



PROMĚNA JAZYKA

A JEHO VÝZKUMU
V DOBĚ NOVÝCH
MÉDIÍ A TECHNOLOGIÍ

SBORNÍK

z Konference studentů českého jazyka 2013

Brno 4. 10. 2013

**PROMĚNA JAZYKA A JEHO VÝZKUMU
V DOBĚ NOVÝCH MÉDIÍ A TECHNOLOGIÍ**
Sborník z Konference studentů českého jazyka 2013
(Brno 4. 10. 2013)

Editorky: Zuzana Děngeová, Pavlína Vališová

Textová úprava: Zuzana Děngeová

Sazba textu, grafické řešení: Olga Pluháčková

Návrh grafického řešení: Jan Šplíchal

Organizační tým konference:

Zuzana Děngeová, Jan Šplíchal, Pavlína Vališová

Odborní garanti konference:

doc. PhDr. Klára Osolobě, Dr., doc. Mgr. Pavel Kosek, Ph.D.

Tento sborník vznikl v rámci projektu Podpora projektů specifického výzkumu zaměřených na organizaci studentských vědeckých konferencí MUNI/B/0796/2012.

Za obsahovou i jazykovou správnost textů odpovídají autoři jednotlivých statí.

Sborník vychází v elektronické podobě ve formátu pdf.

Přístupný z URL <http://www.phil.muni.cz/wucj/> a <http://www.ujc.cas.cz/>.

Neoprávněné užití tohoto díla je porušením autorských práv a může zakládat občanskoprávní, správněprávní, popř. trestněprávní odpovědnost.

Vydal:

Ústav pro jazyk český, Praha

Vydáno v květnu, 2015

ISBN 978-80-86496-87-0

OBSAH

Slovo úvodem	2
Využití corpus driven metod při corpus based výzkumu <i>Přegenerování a podgenerování: Skylla a Charybda automatické morfologické analýzy</i> KLÁRA OSOLSOBĚ	3
Korpus učebnic češtiny pro cizince – jeho tvorba a možnosti využití PAVLÍNA VALIŠOVÁ	13
Tvorba a využití malých specializovaných korpusů na příkladu <i>Korpusu odeslané a přijaté česky psané korespondence Bedřicha Smetany</i> LUCIE RYCHNOVSKÁ	21
Tvorba korpusu novinových titulků a jeho analýza PAVLÍNA SEDLÁŘOVÁ	26
Korpus a jmenný rod ve výuce češtiny jako cizího jazyka BARBORA KOČAŘOVÁ	32
<i>K obtížnému shánění</i> (korpusových dat pro fonetický výzkum) VERONIKA ŠTĚPÁNOVÁ	36
Akademický slovník současné češtiny a softwarový nástroj pro jeho tvorbu MICHAELA LIŠKOVÁ	42
Označování pohlaví u nejnovějších anglicizmů v češtině a ruštině NADEŽDA FRANKE, B.A.	49
Kompozita s prefixoidem <i>euro-</i> v českých a ruských publicistických textech: kontrastivní pohled ZBYNĚK MICHÁLEK	56
Blending v rámci neologické excerpční databáze RENÁTA NEPRAŠOVÁ	60
Internet a nová slovesná pojmenování ZUZANA DĚNGEOVÁ	67
Blog jako žánr potvrzující dynamický vývoj komunikačních možností současnosti JARMIL VEPŘEK	72
Jazyk seznamovacích inzerátů na internetu TEREZA MARKOVÁ	78
Vlastní jména vlakových spojů (porejonyma) na území ČR v letech 1993–2013 MICHAELA FICNAROVÁ	81
Analýza sekularizovaných slovních spojení mezi lexikonem, gramatikou a diskurzem: případ <i>Noemovy archy</i> a <i>Ducha svatého</i> JOHANNES HEINKE	93

SLOVO ÚVODEM



Sborník přináší soubor příspěvků z *Konference studentů českého jazyka 2013*, která se uskutečnila 4. října 2013 na Filozofické fakultě Masarykovy univerzity. Byla realizována v rámci projektu *Konference studentů českého jazyka 2013* (MUNI/B/0796/2012) podpořeného Grantovou agenturou MU: Podpora projektů specifického výzkumu zaměřených na organizaci studentských vědeckých konferencí. Zvolené téma konference *Proměna jazyka v době nových médií a technologií* poukazuje na aktuálně řešené lingvistické otázky. Uveřejněné příspěvky jednak ukazují široké možnosti využití nových technologií ve výzkumu jazyka (využití jazykových korpusů, tvorba lexikálních databází apod.), jednak představují výzkum jazyka nových médií (jazyk na Facebooku, jazyk blogu apod.).

Úvodní statí sborníku je plenární přednáška *Využití corpus driven metod při corpus based výzkumu, Přegenerování a podgenerování: Skylla a Charybda automatické morfologické analýzy*, ve které doc. PhDr. Klára Osolsobě, Dr., zmiňuje a analyzuje některé problémy automatické morfologické analýzy v Českém národním korpusu. Následuje první část sborníku na téma *Tvorba a využití malých specializovaných korpusů, korpusy a výuka jazyků*. Zde naleznete čtyři příspěvky týkající se současného trendu budování malých specializovaných jazykových korpusů. Tyto texty pojednávají o korpusu učebnic češtiny jako cizího jazyka, korpusu česky psané korespondence Bedřicha Smetany, korpusu novinových titulků a o mluveném korpusu, který slouží fonetickému výzkumu. Poslední, pátý příspěvek se věnuje konkrétnímu využití Českého národního korpusu ve výuce češtiny pro cizince. Ve druhé části na téma *Tvorba a využití lexikálních databází, analýzy specifík jazyka nových médií* je věnována pozornost tvorbě lexikální databáze pro výkladový slovník češtiny a tvorbě neologické databáze. Další příspěvky se zabývají novými pojmenováními a také komparativistikou. Jejich témata jsou např. nová slovesná pojmenování na internetu, přechylování v češtině a ruštině a prefix euro- v češtině a ruštině. Sborník uzavírá oddíl s tématem *Proměna jazyka vlivem elektronické komunikace a internetu*, jenž obsahuje příspěvky, které se věnují jazyku blogů, seznamovacím inzerátům na internetu, sekularizovaným slovním spojením „Duch svatý“ a „Noemova archa“ a pojmenováním vlakových spojů.

Rádi bychom poděkovali všem účastníkům konference, a to zejména přednášejícím, za účast a podnětné příspěvky. Děkujeme také odborným garantům projektu doc. PhDr. Kláře Osolsobě, Dr., a doc. Mgr. Pavlu Koskovi, Ph.D., a vedení Ústavu českého jazyka za pomoc s organizací konference. Vzhledem k obohacujícím a podnětným referátům, hojně účasti i příjemné atmosféře konference doufáme, že toto setkání mladých jazykovědců nebylo na půdě Ústavu českého jazyka poslední.

Organizátoři konference

VYUŽITÍ CORPUS DRIVEN METOD PŘI CORPUS BASED VÝZKUMU¹

PŘEGENEROVÁVÁNÍ A PODGENEROVÁVÁNÍ: SKYLLA A CHARYBDA

AUTOMATICKÉ MORFOLOGICKÉ ANALÝZY

KLÁRA OSOLSOBĚ

Ústav českého jazyka FF MU Brno

Abstract

Overgeneration is a property of formal rules which does not cover the exact language data it was designed for. It is equivalent to low precision and occurs when a formal rule (corpus query) is too widely defined. Undergeneration is equivalent to low recall and occurs when a formal rule (corpus query) is too narrowly specified. Both are caused by the ambiguity of natural language. In this article we shall demonstrate how to use corpus driven method in optimization of retrieval technique for corpus based analysis. On a specific example of retrieval of candidates for a word formation model (kutil) we shall show how to use observation of corpus data for progressive specification of corpus query.

Klíčová slova: korpus, corpus based, corpus driven, přegenerování, podgenerování, lemma, tag, slovtvorba

Keywords: corpus, corpus based, corpus driven, overgeneration, undegeneration, lemma, tag, word formation

Úvod

Jedním z podstatných rysů aplikací automatické analýzy přirozeného jazyka je tzv. přegenerování. Formální definici odpovídají jednotky, které tvoří homogenní skupinu (tu, kterou se prostřednictvím formálního zadání snažíme definovat), ale i jednotky, které jsou vůči této skupině heterogenní. Tento jev spadá na vrub obecné vlastnosti přirozeného jazyka, jíž je nejednoznačnost (homonymie) na všech úrovních. Rubem téže mince je tzv. podgenerování, tedy případ, kdy formální zadání je vymezeno příliš úzce, takže nejsou zachyceny jednotky, které se jeho prostřednictvím snažíme definovat.

Na konkrétním příkladu ukážeme postup optimalizace vyhledávání dat pro korpusově založený (corpus based) výzkum slovtvorby, který vychází z korpusově řízené (corpus driven) metody přesňování formálního zadání na základě pozorování přegenerovaných dat.

Korpusově řízená metoda optimalizace postupů vyhledávání dat v jazykových korpusech²

Při vyhledávání jazykových dat používáme při práci s jazykovými korpusemi počítačové programy (korpusové manažery), které umožňují klást dotazy, jimiž se snažíme co možná nejpřesněji definovat vlastnosti (formální) jednotek jazyka, které chceme v korpusu vyhledat.

V tomto příspěvku se snažíme ukázat, jak lze postupným upřesněním formálního zadání dojít k lepším datům, která jsou podkladem pro lingvistický výzkum.

1 Příspěvek vznikl v rámci grantového projektu specifického výzkumu MUNI/A/1165/2014 Čeština v jednotě synchronie a diachronie 2015.

2 Praktické ukázky jsou vybrány ze synchronního korpusu psané češtiny (SYN2010) – <http://ucnk.ff.cuni.cz/>. Při práci s korpusem v ukázkách (obrázky) používáme verzi korpusového manažeru přístupného z webového rozhraní NoSketch Engine na adrese <http://korpus.cz/corpora/>.

Chceme-li minimalizovat přegenerování, pak je třeba formální definici pokud možno zúžit. Naopak chceme-li minimalizovat podgenerování, musíme mít na zřeteli, že příliš úzké zadání může mít za následek ztrátu dat.

Ukažme si, jak funguje pře- a pod-generování na praktickém a celkem jednoduchém příkladu deverbativ tvořených příponou *-tel*. Formální jednoduchá definice zní, že kandidáty lze najít, vyhledáme-li všechna maskulina životná zakončená řetězcem *.*tel*. Při použití takové definice budou případy přegenerování slova jako *datel* nebo proprium *Obrtel*. Přegenerování se nevyhne ani tehdy, zúžíme-li definici na případy, kdy před *tel* předchází *[ai]*. Zůstane přegenerovaný *datel*, a navíc podgenerovaný *přítel* a *neumětel*.

Nyní si ukážeme složitější případ. Budeme používat formální zápis v jazyce **cql** (corpus query language).³

Postup optimalizace vyhledávání substantiv typu *kutil*

Naším cílem je vyhledat v korpusu kandidáty na **jména osob odvozená od l-ových příčestí (konvertovaná)**. Máme-li označovaný korpus, pak se můžeme při zadání dotazu opřít a) o morfológickou značku a b) o lemma. Chceme tedy vyhledat maskulina životná, jejichž základní tvar (lemma) končí na *l*.

Dotaz v CQL bude vypadat následovně: **[lemma="*l" & tag="NNM.*"]** a po jeho zadání do dotazovacího řádku korpusového manažeru získáme konkordanci všech výskytů substantiv, maskulin životných, jejichž základní tvar končí na *l* (obrázek 1).

Obrázek 1

Příběhy Sherlocka Holmese	by asi udivily i jeho	učitele /učitel/NNMP4-----A-----	. " " Neptal ses
Příběhy Sherlocka Holmese	se mohl s tím tvým	přítelem /přítel/NNMS7-----A-----	seznámit ? " " Určitě
Příběhy Sherlocka Holmese	Je schopen podat svému nejlepšímu	příteli /přítel/NNMS3-----A-----	nejnovější rostlinný alkaloid , ne
Příběhy Sherlocka Holmese	přišoupl nohou . " Můj	přítel /přítel/NNMS1-----A-----	hledá byt , a poněvadž
Příběhy Sherlocka Holmese	vzbudit zájem i toho nej povrchnějšího	pozorovatele /pozorovatel/NNMS4-----A-----	. Byl šest stop vysoký
Příběhy Sherlocka Holmese	krásného počasí , neměl jsem	přátel /přítel/NNMP2-----A-----	, kteří by svými návštěvami
Příběhy Sherlocka Holmese	spolubydlící nemá tak jako já	přátel /přítel/NNMP2-----A-----	. Potom jsem však zjistil
Příběhy Sherlocka Holmese	aby si z nich schopný	pozorovatel /pozorovatel/NNMS1-----A-----	neudělal přesný obraz . "
Příběhy Sherlocka Holmese	mlčení , co si jeho	přítel /přítel/NNMS1-----A-----	myslí , to sice vypadá
Příběhy Sherlocka Holmese	. " Lecoq byl ubohoučký	packal /packal/NNMS1-----A-----	, " řekl zlostně .
Příběhy Sherlocka Holmese	pokoje , a podával mému	příteli /přítel/NNMS3-----A-----	dopis . To byla příležitost
Příběhy Sherlocka Holmese	Mohu se vás zeptat ,	příteli /přítel/NNMS3-----A-----	, " řekl jsem libezně
Příběhy Sherlocka Holmese	Yardu , " poznamenal můj	přítel /přítel/NNMS1-----A-----	, " on a Lestrade
Příběhy Sherlocka Holmese	on a Lestrade jsou jednookými	králi /král/NNMP7-----A-----	mezi slepými . Oba jsou
Příběhy Sherlocka Holmese	. " " Můj drahý	příteli /přítel/NNMS3-----A-----	, co z toho mám
Příběhy Sherlocka Holmese	tohohle ! " odpověděl můj	přítel /přítel/NNMS1-----A-----	a ukázal na pěšinu .
Příběhy Sherlocka Holmese	na pěšinu . " Stádo	buvolů /buvol/NNMP2-----A-----	by to tu nemohlo víc
Příběhy Sherlocka Holmese	detektiv . Tvářil se jako	majitel /majitel/NNMS1-----A-----	panoptika , který se vychloubá
Příběhy Sherlocka Holmese	vám , " odpověděl můj	přítel /přítel/NNMS1-----A-----	. " Počínáte si tak
Příběhy Sherlocka Holmese	zahradní pěšině šli jako dva	přátelé /přítel/NNMP1-----A-----	- pravděpodobně byli do sebe

Je patrné, že vidíme téměř samé přegenerované doklady (výjimkou je substantivum *packal* na 10. řádku konkordančního seznamu na obrázku 1). Zobrazíme-li si frekvenční distribuci lemmat

3 Bližší informace k jazyku CQL i k použitému morfológickému značkování lze nalézt zde: <http://ucnk.ff.cuni.cz/bonito/regular.php> a <http://ucnk.ff.cuni.cz/bonito/znacky.php>.

Shrňme tedy, že s ohledem na skutečnost, že není možné jednoduše pomocí morfologické značky vyhledávat podle vzoru, dochází při zvoleném zadání dotazu k masivnímu přegenerování. Dále je vidět, že přegenerováá i disambiguace.

Je tedy žádoucí hledat jiný (lepší) postup. Optimalizace může vycházet z pozorování přegenerovaných dat (corpus driven) a následné formulace pravidel transformovatelných do optimalizovaných formálních definic pro hledání kandidátů reprezentujících příslušný slovtvorný typ.

Pozorujeme, že většina přegenerovaných případů jsou deverbativa tvořená sufixem *-tel*. Položíme si tedy otázku, zda mohou l-ová přičestí v češtině končit např. na řetězec *tel* (ale i např. *rel*, *vel*, *ael*, neboť přegenerovaná jsou i lemmata vlastních jmen jak *Karel*, *Pavel*, *Michael*). Odpověď lze následovně zahrnout do dotazu v cql jako formální podmínku pro vyhledávání.

Popis postupu pro získání odpovědi na základě pozorování korpusových dat

Na otázku, mohou-li l-ová přičestí v češtině končit např. na řetězce *tel*, *rel*, *vel*, *ael*, lze odpovědět takto: a) mohou a v korpusu jsou doložena, b) mohou, ale v korpusu nejsou doložena, c) nemohou, a tudíž v korpusu nejsou doložena.

Hledáme ospravedlnění kladné odpovědi na otázku a). Slovní formulace dotazu pro získání dat z korpusů bude, že hledáme l-ová přičestí taková, že před řetězcem *l*, popřípadě *la*, *lo*, *li*, *ly* (*l*+rodová koncovka) předchází řetězec *tel*, *rel*, *vel*, *ael*.

V cql bude dotaz vypadat takto: `[word="((.*[atrv]el)|(.*[atrv]el[aoiy]))" & tag="V[pq].*"]` a zadáme-li jej ve formě dotazu korpusovému manažeru, bude výsledek vyhledávání nulový (prázdný seznam konkordančních řádků).

Je-li výsledkem prázdný seznam, mělo by platit buď b), nebo c). Prohledáváním ještě větších korpusů by bylo možné pokračovat a snažit se dokázat, že platí za b). Jde o neefektivní postup s malou pravděpodobností úspěchu, neboť ze zkušenosti práce s korpusy je známo, že výjimky u frekventovaných tvarů bývají rovněž frekventované a l-ové přičestí je frekventovaný tvar. Můžeme ovšem hledat nějaké zobecnitelné formální vlastnosti l-ových přičestí a pokusit se tak dokázat, že platí c). Viděli jsme, že relevantní výsledek jsme docílili, když jsme se zabývali otázkou, co předchází před tvarovou koncovkou *-l* v českých l-ových přičestích. Odpovědi mohou být např. tyto: A) libovolná samohláska/souhláska (grafém), B) pouze některé samohlásky/souhlásky (grafémy).

Jednoduchou empirickou evidenci získáme z korpusů, a to tak, že se podíváme, jaká je situace u l-ových přičestí (obrázek 4).

Vidíme, že před tvarovou koncovkou *-l* může předcházet např. dlouhé *á*, krátké *[aeěiyu]*, že se vůbec nevyskytují dlouhé samohlásky (s výjimkou *á*), z krátkých se nevyskytuje *o* (kromě slovenského *bol*). Pokud bychom chtěli zjistit, které souhlásky se vyskytují v českých l-ových přičestích před tvarovou koncovkou *-l*, pak by dotaz v cql vypadal následovně: `[tag="Vp.*" & word!="((.*[aáeěiyu]l)|(.*[aáeěiyu]l[aoiy]))"]`.⁵

5 Projdeme-li seznam tvarů, pak zjistíme, že omezení se bude týkat grafémů *[dʒlnŋqřřtw]*. Ve sledovaném korpusu nenajdeme ani doklady *[gx]*, nicméně l-ová přičestí sloves *grgnout* a *exnout* lze najít např. na internetu.

Obrázek 4

Celkem: 16294 (326 str.)

	lemma	Frekvence
1.	p/n být	803 337
2.	p/n mít	290 303
3.	p/n moci	155 321
4.	p/n říci	154 841
5.	p/n stát	87 888
6.	p/n chtít	74 375
7.	p/n začít	71 930
8.	p/n muset	62 098
9.	p/n dostat	53 171
10.	p/n dát	51 542
11.	p/n přijít	50 572
12.	p/n jít	46 564
13.	p/n vidět	46 101
14.	p/n vědět	42 506
15.	p/n uvést	32 603

Nyní si „pozveme na pomoc“ systematický popis morfolgie českých sloves. Kmenový vokál (KmV) pro tvary od kmene minulého konkrétně l-ových přičestí, mohou být pouze krátké [aieě] a (n)u, nikdy o. Dlouhé á mají pouze nepravidelná slovesa jako *bál* (*se*), *stál* a z pravidelných jen *zdál se* a (*u*)*dál se*. Kromě toho ovšem taky případy, kdy á je kořenový vokál (KoV), tedy sloveso patří ke vzoru *krýt* III. tř. – *hrál, sál, vál, smál se* atd. Všechny ostatní KoV ve III. tř. u vzoru *krýt* jsou krátké (zkrácené oproti infinitivu), vyskytuje se pouze [eěiu] (*klel, děl, žil, zul, myl*) a obdobná omezení nacházíme i u nepravidelných (atematických) sloves *byl, měl, chtěl* atd. Souhlásku mají slovesa I. třídy vzorů *nést, péct* a mohou ji mít (nemají-li kmenotvorné -nu-) slovesa II. třídy vzoru *tisknout* (*tiskl, nadchl se, zestárl*) a některá vzoru *začít* (*zapl, vypl, i moravské rožl*). S ohledem na masivní přegenerování substantiv na *tel* nás zajímá především KoV a KmV [eě].

KmV [eě] mohou mít slovesa I. třídy podle kmene přítomného vzoru *umřít* (*umřel, vytřel, pomlel*), IV. třídy podle kmene přítomného vzorů *trpět, sázet* (*vrtěl, hleděl, probděl*).

KoV [eě] mohou mít slovesa III. třídy podle kmene přítomného vzoru *kryl* (*děl, pěl, klel, zasel*).

Bližší analýza možného okolí e/ě v roli KmV a KoV

Vyhledáme v korpusu všechna l-ová přičestí končící na *el*. V cql bude dotaz vypadat takto: [word="((.*el)|(.*el[aoiy]))" & tag="V[pq].*"]. Část seznamu lemmat podle frekvence vidíme na obrázku 5.

Obrázek 5

Celkem: 5958 (120 str.)

	lemma	word	Frekvence
1.	p/n muset	musel	19 986
2.	p/n přijít	přšel	17 754
3.	p/n jít	šel	11 055
4.	p/n muset	museli	10 507
5.	p/n muset	musela	10 095
6.	p/n najít	našel	8 065
7.	p/n myslet	myslel	7 036
8.	p/n zemřít	zemřel	6 589
9.	p/n odejít	odešel	6 256
10.	p/n otevřít	otevřel	6 217
11.	p/n slyšet	styšel	6 093
12.	p/n vyjít	vyšel	5 257
13.	p/n držet	držel	5 084
14.	p/n pomyslet	pomyslel	4 500
15.	p/n přijet	přjel	4 400
16.	p/n zmizet	zmizel	4 030
17.	p/n otevřít	otevřela	3 851
18.	p/n muset	muselo	3 824
19.	p/n myslet	myslela	3 742
20.	p/n jet	jel	3 737

Na základě pozorování lemmat můžeme formulovat pravidla distribuce *e/ě* v roli KmV a KoV. Před KmV *-e-* stojí v češtině pouze [cčjlrššžž], před KoV *-e-* stojí v češtině pouze [lsz]. Stojí-li před hláskou *e* grafémy [p̥bvt̥dn̥m]⁶, pak se vždy graficky realizuje jako *ě*.

Na základě empirie můžeme tedy tvrdit, že aby mohlo být substantivum konvertovaným l-ovým přičestím, pak musí splňovat tyto podmínky: před *l* může ze samohlásek předcházet pouze [aáeěiuy]; před *l* nemůže předcházet [bpvmdtnkghr]e; před *l* nemůže předcházet [aáeěíioóuúyý] [aáeěiuy].

Formulace dotazu s využitím formálních vlastností pozorovaných dat

Do dotazu v jazyce cql zahrneme podmínky pro vyhledávání substantiv maskulin životných, která mají mít výše uvedené formální vlastnosti (lemma končí na *l* a zároveň před *l* předchází pouze některé přesněji definované kombinace grafémů). S ohledem na fakt, že v korpusech se poměrně často vyskytují chyby v desambiguaci homonymních tvarů l-ové přičestí / tvar konvertovaného jména, omezíme vyhledávání jen na nehomonymní tvary. Přehled homonymních tvarů uvedeme v tabulce 1 (tučně a kurzívou). Jsme si vědomi také toho, že pomineme deriváty od sloves, u kterých se tvarová koncovka *-l* připíná bezprostředně ke kořenové souhláskové finále (slovesa I. třídy slovesné podle vzorů *nést*, *péci* a některá slovesa II. třídy slovesné podle vzorů *tisknout* a *začít*). Od těchto sloves se totiž jména osob sice okrajově tvořit mohou, tvoří se ovšem konverzí substandardního tvaru l-ového přičestí bez tvarové koncovky *-l*, jak dosvědčují např. apelativa *vyklouz*, *zběh* a snad i kompozita *břichopas*, *mrakotřas* nebo propria *Proklouz*, *Skoněšpad*, *Vozembouch* atd. (více Osolsobě 2011: 61n).

V cql bude dotaz vypadat takto: [word="*[aáeěiuy]l((ové)|(ů)|(ům)|(ech)|(e)|(ovi))" & word!="*[bpvmdtnkghr]el((ové)|(ů)|(ům)|(ech)|(e)|(ovi))" & word!="*[aáeěíioóuúyý] [aáeěiuy]l((ové)|(ů)|(ům)|(ech)|(e)|(ovi))" & tag="NNM.*"]

Tabulka 1

N. A.	<i>kutil</i>	kutilové
G.	<i>kutila</i>	kutilů
D.	kutilu/kutilovi	kutilům
V.	kutile	kutilové
L.	kutilu/kutilovi	kutilech
I.	kutilem	<i>kutily</i>

Na obrázku 6 vidíme, že se již na první stránce konkordancí objevuje alespoň jeden relevantní doklad v podobě tvaru *hýřilové*.

6 V seznamu neuvádíme grafém *f*, protože na základě analýzy dat z korpusu SYN se nám nepodařilo najít žádný doklad l-ového přičestí slovesa s l-ovým přičestím na **fěl* a v korpusu czTenTen12 jsme našli pouze překlipy. (Dotaz v cql by vypadal takto: [word="((.*fěl)|(.*fěl[aoiy]))".])

Obrázek 6

Příběhy Sherlocka Holmese	že vojáci pořad prohrávají a	civilové /civil/NNMP1-----A-----	vyhrávají . Rozumíte , nechci
Příběhy Sherlocka Holmese	vznešeného domu Ormsteinů , dědičných	králů /král/NNMP2-----A-----	české země . " "
Příběhy Sherlocka Holmese	velkovévodu z Cassel-Falsteinu a dědičného	krále /král/NNMS4-----A-----	české země . " "
Příběhy Sherlocka Holmese	Saxe-Meningenu , druhou dceru skandinávského	krále /král/NNMS2-----A-----	. Snad je vám známo
Příběhy Sherlocka Holmese	, ' řekla paní Irena	manželovi /manžel/NNMS3-----A-----	a odjela . Nic víc
Příběhy Sherlocka Holmese	. " Jde o rozvod	manželů /manžel/NNMP2-----A-----	Dundasových a zcela náhodou jsem
Příběhy Sherlocka Holmese	to malý dáreček od českého	krále /král/NNMS2-----A-----	jako pozornost za pomoc ,
Příběhy Sherlocka Holmese	a to v případě českého	krále /král/NNMS2-----A-----	a fotografie Ireny Adierové ,
Příběhy Sherlocka Holmese	a křik nějaké opožděné skupinky	hýřilů /hýřil/NNMP2-----A-----	. Po obloze pomalu plynul
Příběhy Sherlocka Holmese	, že je hluboce oddána	manželovi /manžel/NNMS3-----A-----	i synkovi . Ustavičně bloudí
Dr. No	u chodníku , patřila brigádnímu	generálovi /generál/NNMS3-----A-----	karibské oblasti , kingstonskému generálnímu

Při zobrazení frekvenční distribuce lemmat zjistíme, že stále je dost případů, které jsme nehledali (přegenerovávání), nicméně máme už i správné výsledky jako *čumil* a *kutil* (obrázek 7).

Obrázek 7

25. p/n	čumil	98	■
26. p/n	konzul	95	■
27. p/n	federál	93	■
28. p/n	Michal	83	■
29. p/n	kanibal	78	■
30. p/n	Dale	75	■
31. p/n	debil	72	■
32. p/n	kutil	68	■

Ruční analýzou dat (631 lemmat) získáme jak apelativa, tak především propria (mezi nimi je hledaný slovtvorný model bohatě zastoupen). Nacházíme lemmata jako *kutil*, *čumil*, *Nezval*, *Hrabal*, *Pospíšil*, *čmuchač*, *ožrala*, *břídil*, (*Doležal*), *Musil*, *Prášil*, *patolízal*, *štouřal* a další (řazeno podle frekvence).

Nicméně propria jako parasystém ponecháme stranou a zaměříme se pouze na apelativa, a to tak že odstraníme lemmata, která začínají velkým písmenem.⁷ Seznam lemmat se dále redukuje na 173 (obrázek 8).

Ručním tříděním získáme ze 173 lemmat 40 lemmat, která jsou hledanými substantivy konvertovanými z l-ových příčestí. V korpusu SYN2010 lze uvedeným postupem najít tato lemmata: *kutil*, *čumil*, *čmuchač*, *ožrala*, *břídil*, *patolízal*, *štouřal*, *žvanil*, *slídil*, *střádal*, *packal*, *hejkal*, *reptal*, *tlachal*, *rýpal*, *rejpal*, *hýřil*, *chlubil*, *všeuměl*, *kýval*, *kecal*, *mazal*, *škrabal*, *všudybyl*, *skuhral*, *loudil*, *fňukal*, *brblal*, *šukal*, *škudlil*, *šeptal*, *čmáral*, *zlobil*, *koktal*, *mlsal*, *klábosil*, *hrabal*, *hloubal*, *cmrndač*, *cintal* (seřazeno podle frekvence).

Můžeme tedy konstatovat, že ačkoliv více než ¼ (77 %) dokladů jsou případy, které spadají pod pojem přegenerovávání (174-40=134), tak pokud bychom byli zůstali u prvního pokusu, pak by byla míra přegenerovávání vyšší než 98 % (40 lemmat bychom museli ručně vybrat ze seznamu 3023 lemmat). Míru přegenerování jsme tudíž snížili o 21 %.

⁷ Propria lze odstranit pomocí volby negativního filtru vybrat a odstranit lemmata začínající na velké písmeno, tedy [lemma=" [AÁBĀCĀĎĎFGHĪJKLNMNŇŌÓPQRŔSŠTŤUÚVWXYŸZŽ].*"].

Obrázek 8

Celkem: 173 (4 str.)

	lemma	Frekvence	
1.	<u>p/n</u> král	5 799	
2.	<u>p/n</u> manžel	3 291	
3.	<u>p/n</u> profesionál	888	
4.	<u>p/n</u> generál	837	
5.	<u>p/n</u> anděl	551	
6.	<u>p/n</u> radikál	517	
7.	<u>p/n</u> liberál	517	
8.	<u>p/n</u> vandal	481	
9.	<u>p/n</u> rival	426	
10.	<u>p/n</u> kardinál	297	
11.	<u>p/n</u> novomanžel	186	
12.	<u>p/n</u> maršál	116	
13.	<u>p/n</u> mobile	115	
14.	<u>p/n</u> admirál	107	
15.	<u>p/n</u> horal	106	
16.	<u>p/n</u> pedofil	101	
17.	<u>p/n</u> čumil	98	
18.	<u>p/n</u> konzul	95	
19.	<u>p/n</u> federál	93	
20.	<u>p/n</u> kanibal	78	
21.	<u>p/n</u> debil	72	
22.	<u>p/n</u> kutil	68	

Problém podgenerování

Otázkou zůstane, jak velké množství lemmat tímto postupem zachyceno nebylo. Pomineme případy, kdy pádová substantivní koncovka je homonymní s rodovou koncovkou l-ového přičestí (tvarová homonymie *kutil*, *kutila*, *kutily*), které většinou vykazují velké množství chyb v disambiguaci, takže jsme je ze své analýzy záměrně vyloučili, a zaměříme se na případy nezaznamenané ve slovníku morfologického analyzátoru, tedy „odpadkový koš“ (tag=„X.*“).

Zopakujeme dotaz, ale místo maskulin životných budeme hledat slova s označením slovního druhu „X“:

V cql bude dotaz vypadat takto: **[word="*[aáeěiuy]l((ové)|(ů)|(ům)|(ech)|(e)|(ovi))" & word!="*[bpfvmdtnkghr]el((ové)|(ů)|(ům)|(ech)|(e)|(ovi))" & word!="*[aáeěííoóuůyý] [aáeěiuy]l((ové)|(ů)|(ům)|(ech)|(e)|(ovi))" & tag="X.*"]**

Po odstranění proprií⁸ získáme ručním tříděním dalších 10 lemmat z celkového počtu 731 výskytů (*plazil*, *velebil*, *remcal*, *plížil*, *patlal*, *muchlal*, *kousal*, *drnkal*, *cachtal*, *brouzdal*).

Nejde jen o hapaxy (obrázek 9), ačkoliv je třeba připomenout, že doklady pocházejí z jediného dokumentu.

8 Viz předcházející poznámka.

Obrázek 9

Výskytů: 18 i.p.m.: 0,15 (vztaženo k celému korpusu) | ARF: 1

Hvězdotřesení	. Okamžitě ho obklopili tři plížílové /plížílové/X@..... , půl tuctu plížovníků a	
Hvězdotřesení	se neboj . Ti hluční plížílové /plížílové/X@..... už jsou pryč . "	
Hvězdotřesení	všechn ten rámus . Ti plížílové /plížílové/X@..... dokážou rozvíbrovat celý blok .	
Hvězdotřesení	patrně důvodem , proč byli plížílové /plížílové/X@..... nejoblíbenějšími mazlíčky čilů . Skoro	
Hvězdotřesení	Tekutý písek . " Ostatní plížílové /plížílové/X@..... budou rádi , že zase	
Hvězdotřesení	. Pokud vejce nesežrali toulaví plížílové /plížílové/X@..... , mělo mládě slušnou šanci	
Hvězdotřesení	okusoval trávnik , Valík a plížílové /plížílové/X@..... nebyli nikde v dohledu .	
Hvězdotřesení	" Tihle geneticky čistí laboratorní plížílové /plížílové/X@..... vypadají všichni stejně , "	
Hvězdotřesení	kůži pokrytou puchýří . Domácí plížílové /plížílové/X@..... na tom byli stejně jako	
Hvězdotřesení	ji podobné podnikali lovy divocí plížílové /plížílové/X@..... . Někdo o ni pečoval	
Hvězdotřesení	očima stačil všimnout , že plížílové /plížílové/X@..... se vyděšeně rozprchli už dávno	
Hvězdotřesení	v horách , kde se plížílové /plížílové/X@..... chovat nesmějí . Bojím se	
Hvězdotřesení	Čilové z okolních klanů i plížílové /plížílové/X@..... zažili už tucty podobných přistání	
Hvězdotřesení	" " Tady jsou imperátorovi plížílové /plížílové/X@..... chránění před divokými mršty .	
Hvězdotřesení	ani jeden nechybí . Imperátorovi plížílové /plížílové/X@..... tu mají víc krmení .	
Hvězdotřesení	Imperátorovi náleží vše . Všichni plížílové /plížílové/X@..... , každíčký korální ořech ,	
Hvězdotřesení	, kde jsou chlupatí jako plížílové /plížílové/X@..... . " " To je	
Hvězdotřesení	jako by to byli nějakí plížílové /plížílové/X@..... v zoologické zahradě ! "	

Optimalizaci míry podgenerování spatřujeme v automatizovaném vytvoření seznamu kandidátů na hledané jednotky z tvarů nezachycených slovníkem automatického morfologického analyzátoru (tvarů, které splňují podmínky otestované na datech rozpoznávaných automatickou morfologickou analýzou a zároveň mají značku tag=„X.“). Ručním tříděním automaticky pořízených dat se podařilo získat celkem dalších 10 lemmat, která nebyla zachycena výše uvedeným postupem a spadala tak pod pojem podgenerování. Celkový seznam 40 lemmat se takto rozšířil o dalších 10 lemmat (zlepšení o 25 %). Efektivnost uvedeného vyhledávání lze porovnat také s postupy popsanými jinde (srov. Štícha 2011 : 225–226).

Závěr

Ukázali jsme, jak lze při zadání dotazu korpusovému manažeru na základě pozorování přegenerovaných dat postupně optimalizovat dotaz tak, aby se počet dokladů, jež je dále třeba podrobit ruční analýze, pokud možno minimalizoval. V uvedeném příkladu se nám podařilo snížit míru přegenerování o 21 %. Vyzkoušený postup jsme zopakovali na datech, která nejsou rozpoznána automatickou morfologickou analýzou, a spadají tudíž pod pojem podgenerování. Ruční analýzou takto získaných dat se nám podařilo rozšířit počet lemmat o 25 %.

Redukce počtu lemmat pro ruční analýzu přispěla k větší efektivitě a spolehlivosti vyhledání maximálního počtu jednotek; minimalizace problému přegenerování i podgenerování vede obecně k redukci rozsahu ruční práce, která je vždy nákladná (na čas popř. i finanční prostředky, z nichž je třeba hradit školené anotátory) a představuje také zvýšené nebezpečí chyb z nepozornosti, které jsou následně poměrně obtížně detekovatelné.

Uvedený postup je případem metody korpusové lingvistiky, kdy užíváme korpus nejen jako zdroj pro lingvistické observace (corpus based), ale též jako zdroj odhalování pravidel fungování přirozeného jazyka (corpus driven), která mohou být úspěšně využita i k jiným účelům, než jsou ty, jimž slouží v daném případě. Konkrétně si lze představit např. využití uvedených pravidel při tvorbě guesserů (hadačů), tedy programů, které se na základě různých technik snaží „uhádnout“ vlastnosti jednotek přirozeného jazyka (např. interpretace lemmatu a morfologického tagu).

Bibliografie

- DOKULIL, M. a kol. (1986): *Mluvnice češtiny 1*. Praha: Academia.
- KOMÁREK, M. a kol. (1987): *Mluvnice češtiny 2*. Praha: Academia.
- McENERY, T. – HARDIE, A. (2012): *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- OSOLSOBĚ, K. (2011): *Morfologie českého slovesa a tvoření deverbativ jako problém strojové analýzy češtiny*. Brno: Masarykova univerzita.
- OSOLSOBĚ, K. (2011): Korpus jako zdroj dat pro studium slovo tvorby. In: V. Petkevič – A. Rosen (eds.), 3. *Gramatika a značkování korpusů*. Praha: Nakladatelství Lidové noviny / Ústav Českého národního korpusu, s. 10–23.
- ŠTÍCHA, F. (ed) (2011): *Kapitoly z české gramatiky*. Praha: Academia.

Elektronické zdroje

- Korpus SYN2010: Český národní korpus – SYN2010*. Ústav Českého národního korpusu FF UK, Praha 2010. Cit. 6. 11. 2013, dostupný z WWW: <<http://www.korpus.cz>>.
- Korpus SYN: Český národní korpus – SYN*. Ústav Českého národního korpusu FF UK, Praha 2010. Cit. 6. 11. 2013, dostupný z WWW: <<http://www.korpus.cz>>.
- Korpus czTenTen12*. Cit. 6. 11. 2013, dostupný z WWW: <<https://ske.fi.muni.cz>>.

KORPUS UČEBNIC ČEŠTINY PRO CIZINCE – JEHO TVORBA A MOŽNOSTI VYUŽITÍ¹

PAVLÍNA VALIŠOVÁ

Ústav českého jazyka FF MU Brno

Abstract

The aim of this paper is to introduce a project of compilation a small specialized corpus from textbooks of Czech as a foreign language and to present its benefits for language teaching. The use of this type of corpora is divided in a) descriptive – a description of textbook language as a specific genre, and b) didactical – a research of different types of language exercises. The analysis of the data of the first part of this corpus (7 textbook on level A1) and comparison with the descriptor of A1 level of the CERF shows how textbook corpus can improve teaching materials.

Klíčová slova: čeština jako cizí jazyk, korpusová lingvistika, učebnicový korpus, výuka jazyků, Společný evropský referenční rámec pro jazyky (SERRJ)

Keywords: Czech as a foreign language, corpus linguistics, textbook corpus, language teaching, Common European Framework of Reference for Languages (CEFR)

1. Nepřímé využití korpusů

Učebnicový korpus patří k nepřímým využitím korpusu. Tím myslíme využívání korpusů lingvisty, učitelé nebo autory učebnic pro tvorbu výukových materiálů. Korpus je tedy pro ně zdrojem dat, a to jak kvantitativních, tak i kvalitativních. Frekvenční data jsou spolu tématem a budováním celkové komunikační kompetence kritérii pro rozhodování, v jaké fázi výuky jaký gramatický jev předkládat (Hrdlička, 2009, s. 31; Škodová – Štindlová, 2007, s. 57). Dále je možné z korpusu získávat typická slovní spojení, fráze či modelové věty (ať již autentické či upravené) pro tvorbu cvičení (Boulton, 2012; Frankenberg-Garcia, 2012).

Učebnicový korpus sice neslouží jako zdroj dat pro tvorbu cvičení, ale jako soubor všech slov a frází, se kterými student přišel do styku. Samozřejmě, že tento korpus je limitovaný, neboť neobsahuje obrazový materiál učebnic. Hlavně však nemůže zahrnovat interakci ve třídě, ať již učitele a studenta nebo studentů mezi sebou. Přesto však může být přínosný, pokud data z pedagogického korpusu porovnáme s daty z korpusu národního (Gouveneuová–Meunierová, 2009).

2. Analýza učebnic a učebnicový korpus

Tento korpus nahrazuje a automatizuje klasickou analýzu učebnic, při které postupujeme stránku za stránkou. Během posledních dvaceti let, kdy se obor čeština jako cizí jazyk začal rychle rozvíjet a vzniklo mnoho nových učebnic, byla publikována také řada studií opírajících se o analýzu učebnic. Při bližším zkoumání zjistíme, že největší podíl tvoří bakalářské a diplomové práce různých bohemistických pracovišť v ČR, což dokazuje, že se čeština jako cizí jazyk řadí mezi nové disciplíny (viz tabulka č. 1). Za zmínku jistě stojí dvě monografie Milana Hrdličky, které se věnují předloškám (2000) a celé mluvnici (2009). Objevují se však i studie, jež nejsou zaměřené na gramatiku, ale

1 Příspěvek vznikl v rámci grantového projektu specifického výzkumu MUNI/A/1165/2014 Čeština v jednotě synchronie a diachronie 2015.

na lexikum obsažené v učebnicích, metajazyk, tj. instrukce cvičení, nebo sociokulturní dovednosti, tedy tradice a společenské normy v ČR, jež se odráží v komunikačních situacích a taktéž i v jazyce.

Tab. 1 Přehled analýz učebnic češtiny jako cizího jazyka²

Oblast výzkumu	Autor	Téma	Počet učebnic	Úroveň	Typ práce
Gramatika	Hrdlička (2000)	předložky	38	A1–B2	monografie
	Drlíková (2008)	vid	4	A1–B2	bakalářská práce
	Hrdlička (2009)	česká mluvnice	33	A1–B2	monografie
	Lukšija (2010)	místní předložky	2	A1–A2	bakalářská práce
	Vališová (2011)	slovesné třídy	23	A1–A2	diplomová práce
	Bičišťová (2011)	číslovky	–	–	bakalářská práce
	Lukšija (2012)	deklinace substantiv	10	A1–A2	diplomová práce
	Kodajková (2012)	gramatické kategorie slovesa	–	–	diplomová práce
	Kočařová (2013)	rod	5	A1	bakalářská práce
	Koláčková (2013)	imperativ	19	A1–B1	bakalářská práce
Lexikum	Kovaříková (2009)	obecná čeština	–	–	bakalářská práce
	Nezvalová (2010)	slovotvorba	3	A1–A2	bakalářská práce
	Nováková (2010)	obecná čeština	4	B1–B2	diplomová práce
	Ocztošová (2012)	slovní zásoba pro německy mluvící	–	–	bakalářská práce
Metoda	Hrdlička (2009)	komunikační metoda	33	A1–B2	monografie
	Hladíková (2011)	signální gramatika	4	A1	diplomová práce
Metajazyk	Hradilová (2010)	terminologie jmenné flexe	3	A1–A2	odborný článek
	Osolsobě – Vališová (2010)	lingvistická terminologie	17	A1–A2	odborný článek
Text	Volnogradská (2012)	typy textů	3	A1	diplomová práce

2 Uvedené diplomové práce pochází z bohemistických pracovišť MU v Brně, UK v Praze, TUL v Liberci, ZCU v Plzni a UJEP v Ústí nad Labem. Tam, kde jsme neměli možnost nahlédnout, tzn. práce nejsou přístupné on-line, informace o počtu učebnic a jejich úrovni neuvádíme.

Podívejme se nyní na to, jak tuto učebnicovou analýzu zautomatizovat. Chtěli bychom stručně představit již existující korpusy učebnic anglického jazyka. Úplně prvním korpusem byl The TOEFL 2000 First Academic Spoken and Written Corpus z roku 2002, kolekce americké angličtiny o celkovém rozsahu 2,7 miliónů slov, z čehož necelých 800 tisíc slov pochází z učebnic. Tento korpus však sloužil i jiným účelům než pouze pro analýzu učebnic.

Druhým počinem na tomto poli je německý German English as a Foreign Language Textbook Corpus (GEFL TC) z roku 2004. Je tvořen učebnicemi angličtiny určenými pro německé mluvčí a zahrnuje 100 tisíc slov. Jeho tvůrkyně Römerová již publikovala i několik studií, ve kterých srovnává tento učebnicový jazyk s autentickým jazykem z britských národních korpusů (Römerová, 2004, 2005).

Existuje i několik dalších malých korpusů, např. angličtiny pro čínské mluvčí, nyní bychom se však chtěli zaměřit na projekt univerzity v Lovani, neboť ten se liší rozsahem i pojetím. Je jím korpus TeMa³ z roku 2007. Dříve uvedené korpusy zahrnovaly převážně texty nebo dialogy, zatímco korpus TeMa již obsahuje celé učebnice, neboť jeho cílem je výzkum všech částí učebnic. Také jeho rozsah je daleko větší, a to asi 724 tisíc slov a skládá se z 32 učebnic obecné angličtiny. Navíc díky tzv. pedagogické anotaci nabízí mnohem větší pole využití (Gouveneurová – Meunierová, 2009).

Korpus TeMa je rozdělen na 4 subkorpusy: texty, přepisy poslechů, cvičení a instrukce. U každé učebnice je vždy uveden zdroj, a to konkrétně v tomto pořadí: 1. název učebnice, 2. jestli je to učebnice nebo cvičebnice, 3. lekce, strana, cvičení a nakonec 4. typy cvičení. Pro čtvrtou část korpusu autoři vytvořili speciální značky pro pedagogickou anotaci, tzn. označkováno je každé cvičení podle typu, např. doplňování (slova z výběru, slova ve správném tvaru, libovolného slova), spojování apod. Při doplňování je samozřejmě označena i správná forma slova, která se má doplnit.

3. Využití učebnicového korpusu

Využití učebnicového korpusu dělíme na dvě hlavní oblasti: 1. Deskriptivní – popis učebnicového jazyka jako specifického žánru, a 2. Didaktická – výzkum typů cvičení.

Mezi deskriptivní způsoby využití patří:

1. Jak jsme již zmínili, výzkum slovní zásoby a komparace s autentickým územ v národních korpusech (Römerová, 2004, 2005; Gouveneurová – Meunierová, 2009).
2. Učebnicový korpus může být též využit pro studium osvojování jazyka, neboť můžeme vyhledat všechny kontexty, ve kterých se slovo či fráze vyskytují v učebnicích různých úrovní (Hustonová, 2010).
3. Pedagogické značkování však umožňuje i zkoumání učebnic z dalších perspektiv. Při rozdělení na subkorpusy podle úrovní lze zkoumat výběr slovní zásoby pro jednotlivé úrovně. Slovní zásobu specifickou pro konkrétní úroveň lze porovnat s popisem úrovní Společného evropského referenčního rámce pro jazyky (SERR)⁴.
4. Lze porovnat input v podobě textů, poslechů i cvičení, tzn. jaká slovní zásoba se vyskytuje v té které části, a zda se od sebe liší.
5. Porovnáním kolokací slov ve cvičeních lze objevit rozdíly mezi učebnicemi. (Gouveneurová, 2008 v pilotní studii dle korpusu TeMa např. objevila v tomto směru mezi učebnicemi velké rozdíly.)
6. Také lze zkoumat metajazyk, jenž učebnice využívají, a to nejen pro instrukce ke cvičením (Používají obecné výrazy nebo konkrétní? Imperativ či jiný tvar?), ale i při výkladu grama-

3 Viz <http://www.uclouvain.be/en-cecl-tema.html>.

4 Společný evropský referenční rámec pro jazyky (SERR): <http://www.msmt.cz/mezinarodni-vztahy/spolecny-evropsky-referencni-ramec-pro-jazyky>.

tiky. Zde se však často místo češtiny objevuje zprostředkovací jazyk. Rovněž důslednost ve využívání lingvistické terminologie je důležité téma pro výzkum. V učebnicích češtiny nalezneme velkou nekonzistentnost mezi českými termíny a mezinárodními výrazy, a to i v rámci jedné učebnice (Osolsobě – Vališová, 2011).

7. Další oblastí výzkumu je též typ textu, tzn. podíl psané a mluvené komunikace v učebnici a různých jazykových variant, např. zastoupení spisovné/hovorové/obecné češtiny v učebnicích či v jednotlivých částech jedné učebnice.
8. Při možnosti automatického morfologického značkování, lze zkoumat i další jevy spojené s výukou gramatiky.

Z didaktického hlediska je možno zkoumat typy cvičení podle tématu, jevu nebo referenční úrovně. (Např. Gouvernerová ve výše zmíněné pilotní studii objevila, že typy cvičení se velmi liší dle úrovně, a také poukázala na nedostatek pozorovacích aktivit, které podle výzkumu v oblasti osvojování druhého jazyka hrají velmi důležitou roli.) Nakonec je možné navrhnout zlepšení díky výsledkům současného výzkumu osvojování druhého jazyka i za pomoci současných korpusů rodilých i nerodilých mluvčích, tj. národních korpusů nebo korpusů žákovských, které se v současné době budují i pro češtinu jako cizí jazyk.

