

Obsah

Články / Articles

Lexikon korpusu Orální historie (Příběhy)

Lexikon of Oral history corpus /3

Václav Cvrček, František Čermák, Lucie Chlumská, Jiří Mácha

Konkurenčné formy sekundárnych imperfektív v slovenčine

Competitive Forms of Secondary Imperfectives in Slovak language /27

Miloslava Sokolová

Příčinnostní vedlejší věty zájmenně-částicové

Pronominal-particle subordinate clauses expressing causality /46

Josef Štěpán

Zur Aspektualität des modalen Partizips I mit *zu* in attributiver Stellung (in kontrastiver Sicht)

On the aspectuality of the modal participle I with *zu* in the attributive position (a contrastive view) /57

František Štícha

Recenze / Reviews

Stefan Th. Gries:

Quantitative Corpus Linguistics with R /71

Lucie Chlumská

Zprávy / News

Corpus Linguistics 2011 – Discourse and Corpus Linguistics /77

Lucie Chlumská, Olga Richterová

Korpusová lingvistika Praha 2011 - Za vším hledej korpus /83

Lucie Chlumská

**Opolská konference o normě a antinormě ve slovanských jazycích
a mikrojazycích /88**

Josef Šimandl

Co je v ČNK nového I /93

Václav Cvrček

Pozvánka: Gramatika a korpus 2012 /95

Call for papers: Grammar & Corpora 2012 /97

Pokyny pro autory /99

Instructions for authors /101

Objednávka /103

Order form /104

Lexikon korpusu Orální historie (Příběhy)

Václav Cvrček, František Čermák, Lucie Chlumská, Jiří Mácha
Ústav Českého národního korpusu Filozofické fakulty UK, Praha
vaclav.cvrcek@ff.cuni.cz, frantisek.cermak@ff.cuni.cz,
lucie.chlumaska@ff.cuni.cz, jiri.macha.84@gmail.com

Lexicon of Oral history corpus

ABSTRACT: This paper is based on a study which was conducted within the research grant “Institutions in Life Stories. Multilevel Comparative Analysis of Biographical Narratives of Three Groups of Participants in Czech Society in 20th Century”. The aim of this research was both to describe one possible way of using a corpus to identify relevant differences between three types of text (in this case biographical narratives of three groups of speakers: communist officials, dissidents and so-called common people) and to serve as a basis for further analysis (be it a linguistic, sociological or historical analysis). We tried to point out typical features of the language of each group based on the most frequent expressions (nouns, adjectives etc.) and especially collocations. We also compared the corpus Příběhy (Stories) as a whole with the ORAL2008 corpus of synchronic spoken Czech, the SYN2005 corpus of synchronic written Czech and the Totalita corpus (a corpus of communist propaganda).

KEY WORDS: lexikon, keywords, collocations, oral history

KLÍČOVÁ SLOVA: lexikon, klíčová slova, kolokace, orální historie

Konkurenčné formy sekundárnych imperfektív v slovenčine

Miloslava Sokolová

Filozofická fakulta, Prešovská univerzita, Prešov, Slovensko

sokolova@unipo.sk

Competitive Forms of Secondary Imperfectives in Slovak language

ABSTRACT: The category of aspect is more abstract than *actio verbi*. They both form functional-semantic category of aspectuality as its central part. The primary stimulus of aspect formation is connected with the basic function of verbal aspect – to multiply the temporal-modal forms in aspect pairs. It is conjugation type of primary forms that is relevant for distribution of imperfectivization processes and forms. Primary forms are more frequent than secondary forms within aspect correlations. Aspect pairs represent 87 % of aspect correlations (1:1), nevertheless, there are also asymmetrical relations (1:2, 1:3, 1:4). Secondary imperfectives variants are influenced by grammatical, morphological, diachronic and extralingual factors. Slovak National Corpus (SNK) serves as basis for the evidence of co-variants, co-existence of forms and ending of form which lost its function. This dynamics is reflected in the tension between codification, represented by lexicons, and language practice, represented by SNK. The important indicator of variants existence is represented by participles.

KEY WORDS: Slovak language, verbal aspect, secondary imperfectives, co-variants

KLÚČOVÉ SLOVÁ: slovenčina, verbálny aspekt, sekundárne imperfektíva, varianty

Příčinnostní vedlejší věty zájmenně-částicové

Josef Štěpán

Peroutkova 1760/19, 150 00 Praha 5

stepan.josef@centrum.cz

Pronominal-particle subordinate clauses expressing causality

ABSTRACT: The article deals with one type of subordinate clauses expressing causality, namely pronominal-particle clauses in the present Czech. The incorporation of these clauses into the head clause lies in the fact that they express broadly understood causal meanings of adverbial subordinate clauses using primary and secondary prepositions and multiple-word units together with the pronoun *to*, which anticipates the subordinate clause introduced by the connecting particles *že*, *aby*. Data from the Czech National Corpus concerning subordinate clauses of cause (reason), purpose, condition and concession are analysed from the point of view of the Czech language (grammar) system.

KEY WORDS: anticipation pronoun; clauses of cause, purpose, condition, concession; connective particle

KLÍČOVÁ SLOVA: anticipační zájmeno; věta příčinná, účelová, podmínková, přípustková; spojovací částice

Zur Aspektualität des modalen Partizips I mit *zu* in attributiver Stellung (in kontrastiver Sicht)

František Štícha

Ústav pro jazyk český AV ČR, v. v. i., Praha

sticha@ujc.cas.cz

On the aspectuality of the modal participle I with *zu* in the attributive position (a contrastive view)

ABSTRACT: The semantic aspectuality of the German present participle (participle I) is not mentioned in Grammars of German, and in specialized studies it is taken for granted that the aspect of this sort of verbal derivative (e. g. *besuchende* 'visiting') is only imperfective. In Štícha (2009) it is shown that the German attributive participle I can also be used and interpreted in the meaning of the Slavic perfective aspect. In this article, a special attention is paid to the aspectual meaning of the modal usage of the present participle with the free syntactic morpheme *zu* (*zu besuchende* 'to be visited'). Examples of sentences containing this sort of modal attribute are selected from the corpus material (Corpus W - all free accessible corpora - of the Institute for the German Language in Mannheim, Germany) to show that the modal participle must be interpreted as perfective in most of its sentence usages. Some elementary statistics are added to strengthen the arguments.

KEY WORDS: aspect, perfective, participle, attribute, modal meaning, corpus linguistics

K aspektualitě modálního participia I s *zu* v atributivní pozici (z hlediska kontrastivního)

ABSTRAKT: Vidový význam německého modálního participia I v německých gramatikách není zmiňován a ve speciální literatuře se předpokládá, že tento druh verbálního derivátu (např. *besuchende* - *navštěvující/navštívící*) má význam imperfektiva. Ve Štícha (2009) bylo ukázáno, že atributivní participium I bývá v němčině užíváno a chápáno také ve významu slovanského perfektiva. V tomto článku se speciálně zabýváme vidovým významem modálně užívaného participia I s volným syntaktickým morfemem *zu* (*zu besuchende* - 'mající být navštíven'). Na korpusovém materiálu (prohledáván byl Korpus W - všechny veřejně přístupné korpusy - mannheimského Ústavu pro německý jazyk) se ukazuje, že ve většině případů modálního užití participia I s *zu* jde o význam perfektivní. Na podporu teoretických argumentů je připojeno několik elementárních statistik.

KLÍČOVÁ SLOVA: vid, perfektivní, participium, atribut, modální význam, korpusová lingvistika

Stefan Th. Gries: Quantitative Corpus Linguistics with R

Gries, Stefan Thomas: *Quantitative Corpus Linguistics with R: a practical introduction*. New York – London: Routledge 2009, 248 s.

Lucie Chlumská

Ústav Českého národního korpusu Filozofické fakulty UK, Praha

lucie.chlumska@ff.cuni.cz

Mohlo by se zdát, že učebnic a úvodů do korpusové lingvistiky už bylo sepsáno víc než dost. Je proto zcela namístě otázka „Skutečně potřebujeme další?“ A přesně tak se ptá v úvodu své knihy *Quantitative Corpus Linguistics with R* sám její autor, Stefan Th. Gries, a hned si také odpovídá. V čem je tedy tento úvod jiný? Předně se Gries na rozdíl od ostatních nezabývá, či jen velmi okrajově, dějinami korpusové lingvistiky, otázkami sestavování korpusů, jejich složením, reprezentativností či anotací, ani nevěnuje pozornost typům korpusů (v knize rozlišuje pouze dvojici anotovaný – neanotovaný) či výzkumům, které provedli jiní. Hlavním cílem knihy je slovy autora „to teach you *how to do corpus linguistics*.“ Tento smělý plán vyplývá ze samotného postoje Stefana Griesa ke korpusové lingvistice a její roli v rámci jazykovědy, což je velmi aktuální téma, které nyní plní nejen stránky časopisu *International Journal of Corpus Linguistics* (15:3, 2010). Pro Griesa je totiž korpusová lingvistika nikoli samostatným lingvistickým oborem (jako třeba sociolingvistika nebo konstrukční gramatika), ale hlavně *metodou* zkoumání, a tudíž se zabývá především „technickou“ stránkou věci, tj. tím, jak vytvořit frekvenční seznamy, jak uložit výsledky, jak je statisticky ohodnotit či jak z nich udělat graf. K tomu všemu mu slouží jediný software nazvaný jednoduše **R** (<http://r-project.org/>).

