

ČEŠTINA

u n i v e r z á l i a
a

s p e c i f i k a

5

2004

ČEŠTINA – UNIVERZÁLIA A SPECIFIKA

5

Sborník 5. mezinárodního setkání bohemistů
v Brně 13.-15. 11. 2003

Separát

Neil BERMEL: *V korpuse nebo v korpusu? Co nám řekne (a neřekne)
ČNK o morfologické variaci v tvarech lokálu*163-171
*V KORPUSE OR V KORPUSU? WHAT THE CZECH NATIONAL CORPUS
CAN (AND CAN'T) TELL US ABOUT MORPHOLOGICAL VARIATION IN
THE LOCATIVE SINGULAR*

Editors © Z. Hladká, P. Karlík
Masarykova univerzita v Brně

Vydalo Nakladatelství Lidové noviny (NLN, s.r.o.), Praha, v roce 2004
ISBN 80-7106-659-1

Citace: Hladká, Z.–Karlík, P. (eds.): Čeština – univerzália a specifika, 5.
Praha: Nakladatelství Lidové noviny 2004

OBSAH

HOSTÉ

| | |
|---|---|
| Joseph E. EMONDS: Organizace lexikonu | 9 |
| <i>THE ORGANIZATION OF LEXICON</i> | |

FÓRUM

| | |
|--|----|
| Petr KARLÍK: Mají dějová substantiva slovesný rod? | 33 |
| <i>ARE DERIVED ACTION NOMINALS SENSITIVE TO THE ACTIVE - PASSIVE VOICE DISTINCTION?</i> | |
| Aleš SVOBODA: Poznámka o českých zjišťovacích otázkách v budoucím čase ve srovnání s anglickými | 47 |
| <i>A NOTE ON CZECH FUTURE TENSE YES-NO QUESTIONS IN COMPARISON WITH THOSE IN ENGLISH</i> | |
| Josef ŠIMANDL: Korpus jako (jediný) zdroj? | 50 |
| <i>EIN KORPUS ALS (EINZIGE) QUELLE?</i> | |

SYMPOZIUM – REFERÁTY

| | |
|--|-----|
| Markus GIGER: 55 let po Garvinovi: Čeština a standardní průměrná evropština | 58 |
| <i>55 JAHRE NACH GARVIN: TSCHECHISCH UND STANDARD-DURCH- SCHNITTSEUROPAÏSCH</i> | |
| Galina NEŠČIMENKOVÁ: K zintenzivnění tendencí konvergentního vývoje současných slovanských jazyků | 67 |
| <i>K АКТИВИЗАЦИИ ТЕНДЕНЦИЙ КОНВЕРГЕНТНОГО РАЗВИТИЯ СОВРЕМЕННЫХ СЛАВЯНСКИХ ЯЗЫКОВ</i> | |
| Petr SGALL: K obohacování spisovné češtiny | 77 |
| <i>ON ENRICHING STANDARD CZECH</i> | |
| Masako U. FIDLEROVÁ: Zvukový symbolismus v češtině (na pozadí japonštiny a jiných jazyků) | 86 |
| <i>SOUND SYMBOLISM IN CZECH: AN APPROACH BASED ON JAPANESE AND OTHER LANGUAGES</i> | |
| Valerie VASILJEVOVÁ: Mezijazyková funkční asymetrie z hlediska specifiky příbuzných jazyků | 94 |
| <i>FUNCTIONAL ASYMMETRIES IN RUSSIAN AND CZECH FROM THE VIEWPOINT OF RELATED LANGUAGES</i> | |
| František ČERMÁK: Jazyková variabilita: Příklad přísloví | 99 |
| <i>LANGUAGE VARIABILITY: CONSIDERING PROVERBS</i> | |
| Marek NEKULA: Deminutiva a zdvořilost | 110 |
| <i>DIMINUTIVE UND HÖFLICHKEIT</i> | |

| | |
|---|-----|
| Milada HIRSCHOVÁ: Ilokuční slovesa: Mezi pragmatikou a syntaxí | 120 |
| <i>ILLOCUTIONARY VERBS: BETWEEN PRAGMATICS AND SYNTAX</i> | |
| George CUMMINS: Funkce hovorového instr. pl. v komunikační výpovědi | 129 |
| <i>THE FUNCTION OF THE COLLOQUIAL INSTRUMENTAL PLURAL IN THE UTTERANCE</i> | |
| François ESVAN: Budoucí čas a vid v češtině: Několik poznámek na základě korpusových dat | 136 |
| <i>FUTUR ET ASPECT DU VERBE EN TCHÈQUE: QUELQUES REMARQUES SUR LA BASE DES DONNÉES DU CORPUS</i> | |
| Vladimír PETKEVIČ: Využití pravidel pro negaci v automatickém značkování českých korpusů | 143 |
| <i>THE EXPLOITATION OF NEGATION RULES FOR THE AUTOMATIC TAGGING OF CZECH CORPORA</i> | |
| Karel PALA & Radek SEDLÁČEK & Marek VEBER: Vztah mezi tvarotvornými a slovtvornými vzory v češtině | 151 |
| <i>RELATIONS BETWEEN INFLECTIONAL AND DERIVATIONAL PATTERNS IN CZECH</i> | |
| Neil BERMEL: <i>V korpuse nebo v korpusu?</i> Co nám řekne (a neřekne) ČNK o morfoloické variaci v tvarech lokálu | 163 |
| <i>V KORPUS OR V KORPUSU? WHAT THE CZECH NATIONAL CORPUS CAN (AND CAN'T) TELL US ABOUT MORPHOLOGICAL VARIATION IN THE LOCATIVE SINGULAR</i> | |
| Peter KOSTA & Jens FRASEK: Neakuzativita (ergativita) vs. neergativita v češtině, polštině a jiných slovanských jazycích na rozhraní morfologie a syntaxe | 172 |
| <i>UNACCUSATIVITY (ERGATIVITY) IN CZECH, POLISH AND SOME OTHER SLAVIC LANGUAGES AT THE MORPHO-SYNTACTIC INTERFACE</i> | |
| Barbora SKARABELA & Cathy O'CONNOR & Joan MALING: Monolexemické posesory v češtině a jiných jazycích | 195 |
| <i>MONO-LEXEMIC POSSESSORS IN CZECH AND OTHER LANGUAGES</i> | |
| Ludmila VESELOVSKÁ: Rozšířená verbální projekce v češtině: tři druhy slovesa <i>být</i> | 203 |
| <i>THREE KINDS OF THE VERB BE IN CZECH</i> | |
| Hana ŠKRABALOVÁ: Koordinace a struktura nominálních frází v češtině | 213 |
| <i>COORDINATION AND THE STRUCTURE OF NOUN PHRASES IN CZECH</i> | |
| Tobias SCHEER: O samohláskové délce při derivaci v češtině | 224 |
| <i>VOWEL LENGTH IN THE DERIVATIONAL MORPHOLOGY OF CZECH</i> | |