4. Český učebnicový korpus

V rámci své doktorské práce pracujeme na vybudování malého specializovaného korpusu z učebnic češtiny jako cizího jazyka (UčKo). Učebnicový korpus by měl být složen z 15 současných učebnic češtiny pro cizince (viz tabulka č. 2) – všechny byly vydány po roce 2000, některé jsou však původně staršího data, neboť se jedná o další vydání, což ukazuje, že učebnice se stále využívá pro výuku. Úroveň učebnic jsme zvolili A1 až B1 podle Evropského referenčního rámce pro jazyky, a to z toho důvodu, že úroveň B1 je označována za tzv. prahovou, ale také z praktického hlediska – pro tyto úrovně máme nejvíce učebnic, pro úrovně vyšší již málo, proto by nebylo možné srovnání. Celkový počet učebnic je ve skutečnosti o něco málo vyšší, uvádíme pouze 15 učebnic, protože, pokud má učebnice více dílů, ale stejného autora či autory, označujeme ji jako jednu učebnici. Výběr učebnic jsme uskutečnili nejen podle data vydání a úrovně, ale i na základě toho, které učebnice učitelé často využívají (k tomu jsme využili studie Koláčkové (2010) i vlastních zkušeností z praxe), ale i faktu, aby byly zahrnuty učebnice pro různé cílové skupiny studentů, např. anglicky mluvících, Slovanů, azylantů apod.

Tab. 2 Seznam učebnic v korpusu UčKo

1	Adamovičová, A. – Ivanovová, D. (2007): <i>Basic Czech I, II</i> . Praha: Karolinum.
2	Bischofová, J. et al. (2007): <i>Čeština pro cizince a azylanty. Učebnice B1</i> . Brno: SOZE.
3	Bořilová, P. – Holá, L. (2011): <i>Česky krok za krokem 2</i> . Praha: Akropolis.
4	Cvejnová, J. (2010): <i>Česky, prosím</i> . Praha: Karolinum.
5	Čechová, E. – Remediosová, H. (2005): <i>Chcete mluvit česky?</i> Liberec: Harry Putz.
6	Hádková, M. (2005): <i>Čeština pro cizince a azylanty. Učebnice A1, A2</i> . Brno: SOZE.
7	Holá, L. (2006): <i>New Czech Step by Step</i> . Praha: Akropolis.
8	Holá, L. (2011): <i>Čeština Express 1, 2</i> . Praha: Akropolis.
9	Kestřánková, M. et al. (2010): <i>Čeština pro cizince B1</i> . Brno: Cpress.
10	Matula, O. (2007): <i>Český den</i> . Praha: Člověk v tísni o.p.s., Projekt Varianty.

11	Nekovářová, A. (2006): <i>Čeština pro život</i> . Praha: Akropolis.
12	Parolková, O. (2004): <i>Czech for foreigners 1</i> . Praha: Bohemika.
13	Pintarová, M. – Režková, I. (1995): <i>Communicative Czech. Elementary Czech. Intermediate Czech</i> . Praha: Univerzita Karlova.
14	Štindl, O. (2008): <i>Easy Czech. Elementary</i> . Praha: Akronym.
15	Štindlová, B. (2008): <i>Česky v Česku 1, 2</i> . Praha: Ústav jazykové a odborné přípravy UK: Akropolis.

Při samotné tvorbě korpusu je tedy nejdříve nutné učebnice oskenovat, následně soubory pdf zkonvertovat do textového souboru (použili jsme program zabudovaný do Informačního systému Masarykovy univerzity) a poté vyčistit, tj. ručně opravit chyby, které při konverzi vznikly, např. když bylo některé písmeno nebo celé slovo rozeznáno špatně. Pro kompilaci využíváme nástroj WebBootCat vyhledávače Sketch Engine⁵, který umožňuje vložení vlastního korpusu ve formátu txt. Následně je možné vytvořit seznam nejfrekventovanějších slov (wordlist), ve srovnání s ČNK vygenerovat klíčová slova nebo vyhledávat slova/lemmata (Sketch Engine umožňuje lemmatizaci) a jejich kontexty jako v klasickém korpusu. Korpus bychom chtěli rozdělit na 3 subkorpusy dle korpusu úrovně SERRJ (A1, A2, B1)⁶. Záměrem je využití tohoto korpusu pro výzkum slovní zásoby s ohledem na jednotlivé úrovně SERRJ a s porovnáním s autentickým jazykem v ČNK.

Vzhledem k rozsahu korpusu jsme se rozhodli využít pouze texty z učebnic. Zahrnuli jsme však i dialogy, včetně dialogů určených k doplňování (protože je to aktivita vyžadující čtení) a dalších cvičení, které patří k textu (např. otázky). Nakonec jsme přidali i modelové věty prezentující gramatický jev či význam slova, neboť jsou často prezentovány ve formě malého dialogu, a to zvláště u nižších pokročilostí. Tam se ve většině případů kryje text určený k procvičování čtení a modelový text nebo dialog prezentující nové formy.

Jaké problémy při budování nastaly? Již při myšlence rozdělení knih podle SERRJ vzniká u některých učebnic překážka. Často jsou totiž orientovány na výuku více úrovní, např. A1 až A2, nebo dokonce A1 až B1. Je tedy obtížné někdy rozhodnout, ve které části učebnice určit tuto hranici mezi úrovněmi. Nejlépe na tom jsou v tomto ohledu učebnice nejnižší úrovně A1, zřejmě i kvůli tomu, že na této úrovni cizinci skládají Zkoušku z češtiny pro trvalý pobyt. Učebnic úrovně A1 je proto největší množství a často jsou autory cíleně omezeny na tuto nejnižší úroveň.

Další obtíže nastávají při tzv. čištění textového souboru. Učebnice jsou často barevné a obsahují obrázky, fotografie a grafickou úpravu stran, proto některé věty, např. ty, jež jsou otištěny světlým písmem na tmavém pozadí, nejsou programem vůbec rozeznány. Z tohoto důvodu je tvorba korpusu, přestože malého rozsahu, velmi časově náročná.

5. Výběr slovní zásoby v korpusu učebnic ČCJ úrovně A1

V následující části prezentujeme výsledky první studie na základě učebnicového korpusu, kde jsme se zaměřili na výběr sloves v učebnicích úrovně A1 (Vališová, 2013). Sloveso je centrem české věty, a proto je důležité studenty seznámit se slovesnými tvary a vybrat taková slovesa, která budou při komunikaci v základních situacích používat. V této analýze jsme postupovali metodou data-driven. Nejprve jsme tedy získali seznam sloves z první části našeho korpusu a tato data jsme

5 <http://sketchengine.co.uk/>

6 Popisy úrovně pro češtinu jako cizí jazyk: <http://www.msmt.cz/mezinarodni-vztahy/referencni-urovne-pro-cestinu-jako-cizi-jazyk>.

následně porovnali se seznamem sloves z doporučené slovní zásoby v popisu úrovně A1 pro češtinu jako cizí jazyk (Hádková et al., 2005) a některá slovesa také s daty z ČNK (SYN2010). Naše výchozí otázky byly: Která slovesa by měla být prezentována na úrovni A1? Zahrnuje úroveň A1 i slovesa dokonavá?

Úroveň A1 je považována za základní úroveň téměř bez využití gramatiky. Ústní projev studenta na úrovni A1 znamená, že umí „jednoduchými frázemi a větami popsat místo, kde žije, a lidi, které zná“. Nicméně při výuce češtiny jako cizího jazyka se nevyhneme gramatice ani na nejnižší úrovni A1. Tato úroveň je také důležitá z toho důvodu, že z ní musí dělat zkoušku ti cizinci, kteří žádají o trvalý pobyt v České republice. Proto je součástí popisu úrovně i tzv. gramatické minimum. Podle Metodiky ke zkoušce z češtiny pro trvalý pobyt (Cvejnová a kol., 2010) tam patří: minulý, přítomný a budoucí čas nedokonavých sloves, imperativ a kondicionál pouze ve frázích (Ibidem, s. 22), zatímco oficiální popis úrovně doporučuje prezentovat dokonavá slovesa jako zvláštní skupinu sloves a také vyučovat konkrétní tvary dokonavých sloves, nejlépe s kontextem, tzn. v nějaké frázi (Hádková et al., 2005: s. 255).

Pro studii jsme využili první dobudovanou část učebnicového korpusu sestávající z textů, dialogů a modelových vět ze 7 učebnic češtiny jako cizího jazyka na úrovni A1. Tento podkorpus (UčKo-A1) má téměř 62 tisíc slov. Vytvořit jej bylo jednodušší v tom, že všechny tyto učebnice zahrnují učivo pouze základní úrovně A1 podle SERRJ. Z korpusu UčKo-A1 jsme vygenerovali seznam slov, ze kterého jsme následně ručně vytřídili všechna slovesa. Celkem bylo v učebnicích 135 sloves, a z toho 32 (24%) z nich dokonavých. Pokud se podíváme blíže na jejich kontexty, zjistíme, že jsou tato slovesa prezentována vždy v konkrétním tvaru a konkrétním kontextu, studenti se je tedy učí jako frázi a nemusí nic o vidu ještě znát.

Slovesa jsou nejčastěji prezentována v infinitivu (1), a to obvykle po modálních slovesech, dále v imperativu (2), minulém čase (3) nebo v perfektivním futuru (4), což však studenti vnímají jako přítomnost.

- (1) pomoci: 6 tokenů, 4 učebnice
Můžete mi < pomoci > prosím vás?
- (2) prominout: 36 tokenů, 4 učebnice
Aha. < Promiňte >, jak se to píše?
- (3) narodit se: 17 tokenů, 6 učebnic
Jan se < narodil > 13. 5. 1980.
- (4) vzít: 8 tokenů, 5 učebnic
< Vezmu > si čtyři rohlíky.

Výsledky jasně ukazují, že studenti nejsou seznámeni s kategorií vidu, ale učí se konkrétní tvary v rámci fráze potřebné pro danou komunikační situaci. Překvapivé však bylo, že některá z těchto sloves byla zahrnuta pouze ve 2 či 3 učebnicích, což znamená, že autoři se jim snaží co nejvíce vyhýbat. Naše otázka však je: Proč se dokonavým slovesům vyhýbat, když stačí prezentovat pouze jednu jejich formu?

Při srovnání se seznamem sloves z doporučené slovní zásoby popisu úrovně pro češtinu jako cizí jazyk (Hádková et al., 2005) uvidíme velké rozdíly. Tento deskriptor byl napsán v roce 2005 a je vyvěšen na webových stránkách MŠMT, autoři učebnic, které vyšly v rozmezí let 2005–2010, by s ním tedy měli být seznámeni (jedna z autorek je navíc spoluautorkou deskriptoru). Popis úrovně A1 zahrnuje seznam slovní zásoby, kterou je doporučeno vyučovat, povinná slovní zásoba je označena, ostatní je fakultativní, určena pro pasivní znalost. Vybrali jsme tedy pouze slovesa z povinné slovní zásoby, tj. 162 sloves, z čehož 51 (31 %) bylo dokonavých. V učebnicích však bylo dokona-

vých sloves pouze 24 %. Proč tak velké rozdíly? Vzhledem k tomu, že součástí korpusu jsou pouze texty, očekávali bychom, že tam bude dokonavých sloves víc, neboť čtecí aktivity zahrnují i pasivní porozumění neznámé slovní zásoby podle kontextu. Důvod může být ten, že v popisu úrovně jsou zahrnuta některá slovesa nevhodná pro tuto úroveň, např. *proclít*, nebo že autoři vycházeli více ze své praxe než z popisu úrovně (Vališová, 2013).

Řešením může být porovnání učebnic mezi sebou (např. s pomocí dat z učebnicového korpusu) a ve srovnání s deskriptorem zjistit, jaká slovní zásoba se shoduje v různých učebnicích i v deskriptoru, tzn. vybrat nejčastěji prezentovanou slovní zásobu. Pomůckou mohou být také frekvenční data z Českého národního korpusu (ČNK). Lze vyhledat rozdíl ve využití některých variant (viz <http://syd.korpus.cz/>) a z daných údajů vycházet při výběru slovní zásoby nebo vyhledat nejčastější tvary daného slova a na nejnižší jazykové úrovni A1 prezentovat jenom jeden nebo dva vybrané tvary sloves.

Uvedme si jako příklad sloveso *říct*. Bylo sice zahrnuto ve všech učebnicích v korpusu UČKo-A1, ale minulý čas *řekl* pouze v jedné z nich (*Česky v Česku II*), přitom podle ČNK je nejfrekventovanější forma právě minulý čas *řekl*. Příčinou může být fakt, že učebnice většinou prezentují dialogy, a ne souvislé texty. Druhým důvodem může být nepravidelnost tohoto slovesa: *říct* – *řeknu* – *řekl*, kde se jednotlivé formy liší. Studenti toto slovo většinou znají pouze z fráze: *Jak se to řekne česky?*, popř. z několika dalších. Korpus nás však tady upozorňuje na fakt, že bychom se ani na úrovni A1 neměli vyhýbat souvislým textům a čtecím aktivitám a že data z ČNK nám spolu s tématem lekce i referenční úrovní mohou být nápomocná při výběru slovní zásoby do výukových materiálů.

6. Závěr

Na závěr bychom chtěli shrnout, že tento typ korpusu nám otevírá nové oblasti výzkumu učebnicového jazyka i didaktických metod, je však nutné mít na paměti, že učebnicový korpus je i přes svoje výhody velmi limitovaný tím, že učebnice redukuje pouze na text a nelze zahrnout interakci ve třídě, která je nedílnou součástí výuky jazyka. Učebnicový korpus také nabízí komparaci s autentickým jazykem národních korpusů, je však nutné vzít v potaz, že přestože studenti mohou těžit z autentického jazykového inputu, adaptované či vymyšlené texty jsou také nedílnou součástí učebnic, zvláště pro nižší pokročilosti. I přesto se však domníváme, že jeho výhody nad těmito limity převažují.

Učebnicový korpus dodá přesnější data než klasická analýza učebnic a tato data můžeme následně porovnat s daty z národního korpusu nebo s popisem dané úrovně podle SERRJ. S přihlédnutím k frekvenci nám data z ČNK pomůžou vybrat nejčastěji užívaná slova, popř. fráze (v naší studii tvary dokonavých sloves), a mohou tak zlepšit výukové materiály využitím autentického jazyka.

Použitá literatura

Učebnice češtiny jako cizího jazyka

ADAMOVIČOVÁ, A. – IVANOVÁ, D. (2007): *Basic Czech I*. Praha: Karolinum.

CVEJNOVÁ, J. (2010): *Česky, prosím I*. Praha: Karolinum.

HÁDKOVÁ, M. (2005): *Čeština pro cizince a azylanty. Učebnice A1*. Brno: SOZE.

HOLÁ, L. (2011): *Čeština Express 1, 2*. Praha: Akropolis.

MATULA, O. (2007): *Český den*. Praha: Člověk v tísni o.p.s., Projekt Varianty.

ŠTINDL, O. (2008): *Easy Czech. Elementary*. Praha: Akronym.

ŠTINDLOVÁ, B. (2008): *Česky v Česku 1, 2*. Praha: ÚJOP UK, Akropolis.

Sekundární literatura

- BOULTON, A. (2012): Hands-on / hands-off: Alternative approaches to data driven learning. In: J. Thomas – A. Boulton (eds.), *Input, Process and Product. Development in Teaching and Language Corpora*. Brno: Masaryk University Press, s. 152–168.
- CVEJNOVÁ, J. et al. (2010): *Metodika přípravy ke zkoušce z českého jazyka pro žadatele o trvalý pobyt (úroveň A1)*. Praha: VÚP. [online]. Cit. 2013-11-26. <http://cestina-pro-cizince.cz/uploads/Dokumenty/cest_pro_ciz_final.pdf>.
- CVRČEK, V. et al. (2010): *Mluvnice současné češtiny*. Praha: Karolinum.
- CVRČEK, V. – VONDŘIČKA, P. (2011): *SyD – Korpusový průzkum variant*. Praha: FF UK. [online]. Cit. 2013-11-26. <<http://syd.korpus.cz/>>.
- FRANKENBERG-GARCIA, A. (2012): Integrating corpora with everyday language teaching. In: J. Thomas – A. Boulton (eds.), *Input, Process and Product. Development in Teaching and Language Corpora*. Brno: Masaryk University Press, s. 36–53.
- GOUVERNEUROVÁ, C. (2008): The phraseological patterns of high-frequency verbs in advanced English for General Purposes: a corpus-driven approach to EFL textbook analysis. In: F. Meunierová – S. Grangerová (eds.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam & Philadelphia: Benjamins, s. 223–246.
- GOUVERNEUROVÁ, C. – MEUNIEROVÁ, F. (2009): New types of corpora for new educational challenges. Collecting, annotating and exploiting a corpus of textbook material. In: K. Aijmer (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins.
- HÁDKOVÁ, M. – LÍNEK, J. – VLASÁKOVÁ, K. (2005): *Čeština jako cizí jazyk. Úroveň A1*. Olomouc: Univerzita Palackého v Olomouci. [online]. Cit. 2013-11-26. <<http://www.msmt.cz/mezinarodni-vztahy/referencni-urovne-pro-cestinu-jako-cizi-jazyk>>.
- HRDLIČKA, M. (2000): *Předložky ve výuce češtiny jako cizího jazyka*. Praha: Karolinum.
- HRDLIČKA, M. (2010): *Gramatika a výuka češtiny jako cizího jazyka*. Praha: Karolinum.
- HUSTONOVÁ, S. (2010): *Corpora in Applied linguistics*. Cambridge: Cambridge University Press.
- KOLÁČKOVÁ, L. (2010): Hlediska, podle nichž učitelé ČCJ vybírají učebnice češtiny pro cizince. In: K. Hlínová (ed.), *Sborník Asociace učitelů češtiny jako cizího jazyka (AUCCJ) 2010*. Praha: Akropolis, s. 95–100.
- OSOLSOBĚ, K. – VALIŠOVÁ, P. (2011): Tagset korpusů ČNK z hlediska předpokládané znalosti gramatické terminologie u nerodilých mluvčích (Možnosti a meze využívání korpusů češtiny pro nerodilé mluvčí). In: I. Dominiková – M. Lachout (eds.), *Lingua terminologica*. Praha: MUP, s. 141–156.
- RÖMEROVÁ, U. (2004): A corpus-driven approach to modal auxiliaries and their didactics. In: J. Sinclair (ed.), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, s. 185–199.
- RÖMEROVÁ, U. (2005): Looking at looking: Functions and contexts of progressives in spoken English and ‚school‘ English. In: A. Renouf – A. Kehoe (eds.), *The Changing Face of Corpus Linguistics*. Amsterdam: Rodopi.
- Společný referenční rámec pro jazyky: jak se učíme jazykům, jak vyučujeme a jak v jazycích hodnotíme* (2006). [online]. Cit. 2013-11-26. <<http://www.msmt.cz/mezinarodni-vztahy/spolecny-evropsky-referencni-ramec-pro-jazyky>>.
- ŠKODOVÁ, S. – ŠVINDLOVÁ, B. (2007): Modifikace principů přímé metody pro potřeby výuky gramatiky češtiny jako cizího jazyka. In: J. Čemusová – B. Štindlová (eds.), *Sborník Asociace učitelů češtiny jako cizího jazyka (AUCCJ) 2006–2007*. Praha: Akropolis.
- VALIŠOVÁ, P. The Choice of Verbs in Czech as a Foreign Language Textbooks. In: V. P. Zacharov – O. A. Mitrofanova – M. V. Chochlova. *Proceedings of the International Conference “Corpus Linguistics – 2013”*. St. Petersburg: St. Petersburg State University, s. 138–147.

TVORBA A VYUŽITÍ MALÝCH SPECIALIZOVANÝCH KORPUSŮ NA PŘÍKLADU *KORPUSU ODESLANÉ A PŘIJATÉ ČESKY PSANÉ KORESPONDENCE BEDŘICHA SMETANY*¹

LUCIE RYCHNOVSKÁ

Ústav českého jazyka FF MU Brno

Abstract

The creation and use of small specialized corpuses by the way of example of the *Bedřich Smetana's Czech Written Correspondence Corpus*

The paper deals with the *Bedřich Smetana's Czech Written Correspondence Corpus*, it shows how can be used small specialized corpuses for the analysis of some person's or some period's language. It also describes advantages of so-called additional annotations of letters (it means sociolinguistic and territorial tags or information about the date of letter-writing).

Klíčová slova: korpus – soukromá korespondence – Bedřich Smetana

Key words: Corpus – Private Correspondence – Bedřich Smetana

Jazykový korpus je vnitřně strukturovaný a unifikovaný soubor elektronicky uložených jazykových dat. Jeho smysl spočívá v tom, že se v něm textová data nacházejí v přirozeném kontextu, a jestliže je korpus dostatečně rozsáhlý (v řádu milionů slovních výskytů), je schopen zachytit typická užití jazykových jednotek a jejich frekvenční distribuci.² Na příkladu korpusu česky psané korespondence Bedřicha Smetany se pokusíme ukázat, že smysl nepostrádá ani tvorba malých specializovaných korpusů, které výše uvedené šíře nedosahují, a že i tyto korpusy mohou být nápomocny při jazykovém bádání.

Korpus Smetanovy korespondence patří do linie korpusů soukromé korespondence vznikajících na Filozofické fakultě Masarykovy univerzity. Korpusovému zpracování a zkoumání tohoto typu textů se věnuje zejména Zdeňka Hladká, výsledkem její odborné činnosti v této oblasti jsou především práce *Čeština v současné soukromé korespondenci. Dopisy, e-mail, SMS* (2005) a *111 let českého dopisu v korpusovém zpracování* (2013). Součástí publikací jsou v obou případech dvě CD, která vedle korpusového uložení obsahují i plné znění a fotokopie shromážděného korespondenčního materiálu. V korpusové podobě jsou korespondenční texty (v prvním případě z let 1995–2005 a ve druhém z let 1902–2012) uloženy pod manažer Bonito, díky němuž je mj. možné vyhledávat statistické údaje o frekvenci výskytu různých jazykových jevů. Uvedený korpusový manažer je dobře znám, neboť jej využíval také Český národní korpus, nebudeme se proto věnovat popisu všech možností, které nabízí, nýbrž se zaměříme pouze na specifika brněnských korpusů.

Ke všem korespondenčním textům zpracovaným v rámci výše uvedených prací byly připojeny tzv. doprovodné údaje. Těmi se rozumí sociologické charakteristiky pisatele a adresáta, tedy informace o jejich věku, pohlaví, vzdělání a jejich vzájemném vztahu (postižení toho, zda jde o příbuzné,

1 Příspěvek vznikl v rámci grantového projektu specifického výzkumu MUNI/A/1165/2014 *Čeština v jednotě synchronie a diachronie 2015*.

2 Petkevič, V. (2002): Korpus jazykový. In: P. Karlík – M. Nekula – J. Pleskalová (eds.): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, s. 235.

milence, přátele či známé). Významným novem, s nímž přišla už první zmíněná publikace, se stala skutečnost, že kromě toho byla vůbec poprvé reflektována rovněž nářeční příslušnost pisatele (resp. to, v jaké lokalitě se dlouhodobě pohybuje, a jakou teritoriálně příznakovou mluvou tudíž může být ovlivněn). Díky speciální modifikaci manažeru Bonito (kterou vytvořil Pavel Rychlý z Fakulty informatiky MU) lze uvedené charakteristiky libovolně kombinovat a vytvořit tak sub-korpus pouze určitého typu dopisů, v němž má vyhledávka proběhnout.

Sociologické a teritoriální označování dopisů je funkční zejména s ohledem na povahu zkoumaného materiálu: žánr soukromého dopisu stojí na pomezí projevů psaných a mluvených, kontrolovaných a spontánních. Psanost je přitom většinou vázána na spisovnost, a tedy uvědomělý výběr jazykových prostředků, a tak lze díky doprovodným údajům sledovat, do jaké míry (nebo zda vůbec) i přesto do těchto projevů pronikají prvky nespisovné a do jaké míry (pokud vůbec) je jejich průnik vázán na teritoriální příslušnost pisatele. Zkoumat lze dále například i to, jak je výběr konkrétních jazykových prostředků (spisovných i nespisovných) ovlivněn věkem, pohlavím či vzděláním pisatele.

V jisté modifikaci využívá stejného principu rovněž korpus zpracovávající materiál starší, totiž *Korpus odeslané a přijaté česky psané korespondence Bedřicha Smetany* vytvořený na Filozofické fakultě MU ve spolupráci s Fakultou informatiky MU roku 2009. Korpus je rozdělen na dva subkorpusy, a to smet1 obsahující 424 dopisů odeslaných a smet2, který zahrnuje 337 listů přijatých. Materiálovým východiskem korpusového zpracování se staly elektronické přepisy těch dopisů, jež byly v době tvorby korpusu k dispozici v Muzeu Bedřicha Smetany v Praze. Vyloučeny přitom byly jednak listy psané cizí rukou a Smetanou pouze podepsané (které – ačkoli třeba vznikly podle Smetanova konceptu – mohl skutečný pisatel korigovat), jednak dopisy či jejich koncepty z posledních let Smetanova života (které jsou poznamenány Smetanovou psychickou chorobou, a jejichž výpovědní hodnota pro jazykové bádání tudíž klesá). Pro účely korpusového zpracování musely být dopisy částečně jazykově a formálně upraveny, v maximální míře však byla zachována věrnost výsledné verzi zápisu Smetanovou rukou. Nakonec byly ke všem dopisům (stejně jako u korpusů současné soukromé korespondence) připojeny doprovodné údaje. Pro účely zkoumání Smetanovy korespondence se ukázala jako zásadní zejména informace o roce napsání dopisu, dále o adresátovi (u korespondence odeslané) nebo pisateli (u korespondence přijaté) a s tím související určení vztahu těchto osob ke Smetanovi (rozlišení dopisů manželce – značka M, dětem – značka D, ostatním rodinným příslušníkům – značka R, přátelských – značka P, polooficiálních – značka L, oficiálních jednotlivci – značka O a oficiálních institucí – značka I). Díky těmto údajům je možné sledovat rozložení zkoumaných jevů ve Smetanově korespondenci z hlediska časového či sociálního.

Vyvstává samozřejmě otázka, proč zkoumat právě jazyk Bedřicha Smetany a proč věnovat úsilí budování korpusu právě jeho korespondence. Zatímco o zajímavosti dopisů tohoto význačného hudebního skladatele po obsahové stránce asi nebude pochyb, o jejich stránce jazykové to platit nemusí. Přesto si i ta zaslouhuje naši pozornost. Bedřich Smetana (1824–1884) je z hlediska jazykového vyjadřování zajímavý tím, že neměl české školy, tzn. prošel takovými vzdělávacími institucemi, kde se vyučovalo v němčině a kde čeština nebyla ani vyučovaným předmětem. Z poznámek ve Smetanových denících, popř. z některých dalších dokumentů (např. matričních či biřmovacích zápisů), je ovšem zřejmé, že jazykem privátní komunikace byla v rodině Smetanových čeština.³ Smetana byl tedy česko-německý bilingvní mluvčí, češtinu ovšem zprvu pravděpodobně znal

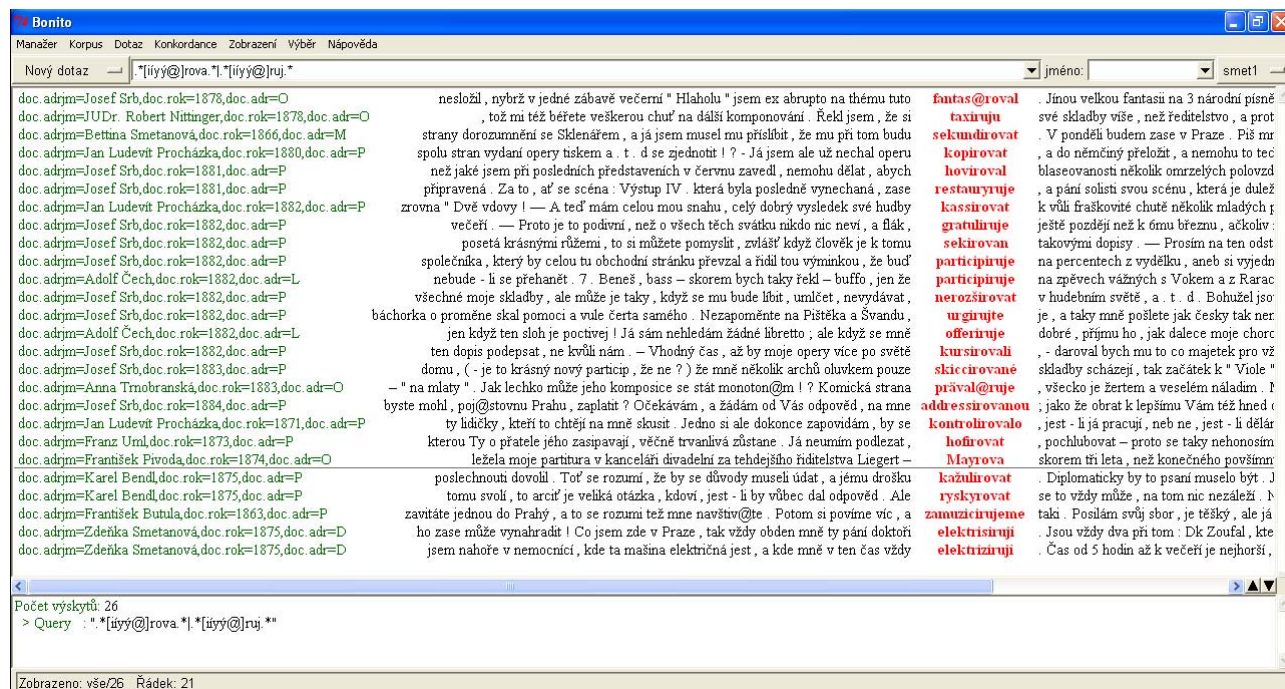
3 K tomu blíže např. NEKULA, M. – RYCHNOVSKÁ, L. (2010): Smetanova čeština v dobovém kontextu. *Hudební věda* 47, č. 1, s. 43–76 nebo NEKULA, M. – RYCHNOVSKÁ, L. (2012): Jakou češtinu užíval Bedřich Smetana? / Bedřich Smetana's Use of the Czech Language. *Musicalia. Časopis Českého muzea hudby / Journal of the Czech Museum of Music* 1–2, s. 6–38.

a užíval pouze v její podobě mluvené. V projevech psaných na ni přechází až r. 1860, tj. ve svých 36 letech. Podnětem pro to se stal česky psaný dopis Jana Ludevíta Procházky z 11. března 1860 zasláný Smetanovi do Göteborgu, kde Smetana v té době působil jako učitel hudby a dirigent. Procházka se na Smetanu obrací s prosbou, zda by nezkomponoval hudbu pro mužské sbory a nepřispěl tím do národně buditeckého procesu. Smetana souhlasí a výrazem jeho zapojení do národního hnutí se stává nejen tvorba hudební, ale také příklon k češtině jednak v korespondenci s českými adresáty, jednak v denících.

Smetanovy dopisy se tak stávají materiálem, na němž je možné sledovat, jaké jevy činí člověku osvojivacím si český pravopis a gramatiku až v dospělém věku potíže, v jaké míře do takového projevu pronikají prvky odchylující se od dobových nároků na kultivovaný psaný projev (tedy prvky typické pro projev mluvený, prvky více či méně teritoriálně zabarvené) a nakolik je český projev bilingvního mluvčího ovlivňován jeho druhým jazykem (v tomto případě němčinou). Výhodou se přitom stává skutečnost, že Smetana byl z hlediska jazykového průměrným člověkem své doby, a tak vzniká předpoklad, že (alespoň některé) jevy, jež se vyskytují u Smetany, byly vlastní i projevům jiných mluvčích 19. století. Díky *Korpusu odeslané a přijaté česky psané korespondence Bedřicha Smetany*, který zachycuje korespondenci obousměrnou, lze jednoduše ověřit, zda tomu tak skutečně je.

Jakým způsobem je tedy korpus při analýze Smetanova vyjadřování v češtině využitelný? Své uplatnění nalézá především při zkoumání jevů systémových, tzn. hláskoslovných, morfologických a časově také pravopisných. V těchto případech se lze vhodně zadaným dotazem vyhnout nutnosti procházet všechny dopisy ručně. Jako příklad můžeme uvést zkoumání frekvence výskytu slovesných tvarů přejímaných z němčiny (resp. přes němčinu), které obsahují německý komponent *-i/ýr-*, k němuž je připojována česká infinitivní přípona *-ova-* nebo přítomní *-uj-*. Dotaz by mohl vypadat následovně: **[íý]rova.*|[íý]ruj.**, při jeho tvorbě je ovšem nutné vzít v úvahu rozkolísanou kvantitu vokálů ve Smetanových projevech (u *i/i* navíc komplikovanou nedbalostí zápisu, takže v některých případech nelze rozhodnout, zda je psána tečka či čárka; v korpusu je takový sporný grafém nahrazován znakem @). Dotaz tudíž musíme upravit na: **[íýý@]rova.*|[íýý@]ruj.** (viz obr. 1). Vzhledem k tomu, že korpus není morfologicky označován, je dále nutné ruční dotřídění, kterým eliminujeme výsledky hledání jako *Mayrova* (viz podtržený výsledek hledání na obr. 1).

Obrázek 1



Uvedená slovesa jsou zajímavá tím, že ve sledovaném období vzniká tlak na jejich odstranění, preferována je podoba bez *-í/ýr-*.⁴ Např. ve slovníku Františka Štěpána Kotta nalezneme většinu sloves objevujících se ve Smetanově korespondenci s *-í/ýr-* pouze bez této přípony. Kott má tedy jen *adressovati*, *fantasovati* (s poznámkou *Ne: fantasirovati*), *gratulovati*, *kassovati*, *participovati*, *restaurovati*, *musikovati/muzikovati*, jako ekvivalent německého *elektrisieren* uvádí *elektřiti*, *elektrovati* atd.⁵

Ukazatelem toho, že podoby prokazující vliv němčiny byly i přesto ještě koncem 19. století zřejmě běžně užívané, je rozložení těchto slovesných tvarů ve Smetanově korespondenci. Z doprovodných charakteristik zjišťujeme (viz obr. 1 informace uvedené zeleně), že jejich výskyt není vázán jen na určitý druh korespondence nebo např. jen na dopisy z dřívější doby, nýbrž se objevují v dopisech od 60. let až po listy z posledních roků Smetanova života a z hlediska typologického se nalézají jak v korespondenci rodinné a přátelské, tak také polooficiální a oficiální. Závěry o obecném rozšíření zkoumaných slovesných tvarů lze navíc snadno potvrdit pohledem do korespondence přijaté. Zadáme-li stejný typ dotazu v rámci subkorpusu smet2, nalezneme tvary jako *urgiroval jsem; by se [...] rentirovalo* (v dopisu Jana Ludevíta Procházky z r. 1882); *aplaudivovat; jest velmi krásně aranžirovana* (v dopisech Josefa Palečka, bez datace) apod. (viz obr. 2).

Obrázek 2

The screenshot shows the Bonito search interface. At the top, there are tabs for 'Manažer', 'Korpus', 'Dotaz', 'Konkordance', 'Zobrazení', 'Výběr', and 'Nápověda'. Below the tabs, there is a search bar with the query '*[íýř@]rova.*|[íýř@]ruj.*' and a dropdown menu for the corpus 'smet2'. The search results are displayed in a table with columns for document information (doc.pis=L, doc.pisjm=Eliška Krásnohorská, doc.rok=1879), the search term, and the context of the term in the original text. The search term is highlighted in red in the original image. The context is shown in a smaller font and is partially obscured by the search term. The search results are sorted by the number of occurrences, with 14 results shown.

V menší míře se naopak korpus uplatňuje při zkoumání roviny lexikální či syntaktické. Z povahy jevů příslušejících k těmto jazykovým plánům plyne, že je lze těžko formalizovat do podoby nějakého dotazu, mnohdy se totiž v jejich případě jedná o jedinečné výskyty, často je pro jejich posouzení nezbytný širší kontext, znalost celé struktury sdělení. Většinou je tedy nutné praco-

4 K tomu srov. KAMIŠ, A.: Citovat × citýrovat, pauzovat × pauzírovat. *Naše řeč* 64, č. 3, s. 162–163.

5 Srov. příslušná hesla in KOTT, F. Š. (1878–1893): *Česko-německý slovník zvláště grammaticko-fraseologický*. Praha: Tiskem a nákladem knihtiskárny Josefa Koláře / Tiskem a nákladem knihtiskárny Františka Šimáčka.

vat přímo s originály dopisů či jejich věrnými přepisy. Korpus nám však může být oporou, bude-li nás zajímat např. zastoupení konkrétní lexikální jednotky či slovesné vazby nebo jejich rozložení v celku Smetanovy korespondence.

Korpusové zpracování materiálu výrazně usnadňuje analytickou část jazykového bádání, a to všude tam, kde by ruční procházení textů při zkoumání jednoho každého jevu nebylo únosné. Prostřednictvím elektronického korpusu lze jednoduše zjistit nejen to, zda vůbec se určitý jev ve zkoumaném typu komunikátů objevuje, ale také to, v jaké frekvenci je v nich zastoupen. Socio-lingvistické a teritoriální označkování materiálu pak vytváří prostor pro rozšíření bádání. Lze sledovat, v jakém typu komunikátů se určitý jazykový jev vyskytuje, na jakou jazykovou situaci a jaké mluvčí je vázán, zaměřujeme-li se na projevy jednoho mluvčího, lze zkoumat, k jakým případným posunům (v čase či s ohledem na adresáta) v jeho vyjadřování dochází atd.

Pokud jsou navíc různé malé korpusy tvořeny dle totožného principu a pokud je tento princip kompatibilní s korpusovým zpracováním jazykového materiálu v rámci rozsáhlých národních korpusů (jako tomu je v případě korpusů vznikajících na Filozofické fakultě MU a Českého národního korpusu), pak se otevírá možnost srovnání výsledků jednotlivých dílčích bádání i možnost učinit z korpusů malých specializovaných součástí korpusů velkých, široce pojatých.

Seznam pramenů a literatury:

Korpus odeslané a přijaté česky psané korespondence Bedřicha Smetany [CD-ROM] (2009). Brno: ÚČJ FF MU.

HLADKÁ, Z. a kol. (2005): *Čeština v současné soukromé korespondenci. Dopisy, e-maily, SMS* [2 CD-ROM]. Brno: Masarykova univerzita.

HLADKÁ, Z. a kol. (2013): *111 let českého dopisu v korpusovém zpracování* [2 CD-ROM]. Brno: Masarykova univerzita.

KAMIŠ, A.: Citovat × citýrovat, pauzovat × pauzírovat. *Naše řeč* 64, č. 3, s. 162–163.

KOTT, F. Š. (1878–1893): *Česko-německý slovník zvláště grammaticko-fraseologický*. Praha: Tiskem a nákladem knihtiskárny Josefa Koláře / Tiskem a nákladem knihtiskárny Františka Šimáčka.

NEKULA, M. – RYCHNOVSKÁ, L. (2010): Smetanova čeština v dobovém kontextu. *Hudební věda* 47, č. 1, s. 43–76.

NEKULA, M. – RYCHNOVSKÁ, L. (2012): Jakou češtinu užíval Bedřich Smetana? / Bedřich Smetana's Use of the Czech Language. *Musicalia. Časopis Českého muzea hudby / Journal of the Czech Museum of Music* 1–2, s. 6–38.

PETKEVIČ, V. (2002): Korpus jazykový. In: P. Karlík – M. Nekula – J. Pleskalová (eds.): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, s. 235.

TVORBA KORPUSU NOVINOVÝCH TITULKŮ A JEHO ANALÝZA

PAVLÍNA SEDLÁŘOVÁ

Ústav českého jazyka FF MU Brno a Laboratoř zpracování přirozeného jazyka na Fakultě informatiky MU Brno

Abstract

The aim of our work was to create and analyze the corpora of tabloid headlines in order to get sufficient amount of data which should be used to develop new algorithm. This algorithm allows us to sort records by submitted criterias (to tabloid headlines and serious headlines). The key issue was to detect the characteristic signs of tabloid headlines. Assuming all knowledge gained in the process, we made a list of the most characteristic attributes of tabloid headlines. This list was changed and rewritten in order to make it readable by computer, such that it could be useful in the machine learning process and in the developing new algorithm for recognition tabloid headlines. In my contribution, I introduce results, which we gained, make readers acquainted with the methodology of work. I also explain, how we proceeded with the testing of rules based on detected attributes of tabloid headlines and outline possible development of our work.

Klíčová slova: elektronická média, online verze českých deníků, rysy bulvárních titulků, tvorba korpusu, morfologická analýza, Support Vector Machines, strojové učení

Key words: electronic media, online version of the Czech journals, building corpus, main features of tabloid titles, morphological analysis, Support Vector Machines, machine learning

Naším cílem bylo vytvořit a analyzovat korpus bulvárních titulků, abychom získali dostatečný objem dat pro vyvinutí algoritmu, který umožňuje filtrování záznamů (na titulky bulvární a nebulvární) na základě předložených kritérií. Stěžejním bodem pro nás tedy bylo zjistit, čím se titulky bulvárních článků vyznačují a co je pro ně charakteristické. Na základě poznatků, které jsme při tomto procesu získali, jsme sestavili seznam nejvýraznějších rysů bulvárních titulků. Tento seznam jsme dle potřeb upravili tak, aby bylo možné jej počítačově zpracovat, a využít ho tak v procesu strojového učení při tvorbě algoritmu, který by mohl automaticky rozpoznávat bulvární titulky. V příspěvku představím výsledky, kterých jsme dosáhli, seznámím čtenáře s metodikou práce, vysvětlím, jak jsme postupovali při testování pravidel založených na vyznačených rysech bulvárních titulků, a nastíním budoucí vývoj naší práce.

Úvod

Internet je v dnešní době plný informací, z nichž většina je pro nás naprosto nezajímavá a neužitečná. Pro velkou část uživatelů internetu mohou být těmito informacemi na příklad bulvární články. Společně s Markem Grácem jsme se rozhodli zkusit navrhnout program, který by nežádoucí data tohoto typu dokázal odfiltrovat. Cílem naší práce tedy bylo přijít na to, čím se titulky takových článků vyznačují a co je pro ně charakteristické. Tyto poznatky pak byly zdrojem informací při tvorbě algoritmu, který bude automaticky rozpoznávat bulvární titulky. Vznikne tak účinný filtr, který by uživatel mohl využít např. na svém účtu na sociální síti, v mailové schránce a na dalších místech, kde mu jsou (mnohdy ve formě reklamy) nabízeny články různých internetových deníků. Vybral by si tak pouze ty texty, které ho zajímají. Tím by alespoň částečně mohl eliminovat množství informací, které se na něj při práci na internetu hrne.

Pro tvorbu funkčních a použitelných dat jsme využili strojové učení, a to tzv. strojové učení s učitelem (dohledem), konkrétně metodu SVM (**Support Vector Machines** – algoritmy podpůrných vektorů; podrobněji dále). Celý proces tvorby korpusu a jeho analýzy se skládal ze tří částí, přičemž v první fázi bylo nutné nasbírat dostatečné množství dat a ručně je označkovat. Tato fáze byla nepostradatelným základním stavebním kamenem celého procesu strojového učení. Bylo potřeba nastavit procesy tak, aby všichni anotátoři pracovali konzistentně. Na základě výsledků této fáze totiž teprve mohla začít ta nejdůležitější a nejobsáhlejší etapa, a to předkládání dat programu, testování funkčnosti a úpravy pravidel vytvořených na základě poznatků získaných analýzou souboru vzorků. Třetí, poslední část pak tvoří výsledky, které jsou připraveny k dalšímu zpracování, avšak tentokrát již nejsou v potaz brány zdrojové soubory. V praxi to znamená, že v klíčové druhé fázi je nutné vypořádat hlavní rysy zkoumaného vzorku, převést je na počítačově srozumitelné příkazy a získat vektory (posloupnosti znaků, které nesou informaci o výsledcích jednotlivých testů – v tomto případě jde o pravdivostní hodnotu **true/false**, která je odpovědí na to, zda daný titulek odpovídá zadanému pravidlu). Tyto vektory pak tvoří třetí část metody strojového učení. Zpracování výsledků, tedy strojové učení, je pak z hlediska programování relativně jednoduché. Klíčovým je samotný sběr dat (a následná analýza), který zpravidla zabere většinu času celého procesu. Zatím jsme se věnovali především prvním dvěma fázím strojového učení. Výstup, který jsme takto získali, poslouží jako podpůrný systém při tvorbě funkčního algoritmu na automatické rozpoznávání (nejen) bulvárních článků. Pokud bude jeho přípravě věnována dostatečná pozornost, může být ve své elementární podobě využitelný i pro jiné operace (vyhledávání, rozpoznávání, filtrace záznamů) v oblasti počítačové lingvistiky.

Značkování

Do korpusu jsme zařadili titulky z internetových serverů **Super.cz**, **Blesk.cz**, **Ahaonline.cz**, **e15.cz**, **iHNed.cz**, **iDnes.cz** a **Novinky.cz** (poslední čtyři internetová média jsou do projektu zařazena jako zástupci hlavního proudu). Nebulvární deníky jsme zařadili proto, abychom získali dostatečný objem dat, na základě kterých jsme mohli získat srovnání, a tak potvrdit naše poznatky. Zaměřili jsme se hlavně na rubriku **Zprávy**. Stažení titulků z internetových online verzí výše zmíněných deníků zajistil Mgr. Marek Grác, PhD. Celkový počet titulků činil 14 653. Já jsem pak data importovala do excelové tabulky, ve které jsem každému titulku přiřazovala atribut **zjevně bulvární**, **spíše bulvární**, **neutrální**, **spíše seriózní** nebo **zjevně seriózní**. Poslední dva jmenované přívlasky titulků se v našem vzorku vyskytovaly velmi výjimečně (2,55 %), do tabulky jsme je ale přesto zařadili. Tabulku jsem vypracovala tak, aby bylo možné filtrovat záznamy dle daných kritérií. Při následném doplňování označovaných částí (na korpusu pracovali současně tři anotátoři, proto bylo nutné dílčí výsledky postupně ukládat do zvláštního dokumentu) však tabulka nezvládala množství dat, které do ní bylo vkládáno, a při ukládání došlo k chybě, která způsobila odstranění veškerého formátování včetně filtru. Bohužel se jednalo o nevratnou akci, program ani po opětovném pokusu o naformátování nezpracoval takový objem dat, jaký tabulka obsahovala. Filtrování vyřešil poté Marek Grác jiným způsobem.

Při porovnání výsledků analýzy jsme kromě toho, že se potvrdil náš předpoklad (že typicky bulvární deníky budou mít zastoupení bulvárních titulků více než 80 %), přišli na pozoruhodný fakt. Deníky, které jsou obecně považovány za mainstreamové až seriózní, měly totiž v některých případech mnohonásobně větší zastoupení bulvárních titulků (dle názoru anotátorů) než titulků neutrálních. Domníváme se, že chyba vznikla zařazením pravidla „pokud titulek začíná jménem následovaným bezprostředně dvojtečkou, jedná se o bulvární titulek“. Tato forma psaní titulků (v internetových verzích deníků, nikoli v tištěné podobě) se totiž hojně rozšířila i mimo oblast bulváru. Např. na serveru *e15.cz* (ekonomické zprávy) je nadpoloviční většina titulků psána

výše zmíněným způsobem. Tím si vysvětlujeme fakt, že zastoupení bulvárních titulků na daném serveru přesáhlo 60 %. Nicméně důležitou roli při vyhodnocování „chybovosti“ anotátorů hrála nemožnost objektivního zhodnocení míry bulvárnosti.

Posoudit, nakolik je nebo není titulek bulvární, je záležitostí velmi subjektivního úhlu pohledu. Při tvorbě seznamu znaků bulvárních titulků jsem se proto snažila řídit těmi nejviditelnějšími a nejsnáze rozpoznatelnými znaky. Z nich jsem poté vytvořila krátký manuál pro další dva anotátory, kteří se značkováním pomáhali. Tento postup je nestandardní a v další fázi práce se pokusíme uchopit značkování jinak. Budeme se snažit získat mnohem více anotátorů, kterým bychom dali pouze obecné zadání. Jejich úkolem by bylo značkovat dle vlastního úsudku. Nejvýraznější rysy, na základě kterých posuzovali bulvárnost, bychom poté vyabstrahovali až zpětně. Vytvoření seznamu jedním člověkem totiž není zcela objektivní. Domníváme se, že právě kvůli této subjektivitě vznikly některé odchylky a chyby. Dalším faktorem, který mohl negativně ovlivnit anotování, byl fakt, že při značkování měl člověk-analyzátor k dispozici pouze celé znění titulku. Často bylo tedy nutné pročíst titulek několikrát, občas bylo dokonce nezbytné vyhledat původní článek k danému titulků, zběžně ho pročíst, prohlédnout fotky nebo alespoň projít úvodní část těsně pod nadpisem. Tento postup celý proces značkování značně zdržoval, nabízí se tedy otázka, zda by nebylo možné práci nějakým způsobem ulehčit a zefektivnit. Jednou z možností, jak usnadnit rozhodování při značkování, je vytvořit značkovací program, ve kterém by se anotátorovi zobrazoval titulek i s úryvkem článku. Rozhodli jsme se tento postup použít v další fázi naší práce, a začít tak se značkováním od začátku (získáme tak další data, která budeme moci porovnat s těmi, která už máme). Rozhraní by vypadalo tak, že v hlavičce by byl titulek, pod ním menším písmem úvodní a závěrečná část článku a v patičce „klikací“ políčka (**zjevně bulvární, spíše bulvární, neutrální, spíše seriózní, zjevně seriózní**). Po označení by anotátor přešel pomocí šipky na další titulek. Další možností je pak obměna tohoto systému zobrazování titulků, a to přidáním fotografií. Nevýhodou obou postupů je dle mého názoru nemožnost celkového pohledu na seznam titulků tak, jak to umožňuje tabulka. V navrhovaném programu by sice byla možnost přejít pomocí posuvníku nebo zadáním čísla na požadovaný titulek, anotátor by však nikdy značkováná data neviděl jako celek. Navíc by nastal problém, pokud by neznal přesné číslo titulku, a musel by jej tedy hledat ručním procházením části seznamu, ve které se dle jeho názoru titulek vyskytoval, což by mu zabralo spoustu času. Situace, kdy je potřeba takto vyhledávat, není častá, ale v momentě, kdy je nutné vrátit se k označovanému titulků z důvodu upřesnění nebo ověření správnosti anotátora úsudku, je tento způsob nevhodný.

