word forms occurring in the SYN corpus (1.2 billion words) only once, twice or three times, the so-called hapax legomena, which provide a basis for the study of potentiality in language. As the material was very large, only 20 samples were chosen, each containing 3 000 forms, i. e. 60 000 forms overall. Approximately 50 % of word forms were various mistakes, especially typing errors, or words from other languages. Therefore only the remaining 30 000 word forms were selected as the basis for this study. The analysis showed that the most relevant suffixes for hapax legomena are *-ovský* (e. g. *jimmyreedovský*), *-ák* (e. g. *medvěďák*), *-ista* (e. g. *havlista*), *-ing/-ink* (e. g. *gardening*, *dancink*), *-ovitý* (e. g. *kladivovitý*), type *po vojensku*, diminutives derived from abstracts (e.g. *minulůstka*) and names of women professions (e. g. *meduprodavačka*). Moreover, compound words with the first parts *douho-* (e. g. *douhorožec*), *gala-* (e. g. *galamenu*), *jino-* (e. g. *jinomluva*), *kino-* (e. g. *kinofajnšmekr*), *mega-* (e. g. *megakatastrofa*), *nízko-* (e. g. *nízkohlučný*), *polo-* (e. g. *poločíselný*), *video-* (e. g. *videokomentář*) were typical for new words.

KEY WORDS: hapax legomenon, potentiality, suffix, diminutive, feminine noun, compound

KLÍČOVÁ SLOVA: hapax legomenon, potencialita, sufix, deminutivum, přechýlené slovo, kompozitum

Konstruktionen mit formalem Objekt im Deutschen und ihre Entsprechungen im Tschechischen

Hana Peloušková

Pedagogická fakulta Masarykovy Univerzity, Brno

pelouskova@ped.muni.cz

Structures with the formal object in German and their equivalents in Czech

ABSTRACT: This article treats German constructions with *es* in the function of formal object, their structures and equivalents in Czech. It tries to relate German structures with formal object to their Czech equivalents by using the referential properties of the formal object occurring in the German structure. The research is based on parallel language data extracted from the Czech-German Corpora InterCorp and ČNPK.

KEY WORDS: contrastive linguistics, parallel corpus, syntactic structures, pronoun *es*, formal object, idioms, reference, equivalents

Konstrukce s formálním objektem v němčině a jejich protějšky v češtině

ABSTRAKT: Článek se zabývá německými konstrukcemi s *es* ve funkci formálního objektu, jejich strukturami a protějšky v češtině. Pokouší se najít vztahy mezi konkrétními strukturami s formálními objekty, mírou reference formálních objektů a výskytem českých protějšků *es*. Výzkum se opírá o autentická jazyková data vytěžená z česko-němec-kých paralelních korpusů InterCorp a ČNPK.

KLÍČOVÁ SLOVA: kontrastivní lingvistika, paralelní korpus, syntaktické struktury, zájmeno *es*, formální objekt, idiomy, reference, ekvivalenty

Recenze / Reviews

Zdeňka Hladká, Olga Martinčová: Slova v soukromých dopisech. Lexikografická sonda.

Brno: Masarykova univerzita, 2012, 138 s.

Zdeňka Hladká a kol.: Soukromá korespondence jako lingvistický pramen.

Brno: Masarykova univerzita, 2013, 185 s.

Ivana Bozděchová

Filozofická fakulta Univerzity Karlovy, Praha

bozdiaff@ff.cuni.cz

Lexikograficky a korpusově zpracované zdroje moderní češtiny se v posledních několika letech obohatily o neobyčejně pestrý, zajímavý, ba (laicky i lingvisticky) lákavý materiál – o soukromou korespondenci. V souvislosti s rozsáhlým projektem zahrnujícím sběr, analýzy a korpusové zpracování osobních dopisů, který probíhá od 90. let 20. století v Ústavu českého jazyka FF MU v Brně, vznikla unikátní knižní publikace Zdeňky Hladké a Olgy Martinčové *Slova v soukromých dopisech. Lexikografická sonda*. Na základě korpusově uloženého souboru ručně psané korespondence, který je součástí Českého národního korpusu (KSK-dopisy; <http://ucnk.ff.cuni.cz>) zkušené autorky zpracovaly slovník příznakového lexika korespondence mladých lidí přelomu 20. a 21. století. Korpus soukromé korespondence obsahuje 2000 ručně psaných dopisů a 1000 e-mailů pocházejících od 3000 pisatelů. Tento materiál byl s dodatkem 2000 SMS publikován v práci *Čeština v současné soukromé korespondenci. Dopisy, e-mail, SMS* (Hladká a kol. 2005); její součástí jsou 2 CD se záznamem všech textů.

Jak autorky úvodem konstatují, k příznakovému lexiku (v jimi analyzovaném korpusu to jsou různorodé, např. nespisovné prostředky typické pro tzv. běžnou mluvu) lze přistupovat z různých hledisek. Pro své záměry zvolily lexikografický pohled a pokusily se zjistit, jaké možnosti může tento přístup nabízet: vzhledem k tomu, že je to první lexikografické zpracování korespondenčních textů v naší lingvistice, jim šlo především o teoretickou a metodologickou sondu, která by mohla sloužit jako východisko pro další lexikografickou reflexi

těchto textů. Zároveň se snažily ukázat, že dopisy jsou vhodným zdrojem pro studium příznakového lexika současné češtiny.

Lexikografické zpracování vyžaduje stanovit obecně teoretická východiska, tedy především vymezit pojetí příznakovosti, příznakových lexikálních prostředků a jejich typologie, a zásady pojetí příkladové části – jak prezentovat užití slova, jeho lexikálně-sémantické okolí, jak naznačit způsob vyjadřování pisatele a jak dokreslit výklad významu slova. Tomu je věnována první rozsáhlejší teoretická část knihy (*Slovník soukromé korespondence. Dopisy mladých lidí z přelomu 20. a 21. století*: s. 7–24). Za ní následuje vlastní slovník (s. 27–94) a tři dokladové přílohy (*Frazémy a idiomy, Charakteristika dopisů a Ukázky dopisů*, s. 96–131). Soupis analyzovaných dopisů (Příloha 2) poskytuje základní sociolingvisticky relevantní údaje o pisatelích všech dopisů a stručnou informaci o tématech jejich textů.

Slovník soukromé korespondence (SSK) je speciální a diferenční výkladový slovník, abecedně uspořádaný. Rozsahem patří mezi tzv. malé slovníky – obsahuje přes 800 hesel. Jako příznakové lexikální prostředky se v něm chápou prostředky s charakteristickým, rozlišujícím rysem (blíže viz s. 7). Příznakovost zařazených jednotek se zakládá na jejich příslušnosti k některému z útvarů (poloútvářů) národního jazyka, k typu komunikace a k jednotlivým stylům a stylovým vrstvám, na vyjádření citového, hodnotícího a volního postoje mluvčího, zahrnuty jsou i výrazy zastaralé a zastarávající, dobové a příznakové frekvenčně. Pro ilustraci uvádíme několik příkladů:

- výrazy hovorové: *koukání, opíkáci, prozvonit, šéfka, teda, ženská, rošťák, semolit* expr.
- výrazy zastaralé: *lont* (= vesnice, venkov) ob.
- výrazy hanlivé: *fašoun*
- výrazy knižní: *povýšeneckost* poněk. knižní, *rovněž*
- expresiva: *bodanec, blábolení, cvokatět, člobrd'a, dovoška* (= dovolená), *foglaromol, holčenčí, rýpnout si, šlichta, tvrd'ák, vyžahnout, zabíjárna*
- slova obecněčeská: *bezďák* hanl., *časák, háklivý, lahváč, podporák, rehabka, šrotit se, zdravka, kalba, kalit, narozky, pokec* expr.
- slova vulgární: *buzerace, kripl, vyflusaný, žrát*
- slangismy: *cigo, dějzl* (= dějepis), *devet'ák* (= dozorčí útvaru), *fest'ák, normal, pakárna, výučňák, žádovka*
- pragmatická deminutiva: *albíčko, brigádička, čajík, dopísek, džemík, filmeček, jídelko*
- pozdravy: *ahojda, čauik, sbohemky*

Materiálovou základnou slovníku bylo 100 ručně psaných dopisů z KSK-dopisy, kritériem jejich výběru byl věk pisatele do 30 let. Dopisy pocházejí z let

1997–2003, jsou čistě osobní (včetně pouze kontaktních). Kritériem výběru heslových jednotek slovníku je příznakovost lexikálního prostředku, zařazena jsou slova plnovýznamová i neplnovýznamová. Výklad jejich významu sleduje cíl informační a charakterizační, ukazuje význam jednotek jako uzuální i individuálně odstíněný, jako jednoznačný, určitý nebo i neurčitý, vágní, případně v různé míře expresivní. Výklady významu v SSK tedy nejsou vždy nutně identické s výklady těchto jednotek ve výkladových slovnících. Tak se zejména kvantifikátory stylových charakteristik neuvádějí mechanicky podle dosavadních slovníků, ale na základě empiricky zjištěných dat se přihlíží k současnému stavu. Příznakové prostředky jsou značně osobité a diferencované hláskoslovně, tvaroslovně i pravopisně, stanovisko k jejich prezentaci v SSK usiluje o řešení poměru mezi autentičností dopisů a požadavkem srozumitelnosti. Tak se texty např. neupravují graficky ani pravopisně (dopisy s velkou koncentrací pravopisných chyb přitom nejsou vzácné). Frazémy a idiomy jsou zpracovány samostatně v Příloze 1 s cílem pouze registračním (zaznamenáno je okolo 120 frazeologických jednotek).