| | |
|--|-----|
| Oldřich ULIČNÝ & Eva HRACHOVINOVÁ: Příspěvek k problematice formalizace v jazykovědě: Možnosti řešení různých případů kongruence | 240 |
| <i>ZUM PROBLEM DER FORMALISIERUNG IN DER SPRACHWISSENSCHAFT: VERSCHIEDENE LÖSUNGEN DER KONGRUENZ</i> | |
| Tilman BERGER: Osud nepravidelných substantiv v češtině | 246 |
| <i>DAS SCHICKSAL DER UNREGELMÄSSIGEN SUBSTANTIVE IM HEUTIGEN TSCHECHISCHEN</i> | |
| Jan ŠTĚPÁN: Otázky a rozkazy v logice | 255 |
| <i>QUESTIONS AND IMPERATIVES IN LOGIC</i> | |
| Milan RŮŽIČKA: Geometrie rysů českého finita | 265 |
| <i>FEATURE GEOMETRY OF CZECH FINITE VERB</i> | |
| Mojmír DOČEKAL: Formální sémantika adnominálního genitivu | 276 |
| <i>FORMAL SEMANTICS OF THE CZECH ADNOMINAL GENITIVE</i> | |
| Pavel CAHA: Konstrukce akuzativu s infinitivem v češtině | 287 |
| <i>ACCUSATIVE + INFINITIVE CONSTRUCTIONS IN CZECH</i> | |
| Petr BISKUP: Slovní pořádek, scrambling a adverbialia v češtině | 302 |
| <i>WORD ORDER, SCRAMBLING AND ADVERBIALS IN CZECH</i> | |
| Jana HOFFMANNOVÁ: „Mlžení“, „kusejrování“ a některé komunikační strategie českých politiků | 312 |
| <i>HOW CZECH POLITICIANS “ANSWER” – OR RATHER AVOID – THE QUESTIONS</i> | |
| Světlá ČMEJRKOVÁ & František DANEŠ: Svě a cizí z hlediska funkčního | 321 |
| <i>“ONE’S ONE” AND “FOREIGN” FROM A FUNCTIONAL POINT OF VIEW</i> | |
| Petr MAREŠ: Hry na neformálnost. K neformálnímu vyjadřování v současných českých odborných textech | 332 |
| <i>SPIELE MIT DEM INFORMALEN. ÜBER DIE INFORMALE AUSDRUCKSWEISE IN GEGENWÄRTIGEN TSCHECHISCHEN FACHTEXTEN</i> | |
| Joanna MIELCZAREKOVÁ: Zvláštnosti obecné češtiny jako poetizačního nástroje (na základě polských překladů novější české prózy) | 340 |
| <i>ON CZECH COLLOQUIAL SPEECH AS A POETIC MEANS OF EXPRESSION (BASED ON POLISH TRANSLATIONS FROM MODERN CZECH LITERATURE)</i> | |
| Markéta LOPATKOVÁ & Jarmila PANEVOVÁ: Valence vybraných skupin sloves (k některým slovesům dandi a recipiendi) | 348 |
| <i>VALENCY OF SELECTED GROUPS OF VERBS (ON SOME CZECH VERBS “DANDI” AND “RECIPIENDI”)</i> | |
| Renata BLATNÁ: Valence víceslovných předložek s nominativem (na materiálu korpusu SYN2000) | 357 |
| <i>MULTI-WORD PREPOSITIONS WITH NOMINATIVE CASE VALENCY (BASED ON MATERIAL FROM THE SYN2000 CORPUS)</i> | |