R je univerzální nástroj, který v sobě spojuje kalkulačku, statistický program, grafický program a vlastní programovací jazyk. Jeví se tedy jako ideální pomocník korpusového lingvisty, ale má to jeden háček. Aby člověk mohl **R** využívat se všemi jeho funkcemi, je třeba se naučit programovat. To jazykovědce nutně staví před další otázku: „Proč se učit programovat, když je k dispozici spousta konkordančních programů, které udělají většinu práce za mě?“ I na to má Gries hned v úvodu připravenou odpověď. V první řadě má člověk využíva-

jící **R** naprostou kontrolu nad tím, co vlastně s daty dělá (nelze nedodat, že pouze v případě, že to s **R** skutečně **umí**). Není tedy odkázán na případná omezení či změny v nastavení konkordančních programů (jak Gries uvádí, výsledky téhož zkoumání se mohou například v MonoConc Pro 2.2, AntConc 3.2 a WordSmith Tools 4.0 podstatně lišit). S tím úzce souvisí další argument, že programovací jazyk je mnohem univerzálnější nástroj než jakýkoli konkordanční software. Pro srovnání: ačkoli český korpusový vyhledávač Bonito dokáže leccos, bohužel uživateli neumožňuje pracovat s vlastním korpusem. Naproti tomu **R** lze použít prakticky na jakýkoli, tedy i doma sestavený korpus. Další nespornou výhodou **R** jistě je, že je zcela zdarma, což může pro mnohé být jedním z rozhodujících faktorů, a funguje bez problémů pod operačním systémem MS Windows, Linux i Mac OS.

Jistým uklidněním pro všechny neprogramátory, které představa černé obrazovky a řádků plných nejrůznějších symbolů děsí, je, že napsat skript je většinou nutné jen jednou a pak jej lze s úspěchem využívat znovu. Naopak lidé, kteří zkušenosti s programováním mají, by se mohli právem ptát, proč se učit nový programovací jazyk, když lze ke stejnému účelu použít třeba Perl (viz například Danielsson 2004). Jak již bylo uvedeno výše, **R** je na rozdíl od Perlu vybaven mnoha funkcemi pro statistické vyhodnocení a grafické znázornění výsledků, tudíž lze veškerou korpusovou práci provádět v rámci jednoho programu. To se velmi hodí například při práci s opravdu velkými korpusem, a tedy velkým objemem dat, jelikož textové editory typu Microsoft Excel či jeho protějšek Calc v OpenOffice.org jsou na to příliš pomalé a velké množství dat většinou v praxi nedokážou zpracovat. Výhody **R** můžeme demonstrovat na příkladu funkce `table`, která umožňuje jednoduché vytvoření frekvenčního seznamu, což je pro korpusového lingvistu obzvláště užitečné. Porovnáme-li jednoduchý příkladový úkol (frekvenční seznam 8 náhodných slov) v **R** a Perlu, vidíme, že v **R** je to nesporně jednodušší, přestože Perl patří mezi ty úspornější programovací jazyky:

R:

```
frequency<-c("in", "the", "bed", "in", "the", "room", "is", "the")
sort(table(frequency))
```

Perl:

```
@frequency=("in", "the", "bed", "in", "the", "room", "is", "the");
foreach $string (@frequency)
{
  ++$counts{$string};
}
foreach $key (sort {$counts{$a} <=> $counts{$b}} keys %counts)
{
  print "$key \t $counts{$key} \n";
}
```


Nemluvě o tom, že mnoho funkcí a příkazů, které se člověk naučí používat v *R*, lze uplatnit i v ostatních programovacích jazycích (například regulární výrazy, které člověk potřebuje znát pro tvoření složitějších dotazů i v českém korpusovém manažeru Bonito). Výtka týkající se rychlosti *R* také není namístě, pro naprostou většinu úkolů je tento software dostatečně rychlý. Sám Gries uvádí, že za svou akademickou dráhu byl nucen Perl kvůli rychlosti použít jen několikrát. Každý korpusový lingvista tedy stojí před rozhodnutím: buď se spokojí s uživatelsky přívětivými korpusovými manažery a konkordančními programy za cenu, že v nich většinou nemůže příliš ovlivňovat nastavení jednotlivých funkcí a výchozí definice, se kterými program pracuje, nebo věnuje víc času a úsilí a naučí se programovací jazyk, který toho dokáže víc a poskytuje uživateli větší svobodu i pro práci s vlastními korpusy. Gries dělá se svou knihou všechno proto, aby zvítězila druhá možnost.

Už při zběžném pohledu na strukturu knihy *Quantitative Corpus Linguistics with R* je jasné, že nejde o publikaci, kterou by si člověk četl pár minut před spaním. Tato učebnice (což je asi nejvhodnější označení) je totiž skutečně interaktivní, vyžaduje aktivní přístup čtenáře a využívá všech výhod, které internetová doba nabízí. V knize najdeme odkaz na webovou stránku, ze které lze stáhnout veškerá data, se kterými se v knize pracuje (soubory s ukázkami korpusů, skripty, příkladová cvičení i řešení úkolů). Tato stránka byla původně k nalezení na adrese groups.google.com po zadání „CorpLing with R“ do vyhledávání, ale v říjnu 2010 se na této stránce objevilo oznámení, že se obsah kvůli problémům s Google Groups přesouvá na následující adresu: <http://www.linguistics.ucsb.edu/faculty/stgries/research/qclwr/qclwr.html>. Pro stažení příslušných souborů je zapotřebí znát hesla, která jsou uvedena v knize na s. 20. Ke stažení jsou také více či méně specializované balíčky funkcí vytvářené uživateli, např. gsubfn od Gabora Grothendiecka, jehož užitečná funkce gsub umožňuje zaměnit v celém textu určitý prvek za jiný (např. při čištění anotovaného korpusu od tagů).

Knihy je členěna do šesti kapitol, přičemž její těžiště spočívá ve třetí a čtvrté kapitole. Pomineme-li úvodní kapitolu, ve druhé jsou čtenáři osvětleny základní pojmy korpusové lingvistiky a tři možnosti zkoumání korpusů, kterými se Gries v knize zabývá: frekvenční seznamy, kolokace a konkordance (jako prostředek zkoumání kontextu). Všechny kapitoly doplňují šedé rámečky, v nichž autor vybízí k dalšímu přemýšlení a studiu (hned v úvodu například pokládá zdánlivě banální otázku, co je slovo a jak by se dalo definovat pro počítač, přičemž vzápětí se zálučně zeptá, jak by pak člověk řešil případy jako *better-suited*, *armchair linguist/armchair-linguist* či *25-year-old*.) Kromě těchto úkolů v knize najdeme tzv. pauzy na přemýšlení (*think breaks*), které vyžadují, aby čtenář zpomalil a utřídil si myšlenky. Jak již bylo zmíněno, tuto knihu nelze číst **jen tak**, je nutné všechny příkazy hned zkusit u počítače, bez toho *R* prakticky nelze pochopit. Gries si je vědom toho, že nejlépe se člověk naučí programovat **programováním**, proto počítá s chybami a někdy je i využívá ve cvičeních (na

s. 76 vybízí k provedení určitého příkazu a hned vzápětí se ptá, proč to nefunguje). Je třeba mít na paměti, že v **R** na rozdíl od hotových konkordančních programů či vyhledávačů neexistuje „tlačítko“ zpět, je proto více než doporučením, aby si uživatel zapisoval skript do některého z textových editorů, např. Tinn-R, které umožňují nastavit barevné zvýraznění syntaxe daného programovacího jazyka.

Kapitola 3 (*An Introduction to R*) je pro laika zásadní – učí uživatele pracovat se softwarem. Gries zde vysvětluje pojmy jako *vector*, *data frame*, *list* či *loop* (smyčka) a učí zacházet s regulárními výrazy, přičemž nepředpokládá jakoukoli vstupní znalost. (Pro české čtenáře je namísto podotknout, že kniha je napsána velmi přístupným způsobem, a to jak obsahově, tak co se jazyka týče. Může ji využít každý s pokročilejší znalostí angličtiny.) Mnoho funkcí, jež se v **R** používají, je společných i pro ostatní programovací jazyky, např. *ls*, *grep*, nebo podmínkové výrazy *if*, *else* či *for*. Hlavní příkazy, které Gries zdůrazňuje nejen v této kapitole, ale v celé knize (např. na s. 68), i na svých workshopech věnovaných **R**, které pravidelně pořádá v Texasu, je „*Know thy corpus!*“. Neexistuje důležitější zásada. I když se člověk rozhodne pro předpřipravené konkordanční programy, musí být velmi dobře obeznámen s charakterem a strukturou korpusu, s nímž pracuje. Je anotovaný, lemmatizovaný, taggovaný? Pokud ano, jakou formu mají tagy? Bez těchto znalostí prakticky nelze žádný software používat. U **R** to platí dvojnásob. Na konci třetí kapitoly už tedy čtenář ví, např. jak vyčistit korpus od nepotřebných tagů, jak změnit anglický formát data na americký (nebo jakýkoli jiný) či jak ukládat a komprimovat různé datové struktury.