Publikace je doplněna seznamem příslušné odborné literatury, anglickým resumé a biogramy obou autorek.

Jak je z hesel SSK patrné, soukromé dopisy skutečně patří k jazykově velmi zajímavým komunikátům. Ilustrujme alespoň dvěma příklady:

čoch, -u m. slang. *zábava, party, přátelské setkání, často ve spojení s pitím alkoholu; mejdan*: Jenže je mi 21 a pořád mě omezují. U [XY] nesmím spát, když je třeba ples nebo čoch, musím být brzo doma... [1963]¹

zvířátko, a- stř. expr. s klad. cit. zabarvením (*malé*) *zvíře*: **1.** Sice, jak Tě znám, Tě trápí spuštění Temelína, války, opuštěný děti a zvířátka, přírodní katastrofy, zlí a škodní lidé atd. – všechno to fuj, co lidstvo dělá... [1899]; Nejprve Ti posílám spoustu pozdravů z té naší „velké“ vesnice čítající tak 500 obyvatel, k tomu i nějaké ty domácí zvířátka (prasátka, slepice...). [1926] **2. označení milé osoby (v oslovení)**: Ahoj zvířátko! [1979] Při oslovení užita zastaralá podoba s krátkým -i-, tj. zvířátko.

Na lexikografické zpracování soukromé korespondence v uvedené knize navazuje v obecnější, teoreticko-metodologické rovině další publikace, monografie Z. Hladké a kol. *Soukromá korespondence jako lingvistický pramen*. Autoři v sedmi samostatných kapitolách přispívají k výzkumu epistolárních textů novými pohledy, zaměřují se především na korpusové zpracování soukromých dopisů a jejich využitelnost pro kvantitativně orientovaný výzkum jazykového úzu. Nově jsou prezentovány moderní metodické postupy z oblasti sociolingvis-

¹ Evidenční číslo dopisu.

tiky a pragmalinguistiky. Zdrojem dílčích analýz tentokrát byla především korespondence významných osobností české kultury.

První tři kapitoly knihy jsou věnovány KSK-dopisy. Úvodní kapitola Z. Hladké *Korpusové zpracování soukromé korespondence v ÚČJ FF MU v Brně* přibližuje historii vzniku tohoto korpusu a metodiku jeho tvorby, zároveň také reflektuje zkušenosti spjaté s jeho užíváním. Technickým pozadím při jeho zpracování se detailně zabývá D. Hlaváčková (*Korpusové zpracování korespondenčních textů: morfologické značkování*), slovtvorné analýze cca 1000 hypokoristik z rodných jmen v něm doložených J. Machalová a K. Osolsobě (*Hypokoristika z rodných jmen v Korpusu soukromé korespondence*). Ve čtvrté kapitole navazuje Z. Hladká na své předchozí výzkumy posuzující korespondenční texty jako materiál pro poznání dynamiky vztahu mezi spisovnou podobou národního jazyka a jeho nespisovnými varietami² (*Teritoriální a sociálně podmíněné diference v jazyce soukromé korespondence*). Autorka zároveň představuje některé metodické postupy kvantitativně orientované analýzy a připojuje lingvistické reflexe starších dopisů (B. Smetany, B. Němcové). Z dopisů 19. století vychází také následující kapitola M. Nekuly a L. Rychnovské *Jazyková biografie Bedřicha Smetany a jazyk jeho korespondence se zřetelem ke gramatickým jevům*, která ilustruje sociolingvistické metody jazykových biografií. Edici rozsáhlé publikace *Němcová, Božena: Korespondence I-IV* (Praha: NLN 2003, 2004, 2006, 2007), metody a výsledky komunikační analýzy a interpretace této korespondence přibližuje kapitola *Dopisy jako výzva. Jazykovědný výzkum korespondence Boženy Němcové* (R. Adam – J. Janáčková – A. Macurová – F. Martínek). Poslední kapitolou (J. Hoffmannová: *Soukromý dopis: mezi psaným textem a mluveným dialogem*) se analýzy představené v předchozích částech publikace doplňují o tradiční stylistický přístup a o aspekty pragmalinguistické (zejména o problematiku zdvořilosti a tzv. teploturu); autorka využívala korespondenci J. Nerudy a K. Čapka i soukromé dopisy mládeže z KSK-dopisy. Monografie je v závěru opatřena biograpy všech autorů a anglickým resumé.

Tato monografie rozšiřuje a naznačuje spektrum možností výzkumu korespondence, poukazuje na její různé aspekty a ukázkově ji prezentuje v různorodých podobách. Předkládá důležité a cenné badatelské výsledky, stejně přínosná je však i její motivační či inspirativní hodnota: v návaznosti na předchozí lexikografickou sondu a spolu s ní jistě poslouží jako zdroj poznatků, odborných pramenů i podnětů kaplikaci toho, co bylo dosud na materiálu soukromé korespondence prozkoumáno. Obě představené publikace nepochybně najdou následovatele mezi badateli nejen ve sféře českého epistolárního stylu, jejich empirické poznatky i dílčí či obecnější závěry bude možné

² Viz seznam literatury na konci recenze.

aplikovat širše pro jiné styly, a to i v jiných jazycích. Přesvědčivě také dokládají, jaký podíl na mapování jazykového úzu mohou mít korpusové zdroje.³

LITERATURA

HLADKÁ, Z. (2001): Spisovnost a nespisovnost v jazyce soukromé korespondence (se zřetelem k teritoriální příslušnosti pisatelů). *Naše řeč* 84, 225–234.

HLADKÁ, Z. (2002): Tvorba a využití korpusů češtiny na FF MU v Brně. In: Z. Hladká, P. Karlík (eds.), *Čeština - univerzália a specifika* 4. Praha, 307–310.

HLADKÁ, Z. (2005): Zkušenosti s tvorbou korpusů češtiny v ÚČJ FF MU v Brně. In: *SPFFBU*, A 53, 115–124.

HLADKÁ, Z. (2006): Pravopis v současné korespondenci mladých lidí (na materiálu tradičních dopisů, e-mailů a SMS). *Naše řeč* 89, 73–88.

HLADKÁ, Z. (2008): Ke spisovnosti a nespisovnosti v současném úzu. In: M. Kopřivová, M. Waclavičová (eds.), *Čeština v mluveném korpusu*. Praha: Nakladatelství Lidové noviny / Ústav českého národního korpusu, 97–104.

³ I proto bylo mj. následně v korpusovém uložení a v celém rozsahu zaznamenáno 2000 ručně psaných soukromých dopisů z let 1902–2012 na dvou CD a vydáno v souboru HLADKÁ, Z. a kol. (2013): *111 let českého dopisu v korpusovém zpracování*. Brno: MU.

Žánr epistolární v korpusu

František Čermák

Ústav Českého národního korpusu Filozofické fakulty UK, Praha

frantisek.cermak@ff.cuni.cz

Korpus psaných soukromých dopisů vydaný interně pro studijní účely dlouholetou péčí Z. Hladké na Masarykově Univerzitě (*111 let českého dopisu v korpusovém zpracování*, srov. však už Hladká, 2005) je reprezentativní vzorek ručně psaných a do elektronické podoby přepsaných soukromých dopisů za dlouhou dobu více než jednoho století. Ta zahrnuje konec Rakouska-Uherska, první republiku, dobu komunistického režimu i první desetiletí 21. století a je zároveň, zčásti nechtěně, záznamem pokračujícího zániku tohoto žánru a tedy i pozoruhodným monumentem dlouhé odcházející tradice.

Takto shromážděný korpus přináší vedle prepisů, které budou plně zveřejněny na stránkách Českého národního korpusu, i původní, především rukopisné listy 2000 různých pisatelů, umožňující srovnání prepisu a obsahu s jejich původní rukopisnou formou.

