

Obsah / Content

Články / Articles

Vybrané hláskové jevy severovýchodočeské nářeční oblasti v korpusu DIALEKT

Phone-level phenomena of the Northeast Bohemia region in the DIALEKT corpus /3

Hana Goláňová, Martina Waclawičová

Ke konkurenci nominativ/instrumentál predikativní: korpusová analýza větných struktur se subjektivními substantivy rodu mužského životného

On the competition of nominative/instrumental predicative: a corpus analysis of sentence structures with subject masculine animate nouns /16

Martin Janečka

Subkategorizace adverbálních významů (hranice mezi obsahem a významem)

Subcategorization of adverbial meanings (boundary between cognitive content and linguistic meaning) /33

Marie Mikulová, Jarmila Panevová

Nekodifikujme, raději popisujme a vysvětlujme normu a úzus Pěstujme jazykovědu, nepěstěme jazyk

Let's not codify language, let's describe and explain it's norms and usage instead Let's practice linguistics instead of refining language /47

Luboš Veselý, Vojtěch Veselý

The Fragmented Paradigm

Fragmentárnost paradigmatu /63

Miloslav Vondráček

Recenze / Reviews

Kateřina Šormová – Karel Šebesta a kol.: Korpusy v jazykovém vyučování

Praha: Univerzita Karlova, Filozofická fakulta, 2019 /74

Barbora Hrabalová

**Jana Hoffmannová – Jiří Homoláč – Kamila Mrázková (eds.):
Syntax mluvené češtiny**

Praha: Academia, 2019 /76

Magdaléna Rysová

Zprávy / News

25 let Českého národního korpusu a nová aplikace /81

Václav Cvrček, Michal Křen

Pokyny pro autory /83

Instructions for authors /84

Vybrané hláskové jevy severovýchodočeské nářeční oblasti v korpusu DIALEKT

Hana Goláňová, Martina Waclawičová

*Ústav Českého národního korpusu Filozofické fakulty UK, Praha
hana.golanova@ff.cuni.cz, martina.waclawicova@ff.cuni.cz*

Phone-level phenomena of the Northeast Bohemia region in the DIALEKT corpus

ABSTRACT: DIALEKT is the newest specialized corpus of spoken Czech. It contains recordings of Czech dialects, along with two layers of transcripts: a dialectological one and an orthographic one. It is thus a useful tool for studying the phonetics of Czech dialects. The present article explores four phone-level phenomena in the Northeast Bohemia region, and the lexicalization process they seem to be undergoing. Remnants of the original bilabial *w*, in the guise of rounded *w* and non-syllabic *y*, are still well-attested in this dialect region, whereas the last vestiges of hard *l* and hard *y* are so scarce that they will in all likelihood disappear with the current oldest generation.

KEYWORDS: spoken corpora, dialect corpora, phonetic features of dialects

KLÍČOVÁ SLOVA: mluvené korpusy, nářeční korpusy, nářeční hláskové jevy

Ke konkurenci nominativ/instrumentál predikativní: korpusová analýza větných struktur se subjektovými substantivy rodu mužského životného

Martin Janečka

*Katedra českého jazyka, Pedagogická fakulta UK, Praha
martin.janecka@pedf.cuni.cz*

On the competition of nominative/instrumental predicative: a corpus analysis of sentence structures with subject masculine animate nouns

ABSTRACT: In this study I follow copious discussion about semantics of case category in Czech. I analyse chosen nouns in the position of nominative predicative or instrumental predicative with the help of the newest synchronic written corpus of Czech language – SYN v7. I try to verify opinions on this problematic presented in literature with the emphasis on the fact that only one study was realized on corpus material base so far (Štícha 2004). I dedicate special attention to masculine animate nouns *pán, muž, chlap, otec* and *syn*. I take into consideration various semantic features of these nouns like “male person”, “husband” or “monarch” within my analysis.

KEYWORDS: nominative, instrumental, case, language corpus, predicative, semantic feature

KLÍČOVÁ SLOVA: nominativ, instrumentál, pád, jazykový korpus, predikativ, sémantický rys

Subkategorizace adverbálních významů (hranice mezi obsahem a významem)

Marie Mikulová, Jarmila Panevová

Ústav formální a aplikované lingvistiky MFF UK Praha

{mikulova, panevova}@ufal.mff.cuni.cz

Subcategorization of adverbial meanings (boundary between cognitive content and linguistic meaning)

ABSTRACT: The necessity to distinguish between cognitive content and linguistic meaning arose in European structural linguistics (Saussure 1916) and was further elaborated in the Prague Linguistic Circle (Mathesius 1942; Dokulil – Daneš 1958; Daneš 1974). In the contribution, we describe the practical aspects of applying the principle of distinguishing meaning and content to the task of delimitation of adverbial meanings, expressed by prepositional groups. We present methodology and main principles we work with in completing the set of meanings of adverbials. We describe how we use the principle of substitutability of synonyms. All the examples in the contribution relate to spatial adverbials but the principles apply to adverbials in general. Our theoretical framework is Functional Generative Description (Sgall et al. 1986).