V kapitole čtvrté se konečně dostáváme od základů programování k jejich praktickému využití v korpusové lingvistice. Čtenář se například naučí vytvořit frekvenční seznam slov či dvojic slovo – tag v rámci anotovaného i neanotovaného korpusu. Na některých příkladech je patrné, že Gries se skutečně zaměřuje především na *metodu*, ale už ne na samotný účel práce. Na s. 118 podrobně vysvětluje několik způsobů, jak z korpusu získat dvojice po sobě následujících slov, ale vůbec se nezabývá tím, proč je zrovna tato dovednost pro korpusového lingvistu užitečná. Předpokládá se tedy, že knihu dostane do ruky člověk, který ví přesně, jaká data potřebuje, jen se chce naučit nový způsob, jak se k nim snadno a efektivně dostat.

Tím nejzajímavějším v knize jsou však větší či menší případové studie. Mezi ty jednodušší patří vyhledání všech potenciálních předložkových sloves (v Quirkově terminologii slovesa, jež musí předložce bezprostředně předcházet) s předložkou *in* (s. 142) nebo postup, jak nalézt v korpusu BNC všechna tranzitivní frázová slovesa, tedy ta, v nichž předmět může stát před nebo za částicí frázového slovesa: *Mary gave away the secret.* / *Mary gave the secret away* (s. 143). Hlavní využití v této studii má, kromě tagů a regulárních výrazů, užitečná funkce *exact.matches*, jejímž autorem je sám Gries. Tímto příkazem dostaneme seznam (*list*) skládající se ze 4 částí (nejdůležitější je vektor, který nám

najde „exact matches“, tedy přesně ty výrazy, jež hledáme, ale bez kontextu, dále vektor s informací, kde se tyto matches nacházejí, a konečně vektor obsahující celý příslušný řádek, tedy exact match i s kontextem). Naproti tomu postup, jak získat z korpusu kolokace v daném okruhu klíčového slova, který zabírá 7 stran (s. 149–156), může leckoho odradit. Je třeba podotknout, že v této studii Gries vůbec neuvažuje o kolokačních statistických mírách, pouze představuje postup, jak v tabulce shromáždit slova vyskytující se v okolí slova klíčového i s jejich frekvencemi. Je jistě chvályhodné, že Gries pamatuje i na jazyky, které využívají jinou sadu znaků než angličtina, tj. cyrilici nebo čínské znaky (viz oddíl Unicode, s. 166).

Předposlední kapitola věnovaná statistickým metodám budí dojem, že autor sice nechtěl toto podstatné téma opomenout (*R* vznikl původně jako statistický software), ale zároveň věděl, že mu věnuje celou svou příští knihu (*Statistics for Linguistics with R*, 2010). V této poněkud hutnější kapitole tedy nabízí jen stručný teoretický úvod do proměnných a formulování hypotéz a poté věnuje pozornost například Pearsonovu testu chi-kvadrát a vizualizaci výsledků. V této souvislosti zmiňme možnosti vizualizace známé jako pie, boxplot nebo barplot, které umožňují jednoduše vytvořit různé druhy grafů podle zadaných parametrů. Dalším statistickým testem, o kterém se Gries zmiňuje, je t-test. K tomu, abychom zjistili, zda naše data splňují podmínky pro použití tohoto testu, slouží v *R* dva další testy: Shapiro-Wilk test a Ansari-Bradley test. První z nich ověřuje, zda se data liší od normální distribuce, druhý, zda mají stejný rozptyl. Pakliže podmínky splněny nejsou, t-test není vhodným kandidátem. Náhradou za něj může v takovém případě být neparametrický U-test (funkce `wilcox.test`). Kapitola jako celek představuje pouhý úvod a čtenář patrně bude nucen dohledat si relevantní statistické informace jinde (třeba v Griesově nové knize nebo u Baayena, 2008).

Poslední, krátká kapitola, jež je věnována případovým studiím (které v knize ovšem nenajdeme, Gries zde odkazuje na výše zmíněnou webovou stránku) a dalším aplikacím, opět potvrzuje Griesovo pojetí korpusové lingvistiky jako metody a představuje její využití v psycholingvistice, kognitivní lingvistice a dalších odnožích lingvistiky.

Závěrem lze říci, že kniha plní slib, který Gries čtenářům v úvodu dává. Učí uživatele, **jak** dělat korpusovou lingvistiku, přičemž důraz je zde kladen na metodu zkoumání. Kniha funguje jako výborný manuál softwaru *R* a úvod do programování, ale už méně jako klasický úvod do korpusové lingvistiky. Gries se příliš nezabývá korpusovým pohledem na jazyk, namísto toho představuje jednu z metod, kterou lze pro korpusový výzkum použít. Je tedy primárně určena především těm lingvistům, kteří mají jasnou představu o tom, co chtějí zkoumat a proč. Je nesporné, že obecné znalosti základů programování (např. již mnohokrát zmiňované regulární výrazy a jejich použití) jsou pro korpusového lingvistu více než užitečné a neomezují se zdaleka jen na software *R*. Stejně tak bez

statistiky a především jejího praktického využití se **dobry** korpusový lingvista neobejde. Na trhu sice najdeme statistické knihy (např. Oakes, 1998 či obecnější Volín, 2007), které poskytují kvalitní a fundovaný teoretický základ jednotlivých statistických postupů a metod, ale málokterý autor dokáže na konkrétních příkladech ukázat jejich skutečné využití v praxi tak, jako to dělá třeba Baayen či právě Gries. Sečteno, podtrženo, Griesovu knihu nejvíce ocení ti, kteří se nebojí učit nové věci a kteří chtějí mít svůj korpusový výzkum pevně v rukou.

LITERATURA

BAAYEN, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

DANIELSSON, B. (2004). Simple Perl programming for corpus work. In *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, p. 225–246.

GRIES, Th. S. (2009). *Quantitative Corpus Linguistics with R*. London: Routledge.

GRIES, Th. S. (2010). Materiály z Bootcampu „Corpus Linguistics with R“, 8.–15. 8. 2010 (prezentace autora knihy a poznámky recenzentky).

OAKES, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

VOLÍN, J. (2007). *Statistické metody ve fonetickém výzkumu*. Praha: Nakladatelství Epoque.

Corpus Linguistics 2011

Discourse and Corpus Linguistics

Lucie Chlumská, Olga Richterová

Institute of the Czech National Corpus, Prague

lucie.chlumska@gmail.com, richterova.olga@gmail.com

On 20–22 July 2011, one of the largest and most prestigious corpus linguistics conferences, *Corpus Linguistics 2011*, convened in Birmingham. This conference takes place every two years and is consecutively hosted by three prominent British universities. This year, more than 250 participants attended the conference with its 3 plenaries, 137 full papers, 16 "pecha kuchas" (short presentations of 20 slides, each shown for 20 seconds – thus giving 6 minutes 40 seconds to each presenter) and 35 posters. Furthermore, 2 colloquia (on *Professional discourse and corpora* and on *Recent Perspectives on Multi-Dimensional Analysis*) and 5 workshops were on offer. The wide selection of different formats enabled matching the topics presented with the form which fitted them best and at the same time preserving what works best for the majority.

The distinctiveness of this year's conference, however, is in a different thing: it was the first time the organizers assigned the whole conference a general theme: *Discourse and Corpus Linguistics* (although the 7 parallel sections were not assigned individual titles). Some presenters followed it, for others the given topic was too narrow. The question, whether providing an overarching theme was rather harmful or helpful, can be discussed in detail elsewhere. It will suffice to say here that many presenters focused in their research on professional/academic discourse and used the (in most cases easily available) data for analyses, which will hopefully in future be expanded to other registers (discourses) too.

The broad variety of topics discussed at the conference can be illustrated by the work represented and presented by the three plenary speakers invited to Birmingham. Instead of praising the organizers for smooth running of the

event or describing the conference venue (the International Conference Centre in Birmingham), let us now focus on these contributions.

Susan Hunston held a talk entitled "*Flavours*" of *corpus linguistics: the case of evaluative language*, and she based it on her most recent monograph: *Corpus Approaches to Evaluation: Phraseology and Evaluative Language* (Routledge 2011). Unfortunately, no handouts were distributed and the power-point presentation was not made available to the attendees, thus we must apologize for possible misinterpretation of her ideas.

The gist of her talk can be summed up as an effort to formulate what the parameters are that constitute the 'flavours' mentioned in the title and thus to offer one possible way of conducting corpus-based linguistic analysis. At the same time, she takes into account the criticism that over-reliance on corpus evidence often leads to de-contextualization due to shorter contexts; the whole of texts is what discourse analysis stresses.

The flavours – parameters or dimensions of achievement – were labelled as **extent** (represented e.g. by word frequency lists), **reach** (an attempt at answering the question what constitutes an item and what the unit of analysis is), **depth** (the detail a study goes into), and **strength** (the implicative power of methodology). Hunston suggested these four criteria as a way of looking at methodology dilemmas and exemplified her approach on the example of investigating how modal meaning is expressed. Reach enables capturing the so-called phrasal intensifiers (e.g. *it's up to you*) and can be seen especially in quantitative studies, depth is a dimension often overlooked in other than qualitative studies since modal-like expressions often function in a way similar to regular clauses (e.g. *for fear of being thought disloyal*). The overall conclusion was that factoring in extent, reach and depth can help achieve strength. Ms. Hunston's talk was well-structured and thought-provoking, yet the above-mentioned lack of handouts made it more difficult to further discuss her proposals.