Celý korpus, zčásti již zveřejněný dříve, začala autorka obětavě shromažďovat spolu s kolegy od konce 90. let minulého století. Data jsou pečlivě sociolinguvisticky vybavena: vedle archívních záznamů, signatur apod. především z hlediska pohlaví, věku pisatele, jeho vzdělání, roku napsání dopisu a formy, v zásadě rukopisné, popř. autorsky jiné. V podobě systematické anotace však zachycují také teritoriální rozložení podle pisatele (v zásadě podle tradičního třídění nářečí), jeho vztah k adresátovi (vztah formální, kamarádský, milenecký či partnerský, rodinný, popř. jiný) i přehled rozvrstvení dat podle časových období (v zásadě podle desetiletí). Celý korpus je v souladu s potřebou zachovat soukromí pisatelů autorsky anonymní. Podle všech těchto parametrů lze daná data analyzovat a dokreslovat tak obrázek češtiny za delší dobu, která svými začátky už patří i do diachronie.

Edice, která vyšla ve velmi malém nákladu a kterou doprovázející dvě analytické knihy, zčásti slovníkové povahy (Hladká, 2012; Hladká – Martincová, 2013), tvoří dvě CD s názvem KSK111 a SSK111. Jedno obsahuje vlastní korpus (s programem Bonito) s doprovodnou dokumentací, charakteristikami všech

textů, možností vytvořit si frekvenční slovníky, ale podle dostupnosti i fotokopiemi aj. Druhé se s prvním obsahově shoduje, zpřístupňuje však pro rozsáhlejší studium plnou podobu textových dokumentů v podobě excelových tabulek.

Je na místě ocenit jedinečnou povahu korpusu, který zřejmě už nikdy nebude mít svého konkurenta (nepočítáme-li s možností zcela jiného dopisování si v e-mailech či esemeskách), nemá ho však ani dnes, ani za hranicemi. Necháme-li stranou mimořádnou možnost studovat sociální vztahy mezi pisatelem a čtenářem, která se jinde příliš nenabízí, jedinečnost tohoto korpusu je především v obsahu a typu jazyka, mj. pro jeho smíšenou psaně-mluvenou povahu ale i záznam minulých dob, reálií, které pisatele zajímaly aj. Závěrem je samozřejmě také třeba ocenit jedinečnost vkladu autorky, bez jejíž dlouholeté obětavosti by tento korpus nebyl.

LITERATURA

HLADKÁ, Z. a kol. (2005): *Čeština v současné soukromé korespondenci*. Dopisy, e-maily, SMS. Brno: Masarykova univerzita.

HLADKÁ, Z. a kol. (2013): *111 let českého dopisu v korpusovém zpracování*. Brno: Masarykova univerzita.

HLADKÁ, Z. a kol. (2013): *Soukromá korespondence jako lingvistický pramen*. Brno: Masarykova Univerzita.

HLADKÁ, Z. – Martinčová, O. (2012): *Slova v soukromých dopisech. Lexikografická sonda*. Brno: Masarykova univerzita.

Gramatika a korpus 2012: 4. mezinárodní konference

28.–30. listopadu 2012

Kateřina Najbrtová

Filozofická fakulta Masarykovy univerzity, Brno

145034@mail.muni.cz

Čtvrtá mezinárodní konference Gramatika a korpus se konala u příležitosti stého výročí narození Miloše Dokulila, význačné osobnosti české lingvistiky. Po předchozích setkáních, která se uskutečnila v Praze (2005), v Liblicích (2007) a v Mannheimu (2009), se konference vrátila zpět do Prahy, kde proběhla ve dnech 28.–30. listopadu 2012. Záštitu nad konferencí převzal Ústav pro jazyk český AV ČR, avšak mezi členy organizačního výboru jsme mohli nalézt mnoho zahraničních lingvistů.

Každý ze tří konferenčních dnů byl zahájen plenárním referátem, s výjimkou prvního dne setkání. Zahájení konference bylo věnováno Miloši Dokulilovi – připomenutí jeho vybraných myšlenek se ujal František Štícha.

V konferenčních příspěvcích byla zastoupena nejrůznější témata, více či méně spojená s korpusy. Kromě referátů zabývajících se všemi tradičně vymezenými rovinami popisu jazyka byly prezentovány i příspěvky s tématem budování korpusů, pragmatických aspektů komunikace nebo příspěvky založené na jiných než českých korpusech. Zcela v duchu setkání početně dominovaly prezentace věnované výzkumu slovtvorby a také morfologie a syntaxe.

Plenární referáty prvního dne přednesli A. Polikarpov (*Regularities of Word-Formation in a System Theory of Language Evolution*) a P. Štichauer, který představil novou potenciální klasifikaci kompozit (*Is a new classification of Czech compounds possible?*). V dalším plenárním referátu (*Measuring Standard German? Frequency and Variation as Indicators for Acceptability*) M. Konopka a U. H. Waßner poukázali na to, že míra frekvence může být dobrým pomocníkem při diskuzi o přijatelnosti jevů, ale ne vždy se jedná o přímou závislost. V závěrečný den konference se A. Rosen věnoval vztahu empirie a teorie (v saussurovské terminologii parole a langue) a potíží při vytváření závislostních stromů, jež

by v ideálním případě měly zachycovat oboje (*Grammar based treebank – a happy marriage of empirism and theory?*).

V příspěvku nazvaném příhodně *Srovnávání nesrovnatelného* předvedl J. Králík, že společně s výrazným zvýšením objemu dat musí být upraveny i zažité metody zacházení s nimi – extrémní růst sledovaných textů nevede k extrémnímu růstu přesnosti získaných údajů.

Na korpus jako zdroj ověřování dat, nikoli jejich získávání se zaměřili P. Karlík, M. Ziková a studenti Masarykovy univerzity (*Osobní zájmena: ověřování hypotéz daty (intuice>KORPUS>dotazník)*).

Naopak to, že v určitých případech nemusí být ani rozsáhlé korpusy plně dostačující, prokázala K. Osolobě (*Korpusy a internet jako zdroje dat pro výzkum produktivity periferního slovtvorného typu: adjektiva typu hrůzoucí (hrůza) v korpusech a na internetu*).

Využití korpusu DIAKORP pro diachronní bádání demonstrovali A. Černá (*Člověk v množném čísle (v historické i v současné češtině)*) či F. Martínek (*Slovtvorné typy českých deverbativ a deadjektiv (diachronní pohled)*).

V příspěvku *Automatické generování slov ve slovanských jazycích na základě jejich společných historických kořenů* upozornili P. Pognan a J. Panevová na česko-francouzský projekt počítačového programu, který na základě vložených údajů utváří současné podoby staroslovanských kořenů v různých slovanských jazycích.

Deriváty adjektiv z hlediska jejich vztahu k základovým adjektivům a následného zachycení v anotaci PDT se zabývala M. Ševčíková (*Deadjektivní deriváty v češtině jako deriváty syntaktické vs. lexikální*). Oblast tvorby korpusů a využívání korpusových nástrojů dále zastoupily příspěvky H. Palátové a M. Gráce (*Segmentace textu na věty*) či M. Hnátkové (*Automatická slovnědruhově desambiguace slova „to“ v ustálených větných výrazech*).

Výzkum synchronních slovtvorných vztahů mezi zadanými vzory slovtvorných paradigmat umožňuje nový nástroj MORFIO, o němž informovali V. Cvrček a P. Vondříčka (*Nástroj pro slovtvornou analýzu jazykového korpusu*).

Nesnáze rozpoznávání dativu při anotaci se staly tématem příspěvku V. Petkeviče (*Automatické rozpoznávání substantivního dativu a jeho syntaktických funkcí v českých textech*). Zmíněný úkol je velmi náročný, jeho (byť ne zcela dokonalé) řešení spočívá především ve využití slovesné i substantivní valence.

Webovou aplikaci Corpus Architect umožňující pohodlné budování textových korpusů včetně případné tokenizace a lemmatizace představili J. Pomikálek a V. Suchomel (*Creating Corpora Using Corpus Architect*).

Zajímavý způsob analýzy jazyka využívající závislosti funkce substantiv ve větě na funkčním stylu nabídl T. Jelínek (*Porovnávání funkčních stylů v korpusu SYN2005 na základě frekvence syntaktických funkcí substantiv*).

Problematické vidu a možnostem a omezením jeho výzkumu v korpusech věnovali svá vystoupení F. Esvan (*K střídání vidu při vyprávění v češtině*), T. Berger (*Funkcionální vidové dvojice v ČNK*) a P. Nádeníček (*Dílčí druhy imperfektivnosti a perfektivnosti. Příspěvek k vidovému systému slovanských jazyků*).