| | |
|--|-----|
| Ludmila UHLÍŘOVÁ: O zdvojené (spřažené, anebo nespojité) deixi v češtině..... | 367 |
| <i>ON DOUBLE DEIXIS IN CZECH</i> | |
| Karel OLIVA: O gramatických a negramatických konstrukcích se vztažnými zájmeny <i>který</i> a <i>jenž</i> | 378 |
| <i>ON GRAMMATICAL AND UNGRAMMATICAL CONSTRUCTIONS INVOLVING THE RELATIVE PRONOUNS KTERÝ A JENŽ</i> | |
| Norbert NÜBLER: Česká participia ze syntaktického hlediska | 384 |
| <i>TSSCHECHISCHE PARTIZIPIEN AUS SYNTAKTISCHER SICHT</i> | |
| František ŠTÍCHA: Opisné pasivní konstrukce s relátorem <i>dostat</i> a trpným participiem v současné češtině: Úzus a gramatičnost..... | 391 |
| <i>PERIPHRASTISCHE PASSIVKONSTRUKTIONEN MIT DOSTAT (BEKOMMEN) UND DEM -N/T-PARTIZIP IM MODERNEN TSSCHECHISCHEN: USUS UND GRAMMATIKALITÄT</i> | |
| Zdena PALKOVÁ: Přízvukový takt ve struktuře češtiny | 399 |
| <i>THE STRESS-GROUP IN CZECH PROSODIC STRUCTURE</i> | |
| Šárka ŠIMÁČKOVÁ: Kvalita české laterály u mladých mluvčích | 409 |
| <i>QUALITY OF THE CZECH LATERAL LIQUID IN THE PRONUNCIATION OF YOUNG SPEAKERS</i> | |
| Bohumil VYKYPĚL: Komutace v glosematice a ve fonologii: Příklad staročeských likvid..... | 420 |
| <i>LA COMMUTATION EN GLOSSÉMATIQUE ET EN PHONOLOGIE: LE CAS DES "LIQUIDES" EN VIEUX TCHÈQUE</i> | |
| Pavel MACHAČ: Stabilita zvukových charakteristik fonémů ve spontánních mluvených projevech..... | 427 |
| <i>STABILITÄT DER SCHALLEIGENSCHAFTEN DER PHONEME IN DER SPONTANREDE</i> | |
| Slavomír ONDREJOVIČ: Z česko-slovenského konfrontačního štúdia..... | 436 |
| <i>ZU DEM TSCHECHISCH-SLOWAKISCHEN KONFRONTATIONSSTUDIUM</i> | |
| Ondřej ŠEFČÍK: Alternace konsonantů v češtině – fonotaktické a morfonologické modelování systému | 441 |
| <i>ALTERNATIONS OF CONSONANTS IN CZECH – PHONOTACTICAL AND MORPHONOLOGICAL MODELING OF SYSTEM</i> | |
| Aleš SVOBODA: Bohemistika a jazykovědný časopis..... | 448 |
| <i>CZECH STUDIES AND A LINGUISTIC E-JOURNAL</i> | |
| Stefan Michael NEWERKLA: Josef Valentin Zlobický (1743-1810) – vídeňský obrozenec a osvícenec | 451 |
| <i>JOSEF VALENTIN ZLOBICKÝ (1743-1810) – ERNEUERER UND AUFKLÄRER AUS WIEN</i> | |

V KORPUSE NEBO V KORPUSU? CO NÁM ŘEKNE (A NEŘEKNE) ČNK O MORFOLOGICKÉ VARIACI V TVARECH LOKÁLU¹

NEIL BERMEL

n.bermel@sheffield.ac.uk

0. Úvod

V tomto příspěvku se pokusím o sondu do variací v 6. pádě jednotného čísla jmen mužského rodu: jde o časté kolísání mezi tvary na *-e* a *-u*. Je na místě se ale zeptat předem, k čemu může vést výzkum v oblasti, která je pro rodilého mluvčího češtiny zcela neproblematická a ani se netýká tak dobře známých otázek, jako jsou dichotomie *spisovnost* × *nepisovnost*, *nářečí* × *standard* atd.

Lokál muž. rodu sg. není uveden v *Slovníku spisovné češtiny pro školu a veřejnost* ani v nových vydáních *Pravidel českého pravopisu*. Člověk musí sáhnout až po osmidílném *Slovníku spisovného jazyka českého* nebo po překladových slovnících, aby měl naději, že se dozví, které varianty se uznávají.²

Tento oríšek pomíjejí i mluvnice češtiny, kde se uvádějí většinou jen obecné pokyny nebo jednotlivé příklady (které jsou ovšem v každé gramatice jiné). F. Trávníček (1951:462-3) se zaměřuje na variaci podle zakončení kmenu a uvádí dlouhé seznamy kolísajících slov. B. Havránek & A. Jedlička (1981:138-9) pro koncovku *-u* přidávají kategorie „jmen dějových a vůbec abstraktních“ a „jmen nově vzniklých (zvláště u odborných názvů)“ a pro koncovku *-ě/-e* (dále jen *-e*) uvádějí pojem „jmen místních... označující... prostor, zvláště ve spojení s předložkou *v*, *na*“. U nich se také zmiňují o vztahu mezi tvarem 6. a 2. pádu sg., že genitiv na *-a* obvykle souvisí s lokálem na *-e*. Nejnovější *Příruční mluvnice češtiny* (1995) zdůrazňuje tzv. „adverbiální“ užití pevných spojení „předložka + podstatné jméno“, která inklinují k užití koncovky *-e*, oproti tzv. „objektovému“ užití koncovky *-u*.

V různých zdrojích se odkazuje na nářeční rozdíly u jednotlivých slov: v Čechách se prý preferuje *-u* oproti moravskému *-e*. Distribuce *-e* a *-u* ovšem patrně nefunguje jako obecně uznávaný diferenční rys nářečních skupin.