Paul Baker stimulated discussion with his contribution on *Discourse, news representations and Corpus Linguistics*, which can be downloaded here: <http://www.ling.lancs.ac.uk/stories/1322/>. The presentation was partly based on the research conducted in cooperation with C. Gabrielatos and A. McEnergy for the forthcoming publication *Discourse Analysis and Media Bias: The representation of Islam in the British Press* (Cambridge 2010).

At the beginning of his talk, Mr. Baker reminded the audience of the horrible vagueness of the term discourse and of the fact that corpus-driven techniques serve to reduce cognitive bias (represented e.g. by the primacy effect or the clustering illusion). The main part of his plenary focused on a nine-stage corpus-assisted critical discourse analysis, while the analysis of the representation of Islam in the British press was taken as an illustration of the proposed approach. 'Critical

discourse analysis' was understood as an ideologically-driven approach enabling to factor in sociological phenomena (corpus-based analysis substantiating the work might have seemed secondary at times).

Probably the most memorable part of his talk was an overview of an inter-analyst consistency experiment conducted on five persons. Experienced as well as novice researches were given the task of finding out how "foreign doctors" were represented in the British press in the period between 2000–2010. They were all provided with the same corpus, but the methods they chose to use were entirely up to their discretion (they applied concordance analysis, or identified collocates, key semantic categories, dispersion plots, clusters, keywords with the help of a reference corpus or diachronic keywords).

The findings of all five researches overlapped in one feature only (whereas the overall 'feel' of the data was interpreted in the same way, though). This was presented as an argument aimed at refuting the myth of objectivity and meant to bring us to reflect on our own stances as researchers (the presenter admitted to having a 'Guardian reader perspective' himself).

On the third day, Stefan Th. Gries presented his view of a statistician-turned-linguist in: *Quantitative and exploratory corpus approaches to registers and text types*. His presentation was well-structured, comprehensive and for a non-statistician mind rather difficult; he is possibly the record holder of the number of words per slide and the number of slides per presentation at the conference. Nevertheless, he tried to make his point clear using graphs and definitions and he is open to sending his presentation to anyone interested. In his speech, Gries addressed the dichotomy of qualitative vs. quantitative research that is so commonly made in many linguistic disciplines claiming that qualitative work is ultimately based on quantitative work and, thus, needs to take lessons from quantitative methods into consideration.

Gries also discussed how the study of any linguistic phenomenon in corpora should feature bottom-up analytical steps in order to identify the most promising registers/text types for a particular linguistic phenomenon. He presented several examples, e.g. the extraction and interpretation of n-gram frequencies in the BNC Baby showing the differences among genres and text types. Finally, he tried to show the similarities between corpus linguistics in general (register/text type-based studies in corpus linguistics in particular), and psycholinguistic theories.

Among the papers presented at the conference were topics as varied as *Key words, pivotal words and leading words in "Harry Potter"* and *Quantitative measures of productivity and their significance*. The popularity of Ms. Rowling's book was reflected in the fact that yet another paper drew on her work: *Why Hermione is not the Hero: Using Corpus Methods to Analyse the Discoursal Representation of Female Physicality in Children's Literature*. Some papers dealt

with discourse and some did not even mention the notion, some paper titles were rather simple (paper *Ageing with the corpus* focusing on the stereotypes in depicting young/old women and men using corpus data), some clearly statistical (*Exploring Zipfian Distributions in English Verb Argument Constructions: Corpus and Psycholinguistic Evidence*). Besides that, many papers showed the frustration of teachers striving to make their students acquainted with corpora. A substantial number of the contributions dealt with translations and with language comparison, too.

Since we have neither space nor desire to describe the whole conference, let us now just provide you with a number of glimpses into what we found interesting. Although any selection is clearly subjective and based on the preferences of the authors, we will try to show you the variety of topics presented at the conference.

Research ethics in corpus linguistics was the topic Tony McEnery addressed at the conference twice (in order to satisfy those who did not fit in the room, the presentation was repeated the next day). He spoke about compiling the spoken part of the BNC and the issues connected with securing sensitive information of the respondents and conducting anonymization. Apart from ethics and **respondents**, he also mentioned ethics and **corpus builders** (the example he used was EMILLE corpus, which was offered a vast amount of texts to be used for its parallel corpus; the researchers had to decline these texts since a religious organization wanted to use EMILLE as a tool for spreading its views.). Furthermore, ethics and **corpus distributors** was addressed from the perspective of copyright legislation (may corpora become illegal in certain jurisdictions?) and ethics and **corpus users** was the concluding point suggesting that even well-meaning linguistic analysis (e.g. of teenage language) may be hijacked by the media. The question to be asked in this connection is: How will others interpret my results?

Re-defining subtext as logical form was the name Bill Louw gave to his presentation drawing on the work of several philosophers and finding inspiration in poetry. He argued that corpus linguistics needed a deeper theoretical basis (according to Louw, it is a paradigm-shifting method and not just a methodology, which is the view e.g. S. Th. Gries holds) and showed the results of **finding quasi-propositional variables**. Basically, Louw substituted lexical items in certain clauses by asterisks – This is no country for old men / This is no * for * (from a poem by Yeats, on Easter uprising) – and analysed the results. This method helped him locate the hidden subtext and the notions of subtext, semantic prosody and contextual prosodic theory could all find their application.

One of the Czech researchers, Václav Cvrček, presented an innovative paper on *How Large is the Core of Language?* His research was motivated by the fact that although lexicographers consider a general corpus a sufficient source

for representation of core lexical units, there are still lexemes missing both in dictionaries and, less frequently, in corpora (peripheral lexemes, scientific terms etc.). The question Cvrček poses is how to determine the size and range of the core of language by using corpus data and methods. Cvrček assumes that the core vocabulary is the part of lexicon common to the majority of texts and speakers. He uses the proportion of hapax legomena to all word-types in relation to the growing corpus size to identify the frequency range in which core elements occur, and presents a pipe-resembling graph of the hapax-type ratio (the core of the language becoming clear in the first part of the graph). What is interesting is that this graph has a similar shape in different languages regardless of the types of texts or their order; Cvrček tested this finding on corpora of Czech, English and Italian.

Translation Universals Revisited by Lukasz Grabowski was a very interesting presentation, particularly for researchers focusing on translation and contrastive linguistics. Grabowski focused on several translation universals described in detail in the work of Mona Baker or Sara Laviosa: explication, simplification, normalization, sanitization and levelling-out. In his research based on the comparison of translated and non-translated Polish texts, he found out among other things that lexical richness is, contrary to general expectations, higher in translated texts, whereas sentences tend to be longer in non-translated texts.

From the other presentations, let us just mention in passing Svetlana Vetchinnikova and her *Semantic prosody as a communicative function of a unit of meaning* or Costas Gabrielatos (*If-conditionals and modality: A corpus-based investigation*). Alex Boulton discussed the topic *corpora in translation for non-translation students* and Matthew Brook O'Donnell & Catherine Smith presented a poster entitled *The Adjusted Frequency List: A new method to extract cluster-sensitive frequency lists from corpora*. Theirs is the last research we would like to go into here. Its main objective was to show how little importance do single words bear. What matters are n-grams, occurrence in structures. The poster showed a method for discovering non-accidental n-grams and evidently showed a much more representative picture of the language register analysed, than the previously shown frequency analysis of individual words.

The conference was of course a networking event, too. The suggestion to create a website and a mailing list of PhD students working with corpora was appreciated by the about thirty representatives of this group present. Students were asked to submit their e-mail addresses and they were told that the platform fostering intra-PhD discussion should be made available within a month. This idea of bringing PhD students of corpus linguistics and related disciplines together is very promising; it will help young researchers connect and share any possible troubles as well as successful results.

To sum up, this year's conference was a great opportunity to share ideas and methods as well as outcomes and results within the broad discipline of corpus linguistics. The organizers in Birmingham did a great job; the conference was well-organized with an inspiring and varied programme. We can only look forward to the next conference in a two-year's time, which will take place in Lancaster, 23–26 July 2013.

Korpusová lingvistika Praha 2011

Za vším hledej korpus

Lucie Chlumská

Ústav Českého národního korpusu Filozofické fakulty UK Praha

lucie.chlumaska@gmail.com

Ve dnech 22.–24. září se na Filozofické fakultě Univerzity Karlovy konala konference Korpusová lingvistika Praha 2011. Jak již sám název napovídá, stěžejním tématem konference byly jazykové korpusy – od jejich výstavby, značkování či anotace přes konkrétní využití ve výzkumu gramatiky a slovníku až po úlohu korpusů v lexikografii a při překladu; zahrnuty byly i korpusové aplikace (např. nástroj na výzkum variability SyD). Konference se během tří dnů zúčastnilo více než 90 účastníků z 25 různých akademických pracovišť z České republiky, ale i ze zahraničí. Za všechny jmenujme následující instituce: Akademie věd ve Vídni, The University of Iowa, Universita Granada, Università di Napoli, ÚLŠ SAV Bratislava, Masarykova Univerzita v Brně, Universita Regensburg, University of Leeds či Varšavská univerzita. Celkem bylo předneseno 67 příspěvků, z toho tři plenární přednášky.

