

40 mm

STUDIE Z KORPUSOVÉ LINGVISTIKY

Myriad pro 14/16

Redakční rada:

40 mm

doc. RNDr. Vladimír Petkevič, CSc. (předseda)

doc. Mgr. Václav Cvrček, Ph.D.

prof. PhDr. František Čermák, DrSc.

Mgr. Anna Čermáková, Ph.D.

Mgr. Jan Kocek

PhDr. Marie Kopřivová, Ph.D.

Mgr. Michal Křen, Ph.D.

prof. PhDr. Karel Kučera, CSc.

Ing. Alexandr Rosen, Ph.D.

Mgr. Pavel Vondříčka, Ph.D.

Myriad pro Italic 10/14

blok 150x225mmm

40 mm

Skladební funkce a pád v korpusu: Frekvenční analýza

40 mm

Myriad Pro Semibold 30/32

Tomáš Jelínek

Myriad Pro Regular 16/18



30 mm



33 mm



15 mm

Tato práce vznikla za podpory grantu Grantové agentury České republiky
Treebank češtiny na základě gramatiky (Grammar-based Treebank of Czech),
reg. č. 13-27184S.

Recenzenti:

PhDr. Marie Kopřivová, Ph.D.

doc. PhDr. Klára Osolobě, Dr.

doc. RNDr. Vladimír Petkevič, CSc.

Cover © Michaela Blažejová, 2015

© Filozofická fakulta Univerzity Karlovy, 2015

Všechna práva vyhrazena.

ISBN 978-80-7422-366-2




14 mm

Poděkování



19 mm

Titulek1:
Myriad Pro Bold 16/19,
linka pod: 3b 25%černé, odsazení 2mm




20 mm


Chtěl bych především poděkovat doc. Vladimíru Petkevičovi za inspiraci a podnětnou spolupráci. Děkuji i ostatním kolegům z Ústavu teoretické a počítačové lingvistiky, dr. Alexandru Rosenovi, dr. Haně Skoumalové a dr. Mileně Hnátkové. Děkuji doc. Kláře Osolobě a dr. Marii Kopřivové za cenné poznámky k textu. A velký dík patří také mé manželce Marii za trpělivost a podporu.

Základní písmo:
Charis SIL 9/11, do bloku

Paginace:
Myriad Pro Bold 9/11



20 mm



5
9,2 mm



0. Úvod	8
0.1 Uspořádání publikace	9
0.2 Seznam zkratk a symbolů	9
1. Syntakticky anotovaný korpus	11
1.1 Korpus SYN2005 a jeho žánrové subkorporusy	11
1.1.1 Tabulka zastoupení subkorporusů v korpusu SYN2005	11
1.1.2 Subkorpus beletrie	12
1.1.3 Subkorpus odborné literatury	12
1.1.4 Subkorpus publicistiky	12
1.1.5 Tabulka podílu slovních druhů	12
1.2 Morfologické značkování korpusu	13
1.3 Syntaktické značkování korpusu podle analytické roviny PDT	14
1.3.1 Syntaktické funkce v Pražském závislostním korpusu	15
1.4 Metoda syntaktické anotace	16
1.4.1 MaltParser	16
1.4.2 Zjednodušení dat pro parsing	16
1.4.3 Opravná pravidla	17
1.5 Spolehlivost údajů, korekce výsledků	17
1.5.1 Metoda korekce výsledků	17
1.5.2 Údaj o spolehlivosti tabulky	18
1.6 Rozlišení prostých a předložkových pádů	18
2. Syntaktické funkce	19
2.1 Přísudek	19
2.1.1 Seznamy tvarů a lemmat funkce Pred	20
2.2 Jmenná část verbonominálního predikátu	20
2.2.1 Seznamy tvarů a lemmat funkce Pnom	21
2.2.2 Tabulka realizace funkce Pnom slovními druhy	22
2.2.3 Tabulka realizace funkce Pnom vybranými pády substantiv a adjektiv	22
2.3 Podmět	23
2.3.1 Seznamy tvarů a lemmat funkce Sb	23
2.3.2 Tabulka realizace funkce Sb slovními druhy	24
2.3.3 Tabulka realizace funkce Sb vybranými pády substantiv a zájmen	25
2.4 Předmět	25
2.4.1 Seznamy tvarů a lemmat funkce Obj v korpusu	26
2.4.2 Tabulka realizace funkce Obj slovními druhy	26
2.4.3 Tabulka realizace funkce Obj pády jmen	27
2.4.4 Seznamy sloves s valencemi v korpusu SYN2005	27
2.5 Příslopečné určení	35
2.5.1 Seznamy tvarů a lemmat funkce Adv	36
2.5.2 Tabulka realizace funkce Adv slovními druhy	37
2.5.3 Tabulka realizace funkce Adv pády jmen	37