„Akademická“ *Mluvnice češtiny II* (1986:305-6) zaujímá vůči lokálu postoj prozrazující skoro zoufalství, zato však výstižně zachycuje stav úzu:

¹ Výzkum této otázky vznikl částečně s podporou British Academy, jejíž granty pro cestovní náklady v rámci meziakademické dohody s AV ČR umožnily dva pobyty v ČR během r. 2003.

² Detailní rozbor slovníkových zpracování lokálu je podán u G. Cumminse (1995). Mezi českými vědci si tato problematika získala jen občasnou pozornost, viz např. J. Kořený (1972), L. Klišeš (1953).

„V dalším pádě – LSg – jsou polymorfní formanty *-u* a *-ě* (*-e*), jejich distribuce není však pevně vyhraněna – na rozdíl od rozložení formantů v GSg. Za týchž podmínek se u téhož jména užívá jedné i druhé podoby; uplatňuje se tu tedy v značné míře tvarová variantnost. Jistou motivací k častějšímu užití jen jedné z podob bývá zřetel syntaktický, tzn. zda jde o objekt, nebo o adverbální určení, někdy má vliv původ slova [...].“

A v bezprostředně následující poznámce se dodává: „O nesnadnosti stanovit distribuci podob v LSg svědčí i praxe jazykových příruček a výkladových slovníků: snaží se popsat současný stav pomocí výčtu jednotlivých jmen a relativně působících činitelů. Vzniká tak komplikovaný obraz dokazující pouze tu skutečnost, že jde o podoby natolik rovnocenné, že užití kterékoli z nich neubírá na zřetelnosti významu ani větné stavbě [...].“

Z těchto popisů vyplývá, že Češi zřejmě většinou s výběrem tvarů v lokálu nemají potíže; kde mají na vybranou ze dvou tvarů, jeden z nich může „znít lépe“ či „vhodněji“, nebo mohou pokládat oba tvary za rovnocenné. V dřívější studii tohoto jevu (N. Bermel, 1993) mě zajímala „estetická reakce“ Čechů na tyto variantní tvary, tj. který tvar, *-e* nebo *-u*, zní v různých situacích přirozeněji a jaké „působící činitele“ se tu uplatňují. Výzkumy ukázaly, že Češi dávali přednost koncovce *-e* v kontextech, které zdůrazňují místní charakter 6. pádu, kdežto přijatelnost koncovky *-u* stoupala naopak v kontextech, kde lokál byl čistě gramatický jev, a v kontextech místních, ale méně typických či méně běžných.

V rámci projektu *Možnosti a meze české gramatiky ve světle Českého národního korpusu* (dále MaM; viz webová stránka mam.ujc.cas.cz) jsem se pokusil o další sondu do této problematiky, tentokrát ale jinak zaměřenou. Český národní korpus (dále ČNK) je reprezentativně sestaven, tj. svou kompozicí odráží psaný svět, který obklopuje průměrně vzdělaného Čecha.³ Nebylo proto nezajímavé v korpusu zjišťovat, jestli estetické volby Čechů zjištěné v mé již citované starší studii odpovídají skutečnému výskytu těchto tvarů v psaném jazyce a jestli nám korpus může poskytnout nějaký širší kontext – nazvěme ho **frekvenční pozadí**, který by částečně vysvětlil volby mluvčích.

1. Parametry variace v korpusu

Jaké jsou obecné parametry variace v lokálu? V korpusu se taková variace hledá pomocí tagů, tj. detailních morfologických popisů, které doprovází každý tvar. Dotaz

`[tag="NNIS6.*[12345679]"]`

vyhledává:

- obyčejná substantiva (NN)
- neživ. rodu (I)
- v 6. pádě j.č. (S6),
- která mají variantní koncovku (číslůvky 1-9 na konci tagu).

³ Viz M. Šulc (2001), J. Králík (2001); o reprezentativnosti korpusu a o jeho využití ve výzkumu mluvnických otázek viz F. Štícha (2001).

Výsledek hledání obsahoval přes 87 000 dokladů. Vytřídění s ohledem na velká písmena vyřadilo hned přes 10 000 místních názvů, osobních jmen atd. Po vynechání duplicit tvoří zbývajících 831 dokladů de facto seznam údajně kolísajících slov evidovaných v korpusu.⁴ Valná většina těchto dokladů eviduje kolísání mezi koncovkami *-e* a *-u*, počet jiných doložených variací – např. se slovem *kmen* – je zanedbatelný.

Háček je ale v tom, že čtvrtina těchto tvarů je chybně otagována a ve skutečnosti dva tvary v lokálu nemají. Např. ze 33 slov začínajících na *v-* se 9 (*velkoklub, velkosklad, veřtat, videoart, videoklub, vigvam, voluntariát, vous, výbrus*) podle slovníku vůbec nedá s koncovkou *-e* spojit. Koncovka *-u* je totiž základní a nejde tu o konkurenci tvarů. Zbývajících 24 slov (*vagón, val, věhlas, velechrám, velín, velkoobchod, vepřín, vinohrad, vjezd, vlas, vokál, vokativ, volant, volejbal, vůz, východ, výjezd, výkaz, výklad, výkres, výmol, výnos, výpis, výraz*) má aspoň podle slovníku variantní tvary na *-e* a *-u*. Přes velký počet tvarů chybně označovaných jako variantních jde o víc než **600 podstatných jmen** s potenciální variací.