KEYWORDS: cognitive content, linguistic meaning, homonymy, vagueness, synonymy, form-function relation, adverbial

KLÍČOVÁ SLOVA: kognitivní obsah, jazykový význam, víceznačnost, vágnost, synonymie, vztah funkce a formy, adverbální určení

Nekodifikujme, raději popisujme a vysvětlujme normu a úzus Pěstujme jazykovědu, nepěstěme jazyk

Luboš Veselý, Vojtěch Veselý

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

{vesely, vvesely}@ujc.cas.cz

Let's not codify language, let's describe and explain it's norms and usage instead Let's practice linguistics instead of refining language

ABSTRACT: The paper raises the question whether Czech linguistic community should make an effort to codify language norms (in a prescriptive way). In our opinion, linguists should first and foremost endeavour to describe and interpret the usage of language; creating an empirically adequate codification would be hardly feasible. Our standpoint is substantiated by an analysis of six cases of discrepancy between standard use of a language unit and the up-to-date codification. They are related to the following topics: 1) aspect of the verb *soustředit se* 'focus, concentrate', 2) agreement of a predicate / verbal attribute with a compound subject containing a noun with neutral declension, 3) competition between the units *zpěčovat se* and *vzpěčovat se* 'defy', 4) ortographic codification of verbs like *zkomponovat* 'compose', *zkombinovat* 'combine', *zkonstruovat* 'construct', 5) competition between the verbs *zešeřit se* and *sešeřit se* 'get dark', 6) competition between so-called purpose adjectives ending with *-ící* and so-called verbal adjectives ending with *-ící*.

KEYWORDS: codification, language usage, competition of language units

KLÍČOVÁ SLOVA: kodifikace, úzus, konkurence jazykových prostředků

The Fragmented Paradigm

Miloslav Vondráček

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

vondracek@ujc.cas.cz

The Fragmented Paradigm

ABSTRACT: Defectibility of the paradigm of flexible word classes is generally described in grammar books within the periphery of a word class: (the limitation of the numerical paradigm of pluralia tantum, substance and abstract nouns, the aggradation of relational adjectives etc.) or within the particular grammatical category (the absence of the 1st and 2nd person of impersonal verbs, imperative, vocative of impersonal nouns etc.). However, the actual corpora data only partly correspond to this rough concept. The purpose of this paper is to discuss such issues as to what extent these irregularities should be taken into account in terms of the automatic text processing, how they should be registered in databases such as Internetová jazyková příručka or to what extent they should be reflected in dictionaries.

KEYWORDS: morphology, defectibility of the paradigm, numerical paradigm,

Fragmentárnost paradigmatu

ABSTRAKT: Defektivnost paradigmatu ohebných slovních druhů bývá obecně popsána v gramatikách v rámci periferie slovního druhu: (omezení číselného paradigmatu jmen pomnožných, látkových, abstrakt, agradace relačních adjektiv ap.), nebo v rámci gramatické kategorie (absen- ce 1. a 2. osoby a personálních sloves, imperativu, vokativu a personálních jmen ap.). Reálná korpusová data odpovídají této rámcové představě jen částečně. Cílem příspěvku je hledání odpovědi, do jaké míry s nepravidelnostmi počítat při automatickém zpracování textu, jak je evidovat v databázích typu Internetová jazyková příručka nebo do jaké míry je promítnout do slovníků.

KLÍČOVÁ SLOVA: morfologie, defektivnost paradigmatu, číselné paradigma

Recenze / Reviews

Kateřina Šormová – Karel Šebesta a kol.: Korpusy v jazykovém vyučování

Praha: Univerzita Karlova, Filozofická fakulta, 2019, 110 s.

Barbora Hrabalová

*Ústav českého jazyka, Masarykova univerzita, Brno
b.kocarova@gmail.com*

Oblast aplikované lingvistiky a jejího výzkumu obohatil v posledních dekadách nový zdroj analýzy – jazykový korpus. Není tomu jinak ani v rámci výuky jazyků, kde má korpus a korpusová lingvistika své pevné postavení a obecně velký přínos, zejména v nepřímém využití pro tvorbu slovníků nebo výukových materiálů. Je ale možné pracovat s korpusem přímo ve výuce? Souhlasně by jistě odpověděli autoři publikace *Korpusy v jazykovém vyučování*. Tato příručka, vydaná v letošním roce, cílí na všechny zájemce o učení jazyků, podává přehled nejen českých korpusových zdrojů a ilustruje, jakým způsobem je možné vytvářet aktivity pro výuku. Na sto stranách se čtenář seznámí s danou problematikou postupně v logickém sledu od teorie k praxi. Propagaci využívání korpusových dat ve výuce napomůže zajisté i fakt, že je publikace dostupná ke stažení zdarma ve formě e-knihy.