2.6	Přívlastek	38
2.6.1	Seznamy tvarů a lemmat funkce Atr	38
2.6.2	Tabulka realizace funkce Atr slovními druhy	39
2.6.3	Tabulka realizace funkce Atr vybranými pády substantiva	39
2.7	Doplňek	40
2.7.1	Seznamy tvarů a lemmat funkce Atv	41
2.7.2	Tabulka realizace funkce Atv slovními druhy	41
2.7.3	Tabulka realizace funkce Atv vybranými pády substantiv, adjektiv a zájmen	42
2.8	Kvantifikace	42
2.9	Koordinace a apozice	43
2.9.1	Tabulka zastoupení koordinovaných a aponovaných funkcí	44
2.10	Ostatní syntaktické funkce v PDT	44
2.10.1	Tabulka frekvence funkce ExD a pomocných funkcí	46
3.	Tabulky	47
3.1	Tabulky vztahu slovního druhu a syntaktické funkce	47
3.1.1	Tabulky vztahu slovního druhu a syntaktické funkce (bez koordinace)	49
3.2	Tabulky vztahu pádu a syntaktické funkce	50
3.2.1	Tabulky vztahu pádu a syntaktické funkce (bez koordinace)	52
3.3	Tabulky vztahu pádu a syntaktické funkce u substantiv	54
3.3.1	Tabulky vztahu pádu a syntaktické funkce u substantiv (bez koordinace)	56
3.4	Tabulky vztahu pádu a syntaktické funkce u adjektiv	58
3.5	Tabulky vztahu pádu a syntaktické funkce u zájmen	60
3.6	Tabulky vztahu pádu a syntaktické funkce u číslovek	62
4.	Závěr	65
	Summary	66
Literatura	Obsah 1 úroveň :Charis Sil bold 9/11	67
	Obsah 2 úroveň: číslo bold, písmo reg., odsazení zleva 4 mm	

0. Úvod

Titulek1:
Myriad Pro Bold 16/19,
linka pod: 3b 25%černé, odsazení 2mm

↕ 3 ř.

Vztahy mezi slovy ve větě se již od Aristotela popisují pomocí skladebních funkcí přiřazených jednotlivým větným členům. V gramatikách se často zkoumá, jaké jazykové prostředky na morfologické rovině se využívají v různých funkcích. Frekvenční analýza těchto vztahů založená na rozsáhlých jazykových datech však pro češtinu dosud chybí.

0.1 Uspořádání publikace

Titulek2:
Myriad Pro Bold 12/13, před 2ř, za 1 ř.
linka pod: 2b 25%černé, odsazení 1,5mm

Tato práce má tři části. V první představujeme výchozí korpus SYN2005 a jeho žánrové subkorporusy a uvádíme metody morfologické a syntaktické anotace korpusu. Ve druhé části popisujeme skladební funkce, s nimiž pracujeme; vymezujeme jejich obsah.

Titulek3: Myriad Pro Bold 9/11, 15b před, 7b za
kostička (odrážka): 25%černé, předsazení 4 mm

■ 1.1.1 Tabulka zastoupení subkorporusů v korpusu SYN2005

Tabulka převzatá ze Statistik češtiny uvádí počet tokenů (textových slov počítaných od mezery k mezeře, včetně číslic a interpunkce) a procentuální zastoupení subkorporusů beletrie, odborné literatury a publicistiky v korpusu SYN2005.

korpus	tokenů	podíl
SYN2005	122 419 948	100 %
BEL	50 011 887	40,9 %
ODB	32 692 293	26,7 %
PUB	39 715 768	32,4 %

Tabulky: Charis SIL 8/9,
linky 0,75b, 0,5, ev. 0,3
uivnitř záhlaví před i za 1ř

Uvedené počty tokenů a podíly subkorporusů platí i pro syntakticky anotovaný korpus SYN2005, z něhož vycházíme. V níže uvedených statistikách však pracujeme jen s těmi tokeny, kterým jsou přiřazeny vybrané syntaktické funkce (nepočítáme například interpunkci).