Vyozorované znaky

Jedním z nejvýraznějších rysů byl přehnaný výskyt interpunkce, zvláště pak vykřičníků a otazníků. Ve snaze zdůraznit obsah titulku a upoutat pozornost potenciálního čtenáře vkládají autoři vykřičník za každou dílčí větu titulku. Efektivita „lákání“ se však tímto nadužíváním nezvyšuje, neboť pravidelní čtenáři přestávají původní funkci výše zmíněných interpunkčních znamének vnímat. Tento rys bulvárních titulků je velmi dobře zpracovatelný počítačem, neboť se jedná o počítání zadaných znaků bez ohledu na kontext. Díky možnosti využít označovaný korpus, ve kterém má interpunkce svou značku (svůj tag), byla aplikace tohoto pravidla snadná. Do tohoto pravidla jsme zařadili i častý formát bulvárního titulku, který se postupně dostává i do „nebulvárních“ tiskovin. Jedná se o formu **Jméno: text!** (*Partyšová: Krade ti manžela!*). Analyzátor tedy hledá současný výskyt vlastního jména, dvojtečky a vykřičníku v jednom nadpisu.

Dalším výrazným znakem bylo užívání příjmení (celebrit, veřejně známých osob, politiků apod.) bez uvedení křestního jména (*Dopita vyvázl u soudu bez trestu!*). V současnosti je možné sledovat průnik tohoto prvku i do serióznějších novinových serverů (potažmo tištěných médií neutrálního

rázu). Často je místo jména uveden titul, funkce nebo pracovní pozice člověka, o kterém se v daném článku píše (*Český premiér navštívil vojáky v Afghánistánu: Bez helmy a vesty to fakt nejde*). Toto pravidlo jsme také předložili programu.

Pravděpodobně „nejbulvárnějším“ jevem bylo užívání deminutiv a domácích podob křestních jmen. V titulcích ostatních tištěných (resp. internetových) médií se tato tendence nevyskytuje (zjištěno na základě pozorování). Aby se čtenář naplno ztotožnil s titulkem, resp. s obsahem daného článku, používají novináři zdobněliny nebo domácí podoby jmen ve velké míře. Čtenář se poté setkává s informacemi, že *Lucka vyrazila na turné!*, *Ivetka se opila na večírku!* a *Ondra zavzpomínal na tátul!*, a je tak bez svého vědomí vtažen „do děje“. Jeho podvědomí ho poté nutí přečíst kromě titulku i celý článek, neboť má pocit, že zmiňovaná osoba je mu blízká a že ji dobře zná.

Bulvární titulky jsou charakteristické zamlčováním a překrucováním informací. Snaha publikovat za každou cenu senzaci, a tak ve čtenáři vzbudit zájem, graduje mnohdy až do absurdních situací. Bulvární noviny pak např. publikují články s šokujícím titulkem *Jan Šťastný: Zítřka spáchá v Ordinaci sebevraždu!*, které však nepíší o tom, že by se daný herec chtěl zabít u doktora, jak titulek napovídá, ale o sebevraždě seriálové postavy, kterou herec hraje. Lidé jsou tak záměrně a zcela promyšleně uváděni v omyl jen proto, aby vydavatelé získali co nejvíce senzacechtivých odběratelů.

O tom, že čtenáři se rádi ztotožňují s těmi, o kterých noviny píšou, svědčí i další významný znak bulvárních titulků, kterým je oslovení (*Agáto, už zase?*). Pokud autor článku v titulku osloví celebrity (nebo jinou veřejně známou osobu) jménem, čtenář podvědomě s osobou sympatizuje (nebo naopak), neboť opět pocituje jakousi „známost“ či „kamarádství“.

Užívání „jednoduššího“ jazyka v bulvárních článcích (titulcích) má přiblížit jejich obsah co nejširšímu spektru publika. Proto autoři používají expresivní a hovorové výrazy (*Prcek Révai dostal super kšeft: Chtěl ve StarDance tančit, místo toho bude porotcem*).

Bulvární titulek, aby zaujal, musí silně působit na city, což mohou mimo jiné zajistit citoslovce a částice. Titulek s výrazem „fůůůj“, „ehm“ nebo „pání“ je pak relativně častý (*Vegetariánka Banášová: Fůůůj, maso!*).

Kromě výše zmiňovaných expresiv a výrazů hovorové češtiny se můžeme setkat i s obecnou češtinou (*Petr Kolář: Jakej kolaps? Jsem jen přepracovanej...*). Nejčastěji jsou to nespisovné tvary sloves (piju) a přídavných jmen (přepracovanej).

Výpustka uvnitř titulku může být samozřejmě součástí i jiných než jen bulvárních titulků, nejčastěji se ale vyskytuje právě v nich (*PROSTŘENO: David šetří na bouřlivý sex... marně!*).

Používání deminutiv se netýká pouze titulků obsahujících vlastní jména, o kterých jsem se zmínila výše. Zdobněliny jsou častým jevem, se kterým se v bulváru setkáváme (*Hodně drahý zedeček. J.Lo vyměnili, milion dostane Maria Carey*). Opět navozují pocit domácího a vlídného prostředí, které má potenciálního čtenáře přinutit číst víc než jen titulek.

S trochou nadsázky lze říci, že bulvární tisk je světem paradoxů. Na jedné straně květnatě komentuje každý krok celebrity, na straně druhé šetří zdánlivě zbytečnými slovy. Nejvíce je to patrné v titulcích a článcích, které obsahují peněžní částku nebo informaci o čase (*Hezounek Šimůnek (34): Ta žena mě podvedla! Zmizelo 17 milionů!*). V těch totiž novináři vynechávají měnu (pokud jde o koruny) a výraz hodiny (a jeho pádové formy).

Výskyt neobvyklých metafor či metonymií byl poměrně častý, což opět souvisí s jednodušším jazykem bulváru a slovní úsporou. Zajímavá a nevšední spojení navíc přitahují pozornost a vzbuzují zvědavost (*První prohibiční večírek slavných: Trhli rekord v šampaňském!*).

Zaujmout se autoři snaží opravdu vším, a to jak volbou slov, tak vizuálními efekty. Zvýrazňování slov pomocí verzálek je pak relativně časté (*EXKLUZIVNĚ! Dara Rolins a Rytmus: Tady si staví hnízdečko!*). Zajímavostí je, že typicky jsou takto zdůrazňována slova týkající se peněz (úspora, daně, ceny apod.) a výrazy typu šok, exkluzivní, video, foto apod.

V některých případech se bulvární novináři nebojí použít ani extrémně vulgární slovo. V rámci cenzury jsou ale nuceni je psát s hvězdičkami místo klíčových písmen daného výrazu (*Ivetu šu*ám já, ty si můžeš leda vyh*nit, řval Rychtář...*).

Aby bulvární deníky zaručily svou exkluzivitu a důvěryhodnost, publikují články s titulky, které jsou psány v první osobě množného čísla (*Smrt tatínka Ivany Gottové: Víme proč zemřel v tak nízkém věku!* – zde je dokonce i gramatická chyba, a to chybějící čárka). Potenciálního čtenáře se tak snaží navnadit přesvědčováním – oni (novináři) totiž všechno vědí, mají exkluzivní informace, natočili unikátní video, viděli, vyzpovídali apod.

Slovo skandál a jeho odvozeniny se v bulvárních titulcích vyskytují téměř vždy (*Skandální odhalení Terminátorovy tchyně: Arnold je impotent!*).

V předchozích odstavcích bylo zmíněno, že bulvární titulky obsahují často pouze křestní jména (nebo jejich zdvojnásobení a domácké tvary) celebrit a jiných veřejně známých osob. Pokud se však v článku píše o někom neznámém („člověk z lidu“), uvedení jména je zbytečné, ba dokonce nežádoucí. Čtenář bulvárního deníku je totiž zvyklý na známá jména, která ho poté nalákají k přečtení článku. V případě, že jde o neznámou osobu, autoři v titulku jméno nepoužívají, dokonce se uchylují k tomu, že uvedou pouze osobní zájmeno místo slovních spojení typu „mladý muž“, „neznámý chlapec“ apod., která jsou typická spíše pro tisk serióznějšího rázu (*Usnul na odbavovacím pásu: Vyfotil ho rentgen!* – čtenář netuší, o koho jde, proto je větší pravděpodobnost, že si související článek přečte).

Již několikrát zmiňované vynechávání jedné části jména (příjmení nebo křestního jména) se znovu objevilo v dalším vyzozorovaném jevu. Tentokrát se jednalo o náhradu křestního jména titulem, pracovní pozicí apod. (např. premiér, zpěvačka, herec apod. – *Premiér Nečas v Afghánistánu: Hrozil mu atentát!*).

Nelze opomenout zmínit i fakt, že v bulvárních titulcích a článcích (stejně tak i v seriózním tisku) není pravopisná nebo gramatická chyba ničím neobvyklým (*Smrt na schodišti! Varování: Zabily jí jehlové podpadky!*).

O nahrazování jmen osobními zájmeny již zmínka padla. Kromě nich se v titulcích často objevovala i zájmena ukazovací, která ve většině případů odkazovala na přiloženou fotografii nebo související video (*Tahle ohnivá koule měla dobýt Měsíc!*). Novinář tak nutí čtenáře, aby zhlédl fotografie a případně přečetl daný článek.

V některých případech je čtenář nucen rozklíčovat jméno celebrity (nebo jiné veřejně známé osoby) jen podle nepřímého popisu typu *černá perla českých přehlídkových mol* apod.

Znakem, který je pro bulvár typický, je chybějící sloveso ve větěném celku (*Iveta Bartošová strachy bez sebe. Bojí se vidět vlastní smrt*). Domnívám se, že i v tomto případě se jedná o úsporu slov. Nejčastěji je vynecháváno sloveso být (ve všech tvarech).

O bulvárnosti titulku nás mimo jiné může přesvědčit i nemalý výskyt slov vyjadřujících pochybnost, jakými jsou výrazy údajně, prý apod. (*Kevin ze »Sám doma«: Zbývá mu prý šest měsíců života...*).

Pokud se v titulku objeví tázací zájmeno cože (případně proč atd.) jako samostatná věta, jde s největší pravděpodobností opět o titulek bulvárního rázu (*Cože? Justin Bieber měl strach z prvního polibku se Selenou Gomez*).

Často se novinářům stává, že ve spěchu na sebe nevhodně navážou větné celky (*Johnny Depp se bojí o maminku. Je na jednotce intenzivní péče*). Zde nemusí být čtenáři jasné, koho se druhá část souvětí týká.

Chyb způsobených psaním v časovém nátlaku není málo, příkladem mohou být např. nevhodně zvolená slovní spojení (*Některé otázky opravných státních maturit z matematiky vzbuzují otazníky*). Občas je v titulcích (nejen bulvárních) užitá zkratka **a spol.** V případě, že tato zkratka figuruje za vlastním jménem (a není tedy součástí názvu společnosti), považují daný titulek za bulvární (*Svědek v kauze Řebíček a spol. očerňoval poškozeného, jak to šlo*).

Noviny i časopisy často v titulcích (i článcích) používají svůj název (*Klaus pro Novinky: Barrosův plán vytvořit z Evropy federaci odmítám*). Tento jev sice není typickým pouze pro bulvár, ale pokud

známe tituly bulvárních deníků, můžeme jejich titulky tohoto typu jednoduše filtrovat. Některé z rysů, které jsou výše, jsou na počítačové zpracování velmi složité a abstraktní. Jedná se zejména o rysy z oblasti sémantiky a kontextu. Vývoj nástrojů na sémantickou analýzu a rozpoznání kontextu je zatím na samém počátku a jeho výsledky nejsou pro kvalitní analýzu dostačující. Proto jsme raději tyto znaky do strojového učení vůbec nezahrnuli. Seznam jsme tedy upravili tak, aby všechna pravidla byla snadno zpracovatelná některým z nástrojů pro morfologickou analýzu, desambiguaci (tzv. zjednoznačnění významu) apod. Pro rozpoznání morfologických znaků, na které se vztahovala některá pravidla, bylo nutné využít morfologický analyzátor Ajka a desambiguační nástroj DESAMB. Marek Grác navrhl na základě mnou vytvořených pravidel jednoduché programy, které jednotlivé znaky (vypozorované rysy) testovaly. V dílčích částech se nacházel jak příkaz k porovnání, tak příklady včetně vyhodnocení pravdivosti. Každý z navržených programů jsem prošla, otestovala a navrhla jeho zlepšení; v případě, že nešlo o složitou operaci, upravila jsem jej sama. Jednalo se hlavně o doplnění testovacích příkladů a rozšíření podmínek testu.

Vyhodnocení míry bulvárnosti titulku

Metodu SVM jsme využili proto, abychom získali data potřebná k vytvoření vektoru, na jehož základě by mohlo automaticky proběhnout rozhodnutí o tom, zda příslušný titulek je nebo není bulvární. Při samotném procesu strojového učení byl seznam titulků rozdělen na segmenty po 10 titulcích, z nichž 9 bylo použito jako data pro učení a poslední sloužil pro testování výsledků učení. Výsledkem druhé fáze strojového učení je tedy posloupnost číslíc – vektor, jenž podává informaci o tom, který z daných jevů se v titulku vyskytuje, resp. zda je výsledek testu daného rysu negativní nebo pozitivní. Počet pozitivních výsledků pak rozhodne o míře bulvárnosti titulku.

Závěr

Výsledky testování nebyly v době konání konference ve stavu, ve kterém by mohly být prezentovány. V práci v současné době stále pokračujeme, snažíme se získat nová, kvalitnější data a zdokonalovat testovací programy, čímž bychom chtěli zvýšit úspěšnost programu. Pracujeme na novém způsobu značkování a přemýšlíme nad lepšími postupy. Dosavadní práce se tak může zdát jako zbytečná, ale pro nás má velký přínos. Poskytla nám potřebný odrazový můstek pro budoucí hlubší analýzu a kvalitnější zpracování získaných dat. Doufáme, že naplníme prvotní cíle a vytvoříme funkční algoritmus sloužící ke kvalitnímu filtrování dat na základě poskytnutých podmínek a pravidel.

Použitá literatura:

- BUCHTELOVÁ, R. (2009): *Nový akademický slovník cizích slov A–Ž*. Praha: Academia.
- FILIPEC, J. – DANEŠ, F. – MEJSTRÍK, V. (2005): *Slovník spisovné češtiny pro školu a veřejnost: s Dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky*. Praha: Academia.
- GILES, D. (2012): *Psychologie médií*. Praha: Grada Publishing.
- McQUAIL, D. (1999): *Úvod do teorie masové komunikace*. Praha: Portál
- MIČIENKA, M. – JIRÁK, J. (2007): *Základy mediální výchovy*. Praha: Portál.
- MUSIL, J. (2007): *Komunikace v informační společnosti*. Praha: Univerzita Jana Amose Komenského Praha.
- SEDLÁČEK, R. (1991): *Morfologický analyzátor češtiny* [online]. Brno: Laboratoř NLP FI MU. Cit. 2013-6-4. <<http://nlp.fi.muni.cz/projekty/ajka/ajka.pdf>>
- STERN, J. (2006): *Média, psychoanalýza a jiné perverze*. Praha: Malvern.
- ŠMERK, P. (2007): *K morfologické desambiguaci češtiny* [online]. Brno: Laboratoř NLP FI MU. Cit. 2013-6-4. <https://is.muni.cz/auth/th/3880/fi_r/teze.pdf>
- ŽIŽKA, J. (2005): *Support vector machines (SVM)* [online]. Brno: Laboratoř NLP FI MU. Cit. 2013-6-4. <http://is.muni.cz/el/1433/podzim2006/PA034/09_SVM.pdf>

KORPUS A JMENNÝ ROD VE VÝUCE ČEŠTINY JAKO CIZÍHO JAZYKA

BARBORA KOČAŘOVÁ

Ústav českého jazyka FF MU Brno

Abstract

Over the past two decades we are confronted with progress of the corpus linguistic. But corpora have not been used only in language research, also in different fields of interest such as teaching and learning foreign languages. Concept of using corpora in teaching is called Data Driven Learning and has two different forms. First one, known as direct method, can be accessed by language learners in the classroom via corpus-derived exercises and their interaction. Second form, known as indirect method, help to teachers mainly with decisions about what to teach and has effects on reference works and development of teaching materials. In this article I wish to demonstrate the usage of second method on the elemental grammatical problem of Czech language such as classification of nouns according to their grammatical gender. Grammatical gender as a basic rule which is presented in opening lessons, but textbooks used for the same level (level A1 of the Common European Framework for Languages) do not follow the same pattern of presentation, neither the different description are given. Corpus based analysis can verify this grammatical principle, after that some tips and recommendation could be given. Nowadays, Czech as a foreign language is linguistic discipline which is very well characteristic for its practical progress.

Klíčová slova: čeština jako cizí jazyk, jmenný rod, prezentace gramatiky, korpus ve výuce cizích jazyků

Keywords: Czech as a foreign language, text corpus, grammatical gender, grammar, Data Driven Learning

Využití korpusu ve výuce cizích jazyků

V dnešních dnech jsou jazykové korpusy stále více a více využívány pro výzkum současné podoby jazyka a jeho proměn. Korpusová data však nemusejí být užitečná pouze pro lingvistiku, ale uplatnění najdou také v oblasti jazykového vzdělávání. Ani zde není korpusový přístup považován za závratnou novinku, pod názvem DDL (tedy anglicky *Data Driven Learning*)¹ obohacuje vyučovací proces již více než dvě dekády. Obecně bychom zmíněný přístup mohli definovat jako způsob výuky skrze cvičení a aktivity na korpusu založené, ale také jako způsob výuky na základě korpusových dat. Takto orientované vyučování tedy představují dvě základní podoby.

První, označovaná jako přímá, se zaměřuje na studenty a na jejich aktivity s korpusy. Ty umožňují studentům sledovat jazyková data v jejich autentickém prostředí, tedy taková data, která nejsou zpracovaná čistě pro didaktické účely. S korpusem mohou pracovat jednak přímo ve výuce, ale díky veřejnému přístupu většiny korpusů i individuálně bez dohledu lektora, a tak doplňovat své studium cizího jazyka o novou perspektivu.

Druhá podoba, označovaná jako nepřímá, se zaměřuje na opačnou stranu výuky, a to konkrétně na učitele a lektory, kterým korpus pomáhá odpovědět na otázku co učit. V takovém případě korpus umožňuje získat informace o kolokacích, významech a užití slov, ale především se učitelé

1 O dalších možnostech užití korpusu ve vyučovací praxi viz práce lingvistek Ute Römer, Susan Hunston, Lynn Flowerdew a dalších.

dostává do rukou nástroj pro ověření platnosti gramatických pravidel. Lektor následně poznatky z korpusu může využít pro doplnění vlastních znalostí o gramatice vyučovaného jazyka nebo přímo pro tvorbu vlastních výukových materiálů a cvičení, jež velmi často doplňují učebnice ve výuce. Tento příspěvek je orientován právě na výše popsany nepřímý způsob užití korpusu. Možnosti jeho praktického využití demonstrujeme na konkrétním problému jmenného rodu v češtině. Čeština jako cizí jazyk představuje odvětví bohemistiky v dnešních dnech rychle se rozvíjející, což dokládá především stále se rozšiřující nabídka jazykových kurzů a publikování nových studijních materiálů.

Ten, ta nebo to?

I my, jako rodilí mluvčí, máme občas problém určit rodovou příslušnost slov, což se týká především vlastních jmen (např. *Bystrc, Bedihošť*). O to těžší je to pro studenty češtiny jako cizího jazyka, kteří přicházejí se znalostmi jazyka nebo jazyků, které touto gramatickou kategorií vůbec nedisponují. Jedná se o základní pravidlo, s nímž jsou navíc konfrontováni již v úvodních lekcích. Gramatická kategorie jmenného rodu představuje základní jazykové paradigma, které má velký vliv na další gramatické kategorie (shoda substantiva s adjektivem, participia, jmenné skloňování aj.). S přirozeným rodem, kde se můžeme opřít i o sémantiku slov, není problém. Ovšem u neživotných substantiv, které v češtině rovněž disponují rodem, ztratil rod svou významovou složku, a stal se především základem pro organizaci systému české deklinace i slovní zásoby. Orientačním vodítkem se tak stává podoba slova, tedy příslušná koncovka.

Studenti češtiny jako cizího jazyka jsou s touto skutečností konfrontováni již v úvodních lekcích, ale i učebnice určené pro stejnou úroveň (úroveň A1 podle Společného evropského referenčního rámce pro jazyky) se ve zpracování tohoto pravidla podstatně liší. Tento gramatický jev nepojímají jednotně a taktéž se neshodují v prezentaci jednotlivých koncovek podstatných jmen v prvním pádě jednotného čísla, čímž se může distribuace jevit jako nesystematická. Některé materiály zcela vynechávají samohlásku *-a/-e* jako možné zakončení mužského rodu životného, tj. vzor *předseda, soudce*, nebo konsonant jako možné zakončení substantiv ženského rodu. V některých dalších případech postupují dokonce protichůdně. Učebnice *Do you want to speak Czech?* (2005, s. 31) charakterizuje vokál *-e* jako velmi časté zakončení ženského rodu, naopak v učebnici *New Czech step by step* (2006, s. 25) je ve stejném případě vokál zařazen do kategorie minoritního zakončení. Podobně vágně se ovšem vyjadřují i dostupné jazykové příručky určené pro rodilé mluvčí. Podle *Příruční mluvnice* (2003, s. 233) je například zakončení *-a, -e/-ě* typické pro feminina (a neutra) kromě typů *píseň, kost* a typů adjektivních *pokojská, vedoucí*. Podobně nespécificky se vyjadřuje i *Mluvnice češtiny* (1986, s. 39), a to pomocí výrazů *převážně, řidčeji, častěji a popřípadě*.

Korpusová analýza

Korpus a jazyková data v něm obsažená nám poskytují vodítko, kterým právě tato nejasná tvrzení můžeme podložit přesnými výsledky. Cílem analýzy bylo vyhledat substantiva v nominativu singuláru² z celkového objemu textových slov v korpusu, dále jejich rodovou příslušnost. Z těchto obecných výsledků byla poté konkrétně spočítána frekvence slov s daným koncovým sufixem a jejich bližší charakteristiky. Pro vyhledávání jsme zvolili SYN2010 jako nejaktuálnější korpus současné psané češtiny. Jako nástroj pro vyhledávání jsme použili korpusový manažer Bonito a jeho funkce, s jehož pomocí lze vyhledat slova dle zadaných kritérií.

2 Při prezentaci tohoto základního jevu českého jazykového systému cizincům je třeba se zabírat právě a pouze nominativem, protože začátečníci jsou se jmennou deklinací seznamováni různě a na odlišných úrovních.

Distribuce koncovek maskulin životných je podle korpusových dat velice jasná a přehledná. Jiným než konsonantickým zakončením disponuje přibližně pouze dvacet procent výskytů. Z vokalickej skupiny je nejvíce zastoupeno *-a* (162 506 výskytů) a *-e* (80 913 výskytů). Slova takto zakončená mají podobnou charakteristiku, týkají se především pracovních pozic a podobných veřejných funkcí. Mezi slova s nevyšší frekvencí patří například *předseda, soudce, starosta, policista* a *zástupce*. Překvapivě nacházíme vokalickej zakončení *i* u mužského rodu neživotného, kde se objevují názvy mezinárodních institucí, programů a značek jako například *software, Apple* nebo *T-Mobile*. Jejich zařazení je motivováno foneticky (srov. *Apple* a [*ˈæpəl*]), a proto tuto skupinu neshledáváme problematickou. Užití variant *chléb/chleba* korpus SYN2010 nereflektuje.

Podle očekávání se větší nesrovnalosti týkají feminin a neuter, kde procentuální zastoupení koncovek není tak jednoznačné jako u maskulin. U feminin dominuje zakončení s vokálem *-a* v téměř 61 procentech výskytů, zbylá procenta poté reprezentuje poměrně vyrovnaně vokál *-e/-ě* (17,7 %), konsonant (přibližně 15 %), v nejmenší míře pak substantiva s adjektivní podobou. Na vysoké frekvenci slov zakončených na *-e/-ě* participují především přechýlené podoby na základě přípon *-ce/-ice/-yně*, a poté slova přejatá zakončená na *-ie*. Mimo ně jsou to jednak slova přejatá (*prezentace, absence*), tak i slova bez této charakteristiky jako *svíce, země, snídaně* aj. Vysoký výskyt konsonantu u feminin zapřičiňují především toponyma zakončená na různé konsonanty, např. *Chotěboř, Budapešť*. Za další se na tom podílí i skutečnost, že součástí korpusu jsou vlastní jména cizího původu například *Sarah, Ellen* atd., proto využijeme funkci manažeru Bonito na odstranění slov s velkými písmeny, následně tak zjistíme, že nejfrekventovanější konsonant u feminin je *-t*, a to díky podstatným jménům reprezentujícím abstraktní pojmy a názvy vlastností zakončených na *-ost* typu *radost, přítomnost, záležitost*.

Jako základní sufix neuter je často v literatuře uváděno *-o*. Současný průzkum v korpusu však ukázal, že tomu tak není, neboť sufix *-í* zabírá více než polovinu výskytů (379 927), a to díky verbálním substantivům zakončených pomocí *-ní* a *-tí* (*spolehnutí, zranění, zaklepání*). Na další pozici je to tedy vokál *-o*, dále pak koncovka *-e/-ě* (55 875 výskytů), *-um* (24 315).

Pedagogická aplikace

Korpusový výzkum potvrdil, že distribuce koncovek je poněkud komplikovanou otázkou a je nutné se k ní opakovaně vracet, což většina studijních materiálů důsledně nedělá. Učební materiály tak podporují trend, o němž hovoří Dana Hánková (2010, s. 41). Podle ní učebnice češtiny pro cizince mají tendence se zaměřovat na jeden konkrétní jazykový jev a na jeho formu. V menší míře se pak zaměřují na jeho význam a kontext, v jakém je jev užíván. V praxi to často vypadá tak, že v rámci jedné kapitoly je prezentován pouze jeden jazykový jev, ve vytvořených cvičeních je sice důkladně procvičen, ale ne v návaznosti na další jevy či kontexty užití. Tvoří tak uzavřenou kapitolu, což může evokovat jeho omezené užití. Zároveň i jmenný rod by neměl být obsahem jedné lekce, a to obvykle první nebo druhé, ale měl by být dále systematicky rozvíjen a prezentován v jiných jazykových situacích. S takovými příklady pracuje studijní materiál *Český den*, kde mají studenti za úkol vytvořit z dané podoby instrumentálu (*za poštou, nad městem, před restaurací* aj.) nominativ a určit jeho rodovou příslušnost. Následně v dalším cvičení musí tyto změny slovně popsat a vytvořit tak vlastní popis pravidla (2007, s. 37).

Současné učebnice také tíhnou k rozvíjení jazykových dovedností skrze tematicky jednotné bloky, proto i my můžeme slovní zásobu rozdělit na základě poznatků získaných z vlastní korpusové analýzy. Například v rámci tématu *Orientace ve městě* by nejprve byly prezentovány základní typy či vzory na příkladech jako *škola, kino, autobus* aj. V jeho pokračování, nazvaném jako *Město a jeho okolí*, poté rozšířený soubor koncovek skrze slova *parkoviště, centrum, kadeřnictví, květinářství* pro neutra, *restaurace, kavárna, cukrárna, drogerie* pro feminina. V tomto případě

se vyhneme poznámkám, že slova přejatá mohou u neuter mít podobu s *-um*, které tyto výjimky často doprovázejí, a docílíme, že student je schopen v komunikačním kontextu správného užití bez hlubší teoretické znalosti problematiky.

Korpusové třídění rovněž potvrdilo silné spojení mezi jmenným rodem a slovo tvorbou, proto je možné vybrané úseky slovní zásoby popsat a upravit pro didaktické účely. Například tak, že jednotlivé základní koncovky rozšíříme, pracovní je pojmenujeme a doplníme jimi schémata uvedená v učebnicích. Uvedeme pouze několik možností: *-ě* u neuter lze rozšířit na *-iště* pro názvy specifických míst jako *hřiště, parkoviště, hlediště*, dále například vokál *-a* můžeme u maskulin životných rozšířit na *-ista*, jenž je sufixem charakteristickým pro jména přejatá, a který zastupuje skupinu protagonistů určitého sportovního odvětví (*basketbalista*), reprezentanty určitého uskupení/hnutí (*komunista*), hráče na hudební nástroj (*klavírista*), představitele životního stylu či názoru (*optimista*).

Závěr

Příspěvek ukázal, jakým směrem se dnes může ubírat výzkum spisovného jazyka v ČNK za účelem jeho zpracování pro nerodilé mluvčí a taktéž pro pedagogické účely uplatnitelné v rámci výuky češtiny pro cizince. Korpus v těchto případech umožňuje lépe vybrat vhodnou slovní zásobu, především však napomáhá doplnit údaje v dostupných gramatikách. S výsledky korpusové analýzy není možné pracovat pouze z odborného lingvistického hlediska, musí se také zohledňovat soudobé lingvodidaktické teorie. Před pedagogem tak vždy stojí problém interpretace a otázka praktického využití, taktéž musí brát v potaz charakter studijní skupiny, jejich znalosti a zájmy. Je tedy pouze na něm samotném, jak se rozhodne tyto poznatky uvést do svých hodin.

Použitá literatura

- ČECHOVÁ, E. – REMEDIOSOVÁ, H. (2005): *Do you want to speak Czech?* Liberec: Harry Putz.
- GREPL, M. – KARLÍK, P. – NEKULA, M. – RUSÍNOVÁ, Z. (2008): *Příruční mluvnice češtiny*. 2. vyd. Praha: Nakladatelství Lidové noviny.
- HÁNKOVÁ, D. (2010): Gramatika jako proces. In: *Sborník asociace učitelů češtiny jako cizího jazyka*. Praha: Akropolis ve spolupráci s Asociací učitelů češtiny jako cizího jazyka, s. 33–50.
- HOLÁ, L. (2006): *New Czech Step by Step*. Praha: Akropolis.
- KOMÁREK, M. (1986): *Mluvnice češtiny 2. Tvarosloví*. 1. vyd. Praha: Academia.
- MATULA, O (2007): *Český den. Kurz českého jazyka pro azylanty navazující na Manuál pro učitele českého jazyka pro cizince bez znalosti latinky*. Praha: Člověk v tísni o. p. s., Projekt Varianty.

K OBTÍŽNÉMU SHÁNĚNÍ (KORPUSOVÝCH DAT PRO FONETICKÝ VÝZKUM)

VERONIKA ŠTĚPÁNOVÁ

oddělení jazykové kultury, Ústav pro jazyk český AV ČR, v. v. i.
Letenská 4/123, Praha 1

Abstract

On the difficult gaining of the corpus data for the phonetic research

The paper presents the corpus Monolog, its applications and also its possibilities and restrictions. This corpus was built as a source for analyses of pronunciation in Czech mass media and its mission is to fill partly the gap in the field of the Czech spoken corpora. The paper also shows the results of two analyses concerning the standard pronunciation of Czech:

1) the occurrence of glottal stops in sequences of non-syllabic prepositions + words with vowels in the initial position; 2) *the pronunciation of consonant groups sh (i.e. in words like shánění, shoda).*

Klíčová slova: výslovnost, čeština, korpus mluveného jazyka, fonetický výzkum

Key words: pronunciation, Czech language, corpus of spoken language, phonetic research

1 Úvod

V současnosti jsou v českém prostředí k dispozici korpusové zdroje zaměřené nejen na psaný, ale rovněž na soudobý mluvený jazyk. Pro tu část lingvistů, kteří zkoumají mluvenou spisovnou/noremní/oficiální komunikaci z fonetického hlediska, je však přístup k potřebnému množství dat (včetně zdrojových nahrávek) dosud značně obtížný. Tento typ diskurzu lze (pokud jde o české korpusové zdroje) analyzovat především prostřednictvím korpusů Dialog a Monolog. Jejich rozsah však nedovoluje realizovat veškeré typy potřebných sond. Tento nedostatek je pocíťován např. při tvorbě nových slovníků češtiny, v nichž má být uvedena i výslovnost jednotlivých lexikálních jednotek.

2 Dnešní možnosti výzkumu mluvené češtiny

2.1 Korpus Monolog

V tomto příspěvku¹ bychom chtěli především stručně představit nově vzniklý korpus Monolog, nastínit jeho možnosti a omezení a popsat práci s ním. Analýza dat z tohoto zdroje představuje jeden z možných postupů, jak získávat informace o dnešním úzu a normě mluvené češtiny.

Korpus Monolog obsahuje monologické nahrávky profesionálních mluvčích veřejnoprávního Českého rozhlasu a ortografické, tedy nikoli fonetické přepisy těchto nahrávek. Jedná se o projevy připravené, které jsou většinou čtené. Materiál byl získáván ze zvukového archivu Českého rozhlasu, z něhož lze volně stahovat záznamy mnoha rozhlasových pořadů ve vysoké kvalitě. Vybráno bylo 16 mluvčích (osm žen a osm mužů), od nichž bylo k dispozici větší množství delších monologických nahrávek, které vznikly po roce 2007. Při výběru mluvčích se zohledňoval mimo jiné i regionální původ (pokud bylo možné jej zjistit), a to tak, aby byli pokud možno zastoupeni mluvčí z různých regionů celého Česka.

1 Článek vznikl v rámci grantového projektu GA ČR č. 13-00372S a s podporou RVO: 68378092.

Kvůli předpokládaným detailním fonetickým analýzám bylo důležité zvolit ty pořady, v nichž není řeč podkreslena hudbou či jinými doprovodnými zvuky. Kvůli této podmínce a kvůli požadavku monologického projevu se repertoár vhodných pořadů značně zúžil a korpus je tvořen z velké části zpravodajskými relacemi. Dostupné jsou pouze ty úseky nahrávek, v nichž mluvčí delší dobu souvisle sami hovoří, korpus Monolog tedy neobsahuje celé rozhlasové pořady. V současnosti se celková velikost korpusu blíží k 300 minutám, v nichž zaznělo skoro 40 tisíc textových slov.

U všech mluvčích se předpokládá, že se (např. v souladu s kodexem Českého rozhlasu) snaží o projev po všech stránkách spisovný a že v rámci veřejnoprávního vysílání působí jako tzv. řečové vzory. Ukázky mluvních výkonů každého z mluvčích v korpusu byly podrobeny testům typu řečový vzor. I když se kvalita mluveného projevu jednotlivých mluvčích podle posluchačského hodnocení pochopitelně mírně liší, žádný z nich nebyl považován za špatného, nebo dokonce nepřijatelného mluvčího. Nahrávky v korpusu Monolog lze tak (a to i z dalších důvodů) považovat za jeden z možných zdrojů zkoumání české noremní výslovnosti.

2.2 Práce s korpusem Monolog

Než přejdeme k ukázkám konkrétních analýz některých výslovnostních jevů, zaměříme se stručně na popis práce s korpusem Monolog. Korpus je dostupný z webové adresy *monolog.dialogy.org*, kde jsou také uveřejněny podrobné informace o něm. Pro vstup do samotného korpusu je nutné se zaregistrovat pomocí jednoduchého formuláře, v němž si zájemce zvolí uživatelské jméno a zadá svoji e-mailovou adresu. Na tu přijde heslo umožňující korpus využívat. Po přihlášení je pak možné v pravém sloupci kliknout na odkaz „vyhledávání pomocí Dialogy.Org“, případně použít cestu přes odkaz „jak hledat v korpusu“. Následně se otevře formulář, v němž lze zadat konkrétní slovo, případně řetěz slov (včetně pravého a levého kontextu). Je možné vyhledávat rovněž podle lemmatu určitého slova či na základě různých gramatických značek (tagů). Program umožní v budoucnu hledat i podle pohlaví mluvčího, druhu pořadu atp., tyto funkce zatím nejsou v provozu; k účelům, k nimž byl korpus především vybudován, nejsou ani nezbytně nutné.

Při hledání konkrétních hláskových spojení je velmi výhodné používat znak hvězdičky, který zastupuje libovolně dlouhý řetězec znaků včetně nulového, případně znak tečky, který zastupuje právě jeden znak. Po zadání hledaného výrazu se v horní části obrazovky vpravo vypíše počet nalezených výskytů a zobrazí se jednotlivé konkordanční řádky. Červeně je vyznačeno hledané slovo, na konci každého konkordančního řádku je uvedeno lemma (základní tvar slova; zvýrazněno žlutě), tag (gramatická značka) a informace o pořadu, z něhož nahrávka pochází (zkratka názvu a datum vysílání; zvýrazněno žlutě).

Na začátku každého konkordančního řádku jsou dva odkazy, na které lze kliknout. První je označený písmenem A a spouští se jím audionahrávka v místě, kde se vyskytuje nalezený text. Druhý odkaz je označen číslem repliky a slouží k zobrazení celého ortografického přepisu dané nahrávky.

2.3 Další možnosti výzkumu mluvené češtiny

Korpus Monolog byl budován především jako zdroj výzkumu výslovnostního úzu v českých médiích a se záměrem alespoň minimálně vyplnit mezeru, kterou pociťují ti, kteří se v současnosti zabývají analýzou zvukové (fonetické) roviny češtiny. Ačkoli byly v nedávné minulosti vytvořeny poměrně rozsáhlé korpusy zaměřené na současnou mluvenou češtinu (korpusy řady ORAL v Českém národním korpusu), tyto zdroje nejsou zatím postačující, zejména pokud jde o analýzy fonetických jevů. K dispozici byly totiž dosud pouze přepisy, nikoli samotné zdrojové nahrávky (stav k říjnu 2013).

Pro potřeby lingvistů, kteří zkoumají mluvenou komunikaci spisovnou/noremní/oficiální, vytvářejí výkladové slovníky atp., pak byl donedávna dostupný pouze korpus Dialog, což je speciální multimediální korpus mluvené češtiny, který slouží především k výzkumu mediální komunikace

jako celku. Shromažďuje nahrávky a přepisy diskusních pořadů českých televizí, tedy veřejných jazykových projevů dialogického typu.

Vznik korpusů vhodných pro výzkum zvukové roviny jazyka je po všech stránkách náročnější než získávání materiálu psaného. Zřejmě především proto lze v současnosti vycházet pouze z nepočtených rozsahově omezených zdrojů.

3 Analýza některých fonetických jevů v korpusu Monolog

Nyní se zaměříme na ukázkou konkrétních fonetických výzkumů, kvůli nimž korpus Monolog vznikl především, ačkoli může být samozřejmě využit i pro zkoumání jiných jazykových rovin, případně i k jiným badatelským účelům.

Zkoumali jsme dva odlišné hláskové jevy, díky nimž lze snadno ilustrovat současné možnosti a omezení tohoto korpusového zdroje. Na oba jevy jsme zaznamenali dotazy při práci v jazykové poradně Ústavu pro jazyk český, někdy přímo i od profesionálních mluvčích z českých médií. Obě fonetické analýzy byly prováděny pomocí sluchové metody. V případě, kdy jsme na základě tohoto postupu nebyli schopni rozhodnout, o jakou zvukovou podobu se jedná, jsme přistoupili ke spektrální analýze nahrávky v programu Praat.

3.1 Sekvence *neslabičná předložka + slovo začínající na samohlásku*

Nejprve se budeme zabývat realizací sekvencí *neslabičná předložka + slovo začínající na samohlásku*. Ortoepická kodifikace (Výslovnost spisovné češtiny I, 1967) v tomto případě považuje za spisovnou pouze realizaci s rázem, tj. např. spojení *k obtížnému* jako [kʔopti:žne:mu], nikoli [kopti:žne:mu] nebo [gopti:žne:mu]). V souladu s návrhem R. Skarnitzla (Palková a kol., 2004, s. 71–72) považujeme i my termín *ráz* za střechový pojem pro akusticky (i fyziologicky) odlišné glotální zvukové události, které však plní stejnou funkci; rázem tedy neoznačujeme pouze realizaci jednoznačné a výrazné laryngální okluzivy, ale také např. třepenou nebo dyšnou fonaci. V případě našeho výzkumu jsme pak jako realizaci s rázem hodnotili všechny případy, v nichž byl sluchovou metodou zaznamenán „předěl“ mezi předložkou a následujícím slovem (ať se již jednalo o laryngální okluzivu, nebo o jiný akusticky zřetelný prvek).² Za realizaci bez rázu jsme pak považovali doklady, v nichž nebyl slyšitelný žádný předěl, u těch pak i spektrální analýza potvrdila plynulý přechod mezi předložkou a následujícím vokálem.

V korpusu Monolog jsme celkem analyzovali 215 dokladů spojení *neslabičná předložka + slovo začínající na samohlásku* od 13 různých mluvčích (6 žen a 7 mužů). Z toho bylo 25 dokladů předložky *k*, 113 dokladů předložky *v*, 41 dokladů předložky *s* a 36 dokladů předložky *z*. Za nimi následoval 37× krátký vokál *a*, 1× dlouhý vokál *á*, 28× krátké *e*, 1× dlouhé *é*, 26× krátké *i*, 4× dlouhé *í*, 63× krátké *o*, 8× dlouhé *ó*, 24× krátké *u* a 23× dlouhé *ú* (viz také následující tabulka 1).

2 Zde lze podotknout, že zřetelnost realizace rázu se mezi jednotlivými mluvčími velice lišila. Proto tehdy, kdy na základě poslechové metody nebylo jisté, zda se ráz vyskytl (především v rychlé řeči), přistoupili vždy ještě k analýze spektrální.

Tab. 1: Počty dokladů jednotlivých neslabičných předložek a následujících vokálů rozlišených podle kvality i kvantity.

Segmenty	<i>a</i>	<i>á</i>	<i>e</i>	<i>é</i>	<i>i</i>	<i>í</i>	<i>o</i>	<i>ó</i>	<i>u</i>	<i>ú</i>	Celkem
<i>k</i>	2	x	1	x	4	1	10	x	5	2	25
<i>v</i>	21	1	19	1	15	3	29	3	11	10	113
<i>s</i>	9	x	3	x	5	x	15	1	2	6	41
<i>z</i>	5	x	5	x	2	x	9	4	6	5	36
Celkem	37	1	28	1	26	4	63	8	24	23	215

Ačkoli korpus neobsahuje veškeré teoreticky možné kombinace a zvláště v případě dlouhých vokálů nelze v porovnání se samohláskami krátkými z pochopitelných důvodů získat srovnatelný počet výskytů, přesto se domníváme, že výsledky tohoto výzkumu lze považovat za relevantní. V našem materiálu se totiž neprokázalo, že by konkrétní neslabičná předložka nebo určitý vokál po ní následující vykazoval vzhledem k ostatním co do (ne)výskytu rázu jakékoli odlišnosti.

Ze získaných 215 dokladů se totiž u pouhých 11 jednoznačně prokázalo, že po neslabičné předložce nenásledovala žádná akusticky zachytitelná událost, kterou by posluchač mohl vnímat jako ráz. V těchto 11 případech³ tedy došlo k plynulému přechodu mezi sledovanými segmenty (tzv. splývavé výslovnosti), kdy mluvčí zvukově nijak nesignalizoval mezislovní hranici. 11 dokladů, v nichž se nevyskytl jakýkoli slyšitelný náznak rázu, pochází od 5 různých mluvčích (4 mužů a 1 ženy). Z takového minima výskytů nelze tvrdit, že např. ženy mluví celkově pečlivěji a dodržují ortoepickou kodifikaci, avšak není bez zajímavosti, že celkově nejvíce dokladů splývavé výslovnosti (4 z celkových 19 dokladů od daného mluvčího) bylo zaznamenáno u hlasatele, který byl v testech řečových vzorů hodnocený spíše jako průměrný mluvčí a s ohledem na ostatní testované se umístil jako druhý nejhorší. Ve všech případech, kdy mluvčí zkoumanou sekvenci realizovali splývavě, se jednalo o výslovnost s neznělou podobou předložky (tj. např. spojení *k událostem* jako [kuda:lɔstem], nikoli [guda:lɔstem]; *v úvodním* jako [fu:vodňi:m], nikoli [vu:vodňi:m]). Tyto výsledky korespondují se staršími zjištěními Z. Palkové (1982, s. 191): „výslovnost bez rázu a se znělou podobou předložky (např. *vitálii*) [...] se u mluvčích pražského vysílání objevuje jen výjimečně“. K tomu je nutno dodat, že náš korpus zahrnuje i mluvčí pocházející např. z Moravy, kde bývají podoby se znělou předložkou běžné, avšak ani u nich jsme takovouto realizaci nezaznamenali. Celkově tak lze konstatovat, že mluvčí vybraní do korpusu Monolog (zvláště ti, kteří jsou posluchači hodnoceni jako dobré až vynikající řečové vzory) dodržují ve svých projevech až na nečetné výjimky v sekvencích *neslabičná předložka + slovo začínající na samohlásku* ortoepickou výslovnost.

Ve výše citovaném článku Z. Palková (1982, s. 191) zmiňuje, že splývavá výslovnost bez rázu s neznělou podobou předložky je ve zkoumaných spojeních „[č]astou chybou některých mluvčích“. Naše, na první pohled odlišné zjištění, však nemusí být s tímto tvrzením nijak v rozporu a rozhodně by nemělo vést k závěru, že se výslovnost rozhlasových mluvčích za posledních 30 let v tomto aspektu „zlepšila“. Zásadním rozdílem mezi oběma výzkumy totiž bude pravděpodobně výběr mluvčích, jejichž řeč byla zkoumána. Jak již bylo řečeno, korpus Monolog obsahuje převážně nahrávky hlasatelů/moderátorů, kteří prokazatelně uspěli v testech řečových vzorů a jejichž mluvní výkony ohodnotil relevantní vzorek české populace převážně pozitivně. Z. Palková se naopak snažila postihnout veškeré typy nedostatků vzhledem k žádoucímu standardu, s nimiž se posluchač

3 Konkrétně se jednalo o kombinace *k+o*, *3× k+u*, *3× v+e*, *v+o*, *v+ú*, *z+e*, *z+u*.

ve vysílání může setkat, a její výzkum nezahrnoval jakékoli posluchačské hodnocení individuálních mluvních výkonů. Je tak pravděpodobné, že v její zkoumané skupině se úroveň jednotlivých mluvcích značně lišila a že vynechávání rázu po neslabičných předložkách sejevilo celkově jako časté, avšak vyskytovalo se pouze u některých mluvcích (jak je patrné z výše uvedené citace).

3.1 Souhlásková skupina zapisovaná jako *sh*

Druhým zkoumaným jevem v korpusu Monolog byla výslovnost souhláskové skupiny zapisované jako *sh*. Jedná se o známý případ, kdy se na území Česka vyskytují dvě různé výslovnostní podoby této skupiny, přičemž neznělá varianta [sx] vzniklá progresivní asimilací znělosti je typická pro mluvčí z oblasti Čech a naopak znělá varianta [zh], v níž probíhá znělostní spodobu regresivní, je obvyklá spíše pro obyvatele Moravy a Slezska (toto rozdělení však podle našich zkušeností nemusí platit absolutně a záleží rovněž na konkrétních výrazech, v nichž se skupina *sh* objevuje, které se nechovají vždy zcela jednotně). Podle kodifikace jsou obě realizace této souhláskové skupiny spisovné (srov. Výslovnost spisovné češtiny I, s. 54), tj. např. slovo *shánění* má dvě ortoepické podoby: [sxa:ňeňi:] i [zha:ňeňi:].

Skupina *sh* však zdaleka není v řeči tak frekventovaná jako spojení *neslabičná předložka + slovo začínající na samohlásku*, na něž poskytoval korpus Monolog i v současném rozsahu poměrně hojný počet dokladů. Pokud tedy jde o sekvenci *sh*, podařilo se nám nalézt pouze 15 relevantních dokladů od 8 různých mluvčích (5 žen a 3 mužů), z nichž 13 bylo realizováno nezněle jako [sx] a 2 zněle jako [zh]. Oba případy znělé výslovnosti se vyskytly u téže mluvčí, jejímž rodištěm je Olomouc, studovala a žije však v Praze. Na základě takového minima dat lze pouze předpokládat, že se ve vysílání Českého rozhlasu lze pravděpodobně setkávat s oběma spisovnými variantami, nejsou zde (např. kvůli stížnostem posluchačů) snahy o jednotnou výslovnost u všech mluvčích a že tento jev je zřejmě v idiolektu mluvčích natolik silně zakořeněn, že jej neovlivní ani dlouhodobý pobyt v oblasti, kde převládá opačná varianta. Tato sonda však naopak potvrdila, že k obdobným fonetickým výzkumům je zapotřebí daleko rozsáhlejších korpusů mluveného jazyka, než jaké jsou nyní k dispozici. Účelně lze tuto problematiku zkoumat tehdy, kdy budou k dispozici (nejméně) řádově desítky dokladů, avšak je nutné brát ohled i na náležité vyvážení mluvčích s ohledem na jejich regionální původ. Pak bude možné např. ověřit platnost tradovaných tvrzení o rozdílech ve spisovné výslovnosti mluvčích z Čech a z Moravy/Slezska.

4 Nedostatečnost současných korpusových zdrojů

Ačkoli by se mohlo zdát, že zdroj obsahující několik stovek minut nahrávek, jejichž podstatnou část tvoří nepřilíš staré zpravodajské relace, bude pro podobné výzkumy dostatečný, není tomu, jak jsme viděli, zdaleka tak. Vyhledávání potřebných (nejméně desítek) dokladů uzuální výslovnosti např. na internetu v on-line archivu Českého rozhlasu nebo České televize, na serverech typu YouTube a podobně, je práce velmi časově náročná a zdoluhavá, metodologicky sporná, a navíc často přináší minimální výsledky. Z tohoto důvodu je nutné korpusy mluveného jazyka neustále rozšiřovat, systematicky doplňovat o nová data a s ohledem na fonetické výzkumy zpřístupňovat i zdrojové nahrávky. Pak mohou sloužit i podobným analýzám, které jsou zajímavé nejen z lingvistického, ale i z aktuálního společenského hlediska.