Úvodem je nutné zmínit, že se nejedná o zcela nový přístup k výuce jazyků. Metoda výuky, která využívá jazykové korpusy, byla v kontextu angličtiny jako cizího jazyka rozpracována již na začátku devadesátých let minulého století. Ze všech přístupů je zřejmě nejznámější tzv. data-driven learning. Autorem tohoto termínu a propagátorem korpusové metody byl anglický lingvista Tim Johns (1936–2009).¹ Jeho myšlenky si našly řadu příznivců i v rámci výuky češtiny jako cizího jazyka, odkazy na tuto metodu najdeme například v práci Jaromíry Šindelářové a Svatavy Škodové (2012) nebo Pavlínny Vališové (2016). Podobný cíl, a to inspirovat učitele pro práci s korpusy, má projekt s jednoduchým názvem *Korpusy pro školy*, což je repositář cvičení na webu Českého národního korpusu. Cvičení jsou určena učitelům češtiny na základních a středních školách a rovněž také pro učitele češtiny jako cizího jazyka. Výše uvedené příklady ukazují, že zájem o využití korpusu ve výuce byl u nás patrný i před vydáním této publikace. Nicméně dosud chyběla kniha, která by poskytla souhrnné informace, ucelený pohled na problematiku a rámec pro další studium. Práce kolektivu autorů z různých pracovišť Filozofické fakulty Univerzity Karlovy tyto funkce určitě zastávat může.

¹ Za klíčový je považován článek již z roku 1991, viz Johns (1991).

Kniha je rozdělena do několika samostatných částí. První dvě se věnují akvizičním korpusům a nastiňují způsoby poznání žakovského jazyka, které tyto korpusy umožňují z hlediska různých lingvistických kategorií (např. gramatika, slovní zásoba či pragmatika). Publikace je určena širokému spektru čtenářů, a z toho důvodu tedy nelze očekávat, že text bude směřovat do hlubin teorie. Přesto však nezůstává zcela na povrchu, nastiňuje výzkum z různých úhlů a nabádá k dalšímu čtení. Z těchto dvou částí publikace jasně vyplývá, že akviziční korpusy mají velký výzkumný i pedagogický potenciál. Tato tvrzení mohou vyznívat možná až příliš optimisticky, neboť korpusový výzkum žakovského jazyka rovněž naráží na jisté limity. Nevýhodám a omezením výzkumu žakovského jazyka se však publikace také věnuje, a to ve třetí části, která popisuje proces budování korpusů a soustředí se na rozdíly mezi obecnými korpusy a těmi akvizičními. Limity poznání jsou tedy zmíněny v souvislosti s parametry jako je velikost, vyváženost, autentičnost, strukturování a značkování korpusu. K obecným korpusům se také váže další, čtvrtá část, která představuje podrobný návod, jak získávat data za pomoci nástrojů Českého národního korpusu. Věnuje se celému postupu, od výběru vhodného korpusu přes podoby vyhledávacích dotazů až k možnostem třídění získaných výsledků. Návodný manuál je doplněn obrazovým materiálem a je rovněž využit pro popis základních termínů korpusové lingvistiky.

Konkrétní ukázky práce s korpusy a korpusovými nástroji tvoří jádro poslední, páté části. Zpracovány jsou aktivity, které zahrnují téměř všechny nástroje Českého národního korpusu, jež byly v době přípravy publikace k dispozici, jmenovitě tedy korpusové rozhraní Kontext a dále aplikace Morfio, SyD a KWords. Databáze překladových ekvivalentů Treq zastoupena není, přestože i ta představuje jistý potenciál pro výuku. Pracovní listy jsou doplněny popisem postupu práce či možnostmi, jak lze cvičení variovat. Důraz je kladen i na práci učitele před zadáním korpusových úkolů i po něm, což je z pohledu didaktiky neméně důležité.

Autoři publikace si stanovili dva hlavní cíle. Představit zejména akviziční korpusy jako nástroj pro výzkum jazykových dat pocházejících od mluvčích, kteří se jazyk teprve učí, bylo cílem číslo jedna. Druhým záměrem bylo poskytnout inspiraci pro práci s obecnými korpusy, které mohou být využity přímo ve výuce. Na závěr můžeme konstatovat, že tyto cíle byly úspěšně splněny. Věříme, že podobně prakticky laděný bude i plánovaný druhý díl publikace a že výuka doplněná o aktivity založené na korpusu bude mít stále více příznivců.