Zajímavé je, že ČNK nesvědčí o tak rozsáhlé variaci, jakou najdeme v SSJČ. Tento slovník nabízí např. pro slova *vjezd* a *podpis* variantní tvary v lokálu: *podpise* či *podpisu, vjezdě* či *vjezdu*. Tvary *vjezdě* a *podpise* ale vůbec nejsou doloženy v korpusu a rodilí mluvčí je už nepovažují za alternativy v současném jazyce přijatelné. (SSJČ totiž, jak známo, obsahuje tvary, které nejsou v běžném psaném úzu, ale které se v době jeho vzniku ještě nepociťovaly jako archaické.)

ČNK rovněž nepostihuje celou škálu variací v mluveném jazyce. Např. slovo *korpus* má v mluveném úzu základní lokál *v korpusu*, ačkoli spojení *v korpusu* můžeme slyšet také. Tato druhá varianta se v ČNK ale vůbec nevyskytuje, což svědčí o tom, že korpus je přece jen soubor textů, které prošly jazykovou redakcí, a variace může být proto o něco užší a i konzervativnější než v běžně mluveném jazyce.

2. Možnosti a meze vyhledávání

N. Bermel (1993) se opíral o zkoumání sedmnácti slov, kdežto současná studie se omezuje na několik z nich: *autobus, balkon, časopis, kabát, pivovar, rybník*. Postup vyhledávání byl v hlavních rysech podobný pro každé slovo. Nejdříve byly vyhledány všechny doklady daného slova v lokálu pomocí jedno-
duchého dotazu, který se zaměřuje na „tvary slova *x* v 6. pádě j. č.“, např.:

`[(lemma="pivovar") & (tag="...S6.*")]`

Pomocí třídění byla pak zjištěna frekvence obou tvarů a jejich kolokace s různými předložkami. Další dotaz se týkal výskytu frází s řetězy přívlasktů:

`[(word="[Vv]")(word="[Vv]e")[tag="[PCA]..S6.*"]]{1,3}
[(lemma="pivovar") & (tag="...S6.*")]`

⁴ Podle *Mluvnice češtiny II* (1986) má variaci v tvarech lokálu muž. rodu kolem 1100 slov v SSJČ (viz také G. Cummins, 1995). Vzhledem k zastaralosti dat v SSJČ není divu, že tagovací program ČNK zachycuje jen část této slovníkové variace.

Uvedený dotaz vyhledal všechny doklady slova *v* nebo *ve*, pak maximálně tři přívlasky a slovo *pivovar* v lokálu sg.

Podle původního výzkumu jsem se zaměřoval na několik sémantických faktorů: **lokálnost, typičnost, známost, druh předmětu, obvyklost spojení**. Některé tyto faktory se vyjadřují pomocí různých formálních prostředků, které se v korpusu dobře vyhledávají. Lokálnost se může vyjádřit užitím předložek *v*, *na* či *po* oproti nelokálnímu *o*. Přítomnost přídavných jmen může zase ukazovat na typičnost, či naopak na neobvyklost určitého spojení.

Některé formální prostředky se však v korpusu vyhledávají obtížně. Slovesná vazba s lokálem není častá, ale je důležitá tím, že je sémanticky prázdná. Dá se v korpusu vyhledat, ale výsledky nejsou příliš informativní. Pomocí následujícího dotazu najdeme podstatnou část výpovědi s vazbou *pátrat po něčem*:

```
([lemma="pátrat"] [] {0,2} [word="po"] [tag="A..S6.*"] {0,2}
[tag="NNIS6.*"])
```

Výsledky zahrnují 96 dokladů, většinou v kolokacích *pátrat po zdroji*, *po původu* nebo *po osudu*. Jsou ale i kolokace se slovy z našeho seznamu kolisajících slov: *pátrání po voze*, *pátral po výmluvném potu*, *pátral po správném obchodě*, *pátrám po životě*, *pátral po celém domě*, *pátral po stropě*, *pátraje po dopisu*, *pátrají po vodopádu*, *pátrá zde po archivu*, *pátral po zápase*. Je ovšem otázka, co nám tyto ojedinělé doklady řeknou bez širšího pozadí pro každé podstatné jméno. Proto není takový postup příliš přínosný, a zůstaneme tedy u vyhledávání podle jednotlivých podstatných jmen.

3. Data z korpusu

Korpusové údaje většinou potvrzují, že se v lokálu nepovoluje jednoduché rozdělení tvarů na základní a okrajové. Další možnost by byla tzv. nulová hypotéza, podle které je variace víceméně úplně volná a vyskytí jedné či druhé koncovky se řídí jen nějakou vágní osobní preferencí. Mezi těmito dvěma póly je ale třetí, snad pravděpodobnější možnost: **škála různých křížících se možností**, do které přispívají určité sémantické, či možná kvantitativní faktory.

3.1 *Autobus*

Situace se jasně ukazuje na příkladu slova *autobus* (429 dokladů v 6. pádě j.č.). Z nich 292 je ve tvaru *autobuse* a 137 ve tvaru *autobusu*. Jedna koncovka se totiž zjevně preferuje, ale druhá je zastoupena také hojně.

Po podrobení výsledků bližší analýze se zjistí, že v kontextech s více než jen několika doklady jde vždy o nějakou proporcii, v žádném případě o přesně vyhraněné užití jedné či druhé koncovky.

Dominantní předložkou je *v*, které se vyskytuje v přímém spojení s podstatným jménem *autobus* 324krát (75,5 % všech výskytů). S předložkou *v* převládá koncovka *-e* (doložena 234krát, či 72 %).