Prezentace byly rozděleny tematicky do několika dílčích sekcí, např. Korpusy a značkování, Slovtvorba, Morfologie a gramatika, Lexikon, kolokace, slovní druhy, Syntax či Aspekty výzkumu korpusů. Vždy probíhaly paralelně tři prezentace z různých sekcí (kromě plenárních přednášek), takže si účastníci mohli zvolit příspěvek podle svého zájmu a zaměření. Jako snad na každé konferenci, i zde se stalo, že někteří přednášející nedodrželi stanovený časový limit (20 minut na příspěvek a 10 minut na diskuzi). Paradoxní je, že větším problémem nebyl delší, ale kratší příspěvek, jelikož pak se stalo, že všechny následující prezentace v dané sekci začaly o něco dříve a účastníci, kteří přebíhali z jedné sekce do druhé, tak nestihli začátek prezentace. K tomu však došlo pouze v jediném bloku, ve všech ostatních směrech šlapala konference jako hodinky a organizátorům nebylo co vytknout.

O plenárních přednáškách lze v první řadě říci, že se od sebe tematicky i formálním pojetím velmi lišily, ale všem se podařilo shrnout důležité poznatky, přinést nové podněty nebo alespoň podnítit diskuzi. První plenární

přednášky se ihned po zahájení konference zhostil ředitel Ústavu Českého národního korpusu František Čermák. Ve své prezentaci na téma *Korpusy včera, dnes a zítra* přehledně a systematicky zmapoval vývoj korpusů a korpusové lingvistiky od samotného počátku (Včera: *Survey of English Usage* Randolfa Quirka, první korpus *Brown Corpus* aj.) přes etapy vývoje velkých elektronických korpusů (Dnes: stamilionové až miliardové korpusy v Mannheimu, Praze a jinde) až po možnosti a desiderata korpusové lingvistiky v budoucnu (Zítra: problém dostupnosti dat, a to především u mluvených korpusů, otázka potenciality jazyka a vztahu korpusu k jazykové realitě). Čermák také shrnul typy korpusů a jejich charakteristiky, upozornil na problémy, které souvisejí s výstavbou korpusů (data, finance, reprezentativnost, vyváženost apod.) a neopomenul ani věnovat několik slov korpusové lingvistice jakožto samostatné lingvistické disciplíně (v protikladu k druhému pohledu, který považuje korpusovou lingvistiku za pouhou metodu).

Druhou plenární přednášku s názvem *Do We Need Very Large Web Corpora?* přednesli Karel Pala a Pavel Rychlý z brněnské Masarykovy univerzity. Ve svém příspěvku se pokusili odůvodnit nutnost budovat velké korpusy textů získaných na internetu. Uvedli několik příkladů, např. srovnání kolokací sloves *vidět* a *see* na stomilionových korpusech SYN2000 a BNC a dvoumiliardových korpusech enTenTen a CzTenTen. Srovnání potvrdilo základní domněnku, že čím větší korpus, tím více slov v něm lze nalézt a tím přesnější mohou být zkoumané kolokační rámce. Autoři dále věnovali část prezentace technické stránce budování webových korpusů a jejich obhajobě. Na úvodní otázku si tak sami odpověděli jednoznačným Ano, ačkoli hlasy z publika při diskusi upozornily na to, že samotná velikost korpusu nestačí, je třeba se pečlivě zamýšlet nad složením textů a v tomto ohledu nepředstavují webové korpusy o nic menší problém než korpusy klasické, ba možná naopak.

Třetí přednášející František Štícha svou originální prezentací obohatil korpusové pojmosloví o nový výraz, hovořil totiž se špetkou nadsázky o „kvantové lingvistice“. Ve své přednášce s názvem *O skrytosti langue v korpusovém parole aneb k definici „kvantové lingvistiky“* se pokusil nastínit princip „kvantových pravidel“, tedy pravidel, jež se postulují na základě kvantitativní analýzy velkého množství dat (kvant doložených jazykových struktur), která nelze získat jinak než pomocí korpusu a jeho nástrojů. Na příkladu slovosledné struktury (*je* + adverbium + adjektivum + infinitiv v. adverbium + *je* + adjektivum + infinitiv) ukázal, že zkoumáním velkého množství korpusových dat lze objevovat zákonitosti, které jsou introspekci nebo jinými nekorpusovými metodami nepostřehnutelné. Některé hodnoty jazykového systému, „langue“, lze pak podle Štíchy zjišťovat v „parole“ pouze pomocí statistických metod aplikovaných na „kvanta“ studovaných jevů.

Tolik k plenárním přednáškám, nyní alespoň krátce k dalším tématům. Ačkoli bylo na konferenci předneseno mnoho zajímavých příspěvků, není možné

se zde věnovat všem. Následující výběr lze tedy považovat za jakousi ochutnávku pestrého programu konference.

Aleš Klégr (i za Markétu Malou a Pavlínu Šaldovou) představil téměř dokončený projekt věnovaný zkoumání anglických nepředložkových ekvivalentů českých předložek na datech z paralelních korpusů. Cílem projektu je vytvořit materiálovou bázi pro účely aplikované lingvistiky a lexikografie. Předložky patří k nejužívanějším, avšak často neprávem opomíjeným slovním druhům v obou zkoumaných jazycích, přičemž nepředložkové ekvivalenty českých předložek tvoří více než 27 % překladatelských řešení. Tyto nepředložkové ekvivalenty lze dále rozdělit na několik podskupin, jimž se Klégr v příspěvku podrobně věnoval. Prezentace byla velmi zajímavá především pro jazykovědce zabývající se kontrastivní lingvistikou a paralelními korpusy, ale i pro překladatele. Dobrou zprávou pro všechny případné zájemce je, že jedním z výstupů projektu bude monografie věnovaná tomuto tématu.

Pavel Vondříčka (i za Václava Cvrčka) představil nový korpusový nástroj na výzkum variability nazvaný SyD (pro **s**ynchronní i **d**iachronní zkoumání). Tento nástroj je volně přístupný na adrese <http://syd.korpus.cz/> a na rozdíl od korpusů ČNK nevyžaduje žádnou registraci. Je uživatelsky velmi příjemný, přehledný a jednoduchý (především díky velkému množství grafů, tabulek a jiných grafických výstupů). Umožňuje porovnávat dvě až osm variant slova či tvaru (např. *bychom* versus *bysme*), a to nejen v rámci psaných korpusů ČNK, ale také mluvených a diachronních. Pomocí nástroje SyD je tedy možné např. sledovat, jak se vyvíjelo užití slova *veliký* a *velký* v průběhu staletí, nebo jak se liší užití výrazů *pořád*, *stále* a *furt* v mluveném a psaném jazyce a v různých žánrech. Kromě obsáhlých informací o textových typech, žánrech, procentuálním zastoupení v rámci let i staletí lze v nástroji SyD porovnávat i další statistické údaje i kolokace variant. SyD představuje vítané obohacení korpusových nástrojů a má v sobě velký potenciál pro využití např. ve školách či při překladu.

Na konferenci zaznělo několik příspěvků o budování korpusů, ať už paralelních (projekt InterCorp), žákovských (CZESL, ROMi) nebo specializovaných (Korpus PSP ČR). Zmíňme se jen krátce o nejdůležitějších poznatcích. Projekt InterCorp oslavil sedmileté výročí a Michal Křen s Alexandrem Rosenem ve svém příspěvku za celý pracovní kolektiv shrnuli dosavadní výsledky a největší změny. V současnosti je v paralelním korpusu InterCorp k dispozici 22 evropských jazyků a celkový rozsah korpusu je 92 290 000 slov v zarovnaných cizojazyčných textech. Hlavním plánem do budoucna je vylepšit rozhraní Park, které je stále ve výstavbě, doplnit další jazyky a především získat další, žánrově pestřejší texty (nově jsou do korpusu zahrnuty pro některé jazyky i publicistické texty Presseurop).

Nové poznatky pro většinu publika přinesla prezentace o korpusech CZESL (Czech as a Second Language – korpus žáků, jejichž mateřštinou není čeština) a ROMi (korpus romských žáků), především pak pasáže věnované systému ano-

tace těchto korpusů a charakteristice chybovosti u romských žáků. Jiří Mácha představil korpus z textů jednání Poslanecké sněmovny Parlamentu ČR z let 2006-2010 volně dostupných na internetu. U těchto textů lze dobře zjistit jméno autora, datum narození, obvod, za který byl zvolen, i politickou příslušnost. Tyto informace pak spolu se samotným textem slouží ke korpusově založené stylistické analýze textů (Mácha uvedl příklad srovnání mluvčích Petra Tluchoře a Pavla Kováčika, oba ve zkoumaném období zastávali pozice předsedů klubů).

Z jiného soudku byl příspěvek Mileny Hebal-Jezierske nazvaný *Kolokační obrazy některých lexémů patřících do sémantického pole cizinec v českém tisku (s metodologickými úvahami)*. Autorka se v něm zaměřila na to, jak česká tištěná média (na základě korpusu PUB2006) zobrazují cizince různé národnosti. Zkoumala přitom slova *cizinec/cizinecký* a názvy národností jako *Slovák, Ukrajinec, Rom* apod. Na základě kontextové i bezkontextové analýzy pak např. došla k závěru, že obraz cizince v médiích je velmi často spojován s kriminalitou, což je v rozporu se statistickými informacemi o kriminalitě cizinců v ČR. Hebal-Jezierska se v příspěvku také věnovala tomu, které národnosti jsou podle korpusových dat Čechům nejvíce a nejméně sympatické. Celkově lze říci, že obraz cizinců v českých médiích bohužel neodpovídá realitě.