LITERATURA

- JOHNS, T. (1991): Should You Be Persuaded. Two samples od data-driven learning matherials. In: Johns, T. – King, P. (eds.), *Classroom Concordancing. English Language Research Journal 4*. Birmingham: Birmingham University, 1–13.
- ŠINDELÁŘOVÁ, J. – ŠKODOVÁ, S. (2012): Práce s korpusy ve výuce. In: *Metodika práce s žáky cizinci v základní škole*. Praha: MŠMT: 96–119.
- VALIŠOVÁ, P. (2016): Korpus ve výuce češtiny jako cizího jazyka – typy cvičení. In: Starý Kořánová Ilona a Tomáš Vučka. *Čeština jako cizí jazyk VIII*. Sborník příspěvků z VIII. mezinárodního sympozia o češtině jako cizím jazyku. Praha: Univerzita Karlova: 129–141.

Jana Hoffmannová – Jiří Homoláč – Kamila Mrázková (eds.): Syntax mluvené češtiny

Praha: Academia, 2019

Magdaléna Rysová

Ústav formální a aplikované lingvistiky, MFF UK, Praha

magdalena.rysova@ufal.mff.cuni.cz

Kniha *Syntax mluvené češtiny*, jejímiž editory jsou Jana Hoffmanová, Jiří Homoláč a Kamila Mrázková, si klade za cíl podat obecnou charakteristiku běžně mluvené češtiny, a to konkrétně v oblasti syntaktické a textové výstavby mluvených projevů. Výzkum této problematiky je pro českou lingvistiku velkým přínosem, protože (jak autoři knihy sami upozorňují) komplexní popis syntaxe autentických mluvených projevů zde dosud schází. Zároveň jde o úkol velmi náročný – autoři se snaží hledat obecné, univerzální charakteristiky syntaxe nepřipravených (zřídka polopřipravených) projevů, které se ze své podstaty vyznačují jistou mírou vágnosti a neuspořádanosti.

Kniha je dílem kolektivu autorů Ústavu pro jazyk český AV ČR, v. v. i. Na jednotlivých kapitolách se podíleli Anna Čermáková, Mirjam Fried, Martin Havlík, Jana Hoffmannová, Jiří Homoláč, Lucie Jílková, Zuzana Komrsková, Marie Kopřivová, Kamila Mrázková, Petra Poukarová, Olga Richterová, Eliška Zaepernicková a Jiří Zeman.

Cennou částí monografie je již samotná předmluva J. Hoffmannové, která se dotýká hned několika důležitých témat. Zdůrazňuje například, že mluvené projevy nejsou anomáliemi nebo odchylkami od textů psaných, ale že „představují specifický svět“ (s. 12). Výzkum mluvených projevů by se měl zaměřovat hlavně na modifikace systémových konstrukcí známých z psaných textů a na další konstrukce specifické pouze pro mluvený projev. Pozornost je přitom třeba věnovat zachycení a analýze jejich pravidelnosti a ustálenosti.

Vzhledem k tomu, že syntax mluvené češtiny není dosud komplexně probádaným tématem, hledali autoři knihy inspiraci také v jiných oblastech lingvistiky, jako je např. nářeční syntax či zahraniční interakční lingvistika a konverzační analýza.

Jedinečnost knihy spočívá mj. v jazykovém materiálu, který autoři využívají pro svůj výzkum. Jedná se o autentické texty mluvených projevů dostupné z různých korpusů. Je zapotřebí ocenit rozmanitost použitých dat – autoři pracovali s celou řadou korpusů, jako je např. ORAL2013, ORAL2006, ORALv1, ORTOFON, Pražský mluvený korpus, Brněnský mluvený korpus nebo DIALOG. Srovnáváním výsledků výzkumu na různých jazykových korpusech autoři zároveň získali poznatky o stylistických diferencích některých typů mluvených projevů. Lze s jistotou konstatovat, že výzkum syntaxe mluvené češtiny provedený na takto rozsáhlých autentických datech je v tomto ohledu unikátní.

Výsledky výzkumu jsou v jednotlivých kapitolách podpořeny ukázkami z využitých korpusů (resp. přepisy mluvených projevů) – za zmínku jistě stojí i to, že jako doplněk publikace vznikl soubor vybraných audionahrávek k jednotlivým kapitolám, který monografii výborně dokresluje.