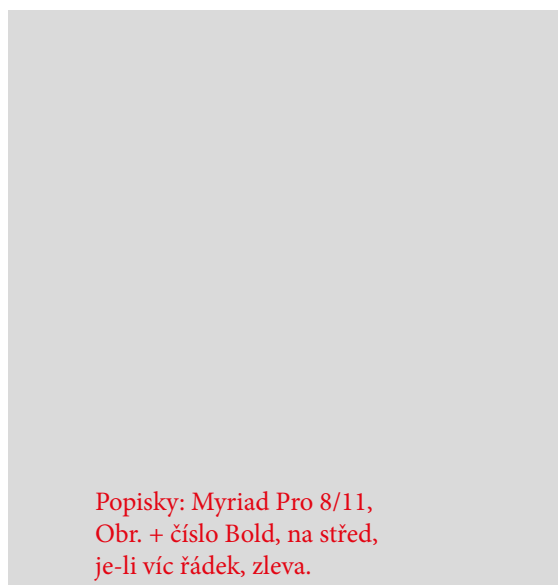
■ 1.1.2 Subkorpus beletrie

Do subkorpusu BEL se řadí především romány a povídky, dále literatura faktu a v menší míře poezie, písně a dramatické texty. Většina subkorpusu sestává z relativně dlouhých textů, v průměru 13 000 tokenů na dokument.¹

V subkorpusu beletrie jsou více než v ostatních subkorpusech zastoupena slovesa, zájmena a adverbia.

Pozn. pod čarou: Charis SIL 7/8,
linka nad první pozn. 05b, 20mm

1 Struktura doc (dokument) v korpusu SYN2005 odpovídá vnitřnímu členění textů, např. jednotlivé povídky v knize povídek, slovníková hesla ve slovníku, jednotlivé články v novinách a časopisech.



obrázky na střed
stránky

Obr. 1. Závislostní strom v Pražském závislostním korpusu

Titulek4: Myriad Pro Bold 9/11, 1ř před,
kostička (odrážka): 25%černé, předsazení 4 mm

■ 1.1.3.1 Subkorpus odborné literatury

Subkorpus odborné literatury zahrnuje vědeckonaučnou literaturu (vědecké práce), populárněnaučnou literaturu, profesní a zájmové časopisy, učebnice a návody, slovníky a administrativní texty. Texty jsou kratší nebo více členěné (např. slovníky) než texty beletristické, na jeden dokument připadá 2400 tokenů.

Ze všech tří subkorpusů jsou v subkorpusu odborné literatury nejvíce zastoupena substantiva a adjektiva.

Titulek5 a dál: Myriad Pro Bold Italic 9/11, 1ř před,
kostička (odrážka): 25%černé, předsazení 4 mm

■ 1.1.3.1.1 Subkorpus publicistiky

Subkorpus publicistiky se skládá z novin a (neodborných) časopisů, v malé míře také z dalších krátkých textů jako korespondence, inzeráty aj. Korpus je převážně sestaven z velmi krátkých textů: novinových a časopiseckých článků, které mají v průměru 350 tokenů na dokument.

Subkorpus publicistiky je co do podílu slovních druhů podobný subkorpusu ODB (relativně vyšší podíl substantiv a adjektiv, nižší podíl sloves a adverbií), rozdíl mezi PUB a BEL jsou ale menší než mezi ODB a BEL.

(1) Kdo se vydá **na cestu**/pA?

(2) Připojil jsem se **k němu**/pD a jeho **přátelům**/pD.

(3) Roku 1980 Aksjonov emigroval **do Spojených**/pG států.

příklady: Charis SIL Italic 8/11, 1ř před, 1ř za

technické písmo (softwarové řádky) Cousine 8,5/11

Kromě textů přímo citovaných v této práci zde uvádíme i další díla, která byla použita při přípravě této publikace a při vývoji metod značkování syntakticky anotovaného korpusu SYN2005.