5 Jiné (budoucí) možnosti výzkumu mluvené češtiny

Budování korpusů zaměřených na mluvený jazyk je činnost poměrně náročná a dlouhodobá. Zvláště pro případy, kdy je třeba pouze co nejrychleji zjistit výslovnostní podobu např. konkrétního slova cizího původu nebo vlastního jména, je nezbytné v současnosti hledat i jiné než klasické lingvistické korpusové zdroje.

Jedním ze zdrojů pro podobné informace o mluvené češtině se snad v budoucnu stane nástroj, který je vyvíjen na Technické univerzitě v Liberci. Projekt NAKI (DF11P01OVV013), kteří tamní vědci od Ministerstva kultury ČR získali, se zaměřuje na zpřístupnění zvukového archivu Českého rozhlasu pro sofistikované vyhledávání pomocí automatického rozpoznávání řeči. Nejedná se o přístup zaměřený primárně jazykovědně, avšak doufáme, že se bude jednat o velice zajímavý a cenný zdroj reálných dat, která by mohla být případně využita i lingvisticky.

Literatura a zdroje

Český národní korpus – ORAL2008 (2008). Praha: Ústav Českého národního korpusu FF UK. On-line na adrese <<http://www.korpus.cz>>.

Korpus DIALOG 1.1. (2012). Praha: Ústav pro jazyk český, AV ČR. On-line na adrese <<http://ujc.dialogy.cz>>.

Korpus MONOLOG 1.0. (2012). On-line na adrese <<http://monolog.dialogy.org>>.

PALCOVÁ, Z. – VEROŇKOVÁ, J. – VOLÍN, J. – SKARNITZL, R. (2004): Stabilizace některých termínů pro fonetický popis češtiny v závislosti na nových výsledcích výzkumu. In: Duběda, T. (ed.). *Sborník z Konference česko-slovenské pobočky ISPhS 2004*. Praha: FF UK, s. 65–74.

PALCOVÁ, Z. (1982): Výslovnost rozhlasových mluvčích. *Naše řeč*, 65, s. 186–196.

Výslovnost spisovné češtiny I. Zásady a pravidla. Výslovnost slov českých (1967). B. Hála (ed.). 2. vydání. Praha: Academia.

AKADEMICKÝ SLOVNÍK SOUČASNÉ ČEŠTINY A SOFTWAREVÝ NÁSTROJ PRO JEHO TVORBU¹

MICHAELA LIŠKOVÁ

oddělení současné lexikologie a lexikografie, Ústav pro jazyk český AV ČR, v. v. i.
Letenská 4/125, 118 51 Praha 1

Abstract

The article presents the preparation of the new monolingual dictionary of contemporary Czech and the new Dictionary Writing System (DWS) that is being developed at the Institute of the Czech Language of the Academy of Sciences of the CR, v. v. i.

Klíčová slova: slovník, software pro psaní slovníku, heslář, zdroje pro slovník, mikrostruktura, význam, exemplifikace, současná slovní zásoba

Keywords: dictionary, dictionary writing system, list of entries, dictionary sources, microstructure, sense, exemplification, contemporary vocabulary

1. Úvod

V oddělení současné lexikologie a lexikografie Ústavu pro jazyk český AV ČR, v. v. i., v Praze vzniká od počátku roku 2012 nový všeobecný výkladový slovník s pracovním názvem *Akademický slovník současné češtiny* (ASSČ). Heslář ASSČ by měl obsahovat 120–150 tisíc lexikálních jednotek². Předpokládaným uživatelem ASSČ je středoškolsky vzdělaný rodilý mluvčí, počítá se však i se zájemci o češtinu jako cizí jazyk.

ASSČ si klade za cíl zachytit slovní zásobu současné češtiny. Jako krajní bod v minulosti byl určen konec 2. světové války; časový úsek od roku 1945 po dnešek by měl obsáhnout tři generace. Ve slovníku popisujeme slovní zásobu všeobecně rozšířenou, užívanou ve veřejné oficiální a polooficiální komunikaci i v komunikaci běžné (tj. neveřejné, neoficiální). V omezeném rozsahu jsou ve slovníku podávány jednotky užívané v profesní a zájmové komunikaci, pokud jsou již užívány také mimo své původní prostředí. Z výrazů nářeční povahy se zachycují jen ty výrazy, které mají širší územní platnost, užívají se především v mluvené komunikaci, příp. i v literatuře. Ve svém stručném příspěvku představíme pouze několik – velmi dílčích – aspektů z komplexní problematiky zpracování slovníku.

2. Software pro psaní slovníku

Pro tvorbu a publikaci slovníků existuje několik kvalitních zahraničních softwarových nástrojů, např. TshwaneLex, IDM DPS, iLEX či open-source program Matapuna Dictionary Writing System. Po pečlivém vyhodnocení výhod a nevýhod těchto aplikací pro naše potřeby jsme se rozhodli pro tvorbu nového softwaru, který zohlední specifika tvorby českého výkladového slovníku

1 Příspěvek vznikl v rámci grantového projektu programu aplikovaného výzkumu a vývoje národní a kulturní identity (NAKI) *Nová cesta k modernímu jednojazyčnému výkladovému slovníku současné češtiny* (DF13P01OVV011).

2 Ke dni 28. 11. 2013 je zpracováno 9 286 hesel (celé hesláře písmen *a* a *b* a část hesláře písmene *c*).

a umožní průběžné a rychlé modifikace programových modulů podle eventuálních změn lexikografické koncepce.

Pro psaní ASSČ tedy vzniká – a dále je zdokonalován – speciální softwarový nástroj (DWS). Byl navržen tak, aby umožnil slovníkové zpracování současné slovní zásoby češtiny a aby s jeho pomocí mohly do budoucna vzniknout i slovníky specializované, např. slovník neologický, slangový, terminologický, homonymický, slovník zkratk nebo cizích slov. Software umožňuje jak tištěný, tak elektronický slovníkový výstup (ASSČ bude postupně uveřejňován na internetu a po dokončení celého díla bude rozhodnuto, zda bude slovník vydán i v knižní podobě). Zájemce o (nejen technické) podrobnosti odkazujeme na článek K. Barbierika a kol. (Barbierik, 2013, s. 9–26)³.

3. Zdroje a heslář slovníku

Heslář slovníku vymezuje J. Filipec jako „lexémovou podmnožinu podloženou množinou výskytů v materiálové kartotéce, korpusu, popř. dílčích kartotékách, vzhledem k cíli jistého typu slovníku“. (Filipec, 1995, s. 22) Heslář ASSČ je vytvářen na základě vyvážených korpusů Českého národního korpusu (SYN 2000, SYN 2005, SYN 2010) s celkovým objemem 300 milionů slov. V souladu se zvoleným způsobem zpracování slovníkových hesel postupujeme i při tvorbě hesláře od počátku abecedy po jednotlivých písmenech, nikoli např. po slovních druzích nebo od nejfrekventovanějších hesel; heslář celého slovníku tedy není předem k dispozici. Vygenerovaný hrubý heslář po odstranění překlepů, chybné lemmatizace ap. následně vyhodnocuje každý zpracovatel přiděleného úseku hesláře. Na jedné straně vyřazuje úzce specializované termíny, které jsou sice početně dostatečně zastoupeny, avšak např. jen v několika málo zdrojích nebo určité odborné oblasti. Na straně druhé jej obohacuje o lexémy, které jsou sice frekventované méně, ale patří do daného slovtvorného hnízda (doplňují se zejména přechýlená jména, adverbia, zdobněliny).

Hnízdování se užívá méně než v předchozích slovnících (srov. *Slovník spisovné češtiny pro školu a veřejnost*, *Slovník spisovného jazyka českého*). Jako samostatná hesla se v ASSČ podávají tradičně hnízdované deriváty (vztahová adjektiva, adverbia, substantiva označující názvy vlastností). Pomocí hnízdování se zpracovávají jen některé lexikální typy – přechýlené názvy (*anglofilka* ← *anglofil*), zdobněliny (*ažurka* ← *ažura*) a opakovací slovesa (*balívat* ← *balit*), a to ty, u nichž se významová struktura derivátu neliší od základového slova.

Ukázka hesláře v softwarovém nástroji pro tvorbu slovníku (DWS):

heslo	hom.	sl. druh	varianty	zpracovatel	vytvořeno	změněno
agentka		ž.		pernicova	19.09.2012	19.06.2013
agentský		přid.		pernicova	19.09.2012	19.10.2012
agentství		s.		pernicova	16.01.2013	23.01.2013
agentura		ž.		pernicova	19.09.2012	19.10.2012
agenturné		přisl.		pernicova	19.09.2012	22.10.2012
agenturní		přid.		pernicova	19.09.2012	19.10.2012
agilita		ž.		pernicova	19.09.2012	06.11.2013
agility		s.		pernicova	19.09.2012	06.11.2013
agilné		přisl.		pernicova	19.09.2012	13.11.2013
agilní		přid.		pernicova	19.09.2012	18.10.2012
agilnost		ž.		pernicova	19.09.2012	13.11.2013
agitace		ž.		pernicova	19.09.2012	06.11.2013
agitačné		přisl.		pernicova	19.09.2012	06.11.2013
agitační		přid.		pernicova	19.09.2012	14.12.2012

3 Pro zajímavost uvedme, že při současné první fázi redakčních úprav velmi oceňujeme snadnou spolupráci autorů a redaktorů slovníkových hesel prostřednictvím prototypu redakčního modulu, který eviduje i historii změn: v případě potřeby je tak možné se vrátit k některé z předchozích verzí hesla.

Při zpracování hesel je pro nás základní korpus SYN, vedle toho však využíváme i dalších zdrojů. Utříděný materiál lze čerpat z interní databáze Pralex, obsahující podrobný popis lexikální, sémantické a syntaktické spojitelnosti heslových slov, zejména substantiv⁴. S tříděním materiálu zejména u tzv. velkých hesel nám velmi pomáhá nástroj Word Sketch aplikace The Sketch Engine. Cenný je pro nás elektronický mediální archiv Newton Media, a to především pro přepisy mluvených mediálních komunikátů. Dále využíváme neologickou databázi Neomat⁵, interní oborovou databázi ODE a jiné zdroje, doplňkově vyhledáváme i na internetu.

4. Mikrostruktura slovníku

V mikrostruktuře se podává všestranná informace o dané lexikální jednotce. Mikrostrukturu hesla v ASSČ tvoří lemma (včetně variant), označení homonymie, výslovnost, původ lexikální jednotky, gramatická charakteristika (slovnědruhové zařazení⁶, tvaroslovné údaje; valence), stylisticko-pragmatická charakteristika⁷, obor užití, u řídkého výskytu lexému též údaj o frekvenci. Dále se uvádí výklad významu, informace o začlenění lexému do synonymních a opozitních vztahů, exemplifikace, příp. rozšiřující etymologická nebo encyklopedická poznámka a odkazy na (sémanticky, gramaticky) souvztažná hesla.

Součástí heslové stati jsou také připojené víceslovné lexikální jednotky a jednotky přihnízdované.

Jako ukázkou zpracování heslového odstavce v připravovaném slovníku uvádíme heslo *bába*, *baba*⁸. Takto v DWS vypadá horní část karty hesla:

4 Databáze Pralex vznikala v letech 2005–2011 v rámci výzkumného záměru *Vytvoření databáze lexikální zásoby českého jazyka počátku 21. století* (AV0Z90610521). Pro její tvorbu byl ve spolupráci s Centrem zpracování přirozeného jazyka FI MU v Brně vyvinut lexikografický software Praled. Tento software je však podle toho, jak o DWS (dictionary writing system) píše např. A. Abel a A. Klosa (Abel, Klosa, 2012, s. 1–23), neúplný a pro tvorbu a publikaci výkladového slovníku by ho nebylo možné použít bez rozsáhlých programových úprav, změn a doplnění (chybí zejména redakční a administrativní nástroje a nástroj pro tiskový výstup odpovídající tradičnímu slovníku, protože o tištěné podobě databáze Pralex se dále neuvažovalo).

5 Neomat je interní databáze provozovaná oddělením současné lexikologie a lexikografie ÚJČ. V omezeném rozsahu je přístupna veřejnosti z URL: <http://www.neologismy.cz>.

6 U mužského rodu se důsledně rozlišuje životný a neživotný rod.

7 Více ke stylisticko-pragmatické charakteristice v ASSČ viz příspěvky ze semináře *Styl a slovník* pořádaného Jazykovědným sdružením, na němž vystoupili pracovníci dvou lexikografických projektů: *Německo-české lexikální databáze* a projektu *Nová cesta k modernímu jednojazyčnému výkladovému slovníku současné češtiny* (Opavská, Kochová, 2013 a Lišková, 2013).

8 Koncepční zásady se v průběhu slovníkové práce upravují a zpřesňují, je tedy možné, že uváděná slovníková hesla budou ještě (z)měněna.

A takto tiskový náhled:

bába I báby, **baba I** baby (2. mn. bab) ž.

1. častěji **bába** kolokv. hanl. stará žena: *stará bába, vesnické báby; vedle v pokoji se ve večerních hodinách vždy sesednou báby a donekonečna klevetí; dědek a bába z perníkové chaloupky* pohádkové postavy □ **bába kořenářka** zast. stará žena sbírající léčivé byliny a připravující z nich léky, masti ap.

2. kolokv. hanl. nepřijemná, zlá žena: *protivná bába u přepážky, nepřijemná baba v komisi; pode mnou bydlí strašně zvědavá baba*

3. **baba** kolokv. expr. žena, ženská, zejména pěkná nebo energická: *pěkná baba, sexy baba; v hospůdce se probere vše od sportu až po baby; jsem pořádná baba, která se nedá jen tak zastrašit*

4. častěji **bába** kolokv. expr. manželka, životní partnerka: *nikdo nedovede uvařit líp než moje bába; nenapadlo mě, že mě kámoš podrazí s mojí babou*

5. **bába** zast. babička 1 (matka jednoho z rodičů): *obě zdědily po své společné bábě vzácnou ušlechtilost a vznešenost*

6. **baba** hanl. (zejména o mužích) bázlivý, zbabělý člověk; syn. slaboch, zbabělec: *ty babo!*; často v přísudkovém postavení *nebuď baba!*

7. **baba** účastník dětské hry, který honí ostatní a svou roli předává zprav. dotekem: *baba honí ostatní dívky ve vymezeném prostoru pro hru* • tato role, úloha hráče: *dát, dostat babu; mít babu; hrát (si) na babu; hra na babu* (i název hry); *baba se předává plácnutím*

□ **baba princmetálová** zast., hanl., řidč. **bába princmetálová** zast., hanl. zlá, hubatá žena, zpropadená, zatracená ženská: *Tatínek svou tchyni moc nemiloval a vždycky říkal: Uvidíš, že nás ta baba princmetálová všechny přežije.* [Blesk 2007]

□ **porodní bába** pomocnice při porodu: *po tisíciletí byl porod záležitostí porodních bab; co si ale v kritické situaci počne s rodičkou samotná porodní bába kdesi v kuchyni?*

□ **slepá bába** účastník dětské hry, který se zavázanýma očima chytá ostatní: *slepá bába se roztočí a hledá kolem sebe s nataženýma rukama* • tato role, úloha hráče: *hra na slepou bábu* (i název hry); *hrát (si) na slepou bábu* (i fraz.)

◇ **(a teď) babo rad'** kolokv. nevím, co teď (budeme) dělat; není jasné, jak tuto situaci vyřešit: *tak a teď babo rad'! Která z těch variant je ta správná?; Velký anglicko-český slovník uvádí dokonce deset významů. Potom, babo rad'!*

◇ **[hádat se, dohadovat se ap. (o něco)] jak(o) báby na trhu** kolokv. expr. obhroublým, nedůstojným způsobem něco dojednat, o něco se hádat: *politické strany se hádají jako báby na trhu; jak báby na trhu se dohadují v Táboře o odbahnění a revitalizaci nádrže Jordán*

◇ **hrát (si) na slepou bábu** skrývat pravé záměry, přetvařovat se, nejednat otevřeně: *pokud jde o naše stanoviska a názory, přestali jsme si hrát na slepou bábu; nemůžete hrát na slepou bábu a tvářit se, že je všechno v pořádku*

◇ **kam čert nemůže, (tam) nastrčí babu / bábu** kolokv. expr. (pořekadlo) příchod, přítomnost (staré) ženy situaci, práci zhoršuje, komplikuje ap.

◇ **svíčková bába** hanl. přepjatě, často neupřímně zbožná žena (řidč. i muž): *podle mínění svíčkových bab se oženil moc brzy po smrti své první ženy*

4. 1. Výklad významu

V rámci mikrostruktury je nejdůležitější výklad významu, tj. slovníková definice. Při podání lexikálního významu hesel se v ASSČ vychází ze základního konceptu určení druhového zařazení – genus proximum – a určení diferenčních významových prvků – differentia specifica (se zřetelem k tomu, že vedle prvků nociónálních je třeba popsat i prvky pragmatické). Za součást lexikálního významu se však považují také ty významové prvky, které sice není možno hodnotit jako nutné distinktivní rysy, ale které odrážejí komplex vědomostí, jež o označované mimojazykové skuteč-

nosti mají uživatelé jazyka na úrovni běžného poznání. V určité míře se tedy při výkladu významu připouštějí „encyklopedické“ údaje (zejména ty, které se objektivně odrážejí v slootovorné struktuře slova, v ustálených přirovnáních a dalších frazémeh a v sémanticky odvozených významech) (srov. Dolník, 2009, s. 173nn.).

Věnujeme péči tomu, aby byl výklad významu srozumitelný a neporušovala se základní lexikografická zásada nevykládat jednodušší složitějším; v definicích se nepoužívají slova, která nejsou v hesláři slovníku. Abychom se vyvarovali definic kruhem, nevykládáme synonymy⁹.

Frustraci nad slovníkovými definicemi zná jistě každý uživatel slovníku. Kromě zmíněných definic kruhem jsou nepříjemné také složité výklady významů, jejichž metajazyk je na příliš vysoké odborné úrovni. Terminologická slovní zásoba pochopitelně klade specifické požadavky, termíny nelze vykládat banálně a zjednodušovat je přes míru¹⁰, ale jestliže čtenář slovníku (který není vzdělán na vysoké úrovni v daném oboru) výkladu neporozumí, pak slovník selhává. Jako příklad uveďme heslo *absorbance* z *Nového akademického slovníku cizích slov*: **absorbance** -e ž <l> fyz. chem. logaritmicky vyjádřená hodnota vnitřní transmisní hustoty, extinkce.

Autoři komerčního *Slovníku současné češtiny* se v tomto ohledu vymezují vůči stávajícím slovníkům; v předmluvě deklarují, že jejich hlavním cílem je srozumitelnost: „Výkladový slovník není slovníkem vědců pro vědce, ale především pro studenty a veřejnost.“ (2011, s. 5) Otázkou je, nakolik se jim to podařilo¹¹ (srov. Čermáková, 2013, s. 200nn.).

4. 2. Exemplifikace

Na zpracovatele klade vypracování exemplifikace nemalé nároky: „Exemplifikace lemmatu musí být především jak typická, tj. ilustrující typický (a nikoli často idiosynkratický autorský) úzus, tak konkretizační vůči veškeré informaci většiny hesla, zvláště však vůči informaci naznačené funkcí, valencí, kontextem a sémantikou.“ (Čermák, 1995, s. 107) Těmto požadavkům se při tvorbě exemplifikační části heslového odstavce snažíme dostat.

Vzhledem k současné mohutnosti jazykových zdrojů se většinou nestává, že bychom nemohli najít vhodný doklad. Poctivé procházení zdrojů s ohledem na jmenované aspekty je však – i přes využívání nejrůznějších sofistikovaných nástrojů – poměrně časově náročné. V exemplifikaci uvádíme jak stručné typické doklady, tak doklady rozšířené a větné, které mohou ukazovat i méně obvyklé užití slova, což se týká především výrazů z periferních oblastí slovní zásoby. Výjimečně, kupř. u synsémantik nebo frazémů, je třeba exemplifikovat pomocí nadvětných celků.

Při revizích exemplifikace jsou nejčastěji odstraňovány tyto nedostatky: exemplifikace je chudá, chybí některé z typických spojení; exemplifikace je příliš obsáhlá, nepřehledná, zbytečně se v ní opakují doklady se stejnou spojitelností a sémantikou; bez dobrého důvodu jsou zařazovány okrajové případy užití lexikální jednotky; exemplifikace je nicneříkající¹²; zpracovatel v exemplifikaci uvádí významové rysy, které by měly být spíše součástí výkladu významu; ad. Tyto lapsy (hodno-

9 Možnou výjimkou je výklad synonymem u stylově příznakových a expresivních lexémů aj.

10 Pro práci na ASSČ budujeme síť odborníků, s nimiž konzultujeme význam, exemplifikaci a případně další údaje hesla daného oboru nebo oblasti. Věříme, že tato spolupráce pomůže dovést podobu terminologických hesel do optimální podoby.

11 Recenzentka slovníku A. Čermáková upozorňuje na obcházení pravidel „slušného lexikografického chování“ (jako jsou požadavky na přehlednost a rychlou srozumitelnost) např. u lexikálních jednotek z botanické oblasti. Kromě definic kruhem recenzentka kritizuje užívání termínů v definicích (některé z nich nejsou ani v hesláři slovníku).

12 Např. u frazému **tichá voda břehy mele**: *No, pěkně se vybarvuje. Ale jak se říká, tichá voda břehy mele.* [V. Kabátová]

cené však do značné míry subjektivně) se snažíme minimalizovat a usilujeme též o to, aby byla exemplifikace čtivá.

Jako ukázkou zpracování významu a exemplifikace v připravovaném slovníku uvádíme ještě tři typově různá hesla: *babybox*, *bacilonosič* a *babrat se*.

babybox [bejbi-] -xu (6. j. -xu) m. než. <angl.>

speciálně vybavená schránka v blízkosti zdravotnického zařízení určená k anonymnímu odložení (novorozeného) dítěte: *odložit dítě do babyboxu*; *ženy, které zvolily babybox, zřejmě řešily těžkou životní situaci*

bacilonosič -če (3., 6. j. -či, -čovi, 1. mn. -či) m. živ.

člověk nebo živočich (zprav. bez příznaků nemoci), který má ve svém organismu choroboplodné zárodky, jejichž vylučováním šíří nákazu, přenašeč nákazy: *bacilonosič žloutenky*; *vyšetření by měla odhalit případného bacilonosiče*; *Také choroby přenášené hmyzem a jinými bacilonosiči se díky globálnímu oteplení mohou rozšířit nebývalou měrou.* [DB 2004]

► **bacilonosička** (2. j. -ky, 2. mn. -ček) ž. přechýl.: *bacilonosička tyfu*; *pacientka se nakazila od 73leté příbuzné, která je bacilonosičkou a nemoc se u ní neprojevuje*

babrat se (3. j. -rá se, rozk. -rej se!, čin. -ral se, podst. jm. -rání (se)) ned. kolokv. expr.

1. (kdo || v čem) dlouho nebo opakovaně, bez účelu se zabývat nějakým problémem; syn. patlat se: *babrat se v problémech*; *babrání se v minulosti nemá smysl*

2. (kdo || s čím, s kým) zbytečně dlouho, s přehnaným úsilím se něčím, někým zabývat; syn. piplat se, párat se: *babráám se s tím hlášením už víc než hodinu*; *s ničím se moc nebabrá, na všechno jde zpříma*; *soud se zbytečně babral s jasnými viníky*

5. Závěr

V tomto příspěvku jsme uvedli několik disparátních poznámek k rozličným aspektům tvorby připravovaného výkladového slovníku: zmínili jsme se o softwarovém nástroji pro tvorbu a publikaci slovníku, dále o zdrojích a hesláři slovníku, z mikrostruktury slovníku pak zejména o výkladu významu a o exemplifikaci. Na ukázkou jsme připojili čtyři heslové stati z připravovaného slovníku (*bába*, *baba*; *babybox*; *bacilonosič* a *babrat se*).

V textu jsme citovali velkého lexikografa a lexikologa Josefa Filipce. Na závěr jej ocitujeme ještě jednou, tentokrát jakožto autora příležitostných veršů, které výstižně zachycují povahu slovníkářské práce: Bujný oř je mluva naše, slovník táhne se jak kaše.

LITERATURA A ELEKTRONICKÉ ZDROJE:

ABEL, A. – KLOSA, A. (2012): The lexicographic working environment in theory and practise. In.: R. V. Fjeld – J. M. Torjusen (eds.), *Proceedings of the 15th EURALEX International Congress*. Oslo: University of Oslo, s. 1–23.

BARBIERIK, K. – HOLCOVÁ HABROVÁ, M. – KOCHOVÁ, P. – LIŠKA, T. – OPAVSKÁ, Z. – VIRIUS, M. (2013): A New Path to a Modern Monolingual Dictionary of Contemporary Czech: the Structure of Data in the New Dictionary Writing System. In: K. Gajdošová – A. Žáková (eds.), *Natural Language Processing, Corpus Linguistics, E-learnig*. Lüdenscheid: RAM-Verlag, s. 9–26.

ČERMÁK, F. (1995): Paradigmatika a syntagmatika slovníku: možnosti a výhledy. In: F. Čermák – R. Blatná (eds.), *Manuál lexikografie*. Jinočany: H&H, s. 90–115.

- ČERMÁKOVÁ, A. (2013): Jaký slovník uživatelé češtiny potřebují? O Slovníku současné češtiny nakladatelství Lingea. *Slovo a slovesnost*, 74, s. 195–210.
- Český národní korpus [online]. Praha: Ústav českého národního korpusu. Cit. 2013-01-24. <<http://www.korpus.cz/>>.
- DOLNÍK, J. (2009): *Všeobecná jazykoveda: Opis a vysvetľovanie jazyka*. Bratislava: Veda. 376 s.
- FILIPEC, J. (1995) Teorie a praxe jednojazyčného slovníku výkladového. In: F. Čermák – R. Blatná (eds.), *Manuál lexikografie*. Jinočany: H&H, s. 14–49.
- IDM DPS [online]. Cit. 2013-01-24. <http://www.idm.fr/products/dictionary_writing_system_dps/27/>.
- iLEX [online]. Cit. 2013-11-24. <<http://www.emp.dk/ilexweb/index.jsp>>.
- KRAUS, J. a kol. (2005): *Nový akademický slovník cizích slov*. Praha: Academia.
- LIŠKOVÁ, M. (2013): Způsob označování stylové příznakovosti ve výkladovém slovníku. *Jazykovědné aktuality*, roč. 50, č. 3–4, s. 154–158.
- Matapuna Dictionary Writing System [online]. Cit. 2013-01-24. <<http://sourceforge.net/projects/matapuna/>>.
- Neomat [online]. Verze 2.0. Praha: Ústav pro jazyk český AV ČR, v. v. i. Cit. 2013-11-24. <<http://www.neologismy.cz/>>.
- Newton Media [online]. Cit. 2013-11-24. <<http://www.newtonmedia.cz/>>.
- Nový akademický slovník cizích slov (2005). Praha: Academia.
- OPAVSKÁ, Z. – KOCHOVÁ, P. (2013): Stylistická charakteristika lexikálních jednotek v českých výkladových slovnících. *Jazykovědné aktuality*, roč. 50, č. 3–4, s. 141–147.
- Slovník současné češtiny (2011). Brno: Lingea.
- Slovník spisovné češtiny pro školu a veřejnost (1978, 1994, 2003). Praha: Academia.
- Slovník spisovného jazyka českého (1960–1971). Praha: Nakladatelství ČSAV.
- The SketchEngine [online]. Cit. 2013-11-24. <<https://www.sketchengine.co.uk/>>.
- TshwaneLex [online]. Cit. 2013-11-24. <<http://tshwanedje.com/tshwanelex/>>.

OZNAČOVÁNÍ POHLAVÍ U NEJNOVĚJŠÍCH ANGLICIZMŮ V ČEŠTINĚ A RUŠTINĚ

NADEŽDA FRANKE, B.A.

Universita Christiana Albrechta v Kielu

Abstract

This paper deals with the actual realizations of sexus marking of Russian and Czech borrowed substantives. It aims at figuring out the process of sexus marking of borrowed nouns, which denote human beings. The aim of the paper is also to represent possible differences or commonalities of the sexus marking of the borrowed nouns in Russian and Czech. A corpus of the 20 most common English borrowings, which denote people, has been compiled for the accomplishment of the linguistic analysis. This corpus is based on the three dictionaries of foreign words in Russian and Czech. During the linguistic analysis corresponding feminine correlates or other possibilities of sexus marking have been found for every example of the corpus. This has been arranged with the help of the national corpora of Russian and Czech including the supplement of the search machines on the Internet.

Klíčová slova: slovotvorba, sexus, přechylování

Keywords: word formation, gender, feminization

Referát „Označování pohlaví u nejnovějších anglicizmů v češtině a ruštině“ se věnuje skutečnému užívání označení pohlaví u výpůjček z angličtiny v českém a ruském jazyce.

Politické, sociální a společenské proměny v Rusku a Česku vedly k razantnímu růstu slovní zásoby těchto slovanských jazyků, především prostřednictvím enormního vzrůstu lexikálních výpůjček. Většina takových výpůjček, a to i výpůjček v oblasti označování osob, pochází z angličtiny, která vykazuje jiné možnosti při označování pohlaví než slovanské jazyky.

Novým označením ženských osob a přechylování v oblasti nové slovní zásoby není v současné lingvistice věnováno mnoho pozornosti. Proto cíl tohoto příspěvku spočívá v představení možností označování pohlaví u výpůjček z angličtiny a v popsání rozdílů a společných rysů češtiny a ruštiny při označování pohlaví u anglicizmů. Za účelem provedení vlastní analýzy obou jazyků byl sestaven pomocí tří slovníků cizích slov v češtině¹ a ruštině² seznam zkoumaných anglicizmů. Tento seznam obsahuje dvacet nejvíce rozšířených anglicizmů, které označují osoby. Přechýlené ženské podoby nebo jiná užívaná označení pohlaví byla dohledávána v Českém národním korpusu³ a v Ruském národním korpusu⁴ a pomocí internetových vyhledávačů.

1 MARTINCOVÁ, O. (eds.) (1998): *Nová slova v češtině: slovník neologizmů*. Praha: Academia.

2 KRYŠIN, L. P. et alii (eds.) (2005): *Tolkovyj slovar' inozazyčnych slov*. Moskva: ĚKSMO. Citován jako TSIS;

SKLJAREVSKAJA, T.N. et alii (eds.) (2007): *Tolkovyj slovar' ruskogo jazyka načala XXI veka. Aktual'naja leksika*. Moskva: ĚKSMO. Citován jako TSRJ.

3 *Český národní korpus – InterCorp*. Ústav Českého národního korpusu FF UK, Praha (2013) [online]. Cit. 2013-06-12. <<http://korpus.cz/index.php>>. Dalé citován jako ČNK.

4 *Nacional'nyj korpus ruskogo jazyka* (2003-2013) [online]. Cit. 2013-06-12. <<http://ruscorpora.ru/index.html>>. Dále citován jako NKRJ.

Vlastní jazyková analýza se skládá z následujících částí:

- určení kodifikované, popř. obecně užívané formy každé z výpůjček,
- dohledání přechýlených podob nebo jiných označení pohlaví u jednotlivých výpůjček,
- systematizace pravidel označování pohlaví v obou jazycích.⁵

Na hranici 20. a 21. století vykazují ruština a čeština určité tendence, které jsou podobné v obou jazycích (Krysin, 2010, s. 210–211). K takovým tendencím patří nárůst výpůjček z různých cizích jazyků.

Za výpůjčky jsou považovány prvky cizího jazyka, které přechází do cílového jazyka v důsledku jazykového kontaktu (Jarceva, 2000, s. 158–159). Od 80. let 20. století je pozorována intenzifikace tohoto procesu v obou zkoumaných jazycích. L. P. Krysin označuje tento proces za „explozi“ výpůjček, která je ve slovanských jazycích pozorována především od roku 1990 (Krysin, 2002, s. 27). Tento vývoj jazyka je možné zdůvodnit především tím, že v obou zemích, Ruské federaci a České republice, vznikly určité politické, ekonomické, sociální a kulturní podmínky, které usnadnily přijetí a používání slovní zásoby cizího původu. Prostřednictvím otevření hranic se zintenzívnily kontakty mluvčích ruštiny a češtiny se západními zeměmi a s nositeli jejich jazyků, také se zvýšila orientace na západ v oblastech politiky, ekonomiky, masmédií, medicíny, komunikace, IT sektoru, sportu. Zvýšila se i prestiž cizích jazyků. Lze tedy konstatovat, že s přebíráním západních fenoménů vznikla určitá prázdná místa v lexikálním systému slovanských jazyků, která byla zaplněna pomocí výpůjček (Zemskaja, 2004, s. 517–519). Aktivní používání vypůjčených lexémů se týká také označování osob, především pracovních titulů a jiných označení osob, které vykonávají určitou činnost.⁶

Na konci 20. a na začátku 21. století je možné zaznamenat prosazování se angličtiny jako „lingua francy“. V Rusku a v Česku neustále roste počet lidí, kteří mluví anglicky. Proto je časový úsek, kterým se zde zabýváme, charakteristický tím, že většina výpůjček pochází z angličtiny (Zemskaja, 2004, s. 518). Morfologie a tvoření slov v ruštině a češtině, které patří k silně flektivním jazykům (Krysin, 2010, s. 239), vykazují strukturální rozdíly ve vztahu k angličtině, ve které převažují izolační a aglutinační rysy (Bauer, 2003, s. 230–238). Před vlastní analýzou označování pohlaví u přejatých substantiv v ruštině a češtině je nutné představit možnosti označování pohlaví u anglických podstatných jmen. Anglický systém rodů rozeznává následující rody, které závisí na přirozeném rodu (sexu) a kategorii životnosti: jména mužského, ženského a středního rodu. Gramatický rod v angličtině není označen morfologicky a může být určen jen prostřednictvím vztahu k osobním zájmenům *he*, *she* nebo *it*. Angličtina má nicméně několik možností přímého označení pohlaví substantiv:

- rozlišení sexu ve slovní zásobě: *man*, *woman*, *boy*, *girl*;
- kompozita, která obsahují komponenty označující sexus: *girl friend*, *gentleman*, *woman doctor*;
- ženské slovotvorné přípony (např.: *-ine*, *-ess*, *-ette*) a mužské slovotvorné přípony (např.: *-er*, *-ist*, *-ian*).

Tato, co do sexu rozlišená, označení osob tvoří ovšem relativně malou část slovní zásoby angličtiny. Většina anglických podstatných jmen patří k nerozlišeným označením osob, která mohou označovat jak muže, tak i ženy a jsou nejčastějšími „dodavateli“ výpůjček. Tradičně jsou taková podstatná jména zařazována k tzv. společnému rodu a využívají fenomén generického maskulina (Brinton, 2010, s. 116–117).

5 Úplný seznam výpůjček najdete v dodatku.

6 Viz tabulku v dodatku.

Na základě takového systému rodů v angličtině se mohou podstatná jména, která označují osoby, v procesu přejímání chovat takto:

- anglické substantivum bez rozlišení sexu může být přejato, přičemž následkem přizpůsobení k systému cílového jazyka dojde k rozlišení sexu buď slovtvorně nebo pomocí kongruence;
- anglické substantivum bez rozlišení sexu je přejato jako podstatné jméno mužského rodu, ke kterému se dotvoří odpovídající ženská podoba;
- anglické substantivum s rozlišením sexu je přejato jako takové (tedy buď jako substantivum mužského, nebo ženského rodu).

Dále je nutné představit a popsat na příkladu češtiny možné varianty označení pohlaví, které vzniknou následkem procesu přejímání.

Analýza dvaceti vybraných českých výpůjček z angličtiny, které označují osoby, ukázala, že čeština důsledně přejímá anglická substantiva bez rozlišení sexu jako substantiva s určitým (a sice mužským) sexem a tvoří od nich následně pojmenování ženských osob pomocí slovtvorné přípony (nejčastěji příponou *-k(a)*).

Příklad: *manažer – manažerka*: [...] *díky jejímu časově náročnému zaměstnání (pracovala jako manažerka penzijního a poradenského oddělení v místní účetní firmě)* [...]

Ang.: [...] *her more-than-full-time employment (as the manager of a pension administration and consulting group at a local accounting firm)* [...] (ČNK).

Další příklady:

Mně se moc nelíbí já bloggerky tohoto typu ani od sebe nerozezná-vám... (<<http://www.modnipeklo.cz/diskuse/bloggerka-jagrova-1988/>>. Cit. 2013-06-12).

Gabriela Nováková – interiérová designérka specializující se na individuální přístup ke každému interiéru. (<<http://www.gadesign.cz/designerka/>>. Cit. 2013-06-12).

Doprava a infrastruktura budou dál „in“, tvrdí čínská insiderka (<<http://www.investicniweb.cz/trhy/iw-v-city/2011/1/27/doprava-infrastruktura-budou-dal-tvrdi-cinska-insiderka/>>. Cit. 2013-06-12).

Věra Krincvajová – dramaturgyně, producentka.

(<http://rozhlas.cz/pardubice/jolana/_zprava/vera-krincvajova-dramaturgyne-producentka-1212126/>. Cit. 2013-06-12).

Seznam analyzovaných příkladů obsahuje však dvě výjimky, u kterých se nedotváří přechýlená podoba a jsou používána jen substantiva mužského rodu: *bojfriend / boyfriend* (dosud nekodifikované)⁷ a *plejboj / playboy* (obě varianty kodifikované)⁸. Pro označování ženských osob se používá český ekvivalent *přítelkyně* pro slovo *bojfriend* a lexém *plejboj* nejsou ženské podoby, což je vysvětlitelné jejich sémantikou. Nejednotnost kodifikace nebo nepřítomnost kodifikace lexému a nemožnost jejich přechylování je vysvětlitelná tím, že tato slova jsou nositeli jazyka ještě vnímána jako cizí slova.

Po analýze českých příkladů lze říci toto: Sexus je v sémantice výrazů označujících osoby v češtině natolik výrazný, že výpůjčky bez rozlišení sexu obdrží od mluvčích češtiny sexus (a sice zpravidla mužský rod) a mohou být dále přechylovány. Analýza dvaceti vybraných ruských výpůjček ukázala, že ruština využívá s ohledem na označování pohlaví osob dva způsoby přejímání. Při prvním způsobu přejímání je anglické substantivum bez rozlišení sexu přejímáno jako substantivum

7 Např.: *A můj bojfriend je pro všechny už od malička Honzka.* (<<http://www.baby-cafe.cz/modules.php?name=Forums&file=viewtopic&t=37241&start=45>>. Cit. 2013-06-12).

8 *Internetová jazyková příručka (2008-2013) [online].* Cit. 2013-06-12. <<http://prirucka.ujc.cas.cz/>>. Např.: *Já se radši beru jako plejboje, jako cosi na způsob Jamese Bonda* (ČNK).

s neurčitým sexem. Ve většině případů se jedná o substantiva se slovotvornou příponou *-er/-or*. Označování pohlaví pro pojmenování ženských osob probíhá takto:

- pomocí kongruence s predikátem: *Podderžala svoju kollegu i dizajner Svetlana Bekareva*⁹ (NKRJ). Ve většině příkladů je kongruence s predikátem jediným označením pohlaví. Při kongruenci s atributem využívají taková substantiva generického maskulina: *Ee priglasila na svoj pokaz mody v ramkach Nedeli mody v Moskve molodoj dizajner Kira Plastinina*¹⁰ (NKRJ).
- tvořením tzv. pseudokomposit: *Ženščina-broker okkupiruet Forex?*¹¹ (<http://www.for-trader.net/articles/other/woman_broker.html>. Cit. 2013-06-12).
- příponou *-š(a)*: *bankirša, dilerša*. Takové odvozeniny jsou jen hovorové a mají pejorativní význam: *Bankirša v znak primirenija podarila vrednomu suprugu dorogoj džip*¹² (NKRJ).

Při druhém způsobu přejímání je anglické substantivum s určitým sexem, který může být jak mužský, tak i ženský, přejato jako takové do ruštiny. Při kodifikaci jsou takové výpůjčky vnímány jako dva různé lexémy (TSIS): *biznesmen – biznesvumen / biznesledi: A včera vyjasnilos', čto neutomimaja biznesledi vnov' skryvetsja ot pravosudija*¹³ (NKRJ).

*Rejting vozglavila biznesvumen iz Šanchaja*¹⁴ (NKRJ).

Seznam ruských výpůjček obsahuje také dva příklady, u kterých se na rozdíl od jiných výpůjček nedotváří ženský protějšek, místo něj jsou používána jen substantiva mužského rodu: *bojfriend* a *plejboj*. Nemožnost dotvoření ženského protějšku k těmto lexémům je vysvětlitelná jejich sémantikou. Pro označování ženských osob u slova *bojfriend* se používá ruský ekvivalent: *podruga (příteľkyně)* nebo *devuška (děvče / slečna)*. Výpůjčka *it-girl* se používá v poslední době jako ženská podoba slova *plejboj*, přestože na sémantické rovině tyto anglicizmy nejsou úplně ekvivalentní.¹⁵ Lze přijmout tezi, že tato substantiva obdržela zvláštní pozici v jazykovém úzu, díky své prestiži jako cizí slova musela zůstat, aby podtrhla výjimečnou společenskou pozici označované osoby.

Na základě analýzy seznamu příkladů bylo zjištěno, že ruština i čeština vykazují tendenci označovat sex osob i u substantiv, která byla přejata z angličtiny. K označování sexu dochází v obou jazycích ale různými způsoby. V češtině je sexus v sémantice výrazů označovaných osob natolik výrazný, že výpůjčky bez rozlišení sexu obdrží od mluvčích češtiny sexus (a sice zpravidla mužský rod) a mohou být dále přechylovány.

Ruština naproti tomu nevykazuje v procesu přejímání moc pevná paradigmata ve vztahu k označování ženských osob. Suffixace se ukazuje jako poměrně neproduktivní, proto bývá sexus označován jinak: kongruencí, tvořením pseudokompozit nebo přejímáním substantiv, která rozlišují sexus už v angličtině. Jediný důsledný způsob označování pohlaví v ruštině představuje kongruence subjektu a predikátu, která ovšem není možná u všech forem sloves (např. u slovesa v přítomnosti). Také bylo zjištěno, že určitá skupina substantiv vykazuje tendenci ke generickému maskulinu.

9 *Svoji kolegyni podržela i designér Světľana Bekareva.

10 *Ji pozvala na svou módní přehlídku v rámci Týdne módy v Moskvě mladý designér Kira Plastinina.

11 Žena-broker okupuje Foreks?

12 Bankéřka na znamení usmíření darovala znepřátelenému choti drahý džip.

13 *Ale včera se ukázalo, že neúnavná byznysmenka je znovu na útěku před spravedlností.

14 *Na špici ratingu se umístila byznysmenka ze Šanghaje.

15 <www.yandex.ru> [online] (83 milionu jednotek).

V rámci příspěvku mohlo být analyzováno jen omezené množství výpůjček a jejich možností při označování pohlaví. Bylo by tedy ještě potřeba prostudovat popsání fenomény na základě rozsáhlejšího materiálu.

Seznam literatury:

- BAUER, L. (2003): *Introducing Linguistic Morphology*. Edinburgh: Edinburgh Univ. Press.
- BRITON, L. – BRITON, D. (2010): *The Linguistic Structure of Modern English*. Amsterdam et al.: Benjamins.
- KRYSIN, L. (2010): *Sovremennyyj russkij jazyk: sistema – norma – uzus*. Moskva: Institut russkogo jazyka im. V. V. Vinogradova: Jazyki slavjanskich kul'tur.
- KRYSIN, L. (2002): Leksičeskoe zaimstvovanie i kal'kirovanie v russkom jazyke poslednich desjatiletij. *Voprosy jazykoznanija* 6, s. 27–34.
- OHNHEISER, I. (2006): Zwischen Möglichkeit, Notwendigkeit und Usus: neue Bezeichnungen weiblicher Personen im Tschechischen mit einem Blick auf das Russische. In: J. Reinhart – T. Reuther (eds.), *Ethnoslavica. Festschrift für Herrn Univ. Prof. Dr. Gerhard Neweklowsky zum 65. Geburtstag. Beiträge des Internationalen Symposiums des Instituts für Slawistik der Universität Klagenfurt in Zusammenarbeit mit dem Institut für Slawistik der Universität Wien, Klagenfurt, 7. – 8. April 2006*. Wien: Gesellschaft zur Förderung slawistischer Studien, s. 231–244.
- ZEMSKAJA, E. (2004): *Jazyk kak dejatel'nost'. Morfema, slovo, reč'*. Moskva: Jazyki slavjanskoj kul'tury.

Slovníky:

- KARLÍK, P. et alii (eds.) (2002): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny.
- KRYSIN, L. P. et alii (eds.) (2005): *Tolkovyj slovar' inojazyčnych slov*. Moskva: ÈKSMO.
- MARTINCOVÁ, O. (eds.) (1998): *Nová slova v češtině. Slovník neologizmů 1*. Praha: Academia.
- SKLAREVSKAJA, T. N. et alii (eds.) (2007): *Tolkovyj slovar' russkogo jazyka načala XXI veka. Aktual'naja leksika*. Moskva: ÈKSMO.

Internetové prameny:

- Český národní korpus – InterCorp. Ústav Českého národního korpusu FF UK, Praha (2013) [online]. Cit. 2013-06-12. <<http://korpus.cz/index.php>>.
- Nacional'nyj korpus russkogo jazyka (2003-2013) [online]. Cit. 2013-06-12. <<http://ruscorpora.ru/index.html>>.
- Internetová jazyková příručka (2008-2013) [online]. Cit. 2013-06-12. <<http://prirucka.ujc.cas.cz/>>.

Dodatek

Tabulka označování pohlaví u nejnovějších anglicizmů v češtině a ruštině

Ruština

angličtina	ruština: maskulina	ruština: možnosti označení ženského pohlaví
banker	банкир (bankir)	женщина-банкир (ženščina-bankir) банкирша (bankirša) kongruence s predikátem
blogger	блогер (blogger)	женщина-блогер (ženščina-blogger) блогерша (bloggerša) kongruence s predikátem
broker	брокер	женщина-брокер (ženščina-broker) брокерша (brokerša) kongruence s predikátem
user	юзер (juzer)	женщина-юзер (ženščina-juzer) юзерша (juzerša) kongruence s predikátem
designer	дизайнер (dizajner)	дизайнерша (dizajnerša) kongruence s predikátem
manager	менеджер (menedžer)	менеджерша (menedžerša) женщина-менеджер (ženščina-menedžer) kongruence s predikátem
dealer	дилер (diler)	дилерша (dilerša) kongruence s predikátem женщина-дилер (ženščina-diler)
distributor	дистрибьютер (distrib'juter)	дистрибьютерша (distrib'juterša) женщина-дистрибьютер (ženščina- distrib'juter) kongruence s predikátem
image maker	имиджмейкер (imidžmejker)	женщина-имиджмейкер (ženščina- imidžmejker) имиджмейкерша (imidžmejkerša) kongruence s predikátem
insider	инсайдер (insajder)	kongruence s predikátem инсайдерша (insajderša)
moderator	модератор (moderator)	женщина-модератор (ženščina-moderator) модераторша (moderatorša) kongruence s predikátem
producer	продюсер (prodjuser)	продюсерша (prodjuserša) kongruence s predikátem
promoter	промоутер (promouter)	промоутерша (promouterša) женщина-промоутер (ženščina-promouter) kongruence s predikátem
reporter	репортер (reporter)	репортерша (reporterša) женщина-репортер (ženščina-reporter) kongruence s predikátem
boyfriend	бой-френд (boj-frend)	

girlfriend		девушка, подруга (devuška, podrugа)
playboy	плейбой (plejboj)	it-girl
businessman	бизнесмен (byznesmen)	
businesswoman		бизнеследи/ бизнесвумен (byznesledi/ byznesvumen)
showman	шоумен (šoumen)	
show woman		шоувумен (šouvumen)

Čeština

angličtina	čeština: maskulina	čeština: feminina
banker	bankéř	bankéřka
blogger	bloger	blogerka
broker	broker	brokerka
user	uživatel / user	uživatelka / userka
designer	designér	designérka
manager	manažer	manažerka
dealer	dealer	dealerka
distributor	distributor	distributorka
image maker	imidžmejker / imagemaker	imidžmejkerka
insider	insajder / insider	insajderka / insiderka
moderator	moderátor	moderátorka
producer	producent	producentka
promoter	promotor	promotorka
reporter	reportér	reportérka
boyfriend	boyfriend / bojfrend	
girlfriend		přítelkyně
playboy	playboy / plejboj	
businessman	byznysmen	
businesswoman		byznysmenka
showman	showmen / šoumen	
show woman		šoumenka/ showmanka

KOMPOZITA S PREFIXOIDEM *EURO-* V ČESKÝCH A RUSKÝCH PUBLICISTICKÝCH TEXTECH: KONTRASTIVNÍ POHLED

ZBYNĚK MICHÁLEK

Ústav slavistiky FF MU Brno

Abstract

The article deals with the usage of prefixoid *euro-* (*евро-*) in contemporary Czech and Russian journalistic texts. The basis for our observations are the corpora of Czech (Czech National Corpus – Český národní korpus) and Russian (Russian National Corpus – Национальный корпус русского языка) languages. We observe in which types of compounds the prefixoids are used. The situation is compared with dictionaries (old and new) of both Czech and Russian languages and with the situation on the Internet.

Klíčová slova: prefixoid *euro-/евро-*, publicistický styl, jazyk internetu, korpus, ČNK, НКРЯ
Key words: prefixoid *euro-/евро-*, journalistic style, Internet language, corpus, ČNK, НКРЯ

V příspěvku sledujeme využití prefixoidu *euro-* (*евро-*) v současných českých a ruských publicistických textech, a to na základě korpusů (pro češtinu Český národní korpus, pro ruštinu Национальный корпус русского языка). Všímáme si, v jakých typech kompozit se vyskytuje. Stav konfrontujeme se slovníky obou jazyků (staršími i novějšími) a se situací na internetu.