Opačná tendence se naproti tomu ukazuje u méně frekventovaných předložek *na*, *o* a *po*. Předložka *na* je doložena 8krát, z toho 5krát s koncovkou *-u*. Předložka *o* se vyskytuje jen dvakrát a *po* dokonce jen jednou, všechny s koncovkou *-u*.⁵

Obdobná situace je ve spojeních, v nichž je mezi předložkou a podstatným jménem jeden nebo více přívlasktů. Tady je frekvence koncovky *-u* o něco vyšší: vyskytuje se ve 30 ze 75 dokladů, tedy ve 40 %.

U daného substantiva se proto nedá říct, že jedna koncovka je základní či centrální a druhá okrajová či periferní. Místo toho je jedna koncovka frekventovanější a druhá méně frekventovaná. Přínos korpusových dat tu spočívá v tom, že jasně ukazují, že pro toto slovo jsou některé **sémantické a syntaktické kontexty** frekvenčně spíše základní a jiné spíše okrajové a že výskyt méně frekventované koncovky je vyšší v okrajových kontextech.

3.2 *Balkon*

U některých dalších slov jde doopravdy o koncovku základní a okrajovou; okrajová koncovka však není v žádném kontextu natolik periferní, aby byla vyloučena úplně. Spodní hranice frekvence je tu něco kolem 5 %. Slovo *balkon* nám poslouží jako příklad, ale podobné jsou např. *časopis*, *hrob* a *kabát*.

V korpusu bylo pro slovo *balkon* (s délkou nebo bez délky) 271 dokladů 6. pádu j. č. Valná většina – 86,7 % – byla s tvary *balko/ón-ě*.

Pokud probereme doklady podle předložek, dominantní je *na*, která má 82,3 % všech dokladů. Tvary *balko/óně* se tu vyskytují o něco častěji: v 90,1 % dokladů.

Okrajové předložky *v* a *o* mají po jednom dokladu, v obou případech s koncovkou *-u*. Lokální předložka *po* měla 12 dokladů, ze kterých 7 bylo ve tvaru *balko/ón-ě* a 5 ve tvaru *balko/ón-u*. Výskyt koncovky *-u* je tu značně vyšší, nacházíme ji ve 41,7 % dokladů.

Výskyt koncovky *-e* je také snížený ve spojeních s různými přívlaskty. Např. z 24 dokladů spojení *na* + přívlaskty + *balko/ón-* má 18 dokladů, tj. jen 75 % oproti celkovému výskytu, koncovky *-e* (86,7 %).

3.3 *Kabát*

Slovo *kabát* poskytuje trochu jiný obraz. Zde je zase tvar *kabát-ě* zjevně základní, má ho 89,3 % všech výskytů (241 ze 270). Dominantní strukturou z kvantitativního hlediska však není přímé spojení předložky s podstatným jménem, nýbrž předložka *v* + přívlaskty + *kabát-*. Takovou strukturu má 58 % všech dokladů a z těchto 157 dokladů jich 141, čili 89,8 %, má tvar *kabát-ě*.

⁵ V případě, že jde o ojedinělé výskyt nějakého tvaru (1-5 dokladů), je nutné ověřit, zda všechny doklady nepochází ze stejného zdroje nebo od stejného autora. Přítomnost delších beletrických textů v korpusu může nálezy značně ovlivnit. Takové výsledky proto prošly kontrolou, aby jednotlivé projevy autorské svéráznosti nebyly mylně uvedeny jako všeobecné „trendy“.

Podobné výsledky zjišťujeme u spojení s předložkou *na*, která jsou kvantitativně okrajová, ale sémanticky typická (snadno si představíme, že něco se nachází *na kabátě*). V přímém spojení s předložkou *na* má všech 20 dokladů tvar *kabát-ě* a se vsunutým přívlaskem jen jeden z devíti dokladů měl tvar *kabát-u*.

Výskyt tvaru variantního, totiž *kabát-u*, je vyšší v přímém spojení s předložkami *v* a *po*. S předložkou *v* má 37 dokladů (či 82,2 %) tvar *kabát-ě* a 8 dokladů (či 17,8 %) tvar *kabát-u*. S předložkou *po* jsou celkem čtyři doklady, dva s tvarem *kabát-ě* a dva s tvarem *kabát-u*.

3.4 Pivovar

U slova *pivovar* převládá koncovka *-u*. Celkový výskyt slova *pivovar* v 6. pádě j. č. je 260 dokladů, ze kterých 27,3 % tvoří tvar *pivovař-e* a zbytek tvar *pivovar-u*.⁶