Monografie je rozčleněna do devíti kapitol. Každou z nich můžeme chápat jako téměř samostatnou studii věnující se jednomu vybranému tématu. První kapitola (jejímiž autory jsou M. Havlík a E. Zaepernicková)¹ se zabývá temporalitou a interakcí v rozhovorech. Autoři rozebírají syntaktické a prozodické aspekty projektivity a „dodatků“. Je důležité zmínit, že termín projektivita je zde užíván jinak než ve formální závislostní syntaxi – projektivitou je míněna skutečnost, že v rozvíjené replice je nějakým způsobem dáváno najevo, kam replika či rozhovor směřuje a co bude následovat. Projektivita je v kapitole zkoumána ve vztahu k závislostní syntaxi, slovosledu, prozódii, sémantice a slovním druhům. V rámci analýzy dodatků je v kapitole představena slovtvorně homogenní skupina nových českých termínů *přílepký, vlepky, přelepky* (srov. anglické termíny *glue-ons, insertables, replacements*).

Autoři druhé kapitoly (K. Mrázková a J. Homoláč) analyzují verbálně signalizované sebeopravy v mluvených projevech. Před vlastním výzkumem sebeoprav v češtině rozebírají autoři také vliv typu jazyka na způsob provedení sebeopravy – jednotlivé jazyky se liší např. v tom, jak velké jsou části výpovědi, příp. jakého druhu jsou jazykové elementy, které se v projevu opakují. Typ jazyka má vliv dále na to, zda a kam se autoři projevu při sebeopravě vracejí. V kapitole jsou dále analyzovány jednotlivé typy sebeoprav v češtině, a to opravy ve slovesné frázi, jmenné frázi, předložkové frázi a opravy s návratem na začátek klauze.

Třetí kapitola (jejími autory jsou Z. Komrsková, P. Poukarová a M. Havlík) pojednává o překryvech replik. Překryvy replik mohou být děleny na kompetitivní a kooperativní (podle jejich průběhu) a úmyslné a neúmyslné (podle jejich počátku). V kapitole jsou analyzovány překryvy v běžně mluvené češtině, konkrétně 20 nahrávek dostupných z korpusu spontánní mluvené češtiny ORAL2013. Vlastní analýza se zaměřuje na tři aspekty: proč překryv replik vzniká, čím je ovlivněn jejich počet a jak vypadá kontext, který po překryvu následuje.

Čtvrtá kapitola (J. Hoffmannová a O. Richterová) se zabývá jednoslabičnými nepřízvučnými začátky výpovědi a dialogických replik. Z analýzy korpusových dat vyplynulo, že nejfrekventovanějším začátkem tohoto typu je výraz *sem*. Dále jsou v kapitole rozebírány výrazy *ses, seš, sme, ste, sí, se, ti*, kondicionálové tvary slovesa *být* (z nichž se na začátku syntaktických jednotek nejčastěji objevuje tvar *by*) a zájmenné výrazy typu *mi, mu, tě* nebo *ho*, které se ovšem v iniciální pozici vyskytují spíše sporadicky. Zkoumané začátky výpovědi se bezprostředně vyskytují po prozodických předělech, heziticích či kontinuátorech typu *hmm*. Redukované začátky výpovědi zároveň velmi často následují po vystřídání mluvčích, a podílejí se tak na celkové výstavbě dialogů.

¹ Autorem podkapitoly *Důraz na konci výpovědi* je J. Zeman.

Pátá kapitola (J. Hoffmannová, Z. Komrsková, P. Poukarová a E. Zaepernicková) pojednává o reprodukci řeči/myšlení v mluvených projevech. Autorky kapitoly upozorňují na skutečnost, že signály reprodukce řeči/myšlení jsou v mluvených projevech značně různorodé. Vedle jasně explicitně vyjádřených případů (především slovesy mluvení) se můžeme setkat se signály zredukovanými, z nichž slovesa mluvení/myšlení mizí (srov. *a von že už musí jít; vona prej musim už jít* aj.). Někdy může být reprodukce řeči/myšlení signalizována např. pouze změnou výšky, barvy či síly hlasu. Obecně tak můžeme vyčlenit signály lexikální a syntaktické a signály prozodické a parajazykové. Autorky kapitoly se dále soustřeďují na výzkum reprodukce řeči s nejfrekventovanějším slovesem mluvení *říkat/řít*, a to na datech korpusu ORAL2013. Zkoumají přitom řeč přímou i nepřímou, vnitřní i vnější. Ve zkoumaném vzorku dat se nejčastěji objevují tvary nedokonavého slovesa *říkat* (v minulém i přítomném čase, v 1. i 3. osobě), pro reprodukci vnitřní řeči se používá téměř výhradně osoba 1.

Jako velmi přínosnou hodnotím kapitolu šestou (J. Zeman), která popisuje specifické syntaktické konstrukce typické pro mluvenou češtinu. Autor kapitoly opět upozorňuje na skutečnost, že při vyjadřování myšlenek v mluveném (nepřipraveném) projevu nemusí být jejich syntaktické uspořádání dokonale organizované či linearizované, jako je tomu v projevu psaném či důkladně připraveném, a přesto výsledný projev plní svoji komunikační funkci. V běžném dialogu se tak často vyskytují specifické modifikace „pravidelných“ syntaktických konstrukcí, např. vynechávání či vyčleňování některých výrazů. Autor v kapitole postupně rozebírá zeugma, anakolut a prostředky nepravé hypotaxe. Upozorňuje přitom na skutečnost, že je zapotřebí rozlišovat mezi obecnými konstrukcemi mluvené syntaxe a konstrukcemi odrážejícími individuální řečové návyky mluvčích.