V poslední době můžeme v češtině, ale i v jiných evropských jazycích sledovat nárůst kompozit, jejichž součástí je prefixoid *euro-*. Nárůst souvisí s celospolečenskými změnami a integračními procesy, které mají počátek po druhé světové válce a které nabraly na intenzitě zejména na začátku 90. let, což souvisí s obecně známými geopolitickými přeměnami: s rozpadem Sovětského svazu, s koncem dělení Evropy na Východ a Západ a zejména se vznikem Evropské unie a přistoupením České republiky k ní.

Je přirozené, že tyto mimojazykové změny se odrážejí v lexiku coby nejdynamičtějším plánu jazyka. Pro ilustraci: Slovník spisovného jazyka českého (HAVRÁNEK, B.; 1960) neobsahuje prefixoid *euro-* vůbec, stejně tak Slovník spisovné češtiny pro školu a veřejnost (FILIPEC, J.; 2003). Nový akademický slovník cizích slov (KRAUS, J.; 2009) zachycuje prefixoid *euro-*, *evro-* jako část složenin ve významu ‚Evropa, evropský‘. Slovník Nová slova v češtině 1 (MARTINCOVÁ, O.; 1998) zachycuje prefixoid *euro-* opět jako část složenin ve významu ‚evropský‘. Nicméně slovník Nová slova v češtině 2 (MARTINCOVÁ, O.; 2004) již reflektuje diferencující se význam prefixoidu *euro-* a uvádí následující významy: 1. ‚evropský (v geopolitickém smyslu)‘, např. eurocentrismus; 2. ‚týkající se Evropské unie‘; např. eurobyrokracie; 3. ‚týkající se eura‘, např. eurobankovka.

Podíváme-li se na excerpce z Českého národního korpusu (<http://ucnk.ff.cuni.cz/>; dále ČNK) a srovnáme-li ji s výkladem významu prefixoidu *euro-* ve slovníku Nová slova v češtině 2, vidíme, že došlo za oněch takřka 10 let od jeho vydání k mírnému posunu, a to ve dvou rovinách:

1. Došlo k prudkému nárůstu užívání složenin s komponentem *euro-* ve významu ‚týkající se Evropské unie‘, a to tak, že tento význam bychom mohli považovat za primární. Vedle dnes již zcela běžných slov jako *eurodotace*, *eurokomisař*, *europarlament*, *europoslanec*, *euroúředník*, *euroústava* či *eurozatykač* jsou toho dokladem i okazionalismy, které jsou pro publicistický styl charakteristické a které dokládají velkou „živost“ prefixoidu *euro-*: *eurobubák* („Kdyby *eurobubák* Klaus

neexistoval, museli by si jej někteří jeho oponenti vymyslet.“), *eurocirkus* („*Eurocirkus* stojí miliardy.“; o stěhování Evropského parlamentu mezi Štarsburkem a Bruselem), *eurodisident* („Včera nakonec dorazil na poradu přímo Declan Ganley, *eurodisident* bezmála stejně populární jako sám Klaus.“), *eurofacka* („Brusel dal *eurofacku* Litvě.“), *eurohujer* („Radost z návratu rozumu do míst, kde by jej hledal už pouze přesvědčený *eurohujer*.“), *eurokřeč* („Premiérovi i dalším členům vlády by víc než prospělo, kdyby se zbavili své *eurokřeče*.“), *euromotor* („*Euromotor* se zadržává.“; o Německu a Francii), *eurořeč* (nebo *eurospeak*, *euroslang*; „V *euroslangu* se tomu říká komitologie.“) či *eurošprajc* a *eurotruc* (název článku o Václavu Klausovi).

2. Za nový význam prefixoidu *euro-* bychom považovali význam ‚splňující (celo)evropské standardy a normy‘. *Euro-* tu sice primárně odkazuje na Evropskou unii, nicméně význam je poněkud posunutý; srovnej slova jako *euroústava* a *europaleta*. Slovník Nová slova v češtině 2 sice uvádí například slovo *eurookno* ‚okno jednoduché konstrukce ze tří dřev lepených k sobě, aby se nekroutilo, splňující evropské normy‘ (dodejme, že konstrukce nemusí být pouze dřevěná, ale častěji je plastová), avšak tento význam prefixoidu *euro-* se explicitně neuvádí. Z dokladů ČNK citujme například *euroklíč* ‚speciální klíč, který umožňuje vozíčkářům na území celé EU používat zejména toalety bez toho, aniž se musejí dovolávat obsluhy‘ („Svůj *euroklíč* má u nás necelá padesátka lidí.“), *europaleta* ‚normovaná dřevěná transportní paleta pro vysokozdvizné vozíky‘ („Ukradl celkem dvě stě *europalet*.“) či okazionální ironické *euroseptik*, *eurostudna*, *eurovčelín* s významem ‚normovaný septik, studna, včelín‘ („Proto nyní při prodlouženém víkendu využijme volných chvil k zařízení všech *euroseptiků*, *eurostuden* a *eurovčelínů*, ať se to pak nenahromadí!“). Jmenujme snad ještě *euroláhev*, *euroškola*, *eurozámek* a sám pojem *euronorma*.

Na tomto místě bych rád upozornil na jakousi podskupinu těchto kompozit, jež sice nese význam ‚splňující evropské standardy a normy‘, který je však doplněn významem vedlejším ‚nevalné kvality a/nebo chuti‘. Nejznámějším zástupcem těchto kompozit je jistě slovo *europivo* ‚chuťově (pod) průměrné pivo vařené ve velkém objemu‘. Tento typ je však velmi produktivní, jak můžeme vidět opět na příkladech okazionalismů, s nimiž se setkáme v ČNK: *eurobuchty* a *euroguláš* („Za patnáct let lidí, který budou jíst *euroguláš* a *eurobuchty*, budou vzpomínat, co jsou to pravý český filmy, přisprostlý a košilatý.“), *europárek* („[...] celý boj proti ústavní smlouvě je pouhým Klausovým *europárkem*.“; narážka na britský seriál Jistě, pane ministře). Z internetu citujme například hojný *eurorohlík* („A nyní pojím výtečný europárek a *eurorohlík*.“) či okazionální *eurojogurt* a *eurovýživa* („*Eurojogurt* – doporučená *eurovýživa* splňující schválené chuťové vjemy.“), *eurobotnony* („Vitamínové *eurobotnony* – schválený doplněk stravy. Nedoporučuje se pro děti do 17ti let.“), *eurosýr* („*Eurosýr* – náhražková poživatina dle směrnice EU. Nejedná se o výrobek sýrového druhu.“), *europochoutka* („Čokoládová *europochoutka* pro děti – dle metodiky EU nelze prokázat pozitivní zdravotní účinky produktu.“) či *eurokabanos*, *eurosalám*, *euroklobásy*, *euroovoce*, *eurozelenina* atd. Z dokladů vidíme, že tento typ kompozit nese výrazný ironický odstín.

Pro úplnost a pro zajímavost uvedme několik excerpt okazionálních kompozit, která mají význam ‚evropský (v geopolitickém smyslu)‘: *eurobabička* („V semifinále soutěže *Eurobabička* si změnilo síly 15 vybraných seniorek z různých koutů Čech.“), *euromanány* („*Euromanány*: EU přestane protežovat dodavatele banánů z bývalých evropských kolonií Afriky, Karibiku a Tichomoří na úkor pěstitelů z Latinské Ameriky.“), *eurosafari* („Za branou *eurosafari* pobíhají mufloni, osli i koně převalského.“) a několik s významem ‚týkající se eura‘: *eurokalkulačka* („Mně se to už těžko přepočítává, a *eurokalkulačka* mi pomůže.“) či *euroměšec* („Plány pomoci a větší *euroměšec* takový slib nenahradí.“).

Podíváme-li se na situaci v ruštině, můžeme konstatovat, že se nijak radikálně neliší od situace v češtině, i když určité odchylky můžeme pozorovat. Slovník Ožegovův (ОЖЕГОВ, С. И.; 2012) ani Ušakovův (УШАКОВ, Д. Н.; 2007) prefixoid *евро-* neobsahují, stejně tak akademický Словарь

русского языка (ЕВГЕНЬЕВА, А. П.; 1981). Novější Толковый словарь русского языка начала XXI века (СКЛЯРЕВСКАЯ, Г. Н.; 2007) obsahuje sice několik kompozit s prefixoidem *евро-* (*европарламент*, *Евросоюз*, *евростандарт*), sám prefixoid však jako slovníkové heslo neobsahuje. Толковый словарь иноязычных слов (КРЫСИН, Л. П.; 2008) naopak obsahuje pouze prefixoid *евро-*, a to ve významech 1. ‚европейский‘, 2. ‚соответствующий стандартам, принятым в странах Европы‘. Význam ‚týkající se eura‘ ani ‚týkající se Evropské unie‘ se neuvádí, ačkoliv taková kompozita se pochopitelně vyskytují.

Pokud se podíváme na excerpta z Национального корпуса русского языка (<http://ruscorpora.ru/>), můžeme říci, že *евро-* ve významu ‚evropský (v geopolitickém smyslu)‘ tu jasně převládá: *евроазиатский* («Россия настаивала на *евроазиатском* сотрудничестве.»), *евроатлантический* («Это отвечало бы интересам укрепления стабильности и безопасности на нашем общем *евроатлантическом* пространстве.»), *еврокубок* («Таких команд в *Еврокубке*, собственно говоря, большинство.»), *европоцентристский* («Отказ от „*европоцентристских* подходов“ в пользу „национальной самобытности“») či *евроамериканский*, *евробизнес*, *евролига*, *еврооблигация*, *евросеть*, *Еврошахсоюз*.

Za sekundární význam, který výše zmíněný slovník neuvádí, bychom považovali ‚týkající se Evropské unie‘: *еврокомиссия* («По словам президента *Еврокомиссии*, в Европе полагают, что регион сильно зависим от российского газа.»), *европарламент* («В июне состоятся выборы в *Европарламент*») či samotný *Евросоюз* («А в *Евросоюзе* все решения ещё надо согласовывать»).

Nutno podotknout, že leckdy se stírá význam mezi ‚evropský‘ a ‚týkající se Evropské unie‘. Jak například chápat komponent *евро-* zejména v názvech obchodních společností, např. *ЕвроПартнер*, *Еврогипс*, *Еврохим* nebo *Евроэкспомебель*? Obchoduje společnost *Еврохим* se všemi evropskými státy, nebo pouze se členy EU? Domníváme se, že takový rozdíl není ani pro běžného Rusa směrodatný, neboť Evropu vnímá spíše jako celek, nikoliv jako Evropskou unii a země mimo ni.

Silné zastoupení mají kompozita s prefixoidem *евро-*, který tu odpovídá našemu významu ‚splňující (celo)evropské standardy a normy‘, který je však mírně posunut. Zatímco české *европřepravка* znamená ‚rozměrově normovaná bedna‘, ruské *еврокафель* neznačí ‚rozměrově normované kachlíčky‘, nýbrž ‚kachlíčky, jež odpovídají (jakési obecné) evropské normě a kvalitě. Příkladně: *евровагонка* ‚palubky‘ («Унесли стройматериалы: *евровагонку* и наличники, сделанные на заказ.»), *евроремонт* ‚taková rekonstrukce, po níž stavba (nejčastěji byt) dostane evropské parametry‘ («Ведется *евроремонт* административного здания.»), *евроотделка* ‚vybavení bytu, odpovídající evropskému standardu‘ («Номера с «*евроотделкой*», сауна, солярий, бар, мини-боулинг, бильярд.»), obecné *евростандарт* ‚obecná evropská norma‘ («Приятно отметить, что цена жилья включает полную его отделку по *евростандарту*»). Domnívám se, že dosti vysoká frekvence těchto kompozit je dána tím, že Rusové Evropu vnímají jako Západ, který má vyšší životní úroveň. Ojedinělé je *евробутылка* ‚normovaná láhev‘ («Завод начнет выпускать первые *евробутылки*»).

Poslední skupinu tvoří kompozita významově spojená s měnou euro, která slovník rovněž neuvádí. Zde nenacházíme odlišností od češtiny: *евробонд* ‚eurobond‘ («Они заработали на *евробондах* порядка 10 процентов в валюте.»), *еврозона* ‚eurozóna‘ («Наиболее значительный скачок цен был отмечен в самой проблемной стране *еврозоны* – Греции.»), *евроцент* ‚eurocent‘ («Мы покупаем по 40 евроцентов за литр сырого молока.»).

Zcela ojedinělé je *еврокотлета*, odpovídající svým významem *europivni*: «Чего б не носить ему спокойно свой ливайс, не пить хороший коньяк и не кушать за обедом *еврокотлету* гордон блю желтого цвета?»

Závěrem můžeme říci, že element *euro-* je živou a stále živější součástí slovotvorných procesů jak v češtině, tak v ruštině, přičemž jejich užití se liší jen velmi mírně. V češtině navíc našel uplatnění v kompozitech okazionálního charakteru, která jsou typická pro publicistický jazyk a taktéž pro jazyk internetu.

Použitá literatura:

- FILIPEC, J. (2010): *Slovník spisovné češtiny pro školu a veřejnost*. Praha: Academia.
- HAVRÁNEK, B. (1960): *Slovník spisovného jazyka českého I, A–M*. Praha: Nakladatelství Československé akademie věd.
- KRAUS, J. (2009): *Nový akademický slovník cizích slov A–Ž*. Praha: Academia.
- MARTINCOVÁ, O. (1998): *Nová slova v češtině: slovník neologizmů 1*. Praha: Academia.
- MARTINCOVÁ, O. (2004): *Nová slova v češtině: slovník neologizmů 2*. Praha: Academia.
- ЕВГЕНЬЕВА, А. П. (1981): *Словарь русского языка I. А–Й*. Москва: Русский язык.
- КРЫСИН, Л. П. (2008): *Толковый словарь иноязычных слов*. Москва: Эксмо.
- ОЖЕГОВ, С. И. (2012): *Толковый словарь русского языка*. Москва: Оникс.
- СКЛЯРЕВСКАЯ, Г. Н. (2007): *Толковый словарь русского языка начала XXI века*. Москва: Эксмо.
- УШАКОВ, Д. Н. (2007): *Большой толковый словарь современного русского языка*. Москва: Альта-Принт.

Internetové zdroje:

- Český národní korpus*. [online]. Cit. 2013-09-28. <<http://ucnk.ff.cuni.cz/>>
- Google*. [online]. Cit. 2013-09-28. <<http://www.google.cz>>

Национальный корпус русского языка. [online]. Cit. 2013-09-28. <<http://ruscorpora.ru/>>

BLENDING V RÁMCI NEOLOGICKÉ EXCERPČNÍ DATABÁZE¹

RENÁTA NEPRAŠOVÁ

oddělení současné lexikologie a lexikografie Ústavu pro jazyk český AV ČR, v. v. i.,
Letenská 4/123, Praha 1

Abstract

This paper focuses on blending as a type of mechanical compounding. The researched material results from the internal neologism excerption database of the current lexicology and lexicography department of ÚJČ AV ČR v. v. i. (Institute of the Czech Language of the AS CR, v.v.i.), whose lexical content arises by excerption of mainly journalistic texts. The paper presents the types of lexemes arised from blending, which is found in the database. The formation and processing of neologism material in the database are based on these examples. The researched material is classified to blends arised in the Czech or foreign surroundings, blends of proper names and geographical names and the unclassifiable puns. The material is analyzed in terms of semantic and by occurrence of lexemes in specific types of text.

Klíčová slova: blending, mechanické skládání slov, neologická excerpční databáze

Key words: blending, mechanical compounding, neologism excerption database

Úvodem

Príspevek se zaměřuje na blending jakožto jeden z typů mechanického skládání slov. Zkoumaný materiál vychází z interní neologické excerpční databáze oddělení současné lexikologie a lexikografie ÚJČ AV ČR v. v. i., jejíž lexikální obsah vzniká excerpčí převážně publicistických textů. V příspěvku jsou prezentovány příklady lexémů vzniklých blendingem, jež se v této databázi vyskytují. Na základě problematiky blendingu je představen vznik a zpracování neologického materiálu v databázi. Zkoumaný materiál je klasifikován na blendy vzniklé v českém nebo cizojazyčném prostředí, blendy z vlastních jmen a geografických názvů a na nezařaditelné slovní hříčky. Materiál je analyzován z hledisek, jež se dají při zápisu lexému do databáze vysledovat, tedy z hlediska sémantického a dle výskytu lexémů v konkrétních druzích textu.

Blending

Pro představení vzniku a využití neologické databáze jsme vybrali jeden z jevů, jež je systematicky excerpován pracovníky excerpčního úseku ÚJČ AV ČR, v. v. i. Tento slovotvorný jev ve shodě se Šimandlem nazýváme blending (Šimandl, 2013). Jde o skládání slov, při kterém se začátek prvního a konec druhého spojí do jednoho slova, jež připomíná slova výchozí. „S mechanickou kompozicí má blending společné to, že nehledí k morfémové stavbě slova. Zachází s východiskovými slovy – prototypicky jsou 2 – jako s řetězci elementárních znaků (fonémů, grafémů) a prolne je (přikloní je) tak, že vznikne slovo nové. Obvykle lze podle začátku výtvaru rozeznat první a podle konce druhé východiskové slovo.“ (Šimandl, 2013, s. 111).

Kromě označení blending lze tento slovotvorný postup nazvat asi třiceti různými způsoby, z nichž

1 Příspěvek vznikl v rámci grantového projektu programu aplikovaného výzkumu a vývoje národní a kulturní identity (NAKI) *Nová cesta k modernímu jednojazyčnému výkladovému slovníku současné češtiny* (DF13P01OVV011).

některé jmenuje Šimandl (2013, s. 111): portmanteau, mot-valise, nakrycie, Mischform, haplogické slovo, kontaminace. Pojmenování Šimandl uvádí podle publikace G. Booije *Morphologie: ein internationales Handbuch zur Flexion und Wortbildung* z roku 2000.

Šimandl ve svém příspěvku (2013) uvádí tradiční dělení blendu na dva typy, a to podle (ne)existence řetězce společného oběma spojeným slovům (tzv. transientu). Blendy obsahující transient se nazývají teleskopy, neobsahující transient pak věšáky. Příkladem věšáku vyskytujícího se ve sledovaném materiálu může být *vlkňátko* (vlk + jehňátko), příkladem teleskopu pak *sh-are-nting* (share + parenting).

Při zápisu lexému do neologické excerpční databáze jsou sledovány různé faktory, které pomáhají snadnějšímu vyhledávání. Jedním z nich je Interní poznámka, která usnadňuje filtrování některých jevů, které jsou excerpční schopni označit. Patří k nim kromě blendů (v databázi označených jako kombinace) také synonyma, antonyma, zdvojnásobení, přejímky z cizích jazyků, slovní spojení, frazémy a další. Excerpované lexémy jsou dále označeny slovním druhem a oborovou charakteristikou. U každého zapsaného hesla je uveden druh textu, z něhož bylo excerpováno, kontext usnadňující porozumění významu lexému a zdroj tohoto kontextu (s datem, rubrikou, stranou, autorem textu a jeho případným postavením).

Na základě faktorů, které lze v databázi sledovat, jsme se rozhodli zkoumaný materiál klasifikovat podle způsobu tvoření na blendy vzniklé v českém prostředí, blendy vzniklé v prostředí cizojazyčném, blendy z vlastních jmen a geografických názvů a nezařaditelné slovní hříčky. Blending je považován za charakteristický pro analytickou angličtinu, kde je produktivní a frekventovaný (Miššíková, 2009). Ze zkoumaného materiálu však vyplývá, že blendy mohou vznikat i v jazykovém prostředí českém.

Neologická excerpční databáze

Neologická excerpční databáze oddělení současné lexikologie a lexikografie ÚJČ AV ČR v. v. i. vzniká od roku 1991. Je rozdělena do čtrnácti vrstev, z nichž deset je přístupno veřejnosti na webové stránce www.neologismy.cz. Lexémy excerpované mezi lety 1991 a 1998 sloužily jako východisko pro *Slovník neologizmů 1*, lexémy excerpované mezi lety 1994 a 2005 pak pro *Slovník neologizmů 2*. V roce 2005 se změnil dosavadní systém excerpce, mimo jiné přibyl popis hesel pomocí Interních poznámek, které usnadňují vyhledávání v databázi. Zkoumaný materiál pro tento příspěvek vychází z posledních pěti vrstev databáze, které vznikají od roku 2005 dodnes a nejsou zatím přístupné veřejnosti. Databáze je denně rozšiřována excerpčním úsekem oddělení současné lexikologie a lexikografie ÚJČ AV ČR v. v. i. Materiál, z něhož vychází tento příspěvek, obsahuje víc jak 74 000 hesel, celá databáze pak víc než 276 000 hesel.

Konkrétní příklady blendů z databáze

Blendy podle prostředí vzniku

Mezi zkoumanými lexémy převažují blendy, jež, jak se domníváme, vznikly v českém prostředí. Do této skupiny v tomto příspěvku řadíme jak lexémy s minimálně jedním komponentem českého původu, tak blendy z lexikálních jednotek zcela zdomácnělých, tzn. přejatých, ale „přizpůsobených českému jazykovému prostředí natolik, že se vytratilo povědomí o jejich cizím původu“ (Příruční mluvnice češtiny, 2008, s. 100). Ačkoliv mají oba komponenty zkoumaných blendů často také anglický původ, je pravděpodobné, že tyto blendy vznikly až v českém prostředí.

Mezi blendy vzniklé v českém jazykovém prostředí patří s určitostí *lavičkoboard* = *lavička* + *billboard* (např. Štěpný slogan, který je třeba okamžitě vytisknout na billboardy a lavičkoboardy, narvat do všech printů, připravit bannery a spoty a ještě najmout padesát papírkářů do metra!

Právo, 13. 7. 2002); *webinář* = *web* + *seminář* (např. Pěkný den, ráda bych Vás jménem portálu Workline.cz pozvala na 2 aktuální webináře – online semináře a upozornila na 2 videozáznamy webinářů z oblastí vhodné pro neziskové organizace. Na webinář nemusíte nikam cestovat – vše se uskuteční přímo z vaší kanceláře nebo doma. Po ukončení dostanete trvale bezplatný přístup k záznamu webináře. Získané znalosti si budete moci snadno připomenout nebo je můžete zpřístupnit také svým kolegům. korespondence, 4. 2. 2010); *šuntolino* = *šunt* + *pendolino* (např. Šuntolino za velké peníze došlo # Chlouba Českých drah jménem Pendolino už není výstavní síň, už není důvodem k radosti a pýše. Vlakům Pendolino kolabuje software, a tak zastavují na každém kroku, skoro jako nějaký obyčejný courák. Listy Strakonicka, 1. 2. 2006); nebo *vlkňátko* = *vlk* + *jehňátko* (např. Prostě řečeno: strčíte-li do stejné klece vlka a ovci, nenarodí se vlkňátko. Stane se to, že vlk ovci sežere. Stejně je to se spalovnami a recyklací. Nový prostor, 17. 4. 2012). Za slovesa zmíníme lexém *čilaxovat* = výslovnost angl. *chill out* + *relaxovat* (např. V souvislosti s třetím vydáním Cameronovy monografie, která má přidanou kapitolu právě o volném čase, se dokonce objevilo nové slovo, chillaxing. Zatím nemá v jazycích pracovitějších zemí ekvivalent. Vzniklo ze dvou slov: *relaxing* – relaxování –, a *chill out*, tedy zhruba „hodit se do pohody“². ... Dosud tento výraz používaly jen děti, a nová, dospělácká verze prý podle odborníků doznala jistého významového posunu. Když čilaxují dospělí, stává se z toho rázem dynamický, aktivní způsob odpočinku. LN, 25. 5. 2012).

Jedním z lexémů inspirujících blending je *terminátor*. Příkladem může být *šrotonátor* = *šrotovat* + *terminátor* (např. Šrotonátor, postava ze vzdálené budoucnosti, má ještě dnes na plzeňském náměstí Republiky přesvědčovat pozemšťany o nutnosti recyklace vysloužilých domácích spotřebičů. Šrotonátora vytváří plzeňský výtvarník Tomáš Kůs právě z odloženého odpadu, který mohou Plzeňané na náměstí nosit až do 17. hodiny. Mladá fronta Dnes, 8. 9. 2009), nebo *sperminátor* = *spermie* + *terminátor*, a *guvernátor* (*guvernér* + *terminátor*).

Dalšími blendy vzniklými v českém prostředí jsou například lexémy *mystikál* (*mystika* + *muzikál*), *písníčkář* (*písnička* + *muzikál*) nebo *lyrikál* = *lyrika* (*lyrický*) + *muzikál* (např. Jako lyrikál označuje Michal Horáček své dílo Kudykam. Premiéru v režii Dodo Gombára bude mít 22. října ve Státní opeře. Na světě už přitom je album s písníčkami z lyrikálu, vyšla o něm i kniha a Horáček dal bezpočet rozhovorů. Zdá se, že už všechno víme – jen to vidět. Proč se na Kudykam vypravit? Nabízí tradici i inovaci, Kudykam je završením pětadvacetileté práce tvůrčího dua Hapka + Horáček. Lyrikál však vznikl opačným způsobem, než jsme u muzikálů zvyklí, což Horáček vysvětlil právě žánrem lyrikálu: v muzikálu je slovo v područí muziky, v lyrikálu je to naopak. Magazín Víkend Dnes, 17. 10. 2009).

Častým blendem je spojení se slovem *celebrita*: *weblebrita* = *webový* + *celebrita* (např. Třetí, asi nejpočetnější skupinu, tvoří začínající umělci, kterým film či televize zatím nedaly šanci. Je jim mezi patnácti a třiceti a světem tradičních médií opovrhují jako světem establishmentu. Z mnohých z nich se stávají globální „weblebrity“, celebrity na webu. Službu YouTube, která umožňuje komukoli přidat na Internet videonahrávku a komukoli jinému ji zhlédnout, spustili tři mladí Američané v listopadu 2005. V říjnu 2006 firmu koupila společnost Google za 1,65 miliardy dolarů. Reflex, 25. 6. 2009), *telebrita* = *televizní* + *celebrita*, nebo *celebutant(ka)* = *celebrita* + *debutant(ka)*.

V databázi se vyskytuje také několik lexémů obsahujících slovo *sex* (*sextáze* = *sex* + *extáze*, *sexkurze* = *sex* + *exkurze*, *sexpertka* = *sex* + *expertka*, *sexploze* = *sex* + *exploze*, *sextovat* = *sex* + *textovat*: např. Weiner provozoval tzv. „sexting“ čili textování o sexu, jemuž se podle různých průzkumů věnuje na Západě alespoň někdy skoro polovina mladé populace, ať již svobodné či sezdané. Weiner sex-

2 Uvozovky uvedené v příkladových textech vycházejí ze zdrojových textů.

toval dosti amatérsky (kvalita jeho snímků byla ubohá) a navíc hazardně (přes Facebook a Twitter), taky už není zrovna nejmladší (46 let). I tak by mezi vrstevníky určitě našel spoustu kolegů, z nichž mnozí v rámci virtuálního flirtu poslali vlastní vyboulené trenýrky. Mladá fronta Dnes, 17. 6. 2011).

Častým komponentem je výraz *automat*. Příkladem může být *modlomat* = *modlitební* + *automat* (např. Padá na vás svět, jste uprostřed nákupů a kostel je daleko? V Německu se můžete odtrhnout od denních starostí v „modlomu“ nebo modlitebním automatu. Najdete ho v Berlíně a ve Frankfurtu nad Mohanem. Budky připomínající kabiny na focení mají v repertoáru na 300 modliteb v 65 řečech. Zatím bez češtiny. Jejich tvůrce však s ní počítá a nevylučuje, že by se v rámci expanze do zahraničí mohl takový moderní minikostel objevit i v Praze. Mladá fronta Dnes, 5. 9. 2011), nebo *kondomat* = *kondom* + *automat* (např. V Itálii má katolická církev velmi silné slovo. Kondomy jsou zde sice běžně dostupné v obchodech, například v reklamních kampaních se ale snímky, na kterých je pánská ochrana, objevují jen zřídkakdy. Církev totiž považuje jakoukoli veřejnou zmínku o kondomech za podněcování k nemorálnímu a požitkářskému způsobu života. Umístění prvního „kondomatu“ v italské střední škole okomentovala i Asociace italských studentů. „Jenom v Itálii může takovýto krok způsobit rozruch,“ píše se v jejím prohlášení. „V západních zemích množství výchovných institucí distribuuje kondomy, stejně tak je tomu i ve školách ve Spojených státech,“ dodala. Lidové noviny, 12. 3. 2010), dále pak *tankautomat* = *tankovací* + *automat*; *platbomat* = *platba* (*platební*) + *automat*; *půjčkomat* = *půjčka* (*půjčkový*) + *automat*; *slevomat* = *sleva* (*slevový*) + *automat*; *nosičomat* = *nosič* + *automat*; *kávomat* = *káva* + *automat*; *burčákomat* = *burčák* + *automat*, *balíkomat* = *balík* + *automat*; *parkomat* = *parkovací* + *automat*; *mlékomat* = *mléko* + *automat*; *zlatomat* = *zlato* + *automat*; *pivomat* = *pivo* + *automat* a další.

Některé z následujících příkladů blendů nespíše také vznikly v českém prostředí, mohly by se však vyskytovat také v prostředí cizojazyčném: *biotekt* = *bio* + *architekt*; *cybrid* = *cytoplazmatický* + *hybrid* (např. Do této v budoucnu legální skupiny spadají různé typy smíšených stvoření. Například buněčné hybridy (tzv. cybridy): K jejich získávání se lidská dědičná substance klonuje ve zvířecích vaječných buňkách obsahujících zhruba jedno procento vlastního genetického materiálu ve formě DNA mitochondrií. Z tohoto spojení vznikají převážně lidská embrya. Od cybridů je třeba odlišovat skutečné chiméry. Ty vznikají spojením lidských spermií například s vajíčkem králíka. GEO, 23. 10. 2008); *banskter* = *bankéř* + *gangster* (např. Například vznikla nová složenina „bankster“, kombinace slov „bankéř“ a „gangster“. Nejde o běžné bankéře; kdyby ano, nebylo by zapotřebí nového slova. Ne, tyto borci nalili miliardy do vysoce spekulativních, neregulovaných investičních podniků, o všechno na burze nevyhnutelně přišli, dostali další bezúročné miliardy od svých vlád, aby zůstali v plusu, odmítli komukoli prodloužit úvěr a místo toho zvedli poplatky za kreditky, aby si mohli vyplatit tučné miliardové bonusy. E15, 11. 1. 2010); *greygler* = *grey* + *googler* (např. Petra proto vítá podnikovou kulturu svého nynějšího zaměstnavatele. „Máme tu tzv. podpůrné skupiny, ve kterých se sdružují zaměstnanci s podobným společenským postavením. Máme tu skupinu gayglerů, což jsou zaměstnanci gayové, máme skupinu černých googlerů, skupiny greyglerů, tedy starších zaměstnanců, ale taky asijských zaměstnanců a samozřejmě ženskou skupinu.“ Pátek Lidových novin, 11. 12. 2009); *improshow* = *improvizace* + *show*; nebo *lovotika* = *love* + *robotika* (např. Vědci z Institutu budoucnosti lidstva na Oxfordu předpověděli, že počítače by mohly dosáhnout lidské inteligence už v roce 2028 (s pravděpodobností naplnění prognózy 10 procent). Až se to stane, bude možné vytvořit roboty podobné lidem. Podle dalšího článku na seriózním serveru prý takoví roboti nebudou moc dobří v konverzaci, ale byly by z nich dobré roboprostitutky. Novou oblast vědy, zvanou „lovotika“ – love + robotika – už prý pár let (!) zkoumá na Národní univerzitě v Singapuru vědec Hooman Samani. Víkend Dnes, 26. 1. 2013).

Mnohé zkoumané blendy vznikly v cizojazyčném (převážně anglickém) prostředí a byly přejaty do češtiny jako hotové lexémy. Za všechny uvádíme několik příkladů: *phablet* = *phone* + *tablet* (např. Chytré telefony jsou rok od roku větší a pomalu se přibližují počítačovým tabletům. Zařízením, která kombinují výhody obou těchto kategorií, se říká phablety a podle prognóz prodejců letos zaplaví trh. Většina chytrých telefonů má displej s úhlopříčkou od tří do pěti palců. Ty nejmenší počítačové tablety jsou zase vybaveny sedmipalcovými dotykovými monitory. Prostor mezi oběma kategoriemi vyplňují již zmiňované phablety. Toto označení vzniklo spojením slov *phone* (telefon) a *tablet*. Právo, 21. 1. 2013), *sharenting* = *share* + *parenting* (např. Co je to sharenting. Nové slovo vzniklé z kombinace *parenting* (rodičovství) a *sharing* (sdílení, ve významu fotek na internetu). Nebezpečí přílišné otevřenosti rodičů je, že fotky jejich dítěte lze snadno zneužít. To, co vám dnes přijde vtipné, nemusí přijít vtipné vašemu dítěti, až začne chodit do školy. Ona Dnes, 20. 5. 2013), nebo *glamping* = *glamorous* + *camping* (např. Mě víc baví bydlet ve stanech než ve zděných srubech. Je to pro mě větší dobrodružství. Luxus glampingu spočívá v perfektním bydlení – usínáte za praskání ohně v krbu nebo před stanem – jídle – ráno vám zaměstnanec přinese kávu do stanu – a bezpečí, ale nijak to neovlivní, co se člověku během pobytu stane. ... Ano, taková glampingová cesta do Afriky je drahá. Ale v té vysoké ceně je zahrnuto třikrát denně vynikající jídlo včetně pikniku v přírodě, auto, průvodce, ochrana, a to všechno perfektní. Proč ne?!, 21. 2. 2013), dále pak *smishing* = *SMS* (*Short Message Service*) + *phishing*, *mockument* = *mock* + *document*.

Objevují se také blendy s komponentem *bikiny*: *facekini* = *face* + *bikini*, *fatkini* = *fat* + *bikini* (např. Letos v létě umístila Gabi Gregg, XXL blogerka z Chicaga, na své stránky fotku v plavkách. Má americkou velikost 18, tedy zhruba českou dvaapadesátku, přesto se (považte!) vyfotila v černobílých proužkovaných bikinách, kterým sama vtipně říká „fatkini“, a hrdě se vystavila na blogu. Ona Dnes, 3. 12. 2012), nebo *burkini* = *burka* + *bikini*.

Blendy z vlastních jmen a geografických názvů

Mezi blendy přináležejícím k životným maskulinům se často objevují lexémy z vlastních jmen: *Medveděputin* = *Medveděv* + *Putin*, *Paroulánek* = *Paroubek* + *Topolánek*; *Toparoubek* = *Topolánek* + *Paroubek*, *Parolánek* = *Paroubek* + *Topolánek*, *Merkozy* = *Merkelová* + *Sarkozy*, *Škromachr* = *Škromach* + *machr* (např. Škromachr # Místopředseda ČSSD Zdeněk Škromach, prototyp českého socialisty, minulý týden pravil: „ODS a její soupeřníci, lidovci a zelení, by měli konečně začít držet hubu, když ČSSD přichází s návrhy, které by měly posílit například jistoty, kdy ne vlastní vinou lidé ztratí zaměstnání, a kdy hovoří, že sociální demokracie něco rozhazuje. Je to holt chlap od rány a už ho celá ta demokracie štvě.“ Týden, 11. 5. 2009), *obamatér* = *Obama* + *amatér* nebo geografických názvů: *Prahajzl* = *Praha* + *hajzl* (např. Pražáky to příliš nepotěší, ale ve Vsetíně je zažitý výraz Prahajzl. Značí člověka kluzkého, povýšeného, který jedno dělá a druhé říká. Slovo nedrží. Své problémy přičítá Čunek kromě jiného právě vlivu tohoto podrazáckého pražského prostředí. Trvalo mu to, ale prý už se přizpůsobil. Mladá fronta Dnes, 31. 5. 2008).

Také mezi substantivy ženského rodu se vyskytují blendy z vlastních jmen, mimo již jmenovaných například *sarkomorfoza* = *Sarkozy* + *metamorfoza* (např. Podivné sarkomorfozy # I bez Dominique Strausse-Kahna mají socialisté velkou šanci příští rok porazit Nicolase Sarkozyho. Ten prezidentské křeslo hájí proti levicové hrozbě jejich vlastními zbraněmi a mnohde se odklání od dosavadního stylu i programu. ... „Osm měsíců před prezidentskými volbami Nicolas Sarkozy započal svou poslední metamorfozu,“ píše Le Parisien. „Plán úspor značí konec jedné epochy.“ I když vystupoval jako přísný pravičák, už od začátku svého panování v Elysejském paláci dal Sarkozy místo také osobnostem z druhé strany spektra. Ministrem zahraničí byl takto v prvních letech Bernard Kouchner (spoluzakladatel organizace Lékaři bez hranic), jehož pak Socialistická strana (PS) promptně vyloučila. Post ministra pro města zase zastávala Fadela Amaraová, rovněž spjatá

se socialisty. I když teď Sarkozy, respektive premiér Francois Fillon, levičáky v kabinetu nemá, kurs se po pravicových počátcích stáčí lehce doleva. Týden, 5. 9. 2011), *bachmosféra* (příjmení něm. hud. skladatele *Bacha* + *atmosféra*), nebo *slotalita* (příjmení politika *Sloty* + *totalita*), *serenátorka* (jméno tenistky *Sereny Williamsové* + *terminátorka*). Lze nalézt také blendy z geografických názvů *kubazuela* = *Kuba* + *Venezuela*, *Eurábie* = *EU (Evropská Unie)* + *Arábie*, *Čimerika* = *Čína* + *Amerika*, *Čínamerika* = *Čína* + *Amerika*, *Čindie* = *Čína* + *Indie*).

Z neuter jmenujme *bártapo* = příjmení politika *Bárty* + *gestapo*; nebo z blendů podle geografických názvů *Pigstugalsko* = akronym *PIGS (Portugalsko, Itálie / Irsko, Řecko, Španělsko)* + *Portugalsko*. Mezi přídavnými jmény jsou to pak například *obamerický* = příjmení prezidenta *Obamy* + *americký* a *čexický* = *český* + *mexický*.

Častým slovem inspirujícím spojování do blendu je *ekonomika*. Z četných příkladů vybíráme pouze několik: *vieweghonomika* = příjmení spisovatele *Viewegha* + *ekonomika* (např. *Vieweghonomika*, jak se přezdívá byznysu okolo knih a filmů nejprodávanějšího současného českého spisovatele, dostala nový rozměr. Po více než dvou desítkách knih a řadě snímků, z nichž *Román pro ženy*, *Román pro muže*, *Nestydu* či *Účastníky zájezdu* lze bez nadsázky označit za kinohity, nachystala jedna z předních pražských scén *Vieweghovi* poctu v podobě šedesátidílné série jeho jevištních adaptací. *Hospodářské noviny*, 22. 6. 2011), *obamonomika / obamanomika* = příjmení amerického prezidenta *Obamy* + *ekonomika*, *spearsonomika* = příjmení zpěvačky *Spears* + *ekonomika*, *hřebejkonomika* = příjmení režiséra *Hřebejka* + *ekonomika*.

Slovní hříčky

V databázi se vyskytují také slovní hříčky, jež nelze zařadit do Šimandlovy typologie věšák x teleskop. Nejsou typickými blendy. V souvislosti s již zmiňovaným lexémem *terminátor* vznikla také slovní hříčka *Tourminátor* – tento blend využívá fonetické podobnosti mezi předponou slova *terminátor* a slovem *Tour*. Dalšími příklady mohou být výrazy *šloufl* = spojení příjmení českého politika *Miroslava Šloufa* + slova *šoufl*, jež je založené na fonetické podobnosti obou slov (např. *Ano, kníže prohrál. Což neznamená, že ti, kdo ho volili, mají propadnout depresím. Je čas zasypat i v sobě vykopané příkopy. A třeba se začít vzájemně bavit vtipy, které se vyrojily. ... Někomu tak může být z volby prezidenta „šloufl“.* Jiný si může přidat do oblíbených stránek tu s názvem *zemancountdown.cz* a utěšovat se tím, jak se novému prezidentovi odečítají dny, hodiny a vteřiny ve funkci. *MfD*, 28. 1. 2013) nebo *eStBlishment* (spojení slova *establishment* a zkratky *StB (dř. Státní Bezpečnost)* založené opět na fonetické podobnosti. Velká písmena uvnitř slova zajišťují srozumitelnost spojení obou slov a zaručují, že nejde o překlep.

Stejně jako u maskulin i u feminin vznikají slovní hříčky jako například *cooltura* (*cool* + *kultura*), *deppka* (příjmení herce *Deppa* + *depka*), *Fišed'* (příjmení politika *Fischera* + *šed'*).

Za neutra jmenujme například *Špatnělsko* = *špatný* + *Španělsko*. Dále pak *ČEZsko* nebo *ČEZko* (*ČEZ (České energetické závody)* + *Česko*) – tento blend je rozpoznatelný pouze v psané podobě, v mluvené by musel být opsán (Šimandl, 2013).

Také přídavná jména obsahují některé slovní hříčky (*coolturní* = *cool* + *kulturní*, *cloudný* – využívající fonetické podobnosti mezi angl. slovem *cloud* a českým *kloudný*, *on-líný* = *on-line* + *líný*).

Vznikají také slovesné slovní hříčky, například *doš(o)urat se* = příjmení fotbalisty *Šurala* + *došourat se* (např. *Zleva centroval Breznaník, Šural ve skluzu zasáhl míč, který se doš(o)ural se štěstím k Blažkovi a ten měl z brankové čáry lehkou práci. Mladá fronta Dnes*, 25. 7. 2012), *blobnout* = *blob (Binary Large Objects)* + *blbnout*.

Závěrem

Vzhledem k postupu zápisu lexémů do neologické databáze je výskyt blendů v ní poměrně nevýrazný. Excerptoři při své práci sledují mnoho jevů, kromě nově utvořených slov jsou excerpovány také neosémantismy, zdrobněliny nezaznamenané ve slovnících, slova přechýlená, přejímky a další. Označené blendy jsou tedy ve zkoumaném materiálu zastoupeny hodnotou přibližně 0,5 procenta.

Frekvence užití zkoumaných blendů není z databáze zjistitelná, jelikož databáze zachycuje pouze dynamiku, nikoliv frekvenci. Excerptoři jsou omezeni tak, že každý nalezený lexém může být do databáze zapsán maximálně v pěti různých kontextech. V materiálu se tedy objevují blendy excerpované pouze jedenkrát a blendy excerpované ze dvou až maximálně pěti kontextů. Frekvenci užití by bylo možné zjistit porovnáním se záznamy v Českém národním korpusu SYN, jež jsme pro potřeby představení neologické databáze neprováděli.

Jak bylo již zmíněno, ve zkoumaném materiálu převažují blendy vzniklé v českém prostředí nad těmi, jež vznikly v prostředí cizojazyčném a byly přejaty jako hotové lexémy. Časté jsou blendy, kde je alespoň jedna jejich část odvozena z vlastního jména nebo geografického názvu.

Jak bylo uvedeno, v databázi je také zaznamenáno, z jakého druhu textu je lexém excerpován. Nejčastěji jsou zkoumané blendy excerpovány z článků³ (29 %). Velmi často se blendy vyskytují v publicistické zprávě (18 %), nebo v komentářích (12 %), dále pak v glosách (7 %) a reportážích, recenzích a rozhovorech (shodně po cca 4 %). Výjimečně se objevují v úvahách a esejích (po 2 %) a ve fejetonech (1 %).

Pozn.: Příspěvek vznikl s použitím interního elektronického neologického archivu Neomat oddělení současné lexikologie a lexikografie ÚJČ AV ČR, v. v. i. (k 23. 9. 2013).

Bibliografie:

- MIŠŠÍKOVÁ, G. (2009): Využitie okrajových slovtvorných postupov v slovenčině. In: *Slovo – Tvorba – Dynamickosť*. Bratislava: Jazykovedný ústav Ľudovíta Štúra SAV, s. 136–142.
- ŠIMANDL, J. (2013): Mechanické krácení a mechanické skládání; blending. In: *ПРОБЛЕМИ НА НЕОЛОГИЯТА В СЛАВЯНСКИТЕ ЕЗИЦИ (Problems of neology in slavic languages)*. Sofia: БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ (Bulgarian Academy of Sciences), s. 109–118.
- ŠMILAUER, V. (1971): *Novočeské tvoření slov*. Praha: SPN.
- Příruční mluvnice češtiny* (2008). Praha: Nakladatelství Lidové noviny.

3 Jako články jsou v databázi označovány také některé druhy textů, které lze žánrově zařadit mezi analytické žánry publicistiky, ale zároveň je excerptor není schopen označit jako typický komentář, úvahu, glosu či další druh textu. Je možné, že proto je jejich počet nejvyšší.

INTERNET A NOVÁ SLOVESNÁ POJMENOVÁNÍ¹

ZUZANA DĚNGEOVÁ

Ústav českého jazyka FF MU Brno

Abstract

The paper focuses on the description of neologisms warranted by proper names, functioning as depropriat words or expressions. The analyzed material comes from journalistic texts on the Internet and comprise new verbs created from the proper names of politicians.

Klíčová slova: neologismus, okazionalismus, depropriální pojmenování, slovtvorné procesy v českém jazyce

Keywords: neologism, occasionalism, depropriat words, derivational process in the Czech language

Úvod

V příspěvku se zaměříme na tvoření nových slov, která mají v pojmenovacím základu vlastní jméno osoby. Představíme lexikální materiál depropriálních sloves, která jsou motivována vlastními jmény, a ukážeme, podle kterých slovtvorných typů (modelů) se daná slovesa tvoří, v jakých komunikátech se vyskytují a jak se uplatňují ve slovní zásobě. Soustředíme se na slovesa odvozená od příjmení politiků, např.: *dodobešovat*, *odjulínkovat*, *langerizovat*, *zemanovatět*; zřídka se jako pojmenovací základ depropriálních sloves uplatňuje také křestní jméno: *obarákovat* (podle jména amerického prezidenta Baracka Obamy)². Tvoření depropriálních pojmenování se jeví jako častý aktualizací prvek v publicistice a hojně se s takto vznikajícími lexémy setkáme zejména v textech na internetu, přičemž internet a nová média se jeví jako významný materiálový zdroj pro zachycování a analýzu daných lexémů.

1 Depropriální tvoření slov

Depropriální tvoření slov je v češtině obvyklý a produktivní slovtvorný jev, při němž se uplatňuje množství slovtvorných prostředků a využívá se široké spektrum různých propriálních motivantů. Toto tvoření se výrazně uplatňuje zejména v publicistice, a to zvláště v několika posledních desetiletích, což souvisí s charakterem porevoluční společnosti jako takové – politická scéna se rychle proměňuje, publicistika na ni musí svižně a aktuálně reagovat a jako vhodné aktualizací prostředky se jeví lexémy tvořené z vlastních jmen.

Zvýšené tvoření a užívání depropriálních pojmenování lze odhalit nejen z excerpce materiálu, ale také např. české slovníky neologismů *Nová slova v češtině 1* (1998) a *Nová slova v češtině 2* (2004) věnovaly depropriálním pojmenováním motivovaným jmény politiků výraznou pozornost (např. slovníky uvádí 31 lexémů motivovaných jménem politika Václava Klause: *antiklausovec*, *antiklausovsky*, *antiklausovský*, *antiklausovství*, *klausianství*, *klausismus*, *klausista*, *klausismus*, *klausostrana*, *klausování*, *klausovat*, *klausovatět*, *klausovatý*, *klausovec*, *klausovitě*, *klausovitý*,

1 Příspěvek vznikl v rámci grantového projektu specifického výzkumu MUNI/A/1165/2014 *Čeština v jednotě synchronie a diachronie 2015*.

2 Křestní jména se jako motivanty uplatňují výjimečně, neboť na rozdíl od příjmení nemohou jednoznačně identifikovat danou osobu.

*klausovka, klausovsko-zemanovský, klausovsky, klausovský, neklausovec, neklausovsky, neklausovský, odklausování, poklausovsku, poklausovský, postklausovský, proklausovsky, proklausovský, proklau-
sovsky, protiklausovský, předklausovský*). Pozornost tomuto tématu je věnována i v odborných
studiích např. Martinová (2003, 2011); Michalec (2013); Opavská (2003); Děngeová (2010, 2011).
Na druhou stranu starší mluvnické a slovtvorné popisy tento jev zaznamenávají jen omezeně.
Na tento fakt upozorňuje i Olga Martinová, která v článku Depropriální neologismy v současné
češtině (2011) poznamenává, že v mluvnicích je zaznamenání depropriálních slov značně mezero-
vitě (pouze adjektivům je věnována výraznější pozornost). Ze slovtvorných popisů se podrobněji
depropriálnímu tvoření v oblasti substantiv věnuje Tvoření slov v češtině 2 (1967), výrazněji se
na tento jev zaměřuje Novočeské tvoření slov (1971), kde je zmíněno i tvoření sloves ze jmen lidí,
jako příklady jsou uvedena slovesa *panáčkovat, koniášovat, švejkovat, pasteurovat* (Šmilauer, 1971,
s. 153).

I přes neúplné zaznamenání tohoto jevu v mluvnických a slovtvorných popisech je zřejmé, že
depropriální slovtvorba je významnou součástí jazyka; nová pojmenování se tvoří jednak podle
produktivních slovtvorných typů (které se běžně uplatňují i při obohacování apelativní slovní
zásoby), jednak depropriální slovtvorba využívá i okrajové (nové) slovtvorné prostředky, srov.
široké slovtvorné hnízdo s pojmenovacím základem příjmení politika Jiřího Čunka zahrnující jak
pojmenování tvořená běžně užívanými slovtvornými prostředky: *čunkovec, čunkovský, čunkovsky,
čunkovat, antičunkovský, protičunkovský* ap.; tak pojmenování tvořená okrajovými slovtvornými
postupy: *čunekgate, čunkoland, čunkistán, čunkofleky, čunkobox, čuntejner* aj. (Děngeová, 2011).
Mnohé z takto vznikajících slov zůstávají na okraji slovní zásoby, užívají se jen po omezenou
dobu, přesto se depropriální tvoření jeví jako výrazný způsob obohacování slovní zásoby češtiny
a považujeme za případné věnovat pozornost popisu specifík depropriální slovtvorby.