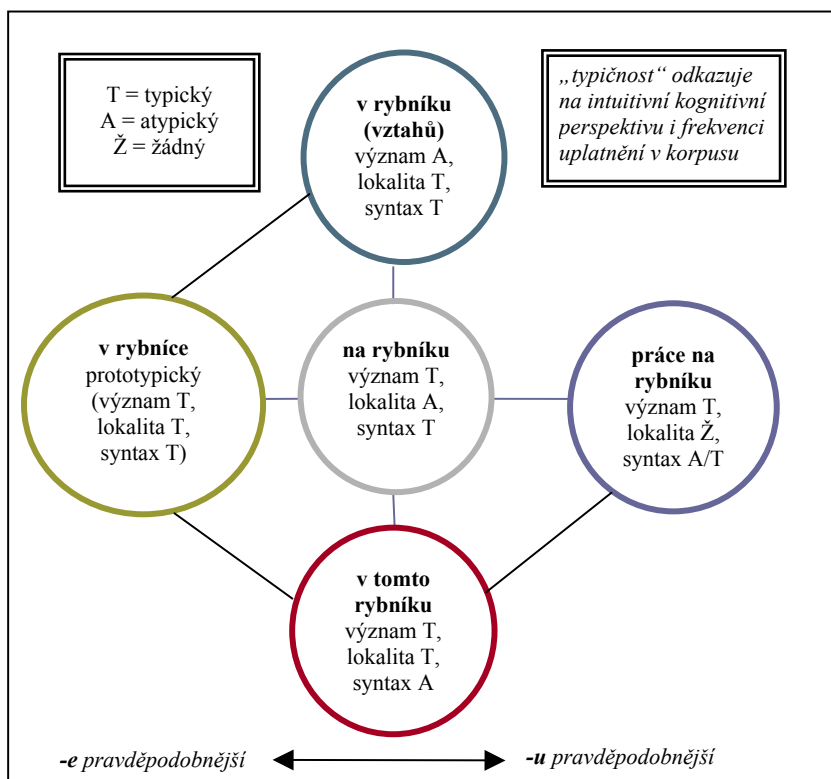
Dominantním kontextem je předložka *v* v přímém spojení s podstatným jménem. Takové kontexty tvoří 54,2 % všech dokladů. Z nich má 30,5 % menšinovou koncovku *-e* a zbytek většinovou koncovku *-u*.

Další početná skupina tvoří kontexty typu „*v* s jedním až třemi přívlasky a podstatným jménem“. Takových dokladů je v korpusu 88, ze kterých 80,7 % má tvar *pivovar-u*. Variace tu nejsou značné, ale zase potvrzují obecný trend, že koncovka *-e* je poněkud lépe zastoupená v kontextech typických pro daný lexém.

Periferní kontexty jsou např. spojení slovesa či podstatného jména se sémanticky prázdnou předložkou *v*, *na*, *o* nebo *po* a 6. pádem. Manuální třídění výsledků poskytuje seznam těchto kombinací: *podíl na pivovaře*, *podíly na pivovaru*, *podíl na největším polském pivovaru*, *vydělat na pivovaru*, *žádali po pivovaru*. Jen jeden z pěti dokladů má koncovku *-e*.

Dále jsou zajímavé nelokální významy předložek. Předložka *o* je podle významu zásadně nelokální a má 4 doklady, všechny s tvarem *pivovar-u*. Dále se vyskytuje 26 dokladů nelokálního užití předložek *v* a *po*. V případě slova *pivovar* jde převážně o kontexty, kde spojení *v pivovaru* znamená „*v pivovarské společnosti*“, např. *podíl v pivovaru*, *účast v pivovaru*, *zlatá akc[i]e v pivovaru*, *získat v pivovaru majoritu*, *investor v pivovaru*. Předložka *po* se užívá v časovém významu: *používat budovu po pivovaru*, *po holešovickém pivovaru nyní Bass uzavírá také pivovar ve Vratislavicích*. Mezi doklady s tímto významem není žádný s koncovkou *-e*.

⁶ Tento příklad vyvolal na sympoziu určité pochyby, a proto je tu na vyzvu kolegů uveden stručný souhrn zdrojů s tvarem *pivovaře*. Ze 71 dokladů celkem bylo 41 z beletrie, 21 z publicistiky a jednotlivé citáty z písní, literatury faktu, vědeckonaučné literatury, populárněnaučné literatury a eferméry. Z beletrie je hojně zastoupen Hrabal (23 dokladů). Zbytek pochází z díla Durycha, Vaculíka, Havla, Škvoreckého, Gruši a z překladů Dickense. I když málem jeden ze tří dokladů je z Hrabalových textů a výskyt tohoto tvaru je značně častější v beletrii (15 % korpusu obsahuje 57,8 % dokladů), koncovka *-e* je také zřejmě používána i jinde.



Obrázek 1. Kognitivní struktura frází se slovem *rybník*

3.5 *Rybník*

Distribuce dvou variantních koncovek je vyrovnanější u slova *rybník*, kde koncovka *-e* je doložena v 62,1 % příkladů a koncovka *-u* ve zbývajících 37,9 %.

Největší počet dokladů je zase u spojení *v rybníku/v rybníce*, doloženého celkem 133krát, tj. v pěti z každých dvanácti dokladů. Právě s tímto spojením je počet dokladů s jednou koncovkou velmi blízký počtu dokladů s druhou koncovkou – *rybníce* 69krát a *rybníku* 64krát. Zde snad hraje nějakou roli frazeologismus *štika v rybníce*. Tento výraz se užívá až na jednu výjimku s tvarem *rybníce* a je doložen v korpusu 12krát.

Naopak spojení „*v* + přívlastek + *rybníce*“ vykazuje jen průměrný výskyt koncovky *-e* (27 ze 70 dokladů, či 38,6 %). Potenciální důvod najdeme v korpusových datech, která jasně ukazují, že slovo *rybník* má i obrazný význam, který není zachycen ve slovnících. Tento význam „uzavřeného dějiště“ se pojí většinou s přívlastkem či dalším komplementem: *klidný život v prvologovém rybníce, stanu se dravou rybou v líném parlamentním rybníku, kal v ryb-*

níku čínsko-amerických vztahů. Užití tohoto méně „lokálního“ významu proto o něco zesiluje tendenci k užívání koncovky *-u* s předložkou *v*.