Součástí bohatého programu byly i referáty zahraničních badatelů zaměřené např. na ruštinu – S. Gorokhova (*The frequency effect in the production of adjective-noun agreement: A corpus based study of Russian speech errors*) či I. B. Itkin a kol. (*Research on stressed prepositions in Russian: a corpus based approach*), německý jazyk regionů Rakouska, Německa a Švýcarska – S. Ueberwasser (*Pluri-Regional German Grammar: A Corpus based Approach*) či slovo tvorbu litevského jazyka – J. Pakerys (*The co-occurrence of denominal and deadjectival verbs with their base words in Modern Lithuanian (a corpus based study)*).

Z. Hladká předvedla, jakým způsobem je možné na příkladech hypokoristik z KSK odhalit slovtvorné postupy (*Využití korpusů korespondence pro účely (nejen) slovtvorných analýz*), deminucí – ať už z pohledu češtiny či čestiny v komparaci s jinými jazyky – se zabývali J. Bílková (*Pseudodeminutiva v češtině*) a T. Káňa (*Česká substantivní deminutiva ve světle korpusových dat*), primární deminutiva odvozená sufixy *-ek* a *-ík* analyzovala z fonologického hlediska M. Ziková (*Maskulinní deminutiva v současné češtině*). Jak interpretovat nejruznější zvláštnosti a inovace slovtvorných procesů v češtině, ukázal J. Šimandl (*Nárůst nepravidelného tvoření slov: o čem svědčí?*). Kompozicí jako slovtvorným postupem se zabývala I. Bozděchová (*Kompozice a její potenciál v současné slovní zásobě češtiny*), vybrané sufixy a jejich funkce spolu s možnostmi spojování byly předmětem zájmu D. Hlaváčkové a K. Paly (*Některé vybrané substantivní sufixy v češtině (-ák, -ec, -ík, -ník) – případová studie*).

Ve vystoupeních tematicky orientovaných na syntax využili jejich autoři korpus jak ke zkoumání spíše okrajových jevů (A. Karčová: *Postponovaný přívlastok v slovenčine a češtině*, M. Šimková: *Predložky kvôli/kvůli a pre/pro v systéme a v použití (na materiáli slovenského a českého jazyka)*), tak jevů ovlivňujících široké úrovně jazyka (M. Dočekal, H. Strachoňová: *Negace a skopus kvantifikátorů*, M. Mikulová a další: *Liší se mluvené a psané texty ve valenci?*), o diachronní náhled se postaral P. Kosek (*Vývoj slovosledných vlastností auxiliáru préterita ve starší češtině*).

Konference byla nejen důstojnou poctou M. Dokulilovi, ale především podnětným setkáním jazykovědců různých věkových skupin a národností, vzájemně spojených tématem korpusového zkoumání jazyka. Autoři referátů obsáhli široké spektrum oblastí výzkumu, přičemž důraz byl kladen na témata slovo-

tvorná a syntaktická. Přednesené příspěvky jsou vynikajícím dokladem toho, jak veliké (de facto neomezené) využití nabízí počítačově zpracovaný jazyk – ještě nikdy v historii neměli lingvisté v rukou zdroje dat v takovém rozsahu. Nutno dodat, že k příjemnému průběhu setkání nemalou měrou přispěla i společná slavnostní večeře, která umožnila neformální setkání (a často i diskuzi) všech zúčastněných.

Příspěvky byly vydány na CD nakladatelstvím Gaudeamus Hradec Králové.

CzeSL – Korpus češtiny nerodilých mluvčích

CzeSL – a Corpus of Non-Native Speakers' Czech

Svatava Škodová

Technická univerzita, Liberec

svatava.skodova@tul.cz

Elektronické jazykové korpusy, definované jako *rozsáhlé soubory elektronických textů, cíleně shromážděných jako referenční zdroj pro vědecké studium* (Čermák 2005), můžeme dnes pokládat za tradiční nástroj vyučování cizích jazyků a výzkumů s ním spojených. Teprve v 90. letech minulého století se však začaly vytvářet korpusy vztahující se k cizím jazykům jako takovým, jsou to tzv. korpusy žákovské.¹ Lze je vymezit jako elektronické korpusy písemných i mluvených projevů nerodilých mluvčích daného jazyka s podrobným záznamem metadat, tj. informací o autorovi, původu textu a okolnostech jeho získání.

Žákovské korpusy se od běžných korpusů synchronních řadou parametrů liší. Zásadní rozdíl představuje povaha shromážděných jazykových dat. Běžné synchronní korpusy směřují k reprezentativnosti ve vztahu k současnému jazyku rodilých mluvčích. „*Korpusová data jsou dnes ve vztahu k jazyku charakterizována jako (1) typická, (2) aktuální, synchronní a věrná, (3) neselektivní, (4) objektivní a realistická, (5) dostatečná, (6) nenáhodně získaná a (7) získatelná a získávaná snadno a rychle*“ (Čermák 2005: 19), zatímco korpusy žákovské na zachycení současného jazyka neaspírají, ale zaznamenávají tzv. mezijazyk, tj. individuální jazyk žáků, který představuje strukturu velmi dynamickou a proměnlivou, která se může od běžně užívaného jazyka výrazně odlišovat. Z toho pak plynou četné další specifické rysy žákovských korpusů a jejich vytváření, např. odlišné pojetí autenticity sbíraných projevů, reprezentativnosti, odlišný způsob sběru jazykových dat apod.²

Žákovské korpusy jsou zásadním inovačním prvkem ve výzkumu osvojování cizích jazyků i jejich vyučování. Jejich badatelská hodnota spočívá v tom,

¹ Learner corpus (S. Granger). V české literatuře se zpočátku objevil termín korpus studijní, termín korpus žákovský se nám jeví jako vhodnější.

² Podrobněji o specifických rysech žákovských korpusů a akvizičních korpusů obecně pojednává Šebesta (2010), Štindlová (2011).

že poskytují širokému okruhu výzkumníků velké soubory jazykových dat pro zkoumání žákovského mezijazyka, jeho vývoje a faktorů, které tento vývoj ovlivňují, dovolují identifikovat s poměrně velkou jistotou korektní i nekorektní jazykové pravidelnosti v mezijazyce a jejich vztah k různým potenciálním činitelům, které ho ovlivňují, jako je věk, první jazyk žáka, okolnosti osvojování druhého jazyka, délka a povaha formální jazykové výuky apod. Žákovské korpusy umožnily nově definovat a rekonstruovat oba základní typy analýz, které se v studiu osvojování druhého/cizího jazyka tradičně uplatňovaly: kontrastivní analýzu a analýzu chybovou.

O žákovský korpus opřena kontrastivní analýza se od analýzy tradiční odlišuje tím, že je zaměřena nikoli na studium výchozího a cílového jazyka, ale právě mezijazyka, a sice na základě srovnání mezijazyka skupiny žáků s jazykem cílovým nebo na základě srovnání mezijazyků dvou různých skupin žáků. Zjišťují se přitom nejen odchylky ve smyslu užití nekorektního, ale i nadužívání nebo podužívání jednotlivých jazykových prostředků a konstrukcí.

Chybová analýza opřena o žákovský korpus je často založena na specifickém typu chybové anotace textů; to s sebou nese systematicčnost a explicitnost v míře tradičními metodami obtížně dosažitelné. Velkou výhodou je i to, že při využití korpusu analyzujeme užití nestandardní či nekorektní na pozadí užití korektních, můžeme si systematicky všimnout funkčního využití nekorektních forem, sledovat aspekt cizorodosti (tzv. *foreign-soundingness*) apod. Výsledky kontrastivních i chybových analýz opřených o korpus se už poměrně dlouho a hojně využívají při tvorbě jazykových slovníků a učebních materiálů, a to především v angličtině.

První známý veřejný žákovský korpus začal být budován v belgické Lovani počátkem 90. let pod názvem International Corpus of Learner English (ICLE). U jeho zrodu stojí S. Grangerová a centrum CECL (Centre for English Corpus Linguistics). CECL stojí u zrodu i dalších žákovských korpusů, např. LINDSEI, LONGDALE, VESPA aj., a jeho činnost byla inspirací pro vytváření korpusů i v jiných zemích, zvl. dálnévýchodních. Vedle těchto korpusů s primárně akademickým určením vznikají od 90. let i korpusy komerční, spojené s působením velkých nakladatelství zaměřených na výuku jazyků, např. Longman Learners' Corpus, Cambridge Learner Corpus.

Další jazyky jsou v žákovských korpusech zastoupeny výrazně méně než angličtina, celkový rozsah jazykových dat zachycených jejich žákovskými korpusy – pokud jde o korpusy veřejně známé či dostupné – zpravidla nedosahuje jednoho milionu slov. Ze slovanských jazyků je zastoupena pouze slovinština korpusem o rozsahu 35 000 slov. Český žákovský korpus se díky svým parametrům řadí ke korpusům ve světovém měřítku unikátním.