2 Analýza materiálu

Depropriální slovesa z našeho lexikálního materiálu³ mají individuální význam, který je realizován
v rámci obecného významu slovtvorného. Slovesa, která jsme zachytili, se tvoří pomocí těchto
formantů: -ova-/-uj- (*bémovat*), primárně se slovesa tvoří ve vzoru *kupovat*, od nich se pak mo-
hou tvořit slovesa ve vzoru *sázet* (*grossovat* → *grossovatět*); pře + -ova-/-uj- (*přeparoubkovat*); od +
-ova-/-uj- (*odjulínkovat*); vy- + -ova-/-uj- (*vybártovat*); -ovatět-/-ovatěj- (*havlovatět*); z- + -ovatět-
/-ovatěj- (*zgrossovatět*); v menší míře se také uplatňuje cizí kmenotvorný formant, který se obvykle
pro tvoření od domácích základů nevyužívá: -izova-/-izuj- (*paroubkizovat*); z + -izovat-/-izuj- (*zp-
aroubkizovat*); de + -izova-/-izuj- (*delangerizovat*).

Nová depropriální slovesa vznikají s individuálním významem a konkrétní význam sloves nemusí
být zřetelný ani z širšího jazykového kontextu. Význam úzce souvisí s mimojazykovým kontextem,
s osobou, jejíž jméno je pojmenovacím základem, s jejími kauzami, s názory, které zastává, s po-
litickou stranou, se kterou je osoba spojována ap. To také způsobuje efemérní charakter daných
sloves, neboť se zánikem mimojazykové skutečnosti, která je součástí pojmenovacího motivu, se
význam sloves často stává nesrozumitelným.

3 Lexikální materiál jsme získali cílenou excerpcí internetových článků (zdroje jsou uvedeny
za příkladovými texty), pro dohledávání materiálu jsme využívali také Český národní
korpus dostupný z <<http://www.korpus.cz>>, využívali jsme také lexikální materiál interního
elektronického archivu Neomat oddělení současné lexikologie a lexikografie ÚJČ AV ČR, v. v. i.,
u příkladů z této databáze je v závorce uvedeno Neomat.

Nejčastěji jsme se v materiálu setkali se slovesy s přirovnávacím významem⁴ ,počínat si, vést si, chovat se jako nositel daného jména', např. u slovesa *bémovat* je obecný strukturní význam ,chovat se jako politik Pavel Bém', individuální / kontextový význam je ,vykrucovat se, lhát, jak je typické pro Pavla Béma': *Dovolili za soutěž o podobu knihovny utratit 28 milionů, přislíbili Vlastimilu Ježkovi, šéfovi NK, peníze na stavbu, Praha mu přiklepla pozemek na Letné, primátor Bém byl pro, ale pak předvedl kotrmelec s tolika vruty, že by se za ně Aleš Valenta nemusel stydět a na olympijské zlato hravě by dosáhl. Otáčet na obrtlíku mohlo by se nově česky říci bémovat..* (www.tyden.cz)

Individuálnost kontextového významu dokládá také sloveso *čunkovat*, které ve dvou různých příkladech nabývá různého významu: *Podrobnosti může laskavý čtenář najít v příslušných hospodářských rubrikách médií, ale zhruba jde o to, že se jaksí nedaří přestat „čunkovat“ (někoho překabátit, v tomto případě EU špatně nastavenými domácími pravidly pro čerpání dotací) a „Evropa“ nám na to nějak nechce skočit.* (SYN); [...] sloveso „čunkovat“: Slovo má přitom dva významy: *vykrucovat se z nepříjemných dotazů, či vystěhovávat romské spoluobčany.* (www.zpravy.idnes.cz)

Setkáme se také se slovesy s obecným strukturním významem ,stávat se, stát se podobným nositeli daného jména, připodobňovat se mu / nabývat, nabýt vlastnosti, jakými se vyznačuje nositel daného jména'. K porozumění individuálnímu významu je opět nutné znát širší textový kontext a mimojazykový kontext, zejména vztah a charakter obou osob, které bývají porovnávány: *Klaus, který nám – zdá se – začíná na Hradě trochu havlovatět.* (www.tyden.cz); *„Pro mě je šokující, že doposud uvážlivý ministr Heger nám nějak julínkovatí,“ řekl stínový ministr zdravotnictví David Rath v narážce na exministra Julínka, který poplatky zavedl.“* (Neomat); *Nyní sociální demokraté vtipkují, že Paroubek „zemanovatí“. Stejně jako kdysi Zeman, i Paroubek po volbách pořád opakoval, že nic nechce.* (Neomat)

Mnohdy se význam sloves vztahuje nejen ke konkrétní osobě, ale celkově k politice, kterou představuje, či k období, v němž politik působil, s nímž je spojován, např. sloveso *zfcovatět* má v daném kontextu význam ,tvořit systém podobným tomu, který představuje vláda vedená Robertem Ficem'⁵: *„Rád bych ČSSD ,zfcovatěl', aby se tu mohlo odehrát něco jako na Slovensku,“ řekl Földyna s odkazem na slovenského premiéra Roberta Fica, který po zisku 44 procent Smeru-SD vede jednobarevnou vládu.* (www.novinky.cz); sloveso *zbursíkovatět* je užito ve významu ,strana bude mít podobný charakter, jako měla za období vedení politikem Martinem Bursíkem': *Skupina zelených, která drží linii zpochybňování měšťáckého života a zaběhnutých autorit, tedy ideje, z nichž zelené hnutí vzešlo, se potýká s obdobným problémem. Aby ovládla nejen republikovou radu, musí přijít s reálnými sliby konformním spolustraníkům. V tom okamžiku by ale zbursíkovatěla a ztratila smysl existence.* (Neomat).

Uplatňuje se také tvoření sloves s obecným významem ,zbavit se (úřad / politickou scénu) vlastností (rázu), s jakým(i) je spojován nositel daného jména' a ,zbavit úřad / politickou scénu nositele daného jména': *Situaci ve zdravotnickém resortu je třeba uklidnit, odjulínkovat, odcikrtovat.* (Neomat); *Když Petr Jan Kalaš přebíral po Liboru Ambrozkovi ministerstvo životního prostředí, dostal od tehdejšího premiéra Mirka Topolánka úkol – „odbaierizovat“ ministerstvo.* (Neomat); *Koalice je křehká jako čínský porcelán, třtina nalomená, a premiér, v jehož silách není vládu „odčunkovat“, nebude plýtvat energií na to, aby se postavil všem regionálním politikům, to jest skalám, na nichž jsou vybudovány partaje.* (www.tyden.cz)

Pro vyjádření významu ,zbavit se (úřad / politickou scénu) vlastností (rázu), s jakým(i) je spojován

4 Při klasifikaci sloves a výkladu strukturního významu vycházíme z textu O. Martincové *Depropriální neologismy v současné češtině* (2011, s. 33–34).

5 Jsme si vědomi, že naše výklady nemusí být vyčerpávající a nepostihují všechny významové odstíny daných sloves.

nositel daného jména⁶ jsou slovesa tvořena cizím prefixem de-: *ČSSD má totiž jasný cíl: odpolitizovat policii, delangerizovat ji.* (www.zpravy.idnes.cz); *Je nejvyšší čas ,derathizovat⁶ všechny naše profesní organizace.* (www.zpravy.idnes.cz). Pro vyjádření zmíněného významu (odstranit daného politika z vedení strany, z funkce ap.) se využívá také domácího prefixu vy-: *Navrhuji Bártu „vybárovat“ raději na druhou stranu zeměkoule, tedy směrem na Východ.* (Neomat)

Výraznou skupinu představují slovesná pojmenování s prefixem pře-, naznačující přehnanou míru. Význam daných lexémů lze vyložit jako ,je přesmíru, přespříliš nositele daného jména / jeho vlivu / názorů‘ (hojně se tento typ uplatňuje v oblasti kulturní publicistiky: *přebecktovat, přečechovat, přečtvrtníčkovat, přegottovat, předrassovat, překrausovat, přemozartovat, přemiüllerovat, přeštreitovat* aj.). Takto tvořené slovesné tvary se vyskytují zpravidla ve tvaru pasivního participia, v textu mají často funkci titulku / mezititulku: *Přebabišováno. Čas od času se stane, že na titulních stranách největších deníků v zemi vyjde v týž den velmi podobná fotografie. [...] Tentokrát se to podařilo Reflexu, Respektu a Euru, které své tituly ozdobily portréty miliardáře Babiše.* (www.zpravy.idnes.cz); *Tiskový mluvčí ODS Jiří Sezemský nejdříve vzkázal, že „strana nepovažuje za příliš produktivní, aby bylo přeparoubkováno a přetopolánkováno, a zvláště nebude podporovat předvolební divadla, odhalující monarchistické sklony současného pana předsedy vlády“.* (Neomat); *Volali mi z Reflexu, že o Richtrovi a situaci na magistrátě psát nemám, protože už prý mají jaksi přerichterováno a zamagistrováno.* (Neomat).

3 Depropriální slovesa a jejich postavení ve slovní zásobě

Depropriální slovesa jsou z hlediska slovní zásoby lexikálními prostředky periferními, zpravidla se nestávají trvalou součástí slovní zásoby (na rozdíl od mnohých depropriálních adjektiv a substantiv). V porovnání s depropriálními jmény se také tvoří mnohem méně často a ponechávají si okazionální charakter.

Charakteristická je pro ně také krátkodobost užití a výrazná vázanost na dobový kontext – životnost těchto lexémů závisí na životnosti označovaného jevu a se zánikem mimojazykové skutečnosti, kterou označují, zaniká i dané slovo a schopnost porozumění jeho významu. Je jisté, že význam některých sloves, uvedených v tomto příspěvku, již nemusí být pro čtenáře srozumitelný.

Závěr

V příspěvku jsme představili jeden typ tvoření nových slovesných pojmenování, popsali jsme slovtvorné postupy, kterými se slovesa tvoří, jakých významů nabývají, v kterých komunikátech se vyskytují a jak se uplatňují ve slovní zásobě. Analýzou materiálu jsme chtěli upozornit na vznikání těchto sloves s individuálním významem a na jejich efemérní charakter. I přes krátkodobost užití a okazionální charakter depropriálních sloves jsme se pokusili dokázat, že depropriální slovtvorba představuje výraznou tendenci v oblasti obohacování slovní zásoby, jejíž užívání vzrůstá a je záhodno tomuto způsobu tvoření nových slov věnovat pozornost.

6 Expresivnost slovesa ,derathizovat‘ je zvýšena podobností se slovem deratizace ,hubení škodlivých hlodavců‘.

Použitá literatura

- BUZÁSSYOVÁ, K. (2008): Jazyk a spoločnosť: krátka „kariéra“ niektorých slov a významov slov (Príspevok k vymedzeniu pojmu efemérne slovo). In: *Jazyk a jazykoveda v pohybe*. Na počesť Slavomíra Ondrejoviča. Ed. S. Mislovičová. Bratislava: Veda, s. 245–255.
- DĚNGEOVÁ, Z. (2010): Nové lexikální jednotky motivované jmény českých politiků. *Bohemistika*, č. 3, s. 167–186.
- DĚNGEOVÁ, Z. (2011): Nová pojmenování motivovaná příjmením Jiřího Čunka. *Naše řeč*, 94, č. 5, s. 267–272.
- DOKULIL, M. a kol. (1967): *Tvoření slov v češtině 2*. Praha: Academia.
- Korpus SYN: Český národní korpus – SYN [online]. Praha: Ústav Českého národního korpusu FF UK. Cit. 2013-11-20, dostupné z <<http://www.korpus.cz>>.
- MARTINCOVÁ, O. (2011): Depropriální neologismy v současné češtině. In: *Jazykovedné štúdie XXIX (Život medzi apelatívami a propriami)*. Bratislava: Veda, s. 20–37.
- MARTINCOVÁ, O. a kol. (2005): *Neologizmy v dnešní češtině*. Praha: Ústav pro jazyk český Akademie věd České republiky.
- MARTINCOVÁ, O. (2003): Propria v neologické slovo tvorbě. In: *Dynamika a inovace v češtině a bulharštině 90. léta 20. století*. Praha: Praha: Ústav pro jazyk český, 2003, s. 19–25.
- MARTINCOVÁ, O. a kol. (1998): *Nová slova v češtině: slovník neologizmů*. Praha: Academia.
- MARTINCOVÁ, O. a kol. (2004): *Nová slova v češtině: slovník neologizmů 2*. Praha: Academia.
- Neomat* [online interní databáze]. Verze 2.0. Praha: Ústav pro jazyk český AV ČR, v. v. i. Cit. 2013-11-20.
- МИЧАЛЕЦ, В. (2013): Depropriální neologismy tvořené z příjmení. In: *ПРОБЛЕМИ НА НЕОЛОГИЯТА В СЛАВЯНСКИТЕ ЕЗИЦИ (Problems of neology in slavic languages)*. Sofia: БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ (Bulgarian Academy of Sciences), s. 153–163.
- ОПАНСКА, З.: Některá pojmenování užívaná odpůrci a příznivci. In: *Varia X*. Bratislava: Slovenská jazykovedná spoločnosť pri SAV, 2003, s. 9–12.
- ŠMILAUER, V. (1971): *Novočekské tvoření slov*. Praha: Státní pedagogické nakladatelství.

BLOG JAKO ŽÁNŘ POTVRZUJÍCÍ DYNAMICKÝ VÝVOJ KOMUNIKAČNÍCH MOŽNOSTÍ SOUČASNOSTI¹

JARMIL VEPŘEK

Ústav českého jazyka FF MU Brno

Abstract

Until today, my linguistic works were concerned with blogs as a specific kind of texts in the age of internet. This study – based on my undergraduate thesis – is an attempt to sketch them briefly, with respect to suggested length of this paper.

The specificity of blogs as texts is that – even though they appear in communication sphere without any rules (norms) of expression – several features of style similarity can be found if they are compared in adequate quantity. However, this possibility depends on the way these texts are selected and analyzed; both need to be explained here preferably. Subsequently, particular analytic results will be presented to support the main hypothesis of this work which says – blog can be understood as a new genre of journalism and as a proof of dynamic evolution of today's communication possibilities.

Klíčová slova: blogy, subjektivita stylu, persvaze, internetová publicistika, komunikace na internetu, analýza textů, stylové faktory, stylistické aspekty výstavby textu

Key words: blogs, subjectivity of style, persuasion, internet journalism, internet communication, text analysis, style-forming factors, stylistic aspects of text composition

Úvod

Blogy, speciálně v českém prostředí, dnes mají za sebou vývoj sotva přesahující jedno desetiletí; speciálně pokud hovoříme o blížích mediálních (viz dále). Přesto se problematice blogování z různých hledisek věnovalo již mnoho autorů (viz Bibliografie). Cílem tohoto příspěvku je prezentovat první ucelený stylistický pohled na tento fenomén internetové publicistiky a prokázat, že v současnosti je možné chápání blogu jako analytického žánru internetové publicistiky. Nejprve se zaměřím na samotnou charakteristiku blogu, dělení blogů a hypotetický posun (bude upřesněno), který nastal v jejich využívání. Poté prezentuji novou metodu a pomocí té následně analyzuji vybrané texty, čímž své hypotézy potvrdím.

1 Dělení blogů, nový vývoj blogů a výběr textů k analýze

Blogy lze dělit zaprvé bez ohledu na charakter textu, a to dle místa publikování. Může jít jednak o před veřejností publikované texty (1a), které předpokládají neomezený počet čtenářů (např. mediální blogy; největší v ČR je *Blog.iDnes.cz*); jednak o soukromě publikované texty (1b) s omezeným či malým okruhem čtenářů (osobní webové stránky, sociální sítě → Facebook). Druhé dělení spočívá v charakteru publikovaných textů (bez ohledu na místo). Při této diferenciaci se jedná za prvé o texty obracející se k adresátům s naléhavostí (2a), přesvědčující veřejnost o autorově pravdě, vyznačující se působením persvazivní funkce (stejně jako komentář a glosa, analytické žánry publicistiky); případ druhý zastupují texty postrádající působení persvazivní funkce (2b); informující

1 Příspěvek vznikl v rámci grantového projektu specifického výzkumu MUNI/A/1165/2014
Čeština v jednotě synchronie a diachronie 2015.

například o soukromí autora bez snahy o ovlivnění adresáta a bez jakéhokoli vyššího cíle, který by mohl text v komunikační sféře plnit.

S nástupem Facebooku a jiných sociálních sítí, a tedy s dalším masovým rozvojem komunikačních možností, jež internet lidem přináší, došlo ve vývoji blogů k posunu umožňujícímu uvažování o blogu jako o novém analytickém žánru publicistiky, což ovšem nezbytně vyžaduje specifický pohled na danou problematiku. Nelze totiž prohlásit, že veškeré blogy začínají nabývat stylově stejného charakteru. Jde pouze o podmnožinu textů, jejíž výběr je však logický a relevantní (viz dále). Zatímco texty soukromé (2b) se přesouvají do prostředí soukromého (1b), zejména na Facebook, publikování textů veřejných (2a) má logicky největší smysl v mediálním prostředí veřejném (1a), kde tyto texty mají potenciál získat skutečný ohlas velké masy čtenářů. Samozřejmě, že tato skutečnost není zdaleka absolutní pravdou. Cílem však nebylo prohlásit blog obecně za analytický žánr, nýbrž prokázat formování tohoto žánru uvnitř množiny všech textů jako blog označených.

Pro analýzu bylo vybráno 50 nejčtenějších textů z českého mediálního blogu – *Blog.iDnes.cz* – publikovaných v týdenním časovém intervalu, a to z následujících důvodů; *Blog.iDnes.cz* je nejpopulárnější český mediální blog s největším počtem blogerů a stejně tak čtenářů (vyskytuje se na největším žurnalistickém serveru *Mladé fronty Dnes*); právě čtenost je důležitý prvek výběru. Interakce mezi blogery jako autory, blogery jako čtenáři a samotnými médii (žurnalisty a opět čtenáři, kterými jsou rovněž právě blogeři) je nesmírně podstatná; právě ta vede k ustálenosti jejich vyjadřování. Právě tato interakce (autor – čtenář) je procesem, který je pro stylistiku (a samozřejmě také pro literární teorii) přirozeně podstatný; intenzita tohoto procesu je však stále vyšší a s nástupem internetu dnes nesrovnatelná s minulostí a s dobou klasického tisku (kdokoli může informace poskytovat a kdokoli je také může díky internetu vyhledávat, od diskusí až po romány, a to zdarma).

Možnost chápání blogu jako žánru je pak pouze jedním z důsledků tohoto a podobných či souvisejících procesů, jež vznikly v novodobé komunikační sféře všeobecně.

2 Metoda uchopení žánru a nové termíny

Množina všech textů označených jako blog by byla ve smyslu hledání žánru stylisticky neuchopitelná, neboť blog obecně je se stylovou variabilitou neodmyslitelně spojen; tento problém byl odstraněn výběrem pouze té podmnožiny všech reálně existujících blogů, která může posloužit zvolenému cíli, přičemž tento výběr je zcela relevantní (viz kap. 1). Zvolenou (pod)množinu textů tvoří 50 reálně existujících blogových příspěvků. Cílem je definovat nový žánr (abstraktní centrální text, viz dále) na základě stylistické analýzy této množiny.

Pokud by měl být definován blog obecně, povědomí o centrálním textu by vůbec nemohlo vzniknout, jelikož na blozích se obecně vyskytují texty stylově spadající do sfér prakticky všech funkčních stylů a tato stylová nestálost jednoduše zobecnění neumožňuje. Odlišná je pak situace, kdy byla z celkové množiny všech blogů zvolena (nikoli účelovým způsobem) pouze ta její podmnožina schopná smysluplné výpovědi (za předpokladu stylové blízkosti zvolených textů). Povědomí o žánru může vzniknout pouze za předpokladu, že se v určité komunikační sféře objevují komunikáty, jež mají podobné stylové vlastnosti; všechny takové komunikáty se pak na formování abstraktního modelu podílejí společně.

Na základě seznámení se se zvolenými texty a na základě nabytí přesvědčení, že mezi nimi existuje stylová homogenost, lze následně přistoupit k analýze samotné.

Ta se soustředila na popis působení objektivních stylovtvorných faktorů a na vybrané rysy výstavby textu, stejně jako na výskyt vybraných jazykových prostředků (na stylovtvorné faktory subjektivní nebyl brán logicky ohled, neboť působí vně textu); podstatný je zde termín stylová dominanta, jenž byl stylistikou (Hausenblas, 1996; Křístek, 2009) vždy chápán jako ten výsledný rys textu či proje-

vu, který lze považovat za konstituující či nejzřetelnější; zde je pojata částečně odlišným a z části podobným způsobem, nepochybně však způsobem novým.

Pro potřeby své práce jsem zavedl termíny statistická většina a malá dominanta (viz dále), přičemž existence statistické většiny je podmínkou vystižení malé dominanty. Je-li například zjištěno, že v 80 % textů (→ statistická většina) jakékoli analyzované množiny je užitým kódem obecná čeština, je tím definována jedna malá dominanta dané množiny textů na úrovni jednoho stylového faktoru (kódu). Procentuální síla každé statistické většiny pak určuje váhu jí určené malé dominanty. Odhalení, či naopak neodhalení malých dominant poté umožňuje (nebo naopak neumožňuje) formulování definice centrálního textu, a vede tudíž k možnosti odhalení žánru nebo alespoň procesu jeho zrodu a formování.

Neméně důležité je následně vzít v potaz také vztahy mezi malými dominantami, tj. zda a do jaké míry ta která dominanta ovlivňuje dominanty zbylé; což je v podstatě korespondující se vzájemným působením stylových faktorů. Nejvýraznějším stylovým faktorem je zpravidla funkce a u zkoumaných textů nebylo náročné již před analýzou vypořádat působení funkce persvazivní; toto působení však bylo nutné analyzovat, stejně jako faktory ostatní, aby mohla být finální definice centrálního textu co nejkomplexnější a nejobektivnější.

3 Výsledky analýzy – malé dominanty analyzované množiny textů

Předem je třeba říci, že podrobné analýzy jsou a stále budou dostupné v souborech uložených v archivu diplomové práce (Vepřek, 2013; viz Bibliografie) a zde uvádím jen souhrn výsledků.

Prvním analyzovaným faktorem bylo prostředí a v případě zvolené množiny je výsledek jednotný. Jde o prostředí veřejné a známé, časově a prostorově neomezené (blogosféra) a nenormované; přesto je blogosféra prostředím s ustálenými prvky vyjadřování (viz dále).

1) prostředí – veřejné, známé, neomezené, nenormované

100 % 

Druhým nahlíženým faktorem pak byla forma. Ve všech případech se jedná primárně o formu psanou (analyzované blogy jsou texty psanými a publikovanými), ovšem ani sekundárně se nevyskytují prvky mluvenosti, jako např. tendence k nespisovnosti či projevy nepřipravenosti (k tomu opět blíže u dalších faktorů).

2) forma – psaná, bez prvků indikujících formu mluvenou

100 % 

Složitější pak bylo posoudit ráz všech textů. Jelikož analyzované komunikáty jsou texty veřejně publikované a rozsah jejich čtenářů se pohybuje mezi 2 500 – 40 000 (každé PC se započítává pouze jednou) a jelikož záměrem každého autora je, bez ohledu na cíle další, publikovat svůj text před veřejností, ráz všech textů je primárně veřejný. U 94 % komunikátů pak lze hovořit o rázu oficiálním až polooficiálním; pouhé 3 texty mají ráz neoficiální.

3) ráz – veřejný, texty veřejně publikované před předem neomezeným počtem čtenářů

100 % 

Velmi pozoruhodné zjištění přišlo při analýze užitého kódu v rámci variet češtiny. Ukázalo se totiž, že ve všech textech je primárně voleným kódem čeština spisovná (bez ohledu na míru jejího ovládnutí autory textů); to samo o sobě potvrzuje jejich veřejnost a oficiálnost; stejně tak tento fakt souvisí s formou textů, je dán prostředím publikování, poukazuje na funkci a podobně (vzájemné působení stylových faktorů je zcela zřetelné).

4) kód – primárně voleným verbálním kódem všech textů je spisovná čeština

100 % 

Míra připravenosti textů je pak odvoditelná právě z působení faktorů předchozích. Spisovná čeština jako zvolený kód, psaná forma, veřejný ráz a veřejné prostředí (plus faktory další) – tyto rysy,

projevující se u všech textů, jasně poukazují na připravenost textů. Občasné překlepy jsou pouze výjimkami potvrzujícími pravidlo.

5) **připravenost** – komunikáty připravené (dáno ostatními faktory)

100 % 

Malé dominanty tvořené statistickými většinami doposud uvedených stylových faktorů nejsou rozhodující pro hodnocení postavení blogu mezi analytickými žánry (komentář, glosa), ale jsou podstatné pro demonstraci samotné možnosti uvažování o blogu jako o publicistickém textu (žánru); např. sama nepřipravenost by takové uvažování vyloučila.

Velice podstatným faktorem pro blogosféru je téma. Tématem 46 % textů je politika, 28 % finance atd. Taková statistika by však nebyla ideální. Důležité je, že blogeři (zejména ti, jejichž texty jsou čtenářsky úspěšné) reagují na aktuálně kontroverzní témata, která jsou zároveň probírána v médiích. Často ostře reagují nejen na témata samotná, ale rovněž na jejich zpracování v médiích. Nejfrekventovanější téma dne či týdne přitom nemusí s politikou vůbec souviset; důležitá je pouze závažnost tématu pro veřejnost či zájem veřejnosti o téma. Následuje tak zjištění, že v 94 % textů jde o téma společensky závažného či závažnějšího dosahu (politika, finance, školství, bezpečnost aj.).

6) **téma** – společensky závažné a podstatné

94 % 

Jak již bylo řečeno, při analýze mne nezajímá vliv subjektu na objektivně viditelnou podobu textu. Nemohl však být opomenut explicitní výskyt subjektu v textu, neboť je podstatným a snadno identifikovatelným rysem blogosféry; zejména podstatným pro její odlišení od publicistiky profesionální. Osoba pisatele, jako aktivní či pasivní účastník obsahu, vstupuje do 88 % textů (věty typu *Já jeho názor nesdílím!*; *To mě při mém rozhodování nikterak neovlivnilo.*); v 66 % případů jde dokonce o min. 4 výskyty. Tato explicitní subjektivita textu poukazuje na fakt, že bloger vždy mluví sám za sebe, nikoli např. za deník (médiu), jehož je zaměstnancem, jako je tomu u žurnalistů. Přesto může mít stejné vyjadřovací schopnosti a znalosti, není však při své výpovědi omezen.

7) **explicitní výskyt subjektu** – přímý výskyt subjektu v textu

88 % 

N nejen přímý výskyt subjektu, ale také přímý kontakt s adresátem je pro blogosféru zcela typickým rysem. V 1. os. plurálu se tento kontakt objevuje v 78 % textů (např. *Nesmíme to dopustit!*), v 2. os. plurálu pak v 52 % textů (*Zamyslete se nad tím, co se děje.*), v průniku pak v 84 % komunikátů. U 52 % textů se navíc objevuje kontaktní imperativ. Kontaktovost je signálem působení persvazivní funkce.

8) **explicitní kontakt s adresátem** – ztotožnění se se čtenáři, výzva k nim směřující atd.

84 % 

Všechny doposud vyjádřené dominanty jsou ovlivněny funkcí komunikátů; již zřetelná kontakto-
vost, častý imperativ, závažná témata, či vědomě užitá spisovná čeština jako kód poukazují na očividné působení funkce persvazivní; její vliv byl dále analyzován na vybraných rysech výstavby textů a užitých jazykových prostředků. Projevy této funkce jsou ve zvolené množině textů jednak zcela očividné, jednak právě jejich existence má být směrem k analytické publicistice prokázána; jde tedy opět o postup relevantní, nikoli účelový (nesnažím se prohlásit projevy persvaze za pro blogosféru stěžejní, ale poukázat na fakt, že o působení této funkce v analyzovaných textech nelze pochybovat).

9) **funkce** – persvazivní, tento údaj vychází jak z již uvedených, tak následujících statistik

94 % 

Následující subdominanty poukazují na stylové rysy, kterými je persvaze v uvedeném procentu textů posílena. Jde o vybrané prvky výstavby textu a jazykové prostředky.

Zde – tj. při jejich výběru – jsem byl nucen postupovat již poněkud cíleně, ovšem vzhledem ke sta-

novenému cíli jde o postup logický.

Nejprve jsem se snažil ve všech textech najít delší výpovědní celky (souvětí až části odstavců) nesusící zřetelně argumentační ráz (příčinné větné vazby atd.); podobné výpovědní úseky se vyskytly v 74 % komunikátů.

argumentace – delší výpovědní celky argumentačního charakteru

74 % 

Druhým vyhledávaným prvkem persvaze byly hodnotící výrazy (adjektiva, adverbia, pronomina), které jsem samostatně vyhodnotil jako výrazy posilující persvazi. Minimálně 8 výskytů takových výrazů se objevilo u všech textů, ale samozřejmě zde byl můj přístup k analýze značně subjektivní (např. *Jednání této vlády je absolutně nepřipustné!* = 2 výrazy tohoto typu; stejně tak ale *Vůbec se nezadýcháte!*. – přestože jde o jeden z textů soukromého rázu).

hodnotící výrazy – adjektiva, adverbia, pronomina

100 % 

Zaměřil jsem se tedy posléze navíc na vyhledání konkrétních výrazů, a to tzv. totalizátorů; zájmen *všichni, každý, nikdo a žádný*. Alespoň jeden z nich se vyskytl v 84 % textů.

totalizátory – *všichni, každý, žádný, nikdo*

84 % 

Dalším častým rysem analyzovaných komunikátů se ukázala být tázací forma jako další projev kontaktu se čtenářem a rovněž projev posilující působení persvazivní funkce. Adresát je otázkami vtažen do textu. Otázky směřující přímo na adresáty se objevily v 78 % textů.

tázací forma – další projev kontaktu se čtenářem

78 % 

Posledními dvěma analyzovanými rysy pak byly prvky vertikálního členění textu a intertextovosti jako rysů výstavby textu.

Alespoň 3 výskyty prvků vertikálního členění textu (změna velikosti písma, barvy písma, ztučnění písma, publicistické uvozovky, větší počet vykřičníků aj.) byly objeveny u 66 % textů; alespoň 1 výskyt u 96 %.

vertikální členění textu – velikost a barva písma, tučné písmo aj.

96 % 

Hypertextové odkazy na texty jiného nebo téhož blogera, či na texty mimo blogosféru byly identifikovány u 74 % textů. Blogeréři používají jiné texty často právě pro podporu názorů svých, či poukazují na skutečnost, že se tématem nezabývají poprvé apod.

intertextovost – hypertextové odkazy na jiné texty

74 % 

4 Definice centrálního textu a blogu jako analytického žánru

Analýzou určené výrazné statistické většiny a pomocí nich vystižené malé dominanty prokázaly nepopíratelnou existenci ustálenosti vyjadřování v dané množině textů. Komplexní definice blogu jako žánru na základě malými dominantami tvořeného abstraktního centrálního textu může být pak následující:

Připravený a kultivovaným jazykem psaný text oficiálního nebo polooficiálního rázu publikovaný v daném veřejném prostředí (blogosféra) s analytickými intencemi autora (blogera). Ten za využití variabilních slohových postupů prezentuje svůj subjektivní, avšak argumenty podložený názor na společensky podstatná témata, přičemž dává frekventovaně explicitně najevo svoji přítomnost jako přítomnost autorského subjektu a rovněž frekventovaně explicitně kontaktuje adresáta formou 2. os. plurálu, ztotožněním (1. os. pl) či tázací formou; což jej stejně jako celková subjektivita a (primárně) nenormovanost odlišuje od komentáře či glosy. Na adresáta autor dále působí jak

v rovině vyjadřovacích prostředků, tak v rovině výstavby textu, kde k posílení persvaze v obou případech využívá specifických možností internetu a především nenormovanost vyjadřování v blogosféře. Blogger nejenže prezentuje svůj názor, ale zároveň se snaží adresáty přesvědčit o jeho objektivitě a správnosti, a to obsahově i formálně primárně neomezenými způsoby, u kterých však došlo k ustálení.

Závěr

Při pohledu na výsledná zjištění vidíme něco, co nepochybně zasluhuje pozornost stylistiky. Blog lze totiž chápat jako první žánr, který vykrytalizoval z absolutně subjektivního a nenormovaného prostředí směrem k objektivitě. Zcela protikladně se naopak vyvíjí profesionální publicistika, dnes jednoznačně směřující k opačnému pólu, tedy k pólu subjektivnímu. Obojí je umožněno protínáním komunikační sféry prostředělovací a publicistické. Zatímco laikové jsou dnes schopni (díky enormnímu příjmu informací) a mají možnost (díky internetu) publikovat, publicisté se naopak v boji o čtenáře přibližují svým adresátům, jejich jazyku a potřebám. Tím je vystižen dynamický a prudký vývoj komunikačních možností současnosti, o němž analýza blogů za pomoci nové metody podala názorný příklad.

Bibliografie

- BEDNÁŘ, V. (2011): *Internetová publicistika*. Praha: Grada Publishing.
- BYRON DL. – BROBACK S. (2008): *Blogy: Publikuj a prosperuj*. Přeložil J. Hynek. Praha: Grada Publishing.
- GORMAN, M. (2005): BackTalk: Revenge of the Blog People! In: *LibraryJournal.com* [online]. 2005-02-15. [cit. 2013-03-15]. <<http://www.libraryjournal.com/article/CA502009.html>>
- HAUSENBLAS, K. (1996): K celkové stylistické charakteristice textů (J. Neruda – U tří lilií – styl a smysl povídky). In: *Od tvaru k smyslu textu: stylistické reflexe a interpretace*. Praha: Karlova Univerzita, Filozofická fakulta, s. 143–151.
- KŘÍSTEK, M. (2009): Stylistický rozbor textu a stylová dominanty. In: E. Lehečková – J. Tábořský, eds. *Struktura, variety, funkce. Acta Universitatis Carolinae, Philologica 2* (2006), Slavica Pragensia XLI. Praha: Karolinum, s. 65–70.
- RINGMAR, E. (2007): *A Blogger's Manifesto: Free Speech and Censorship in the Age of the Internet*. London: Anthem Press. <<http://ia700401.us.archive.org/12/items/ABloggersManifestoFreeSpeechAndCensorshipInTheAgeOfTheInternet/ErikRingmarABloggersManifesto.pdf>>
- SIVEROVÁ, L. (2011): *Čeština dívčích blogů*. Brno. Diplomová práce. Masarykova univerzita, Filozofická fakulta, Ústav českého jazyka. <https://is.muni.cz/th/342448/ff_b/>
- VEPŘEK, J. (2013): *Blogy jako žánr potvrzující dynamický vývoj komunikačních současnosti*. Brno. Diplomová práce. Masarykova univerzita, Filozofická fakulta, Ústav českého jazyka. <https://is.muni.cz/auth/th/342423/ff_m/>
- VEPŘEK, J. (2011): *Blogy a stylistické aspekty výstavby textu*. Brno. Diplomová práce. Masarykova univerzita, Filozofická fakulta, Ústav českého jazyka. <http://is.muni.cz/th/342423/ff_b/>
- VOZÁROVÁ, M. (2009): *Čeština blogu*. Brno. Diplomová práce. Masarykova univerzita, Filozofická fakulta, Ústav českého jazyka. <https://is.muni.cz/auth/th/215351/ff_b/>
- VOZÁROVÁ, M. (2011): *Obecná čeština na vybraných blozích*. Brno. Diplomová práce. Masarykova univerzita, Filozofická fakulta, Ústav českého jazyka. <https://is.muni.cz/auth/th/215351/ff_m/>
- (další zdroje a kompletní výsledky analýz jsou uvedeny v diplomové práci a příložených souborech → 1. bibliografický údaj a odkaz na snadno dostupný elektronický archiv práce)

JAZYK SEZNAMOVACÍCH INZERÁTŮ NA INTERNETU

TEREZA MARKOVÁ

Ústav českého jazyka FF MU Brno

Abstract

The thesis deals with results of the elementary language analysis of more than two thousands of dating advertisements on the Internet portals such as Seznamka.cz and Stesti.cz. The analyzed elements were chosen from a previous observation which was based on finding out the typical features of the language of the dating advertisements. The advertisements were searched for their orthography, selected vocabulary, importance of the introductory phrases, meaning of questions and using of various emoticons. Particular results are compared according to the differences between men and women.

Klíčová slova: seznamovací inzeráty, jazyková analýza, úvodní věta, deminutiva, grafické prostředky

Keywords: dating advertisements, language analysis, introductory phrase, diminutives, graphics resources

Jazyk seznamovacích inzerátů na internetu

Ve 21. století se svět v mnoha oblastech rychle mění a vyvíjí. Vyspělé technologie nás provází každodenním životem, a tím pádem se i komunikace mezi lidmi přesouvá do elektronické sféry. Díky různým typům komunikátorů jsme schopni se spojit a dorozumívat se téměř s kýmkoli po celém světě. Ani oblast partnerských vztahů tohoto pokroku nebyla ušetřena, a tak se nám otevírá příležitost například pro studium a bádání synchronního jazyka seznamovacích inzerátů.

Tomáš Novák ve své knize uvádí, že termín seznamovací inzerát vzniká v Anglii v roce 1695, kdy poprvé inzerát tohoto typu vychází. V českých zemích se tomu stalo až o necelé století později, 8. března 1794. První seznamovací inzerát byl zveřejněn v *Poštovských novinách* a jeho autorem byl čtyřicetiletý vdovec, který hledal manželku pro sebe a matku svým dvěma dětem (Novák, 2003). Inzerát byl původně pouze krátké a placené oznámení v tisku, nyní pojmenovává i sdělení na veřejných vývěskách, na internetu, reklamu ve sdělovacích prostředcích. S technickým pokrokem také postupně klesá počet tištěných seznamovacích inzerátů v periodikách a naopak se rapidně zvyšuje jejich počet na internetových seznamovacích portálech.

Seznamovací inzerát je zcela kontextově nezávislý komunikát monologického rázu určený pro potenciální zájemce. Je pro něj typické zejména modelové vyjadřování, většinou až stereotypní, ale zato veskrze explicitní, což dokážeme v jazykové analýze níže.

Nejsnazší cesta, jak se seznámit přes internet, vede přes seznamovací portály. V České republice je hned několik serverů zabývajících se touto problematikou, největší z nich jsou Seznamka.cz, Lide.cz a Rande.cz. Pro analýzu jazyka seznamovacích inzerátů na internetu jsme zvolili inzeráty z portálů Seznamka.cz a Stesti.cz. Seznamka.cz existuje již od roku 1996 a je nejstarším a největším seznamovacím portálem v České republice. Možnost seznámení poskytuje zdarma a v několika desítkách kategorií, od vážného vztahu přes flirt až po společné zájmy. Průměrná návštěvnost je téměř 30 000 uživatelů denně a pro možnost přidání inzerátu je třeba se zaregistrovat, a to prostřednictvím e-mailu a zvoleného hesla.

Na portálu Stesti.cz je zaregistrovaných přibližně 120 000 uživatelů a nabízí podobné služby jako

Seznamka.cz. Denní návštěvnost se pohybuje okolo 17 000 uživatelů, tento nižší počet může být způsoben zpoplatněným způsobem registrace, čímž zřejmě portál přichází o část zájemců, ale na druhou stranu lze v souvislosti s tím uvažovat o zvýšení kvality a věrohodnosti seznamovacích inzerátů.

Tyto dva portály byly zvoleny jako vhodné kvůli stěžejním informacím, které inzerent musí zadat při vytvoření svého profilu, a to pohlaví, věk a nejvyšší dosažené vzdělání. Výsledky vychází z analýzy celkem dvou tisíc seznamovacích inzerátů heterosexuálních mužů a heterosexuálních žen. Průměrný věk mužů, kteří hledají své partnerky přes internet, je 37 let, u žen pak 39 let. Nejvyšší zastoupení mají absolventi středních škol, ale je třeba podotknout, že více vysokoškolsky vzdělaných žen hledá partnera touto cestou oproti vysokoškolsky vzdělaným mužům.

Nyní se dostáváme již k samotné jazykové analýze, která se většinou ponese v duchu komparace způsobu vyjadřování mužů a žen, což budou potřebné informace pro další práci strojového učení a automatickému rozlišení pohlaví nebo jiných charakteristik pisatele. Výběru jednotlivých analyzovaných jevů předcházelo dlouhé pozorování typických rysů seznamovacích inzerátů.

Charakteristické pro seznamovací inzeráty se ukázalo užívání obecné češtiny, což se projevilo zejména diftongizací *í > ej*, hlavně tedy v koncovkách tvrdých adjektiv mužského rodu (např. *kterej*), méně frekventovaně pak v substantivech (např. *rejpalka*). Častěji se tento jev vyskytoval v inzerátech mužských pisatelů. Zaznamenali jsme také zvýšenou četnost úženy *é > í*, a to pouze v adjektivních koncovkách nebo v zpodstatněných přídavných jménech (např. *hledám mužskýho*).

Vzhledem k délce a praktickému významu seznamovacích inzerátů by se dala očekávat snaha pisatelů o jazykovou správnost, avšak opak je pravdou. Inzerenti chybovali nejčastěji v užití osobního zájmena *já* v genitivu, dativu, akuzativu a lokálu singuláru, a to až ve třetině případů. Z analýzy vyplývá, že pokud inzerent neměl dostatečné znalosti o zákonitostech této problematiky, tak užil v celém inzerátu variantu *mě*. Dále lze pozorovat chybné zápisy slov, které inzerenti píší tak, jak se vyslovují (např. *Trvalý vztah pořát nenajdu?*) nebo nezapisují kvantitu vokálů, zpravidla pak u sloves v první osobě singuláru (např. *... zatím jsem neměl štěstí, tak to zkusím tady*).

Z hlediska lexika hrají podstatnou roli deminutiva. Ta ženy používají až o 50 % častěji než muži. A ačkoli si inzerenti uvědomovali příznakovost těchto výrazů a byli ve výběru obezřetní (většinou jedno deminutivum v inzerátu), tak je zřejmé na příkladu, že ani pro oblast seznamovacích inzerátů není vhodné nadměrné užívání tohoto lexika (např. *Zdravím všechny..jsem společenský tvoreček, který má lidi, tanec, posezení u ohničku s nezbytnou kytarou a vše krásné co k životu patří. Jsem romantická dušička, která ještě touží po tom někoho potkat*).

Vzhledem k opravdu velkému množství seznamovacích inzerátů na internetu lze předpokládat, že se inzerenti budou snažit zaujmout hned prvními slovy svého inzerátu tak, aby recipient inzerát dočetl, popřípadě na něj reagoval. Projevila se ovšem přesně opačná tendence. Více než 60 % seznamovacích inzerátů začíná jedním z následujících způsobů: tvarem slovesa *být* (častěji ženy), pozdravem (častěji muži) nebo výčtem vlastností, ať už svých nebo požadovaných od hledaného partnera (nezávisle na pohlaví).

Dalším charakteristickým rysem textu seznamovacích inzerátů je časté pokládání otázek. Tázací větou se inzerent snaží aktivizovat čtenáře, nutí jej k reakci a v podstatě se pokouší o dialogičnost. Inzerenti většinou otázky situují na začátek nebo konec inzerátu a jejich znění se ukázalo opět velmi stereotypní. Nejčastěji jsou to pak zejména dotazy typu *najdu tě/lásku/tady?, jsi tady?, existuješ?*. Tázací věta se objevila přibližně ve 20 % seznamovacích inzerátů a je třeba podotknout, že oproti genderovým předpokladům častěji otázky kladli muži, a dokonce ženy byly ekonomičtější, co do počtu otázek v jednom inzerátu.

I uplatnění grafických prostředků může být zásadním faktorem úspěšnosti inzerátu. Užitím emotikonů, velkých písmen nebo řetězením interpunkčních znamének inzerent působí větší měrou

na vizuální smysly recipienta, tím pádem na sebe více upozorní a možnost, že na inzerát někdo zareaguje, je vyšší. Ke specifickým elektronické komunikace patří ignorace diakritiky, seznamovacích inzerátů bez diakritiky bylo přibližně 10 % a častěji bez diakritických znamének psali muži oproti ženám. I v oblasti interpunkce docházelo v seznamovacích inzerátech k častým nekorektnostem, pisatelé chybovali jak ve správném grafickém zápisu interpunkce, tak v jejím užití. Zcela bez interpunkce bylo přes 8 % inzerátů, ale šlo zejména o krátké inzeráty většinou složené ze dvou vět. K řetězení interpunkčních znamének docházelo až třikrát častěji v seznamovacích inzerátech žen. Poslední významnou složkou grafických prostředků v inzerátech je užívání emotikonů, tedy symbolů určených pro vyjádření pocitů, nálady a dalších duševních stavů pisatele (Martincová, 2004). Inzerenti nejčastěji psali emotikon, který značí úsměv :-), vyskytl se ve více než třech stech inzerátech, zatímco jeho opak :-(jen v osmi inzerátech, což svědčí i o tom, že se pisatelé snaží, aby na čtenáře působili šťastně a pozitivně. Častěji emotikony užívali muži, a je třeba podotknout, že emotikony jsou stále doménou spíše mladých lidí (26 až 30 let).

Výsledky z této elementární jazykové analýzy seznamovacích inzerátů budou v budoucnu použity k další práci v oblasti strojového učení, a to za pomoci algoritmu podpurných vektorů (metodou Support vectore machine), pro automatické rozlišení pohlaví pisatele bez ohledu na rodové koncovky. Již nyní lze předpokládat, že pokud je inzerát bez diakritiky, obsahuje emotikon a více než dvě tázací věty je velmi pravděpodobné, že autorem bude muž. Naopak pokud inzerát začíná slovem *jsm*, obsahuje deminutivum a dochází v něm k řetězení interpunkčních znamének, půjde o inzerát ženy.

Již více než tři století lidé k seznamování využívají inzeráty, v dnešní technicky pokročilé době lze získat nepřeberné množství těchto seznamovacích inzerátů na internetu a čerpat z nich znalosti například o současném jazyce, o což jsme se pokusili i v této elementární analýze. Rozbor vybraných jazykových jevů, založený na komparaci jednotlivých výskytů v inzerátech mužů a žen, začíná charakteristikou užívání obecné češtiny v inzerátech, což je dokázáno zejména na diftongizaci *í > ej* a úzení *é > í*. Dále se zabývá jazykovou správností inzerátů, kde se ukázaly největší nedostatky při psaní *mě/mně*. Projevilo se, že podstatnou roli hrají v seznamovacích inzerátech deminutiva, jejichž užití je doménou žen a také, že stereotypnímu vyjadřování podléhají (nezávisle na pohlaví) znění úvodních frází nebo tázacích vět v inzerátech. Analyzovali jsme i užívání různých grafických prostředků jako jsou interpunkce, diakritika nebo emotikony. Jednotlivé výsledky budou konfrontovány v dalším strojovém zpracování, které by mělo vést k automatickému rozlišení pohlaví pisatele a dalších charakteristik.

Seznam literatury

- MARTINCOVÁ, O. a kol. (2004): *Nová slova v češtině: slovník neologizmů 2*. Praha: Academia.
 NOVÁK, T. (2003): *Jak loví ženy? Jak loví muži? aneb Být sám sebou a nebýt sám – seznámení možné i nemožné*. Líbeznice: Víkend.

VLASTNÍ JMÉNA VLAKOVÝCH SPOJŮ (POREJONYMA) NA ÚZEMÍ ČR V LETECH 1993–2013

MICHAELA FICNAROVÁ

Ústav českého jazyka FF MU Brno

Abstract

Proper Names of Train Lines in the Czech Republic (1993–2013)

The presented article deals with proper names of train lines operating in the Czech Republic between 1993 and 2013. Most of the 357 proper names, which represent the material basis of research, were gathered especially by excerption from available timetables of Czech Railways. The introductory part of article is devoted to the brief presentation of the position of proper names of train lines in the onomastics. This part is followed by the summary of existing bibliographic sources relating to the matters of the proper names of train lines and to the presentation of classification, which is based on motivational and structural models. The most important section of this article is the summary of proper names of train lines classification results. Attention is concentrated on the developmental changes that the names of train lines have undergone through in our territory since the mid-1960s.

Klíčová slova: vlastní jména vlakových spojů, porejonyma, chrématonyma, propria, onomastika, klasifikace porejonym, motivační modely, strukturní modely, vlaky, Česká republika

Key words: proper names of train lines, chrematonyms, proper names, onomastics, classification of proper names of train lines, motivation models, structural models, trains, Czech Republic

Příspěvek si klade za cíl představit soubor vlastních jmen vlakových spojů (v onomastické terminologii **porejonym**), který se vyskytoval na území České republiky v letech 1993–2013.¹ Materiálovou základnu tvoří **357 názvů** získaných excerpcí z jízdních řádů Českých drah a elektronických zdrojů (internetových portálů dopravců, zájmových stránek apod.). Pozornost je soustředěna nejdříve na postavení porejonym v rámci onomastické terminologie a na stručné seznámení s literaturou vztahující se k problematice porejonym. Následně je představena modifikovaná vztahově-modelová klasifikace, jež byla na excerpovaný materiál aplikována, a jsou prezentovány výsledky klasifikace názvů vlakových spojů. V závěru příspěvek sleduje kvantitativní nárůst porejonym, proměny jejich motivací a příčiny těchto proměn.