Ve spojeních s předložkami *na*, *po* a *o* má značně vyšší výskyt koncovka *-u*. Se vsunutým přívlastkem je koncovka *-e* velmi řídká – setkáváme se s ní jen v mála případech. To je pozoruhodné proto, že slovo *rybník* se pojí v čistě lokálním smyslu s třemi předložkami: *v*, *na* a *po*. Podle četnosti výsledků se dá soudit, že nejčastější pojem lokálnosti je *rybník* jako **prostor trojrozměrný**, ve kterém plave člověk, ryba nebo jiné zvíře. Méně rozšířený je *rybník* jako **prostor dvojrozměrný**, kde pozorujeme jenom jeho hladinu. Frekvence předložek tedy odpovídá přesně výskytu koncovek *-e* a *-u*, kde *-e* je častější u „nejtypičtějších“ lokálních prostorů.

4. Závěry

Na závěr je na místě zeptat se, jestli se vyskytly obecné korespondence mezi korpusovými údaji a nálezy dřívější studie, a zhodnotit příspěvek korpusových dat k této problematice vůbec. Následující výklad se opírá o přístupy kognitivní lingvistiky (viz např. L. Janda, 1993, či L. Janda & S. Clancy, 2002), nebyl by ale cizí pojmům Pražské školy, jako např. „centrum a periferie“.

Jako příklad uveďme znovu slovo *rybník*. Adekvátní definice rybníka představuje souhrn jevů nutných, typických, atypických a vyloučených. Rybník, který obsahuje jen typické rysy, je tzv. prototypický. Pro ostatní rybníky se typičnost měří jejich blízkostí či vzdáleností k těmto typickým parametrům – lokalitě, významu, syntaxi. Kontext zvýrazňuje nebo ztlumuje různé aspekty těchto jevů, a proto můžeme používat kontextuální rysy, abychom zjistili míru typičnosti (viz obrázek 1).

Odpověď na první otázku tedy zní, že se korpusová data v hlavních rysech shodují s daty z dotazníkového výzkumu, tj. že se estetické reakce odrážejí v psaném světě našich mluvčích.

Druhá otázka se týkala souvislosti mezi frekvencí v korpusu a kognitivním vysvětlením těchto jevů. Kognitivní lingvisté často určují typičnost podle svých vlastních intuitivních představ. V korpusové lingvistice by typičnost měla spíše vyplývat z frekvence různých jevů. Ve většině případů z ČNK jdou sémantická a frekvenční typičnost ruku v ruce. Jediné slovo, u kterého se rozešly, je *kabát*, a v tomto případě distribuce koncovek sledovala spíše frekvenční typičnost než typičnost sémantickou. Korpusová data totiž naznačují, že frekvence by mohla být uvedena jako čtvrtý parametr v typičnosti, vedle tří parametrů uvedených v diagramu, a právě tuto frekvenci můžeme nejlépe zjistit v korpusu – nebo v korpusu.

LITERATURA

BERMEL, N. (1993): Sémantické rozdíly ve tvarech českého lokálu. *NŘ*, 76, s. 192-198.

- CUMMINS, G. (1995): Locative in Czech: -U or -ě? Choosing locative singular endings in Czech nouns. *Slavic and East European Journal*, 39, s. 241-260.
- HAVRÁNEK, B. & A. JEDLIČKA (1981): *Česká mluvnice*. Praha: SPN.
- JANDA, L. (1993): *A geography of case semantics*. Berlín: Mouton de Gruyter.
- JANDA, L. & S. CLANCY (2002): *The case book for Czech*. www.seelrc.org
- KRÁLÍK, J. (2001): Vyvážení zdrojů Synchronního korpusu češtiny SYN2000. *SaS*, 62, s. 38-53.
- KOŘENSKÝ, J. (1972): Výskyt variantních tvarů podstatných jmen v češtině. *NŘ*, 55, s. 17-18.
- KLIMEŠ, L. (1953): Lokál singuláru a plurálu vzoru „hrad“ a „město“. *NŘ*, 36, s. 212-219.
- Mluvnice češtiny II* (1986) = Komárek, M. & J. Kořenský ad. (eds.): *Mluvnice češtiny. Tvarosloví*. Praha: Academia.
- Příruční mluvnice češtiny* (1995) = Karlík, P. & M. Nekula ad. (eds.): *Příruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.
- ŠTÍCHA, F. (2001): Kritéria dramatickosti (korpus jako argument a inspirace). *SaS*, 62, s. 161-165.
- ŠULC, M. (2001): Tematická reprezentativnost korpusů. *SaS*, 62, s. 53-61.
- TRÁVNÍČEK, F. (1951): *Mluvnice spisovné češtiny*. Praha: Slovanské nakladatelství.

V KORPUSE OR V KORPUSU? WHAT THE CZECH NATIONAL CORPUS CAN (AND CAN'T) TELL US ABOUT MORPHOLOGICAL VARIATION IN THE LOCATIVE SINGULAR

This paper investigates a particular example of morphological variation where the competing forms do not function as well-defined semantic units or as stylistic or dialect markers. Instead, they are usually construed as representing “free variation” for a large class of words. Following on an earlier study that mapped aesthetic reactions to these variant forms, this paper attempts a quantitative look at the scale of the problem and its instantiation in a small set of words. Its findings, which are based on data from the representatively-constructed CNC, mirror the aesthetic judgments made by users in the earlier study, and suggest that a cognitive model of this variation that depends on ideas of “typicality” and “prototypicality” should be expanded to include the frequency-based “typicality” demonstrated in the Corpus data.