- Baayen, H. R. (2008). *Analyzing Linguistic Data*. Cambridge: Cambridge University Press.
- Ballesteros, M. & Nivre, J. (2012). MaltOptimizer: A System for MaltParser Optimization. In N. Calzolari et al. (eds), *Proceedings of LREC 2012*, 2757–2763. Istanbul: ELRA.
- Bartoň, T., Cvrček, V., Čermák, F., Jelínek, T. & Petkevič, V. (2009). *Statistiky češtiny*. Praha: Nakladatelství Lidové noviny.
- Bohnet, B., Nivre, J., Boguslavsky, I., Farkas, R., Ginter, F. & Hajič, J. (2013). Joint Morphological and Syntactic Analysis for Richly Inflected Languages. In N. Galatos, A. Kurz & C. Tsinakis (eds), *Proceedings of TACL 2013*, 415–428. Nashville, TN: ACL.
- Cvrček, V., Kodýtek, V., Kopřivová, M., Kovářiková, D., Sgall, P., Šulc, M., Táborský, J., Volín, J. & Wacławicová, M. (2010). *Mluvnice současné češtiny I. Jak se píše a jak se mluví*. Praha: Karolinum.
- Čechová, M., Dokulil, M., Hlavsa, Z., Hrbáček, J. & Hrušková, Z. (2000). *Čeština – řeč a jazyk*. Praha: Nakladatelství ISV.
- Čermák, F. (2001). Language corpora: the Czech case. In V. Matoušek, P. Mautner, R. Mouček, R. & K. Taušer (eds), *Lecture Notes in Computer Science, Proceedings of TSD 2001*, 21–30. Berlin/Heidelberg: Springer Verlag.
- Čermák, F. – Doležalová-Spoustová, D. – Hlaváčová, J. – Hnátková, M. – Jelínek, T. – Koček, J. – Kopřivová, M. – Křen, M. – Novotná, R. – Petkevič, V. – Schmiedtová, V. – Skoumalová, H. – Šulc, M. – Velíšek, Z.: *SYN2005: žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2005. Dostupný z WWW: <http://www.korpus.cz>
- Daneš, F. (1968). Sémantická struktura větného vzorce. In J. Bauer (ed.), *Otázky slovan-ské syntaxe II*, 45–49. Brno: Univerzita J. E. Purkyně.
- Daneš, F., Hlavsa, Z., Jirsová, A., Macháčková, E. & Prouzová, H. (1981). *Větné vzorce v češtině*. Praha: Academia.
- Grepl, M. & Karlík, P. (1998). *Skladba češtiny*. Olomouc: Votobia.
- Hajič, J. (2004a). *Complex Corpus Annotation: The Prague Dependency Treebank*. Bratislava: Jazykovedný ústav L. Štúra, SAV.
- Hajič, J. (2004b). *Disambiguation of Rich Inflection*. Praha: Karolinum.
- Hajič, J., Hajičová, E., Panevová, J. & Sgall, P. (1998). Syntax v Českém národním korpusu. *Slovo a slovesnost*, 59, 168–177.
- Hajič, J., Panevová, J., Buráňová, E., Uřešová, Z. & Bémová, A. (1999). *Anotace Pražského závislostního korpusu na analytické rovině: pokyny pro anotátory*. Praha: MFF UK.
- Hajič, J., Böhmová, A., Hajičová, E. & Vidová Hladká, B. (2003). The Prague Dependency Treebank: A Three-Level Annotation Scenario. In A. Abeillé (ed.), *Treebanks: Building and Using Parsed Corpora*, 103–127. Amsterdam: Kluwer Academic Publishers.
- Hajič, J., Panevová, J., Hajičová, E., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J., Mikulová, M., Žabokrtský, Z., Ševčíková-Razimová, M. & Uřešová, Z. (2006). *Prague Dependency Treebank 2.0, CD-ROM*. Philadelphia: Linguistic Data Consortium.
- Holan, T. & Žabokrtský, Z. (2006). Combining Czech Dependency Parsers. In P. Sojka, I. Kopeček & K. Pala (eds), *Lecture Notes in Computer Science, Proceedings of TSD 2006*, 95–102. Berlin/Heidelberg: Springer Verlag.
- Hrbáček, J. (1995). Několik poznámek o syntaktickém vztahu závislosti. *Český jazyk a literatura*, 46, 5–13.

Myriad Pro Bold 20/23,
linka pod: 3b 25%černé, odsazení 2mm

V ŘADĚ STUDIE Z KORPUSOVÉ LINGVISTIKY DOSUD VYŠLO

Titul: Myriad Pro Bold 9/11, 1ř před,
kostička (odrážka): 25%černé, předsazení 4 mm

■ Korpusová lingvistika: Stav a modelové přístupy

František Čermák – Renata Blatná (eds) Autor: Myriad Pro Italic 9/11

Tento svazek předkládá čtenářům korpusově založené studie (tzv. case studies), které se věnují takovým základním jazykovým jevům a problémům, jako je např. hranice mezi lexikonem a gramatikou, hranice mezi psaným a mluveným jazykem apod., jejichž řešení lze chápat jako modelové pro zkoumání obdobných případů, tj. jejichž závěry lze extrapolovat.