Postavení porejonym v rámci onomastiky

V onomastice se podle druhu pojmenovávaného objektu tradičně rozlišují tři základní skupiny vlastních jmen (VJ), a to **antroponyma** (VJ jednotlivého člověka nebo skupiny lidí), **toponyma** (VJ zeměpisná; VJ objektů, které leží v krajině a jsou fixovatelné v mapách) a **chrématonyma** (VJ

1 Příspěvek vznikl v rámci grantového projektu specifického výzkumu MUNI/A/1165/2014 Čeština v jednotě synchronie a diachronie 2015. Jeho východiskem je magisterská diplomová práce *Vlastní jména vlakových spojů (porejonyma) na území ČR v letech 1993–2011* obhájená v roce 2012 na Filozofické fakultě Masarykovy univerzity a také publikace *Vlastní jména vlakových spojů (porejonyma) na území ČR v letech 1993–2013* (v tisku).

lidských výtvorů, které jsou zakotveny nikoli v přírodě, nýbrž ve společenských, ekonomických, politických, kulturních aj. vztazích).²

Porejonyma (VJ konkrétních dopravních spojů a linek) náleží do systému **chrématonym**. Na další zařazení porejonym však nepanuje jednotný názor; někteří badatelé je řadí k **sérionymům** (k pojmenováním **sériově vyráběných průmyslových výrobků**; např. M. Knappová);³ jiní je naopak přiřazují k **unikátonymům** (k VJ jedinečných výrobků; produktů vyráběných jednotlivě). Mezi **tzv. technická unikátonyma** řadí porejonyma spolu s názvy lodí, kosmických sond, vzducholodí apod. např. L. Garančovská.⁴ Dle našeho názoru náleží porejonyma k unikátonymům, neboť pojmenovávají dopravní spoje, jejichž jedinečnost je dána jejich trasou a časem.

Shrnutí dosavadní literatury k tématu

Ačkoli názvy vlakových spojů existují na našem území poměrně dlouhou dobu,⁵ zůstávaly zpočátku stranou soustavnějšího jazykovědného bádání.

Jedním z prvních autorů, jenž se porejonymy začal zabývat, byl **Jaroslav Hubáček**. V článku **K názvům vlaků**, publikovaném v roce 1966 v časopise *Naše řeč (dále NR)*,⁶ upozornil na výskyt těchto názvů v tehdejší jízdním řádu ČSD. Autor se v příspěvku soustředil pouze na výklad a hodnocení porejonym oficiálních, porejonyma neoficiální, jež náleží především k pojmenováním slangovým, zmínil okrajově. Předmětem článku byly také důvody vedoucí ke vzniku názvů vlakových spojů a k jejich oficiálnímu užívání.

V rámci onomastického semináře vedeného Vladimírem Šmilauerem přednesl svůj příspěvek **O jménech vlaků Ivan Honl**.⁷ Honl se v něm sice omezil na situaci v ČSSR, upozornil však na šíření vlastních jmen vlaků ze zahraničí na evropské území včetně našich zemí. Věnoval se především pojmenovacím motivům existujících porejonym. Stejně jako Hubáček poukázal Honl na odlišný původ železničních pojmenování, který přinesl vznik jednak pojmenování oficiálních, jednak slangových.

V roce 1975 publikoval ve *Zpravodaji místopisné komise (dále ZMK)* stručný příspěvek nazvaný **Jména rychlíků Vladimír Šmilauer**.⁸ Reagoval na podnět Marie Novákové-Šlajsové, která si povšimla, že se názvy vlakových spojů těší velké oblibě a jejich počet každoročně narůstá. Šmilauer rozdělil existující porejonyma podle jejich typu do tří skupin čítajících celkem 45 názvů. První z nich zahrnovala jména utvořená z názvů zeměpisných, druhá ze jmen osobních, poslední pak ze jmen abstraktních.

2 KARLÍK, P. – NEKULA, M. – PLESKALOVÁ, J. (2002): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, s. 204.

3 KARLÍK, P. – NEKULA, M. – PLESKALOVÁ, J. (2002): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, s. 505.

4 GARANČOVSKÁ, L. (2009): Vymedzenie chrématonym v onymickom systéme. *Acta onomastica*, 50, s. 97.

5 Názory na to, kdy se na našem území vyskytlo první oficiální jméno vlaku, nejsou zcela jednotné. Ludvík Kuba či Jaromír Spal uvádí, že se první takový název objevil v ČSR roku 1936, přičemž šlo o dieselelektrický motorový rychlík, jenž svému názvu **Slovenská strela** ostudu rozhodně nedělal, neboť dokázal urazit trasu Praha – Bratislava za pouhé 4 hodiny a 51 minut. Ivan Honl naproti tomu považuje za první oficiální jméno vlaku u nás tzv. **Balkanzug, který vyjel** již během první světové války v r. 1915.

6 HUBÁČEK, J. (1966): K názvům vlaků. *Naše řeč*, 49, s. 151–154.

7 HONL, I. (1970): *O jménech vlaků*. *Zpravodaj místopisné komise*, 11, s. 9–15.

8 ŠMILAUER, V. (1975): *Jména rychlíků*. *Zpravodaj místopisné komise*, 16, s. 555–557.

Širší kontext železničních pojmenování představil v článku z r. 1982 *Z onomastiky železniční Jaromír Spal*.⁹ Spal upozornil na skutečnost, že s výjimkou krátkého Šmilauerova příspěvku zůstávají železnice a zařízení na nich mimo pozornost českých onomastiků.

V uvedeném příspěvku sledoval vývoj pojmenování od názvů železničních společností a tratí přes vlastní jména lokomotiv a vagonů až k pojmenování jednotlivých rychlíků. Nutno podotknout, že se neomezil pouze na prostředí české, nýbrž je začlenil do souvislostí světových.

První komplexní zpracování porejonym přinesla diplomová práce **Martiny Galečkové** z roku 1988 nazvaná *Názvy vlaků*.¹⁰ Pod vedením Jaroslava Hubáčka autorka shromáždila celkem 79 názvů vnitrostátních expresů a rychlíků a analyzovala je z hlediska motivačního i onomaziologického. Mimo názvy vyskytující se na území ČSSR Galečková reflektovala jména dalších 21 evropských dálkových spojů, které našim územím neprojízďely. Onymický materiál získala především z domácích jízdních řádů z 80. let, názvy dálkových spojů pak z publikace Stanislava Antonického *Evropské železnice*. Autorka provedla také sociologický výzkum (u 40 respondentů), který měl ukázat, jaké je povědomí veřejnosti o názvech spojů. Zajímalo ji, zda si cestující pojmenování uvědomují, zda jsou schopni vysvětlit jejich původ a motivaci a v neposlední řadě také to, jaký je jejich názor na zavádění porejonym. Z hlediska teoretického Galečková v diplomové práci navázala na články Jaroslava Hubáčka publikované v *Naší řeči*¹¹ a v *Onomastickém zpravodaji*.¹²

Další prací, která byla věnována problematice porejonym na našem území, je diplomová práce nazvaná *Vlastní jména dopravních spojů (porejonyma) a jejich jazykové zpracování* od **Ladislava Pokorného**.¹³ Ta vznikla roku 1995 pod vedením Ludvíka Kubu na Pedagogické fakultě Univerzity Jana Evangelisty Purkyně v Ústí nad Labem.¹⁴ Jak sám autor uvádí, podnětem a inspirací mu byl článek Vladimíra Šmilauera z roku 1975 a jeho cílem bylo dovést problematiku porejonym do autorovy současnosti, tj. do roku 1994. Obdobně jako v případě diplomové práce Martiny Galečkové byl jazykový materiál získán prostudováním jízdních řádů. Pokornému se tímto způsobem podařilo shromáždit celkem 146 názvů vlakových spojů, jež se vyskytly na území Československa (resp. České republiky) v letech 1969–1994. Nejdůležitější a nejrozsáhlejší část práce představuje rozbor všech zjištěných porejonym. Jednotlivá hesla obsahují technické informace o pojmenovaném objektu, dále rozbor lexikálně-sémantický, slovtvorný, popř. etymologický. Některé názvy jsou také doplněny o zajímavosti tvaroslovné, zvukové či grafické.

V návaznosti na diplomovou práci publikoval **Ladislav Pokorný** roku 1996 článek nazvaný *Vlastní jména dopravních spojů (porejonyma) a jejich jazykové zpracování (diplomová práce na PF UJEP Ústí nad Labem)*, v němž přehledně rozdělil analyzovaná porejonyma dle jejich struktury a motivace pojmenování.¹⁵

Problematika porejonym zůstává aktuální i po roce 2000. Pozornost jim věnují dva autoři, jimž je železniční onomastika velmi blízká. Prvním z nich je **Ludvík Kuba**, jenž v obsáhlém článku nazvaném *Proces nominace dopravních prostředků provozovaných na území našeho státu*¹⁶ detailně

9 SPAL, J. (1982): *Z onomastiky železniční*. *Zpravodaj místopisné komise*, 23, s. 343–348.

10 GALEČKOVÁ, M. (1988): *Názvy vlaků*. Ostrava, 66 s.

11 HUBÁČEK, J. (1966): K názvům vlaků. *Naše řeč*, 49, s. 151–154.

12 HUBÁČEK, J. (1985): Zur Problematik der Zugnamen. *Onomastický zpravodaj*, 26, s. 91–100.

13 POKORNÝ, L. (1995): *Vlastní jména dopravních spojů (porejonyma) a jejich jazykové zpracování*. Ústí nad Labem, 180 s.

14 Práce je dostupná ve formě prezenční výpůjčky v Ústřední knihovně PF UJEP.

15 POKORNÝ, L. (1996): *Vlastní jména dopravních spojů (porejonyma) a jejich jazykové zpracování: diplomová práce na PF UJEP Ústí nad Labem*. *Acta onomastica* 37, s. 139–146.

16 KUBA, L. (2007): *Proces nominace dopravních prostředků provozovaných na území našeho státu*. *Acta onomastica*, 48, s. 79–107.

klasifikuje motivace pojmenování pro kolejnicové dopravní prostředky (lokomotivy i vlakové spoje), dále pak pro názvy vodních dopravních prostředků, lodí námořní dopravy, automobilů a letadel.

Prostřednictvím příspěvku ve sborníku věnovaném životnímu jubileu významného českého onomastika Rudolfa Šrámka se pak k tematice porejonym vrátil r. 2009 **Jaroslav Hubáček**. Jeho článek *K propriím na železnici*¹⁷ podává klasifikaci názvů lokomotiv a dálkových vlakových spojů. Od publikování autorova prvního příspěvku zaměřeného na vlastní názvy vlakových spojů uplynulo celých 33 let a tento znatelný časový odstup umožňuje sledovat, jak výrazné změny v porejonymech vyskytující se na našem území nastaly.

Názvy vlakových spojů však nejsou atraktivní pouze pro jazykovědce. Velké množství článků k tomuto tématu lze v současné době nalézt v prostředí internetu. Zvýšený zájem o vlakové spoje a jejich názvy ze strany celostátních i regionálních deníků (resp. zpravodajských portálů) je možné sledovat zejména v době, kdy vchází v platnost nový jízdní řád (dále JŘ).¹⁸ K takovým patří např. článek uveřejněný v ekonomické rubrice portálu **iDnes.cz** *České dráhy změni jména vlaků, Šohaj už nepojede*,¹⁹ jenž se jednak zabýval změnami porejonym vyplývajícími z návrhu nového JŘ, jednak názvy spojů představil jako reklamní nástroj. Obdobné motivace měl i příspěvek publikovaný na portálu **Denik.cz** nazvaný *Emanuel Rádl, Kiss, Šumava. To jsou české vlaky*²⁰ nebo článek *Netradiční způsob zviditelnění firmy: pojmenujte po ní vlak* ze serveru **Novinky.cz**.²¹

Motivačně-modelová klasifikace porejonym

Na základě zkušeností s modelovou klasifikací pomístních jmen (dále MKPJ)²² jsme také na porejonyma chtěli aplikovat třídění, jež by respektovalo a odráželo jednotlivé fáze pojmenovacího procesu a zohlednilo specifika, jimiž se onymický materiál vyznačuje oproti apelativům.

Mezi klasifikace, jež tato kritéria splňují (tzn. klasifikují komplexně, v první fázi podle pojmenovacích motivů vyjádřených formou vztahových, resp. motivačních modelů, v druhé fázi dle způsobu vyjádření těchto motivů prostřednictvím slovtvorných, resp. strukturních modelů a typů), patří např. modelová klasifikace antroponym V. Blanára, dále modelová klasifikace oikonym R. Šrámka a modelová klasifikace pomístních jmen J. Pleskalové.

Jak uvádí J. Pleskalová, modelová klasifikace není omezena pouze na určitý druh proprií, nýbrž „jak toponymické, tak antroponymické modely jsou obecně platné a v různých jazycích navzájem srovnatelné. Postihují základní rysy pojmenovacího systému“.²³ Pro klasifikaci porejonym jsem

17 HUBÁČEK, J. (2009): K propriím na železnici. In: M. Harvalík – E. Minářová – J. M. Tušková (eds.), *Teoretické a komunikační aspekty proprií: Prof. Rudolfovi Šrámkovi k životnímu jubileu*. Brno, s. 283–285.

18 Tj. přibližně polovina prosince.

19 LYSONĚK, T. (2006): České dráhy změni jména vlaků, Šohaj už nepojede. *iDnes.cz* [online]. Cit. 2013-10-30. <http://ekonomika.idnes.cz/ceske-drahy-zmeni-jmena-vlaku-sohaj-uz-nepojede-frg-/ekonomika.aspx?c=A060914_202142_ekonomika_ad>.

20 ŠKAPÍKOVÁ, Š. (2009): Emanuel Rádl, Kiss, Šumava. To jsou české vlaky. *Denik.cz* [online]. Cit. 2013-10-30]. <http://www.denik.cz/ekonomika/jmena_vlaky20090222.html>.

21 Netradiční způsob zviditelnění firmy: pojmenujte po ní vlak (2008). *Novinky.cz* [online]. Cit. 2013-10-30. <<http://www.novinky.cz/kariera/144878-netradicni-zpusob-zviditelneni-firmy-pojmenujte-poni-vlak.html>>.

22 FICNAROVÁ, M. (2009): *Pomístní jména na katastrálním území Bohutín a Nový Malín (okres Šumperk)*. Brno.

23 PLESKALOVÁ, J. (1992): *Tvoření pomístních jmen na Moravě a ve Slezsku*. Jinočany, s. 15.

jako vzor zvolila MKPJ, současně však bylo třeba brát na zřetel specifika, kterými se tato skupina vlastních jmen vyznačuje vůči pomístním jménům (dále PJ),²⁴ a na základě rozdílů mezi oběma skupinami VJ modelovou klasifikaci poněkud modifikovat. V následujícím teoretickém výkladu vycházíme z MKPJ,²⁵ aplikujeme ji však na shromážděný porejonymický materiál.

Základní pojmenovací motivy v klasifikaci porejonym vystihuje **motivační model** (dále jen **MM**), v němž se rozlišují čtyři kategorie:

MM A – poloha nebo směr pojmenovávaného objektu (*Bečva, Hungaria, Praděd*)

MM B – druh pojmenovávaného objektu (*Karlex*)²⁶

MM C – vlastnost pojmenovávaného objektu, a to inherentní (*Excelsior*), nebo vnější, založená na vztahu k jinému objektu (případně okolnosti) (*Beruška, Csárdás, Salubia*)

MM D – vztah k osobě, a to i právnické (včetně výrobků, jimiž jsou právnické osoby reprezentovány, a akcí, jež právnické osoby pořádají a organizují) (*Budvar, Smetana*)

V případě názvů dvouslovných a víceslovných jsou tyto jednoduché MM kombinovány do MM dvoučlenných (*Křivoklátský expres* → A+B) a vícečlenných (*Výletní motoráček KŽC* → C+B+D), přitom člen základní je ve vzorci podtržen, vedlejší je vyjádřen jako +A+, +B+, +C+, +D+.

Ve shromážděném materiálu jsou zastoupena také porejonyma, u nichž není možné určit, který člen je hlavní a který vedlejší (např. *Balt-Orient*). V takovém případě klasifikují oba členy jako hlavní a tuto skutečnost značím podtržením obou členů (*Balt-Orient* → MM A+A).

Velmi početnou skupinu představují porejonyma, jež byla vytvořena transonymizací, tedy procesem, v němž vzniká proprium beze změny slovtvorné struktury fundujícího propria (nebo proprií).²⁷ V případě takto vytvořených dvouslovných a víceslovných porejonym (např. *Děčínský Sněžník, Gustav Klimt, Johann Gregor Mendel, Rádio Kiss Jižní Čechy*) uvádím MM pouze jednočlenný (*Děčínský Sněžník* → A, *Gustav Klimt* → D, *Johann Gregor Mendel* → D, *Rádio Kiss Jižní Čechy* → D), neboť bez ohledu na počet členů odkazují tato propria vždy pouze k jedné osobě, jednomu místu, jedné společnosti.

Aby od výše uvedené skupiny porejonym bylo možné odlišit názvy spojů, jež sice také odkazují k jedné osobě, avšak jejich vznik byl motivován spojením apelativum – proprium (tzn. při jejich tvoření bylo využito jak transonymizace, tak i proprializace; typ *Primátor Zenkl, Praotec Čech*), klasifikují tuto skupinu jako D+D.

Rovněž MM je podřazena rovina strukturních modelů (SM) a strukturních typů (ST), jež vystihuje způsob vyjádření jednotlivých pojmenovacích motivů. Strukturní model tedy postihuje konkrétní pojmenovací postup.

24 Pomístní jména označují terénní úvary, které jsou v krajině pevně fixovány. Porejonyma naproti tomu pojmenovávají objekty pohybující se po určité trase, směřující z výchozího bodu do bodu cílového. Prakticky to znamená, že takový objekt může být pojmenován podle jakéhokoli znaku, který souvisí (mnohdy i velmi volně) právě s jeho trasou (*Bečva, Bohemia*), cílovou či výchozí stanicí (*Cassovia, Děčínský Sněžník*), s určitou vlastností, např. rychlostí, kvalitou (*Slovenská strela, SC Pendolino Sprinter, Excelsior*). Některá porejonyma (zejména ta nově vytvořená v posledních letech) vzbuzují dojem, že vznikla zcela náhodně (*Kahan, Beruška*) či pouze z důvodů komerčních (*Rádio Kiss Jižní Čechy, www.pohadkovekralovstvi.cz*).

25 Ačkoli v MKPJ pracuje J. Pleskalová s pojmem vztahový model (VM), považuje za vhodnější termín motivační model (MM), jenž je užíván v antroponymii. J. Pleskalová tak respektuje v toponymické klasifikaci již zažitý pojem vztahový model, který pro oikonyma zavedl R. Šrámek. Při klasifikaci porejonym pracuji s termínem motivační model.

26 Porejonymum *Karlex* vzniklo ze spojení *Karlovarský expres*.

27 KARLÍK, P. – NEKULA, M. – PLESKALOVÁ, J. (2002): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, s. 347.

Při tvoření porejonym se uplatňuje:

- transonymizace (*Budvar, Doupovská dráha, Jan Perner, Sport Karta Vltava-Dunaj*)
- proprializace (*Amicus, Klidná síla, Manažer*)
- derivace (*Boleslavský motoráček, Šenovský motoráček*); sem jsou řazena také subst. v pl. (*Posázavské linky*), porejonyma vytvořená konverzí (*Beskyd*) či smíšeným způsobem prefixálně-sufixálním (*Podbrdské linky, Podřipský motoráček*).
- abreviace (*Karlex*)

Porejonyma typu *Balt-Orient* (MM A+A) a typu *Primátor Zenkl, Praotec Čech* (MM D+D) na úrovni strukturních modelů plně odpovídají MM. Klasifikuji je proto jako S+S (*Balt-Orient*) a S+S (*Primátor Zenkl, Praotec Čech*).

U dvoučlenných a vícečlenných porejonym vytvořených transonymizací vyjadřují počet členů číselně formou dolního indexu (*Děčínský Sněžník* → MM A, SM S₂, *Johann Gregor Mendel* → MM D, SM S₃, *Rádio Kiss Jižní Čechy* → MM D, SM S₄).

SM je potom vyjádřen konkrétním ST tvořeným určitým formantem (*Beskyd* → ST -0).²⁸

Výsledky klasifikace porejonym**Motivační modely**

Nejpočetněji zastoupeným motivačním modelem je MM A, z celkového počtu 357 porejonym je polohou motivováno 153 názvů (43 %). Nepatrně za MM A zaostává MM D (vyjadřující vztah k osobě) se 132 porejonymy (37 %). Počet porejonym patřících k MM C (vlastnost pojmenovávaného objektu) a MM B (druh pojmenovávaného objektu) je shodný a v porovnání s dvěma předchozími modely výrazně nižší; k MM B i MM C patří 36 názvů (10 %).

Porejonyma z hlediska struktury

Ve shromážděném materiálu je obsaženo nejvíce porejonym **jednoslovných** (207 jmen; 58 %), porejonym **dvouslovných** je o více než třetinu méně (132 jmen; 37 %) a porejonyma **víceslovná** představují s 18 výskyty pouhých 5 %.

Pokud přihlídneme také k MM, můžeme konstatovat, že **jednoslovná** porejonyma naprosto převažují u MM A, neboť ze 153 porejonym je celkem 140 jednoslovných (tj. 92 %). K MM C náleží jednoslovných názvů 14 (39 %) a MM D obsahuje 52 takových jmen (39 %).

Dvouslovná porejonyma jsou nejvýrazněji zastoupena v rámci MM B, kde při počtu 28 jmen tvoří 78 %. Následuje MM C, v němž je dvouslovných porejonym 21, což je více než polovina všech názvů (58 %), a MM D se 70 porejonymy (53 %).

Nepočtená **víceslovná** porejonyma zcela chybí u MM A, u MM B tvoří téměř jednu pětinu (7 jmen, 19 %) a u MM D 8 % (10 jmen).

Strukturní modely

Počet SM, jež se vyskytují v našem souboru porejonym, není příliš vysoký. To je způsobeno zejména skutečností, že se shodné SM objevují napříč různými MM (např. S, S₂, S+S, ADd+S) a že některým SM náleží velký počet porejonym (bohatě jsou zastoupena především porejonyma MM A a MM D, jež vznikla transonymizací a patří ke SM S a S₂).

Motivace **polohou** je (s výjimkou názvů *Beskyd, Krušnohor, Balt-Orient, Kolečovka 2012, Kolečovka 2013*) u všech 152 jednoslovných (SM S) i dvouslovných (SM S₂) porejonym realizována pře-

28 PLESKALOVÁ, J. (1992): *Tvoření pomístních jmen na Moravě a ve Slezsku*. Jinočany, s. 13–22.
KARLÍK, P. – NEKULA, M. – PLESKALOVÁ, J. (2002): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, s. 213–214 a s. 318.

nesením již existujícího názvu na vlakový spoj. Ten je nejčastěji pojmenován podle **vodního toku**, neboť trasa vlaku v mnoha případech vede právě podél potoků a řek a jejich názvy je tak možné přenášet na vlakové spoje (porejonyma *Agara, Bečva, Berounka, Bílina, Bobrava, Cidlina, Doubava, Dřevnice, Hornád, Hron, Hruškovice, Chomutovka, Chrudimka, Jevišovka, Jihlava, Jizera, Kamenice, Kyjovka, Kysuca, Labe, Laborec, Litava, Lužnice, Malše, Metuje, Nežárka, Nisa, Nysa, Odra, Ohře, Olšava, Olše, Olza, Orlice, Ostravica, Otava, Ploučnice, Punkva, Rokytka, Rokytná, Sázava, Svatava, Svitava, Třebovka, Úpa, Vlára, Vltava*). V rámci MM A jsou však využívány také další významné orientační body nacházející se na nebo v blízkosti trati, po níž vlak jede, a to **pohoří, hory, vrcholy, kopce, propasti** (*Avala, Blaník, Bradlo, Děčínský Sněžník, Fatra, Javorina, Javorník, Javořina, Ještěd, Křížová hora, Kunětická hora, Laborec, Landek, Libín, Macocha, Makyta, Milešovka, Pancíř, Porta Bohemica, Praděd, Radhošť, Radobýl, Říp, Santon, Sněžka, Sněžník, Špičák, Šumava, Tanečnicka, Vihorlat, Vysočina, Žuráň*), **hrady, zámky, kláštery** (*Bezděz, Bouzov, Buchlov, Cimburk, Devín, Drábské světničky, Hedeč, Helfštýn, Hukvaldy, Choustník, Karlštejn, Konopiště, Kotnov, Kyšperk, Lednice, Nový Světlov, Pernštejn, Petrov, Radyně, Sanssouci, Strékov, Svojanov, Špilberk, Švihov, Točnick, Větruše, Vranov*). Méně často **města, městské části, obce a osady** (*Budatín, Cassovia, Hradčany, Klokoty, Kozlov, Ležáky, Luna, Nesebar, Praha, Slavkov, Varsovia, Velehrad, Vindobona, Vyšehrad*), **vodní plochy**, např. **nádrže, rybníky, moře** (*Bezdrev, Hejtman, Jadran, Jordán, Rozkoš, Rožmberk, meandr Srdce Vltavy, Širava, Vajgar*), **tratě a dráhy** (*Doupovská dráha, Švestková dráha*) či **městský park** (*Javorka*). Specifickou kategorií představují vlaky pojmenované latinskými **názvy správních území a celků**, jež tvoří jednu z nejstarších skupin porejonym (*Bohemia, Hungaria, Moravia, Pannonia, Polonia, Saxonia, Silesia, Slovakia*). V posledních letech je možné sledovat nárůst porejonym pojmenovaných podle (euro)regionů, jež vznikají na objednávku právnických osob a slouží k propagaci daného (euro)regionu (řadíme je k MM D, viz níže; např. *Donau Moldau, Dunaj Vltava, Egrensis, Glacensis, Kahan, Silva Nortica*).

Nejčastější strukturní modely vyskytující se u porejonym motivovaných primárně **druhem pojmenovávaného objektu** (MM B) jsou ADd+S_d (porejonyma *Babylonský motoráček, Bezdězský motoráček, Boleslavský motoráček, Kamenický motoráček, Lužický motoráček, Morávecký motoráček, Polabský motoráček, Polenský motoráček, Posázavské linky, Posázavský motoráček, Pražský motoráček, Roudnický motoráček, Sněžnický motoráček, Středohorský motoráček, Šenovský motoráček, Zlonický motoráček*), dále ADd+S (Chomutovský expres, Královéhradecký expres, Zlínský expres) a dADd+S_d (Podbrdské linky, Podbrdský motoráček, Podlipanský motoráček, Podřipský motoráček). Základem všech těchto názvů je apelativní vyjádření druhu pojmenovávaného objektu (*expres, motoráček, linky*), které je blíže specifikováno adjektivem utvořeným ze zeměpisného názvu (resp. z logonyma u porejonyma *Babylonský motoráček*) a vyjadřujícím polohu (resp. vztah k právnícké osobě u *Babylonského motoráčku*). Zbývající dvouslovná porejonyma pak patří k SM S+S (*Cyklovlak Lipno, Cyklovlak Okoř, SC Pendolino, SC Smetana*) a S+S (*D1 Express*). U sedmi víceslovných porejonym náležejících k MM B je škála SM pestrá, avšak jejich početní zastoupení nízké (S+S₂ SC *Antonín Dvořák, SC Slovenská strela*; S+S₃ SC *Johann Gregor Mendel*; S+P+S *Vlak na Dobříš*; S+S+S SC *Pendolino Sprinter*; ADd+S_d+S *Výletní motoráček KŽC*). Jediný zástupce jednoslovných porejonym *Karlex* patří k SM S₂abr.

Také u porejonym zařazených k MM C (vyjadřujících **vlastnost či okolnost** spjatou s pojmenovávaným objektem) je častý SM ADd+S (*Brněnský drak, Červnová noc, Dětský drak, Jemnická dráha, Křivoklátský expres, Kutnohorský expres, Mikulovský expres, Mladějovský výlet, Pálavský okruh, Slovácký vinohrad, Slovenská strela, Švýcarská kvalita, Vánoční kometa, Žatecký chmel*). U dvouslovných porejonym se dále objevují nepoččetně zastoupené SM S+S (*Generál Laudon, Velo Podyjí*), S+NUM (*Zubačka 2011, Zubačka 2013*), AD+S (*Klidná síla*), Sd/S+S (*Křivoklát Ex*) a SM S₂ (*Svatý Jiří*). Jednoslovná porejonyma patří nejčastěji k SM S, jež sdružuje názvy vytvořené

buď proprializací (*Beruška, Csárdás, Cykloturistika, Manažer, Večernice, Velo*), či transonymizací (původní kosmonyma *Canopus, Phoenix*, antroponymum *Mikuláš* či theonymum *Neptun*). Pouze malé zastoupení má SM SADd (lat. porejonyma vytvořená transonymizací *Amicus, Excelsior*) a SM Sd (*Metropol, Salubia*). Jediné víceslovné porejonymum *Karlovarský Chodovar Ex* náleží k SM ADd+S+S.

Jednoslovná, dvouslovná i víceslovná porejonyma MM D, jež reflektují široce chápaný **vztah k osobě**, vznikla v naprosté většině případů transonymizací. Z celkového počtu 132 jmen je vytvořeno jiným způsobem pouze 10 porejonym (SM S+S *ČD Wellness, ČD Yetti, Praotec Čech, Primátor Zenkl*; SM ADd+S *Náchodský Primátor, Přerovský Zubr, Vltavský vodák*; SM AD+S *Svijanský speciál*, SM S+NUM *SKUPINA PĚTI* a SM S+S+S *Muzeum ČD Lužná*). Nejfrekventovanějšími SM u jednoslovných a dvouslovných porejonym v rámci MM D jsou tak SM S, S₂ (obdobně jako u MM A). K těmto SM patří názvy transonymizované ze **jmen významných osobností**, a to ze jmen **hudebních skladatelů** (*Andreas Hartauer, Anton Bruckner, Antonín Dvořák, Franz Schubert, Gustav Mahler, Chopin, Johannes Brahms, Leoš Janáček, Matalík, Smetana, Zdeněk Fibich*), **panovníků, šlechticů a vojevůdců** (*Báthory, Bem, Jan Žižka, Karel I., Mathias Corvinus, Matúš Čák, Mojmír, Petr Vok, Rastislav, Sobieski, Svatopluk, Valdštejn*), **spisovatelů** (*Franz Kafka, Goethe, Heinrich Heine, Jaroslav Hašek, Karel Čapek, Karel Klostermann, Rudolf Těsnohlídek*) a **dalších umělců** (malířů, sochařů, architektů, loutkoherců; *Antonín Slavíček, Gustav Klimt, Josef Skupa, Jóže Plečnik, Matthias Braun, Santini*), **vědců a vynálezců** (*Albert Einstein, Emanuel Rádl, František Křižík, Galileo Galilei, Jan Kašpar, Johannes Kepler, Josef Ressel, Kopernikus, Metoděj Vlach, Viktor Kaplan*). Z hlediska motivace je velmi zajímavá skupina jmen sdružující **osobnosti, jejichž činnost byla spjata s železnicí a železniční dopravou** (*Alois Negrelli, Jan Perner, Josef Rosenauer, Matthias Schönerer, Vojtěch Lanna, Wenzel Draxler*). Dále byla porejonyma motivována jmény významných, **pedagogů** (*Comenius, Ferdinand Kindermann*), **studentů** (*Jan Palach*), **lékařů** (*Jan Jesenius*), **kazatelů** (*Jan Hus*), **sportovců** (*Emil Zátopek*), **rybníkářů a lesníků** (*Jakub Krčín, Josef John*), **zbojníků a selských hrdinů** (*Jakub Kubata, Ondráš*). Specifickou skupinu představují názvy utvořené z **nepravých antroponym, jmen bájných či literárních postav** (*Josef Švejk, Krakonoš, Peklík, Praotec Čech, Radegast*). Zastoupena jsou také porejonyma utvořená ze **jmen obyvatelských** (*Detvan, Horehronec, Hradišťan, Košičan, Moravan, Ostravan, Podlužan, Slezan, Vsacan*) a **etnonym** (*Slovan*). Proprializací apelativního označení osob vznikla porejonyma *Galán, Hutník, Junák, Loutkař, Ogar, Permoník, Portáš, Šohaj a Templář*.

V posledních letech se stále častěji objevují pojmenování motivovaná **právníckými osobami, institucemi a společnostmi** (např. *Ohrada, Peklák*) nebo **produkty**, jež tyto subjekty vyrábí a nabízí (např. pragmatonyma *Mladý svět, Vltava Dunaj*), či **akcemi**, které pořádají (akcionyma *Den železnice, Euroorando 2011, Kopečková pouť, Pernštejnské panství*). V rámci této kategorie se vyděluje několik výrazných podskupin. První z nich tvoří porejonyma motivovaná jmény **dopravních společností** (*Arriva, LEO Express, Posázavský pacifik, RegioJet, RT Express, Trilex*), další pak jmény **pivovarů a jejich produktů** (*Budvar, Chodovar, Královský lev, Náchodský Primátor, Rychtář, Samson*), **euroregionů** (*Donau Moldau, Dunaj Vltava, Egrensis, Glacensis, Kahan, Silva Nortica*) či **internetových portálů** (*www.lipnoservis.cz, www.mestodobris.cz, www.pohadkovekralovstvi.cz*). Nepočtená víceslovná porejonyma patřící ke SM S₃ jsou (s výjimkou názvu spoje převzatého ze **jména dopravce** *WAGON SERVICE travel*) vytvořena na základě **jmen osobních** (*F. A. Gerstner, Jan Nepomuk Neumann, Johann Gregor Mendel, Josef Václav Myslbek, Karel Havlíček Borovský*). U SM S₄ se **jméno významné osobnosti** (konk. hudebního skladatele) objevuje pouze jedenkrát (*Carl-Maria von Weber*). Jména motivovaná **právníckou osobou** (resp. názvem jejího produktu) jsou doložena dvě (*Rádio Kiss Jižní Čechy* a *Sport Karta Vltava-Dunaj*).

Pojmenování vlakových spojů výše uvedenými jmény je možné díky **prostorové souvislosti**, neboť

naprostá většina fyzických i právnických osob má spojitost s místem na trase vlaku či v blízkosti trati, po níž vlak jede. Výjimku představují pouze některé reklamní názvy (např. *Eurorando 2011*). Jak je patrné z uvedených příkladů, jako motivace porejonym slouží jména osob světově proslulých (*Albert Einstein, Comenius, Franz Kafka, Galileo Galilei, Goethe*), ale i osob známých pouze v ČR, resp. spjatých s určitou oblastí či regionem ČR (např. *Jakub Krčín, Matalík, Ondráš, Rožmberk, Vojtěch Lanna*). Zatímco první skupinou názvů se honosí spoje mezinárodní (nejčastěji kategorie EC), druhá skupina porejonym doprovází vlaky méně významné, a to vnitrostátní, regionální či výletní a nostalgické (kategorie R, Sp, Os).

Proměny porejonym v čase

V této části se zaměřujeme na vývoj porejonym od pol. 60 let až do současnosti. Během posledních padesáti let došlo nejen k výraznému **nárůstu porejonym**, ale také ke **změnám v jejich motivaci**. Nárůst porejonym v uvedeném období vyplývá především z proměn vnímání názvů vlakových spojů. Zatímco v **60. letech** byly vlastním jménem označovány pouze vlaky **prestižní**, většinou mezinárodní, jež nad ostatními vynikaly kvalitou, rychlostí či komfortem (např. *Vindobona, Hungaria, Saxonia*), v **současné době** jsou hojně pojmenovávány také **vlaky osobní či spěšné** (nostalgické a výletní) jedoucí pouze v určité dny či období (*Cykloturistika, Cyklovlak Lipno, Cyklovlak Okoř, ČD Yetti, Pálava, Pálavský okruh, Vánoční kometa, Velo, Velo Podují, Vltavský vodák*) nebo dokonce v den jediný (*Červnová noc, Drábské světničky, Karlovarský Chodovar Ex, Praotec Čech, Svijanský speciál*). Rozšíření názvů na vlaky „nižších“ kategorií (Os, Sp) je spjato s vnímáním porejonyma jakožto vhodného **marketingového nástroje**. Jestliže ještě z poloviny 90. let nejsou reklamní porejonyma vůbec doložena, dnes takové názvy představují progresivní skupinu a objevují se prakticky v každém novém JŘ. „Zakoupit si“ na určitou dobu svůj vlak a pojmenovat jej může dnes v podstatě každý, kdo je ochoten vynaložit za tímto účelem finanční prostředky.²⁹ Většinou však této možnosti využívají **právnické osoby**. Frekventovaná jsou porejonyma nesoucí název konkrétního pivovaru či jeho produktu (*Budvar, Chodovar, Královský lev, Náchodský Primátor, Rychtář, Samson*), porejonyma reflektující dopravce daného spoje (*Arriva, ČD Wellness, ČD Yetti, LEO Express, Posázavský pacifik, RegioJet, RT Express, Trilex, WAGON SERVICE travel*), nebo již zmiňované transonymizované názvy euroregionů (*Dunaj Vltava, Donau Moldau, Egrensis, Glacensis, Kahan, Silva Nortica*). V posledních letech se objevila také porejonyma nesoucí název zoologické zahrady (*Ohrada*) či střediska sdružujícího lyžařský areál (Ski areál *Peklák*), lanový park (Lanový park *Peklák*) a bike park (Bike park *Peklák*).

Výše popsany jev není typický pouze pro ČR, neboť i v **okolních státech** existuje v současné době velké množství reklamních názvů.

Na **Slovensku** v posledních šesti letech jezdily např. tyto vlaky: R *Obchodné centrum Mirage*, R *Aupark*, R *Kúpele Lúčky*, Zr *Electro world*, IC *Gorenje*, R *Mall.sk*, REx *SLOWAKISCHEN NATIONALMUSEUM*, REx *Technisches Museum Wien*, REx *IHR ZUG ZUR THERME WIEN*, REx *eti-slovensko.sk*, REx *Twin City* (všechny vlaky REx jedou na trase Wien Südbf. – Bratislava hl. st.), R *Avízo*, IC *Zelmer*, R *Quelle*, IC *Slovenka* a IC *UNION POISŤOVŇA*. Situace zde zašla tak daleko, že názvy vlakových spojů, jejichž původním účelem bylo přepravu usnadňovat, začaly být pro cestující matoucí, neboť vlakový spoj se jmenoval v jednom směru jinak než v druhém. Trendem pojmenování vlakových spojů se počátkem r. 2013 zabývalo Ministerstvo dopravy SR a jednalo

29 Např. název *Klidná síla* si politická strana KDU-ČSL zakoupila za 80 000 Kč; (2005): Rychlík Klidná síla vlakem KDU-ČSL. *ŽelPage* [online]. Cit. 2011-11-30. <<http://www.zelpage.cz/zpravy/2104>>.

o prodeji některých tradičních názvů s Železničnou společností Slovensko, jíž prostředky z prodeje zvyšují tržby a současně snižují nároky na dotaci od státu. Přestože původní informace zněla, že k nahrazení tradičních názvů jako *Tatran*, *Gerlach*, *Rysy*, *Kriváň*, *Váh*, *Liptov*, *Ružín* nedojde, média uvádí, že v novém JŘ budou tato porejonyma nahrazena reklamními názvy *Martinus.sk*, *Lyonesse*, *Doxxbet*, *Horalky Sedita*, *Shooppie.s*, *Okay Elektro*).³⁰

Poměrně kuriózním způsobem pak vznikl v r. 2009 název jednoho **polského rychlíku**. Jméno vlaku *Anastazja Skopowska* neoznačuje žádnou významnou osobnost, jak bývá u porejonym tohoto typu zvykem, nýbrž je historicky prvním polským názvem, který byl vydražen v on-line aukci. Rozhodující příhoz učinil *Janusz Skopowski*, jenž za částku převyšující 100 000 Kč získal možnost pojmenovat vybraný vlak a zvolit si také jeho trasu. Pan *Skopowski* připravil tímto počinem velmi originální dárek k osmnáctým narozeninám své dcery *Anastazje* a současně vydraženou částkou přispěl na zakoupení diagnostických přístrojů pro včasné odhalení rakoviny u dětí.³¹

Pokud se podíváme na to, jaké **motivační modely** se uplatňovaly v souborech porejonym shromážděných jednotlivými badateli od r. 1966 až do r. 1994, dojdeme k závěru, že vždy zcela převažovaly názvy vyjadřující polohu. V roce 1966 byly frekventované lat. názvy správních území (*Bohemia*, *Hungaria*, *Pannonia*) a jména typu *Balaton-Jadran*, *Balt-Orient*, *Nord-Orient* vyjadřující výchozí stanici a směr jízdy vlaku. Vztah k osobě motivoval vznik pouze tří porejonym (*Chopin*, *Ostravan* a *Tatran*). V roce 1975 se u porejonym motivovaných polohou nově objevují např. názvy vodstev (*Hron*, *Lipno*), hradů (*Devín*, *Sanssouci*), hor a pohoří (*Vihorlat*) atd. Jen málo názvů vzniká transonymizací nebo proprializací jmen osobních (*Báthory*, *Horník*, *Hutník*). Z tohoto roku jsou doloženy také metaforické názvy vyjadřující vlastnost pojmenovávaného objektu (*Brněnský drak*, *Slovenská strela*, *Salubia*). Přestože v dalších letech počet nových jmen neustále narůstal, soubor porejonym z r. 1988 obsahuje stále nejvíce názvů motivovaných polohou a tento stav přetrvává i v roce 1994 (Pokorný eviduje celkem 70 názvů převzatých ze jmen zeměpisných). V r. 1994 jsou poprvé ve větším počtu zaznamenána pojmenování spojů podle významných osobností (celkem 22 názvů) a podle jmen obyvatelských (20 názvů). Tyto údaje ukazují, že jména osobní jakožto zdroj pro pojmenování vlakových spojů byla v polovině 90. let na vzestupu, což ostatně potvrzují výsledky, k nimž jsme dospěli my – porejonym patřících k MM D je v našem souboru obsaženo 132 (37 %), což je o 21 názvů méně v porovnání s nejfrekventovanějším MM A (153 jmen; 43 %). Kromě transonymizace jmen významných osobností a jmen obyvatelských se na tomto nárůstu podílí již zmiňovaná jména právnických osob a jejich produktů.

30 PORACKÝ, M. (2013): *Železnice chcu zarobit. Z vlaku budú Horalky*. HNonline.sk [online]. Cit. 2013-10-30. <<http://hn.hnonline.sk/ekonomika-a-firmy-117/zeleznice-chcu-zarobit-z-vlaku-budu-horalky-583377>>.

Aspirín ľudí nepovezie, vlaky Kriváň, Tatran či Spišan si ponechajú historické názvy (2013).

Aktuálne.sk [online]. Cit. 2013-11-30. <<http://aktualne.atlas.sk/aspirin-ludi-nepovezie-vlakry-krivan-tatran-ci-spisan-si-ponechaju-historicke-pomenovania/slovensko/spolocnost/>>.

31 *Anastazja Skopowska – nový název rychlíku Praha – Wrocław* (2009). *ŽelPage* [online]. Cit. 2011-11-30. <<http://www.zelpage.cz/zpravy/6865>>.

Dodatek

ADd	adjektivum derivované
ČD	České dráhy
ČR	Česká republika
dADd	adjektivum utvořené prefixálně-sufixálním způsobem
EC	EuroCity
IC	InterCity
JŘ	jízdní řád
MKPJ	modelová klasifikace pomístních jmen
MM	motivační model
NUM	numerale
Os	osobní vlak
PJ	pomístní jméno
R	rychlík
r.	rok
REx	Regional Express
S	substantivum nederivované
S ₂	substantivum dvouslovné vytvořené transonymizací
S ₃	substantivum tříslavné vytvořené transonymizací
S ₄	substantivum čtyřslavné vytvořené transonymizací
SADd	substantivizované adjektivum derivované
Sd	substantivum derivované
SM	strukturní model
Sp	spěšný vlak
ST	strukturní typ
VJ	vlastní jméno
VM	vztahový model
Zr	zrychlený vlak

Seznam literatury

- FICNAROVÁ, M. (2009): *Pomístní jména na katastrálním území Bohutín a Nový Malín (okres Šumperk)*. Brno.
- FICNAROVÁ, M. (2012): *Vlastní jména vlakových spojů (porejonyma) na území ČR v letech 1993–2011*. Brno.
- GALEČKOVÁ, M. (1988): *Názvy vlaků*. Ostrava.
- GARANČOVSKÁ, L. (2009): *Vymedzenie chrématonym v onymickom systéme*. *Acta onomastica*, 50, s. 88–101.
- HONL, I. (1970): *O jménech vlaků*. *Zpravodaj místopisné komise*, 11, s. 9–15.
- HUBÁČEK, J. (2009): *K propriím na železnici*. In: M. Harvalík – E. Minářová – J. M. Tušková (eds.), *Teoretické a komunikační aspekty proprií: Prof. Rudolfovi Šrámkovi k životnímu jubileu*. Brno, s. 283–285.
- HUBÁČEK, J. (1985): *Zur Problematik der Zugnamen*. *Onomastický zpravodaj*, 26, s. 91–100.
- HUBÁČEK, J. (1966): *K názvům vlaků*. *Naše řeč*, 49, s. 151–154.
- KARLÍK, P. – NEKULA, M. – PLESKALOVÁ, J. (2002): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny.
- KUBA, L. (2007): *Proces nominace dopravních prostředků provozovaných na území našeho státu*. *Acta onomastica*, 48, s. 79–107.
- LYSONĚK, T. (2006): *České dráhy změni jména vlaků, Šohaj už nepojede*. *iDnes.cz* [online]. Cit. 2013-10-30. <http://ekonomika.idnes.cz/ceske-drahy-zmeni-jmena-vlaku-sohaj-uz-nepojede-frg-/ekonomika.aspx?c=A060914_202142_ekonomika_ad>.
- PLESKALOVÁ, J. (1992): *Tvoření pomístních jmen na Moravě a ve Slezsku*. Jinočany.
- POKORNÝ, L. (1996): *Vlastní jména dopravních spojů (porejonyma) a jejich jazykové zpracování: diplomová práce na PF UJEP Ústí nad Labem*. *Acta onomastica* 37, s. 139–146.
- POKORNÝ, L. (1995): *Vlastní jména dopravních spojů (porejonyma) a jejich jazykové zpracování*. Ústí nad Labem.
- PORACKÝ, M. (2013): *Železnice chcu zarobit. Z vlaku budú Horalky*. *HNonline.sk* [online]. Cit. 2013-10-30. <<http://hn.hnonline.sk/ekonomika-a-firmy-117/zeleznice-chcu-zarobit-z-vlaku-budu-horalky-583377>>.
- SPAL, J. (1982): *Z onomastiky železniční*. *Zpravodaj místopisné komise*, 23, s. 343–348.
- ŠKAPÍKOVÁ, Š. (2009): *Emanuel Rádl, Kiss, Šumava. To jsou české vlaky*. *Denik.cz* [online]. Cit. 2013-10-30. <http://www.denik.cz/ekonomika/jmena_vlaky20090222.html>.
- ŠMILAUER, V. (1975): *Jména rychlíků*. *Zpravodaj místopisné komise*, 16, s. 555–557.
- Aspirín ľudí nepovezie, vlaky Kriváň, Tatran či Spišan si ponechajú historické názvy* (2013). *Aktualne.sk* [online]. Cit. 2013-11-30. <<http://aktualne.atlas.sk/aspirin-ludi-nepovezie-vlak-krivan-tatran-ci-spisan-si-ponechaju-historicke-pomenovania/slovensko/spolocnost/>>.
- Anastazja Skopowska – nový název rychlíku Praha – Wrocław* (2009). *ŽelPage* [online]. Cit. 2011-11-30. <<http://www.zelpage.cz/zpravy/6865>>.
- Netradiční způsob zviditelnění firmy: pojmenujte po ní vlak* (2008). *Novinky.cz* [online]. Cit. 2013-10-30. <<http://www.novinky.cz/kariera/144878-netradicni-zpusob-zviditleni-firmy-pojmenujte-poni-vlak.html>>.
- Rychlík Klidná síla vlakem KDU-ČSL* (2005). *ŽelPage* [online]. Cit. 2011-11-30. <<http://www.zelpage.cz/zpravy/2104>>.

ANALÝZA SEKULARIZOVANÝCH SLOVNÍCH SPOJENÍ MEZI LEXIKONEM, GRAMATIKOU A DISKURZEM: PŘÍPAD *NOEMOVY ARCHY* A *DUCHA SVATÉHO*

JOHANNES HEINKE

Humboldt-Universität zu Berlin

Abstract

This paper provides a side view into the ongoing work on the linguistic research project *The Secularization of Religious Vocabulary against the Background of Social Change – Polish, Czech, Slovak and German* an English description of which is to be found online: <<http://www.slawistik.hu-berlin.de/member/anagorko/secularization>>. I reflect methodical and theoretic questions regarding the influence of other layers of description as grammar and discourse on lexical and semantic research by studying two complex sacral nouns as *Noah's ark* and *The Holy Spirit*. In order to analytically sharpen the problem, I raise the question if *Noah's ark* can mean a *gene bank* in contemporary Czech.