Český žákovský korpus CzeSL (Czech as Second Language) vznikl ve spolupráci Technické univerzity v Liberci a Univerzity Karlovy v Praze. Jeho vznik je

spjat s projektem Inovace vzdělávání v oboru čeština jako druhý jazyk,³ koncepčně se řadí do skupiny akvizičních korpusů AKCES (Akviziční korpusy češtiny)⁴ budovaných od r. 2005 Ústavem českého jazyka a teorie komunikace FF UK⁵. Texty pro český žakovský korpus byly sbírány v letech 2009–2012, a to výhradně ve formálním školním prostředí kurzů různého typu. Vyjma tzv. odborné kvalifikační práce, byly všechny školní písemné práce, testové texty a eseje získávány v originální rukopisné podobě a následně pak systematicky přepisovány do podoby elektronické; tímto způsobem byl jednoznačně vyloučen vliv automatického jazykového korektoru na výslednou podobu textů a zároveň bylo možné zachytit všechny autorské opravy prováděné v textu. Pro tento korpus nebylo možné důsledně vycházet z dosavadních zkušeností při budování Českého národního korpusu. Nově bylo potřeba vyvinout např. koncepci metadatového pozadí textů, pravidla pro přepis textů nerodilých mluvčích, zvláště však anotační program FEAT (autor: J. Hana) a systém chybové anotace.

Korpus CzeSL je v současné době přístupný v ČNK.⁶ Svým rozsahem (cca 2,5 miliony pozic) patří mezi neanglickými žakovskými korpusy k největším. Pokud jde o první jazyky respondentů, skládá se z jazykových dat získaných od čtyř skupin mluvčích:

(a) Od mluvčích s prvním jazykem blízce příbuzným, tedy slovanským. Velkou převahu mezi nimi mají mluvčí s ruštinou nebo jiným východoslovanským jazykem.

(b) Od mluvčích s jiným (neslovanským) indoevropským jazykem. V této skupině není výrazná dominance žádného jednotlivého jazyka.

(c) Třetí skupinu představují mluvčí s prvními jazyky nepřibuznými.

Skupiny a) – c) zahrnují 8 863 textů v rozsahu 1 314 901 pozic.

³ Projekt (reg. číslo CZ.1.07/2.2.00/07.0259) se řeší v rámci OP Vzdělávání pro konkurenceschopnost, je financován ze zdrojů Strukturálních fondů EU – Evropského sociálního fondu a ze státního rozpočtu ČR. Příjemcem dotace je Technická univerzita v Liberci, na řešení se jako partneři podílejí Univerzita Karlova v Praze a Asociace učitelů češtiny jako cizího jazyka. UK jako partner zajišťuje vznik korpusu. Hlavním koordinátorem projektu je K. Šebesta, sběr a primární zpracování materiálu koordinují M. Hrdlička, T. Hrdličková, Z. Bedřichová a K. Šormová, zajištění korpusu včetně anotací V. Petkevič, anotační program vytvořil J. Hana, koordinátorkami dílčích aktivit projektu jsou S. Škodová a K. Váňová. Řešení se účastní několik pracovišť obou univerzit (KČL FP TUL, ÚBS, ÚČJTK a ÚTKL FF UK, ÚJOP UK), řada studentů doktorského, magisterského i bakalářského studia a četná pracoviště neakademická, především školy a občanská sdružení.

⁴ Viz <http://ucho.ff.cuni.cz/czesl-plain.php>.

⁵ Blíže viz Šebesta (2010).

⁶ Pro další informace srov. Škodová (2009).

(d) Čtvrtou skupinu tvoří texty od romských žáků, celkem 4 420 textů v rozsahu 428 161 pozic. Pro respondenty tohoto subkorpusu, označeného a publikačně prezentovaného jako ROMi, je typické, že ne vždy lze jednoznačně rozhodnout, zda je čeština jejich jazykem prvním, nebo druhým; sociokulturní odlišnosti mezi českou neromskou komunitou a některými komunitami romskými jsou však takového druhu, že lze u jazykového vývoje romských dětí očekávat některé rysy připomínající osvojování češtiny jako druhého jazyka. Romský subkorpus je budován v některých bodech odlišně a jsou u něj zaznamenány i zčásti odlišné parametry.

Ačkoliv byl CzeSL budován s úsilím o maximální možnou úplnost a vyváženost, nebylo vždy možné tohoto stavu dosáhnout:

(a) Je založen na sběru psaných i mluvených projevů žáků, psaná složka však výrazně převažuje. Texty byly sbírány v rukopisné podobě a přepisovány podle podrobně stanovených pravidel, která zajišťují, aby bylo z původního textu zachováno maximum informací (např. rektifikační zásahy studenta, korektury učitele apod.). Výjimku představují kvalifikační práce vytvářené, a tedy i sbírané v podobě elektronické.

(b) Pokrývá všechny úrovně znalosti jazyka podle SERR. V tom se odlišuje od většiny jiných světových žákovských korpusů, které zachycují často pouze jazyk žáků jedné či dvou úrovní znalosti, zpravidla středně pokročilých. V tomto parametru CzeSL nebylo možné dosáhnout vyváženosti: početně převažují práce studentů úrovně B1 a B2, úrovně nižší jsou zastoupeny méně. Tato nevyváženost je dána reálným zastoupením studentů češtiny pro cizince v ČR.

(c) Žánrově a tematicky jde rovněž o texty různorodé. Světové korpusy se někdy omezují na argumentativní a úvahové texty (srov. např. jmenovaný ICLE), korpus CzeSL k takovému omezení nesahá. Největší jeho část tvoří eseje psané jako součást zkoušky, podobně jako je tomu u většiny světových korpusů. 731 816 pozic korpusu však tvoří kvalifikační práce, zvláště bakalářské, magisterské a doktorské. Protože jde o práce kvalitativně jiné než ručně psané eseje a také podmínky sběru jsou u nich poněkud odlišné, tvoří samostatný subkorpus.

(d) Všechny texty jsou vybaveny podrobnými metadaty o mluvčích a textech, zvl. o podmínkách jejich vzniku a sběru. Ve srovnání s běžnými korpusy synchronními je u žákovských korpusů uvádění co nejdetailnějších a přesných dat o mluvčích a textech naprosto nezbytné. V CzeSLu je u mluvčích zaznamenán jejich věk, první jazyk, další jazyky, které zná, pobyt v České republice, případné kontakty s češtinou (např. je-li rodina bilingvní), úroveň znalosti češtiny a podrobně se zaznamenává doba a způsob jejího osvojování a učení (jak dlouho se žák učí češtinu, jak intenzívně, s využitím jakých učebnic).

U textů jsou zaznamenány vedle tématu, žánru a rozsahu podrobně zejména podmínky jejich vzniku, tj. míra řízenosti jejich tvorby učitelem (téma ne/zadáno, žánr ne/zadán, velikost ne/zadána, čas ne/zadán), velikost a povaha opory (ne/možnost využít slovníku, event. přípravné aktivity při zadání apod.) a okolnosti sběru (psáno pro korpus, jako součást zkoušky apod.). S uvedenými parametry se pracuje při vyhledávání v korpusu; hodnota žákovského korpusu a akvizičního korpusu obecně s počtem zaznamenaných metadat roste.

Žákovské korpusy mohou být vybaveny rovněž lingvistickými anotacemi a zvláště specifickým anotačním typem, tzv. chybovou anotací. Anotace těchto korpusů je však mimořádně náročná, především proto, že jde o texty chybové, u nichž nelze zpravidla využít automatické anotační nástroje. Proto zůstává většina žákovských korpusů bez lingvistických anotací.

Korpus CzeSL je – podobně jako další akviziční korpusy AKCES – zčásti vybaven standardní morfosyntaktickou anotací a lemmatizován a bude rovněž vybaven značkami chybovými,⁷ v současné době je však veřejnosti zpřístupněn v ČNK pouze v plain verzi a na anotaci se dále pracuje. Chybová anotace vzhledem ke své náročnosti nepokryje korpus celý, jen jeho část.

Řešitelský tým⁸ se musel vyrovnat s problémy plynoucími z faktu, že jde v mnohých případech o texty vysoce chybové, a rovněž s problémy spojenými s vysokou flektivností češtiny a jejím specifickým slovosledem. Anotační schéma, které tým vypracoval, poskytuje několik výhod, z hlediska češtiny velmi podstatných:

(a) dává možnost zachytit a opravit chyby v nespojitých řetězcích slov;

(b) dává možnost zachytit některé syntaktické chyby – ve shodě, jiném způsobu vyjádření závislosti, v zájmeném odkazování apod.;

(c) dovoluje kombinovat manuální a automatické chybové značkování (automatické značkování je založeno na porovnání chybných a korigovaných slov a tvarů);

(d) dovoluje vnést i do chybového textu morfosyntaktické značkování s využitím automatických nástrojů.