ISBN 80-7106-861-6

ISBN: Myriad Pro Italic 9/11

■ Kolokace

František Čermák – Michal Šulc (eds)

Svazek Kolokace je prvním českým tematickým sborníkem věnovaným dané oblasti. Téma kolokací, „společného výskytu slov“, je dnes centrálním pojmem korpusové lingvistiky, jehož rozmanitost a rozpětí od víceslovných termínů a idiomů až k náhodným kombinacím klade na uchopení problematiky i její zpracování velké nároky.

ISBN 80-7106-863-2

■ Víceslovné předložky v současné češtině

Renata Blatná

Tento svazek představuje všeobecný popis víceslovných předložek, např. v rámci něčeho, vzhledem k něčemu, spolu s něčím, kterých jsou v jazyce řádově stovky (kolem čtyř set). Od jednoslovných předložek, např. v, na, proti apod., kterých jsou řádově desítky, se víceslovné předložky liší především sémanticky. Uvedený popis víceslovných předložek vychází z velkého korpusu SYN2000 a navazuje na zpracování těchto jednotek ve Slovníku české frazeologie a idiomatiky. Výrazy neslovesné.

ISBN 80-7106-865-9

■ Valence českých adjektiv

Marie Kopřivová

Tato práce se věnuje valenci českých adjektiv a využívá při tom rozsáhlého korpusového materiálu, který poskytuje korpus psané češtiny SYN2000. Jejím východiskem je lexikologické pojetí valence a zaměřuje se na popis formálních exponentů adjektivní valence. Nejedná se o vyčerpávající popis valence adjektiv, spíše o naznačení možností, jak k jejímu zkoumání využít korpusový materiál.

ISBN 80-7106-862-4

■ Vidová morfologie českého slovesa

François Esvan

Tento svazek představuje podrobný popis vidové morfologie českých sloves z hlediska jejich tvoření, a to na základě stomiliónového korpusu současné češtiny SYN2000. Tento popis je zachycen v podobě relační databáze obsahující údaje o frekvenci všech sloves nacházejících se v korpusu SYN2000 a o vztazích mezi těmito lexémy z vidového hlediska. Detailně jsou zpracovány a frekvenčně zachyceny způsoby tvoření vidových derivátů: prefixace, sufixace a iterativní sufixace.

ISBN 978-80-7106-913-00

■ Morfologie mluvené češtiny: Frekvenční analýza

Jitka Šonková

Tento svazek podává první soustavnou charakteristiku skloňování a časování v mluvené češtině. Studie vychází z kvantitativní analýzy Pražského mluveného korpusu, tvořeného přepisy více než 304 nahrávek z Prahy a okolí, a zaměřuje se především na konkurenci spisovných a nespisovných tvarů v běžné komunikaci českých mluvčích.

ISBN 978-80-7106-956-0

■ Čestina v mluveném korpusu

Marie Kopřivová – Martina Waclawičová (eds)

Svazek přináší soubor příspěvků z mezinárodní a interdisciplinární konference Čestina v mluveném korpusu konané v roce 2007 na Filozofické fakultě Univerzity Karlovy v Praze. Texty se shodně věnují velmi aktuální a dříve opomíjené mluvené formě jazyka z hlediska řady aspektů. Představují ji tak, jak je zachycena v různých autentických mluvených korpusech, a podrobněji ji zkoumání korpusovými metodami. Témata knihy, první svého druhu, zahrnují široké pole od problematiky budování mluvených korpusů přes zkoumání obecných rysů mluvenosti a variantnosti v mluveném jazyce až po různé aspekty gramatického popisu mluveného jazyka.