Autoři sedmé kapitoly (J. Zeman, J. Hoffmannová a L. Jílková) se zabývají problematikou otázek v mluvené češtině. Na vybraných korpusových datech (zahrnujících také nahrávky autentické komunikace pořízené J. Zemanem) zkoumají fungování otázek v běžném mluveném rozhovoru, zaměřují se přitom na strukturní vlastnosti otázek (jako je kontextualizace či integrace v rozsáhlejší replice), analyzují také jejich vlastnosti zvukové. Autoři rozebírají otázky pravé i nepravé. U jednotlivých typů otázek mj. sledují, do jaké míry je vyžadována reakce partnera rozhovoru a jak tato reakce může vypadat.

Osmá kapitola (M. Fried) je věnována interakčnímu dativu v mluvené češtině. Autorka v kapitole rozlišuje dativy valenční a nevalenční, nevalenční dělí dále na dva podtypy podle toho, zda vyjadřují propoziční obsah. Valenční a nevalenční typy, které propoziční obsah vyjadřují, autorka nazývá dativy sémantickými. Ve svém výzkumu se dále zaměřuje na nevalenční dativy vyjadřující významy pragmatické – tyto dativy jsou v kapitole označovány jako interakční. Interakční dativy jsou děleny podle toho, zda odkazují k adresátovi či k mluvčímu. Na základě korpusových dat jsou v kapitole dále popisovány jednotlivé významy interakčních dativů – vyjádření vysoké/neobvyklé míry, výraz překvapení, vyjádření subjektivní emocionality či vyjádření hodnotícího svědeckého komentáře. Zvláště zajímavá je část kapitoly zabývající se analýzou diskurzivních vlastností inkluzivního *nám*. Při popisu základních funkčních typů

tohoto výrazu se přitom autorka neomezuje pouze na rozbor z hlediska významu, ale rozebírá také specifika z hlediska textového a kolokačního.

Nejvíce prostoru je v knize věnováno poslední, deváté kapitole nazvané „Diskurzivní markery“ (jejími autorkami jsou A. Čermáková, L. Jílková, Z. Komrsková, M. Kopřivová a P. Poukarová). Vzhledem k tomu, že tato kapitola je výrazně rozsáhlejší než kapitoly ostatní a zahrnuje několik případových studií, zastavím se u ní blíže. V kapitole se pracuje s různými typy markerů, autorky nabízejí jejich dělení podle nejpřičtějších funkcí na markery interakční, modální a konektivní. Tyto typy markerů odrážejí vztah k posluchači, ke sdělení a k organizaci projevu. Interakční markery jsou dále děleny na kontaktní (v iniciální – srov. *hej, počkej, hele*, či terminální pozici – srov. *chápeš, vid', rozumíš*) a responzní (*jasně, no, ano, jo*). Jako příklady modálních diskurzivních markerů autorky uvádějí výrazy typu *myslím, řekněme*, za konektivní markery můžeme považovat např. *jako, ale, teď, protože* či *vyložene*. Ke každému typu diskurzivních markerů autorky přináší rozbor dosavadního bádání v českých i zahraničních studiích.

Za méně obratnou považuji formulaci o konektorech týkající se pojetí J. Hrbáčka: „Konektivní funkci plní obvykle spojky a spojovací výrazy, ale mají ji i další výrazy s jinou slovnědruhovou charakteristikou. J. Hrbáček (1994: 57) navíc za konektory považuje i výrazy, které nejsou zapojeny do věty jako větné členy, což je jeden z často uváděných rysů DM.“ (s. 260) Hrbáček (1994) označuje jako konektory výrazy se spojovací funkcí bez větněčlenské platnosti. Spojovací funkci mohou mít v textu i výrazy ve funkci větných členů – tyto prostředky „jsou větnými členy svých vět a současně vyjadřují svým lexikálním významem vztah mezivýpovědní (tím, že ho pojmenovávají)“ (Hrbáček 1994: 54), J. Hrbáček je ovšem termínem konektory neoznačuje. Vhodnější formulace by tedy podle mého názoru byla, že J. Hrbáček považuje za konektory právě (tj. nikoli „navíc... i“) *výrazy, které nejsou do věty zapojeny jako větné členy*. Jedná se však o formulační drobnost.