Emendace (opravy) textu se provádějí při anotaci na dvou rovinách. Ze základní roviny R0 obsahující obsahově originální verzi textu, do níž anotátor nezasahuje, se text přenáší do roviny R1, na níž se provádějí takové zásahy, aby

⁷ Podrobně k této problematice Hana – Rosen – Škodová – Štindlová 2010.

⁸ Na vypracování anotačního schématu se podíleli prof. K. Šebesta, členové ÚTKL pod vedením doc. Petkeviče, dr. Štindlová a dr. Škodová z TUL a pro část ROMi Mgr. Šormová a Mgr. Bedřichová.

jejich výsledkem byl řetězec v češtině existujících slov (vyjma autorských novotvarů, neidentifikovatelných lexémů apod.) a slovních tvarů.

Takto upravený text se pak přenáší do roviny R2, kde se provádějí zásahy, jejichž výsledkem je gramaticky korektní věta, ojediněle (u zájmeného odkazování) text. Jde tedy o chyby syntaktické, slovosledné, o chyby v negaci, zájmeném odkazování apod. Na této rovině se opravují i některé chyby další, např. v složených slovesných tvarech.

Manuální emendaci a anotaci každého textu prováděli nezávisle na sobě dva anotátoři, výsledek jejich práce sjednocuje a hodnotí supervizor. Anotátoři se při emendacích řídí zásadou minimální intervence do původního textu. Samotná anotace využívá i automatických postupů přejatých z korpusů národních.

Anotační značky (označení typu chyby) se vnášejí do textu na dvou mezerovinách (mezi R0 a R1, mezi R1 a R2) do spojnic mezi příslušnými slovy či sekvencemi slov, v první fázi rovněž manuálně, následně pak, kde je to možné, automaticky. V první mezerovině (mezi R0 a R1) může být v této fázi vyznačeno 9 kategorií chyb, z toho 7 manuálně, 2 automaticky, v druhé mezerovině 19 kategorií chyb, z toho 12 manuálně a 7 automaticky; počet vyznačovaných chyb i jejich vymezení se ale v průběhu anotace podle potřeby upřesňují.

Ačkoliv anotace CzeSLu vyžaduje ještě mnoho usilovné práce, může korpus sám o sobě již nyní přinést materiál v podobě a rozsahu unikátních pro zkoumání osvojování češtiny jako druhého/cizího jazyka, pro autory učebních materiálů i pro tvůrce standardizovaných zkoušek.

LITERATURA

ČERMÁK, F. (2005): Korpus, informace a lingvistika. In: *Přednášky z XLVIII. běhu LŠSS UK*. Praha: Karolinum, 19–20.

GRANGER, S. (2002): A Birds-eye view of learner corpus research. In: S. Granger, J. Hung, S. Petch-Tyson (eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Université Catholique de Louvain, 3–35.

GRANGER, S. (2003): Error-tagged learner corpora and CALL: a promising synergy. *CALICO*, 20, 3, 465–480.

GRANGER, S. (2008): Learner Corpora. In: A. Lüdeling, M. Kytö (eds.), *Corpus Linguistics. An International Handbook*. HSK 29. 1. VOL. 1. Berlin/New York: Mouton De Gruyter, 259–274.

GRANGER, S. (2003): The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. In: *TESOL Quarterly*, 37, 3, 538–545.

JAMES, C. (1998): *Errors in Language Learning and Use*. Longman.

NESSELHAUF, N. (2004): Learner corpora and their potential in language teaching. In: J. Sinclair (ed.), *How to Use Corpora in Language Teaching*. Amsterdam/Philadelphia: Benjamins, 125–152.

SEEGMILLER, M. S. – FITZPATRICK, E. (2003): Practical aspects of corpus tagging. In: B. Lewandowska-Tomaszczyk (ed.), *Palc 2001: Practical Applications in Language Corpora*. Peter Lang Pub Inc.

ŠEBESTA, K. (2010): Korpusy češtiny a osvojování jazyka. *Studie z aplikované lingvistiky/Studies in Applied Linguistics*. roč. 1, č. 2, 11–34.

ŠTINDLOVÁ, B. (v tisku) Žákovský korpus. Budoucnost pro poznávání akvizice cizího jazyka. In *Minulost, přítomnost a budoucnost v jazyce a v literatuře. Ústí nad Labem 1.–3. 9. 2010*. PF UJEP: Ústí nad Labem, 2010.

XIAO, R. (2008): Well-known and influential corpora. In: A. Lüdeling, M. Kytö (eds.), *Corpus Linguistics. An International Handbook*. HSK 29.1. Vol. 1. Berlin/New York: Mouton de Gruyter, 383–457.

LITERATURA A WWW ZDROJE TÝKAJÍCÍ SE CzeSLu

BEDŘICHOVÁ, Z. – ŠEBESTA, K. – ŠKODOVÁ, S. – ŠORMOVÁ, K. (2011): Podoba a využití korpusu jinojazyčných a romských mluvčích češtiny: CZESL a ROMi. In: *Korpusová lingvistika Praha 2011 – 3 Gramatika a značkování korpusů*. NLN: Praha, 93–104.

BEDŘICHOVÁ, Z. – ŠEBESTA, K. – ŠORMOVÁ, K. (2011): ROMi – první rozsáhlá databanka romského etnolektu češtiny. In: *Lidé města*, 13, 160–163.

HANA, J. – ROSEN, A. – ŠKODOVÁ, S. – ŠTINDLOVÁ, B. (2010): Error-tagged Learner Corpus of Czech. In: *Proceedings of The Fourth Linguistic Annotation Workshop (LAW IV), Uppsala University 15. – 16. 7. 2010*. ACL: Uppsala, 11–19. Dostupné z: www.aclweb.org/anthology/W/W10/W10-18.pdf.

HRADILOVÁ, D. (2012): Užitečná publikace o užitečném projektu. *Bohemica Olomouciensia*, roč. 4, č. 3, 237–239.

ŠEBESTA, K. (2011): Akviziční korpusy. In: *Minulost, přítomnost a budoucnost v jazyce a v literatuře*. Ústí nad Labem 1.–3. 9. 2010. Ústí nad Labem: PF UJEP.

ŠEBESTA, K. (2010): Korpusy češtiny a osvojování jazyka. *Studie z aplikované lingvistiky/Studies in Applied Linguistics*, sv. 1, č. 2, 11–34.

ŠEBESTA, K. – ŠKODOVÁ, S. (eds.) (2012): *Čeština – cílový jazyk a korpusy*. Liberec: Technická univerzita v Liberci. Dostupné z: <http://kcl.fp.tul.cz/cs/cestina-cilovy-jazyk-a-korpusy>.

ŠKODOVÁ, S. (2009): Možnosti zachycení chyb v tzv. žákovských korpusech. In: O. ULIČNÝ (ed.), *Eurolingua a Eurolitteraria : sborník příspěvků mezinárodní konference Nepřipustnost, nepatřičnost a základy v jazykové a literární komunikaci*, Liberec 2010. Liberec: TUL, 197–205. Dostupné z: <http://kcl.fp.tul.cz/cs/sbeevlevo>.

ŠKODOVÁ, S. – ŠTINDLOVÁ, B. – HANA, J. – ROSEN, A. (2011): Víceúrovňová anotace českého žákovského korpusu. In: V. Petkevič – A. Rosen (eds.), *Korpusová lingvistika Praha 2011 – 3 Gramatika a značkování korpusů*. NLN: Praha.

ŠTINDLOVÁ, B. (2011): *Evaluace chybové anotace v žákovském korpusu češtiny*. Disertační práce. Filosofická fakulta University Karlovy v Praze.

ŠTINDLOVÁ, B. (2011): *Manuál pro přepis psaných materiálů*. Dostupné z <http://utkl.ff.cuni.cz/~rosen/public/transkripce.pdf>, http://utkl.ff.cuni.cz/~rosen/public/transkripce_doplnek.

ŠTINDLOVÁ, B. – ŠKODOVÁ, S. (2011): Žákovské korpusy, CzeSL a čeština jako druhý jazyk. In: *Sborník vědecké konference s mezinárodní účastí Sapere Aude – evropské a české vzdělávání*. Hradec Králové 21.–25. 3. 2011. Hradec Králové: Magnanimitas, 664–674.

ŠTINDLOVÁ, B. – ŠKODOVÁ, S. – ROSEN, A. – HANA, J. (2012): Annotating foreign learners' Czech. In: M. Ziková, M. Dočekal (eds.): *Slavic languages in formal grammar : proceedings of FDSL 8.5, Brno 2010*. Peter Lang GmbH, Frankfurt am Main, Germany, 205–219.

ŠTINDLOVÁ, B. – ROSEN, A. (2012): Návod k anotaci chybového korpusu, verze 5. Dostupné z <http://utkl.ff.cuni.cz/~rosen/public/anotace.pdf>.