ISBN 978-80-7106-982-9

■ Regulace jazyka a Koncept minimální intervence

Václav Cvrček

Tento svazek představuje první pokus, jak na základě korpusových metod vystavět plán jazykově regulační činnosti založené na minimální intervenci do jazyka a řečové činnosti – Koncept minimální intervence (KMI). Přijatým předpokladem tohoto konceptu je fakt, že jazyk se samovolně vyvíjí v účelný nástroj dorozumívání bez pomoci ze strany lingvistů, a není proto důvod, proč by jazykověda měla do vývoje jazyka svými preskriptivními intervencemi za-

sahovat. V polemice s předcházejícími koncepty jazykově regulační činnosti (Teorie jazykové kultury, puristický koncept apod.) vychází KMI striktně z informací, které nám o jazykových prostředcích může poskytnout korpus, mezi něž nepatří kritéria hodnocení na ose spisovný – nespisovný (nebo správný – nesprávný). Zároveň tato publikace přináší obecnější Teorii intervencí, která slouží jako zastřešující teoretický model nad jednotlivými koncepty jazykové regulační činnosti, a korpusové sondy měřící míru preskriptivismu v českém prostředí a dopad kodifikačních intervencí na jazyk.

ISBN 978-80-7106-600-2

■ **Valence českých substantiv**

Anna Čermáková

Valence substantiv se v knize pojímá jako jev lexikologický pro substantiva svébytná, a nikoli (nutně) odvozený od sloves, mající ve svém úhrnu povahu odlišnou a výrazně sémantickou, což naznačuje i řada uváděných subklasifikací. Pojetí valence a zvláště jejích formálních exponentů se přitom soustřeďuje na všechny formální a kategoriální rysy, které ji v korpusu vyjadřují, bez apriorní (a zpravidla omezující) teorie.

ISBN 978-80-7106-426-800

■ **Dnešní skloňování substantiv typů *kámen, břímě***

Josef Šimandl

Desátý svazek řady Studie z korpusové lingvistiky mapuje skloňování skupiny substantiv, která po historicky dlouhou dobu kolísají ve třech pádech singuláru mezi měkkým a tvrdým skloňováním. Práce přináší i sondy do elektronicky přístupného diachronního materiálu. Při popisu jsou konfrontována data podle různých korpusů i podle jiných zdrojů, především podle internetu.

ISBN 978-80-7422-008-1

■ **Perfektum v současné češtině**

Mira Načeva-Marvanová

Perfektum v současné češtině nabízí svým čtenářům možnost podívat se spolu s autorkou do 500-milionové synchronní textové databáze Českého národního korpusu a na základě více než 20 000 dokladů, získaných z této databáze, poznat jedinečnost a překvapující rozmanitost složených forem českého perfekta, které přitom dokládají integraci češtiny do Evropského lingvistického areálu (ELA), respektive i do evropského areálu slovesa *mít* (tzv. „H-areálu“). Souběžný strukturní popis perfektních forem a konstrukcí je zároveň podroben testováním a analýze, zjišťujícím stav a proces jejich gramatikalizace, která je založena především na auxiliaci slovesa *mít*, které vystupuje jako iniciátor gramatikalizačního procesu báзовých konstrukcí [*mít* + *n-/t-participium*], srov. např. *máme napsáno, mám natrénováno, máme zaseto, mám to připraveno* aj. Práce je opatřena množstvím příkladů, tabulek a frekvenční analýzou jednotlivých typů této konstrukce a jejích struktur.

ISBN 978-80-7422-041-8

■ **Mnohojazyčný korpus InterCorp: Možnosti studia**

František Čermák – Jan Koček (eds)

Tento svazek je prvním výstupem z mnohojazyčného paralelního korpusu přinášejícím studie a srovnávací analýzy celkem třinácti jazyků (vůči češtině) v oblasti lexikonu, gramatiky, slovtvorby, frazeologie, syntaxe a dalších.

ISBN 978-80-7422-058-6

■ **InterCorp: Exploring a Multilingual Corpus**

František Čermák – Patrick Corness – Aleš Klégr (eds)

Exploration of grammar, lexis, translations, applications, and methodological issues are studied and illustrated on language pairs or on a group of more languages. This is supplemented by broad and general contributions delineating the field of comparative multilingual corpus linguistics showing possible directions of comparative research based on a multilingual parallel corpus.

ISBN 978-80-7422-042-5

■ **Korpusová lingvistika Praha 2011 – 1 InterCorp**

František Čermák (ed.)