Z příspěvků vycházejících z paralelního korpusu InterCorp bylo zjevné, že autoři se často potýkají s videm, byly mu věnovány hned tři příspěvky. V několika prezentacích se autoři zabývali mluveným jazykem, v tomto ohledu patřil k nejzajímavějším příspěvek Lucie Válkové o příklonném *-s* *Za deset litrůs říkal? Příklonné -s a jeho distribuce ve spontánním dialogu*. Hojně navštívená byla rovněž prezentace Marie Mikulové a Jany Hoffmannové *Korpusy mluvené češtiny a možnosti jejich využití pro poznání rozdílných „světů“ mluvenosti a psanosti*, pojednávající o rekonstrukci standardizovaného textu z mluvené řeči, jež vzbudila u publika rozporuplné reakce. Srovnání rysů mluveného a psaného textu každopádně potvrdilo, že mluvený jazyk a psaný jazyk skutečně představují rozdílné světy. Další otázkou je, jak bude možné mluvený materiál a jeho psané úpravy využít pro automatické rozpoznávání významu mluvené řeči.

Dále je nutno zmínit Hanno Bibera z vídeňské Akademie věd, který se ve svém příspěvku věnoval potenciálu akademických digitálních edic vytvářených v rámci velkých textových korpusů. Nelze opomenout ani příspěvky techničtějšího rázu věnované značkování korpusů, za všechny jmenujme příspěvek Hany Skoumalové *Porovnání úspěšnosti tagování korpusu* či prezentaci Tomáše Jelínka a Vladimíra Petkeviče *Systém jazykového značkování korpusů současně psané češtiny*.

Na závěr zmiňme příspěvek Dominiky Kovářikové s provokativním názvem *Gramatická kategorie negace*, který ukázal, že negace u sloves by měla být podle obecně přijímaných kritérií a na základě korpusových dat považována za gramatickou kategorii. U adjektiv je situace poněkud složitější, ale přesto tento pří-

spěvek, jež lze označit za corpus-driven, přinesl nové poznatky a otevřel pole pro nová zkoumání.

Nezbývá než dodat, že z této korpusové konference vzešly tři sborníky (1. InterCorp, 2. Výzkum a výstavba korpusů, 3. Gramatika a značkování korpusů), které vydalo v září Nakladatelství Lidové noviny v rámci řady Studie z korpusové lingvistiky. Tyto sborníky obsahují plné znění všech příspěvků, které na konferenci zazněly, proto je lze všem zájemcům o tuto oblast jen doporučit.

Opolská konference o normě a antinormě ve slovanských jazycích a mikrojazycích

Josef Šimandl

Ústav pro jazyk český AV ČR, v. v. i.

simandl@ujc.cas.cz

Institut polské filologie a katedra slavistiky spolu s komisí pro slovanské mikrojazyky při Mezinárodním komitétu slavistů uspořádaly ve dnech 15.–16. září 2011 mezinárodní vědeckou konferenci *Norma i antynorma w wielkich i małych językach współczesnej Slavii*. 22 referátů zaznělo v nově upraveném prostředí prosklených střešních ateliérů Collegia Maius Opolské univerzity.

Tematickou orientaci podávalo členění na sekce – A: jazyky západoslovanské, B: jazyky východoslovanské a jihoslovanské, C: slovanské mikrojazyky. Řazení sekcí po sobě osvobodilo účastníky od přebíhání, kolizí a nestíhání, kterými trpí velké konference. Stejně auditorium jako v sekcích měly i úvodní a závěrečné plenární referáty, které se dotýkaly obecných otázek.

Prvním ze tří referátů úvodního pléna se stal referát Jany Hoffmannové *České mluvené projevy: norma a antinorma*. Autorka v něm upozornila na složitou situaci pojmání normy, tím víc pak antinormy, v nepřipravených mluvených projevech. Mj. porovnávala jejich texty se standardizovanými podobami, jaké vznikají při anotaci Pražského závislostního korpusu mluvené češtiny:¹ co je pak normou a co antinormou – „normální“ hezitace, chaotičnost výstavby aj., anebo dodatečná neautorská reformulace? Josef Šimandl v referátu *Antinorma a korpusová lingvistika*² poukázal na různost norem jako modálně chápaných

¹ Příklad: no tak, tak EHM otec ten umřel normálně v roce vosumdesát BREATH EHM EHM byl vyučený kovář pracoval normálně potom jako dlouhou dobu na dráze BREATH -> *Otec umřel v roce 1980. Byl vyučený kovář. Dlouhou dobu pracoval na dráze.* Ke korpusu a způsobu jeho zpracování lze na internetu vyhledat text Hajič, J. et al., *Pražský závislostní korpus mluvené češtiny. Rekonstrukce standardizovaného textu z mluvené řeči* (elektronická adresa se nezobrazuje ani nechová standardně).

² Není to zvykem, ale tomuto příspěvku je věnován poněkud větší prostor než ostatním. Důvodem je, že právě v něm se setkává zaměření konference se zaměřením našeho časopisu.

normativů (norma jazyka, jeho útvarů, normy pro komunikační situace) a na empirickou nepřipadnost představy jejich vzájemného antagonismu: normy spisů koexistují a bastardizují se.³ Antinormu by bylo lépe chápat jako popírání užitečnosti stanovování norem a jejich kodifikace, jak k němu dochází při bezhodnotové deskripci úžu v korpusové lingvistice – což ovšem není jediné možné využití bezpochyby vysoce cenných korpusových dat. Většina ostatních referátů se přesto držela jednoduchého zadání, že antinorma je vše, co působí proti normě, neboli vše, co není norma sama.⁴ Referát Alexandra D. Ďuličenka *Норма и антинорма: К методологическим принципам изучения* ukázal otázky pojímání (anti)normy na hláskoslovném vývoji moskevské mluvy během 1. poloviny 20. století. Z řady obecných otázek, které řešil, uvedu inspirativní připomenutí, že spisovná norma má charakter aposteriorního jazyka,⁵ takže mezi ní a živým jazykem je stále a zákonitě napětí.

V úvodním referátu sekce A přednesl Oldřich Uličný referát *Pojetí normy spisovné češtiny a její kodifikace v připravované Moderní mluvnici češtiny*. Správnost prostředků (podobnou technické normě) prý vyžadujeme nehodnotově, hodnotíme pak vhodnost jejich užití podle komunikační situace. Jiří Zeman v referátu *Tematizace střetu češtiny a moravštiny v neoficiální komunikaci* rozebral úsek konverzace s nesnadným domlouváním o tom, že pro „Morávaka“ *kotouč* je *kolečko* na převážení hlíny. Irena Bogocz(ová)⁶ představila referátem *Refleksja literackiego języka czeskiego i interdialektu potocznego, tzw. obecnej češtiny* některá data ze sociolingvistického výzkumu postojů ke spisovné češtině.⁷ Natalia Długosz v referátu *Normatywny aspekt nowych polskich i bułgarskich rzeczowników złożonych* věnovala pozornost slovům s komponenty *euro-, eko-, bio-, narko-, gej-* aj.

Sekci B otevřel referát *Проблемы нормы [в?] обучении русскому языку как иностранному* A. Miljutiny. Následoval referát Philippa Hofenedera *Norma języka w (post)totalitarnym społeczeństwie. Radzieckie i postradzieckie wpływy*

³ Bastardizace tu není míněna hanlivě; navazuje na významy v genetice (mezidruhové křížení) a technice (funkční integrace nepůvodního vybavení). Srov. obohacování spisovného jazyka (inter)dialektismy nebo slangismy.

⁴ V pozvánce stálo: „Zasób czynników przeciwstawiających się normie literackiej stopniowo rozszerza się, tworząc w efekcie zjawisko, które wstępnie proponujemy nazwać *antynormą*.”

⁵ Prof. Ďuličenko se mj. věnoval umělým jazykům. Ty se dělí na apriorní, v nichž je vymyšleno od začátku vše, a aposteriorní, vybudované na základě jednoho nebo více živých jazyků (např. volapük nebo esperanto). Za vysvětlení děkuji Jiřímu Zemanovi.

⁶ Na domácí Ostravské univerzitě užívá podobu příjmení s *-ová*, v polských textech figuruje příjmení bez *-ová*.