Jako obzvláště přínosnou hodnotím část kapitoly věnovanou vztahu diskurzivních markerů k syntaxi. Autorky upozorňují na skutečnost, že ve většině dosavadních studií jsou markery chápány jako výrazy syntakticky nezačleněné (resp. některé z nich jsou začleněny do 3. syntaktického řádu). Autorky ovšem navrhují podívat se na problematiku diskurzivních markerů z opačného pohledu – syntaktickou nezačleněnost markerů lze chápat jako dosud neúplně popsanou syntaktickou začleněnost. V tomto ohledu popisují diskurzivní markery jako výrazy, které mohou „tvorit prozodické úseky vyjadřující syntaktickou funkci“ (s. 255), tj. mohou např. signalizovat začátek či konec promluvy. Diskurzivní markery lze popsat také z hlediska tzv. dialogické syntaxe (zaměřující se na syntaktické jevy v rozhovorech, které nejsou popsateľné v rámci tradiční syntaxe) jako výrazy, které přispívají ke strukturaci rozsáhlejších úseků dialogu. Obecně tedy mohou být diskurzivní markery chápány jako složky specifických syntaktických struktur vyskytujících se v mluvených projevech.

Součástí kapitoly je dále několik případových studií o vybraných diskurzivních markerech. Předmětem výzkumu je množina (především) polyfunkčních výrazů, které byly vyčleněny na základě frekvence, pozice a kvalitativní analýzy. Autorky se ve

svých studiích primárně zaměřují na jednoslovné diskurzní markery, v jednotlivých studiích jsou okrajově zmiňovány také kolokace, jejichž výskyt ve funkci diskurzních markerů není zanedbatelný. Ustálená spojení v některých případech „potvrzují zjištěné funkce příslušných DM“ (s. 258) a na plnění určitých funkcí se výrazně podílejí. I z těchto důvodů by bylo podle mého názoru zajímavé věnovat ustáleným slovním spojení (resp. víceslovným diskurzním markerům) více pozornosti, i když je zcela pochopitelné, že prostor v monografii je omezený a bylo nutné výběr zkoumaných markerů zúžit.

U diskurzních markerů, které si autorky kapitoly vybraly pro případové studie, je uvedena četnost výskytů v korpusu neformální češtiny ORTOFON a korpusu mediální češtiny DIALOG. Pro oba korpusy je vypočítána také hodnota souhrnné míry výskytů daných markerů na začátku a konci repliky. Ukazuje se například, že (podle tabulky 1 „Přehled analyzovaných výrazů“) pro diskurzní markery *jasně* a *vyloženež* vycházejí hodnoty souhrnné míry záporně. Znamená to tedy, že u nich v mluveném projevu převažuje výskyt na konci repliky. To je zajímavá skutečnost, protože v psané češtině mají diskurzní markery příslovečného charakteru tendenci vyskytovat se ve větě na první pozici. V tomto ohledu se odhalují možné rozdíly mezi projevem psaným a mluveným a bylo by jistě přínosné se na tyto rozdíly v dalším výzkumu zaměřit.

Závěrem mohu konstatovat, že kniha *Syntax mluvené češtiny* přináší velmi zajímavý a cenný rozbor mluvených projevů. V jednotlivých kapitolách se objevují různé přístupy, a to zejména přístup gramatický, stylistický, fonetický, využívány jsou metody konverzační analýzy, analýzy diskurzu, dále interakční a korpusové lingvistiky.

Jak autoři sami naznačují, výzkum syntaxe mluvené češtiny by se dále mohl ubírat směrem koexistence a interakce verbálních prostředků s prostředky neverbálními (gesta, pohledy, mimika aj.). Téma začleňování neverbálních prostředků do syntaktického popisu jazyka je velmi zajímavé a lingvisticky lákavé. Jako čtenáři se můžeme těšit, že se objeví v dalších studiích autorského kolektivu o mluvené češtině.

LITERATURA

HRBÁČEK, J. (1994): *Nárys textové syntaxe spisovné češtiny*. Praha: Trizonia.

25 let Českého národního korpusu a nová aplikace

Václav Cvrček, Michal Křen

*Ústav Českého národního korpusu, Filozofická fakulta UK Praha
{vaclav.cvrcek, michal.kren}@ff.cuni.cz*

V pondělí 9. září 2019 uplynulo 25 let od doby, kdy byl na FF UK oficiálně založen Ústav Českého národního korpusu. Samotnému vzniku ÚČNK předcházelo úsilí několika nadšenců v čele s prof. Františkem Čermákem, jejichž společným cílem bylo vybudovat pro češtinu národní korpus, který by ji zachycoval v různých podobách a formách.