Příspěvky přednesené na konferenci *Korpusová lingvistika Praha 2011* pořádané Ústavem Českého národního korpusu Filozofické fakulty Univerzity Karlovy. Tento svazek se prostřednictvím korpusu InterCorp věnuje srovnávání jazyků na základě paralelních korpusů a navazuje tak na svazek z konference InterCorp z roku 2010. Z více než dvaceti v korpusu zastoupených jazyků se zde ve velmi širokém záběru představují příspěvky týkající se patnácti z nich. Kontrastivní pohled zvláště z oblasti slovtvorby, lexikonu i gramatiky, založený na autentických datech InterCorpu, tak nabízí významný stimul k dalšímu srovnávacímu výzkumu mnoha jazyků. Ten byl donedávna opomíjen, přestože studium druhého jazyka se stále nejčastěji děje na základě srovnání s jiným jazykem, převážně mateřským, jehož vlastní obraz se zároveň prizmatem druhých jazyků může objektivizovat a zpřesňovat.

ISBN 978-80-7422-114-9

■ **Korpusová lingvistika Praha 2011 – 2 Výzkum a výstavba korpusů**

František Čermák (ed.)

Příspěvky přednesené na konferenci *Korpusová lingvistika Praha 2011* pořádané Ústavem Českého národního korpusu Filozofické fakulty Univerzity Karlovy. Tento svazek se v prvním tematickém celku věnuje aspektům výstavby korpusů různého druhu, českým i zahraničním, jazykově obecným i specifickým (např. z fonetického, akustického či stylistického hlediska). Vedle převládajícího synchronního pohledu se tu však objevuje i příspěvek věnovaný problematice výstavby diachronního korpusu češtiny. Pozornost je věnována i korpusu romštiny, v rámci obecných příspěvků se tu podává také přehled aktuálního stavu a problematiky výstavby multilingválního korpusu InterCorp.

Výsledky vlastního výzkumu převážně synchronních dat představuje druhý tematický celek nabízející korpusově podepřený pohled na aspekty fonologické, morfologické a okrajově i syntaktické. Vedle příspěvků orientovaných na lexikální pragmatiku, možné výukové aplikace, regionální specifiku mluvených kor-

pusů a aspekty oikonym je tento pohled doplňován i příspěvkem věnovaným novému nástroji pro korpusově založený synchronní i diachronní výzkum jazykové variability. Do tohoto svazku jsou také zařazeny tři plenární přednášky.

ISBN 978-80-7422-115-6

■ **Korpusová lingvistika Praha 2011 – 3 Gramatika a značkování korpusů**

Vladimír Petkevič – Alexandr Rosen (eds)

Příspěvky přednesené na konferenci *Korpusová lingvistika Praha 2011* pořádané Ústavem Českého národního korpusu Filozofické fakulty Univerzity Karlovy. Tento svazek se věnuje dvěma tematickým oblastem. První z nich je věnována výzkumu gramatiky v převážně synchronních korpusech, obsahově navazuje na příspěvky ve druhém svazku, především na oddíl Aspekty výzkumu korpusů, a přináší kvalifikované vhledy do oblasti slovtvorby, morfologie, ale i do různých oblastí syntaxe. Na ně navazují i aplikovaně orientované příspěvky s orientací na valenci a další. Druhá oblast je orientována spíše technicky a věnuje se různým aspektům syntaktického značkování korpusů. Tematicky ji doplňují příspěvky zabývající se identifikací frazémů a termínů a specifickým žakovským korpusem.

ISBN 978-80-7422-116-3

■ **Deklinační systém femininních oikonym v češtině**

Jana Marie Tušková

Práce J. M. Tuškové, kterou přináší tento svazek Studií z korpusové lingvistiky, je vítaným vhladem do málo a nepřesně poznané oblasti vlastních jmen ženského rodu a jejich skloňování v ČNK, tedy takových jmen jako Kourim, Čáslav, Třeboň, Náměšť, Hluboká, Nusle, Stodůlky, Kateřinky, Rokycany či Pardubice. Kvůli množství dat se tu popisují jen necelé dvě tisícovky těch nejfrekventovanějších. Proti dosavadním pracím a slovníkům, založeným na manuálním a ne nutně vždy systematickým sběru, autorka zvolenou oblast popisuje nejen vyčerpávajícím způsobem, ale dává i představu o jejím jádru i případech periferních.