ŠTINDLOVÁ, B. – ROSEN, A. – HANA, J. – ŠKODOVÁ, S. (2012): CzeSL – an error tagged corpus of Czech as a second language. In: Pězik, P.(ed.), *Corpus Data across Languages and Disciplines*, volume 28 of Łódź Studies in Language. Frankfurt am Main. Peter Lang. PALC 2011 – Practical Applications in Language and Computers, 21–32.

TONO, Y. Learner corpora: design, development and applications. In: *Proceedings of the Corpus Linguistics 2003 Conference*. United Kingdom: Lancaster, 2003, 800–809.

WWW ZDROJE

PRAVEC, N. A.: Survey of learner corpora. ICAME Journal [online], 2002, č. 26, 81–114. Dostupné z: <http://icame.uib.no/ij26/pravec.pdf>.

SINCLAIR, J. EAGLES: Preliminary recommendations on Corpus Typology. EAG--TCWG--CTYP/P. Version of May, 1996. Dostupné z: http://www.ilc.pi.cnr.it/EAGLES96/corpus_typ/corpus_typ.html.

VALIŠOVÁ, P. Korpus jako zdroj systémového popisu české konjugace v učebnicích češtiny jako cizího jazyka. DP, FF MU, 2009. Dostupné z: https://is.muni.cz/auth/th/75420/ff_m_bl/?fakulta=1421;obdobi=4703;studium=499045.

THOMAS, J. Using Corpora in Language Teaching and Learning. Teaching English with Technology, Journal for Teachers of English. 2005, č. 6/1. Dostupné z http://www.iatefl.org.pl/call/j_soft23.htm.

Co je v ČNK nového IV

(Zprávy z Českého národního korpusu)

Lucie Benešová

Ústav Českého národního korpusu Filozofické fakulty UK v Praze

Lucie.Benesova@ff.cuni.cz

Výsledkem usilovné práce na jednom z projektů realizovaných v ÚČNK je v prosinci 2013 nově zveřejněný **korpus spontánní mluvené češtiny ORAL2013**. Hlavními koordinátorkami tohoto projektu byly Lucie Benešová a Martina Waclawičová, technickou podporu zajišťoval Michal Křen. Korpus ORAL2013 je přístupný všem uživatelům ČNK ve standardní nabídce synchronních mluvených korpusů. Podrobné informace o novém korpusu najdou uživatelé na stránkách <<http://korpus.cz>>. Přestože tento korpus svou koncepcí navazuje na řadu mluvených korpusů ORAL, tj. na korpusy neformální mluvené češtiny ORAL2006 a ORAL2008, v mnoha podstatných ohledech doznal změn a vylepšení, jimž věnuji tuto zprávu.

Stejně jako předcházející zmíněné korpusy obsahuje i ORAL2013 data z ryze spontánních, neformálních komunikačních situací. Maximální možné míry autenticity jednotlivých nahrávek jsme se snažili dosáhnout tak, že mluvčí nebyli o nahrávání informováni předem, ale až po jeho skončení. Hlavním cílem bylo získání materiálu představujícího prototypický spontánní mluvený jazyk, jehož sběr patří k nejnáročnějším. Mezi důležité aspekty, které byly během sběru dat zohledňovány, patří zejména dialogičnost promluv (muselo jít nutně o rozhovor dvou nebo více mluvčích), jejich bezprostřední kontakt (fyzická přítomnost všech mluvčích na jednom místě; záměrně se nesbírala data z telefonních rozhovorů, z komunikace přes Skype atp.), vzájemný blízký vztah mluvčích a v neposlední řadě nevěřejnost a neoficiálnost dané komunikační situace.

Korpus ORAL2013 se skládá z 836 nahrávek z let 2008–2011 a obsahuje více než 2,7 milionu textových slov, v sondách dlohromady vystupuje více než 2 500 mluvčích (z toho zhruba polovinu tvoří mluvčí unikátní). Nejzásadnější změnou oproti stávajícím korpusům řady ORAL je to, že všechny nahrávky byly manuálně přepsány v transkripčním programu Transcriber, který umožňuje propojit přepis se zvukovou stopou. Uživatel si tak nyní může ve webovém roz-

hraní (v původním Bonitu 1 to není z technických důvodů možné) ke každému zobrazenému výrazu přehrát odpovídající část zvukové stopy a poslechnout si tak jeho skutečnou realizaci. Přestože jsme se snažili o pořízení co nejkvalitnějších nahrávek, ne vždy bylo možné vyhnout se ruchům a šumům přirozeného prostředí, které k neformálním komunikačním situacím a spontánnímu dialogu bohužel nutně patří.

Na rozdíl od předcházejících mluvených korpusů pokrývá ORAL2013 svými nahrávkami a přepisy celé území České republiky (tj. Čechy, Moravu i Slezsko). Tohoto kompletního pokrytí se podařilo dosáhnout díky spolupráci více než 200 přepisujících, pocházejících především z řad vysokoškolských studentů Filozofické fakulty Univerzity Karlovy v Praze, studentů pod vedením svých pedagogů z Univerzity Hradec Králové, Masarykovy univerzity, Západočeské univerzity v Plzni a Univerzity Palackého v Olomouci.

Při sběru dat jsme usilovali o maximální regionální pokrytí a různorodost mluvčích, materiál byl proto průběžně vyvažován v těchto hlavních sociolingvistických kategoriích: pohlaví, věk, vzdělání a oblast pobytu mluvčího v dětství. Na rozdíl od korpusu ORAL2008 jsme však nepřistoupili k závěrečnému výběru, jehož výsledkem by bylo plné vyvážení korpusu. Znamenalo by to totiž zbafování se cenného materiálu v situaci, kdy je korpus již dostatečně reprezentativní, zatímco jeho hypotetická „ideální“ vyváženost je přitom sporná (zvláště její regionální faktor) a není ani potřebná, protože nové webové rozhraní umožňuje práci s relativními frekvencemi.

Nový způsob přepisu měl dopad také na další změnu: tradiční syntaktická interpunkce použitá v předchozích korpusech řady ORAL byla nahrazena interpunkcí pauzovou. Rozlišují se celkem tři typy pauz (krátká pauza / delší pauza / odmlčení), které přepisující zaznamenávali podle individuálního tempa každého z mluvčích. Jednotlivé repliky jsou členěny na segmenty představující sémanticky, prozodicky i syntagmaticky ucelenou sekvenci v průměru o 5–10 slovech (maximálně však 15). Nově jsou zachyceny také simultánní úseky, kdy mluví dva mluvčí současně. Vlastní přepisovací pravidla navazují na koncepci pravidel předcházejících mluvených korpusů řady ORAL, byly pouze rozšířeny o popis toho, jakým způsobem zachycujeme vybrané specifické jazykové jevy vyskytující se na území Moravy a Slezska. Kompletní znění přepisovacích pravidel je k dispozici na stránkách <<http://korpus.cz>>.

Kvůli ochraně osobních údajů jsou veškerá příjmení a telefonní čísla v přepisech kódována anonymizačními zkratkami, kódování jiných vlastních jmen, jako jsou např. přezdívky, rodná jména, názvy firem, případně jiné citlivé údaje bylo ponecháno na vůli a přání přepisujících nebo samotných mluvčích. Anonymizovány byly samozřejmě i příslušné úseky v odpovídajících zvukových souborech.

Nově jsou v korpusu ORAL2013 označeni mluvčí, kteří vystupují ve více nahrávkách. Toto označení je uživatelům přístupné ve formě náhodně vygenerované „přezdívký“, která je pro daného mluvčího v celém korpusu stejná. Maximální počet slov jednoho mluvčího byl při sběru dat omezen na 15 tisíc textových slov, což při běžném mluvním tempu odpovídá přibližně jedné hodině monologického projevu.

Věříme, že se nový korpus stane důstojným pokračovatelem řady ORAL a vhodnou datovou základnou pro výzkum řady aspektů mluveného jazyka.

Pokyny pro autory

Redakce přijímá pouze původní, dosud nepublikované práce v češtině, slovenštině, angličtině nebo němčině. Pokud je anglický nebo německý příspěvek napsán nerodilým mluvčím, redakce autora vyzývá, aby svůj text nechal korigovat rodilým mluvčím-lingvistou.

Redakce žádá autory, aby své příspěvky ve formátu *.doc zaslali elektronicky na adresu **kga@ujc.cas.cz**. Zároveň je třeba na adresu redakce (**Korpus – gramatika – axiologie, Univerzita Hradec Králové, KČJL PdF, Rokitanského 62, 500 03 Hradec Králové**) zaslat jeden anonymizovaný vytištěný exemplář.