⁷ Postoje ke spisovnému jazyku zkoumaly týmy z Ostravské univerzity od poloviny 90. let už v několika projektech. Výsledky jednoho z dotazníkových výzkumů představila 3. června 2011 také Jana Svobodová na konferenci *Čeština v pohledu synchronním a diachronním (Stoleté kořeny Ústavu pro jazyk český)*.

polityki językowej na stan ukraińskiego. Na časově různých překladech téhož textu do ukrajinštiny, dokonce i na edicích ukrajinských textů (např. Ivan Franko) ukázal, jak se v sovětské éře postupně prohlubovalo stírání národních rysů ukrajinského jazyka. Katarzyna Bednarska a Kamil Szafranec v referátu *Słoweński język literacki – norma czy brak normy?* připomněli, že v malém Slovinsku přetrvává udivující různost (inter)dialektů. Dva referáty přiblížily situaci v Chorvatsku – Mihaela Matešić: *Značajke kodificirane norme i uporabne norme u suvremenome hrvatskom standardnom jeziku*; Teodora F. Cvijanović a Vanessa Vitković: *Odnos sintaktičke norme u hrvatskim normativnim priručnicima i u mladih*; druhý z referátů vycházel z terénního výzkumu.

Do sekce C přešlo jednání referátem Tadeusze Lewaszkiwicze *Spory/polemiki ortograficzne w historii języków łużyckich (na tle słowiańskich sporów/polemik ortograficznych)* mj. připomněl, jaké zmatky a nevoli mohou vyvolat požadavky vzdělavců reetymologizovat psaní slov a „vrátit“ mu podoby, které mluvená řeč dávno opustila. Co představil referát Edwarda Brezy, to říká jeho název: *Prace Rady Języka Kaszubskiego nad kształtowaniem normy w kaszubszczyźnie literackiej*.⁸ Kašubštině se věnovaly i další dva referáty. Dušan Vladislav Paždžerski⁹ v referátu *Norma w języku kaszubskim i jej brak. Gramatyka, pisownia, język literacki* podal přehled popisů kašubštiny a porovnával se stavem v srbsštině; avizoval také příklady problémů těmito popisy reflektovaných a naopak nezaznamenaných, ale zmínil je nakonec jen obecně. Marek Roman Cybulski nazval svůj referát *O normie morfonologicznej w kaszubszczyźnie*. Připomněl mj., jak Eugeniusz Gółąbk své příručce *Kaszëbsczi Słoworz Normatiwny* předeslal, že podává svou normu, a komu se nelíbí, může psát po svém; je tedy koexistence norem, resp. nepevnost normy, pro kašubštinu příznačná. Po kašubologických příspěvcích následovaly dva k situaci ve Slezsku. Bogusław Wyderka v příspěvku nazvaném *Na pograniczu samodzielności. O początkach języka śląskiego* připomněl, že diskurs o uznání slezštiny, popř. Slezska, nemá povahu jazykovou (ve Slezsku se užívá směs polsko-německo-česká), ale politickou. Artur Czesak referátem *Antynormatywne teksty górnośląskie – anarchia a innowacja* představil nejstarší záznamy jazyka tzv. vasropoláků,¹⁰ které vznikaly s funkcí pobavit, a z teorie komunikace připomněl procesy folklorizace, exotizace a s nimi spojené diskriminace (nebo řekněme dehonestace). Žánrů postupně přibývalo, ale dodnes v hornoslezských textech zůstává jejich spektrum poměrně chudé a ve skripici, kterou se manifestuje odlišnost, je

⁸ Rada Języka Kaszubskiego neboli Radzëzna Kaszëbszczégò Jãzëka vytváří pandán známé Rady Języka Polskiego a usměrňuje jazyk mj. vydáváním seznamů schválených podob slov, zejména nových. Její bulletin/*biuletin* lze najít i na internetu (doména www.kaszubi.pl); čísla bohužel nejsou ani kompletní, ani aktuální.

⁹ Toto je polská podoba příjmení. Adresa ve městě Novi Sad měla podobu příjmení *Pažderski*. K nesamozřejmě identitě příjmení v různých slovanských prostředích srov. i pozn. 6.

¹⁰ Např. Ten „Kampf z tem Drachem“ (...) od Schillera.

dost kolísání (i v jednom textu různě psaná táž koncovka).¹¹ Sekci uzavřel Michal Fejsa, příslušník asi 16tisícové rusínské menšiny v srbské Vojvodině,¹² referátem *Стан и тенденції унапредзованя норми руского языка*.¹³ Problémy vzniklé střetáváním pestré palety poměrně silných vlivů (ukrajinština, polština, maďarština, srbština, nověji samozřejmě i angličtina) se řeší přibližováním podob přijatých slov zdrojovému jazyku.¹⁴

Závěrečné plénium se opět věnovalo obecným problémům. Stanisław Gajda v první části svého referátu *Norma a polityka językowa*¹⁵ představil vztahy mezi langue, parole, strukturou a uзуální normou, mezi jejím popisem a normou jednak kodifikovanou, jednak internalizovanou. Druhou část referátu věnoval jazykové politice (na obecné úrovni) a připomněl, že v zacházení s jazykem se úzus posouvá od *wartości* (hodnoty) k *powszechnej ekwiwalencii* v procesu, který lze nazvat *odwartośćowanie*; hodnoty těžko rozpoznáváme, obecně přijímané základní hodnoty nejsou. Normy se změkčují, jejich závaznost slábne, vytvářejí se nové; působí i vzpoura proti normám, preferování chaosu a (pouhé) hry. Ale procesy glocalizace – neboli globální lokalizace – jsou složité; nemusíme rezignovat. Jaroslav Lipowski v příspěvku *Jak traktować antynormę w ortoepii?* představil osm výslovnostních, převážně suprasegmentálních jevů a jejich (ne)príznačnost v mluvě Čechů, Poláků a Slováků. Jen tři z nich našel ve všech třech jazycích, mj. nadměrnou otevřenost vokálů. Přehled zpochybňuje možnost zobecnění ortoepické antinormy (uznáme-li její existenci) pro více jazyků, ba i pro jeden jazyk v různých komunikačních situacích; navíc by se o antinormě mělo mluvit jen při záměrné činnosti „proti“, ne při bezděčných chybách a odchylkách. Jacek Baluch vyšel v referátu *Estetyczna funkcja, norma i wartość w perspektywie językowej* z Mukařovského (1936) a Jakobsonových (1961) funkcí jazyka a připomněl dobové debaty o Hrabalově jazyku, mj. v kontextu debat, zda literární obraz řeči (sovětského) dělníka může obsahovat vulgarismy.¹⁶

¹¹ V diskusi byl připomenut případ *góralů/Góralů*. Jejich dialekt má víc mluvčích než kašubština (nemluvě o slezštině nebo dokonce hornoslezštině), ale nedeklarují se jako svébytný národ s vlastním jazykem.

¹² Rusíni, kteří žili v Haliči na pomezí Ukrajiny, Polska (Lemkové) a Slovenska (Šariš), prošli zčásti – na horní Tise – maďarizací a ve stalinských časech procesem na pomezí genocidy a diasporizace. V Srbsku, vzdor nevelkému počtu, požívají lepšího právního stavu než leckterá větší minorita v leckteré vyspělejší zemi: mají mj. vlastní univerzitu. Proto i úsilí popsat jazyk má bohatší výsledky než u jiných mikrojazyků.

¹³ V rusínštině *русского* : rusínského × ruského : *российского/ московского*.

¹⁴ V tom je rusínský postup odlišný od tradičního českého, kde se ještě na přelomu 80./90. let uvažovalo, že až se slovo dostatečně vžije, mělo by se místo *hardware* zavést psaní *hardvér* apod. Tyto předpovědi se ani v češtině nevyplnily a korpusy ukazují spíše opak (*jazz, puding*).

¹⁵ Také tomuto příspěvku je věnováno více místa. Z triády *korpus – gramatika – axiologie* se dotýká pojmu třetího.

¹⁶ Tyto debaty začaly polemikou nad Káňovou hrou *Parta brusiče Karhana*.

Připomněl tím, že nedávná historie zasluhuje spíš kvalitní reflexi a vyzdvižení poučných momentů než zapomenutí.

Opolští pořadatelé prokázali zdaleka ne poprvé, že umějí uspořádat dělnou konferenci. Od účastníků jim za to patří vřelý dík. Těšíme se na sborník příspěvků, kde bude zájemce moci podrobněji studovat, co tato zpráva mohla jen naznačit.

Co je v ČNK nového I

(Zprávy z Českého národního korpusu)

Václav Cvrček

Ústav Českého národního korpusu Filozofické fakulty UK Praha

vaclav.cvrcek@ff.cuni.cz

Touto první zprávou bychom rádi začali s pravidelným informováním o novinkách v projektu *Český národní korpus*, který v sobě soustředí hlavní infrastrukturu korpusového výzkumu v ČR. Na stránkách KGA by se tak postupně měly objevovat informace o nových korpusech, akcích, publikacích a nástrojích, které mohou být zajímavé pro všechny uživatele českých a paralelních korpusů, pro něž je tento časopis logickou primární disciplinární platformou.

V září na konferenci *Korpusová lingvistika 2011* (sborníky vyšly v Nakladatelství Lidové noviny *Korpusová lingvistika Praha 2011* na podzim 2011) byla oficiálně do ostrého provozu uvedena aplikace **SyD**, která slouží k všestrannému průzkumu morfologických, lexikálních, syntaktických i pravopisných variant. Aplikace, jejíž podrobný popis a možnosti užití jsou vtěleny do příspěvku V. Cvrčka a P. Vondříčky *Výzkum variability v korpusech češtiny*,¹ slouží k synchronnímu i diachronnímu (odtud název SyD) pohledu na jazykové varianty. Aplikace, která doplňuje současné a běžně užívané nástroje o netradiční pohled na korpusová data, je přístupná pro všechny zájemce na adrese <http://syd.korpus.cz>.