ISBN 978-80-7422-138-5

■ **Kvantitativní analýza kontextu**

Václav Cvrček

Kvantitativní analýza, popsaná v této knize, se zabývá tím, jaké jsou obecné charakteristiky kontextu, co můžeme považovat za anomální a co za běžné a jak je možné hodnoty v kontextu slov naměřené lingvisticky interpretovat. O klíčové povaze kontextu při analýze jazykových jevů se v jazykovědě obecně nepochybuje. Bezesporný je jeho význam při analýze sémantiky slov. Výsledky zde prezentované tak můžou posloužit jako východisko pro jakákoli jazyková bádání, která se snaží vedle pragmatického aspektu uchopovat i dosud podceňovaný rozměr syntagmatický.

ISBN 978-80-7422-264-1

■ **Odraz jazykových změn v synchronních korpusech**

Michal Křen

Tento svazek představuje diachronní pohled na synchronní psané korpusy řady SYN, které zachycují blízké stavy jazyka. Cílem práce bylo především zjištění možností a mezí detekce vývojových tendencí v jazyce na materiálu korpusů této řady, mezi vedlejší výstupy patří také vyhodnocení jejich složení. Popisovaná metoda je aplikována v několika variantách na různě definované subkorpusy a podrobně vyhodnocena na úrovni lemmat a lexikálních kombinací. Protože je obtížné odlišit zárodky diachronních posunů od přirozeně existující synchronní variability, je statisticky zjištěná významnost frekvenčních rozdílů jednotlivých výrazů zpětně ověřována na korpusech a interpretace výsledků je korigována informacemi o jejich přesném složení.

ISBN 978-80-7422-265-8

■ **Radost z jazyků – Sborník příspěvků k 75. narozeninám prof. Františka Čermáka**

Vladimír Petkevič – Ana Adamovičová – Václav Cvrček (eds)

Jubilejní dvacátý svazek řady Studie z korpusové lingvistiky jsme věnovali jejímu zakladateli prof. Františku Čermákovi, který se v roce 2015 dožívá významného životního jubilea. Prof. PhDr. František Čermák, DrSc. (narozen v Praze 30. ledna 1940) je doma i ve světě uznávanou vědeckou kapacitou. Je autorem či spoluautorem jedenácti monografií, sedmnácti slovníků (překladových, výkladových, frekvenčních), autorem bezpočtu odborných recenzí, editorem a překladatelem. Jeho jméno je spojováno s koncepcí a zpracováním rozsáhlého a mezinárodně uznávaného Slovníku české frazeologie a idiomatiky. Hlavním životním vědeckým počinem je však založení a dvacetileté ideové vedení Ústavu Českého národního korpusu FF UK, čímž položil zcela nový základ české empirické lingvistiky v ČR. Předkládaný sborník je tematicky velmi pestrý, obsahuje příspěvky zabývající se lexikologií, frazeologií, korpusovou lingvistikou, ale také jazykovou kulturou či morfologií, čímž reflektuje košatost vědecké osobnosti prof. Čermáka a jeho široký rozhled po jazycích a jazykovědě.

ISBN 978-80-7422-361-7

■ **Skladební funkce a pád v korpusu: Frekvenční analýza**

Tomáš Jelínek

Tento svazek představuje rozbor vztahu mezi skladebními funkcemi na jedné straně a slovními druhy a pády slov na straně druhé. Práce vychází z korpusu SYN2005, automaticky syntakticky anotovaného podle vzoru analytické roviny Pražského závislostního korpusu. Ukazuje, jak často se které skladební funkce realizují různými slovními druhy a různými pády flektivních slov, a umožňuje také srovnání frekvence jevů podle žánrových subkorpusů. Navazuje na publikaci Statistika češtiny, kterou doplňuje o syntaktický pohled.

ISBN 978-80-7422-366-2

Skladební funkce a pád v korpusu: Frekvenční analýza

Myriad Pro Semiold 21/23,
linka pod: 3b 25%černé, odsazení 2mm

Tomáš Jelínek Myriad Pro Regular 12/12

Vydalo NLN, s. r. o.,
Nakladatelství Lidové noviny,
Dykova 15, 101 00 Praha 10,
v roce 2015
Vydání první

tiráž: Charis SIL Italic 9/11,
zarovnání vlevo bez dělení

Jazyková a odborná redakce Vladimír Petkevič
Obálku navrhla Michaela Blažejová
Knihu graficky upravila Marie Tvrďá
Technická redakce Josef Kupka
Sazba písmem Utopia Marie Tvrďá
Tisk Art D – Grafický ateliér Černý s. r. o., Praha

Doporučená cena včetně DPH 129 Kč

ISBN 978-80-7422-366-2