Povinnou součástí příspěvku psaného v jakémkoli jazyce je anglický abstrakt v rozsahu 100 až 150 slov a klíčová slova v jazyce příspěvku a angličtině. U anglicky nebo německy psaných příspěvků o češtině žádáme též název, abstrakt a klíčová slova v češtině.

Základní struktura příspěvku:

- název příspěvku;
- jméno a příjmení autora;
- pracoviště, e-mailová adresa;
- anglický abstrakt, klíčová slova;
- text;
- reference

Úprava příspěvku:

- zejm. pokud chcete do textu vkládat tabulky či grafy, doporučujeme nastavit už pro rukopis velikost stránky A5, popř. užít písmo Cambria vel. 9 pro základní text, vel. 8 pro ostatní texty;
- v případě tabulek a grafů je třeba zohlednit černobílý tisk časopisu (resp. tisk v odstínech šedi);
- příspěvek by měl být členěn na kapitoly (a event. subkapitoly); ty se označují čísly desetinné soustavy do maximálně čtyř číslic (1; 1.1; 1.1.2; 1.1.2.1); poté, je-li to nutné, lze užívat římských číslic (I, II) nebo písmen (A, a);
- doklady analyzovaných jevů je vhodné průběžně číslovat a číslo z obou stran uzavorkovat: (1); jednotlivé doklady se píší zkraje strany vždy v samostatném odstavci; pokud se jako doklad daného jevu uvádí několik vět (kvůli žánru, stylu, kontextu atp.), není nutno je číslovat, ale píší se v jednom bloku, shora i zdola vyčleněném z textu, jiným typem písma;
- doklady jsou na konci opatřeny citací zdroje; je-li doklad převzat z korpusu, stačí citovat tento korpus; má-li zavedenou zkratku, pak tuto zkratku, např. SYN2006pub;
- případná morfologická anotace příkladových vět by se měla dít v duchu pravidel Lipské konvence (Leipzig Glossing Rules).
- tabulky je třeba průběžně číslovat, opatřit názvem a umístit na stránce vlevo;
- poznámky jsou umístěny pod čarou;

- citace v textu: (Daneš, 1969, s. 86) / ...jak uvádí Daneš (1969, s. 86);
- úpravu referencí provádějte podle následujícího vzoru:

DUŠKOVÁ, L. (1984): Some contrastive notes on interrogative sentences in Czech and in English. *Philologica Pragensia*, 72–82.

BIRJULIN, N. A. – CHRAKOVSKIJ, V. S. (2001): Imperative sentences: theoretical problems. In: V. S. Chrakovskij (ed.), *Typology of Imperative Constructions*. München: Lincom Europa, 3–50.

ZIFONUN, G. – HOFFMANN, L. – STRECKER, B., et al. (1997): *Grammatik der deutschen Sprache*, Bd. 2. Berlin – New York: Walter de Gruyter.

Redakce si vyhrazuje právo příspěvek odmítnout nebo vyžádat jeho úpravu.

Příspěvky procházejí posouzením nejméně dvou nezávislých oponentů, kteří nejsou členy redakce.

Instructions for authors

The journal solicits only original, previously unpublished work. Articles can be written in Czech, Slovak, English, or German. Please be advised that if the article is written in a language that is not the author's native language, authors are asked to have their manuscript carefully checked by a native speaker.

Manuscripts should be submitted both electronically and in one hard copy. Please email the article as an attached file to the address **kg@ujc.cas.cz** and send one hard copy, without showing the author's name, to the postal address of the journal: **Korpus – Gramatika – Axiologie, Univerzita Hradec Králové, KČJL PdF, Rokitanského 62, 500 03 Hradec Králové**. The printed copy should be written in Microsoft Word, font size 12 pts., line spacing 1.5, and with sufficient margins.

Each submission must be accompanied by an abstract of 100–150 words, written in English.

The file with the electronic submission should contain the following:

- submission title;
- author's first and last name;
- affiliation and the author's email address;
- abstract in English;
- text of the article;
- references.

Formatting guidelines

The paper should be divided into **sections** (and, possibly, **subsections**). These are numbered in a decimal system, up to four levels (e.g. 1; 1.1; 1.1.2; 1.1.2.1). If additional subsections are needed, use Roman numerals (I, II) or letters (A; a).

Attested examples should be numbered sequentially, by using numbers enclosed in parentheses: (1). Examples are not indented but each example forms a separate paragraph. If an example consists of multiple sentences (for stylistic, contextual, etc. purposes), they are written in one contiguous sequence, not as separate (and numbered) paragraphs. The examples should be separated from the main text by a single line and typed in a different font, preferably in Verdana, size 9 pts, single spaced.

Each example is accompanied by reference to its source. If the example is taken from an electronic corpus, it is sufficient to name the corpus; if the corpus has a standard abbreviation, use the abbreviation, e.g. ČNK.

Tables must be numbered and each table must be accompanied by a caption. Tables are left-justified.

Footnotes are numbered sequentially and placed at the bottom of the page.

Literature cited in the text: (Daneš, 1969, 86) / ... Daneš (1969, 86)

Bibliographic references are formatted as follows:

DUŠKOVÁ, L. (1984): Some contrastive notes on interrogative sentences in Czech and in English. *Philologica Pragensia*, 72–82.

BIRJULIN, N. A. – CHRAKOVSKIJ, V. S. (2001): Imperative sentences: theoretical problems. In: V. S. Chrakovskij (ed.), *Typology of Imperative Constructions*. München: Lincom Europa, 3–50.

ZIFONUN, G. – HOFFMANN, L. – STRECKER, B., et al. (1997): *Grammatik der deutschen Sprache*, Bd. 2. Berlin – New York: Walter de Gruyter.

The editors reserve the right to reject unsuitable submissions and to request revisions. Each submission will be reviewed anonymously by two reviewers, not by the journal's editors.

Objednávka

Objednávka časopisu Korpus - gramatika - axiologie				
Odběratel	Název instituce:			
	Dodací adresa:			
	Fakturační adresa:			
	IČO / DIČ / NIP / VAT number:			
	IBAN:			
BIC (SWIFT):				
Objednávám	- zasílání kusu/kusů počínaje číslem			
	- jednorázové zaslání čísla v počtu kusu/kusů			
	s úhradou v měně:	Kč	€	USD
Další informace pro dodavatele:				
Dodavatel	Název instituce: Univerzita Hradec Králové, Pedagogická fakulta			
	Adresa: Rokitanského 62, 500 03 Hradec Králové, ČR			
	IČO: 62690094, DIČ: CZ62690094			
	Bankovní spojení: Česká spořitelna, a.s., číslo účtu: 2733582/0800			
	IBAN: CZ65 0800 0000 0000 0273 3582			
BIC (SWIFT): GIBACZPX				
Datum:	Za odběratele:			
<p>Objednávku odešlete na adresu redakce: Korpus – gramatika – axiologie, Univerzita Hradec Králové, Pedagogická fakulta, KČJL, Rokitanského 62, 500 03 Hradec Králové, Česká republika, nebo elektronickou formou na adresu: kga@ujc.cas.cz.</p> <p>Cena bude objednateli fakturována ve zvolené měně: 60 Kč/ks + poštovné a balné (vč. DPH), nebo 11 €/ks včetně poštovného, balného a bankovního poplatku (bez DPH), nebo 15 USD/ks včetně poštovného, balného a bankovního poplatku (bez DPH).</p>				

Order form

ORDER FORM for <i>Korpus – gramatika – axiologie</i>				
Subscriber	Institution:			
	Delivery address:			
	Billing address:			
	IČO / DIČ / NIP / VAT number:			
	IBAN:			
BIC (SWIFT):				
Ordering information	- regular delivery of copy/copies, starting with issue No.			
	- single issue No., number of copies			
	currency of payment:	Kč	€	USD
Additional information for distributor:				
Distributor	Institution: Univerzita Hradec Králové, Pedagogická fakulta			
	Address: Rokitanského 62, 500 03 Hradec Králové, Česká republika			
	IČO: 62690094, DIČ: CZ62690094			
	Bank: Česká spořitelna, a.s.; Account no.: 2733582/0800			
	IBAN: CZ65 0800 0000 0000 0273 3582			
BIC (SWIFT): GIBACZPX				
Date:	Subscriber signature:			
<p>Order form should be mailed to: Korpus – gramatika – axiologie, Univerzita Hradec Králové, Pedagogická fakulta, KČJL, Rokitanského 62, 500 03 Hradec Králové, Česká republika, or electronically to the following address: kga@ujc.cas.cz.</p> <p>The price on the invoice will be quoted in the currency selected by the subscriber: 60 Kč/copy + shipping and handling (incl. DPH), or 11 €/copy, including shipping, handling, and a bank transfer fee (without DPH), or 15 USD/copy, including shipping, handling, and a bank transfer fee (without DPH).</p>				