V diachronní části může uživatel po zadání dvou variant sledovat jejich konkurenci v rámci sedmi století vývoje češtiny. Pro tyto účely byl sestaven speciální korpus *Diakon*, který zahrnuje texty od nejstarších památek ze 14. století (nejstarším je *Alexandreida*, datovaná do roku 1310) z diachronní sekce ČNK až po publicistiku z let 1991–2009 z korpusu SYN. Vzhledem k proměnám písemné produkce, kterými čeština za 7 století své existence prošla, není možné klást na takový korpus stejné požadavky stylové a žánrové vyváženosti, jaké máme na korpusy češtiny současné, nicméně i tak podává zajímavé svědectví

¹ In: Čermák, F. (ed.) (2011): *Korpusová lingvistika Praha 2011 – 2. Výzkum a výstavba korpusů*. Praha: Nakladatelství Lidové noviny, s. 184–195.

o jazykových změnách. Vedle proměn vzájemného poměru frekvencí jednotlivých variant je možné v této části SyDu sledovat i vývojové tendence (pomocí lineární regrese), prohlížet si detailněji různá časová okna a zobrazovat ukázky textů z korpusu podle jednotlivých let.

Synchronní část – pro laického uživatele možná trochu přitažlivější – prezentuje frekvenční údaje o jednotlivých variantách z různých oblastí současného jazyka. Za základní je zde považována dichotomie psaný a mluvený jazyk, která je reflektována opozicí korpusů SYN2010 na jedné straně a sloučenými korpusy mluveného jazyka Oral2006 a Oral2008 na straně druhé. Zde se často ukazuje – dříve poněkud přehlížená – tendence češtiny výrazně preferovat jeden prostředek v psaném jazyce a konkurenční podobu v jazyce mluveném. V jednotlivých záložkách pod tímto základním rozložením má uživatel možnost pomocí přehledných grafů detailněji nahlédnout do distribuce variant v jednotlivých textových typech a žánrech psaného jazyka, příp. do sociolingvistického profilu mluvčích užívajících jednotlivé podoby. Vedle této informace o rozložení podob se zde prezentuje i jednoduchá kolokační analýza ukazující pomocí tzv. term clouds na rozdíly a shody v užití a kolokabilitě jednotlivých variant. V relativně krátkém čase a bez většího úsilí se tak uživateli dostane komplexní a poměrně přehledná informace o užívání variant v jednotlivých oblastech jazyka a napříč různými osami.

Výhodou webové aplikace SyD, kromě toho, že není nutné ji instalovat, je také možnost snadného předávání výsledků mezi uživateli. Každému dotazu odpovídá unikátní klíč, který je možné ve formě internetového odkazu snadno předávat a citovat a který jednoduše nasměruje každého dalšího uživatele k totožným výsledkům (např. synchronnímu porovnání variant *abychom* a *abysme* odpovídá odkaz: <http://syd.korpus.cz/qhq88Zz4.syn>).

Aplikace SyD, která si od začátku klade nelehký úkol být užitečným nástrojem pro badatele i zajímavou popularizací pro laické uživatele, je prvním pokusem svého druhu. Autoři proto očekávají, že na základě ohlasu uživatelů bude nutné její možnosti přizpůsobit reálným potřebám. Zároveň se už nyní rýsují další možnosti rozšíření, které budou postupně přibývat a o kterých vás třeba právě v této rubrice budeme v budoucnu informovat.

Pozvánka

Gramatika a korpus 2012: 4. mezinárodní konference

U příležitosti stého výročí narození Miloše Dokulila
28.–30. listopadu, Praha
(1. cirkulář)

Konference je součástí zavedené řady setkávání, která začala před šesti lety a slouží jako fórum pro korpusovou lingvistiku a gramatický výzkum. Předchozí konference se konaly v Praze (2005), Liblicích (2007) a Mannheimu (2009).

TERMÍNY

Zaslání abstraktu a přihlášky	30. dubna 2012
Oznámení o přijetí referátu	31. května 2012
Termín pro konferenční poplatek	30. září 2012

ZAMĚŘENÍ

- korpusový výzkum slovtvorby v češtině a/nebo v jiném jazyce/jazycích; teorie tvoření slov;
- korpusový výzkum gramatiky češtiny a/nebo jiného jazyka/jazyků (morfologie a syntax včetně slovosledu)

K PŘÍSPĚVKŮM

Konferenčními jazyky jsou čeština, slovenština, angličtina.
V jednom z těchto jazyků se píše i abstrakt; délka 20–30 řádků.
Délka příspěvku: 20 minut + 10 minut diskuse.

Vybrané příspěvky budou publikovány v tomto časopise.

KONFERENCEŇNÍ POPLATEK

poplatek v plné výši 2000 Kč
zlevněný poplatek (studenti, emeritní badatelé) 1000 Kč

Poplatek zahrnuje konferenční materiály, pohoštění během konference a pronájem konferenčních prostor.

PROGRAMOVÝ VÝBOR

František Štícha (Akademie věd ČR), předseda
Tilman Berger (Univerzita Tübingen)
František Čermák (Univerzita Karlova Praha)
Mirjam Fried (Akademie věd ČR/Univerzita Karlova Praha)
Christian Mair (Univerzita Freiburg)

Patrice Pognan (INALCO, Paříž)
Josef Šimandl (Akademie věd ČR)
Mária Šimková (Slovenská akademie věd, Bratislava)
Pavel Štekauer (Univerzita Košice)
Miloslav Vondráček (Univerzita Hradec Králové/Akademie věd ČR)

ORGANIZAČNÍ VÝBOR

František Štícha, Mirjam Fried, Václava Holubová, Jana Klímová, Josef Šimandl,
Miloslav Vondráček

Druhý cirkulář s informacemi o platbách a ubytování bude zaslán po 31. květnu 2012.

Kontaktní adresa: gramkorp2012@ujc.cas.cz

PŘIHLÁŠKA

na 4. mezinárodní konferenci
Gramatika a korpus 2012
Akademie věd ČR, Národní třída 3, Praha 1
28.–30. listopadu

Přihlášku pošlete prosím na adresu gramkorp2012@ujc.cas.cz **do 30. 4. 2012.**

Jméno, příjmení, tituly:

Organizace:

Forma účasti – vyznačte prosím křížkem:

- přikládám abstrakt referátu/posteru
- zúčastním se bez příspěvku

Kontaktní adresa:

E-mail:

Call for papers

Grammar & Corpora 2012: 4th International Conference

On the 100th anniversary of Miloš Dokulil
November 28–30th, Prague, Czech Republic
(1st circular)

This conference is part of an established series of gatherings that was started six years ago and that serves as a forum for bringing together corpus linguistics and grammatical research. The previous conferences in the series took place in Prague (2005), Liblice (2007), and Mannheim (2009).

DATES

Deadline for abstract submission and registration	April 30 th , 2012
Notification of acceptance	May 31 st , 2012
Deadline for conference fee payment	September 30 th , 2012

FOCUS

- corpus-based research on word-formation in Czech and/or other language(s); theory of word-formation;
- corpus-based grammar research on morphology and syntax (including word-order) in Czech and/or other language(s)

SUBMISSION GUIDELINES

Conference languages: English, Czech, and Slovak.

The abstract should be written in one of those languages; abstract length: 20–30 lines.

Paper length: 20 minutes for presentation plus 10 minutes for discussion.

Selected papers will be published in this journal.

CONFERENCE FEE

Full amount 80 Euros

Reduced amount (students, emeriti) 40 Euros

The conference fee covers conference materials, coffee breaks refreshment, and rental fees for the conference space.

PROGRAMME COMMITTEE

František Štícha (Academy of Sciences, Prague), Conference Chair

Tilman Berger (University of Tübingen)

František Čermák (Charles University, Prague)

Mirjam Fried (Charles University/Academy of Sciences, Prague)
Christian Mair (University of Freiburg)
Patrice Pognan (INALCO, Paris)
Josef Šimandl (Czech Academy of Sciences, Prague)
Mária Šimková (Slovak Academy of Sciences, Bratislava)
Pavel Štekauer (University of Košice)
Miloslav Vondráček (University of Hradec Králové/Academy of Sciences, Prague)

CONFERENCE ORGANIZING COMMITTEE

František Štícha, Mirjam Fried, Václava Holubová, Jana Klímová, Josef Šimandl, Miloslav Vondráček

2nd circular with information on payment and accommodation will be sent out after may 31st 2012.

Contact information: gramkorp2012@ujc.cas.cz

REGISTRATION

for the 4th International Conference
Grammar & Corpora 2012
Prague, Czech Republic
November 28–30, 2012

Please send by e-mail to gramkorp@ujc.cas.cz. Deadline: **April 30th, 2012.**

Name, including titles:

Affiliation:

Participation (mark with an **x** the option that applies):

- I'm attaching an abstract to be considered for presentation – oral/poster.
- I plan to attend without a paper

Contact address:

E-mail:

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 **kg@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;
- příspěvek by měl být členěn na kapitoly (a event. subkapitoly); ty se označují číslly 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 uzávorkovat: (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: např. (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, 354 pp.

ZIFONUM, 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, s. 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, 354 pp.

ZIFONUM, 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>				