Klíčová slova: *Duch svatý, Noemova archa, banka*; kognitivní sémantika, teolingvistika, kulturní lingvistika, (meta-) lexikografie, metodologie, korpusová lingvistika, hypertextová lexikální databáze

Keywords: *Noah's ark, The Holy Spirit, Bank*; cognitive semantics, theolinguistics, cultural linguistics, (meta-) lexicography, methodologies, corpus linguistics, hypertextual lexical databases

Úvod

Ve svém příspěvku se budu zabývat několika metodologickými otázkami vzniklými během práce na berlínském lingvistickém slovníku *Sacrum und Profanum* (SuP), který zachycuje sekularizaci náboženské slovní zásoby ve srovnání češtiny, němčiny, polštiny a slovenštiny.¹ Budu zde pojednávat především o české sekci projektu. Pod pojmem jazyková sekularizace zde operativně rozumím veškeré užití jazykových prostředků a veškerou aktualizaci jejích analytických částí, komponentů a rysů, které okazionálně a zčásti anebo pravidelně, naplno a ustáleně eliminují náboženskou povahu lingvistických jednotek na alespoň jedné z úrovní jazykového bádání. V popředí zájmu stojí otázka, jestli se náboženský význam v rámci celku jazykové komunikace (věta, výpověď, text, dis-

1 SuP: *Sacrum und Profanum*, Religiöse Lexik in der Allgemeinsprache (Deutsch – Polnisch – Slowakisch – Tschechisch) – © Humboldt-Universität zu Berlin, 2009–2013; DFG – Projekt. Vypracovává tým pod vedením prof. Alicji Nagórko na Katedře západoslovanských jazyků v rámci badatelského záměru „Säkularisierung religiöser Lexik (Kurztitel/ zkrácený název: Sekularizace náboženské slovní zásoby)“, podporovaným německou grantovou agenturou DFG (Deutsche Forschungsgemeinschaft). Slovník bude zveřejněn v roce 2014, odkaz a další aktuální informace o projektu, slovníku a třech mezinárodních konferencích jsou vyvěšené na stránkách projektu: <<http://www.slawistik.hu-berlin.de/member/anagorko/saekularisierung>>.

kurz atd.) posouvá vzhledem k jeho sakrálnosti a jestli ano, a jak, se tyto modifikace lexikalizují.² Příspěvek se týká především konferenčního tématu *tvorby a využití lexikálních databází* a jejich vztahu k jazykovým korpusům, poněvadž berlínský slovník vzniká jako hypertextová databáze a poskytuje jazykový materiál získaný rešeršemi v ČNK a konfrontovaný s výkladovými slovníky češtiny.³ Nejde zde však o technologie, které by používaly automatický transfer dat, ale o hypertextový specializovaný slovník, který zachycuje ve své extenzi malou část slovní zásoby intenzivním popisem a vzniká „ruční“ lexikografickou prací. Výsledkem je synteticky prezentovaný upravený a komentovaný materiál, který představuje výsledky sémantické analýzy, ale také může sloužit dalším, navazujícím analýzám a v tomto smyslu není „hotový“.⁴

Chtěl bych nastínit, kterým směrem by další analýzy mohly pokračovat, a proto se soustředím na koncepční otázky. Také bych chtěl umožnit vhléd do „slovníkářské dílny“ a poukázat na některá metodická úskalí a jejich řešení. Touto cestou se pokouším vysvětlit poněkud odlišnou koncepci berlínského slovníku naproti známým výkladovým a kodifikačním dílům. K tomuto účelu nejdříve uvádím několik údajů o slovníku všeobecně a jeho makrostrukturu. Dále se dostanu k popisu mikrostruktury na příkladu zvláštního případu dvouslovných lexémů typu *Noemova archa* a *Duch svatý*. V berlínském slovníku je tento typ v pozici heslového slova výjimkou, jeho podrobný výzkum však přináší poznatky nejen pro případné zpracování dalších slov podobného typu, ale dovoluje sondu do vztahu gramatické a pragmaticko-sémantické struktury náboženské (resp. kulturní) slovní zásoby. Posléze se pak přibližuje nečekanému významu slovního spojení *Noemova archa*, který může v určitých kontextech znamenat „genetická banka“. Zde kladu důraz na empirický přístup k neosémantismům a jejich zachycení v takzvaných profilech. Poslední otázka závěrem pak zní, jestli podléhají obě slovní spojení v dnešním užití procesu jazykové sekularizace. Zdůrazňuji, že se zde nepokouším o reprezentativní prezentaci celého projektu, ale že zpracuji vlastní průnik do problematiky. K zdůraznění tohoto faktu nepoužívám autorský plurál v mezích žánru, ale opravdu jenom tam, kde mluvím o týmové práci. Poněvadž podrobnější informace o projektu v češtině zatím chybí, doufám, že příspěvek může zatím vypomáhat i v této funkci.⁵

Slovník *Sacrum und Profanum*

Projekt o sekularizaci a jeho finální produkt, slovník *Sacrum und Profanum* sledují především dva badatelské cíle: kontrastivní a diachronní výzkum polysémie na straně jedné, teoretický a srovnávací výzkum vztahu jazyka a sociokulturního systému na straně druhé, a to na rovině historické sémantiky a jazykového obrazu světa, dynamiky jeho proměn.⁶ Cílový jazykový útvar je všeobecný

2 K rozsáhlé diskusi o definicích viz příspěvky ve sbornících vydaných v rámci projektu – třetí vyjde 2014: NAGÓRKO 2011b, NAGÓRKO 2012 a NAGÓRKO 2014, pro stručný přehled viz zprávu o konferenci (HEINKE 2012).

3 Plné znění použitých zkratk včetně korpusů, slovníků a internetových zdrojů uvádím v seznamu použité literatury na konci článku.

4 Samotný hypertextový slovník je programován tak, aby se dal rozšířit o další lemmata a případně i o další jazyky. Mám zde na mysli ale spíše možnost další práce s poskytovaným materiálem anebo metodikou.

5 Krátká charakteristika se najde v GRYGGERKOVÁ 2013. Danou tematikou se v české bohemistice věnují především E. Minářová, M. Lašovičková a M. Grygerková, v berlínském týmu pracují/ pracovali na české sekci Jana Marková a Jakub Kopecký.

6 Pro všeobecnější informace o teoretickém pozadí projektu a badatelském směru kognitivní sémantika zde odkazují na práce Alicji Nagórko (především 2007 a 2011a).

jazyk (psaný i mluvený) z poválečného období až do současnosti. Danou slovní zásobu projekt zachycuje formou lexikografie. Východiskem analýzy jsou význačné lexémy s alespoň jedním primárním náboženským a jedním nenáboženským významem (primárním anebo sekundárním). Hlavní zřetel analýzy se klade na sémanticko-pragmatickou sekularizaci konkrétních dílčích významů a sémantických aspektů náboženské povahy.

Byl sestaven seznam pojmově klíčových slov, označený jako *Berlínský seznam*⁷, který čítá 50 slov. Ke konci projektu bude zveřejněn rozšířený seznam, který obsahuje i periferně relevantní lexikum – pro polštinu čítá aktuálně až 600 lexikálních jednotek. Z 50 *klíčových slov* je devět sloves (*bičovat, křtít, modlit se, rouhat se, spasit, světít, věřit, zpovídat se, žehnat*), příslovce *košer*, jedno přídavné jméno (*svatý*), zbytek jsou jména podstatná, z nich tři deverbální. Převažují tedy jména a nominální výrazy. Mimo křesťansko-biblická slova jsou v menší míře zastoupeny také lexémy z jiných náboženství jako *karma* a *guru*. Slouží jako kontrastní příklady a seznam jenom doplňují. Ještě se zastavím u podstatných jmen, kde se pokouším alespoň o nástin charakteristiky z pohledu makrostruktury. Z formálních a strukturních zvláštností uvádím prozatím jenom několik základních faktů. U dvou se vyskytuje knižně-archaický genitiv plurálu *nebes* a *pekel*, tři začínají majuskulí (*Antikrist, Kristus* a *Noemova archa*), jenom jedno však patří do kategorie vlastních jmen, totiž *Noemova archa* – fakt, který působí podezřele vzhledem k našemu vymezení tematiky. Není k tomu ale důvod – *Janovo zjevení* najdeme v hesle *Apokalypsa*, *Ježíše* najdeme jako *Krista, Mesiáše* v hesle *spasit* v derivátech, tj. *spasitel* atd. Bez pečlivě stanovených kritérií zde není možné uvádět kvantitativní údaje o tom, která slova jsou domácího původu, zdomácnělá a cizí. Na příkladech ukážu později, že tyto i další nesémantické vlastnosti mají svůj dopad na sémantickou analýzu a zabývám se dalším gramatickým jevem, *singularii tantum*.

Co se týče sémaziologicko-lexikografické perspektivy, která stojí v popředí našeho zájmu, zbývá zde zdůraznit, že nepracujeme s obvyklým dělením lexému/lemmatu do významových skupin, významů apod. Namísto pojmu sém používáme pojem profil, pod kterým berlínský tým rozumí výběr různých sémantických vlastností/příznaků, které mluvčí chápe jako celost a přes určitý interpretační rámec (kategorii).⁸ Exploraci struktury a koncepce lexému provádíme vyčleňováním profilů, tj. identifikací významů a aspektů, které profily tvoří. Profily v jednom jazyku porovnáváme s dalšími třemi jazyky a se slovníky. Paralelně sondujeme české lexémy jako *key word in context* (KWIC) v dostupných digitálních korpusech. Kvůli těmto zdrojům empirického jazykového materiálu bereme více v úvahu současný žurnalistický a publicistický jazyk a předběžné analýzy stavěné na excerpční metodě, kde hraje větší roli i literární jazyk. Abychom se více přiblížili své původně vymezené cílové jednotce, doplníme tyto postupy o rešerše v českém internetu a mluvených korpusech, mnohem důležitější je však komparace s druhými jazyky a jazyková kompetence v projektu zastoupených rodilých mluvčích.

Zde je nutné vsunout poznámku, že ČNK podle rozlišení Belici a Steyerové (2008: 7–12) především konzultujeme a jenom v menší míře i analyzujeme, tj. ověřujeme údaje z lexikografické literatury a testujeme profilaci. Proto, že ve svých zde prezentovaných analýzách sleduji s teoretickým zájmem stopy neustálených sémantických reflexů v korpusových konkordancích, přistupuji k materiálu spíše kvalitativně a heuristicky než kvantitativně. To znamená mj., že statistické argumenty nejsou rozhodující. Proto upřednostňuji korpusové konkordance z nereferenčního korpusu SYN,

7 Tento pojem uvádí ve svém příspěvku na třetí mezinárodní badatelské konferenci germanista Albrecht Greule, srov. NAGÓRKO (2014).

8 Ke specifickému pojetí pojmu profil ve srovnávání s jinými koncepcemi viz Nagórko 2007 a 2011a. Srov. také Saicovou Římalovou (2010), která představuje profily a další termíny kognitivní lingvistiky v českém prostředí.

kteřé představují větší celek. Konkordance samotné se nemění, dokladů může jenom přibývat, což má vliv na makrostrukturu a statistiku, nikoliv na dostupnost a vnitřní strukturu konkordancí. Slovník *Sacrum und Profanum* nabízí svým uživatelům jak ověřený jazykový materiál z různých zdrojů, tak i teze k povaze sémanticko-gramatické struktury jazyka a jejich proměn. Vedlejší, ale podle mého soudu neméně zajímavým výsledkem jsou poznatky o vnitřní struktuře lexému, kterým jsou věnovány následující úvahy.

Případ *Ducha svatého* a *Archy Noemovy*

V rámci příkladové analýzy bych se nyní chtěl podrobněji zabývat dvěma klíčovými slovy z berlínského seznamu, lexémy *Noemova archa* a *Duch svatý*. Kromě reflexivních sloves *modlit se*, *rouhat se* a *zpovídat se* představuje *Noemova archa* jediné heslové slovo obsahující dvě grafické jednotky oddělené mezerou. Všechna zmíněná klíčová slova považujeme za samostatné lexikální jednotky, které zakotvují celou řadu významů a forem v makrostruktuře slovníku SuP (slovotvorné hnízdo, sémantické pole, etymologie atd.). *Duch svatý* se na berlínském seznamu nevyskytuje jako samostatné heslové slovo, ale jako frazém a také synonymum v rámci heslové stati *duch*. Původně však byly oba lexémy vybrány jako heslová slova vzhledem k tomu, že primární náboženský význam spočívá ve složeném lexému, *archa* a *duch* se ve vztahu k těmto lexémům chovají jako elipsy. Zachycená heslová slova představují pouze špičku ledovce, tedy umožňují badatelský přístup k různým složkám lexika a sémantiky jazyka, jakými jsou slovotvorná hnízda a synonymie. Distribuce sémantických vlastností na těchto rovinách je často ve všech jazycích dost různá. Proto mě zde nezajímá především formální lexikografický status a můj článek neodpovídá nutně finálnímu znění heslového výkladu ve slovníku. Vybral jsem je k analýze, poněvadž jsou jedinými námi zachycenými případy, kde komplexnější forma slova dominuje sémantice a tvoří hlavu hesla či alespoň profilu. Zajímají mě její vlastnosti a dopad těchto vlastností na lexikologický popis.

Sémantika *Ducha svatého* a kategorie množného čísla

Nutno dodat, že se v případě *Ducha svatého* konceptuální zacházení s tímto lexémem posouvalo souběžně se vznikem tohoto článku. Původně jsme uváděli jako hlavu prvního (trinitárního) profilu hesla *duch* reprezentativní formu *Duch svatý*, nyní zastřešuje výklad tohoto profilu už jenom forma *duch*. Tato změna je opodstatněná skutečností, že podobné koncepce „původní, věčné a životodárné síly“ nacházíme i v jiných náboženstvích, přičemž nás zajímají hlavně jiné biblické formy (*Duch Páně*, *Ježíšův Duch* apod.). Sevřenější novozákonní pojetí *ducha* jako části svaté trojice (*Duch*) ovládá český diskurz a dnešní synchronní polysémii z (kognitivního) hlediska mluvčího. Širším pojetím tohoto profilu se však lépe podařilo zachytit i periferii diskurzu a sémantickou souvislost s jinými profily. Následujících konkordance dokládají profil v celé šíři:

(1) *Země byla pustá a prázdná a nad propastnou tůňí byla tma. Ale nad vodami vznášel se duch Boží.* (Genesis 1, 2 podle ČEK)

(2) *Neobyčejným vzácným darem, kterého se mi dostalo, je to, že patronem naší katedrály není, jak to bývá obvykle, některý světec, ale sám Duch Svatý, Duch Boží. Jsem za to Bohu nesmírně vděčný a prosím tohoto našeho vznešeného Tituláře, aby mě vedl a aby provázel dál také celé milé východní Čechy.* (SYN)

(3) *Duch svatý není neurčitá duchovní energie, projev božské moci, ale Boží osoba. Spolu s Otcem a Synem patří Duch svatý do společenství tří božských osob – Nejsvětější Trojice – trojjediného Boha.* (<www.google.cz>, Cit. 2010-10-06.)

(4) *Posledně jsme hovořili o tom, kdo je Duch svatý a jaké je jeho působení a dílo. Duch sva-*

tý je darem a zároveň i dárcem. Dává duchovní dary. Bůh si přeje, abychom byli plni Ducha svatého, obdarováni jeho dary a aby bylo v našem životě vidět ovoce Ducha svatého. (<www.google.cz>, Cit. 2013-07-30)

(5) *Do devíti hodin vyčistíme bitevní pole z nejhoršího, doplníme zásoby vody a s důvěrou budeme očekávat posilu. K tomu nám dopomáhej Bůh Otec, Bůh Syn a **Duch svatý!*** (SYN)

(6) *Je jí šestatřicet, a bude babičkou! Ježíši, panno Marie, **Duchu svatý** a ty třetí ze svaté trojice, na jehož jméno si nikdy nemůžu vzpomenout, dejte, aby to dítě bylo zdravé a aby život jeho rodičů byl navzdory jejich přehmatu procházkou růžovým sadem.* (SYN)

(7) *Ale kde vzít a nekrást? Už už jsem se smiřoval s tím, že právě k takovým veličinám se já jako obyčejný smrtelník vůbec proboujet nemůžu. Tu mě však náhle **osvítil Duch**, sice ne **svatý**, jen můj vlastní, který mě ovšem ne vždy navštěvuje v pravou chvíli.* (SYN)

(8) *Kovář – jenž ostatně postupně plnil i další vyslovená dezideria – ve své studii ohledával náboženské otázky své doby a formuloval základy živé víry „nové církve“, jíž Karel Farský v diskusi se synodním kurátorem českobratrské církve evangelické označil za „liturgickou,, a T. G. Masaryk nazval „církví husitskou“ (Světová revoluce, hlava 137, Praha 1925). Tato církev záhy zaujala systematické stanovisko „vůči všem složkám náboženské otázky“. Karel Farský osobně vytyčil základní propria církve, platné dodnes: je to **Duch Kristův** jako nejvyšší norma učení a Zpřítomnění jako pramen a norma liturgické zbožnosti.* (SYN)

(9) *A vnuknuto mi bylo jen, že varovatelem zjevným jsem, a neměl jsem vědomost žádnou o shromáždění nejvyšším, když ve sporu byli v něm, když Pán tvůj pravil andělům: „Já smrtelníka z hlíny vytvořím, a až jej vyrovnám a vdechnu mu něco z **ducha Svého**, padněte na zem klaníce se před ním!“* (SYN – z českého překladu koránu)

(10) *Lidé jako vy nikdy nerozevrou rty, aby se ohnivé vody napili. **Velký duch** vás obdaroval moudrostí. „Co mám tedy dělat pro nápravu?“ „Slyšte mě,“ opakoval Indián a zase se zatvářil vážně.“* (SYN)

Narazíme v těchto dokladech na metajazykové komentáře typické mj. pro křesťanský misijní diskurz a komunikaci mezi různými konfesemi a církvemi (doklady 3 a 4).⁹ Prvky jazykové sekularizace nacházíme v dokladech 6 a 7.

Lexém *duch* má velmi komplexní sémantiku, která se odráží v bohaté a složité etymologické, slovtvorné, morfosyntaktické a konceptuální struktuře. Slovo *duch* patří k soustavě ústředních kulturních pojmů, které se uskutečňují pomocí formativu domácího (slovanského) původu a dlouhodobě hraje roli v různých diskurzech (viz výskyt v pojmových opozicích: *duch a tělo*, *duch a litera* a triádě novozákonního *Boha: otec, syn, duch*). Jeho integrální součástí je nábožensko-profánní polysémie. Polysémii však necharakterizuje jenom aspekt sakrality. Zajímavé je třídění významů podle toho, jestli připouští gramatický plurál a plurálovou sémantiku. *Duchy* není plurál profilu, který jsme charakterizovali jako „původní, věčnou a životodárnou sílu“. Je pravděpodobné, že uživatel jazyka je vnímá jako různé lexikální formanty a vyskytují se v jiných typických spojeních (např. *řadí tam duchové*). Také není plurální formou psychologického významu, který jsme zachycovali ve čtvrtém profilu hesla *duch*. Tento profil však obsahuje metonymicky sémantizovaný význam *duch* ve smyslu „vynikající osobnost“, ke kterému je možné tvořit plurál *duchové*, viz kolo-

9 Pastorační, teologické, laické výklady významu, lidové etymologie atd.– k těmto aspektům viz referát Michaely Lašťovičkové a Marcely Grygerkové, který vyjde v NAGÓRKO (2014). Ve svém referátu na druhé berlínské konferenci 2011 zdůrazňoval Jozef Pavlovič, že se jazyková sekularizace odehrává i vně církevních, reformačních interkonfesijních diskurzech – srov. HEINKE (2012).

kaci *největší duchové lidstva*. Také ve třetím profilu o „netělesných bytostech“ je viditelná tendence k rozdělení významu v souvislosti s plurálem, jak ukazuje tabulka č 1.

Tab. 1 Přehled významů slova *duch*

<i>PROFIL</i>	<i>Konkrétní významy</i>	<i>Struktura</i>
1. trinitární	<ul style="list-style-type: none"> • <i>původní síla</i> • <i>osoba svaté trojice</i> • <i>svatodušní svátky</i> • <i>motiv pojmenování</i> 	<i>ve frazémeh a elipticky; singulare tantum; mask. anim.</i>
2. nad-přirozený	<ul style="list-style-type: none"> • <i>nelidské bytosti (s biblickým původem)</i> 	<i>často v plurálu</i>
3. magický	<ul style="list-style-type: none"> • <i>netělesné bytosti</i> • <i>negativní, neosobní síly</i> 	<ul style="list-style-type: none"> • <i>plurál netypický</i> • <i>spojeno s plurálem</i>
4. psycho-logický	<ul style="list-style-type: none"> • <i>netělesná sféra člověka</i> • <i>výrazná osoba/ osobnost</i> 	<ul style="list-style-type: none"> • <i>sg. tant.; mask. anim.</i> • <i>možný plurál</i>
5. relevance	<ul style="list-style-type: none"> • <i>charakteristické vlastnosti hnutí, epochy atd.</i> • <i>smysl textu, zákona apod.</i> 	<i>sg. tant.; <duch + [subst. v. gen.]> duch doby, duch místa, duch křesťanství, duch zákonu</i>
6. sociální	<ul style="list-style-type: none"> • <i>atmosféra, charakter sociologického celku „v radostném duchu“</i> 	<i>sg. tant.; mask. anim. různé kolokační vzory olimpijský duch, duch solidarity</i>
7. filozofický	<ul style="list-style-type: none"> • <i>filozofický terminus, opositum hmota; „umění jako projev ducha“</i> 	<i>sg. tant.; mask. anim., někdy však mask. inanim. filozofie ducha, duch křesťanství</i>

Sémantika *Archy Noemovy* a srovnávání

V případě *Noemovy archy* byl vybrán jako vstup do polysémie přímo víceslovný lexém. Tímto rozhodnutím zůstaly stranou pozornosti další lexémy, resp. formativy vztahující se ke slovu *archa*. Česká výkladová a kodifikační lexikografie zachycuje pod heslovým záhlavím *archa* celkově následující čtyři významy (v různém uspořádání): 1. *Archa úmluvy*, 2. *Archa* terminologicky „křídlový oltář“, 3. *Archa Noemova / Noemova archa*, 3a) *Noemův koráb*, 3b) „staré nemotorové vozidlo“. Všechny základní semy mají náboženský, resp. církevní význam, vybočuje přenesené označení pro dopravní prostředky (3b) přinášející různé pragmaticko-sémantické aspekty.¹⁰

Oproti těmto lexikografickým výkladům se projevuje odlišné pojetí lexikálně-pragmaticko-sémantického výkladu v rámci SuP. Berlínský slovník přistupuje k sémantickému vějíři lexému *Noemova archa* skrz tři profily: profil „náboženský“, profil „záchrany“ (Rettung) a profil „nashromáždění“ (Ansammlung). První profil odpovídá sémantické intenci 3a), z něho jsou odvozeny přenesené významy obsažené v dalších profilech. Profil záchrany reprezentuje „velké vozidlo nebo místo, zařízení, instituci, která přijímá, chrání a zachovává pro budoucnost ohrožené živočichy a kulturní objekty“. Podle typického užití v sémantickém rámci tohoto profilu se jedná o „projekty

10 Těmto aspektům bude věnován příspěvek v NAGÓRKO 2014.

k záchově biologické a druhové rozmanitosti (biodiverzita), například genetické banky a zoologické zahrady¹¹. O status třetího profilu „nashromáždění“ pro češtinu ještě není redakčně definitivně rozhodnuto, indikace, která vedla k založení tohoto profilu, pramení ve výzkumu slovenského materiálu. Tam se námi popsané užití vyskytuje okazionálně. Platnost profilu se potvrdila i na polských a německých datech, pro češtinu se objevil zatím jenom jeden zdroj z překladové beletrie¹², pro znázornění uvádím příslušné konkordance:

(11a) *A teď jsem se dokonce rozběhla ke dveřím a prudce je otevřela pro případ, že by za nimi čekala celá archa traumatizovaných mazlíků, seřazených na chodbě v třetím patře a naprosto nechápajících, proč jim nedovolím vlézt dovnitř a rozvalit se v mém pokoji jako předoucí a chrápající klubíčka. Takže jsem rozrazila dveře a tam – trpělivě – čekal nikdo (SYN); (11b) Než skončil včerejší večer, kuchyňský stůl připomínal Noemovu archu plnou zvířátek vyrobených z roliček od toaletáku na Torův sobotní prodejní stánek. (SYN)*

Aplikace metafory na tomto pozadí má charakter přenosu *ad hoc*, předpokladem ale je sémantický, nikoliv slovtvorný potenciál takového užití. Ještě se tomuto problému věnuji v souvislosti s třetím profilem *Noemovy archy*.¹³

Proto, že se v náboženských významech slova *duch* (singulare tantum) a *archa* a významech od nich odvozených většinou nevyskytují samotné, ale vázaně v slovních spojeních jako *Duch svatý*, *Velký duch*, *Duch Páně*, *Ježíšův Duch* apod.; *Noemova archa*, *archa Noemova*, *Noemův koráb*; *archa úmluvy*, *archa úmluvy*, lze hovořit o reprezentantech potencionálního užití. V konkrétních realizacích se aktualizují buďto v dvoučlenných jednotkách, anebo elipticky (textová nebo kontextová elipsa). V případě *archy* byla vybrána část polysémie vázaná s frazémem *Noemova* a popsaná ve třech profilech, syntaktické a lexikální varianty slovního spojení (*Archa Noemova*, *Noemova archa*, *Noemův koráb*) se liší jenom z pragmatického hlediska (časovost, stylistika). Varianty *ducha* (*Duch svatý* vs. *Velký duch*) jsou zachyceny v jednom profilu a mají také svou sémantickou relevanci, přesto jsou si tak blízké, že se dají explikovat jedním profilovým výkladem. Nejdál od sebe to mají frazémy, resp. slovní spojení *Noemova archa* a *archa úmluvy*, i zde se však najde společný jmenovatel v sémantice skládání a zachránění něčeho cenného v něčem pevném ve tvaru truhle.

Výsledky rešerší v korpusu: sémantika záchrany

Cesta indukce a autointrospekce vede k zachycení významů v dosavadní lexikografii nepopsaných. Rešerše v ČNK přinášely zajímavý a někdy i překvapivý materiál užití některých lexémů. Profil „záchrany“ v heslové stati o *Noemově arše* vznikl podobně jako výše komentovaný profil „nashromáždění“ na základě empirického bádání. Pro češtinu tvořilo základ analýzy cca 180 vý-

11 Německé znění v SuP: „Ein großes Gefährt oder ein Ort, der gefährdete Lebewesen oder Kulturobjekte aufnimmt, schützt und für die Zukunft bewahrt. Übertragen auch eine Institution, die dasselbe leistet. Typischerweise werden so Projekte zur Erhaltung der Artenvielfalt bezeichnet, z.B. Genbanken oder zoologische Gärten.“

12 McCOMBIEOVÁ, K. (2006), původně Karen McCombie 2004. Bibliografické údaje pramenů uvádím v seznamu literatury podle údajů v ČNK.

13 Zajímavé by bylo srovnávání slovenských a českých kolokací typů *hotová/úplná archa* a *hotový/úplný babylon* vs. *hotová apokalypsa*, *hotové peklo*, *hotový očištec*, srov. SuP, lemmata *archa*, HORNIK, J. – ORLOŠ, T. Z. (1996), lemma *archa*.

skytů v korpusu syn2006pub (zde SYN 2006). Konkordance byly tříděné podle domén užití, gramatických a formálních zvláštností a syntaktických a kolokačních vzorů užití; byly vyexcerpované typické konotace, atributy a sémantické aspekty. Na této půdě vznikal rozvrh sémantiky českého lemmatu s dvěma profily, který byl konfrontován s dalšími jazyky. Ve srovnávání profilů heslového slova *duch* jsme viděli, že profily mohou zastřešit různě organizované a široké soubory sémantických jednotek a vlastností. Některé profily můžou být terminologické, okazionální a stavěné na stupni sakrálnosti, doméně užití, typu diskurzu apod. V případě *Noemovy archy* získané profily zahrnují celé metaforické pole pod střešou jednoho sémantického aspektu, sémantiky „záchran-y“, která převádí elementy biblického příběhu o Noemovi, jeho rodině a lodi do úplně jiných souvislostí. Tento souhrn je výsledkem hlavního analytického kroku, profilace. Uvádím několik příkladů pro tento profil:

(12) *Moderní zoologické zahrady už nejsou ponurá vězení, „monumenty lidské krutosti“, ale spíš záchranné archy Noemovy. Zvířata se čím dál méně chovají v nevábných klecích za pevnými mřížemi. Pražská zoo má například v plánu, že se na celém svahu pod bohnickým sídlištěm budou v dohledné době prohánět nosorožci.* (SYN)

(13) *Nervy obyvatel evakuačního střediska v uherskohradištské Základní škole Východ jsou po uplynulých dnech napjaté k prasknutí a včera dostaly opět pořádně zabrat. Žijí tu již devátý den a nyní se dozvěděli, že Noemovu archu, jak tu začali říkat nejvýše položené budově města, kvůli stoupající vodě ještě nemohou opustit. Na Mařatický kopec se dokonce začali vracet někteří z těch, kdo jej již jednou opustili.* (SYN)

(14) *„Směřujeme ke stále užšímu vymezení podstaty naší poezie a jejího postavení ve světě,“ komentuje cíle Bítova Jiří Kuběna, druhý duchovní otec ojedinelého setkání, jenž ovšem zdůrazňuje, že celá akce není archou Noemovou na záchranu veršotepců a jejich produkce, ale představuje spíše pomyslnou samoobsluhu.“* (SYN)

(15) *Celkem dvě stě psů u ní našlo domov. „Zřejmě to pro mě byla ideální terapie, protože stále žiju,“ říká Silke. Ovšem idyla končí. „Krétská archa Noemova“, jak se útulku přezdívá, skončí z prostého důvodu. Pronajímatel odmítá Silke obnovit smlouvu, chce pozemek prodat.* (SYN)

Kontexty užití, témata a příslušné domény jsou různorodé: zoologická zahrada (biologie, zvířata, volný čas a turismus, městská zařízení), přírodní katastrofa (události, bezpečnost, ekologie), literární událost (současná poezie, kultura a veřejnost, literární život) a psí útulek (záchrana přírody a zvířat, etika, civilizace). Jaksi specifitěji než tato profilace může působit konkrétně uchopitelný význam, ke kterému nejdříve uvádím doklady:

(16) *Kdyby se potom ohřálo, mohlo by se ho kdykoliv použít pro oplodnění žen – ať v blízké budoucnosti nebo až za několik generací. Natálie napsala: Kdyby byl Noe znal zmrazená spermata, mohla ta archa být malinká jak chata. Stačí prý teď otevřít dvířka od lednic a děti můžeš mít bezpočet, bez hranic. Zato ti dva kluci z té naší chásky, jsou dílem, zaplaťbůh, vášně a lásky.* (SYN)

(17) *Tak začíná celovečerní animovaná sci-fi Titan A. E. – což znamená After Earth, po zničení Země, a jméno Titan má bájná kosmická loď, jakási Noemova archa, jež nese místo tvo-*

rů z kumavky s jejich DANN a zůstává poslední nadějí lidstva. Klíč k této spáse dostal osiřelý klučina, kterého vrací do hry podtitul O patnáct let později a údajný otcův přítel s výzvou „lidstvo tě potřebuje“. (SYN)

(18) *Po celém světě se nebezpečně scvrkává druhová rozmanitost zemědělských plodin. Ve snaze zabránit ztrátě tohoto pokladu navždy založily nespočetné země genetické banky – jakési moderní **Noemovy archy**, v nichž se skladují cenné zdroje výživy lidstva jako například rajčata, pšenice nebo čirok. (SYN)*

(19) *Stavbu financoval miliardář Edward P. Bass, muž, jehož snem bylo učinit první krok na cestě k budoucí kolonizaci Marsu. Na svou **Noemovu archu**, zcela izolovanou od okolního světa, nalodil 3800 různých druhů rostlin a živočichů včetně člověka. Dvouleté „plavby“ se zúčastnilo osm dobrovolníků z řad vědecké komunity (čtyři muži a čtyři ženy). Potraviny jim měla poskytnout především úrodná pole Biosféry. (SYN)*

(20) *Veškeré abstráky se staly věcí minulosti, nejasnou měrou ze středověku. Budoucnost se změnila v blaho. Dětičky narkomanů se nocí početí stávaly jinými bytostmi, lidskými **archami** měněné genetické informace. Od doby, kdy Vladimír z bytu vypoklonkoval poselstvo ze Sultanátů, neznala proslulost Trojice z Východu mezí. (SYN)*

(21) *Nadějí ohrožených živočichů jsou zmrazené geny. Chování se zmrazit nedá, tvrdí odborníci. Je to příliš nákladné, dodávají jiní. **Noemova archa** pro 21. století obsahuje zmrazený genetický materiál celých stád nosorožců a slonů, tygrů goril i zeber. Proslulý americký biolog Edward Wilson odhaduje, že každý den zmizí ze zemského povrchu 137 druhů rostlin a zvířat. (SYN)*

(22) *V době, kdy nejen devastace přírody, ale i lidská chudoba a chamtivost přispívá k vymírání jednotlivých druhů, plní zoologické zahrady funkci genetických bank a jsou novodobými **archami Noemovými**. Stále více přibývá úspěšných reintrodukcí (zpětné vysazení), do kterých se již zapojily naše zoo. Bohužel většinou není kam, téměř vše na zemi již zdevastoval nebo obsadil člověk. (SYN)*

(23) *Britští vědci ze Zoologické společnosti v Londýně, Muzea přírodních dějin a z univerzity v Nottinghamu chtějí přispět k záchraně ohrožených zvířecích druhů tím, že zmrazí jejich tkáň a genetický kód. Svému projektu říkají „Zmrazená **archa**“. Jejich sbírky mají v případě potřeby umožnit i oživení zaniklých druhů. (SYN)*

Znamená archa „genetická banka“?

V těchto konkordancích se používá slovní spojení *Noemova archa*, někdy v inverzní a eliptické podobě ve významu *genetická banka*.

Nespecifický formativ *banka* jako součást analytického anebo kompozičního lexému¹⁴ je v čes-

14 V základních vzorech 1) <[přídavné jméno vztahové] + banka>; 2) <banka + [nominální lexém jedno- nebo vícečlenný v genitivu]>, 3) <[~+o/a+banka]> a vzorech kombinujících; na rozdíl od dvoučlenných vlastních jmen typu *Komerční banka* (typické pro užití ve významu „peněžní ústav“).

ké lexikografii považován za sémantický neologismus v polysémním vztahu k prvnímu významu „peněžní ústav“, popřípadě i k druhému významu *banka dat*, jak ukáží výkladové definice: „Nově ve vyzn. (vědecké, odborné) zařízení k uchování velkého množství něčeho velmi cenného: krevní banka; b. kostní dřeně; lexikální banka Ústavu pro jazyk český; poč. tech. b. dat počítačový soubor konkrétních dat (báze dat) spolu s programovým vybavením pro jejich zpracování; databanka“ (POŠTOLKOVÁ, B. – SOCHOVÁ, Z. 1994, lemma *banka*); „Zařízení k shromažďování a úschově něčeho hodnotného nebo užitečného pro opět (budoucí) využití: banka hudebních nahrávek, audiovizuální banka; národní banka vín (původně v medicíně, v genetice ap.) > kilometrská banka > spermiová banka > záruční banka“ (SN2). SSČ1 uvádí jako druhou část polysémie: „banka dat, sbírka prvotních informací, databanka“, SSČ2 „nemocniční oddělení shromažďující a uskladňující odebraní tkáň, tekutiny, orgány ap. tkáňová/ krevní/ kostní/ genová banka“.

Daná polysémie by se dala jednoduše pochopit třemi profily: Prvním profilem by byl „peněžní ústav“, který měl starou formu *bank* – tato forma dnes reprezentuje speciální případ jackpotu, „peněz vložených do hry“ s frazeologickou kolokací *rozbít bank*, jakási zmenšenina peněžního ústavu, srov. herní banky v hrách typu *Monopoly*. I *kilometrská banka* Českých drah by patřila k „malým“ bankám, pokud uschovává bodový kredit reprezentující cenu za dopravu na jeden železniční kilometr. Druhý profil by byla *banka dat*, včetně sbírek hudebních nahrávek. Důležité je zde, že se jedná o uchování digitálních dat v počítačových systémech. Variantou je formativ *~báze* v synonymu *databáze*. Třetí by reprezentoval genetické banky, které slouží biologickým vědám, medicíně, ekologii a potravinářství. Všechny tři jsou jednoduše přiřaditelné k příslušným sférám užití: finančně-ekonomická sféra, kulturní sféra a biotechnologická sféra včetně oblastí jejího praktického užití, sem patří i slovní spojení jako *krevní banka*, *kostní banka*, *genová banka*, *spermiová banka*; *kryobanka*, *ledobanka* apod., a podle všeho i *banka vín*.

Sémantická analýza diskurzu

Poslední dva zde nastíněné profily lexému *banka* v současné češtině s příslušnými sférami pokrývají velkou část sémantické extenze heslového profilu „záchraný“ lexému *Noemovy archy*, čímž mají vztah se značným stupněm synonymie. Reprezentují oblast užití jazyka, která je poznamenaná podobným diskurzem, včetně nátlaku překládat terminologii, koncepty, metafory a výrazy především z anglofonního světa. Slangové prostředí těch (málo) lidí, kteří v praxi mají častý kontakt s novými reprodukčními technologiemi a potřebu se o nich dorozumívat mluveným jazykem, by si rozhodně zasloužily badatelskou pozornost. Zde je možné konstatovat, že bohatá metaforika *Noemovy archy*, na rozdíl od neutrálního a spíše bezpříznakového, cizího slova terminologického rázu *banka*, méně přirozeně splývá s jazykovým prostředím. Užití v konkordancích většinou explicitně signalizuje, že se jedná o hotovou metafору, která nese s sebou biblický kolorit. Uvozovky jakožto „signál okazionality“¹⁵ jsou jenom jedním z mnoha jazykových a komunikačních prostředků charakterizujících „hranatosť“ lexému *archa* v češtině – ovšem při značném výskytu konceptu *archy* v některých nadnárodních, ale také českých diskurzech a doménách jazykového užití. Test synonymem však také ukáže stabilizující se sémantickou oblast, kde plní diferencované užití obou slov různé jazykové a kontextuální funkce, přičemž souvislost je intuitivně jednoduše pochopitelná.

Sdílené významy vedou k velice širokým sémantickým rysům, *archa* znamenala v latině závěrací skříň.¹⁶ Jde o sklad, respektive o místo a přeneseně o zařízení anebo instituci k uložení, a tím

15 VACHKOVÁ 2007: 79.

16 Viz poznámka 10.

i zachránění cenných a ohrožených bytostí a věcí. Aspekt pohybu (dopravní prostředek) přibyl dynamikou toho ohrožení v narativních souvislostech konkrétního příběhu a vrcholí v označení kosmických lodí se záchrannou misí v postapokalyptických scénářích a futurologických projektech jako *Noemovy archy*. Jazyková platnost se zvyšuje tím, že pomocí formativu překryje další domény, obě slova je dobývají paralelně, ale v různé míře a různým způsobem.

Pokoušel jsem se získat údaje o shodě těch dvou synonym v SYN pomocí KWIC [lemma: *archa*, 10.962 výskytů] a pozitivním filtrem [lemma: *genetický*]¹⁷. Osm z jedenácti získaných konkordancí (necelých 0,1 procent) jsou součástí zpravodaje o projektu Zmrazená archa Zoologické společnosti z roku 2004, uváděl jsem výše jenom jeden z nich. Alespoň podle doposud získaných dat otevřel diskurz, který si vyžadoval nové pojmenovací prostředky český překlad knihy Arthura Haileya z roku 1973 pod názvem *Letiště*, v domácí produkci navazoval 1995 Jáchym Topol s *Andělem*, kde vyzní užití v prostředí fantastické prózy možná nejpřirozeněji (konkordance 20). K další diskurzivní analýze tohoto slibného (sémantického) pole by velice pomohly nástroje, které by dokázaly efektivněji nacházet v korpusech shody kritérií, jakými jsou diskurz, žánr, tematika, doména, paradigmatika a odborná oblast. Kombinace dvou a více kritérií s jistou výpovědní silou by mohla nahradit ve srovnání s lexémem *Duch* chybějící gramatické prostředky, které by pomohly určovat i sémantické vlastnosti a příslušnosti automatickým sestavením hrubého výběru relevantního materiálu. Tento problém je zesilován možností elipsy *archa*, která výrazně zvyšuje počet konkordancí, pokud je zadána jako KWIC. Exkurzem do domény reprodukční medicíny a dalších jsem chtěl také předvést úskalí prací se sémantickým materiálem mezi vágností a šíří konceptuálního přístupu na jednom pólu a úzce vymezeným, často okazionálním lexikografickým statusem na pólu druhém. Gramatické vlastnosti, jako například status *pluralia tantum* napomáhají kognitivní sémantice při profilaci lexikálního materiálu, přidávají konkrétní tvar, kde je sémantika spleťtá a těkavá. Podobně může analýze napomáhat test synonymem a navazující korpusem řešerše.

Jazyková sekularizace na případu dvoučlenných slovních spojení

Srovnávání obou lexémů, jednoho heslového a jednoho profilového slova, přineslo různé případy vztahu sémantiky s formálními, gramatickými a diskurzivními vlastnostmi. Dost různý je také dopad procesu jazykové sekularizace. V celém lexému *duch* funguje dávno zaběhlá sakrálně-sekulární polysémie, novější vlny sekularizace zasáhly daný lexém jenom ve velice jemné míře.

V případě *Ducha svatého* jsou to především modifikace v rámci církevního užití: částečné psaní malým písmenem ve frazému *osvítí/osvítíl duch svatý*, stabilizaci významové diferenciaci ve slovosledu (*svatý Duch* se používá už pouze ve významu svátku a vyskytuje se ve valné většině frazematicky vázaně, drží se však majuskula), kontexty vyjadřující ironii anebo distanci mluvčího. U některých dokladů bylo ze zde uváděných úzkých kontextů nepříliš jasné, jaký (ne-) konfesijní postoj mluvčího byl vyjádřen.

Noemova archa zato byla naplno zasažená vlnou přenášení sakrálních lexémů do nesakrálních sfér, jakou je například reprodukční medicína. Zde se sekularizace odehrává v sémantickém jádru slova a definuje druhý profil, který donedávna vůbec neexistoval. Otázkou zůstává, jak se celý profil vyvíjí, pokud nevzniknou stabilizované konkrétnější významy, které by se dočkaly podobného lexikografického uznání jako dva novosémantismy v heslovém slově *banka*. Důležitým tématem zůstává tlak k překladům na úrovni metodologické (potřeba ekvivalentů) a diskurzivní (berlínský projekt z kapacitních důvodů nemůže soustavně sledovat a zkoncipovat vliv anglického jazyka

17 Při toleranci 5/10/15 pozic na obě strany vyšlo 7/11/11 výskytů.

na vymezenou oblast), jak jsem zde uváděl na dvou příkladech. Také fakt, že se jedná o dva jediné dvoučlenné lexémy na místě lexikografického průřezu danou slovní zásobou, s sebou nese některé metodické implikace, ovšem vzhledem k sémantickým procesům máme před sebou díky dalším vlastnostem dva skoro diametrálně rozlišné případy. Srovnávání jsem zde nemohl vyčerpat, zbývá třeba možnost textologického srovnávání chování těchto dvou lexémů/frazémů s jinými syntaktickými typy frází. Podobně by se daly brát další lexémy z berlínského seznamu, jejich profily, deriváty, frazémy, synonyma atd. jako výchozí body budoucích analýz. Před každou analýzou by se měla zvážit její formální, strukturální, sémantopragmatická a sociopragmatická charakteristika, protože ovlivňuje možnost korpusových rešerší a postavení v lexikonu a diskurzech. Vzhledem k makrostruktuře dané složky slovní zásoby se dopracujeme takovou cestou k odhalování všeobecných rysů jazykové sekularizace. Doufám, že podobně intenzivní výzkum sice okrajového, ale neméně zajímavého materiálu, přináší podněty k dalším analýzám a vede k začlenění podobných úvah a sond jak z oblastí sémantiky, tak i analýzy diskurzu (kterou jsem zde mohl jenom nastínit) do lexikografie, obzvláště v době, kdy vznikají zásadní velké slovníky a rychle se rozvíjejí korpusy a technologie jim blízké.

Použitá literatura

- Akademický slovník cizích slov A–Ž* (1995). Praha: Academia. (ASCS)
- BELICA, C. – STEYER, K. (2008): Korpusanalytische Zugänge zu sprachlichem Usus (Korpus a zkoumání německého jazykového úzu). In: M. Vachková (ed.): *Beiträge zur bilingualen Lexikographie*. Praha: Univerzita Karlova v Praze, Filozofická fakulta, s. 7–26.
- Český ekumenický překlad (2008). In: BÍČ, M. – MÁNEK, J. – SOUČEK, J. B. (2008): *Bible. Český ekumenický překlad – Bible kralická: česká synoptická Bible (v rozsahu celého vydání Bible kralické z roku 1613)*. Praha: Česká biblická společnost. (ČEK)
- Český národní korpus (2000-2013 ff.) [online]. Praha: Ústav Českého národního korpusu FF UK. Cit. 2013-12-13. <<http://www.korpus.cz>>. (ČNK)
- GOOGLE.CZ (bez roku) [online]. Cit. 2013-12-13. <www.google.cz>.
- GRYGERKOVÁ, M. – LAŠŤOVIČKOVÁ, M. (2013): K žánru modlitby (na příkladech z katolického Českého misálu). In: *Stylistika v kontextu historie a současnosti, Docentce Evě Minářové k životnímu jubileu*. Brno: MU, s. 155–162.
- GRYGERKOVÁ, M. (2013): *Teolingvistika, uvedení do problematiky (Studijní materiály pro distanční kurz: Teolingvistika)*. Ostrava: Ostravská univerzita v Ostravě (Scriptum).
- HAILEY, A. (1973): *Letiště*. Praha: Riosport-press.
- HEINKE, J. (2012): *Tagungsbericht zur internationalen Tagung „Wortsemantik im Spannungsfeld zwischen Säkularisierung und (Re)Sakralisierung der öffentlichen Diskurse – Deutsch, Polnisch, Slowakisch, Tschechisch“* (Berlin, 4.–7. Oktober 2011). *Zeitschrift für Slawistik* 57 (2012), 3, s. 353–356.
- HORNIK, J. – ORŁOŚ, T. Z. (1996): *Czesko-polski słownik skrzydlatych słów*. Kraków: Universitas.
- LEXIKO (2005–2013) [online]: *Lexiko – webové hnízdo o novodobé české slovní zásobě a výkladových slovnících*. Praha: Oddělení současné lexikologie a lexikografie ÚJČ AV ČR, v. v. i. Cit. 2013-12-13. <<http://lexiko.ujc.cas.cz/index.php?page=3>>.
- McCOMBIEOVÁ, K. (2006): *Dóóóst dobrej hepáč*. Havlíčkův Brod: Fragment.
- NAGÓRKO, A. (2007): *Lexikologie des Polnischen*. Hildesheim et al.: Georg Olms Verlag.
- NAGÓRKO, A. (2011a): Weltanschauung vs. das sprachliche Weltbild, säkularisierte Lexik im westslawisch-deutschen Sprachvergleich. In: Kuße, H. – Woldt, C. (eds.): *Tschechisch und Slowakisch, Nähe und Distanz. München: Verlag Otto Sagner (Specimina Philologiae Slavicae 163)*, s.

121–133.

NAGÓRKO, A. (2011b) (ed.): *Sprachliche Säkularisierung* (Westslawisch-Deutsch). Hildesheim et al.: Georg Olms Verlag.

NAGÓRKO, A. (2012) (ed.): *Wortsemantik zwischen Säkularisierung und (Re)Sakralisierung öffentlicher Diskurse*. Hildesheim et al.: Georg Olms Verlag.

NAGÓRKO, A. (2014) (ed.): *Semantik und Pragmatik der Säkularisierung*. Hildesheim et al.: Georg Olms Verlag.

Nový akademický slovník cizích slov A–Ž (2005). Praha: Academia. (NASCS)

POŠTOLKOVÁ, B. – SOCHOVÁ, Z. (1994): *Co v slovnících nenajdete, Novinky v současné slovní zásobě*. Praha: Portál.

Příruční slovník jazyka českého (1935–1957). Praha: Česká akademie věd a umění, Státní nakladatelství (učebnic)/ Československá akademie věd: Státní pedagogické nakladatelství. (PSJČ)

SAICOVÁ ŘÍMALOVÁ, L. (2010): *Vybraná slovesa pohybu v češtině, Studie z kognitivní lingvistiky*. Praha: Univerzita Karlova (Acta Universitatis Carolinae, Philologica, Monographia 161).

SÄKULARISIERUNG RELIGIÖSER LEXIK (2009–2014) [online]: *Säkularisierung religiöser Lexik vor dem Hintergrund gesellschaftlicher Veränderungen (Polnisch – Slowakisch – Tschechisch – Deutsch)*. Cit. 2013-12-13.

<<http://www.slawistik.hu-berlin.de/member/anagorko/saekularisierung>>.

MARTINCOVÁ, O. a kol. (1998): *Nová slova v češtině. Slovník neologizmů 1*. Praha: Academia.

MARTINCOVÁ, O. a kol. (2004): *Nová slova v češtině. Slovník neologizmů 2*. Praha: Academia.

Slovník spisovné češtiny pro školu a veřejnost (2007). S Dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky, studentské vydání. Praha: Academia. (SSČ1)

Slovník současné češtiny (2011). Brno: Lingea. (SSČ2)

Slovník spisovného jazyka českého (1960–1971; 1989). Praha: Academia. (SSJČ)

SuP, Sacrum und Profanum, Religiöse Lexik in der Allgemeinsprache (Deutsch – Polnisch – Slowakisch – Tschechisch) (2009–2014). Berlin: Humboldt-Universität zu Berlin.

SYN, Český národní korpus – SYN (2010) [online]. Praha: Ústav Českého národního korpusu FF UK. Cit. 2013-12-09. <<http://www.korpus.cz>>.

SYN, Český národní korpus – SYN2006PUB (2006) [online]. Praha: Ústav Českého národního korpusu FF UK. Cit. 2013-12-09. <<http://www.korpus.cz>>.

TOPOL, Jáchym (2006): *Anděl*. Praha: Torst.

VACHKOVÁ, M. (2007): *Kapitoly k německo-české metalexikografii II*. Praha: Univerzita Karlova v Praze, Filozofická fakulta.

VACHKOVÁ, M. (2008) (ed.): *Beiträge zur bilingualen Lexikographie*. Praha: Univerzita Karlova v Praze, Filozofická fakulta.



THE END

**...OR TO BE
CONTINUED?**



Sledujte nás na Facebooku

Brněňští lingvisté