ÚČNK začínal ve velmi skromných podmínkách, mnozí si ještě vzpomenu na první prostory ve sklepech FF UK. Za čtvrtstoletí své existence se ÚČNK rozrostl, a to nejen prostorově a personálně, ale především v nabídce jazykových zdrojů: rozsahem a různorodostí nabízeného jazykového materiálu, uživatelských rozhraní a souvisejících služeb se ČNK řadí mezi přední projekty svého druhu na světě. ČNK má v současné době více než 7 500 registrovaných uživatelů a násobky uživatelů neregistrovaných, kteří v průměru položí zhruba 3 000 dotazů za den (1,1 mil. dotazů za rok), což svědčí o tom, že infrastruktura pro korpusovou lingvistiku je u nás etablovaná a potřebná. Oceněním kvalit projektu je také udělení statusu českého K(nowledge)-centra CLARIN pro korpusovou lingvistiku v prosinci 2018.

U příležitosti tohoto jubilea plánuje ÚČNK zveřejnit několik nových nástrojů a aplikací. První z nich, která se na portálu www.korpus.cz objevila přímo v den narozenin, je zbrusu nová aplikace Slovo v kostce (dostupná jak z vyhledávacího okna na portálu, tak přímo na adrese <https://www.korpus.cz/slovo-v-kostce>), která slouží k vytvoření rychlého a přehledného profilu zadaného slova ze dvou hlavních perspektiv: jednojazyčné a překladové. V obou případech aplikace skládá informace o zadaném slově výhradně z korpusových dat ČNK a dává je dohromady tak, aby o něm poskytla rychlý a komplexní obrázek. Slovo v kostce se stává důležitou aplikací ČNK, na níž je nyní nově napojeno vyhledávání na hlavní stránce webového portálu, a zároveň pomyslným rozcestníkem, z něhož se nabízejí upřesňující prokliky do ostatních webových rozhraní ČNK, poskytujících o vybraných aspektech zadaného slova podrobnější informace. Jednojazyčná část aplikace shrnuje informace, které lze o hledaném slovu získat v synchronních jednojazyčných korpusech, zejm. SYN2015 pro psaný jazyk a Oral v1 pro jazyk mluvený. Při prezentaci frekvenční charakteristiky je maximální péče věnována tomu, aby tyto údaje – působící na laické uživatele někdy poměrně technicky a neintuitivně – byly zasazeny do kontextu a bylo jim možné

přířknout nějakou smysluplnou interpretaci (např. pomocí uvedení seznamu slov podobně frekventovaných, zařazením do frekvenčního pásma či uvedením četnosti v hlavních typech textu: beletrii, oborové literatuře, publicistice a konverzaci). Poté následuje v jednojazyčné části oddíl věnovaný psanému jazyku, kde jsou prezentována jak data o kolokacích hledaného slova, tak o jeho frekvenčním vývoji v čase (pro tyto účely je využívána publicistická část korpusu SYN v7 a sleduje se vývoj v datově velice dobře pokrytém období od konce 90. let do současnosti). Kolokace jsou prezentovány jak formou word cloudu, která dává představu o míře asociace mezi kolokátem a hledaným slovem, tak formou ukázkových vět. Oddíl pro mluvený jazyk pak shrnuje výskyty slova z různých sociodemografických (pohlaví, věk a vzdělání) a regionálních hledisek (mapa republiky s naznačením frekvence slova v jednotlivých regionech). Vše je doplněno ukázkou autentické mluvy spolu s možností přehrát audio ukázkou.

Překladová část aplikace Slovo v kostce umožňuje na základě dat z paralelního korpusu InterCorp zjišťovat, jaké jsou pravděpodobné překlady zadaného slova, podobně jako je to v nástroji Treq. V současné podobě aplikace lze zjišťovat překladové ekvivalenty z češtiny do angličtiny a několika dalších velkých evropských jazyků. Kromě word cloudu, naznačujícího prominenci jednotlivých překladových ekvivalentů, se pomocí grafu znázorňuje, jak časté je takové překladatelské řešení v jednotlivých hlavních částech InterCorpu – v beletrii, publicistice, parlamentních debatách, evropské legislativě a filmových titulcích. Dominantní výskyt překladového řešení v jednom typu textu může leccos naznačit o jeho stylové platnosti. Vše je doplněno příkladovými větami, na nichž si uživatel může ověřit, v jakých kontextech se překladové ekvivalenty užívají.

Zveřejněná verze aplikace Slovo v kostce je prvním pokusem integrovat různé nástroje a data ČNK na jedno místo. Je přitom zřejmé, že funkce i repertoár zobrazovaných informací se budou s dalšími verzemi rozšiřovat a zpřesňovat. Výhledově by tak Slovo v kostce mělo hrát roli jakési vstupní brány do světa jazykových korpusů ČNK. Budeme proto rádi, pokud Slovo v kostce vyzkoušíte a případně nám prostřednictvím uživatelské podpory ČNK sdělíte své názory na jeho fungování.