

[nová] Čeština
doma
& ve světě

1 2015

Nová čeština doma a ve světě

Filozofická fakulta Univerzity Karlovy v Praze

1/2015

Redakce:

Vedoucí redaktorka – Mgr. Kateřina Romaševská, Ph.D.

Výkonná redaktorka – Mgr. Katarzyna Vaculová

Členové redakce – Mgr. Adriana Filas, Mgr. Oleksandr Sukhanek,

Mgr. Maria Simeunovich-Skvortsova

Korektorky – Mgr. Olga Fojtíková, Bc. Jana Georgievová, Bc. BcA. Markéta Kraevská,

Mgr. Evgenia Ulyankina

Adresa redakce:

Katedra jihoslovanských a balkanistických studií FF UK

nám. J. Palacha 2, 116 38 Praha 1

Tel. (+420) 221 619 347, e-mail: ncds@ff.cuni.cz

Objednávky vyřizuje:

Vydavatelství Filozofické fakulty UK v Praze,

nám. Jana Palacha 2, 116 38 Praha 1, e-mail: books@ff.cuni.cz

Návrh grafické úpravy a obálky: Studio Lacerta (www.sazba.cz)

Sazba: Lukáš Vavrečka, nakladatelství Pavel Mervart (www.pavelmervart.cz)

Obsah

Úvodní slovo	9
Recenze	11
Dana Hánková Valková, J. (2014): <i>Komunikační přístup a učebnice češtiny pro nerodilé mluvčí</i> . Praha: Filozofická fakulta Univerzity Karlovy.	13
Marie Kopřivová Čermák, F. (2014): <i>Periferie jazyka. Slovník monokolokabilních slov</i> . Praha: Nakladatelství Lidové noviny.	18
Kateřina Sachrová Mészáros, A. – Kiss Szemán, R. (2014): <i>Na úprku před sebou samými. Češi a jejich literatura očima současné maďarské bohemistiky</i> . Praha: Akropolis.	22
Články	28
Aksana Schillova <i>Homonymní příslovce a předložky a problematika jejich disambiguace v Českém národním korpusu</i>	30
Václav Velčovský <i>Pokusy o zavedení jazykového zákona v České republice</i>	41
Jitka Veroňková <i>Na českém znění spolupracovali... aneb Hodnocení projevu dabérů přírodovědných dokumentů</i>	50
Marie Podobová <i>Bezobjektové posesivní rezultativní konstrukce od sloves tranzitivních a bezobjektové i objektové posesivní rezultativní konstrukce od sloves intransitivních – jejich tvorba a funkce v českém jazyce</i>	82

Zdeněk Salzman	
<i>Z české kuchyně, domácí i cizí: etymologický výlet</i>	115
O autorech tohoto čísla	121

Homonymní příslovce a předložky a problematika jejich disambiguace v Českém národním korpusu¹

AKSANA SCHILLOVA

Abstrakt: Příspěvek je zaměřen na homonymní příslovce a předložky jako např. *blízko*, *dovnitř*, *vstříc* aj. V příspěvku je uveden jejich seznam, je zde analyzována distribuce těchto homonym v korpusu SYN2010 a úspěšnost jejich automatické disambiguace v tomto korpusu a jsou zde systematizovány a analyzovány chyby, které se vyskytují v korpusové disambiguaci daných jednotek. Příspěvek také ukazuje, že problematika automatické disambiguace homonymních příslovcí a předložek vzniká na pozadí rozsáhlé morfologické homonymie v češtině a často se komplikuje homonymií pádových tvarů jmen. Prezentovaný výzkum je základem pro následné vypracování konkrétních lingvistických pravidel, která by mohla úspěšnost disambiguace daných jednotek v Českém národním korpusu zvýšit.

Klíčová slova: příslovce, předložka, slovnědruhá homonymie, automatická disambiguace

Abstract: The article deals with homonymic adverbs and prepositions such as *blízko*, *dovnitř*, *vstříc*, etc. There is a list of them in the article, besides the distribution of these homonyms in the corpus SYN2010 and success rate of their automatic disambiguation in this corpus are analysed – there is also classification and analysis of the mistakes in the disambiguation of these units. It is concluded that problems of the automatic disambiguation of homonymic adverbs and prepositions arise on the background of wide grammatical homonymy in the Czech language and are mainly complicated by homonymy of case forms of nouns. This research is considered to be the basis for the follow-on development of particular linguistic rules, necessary for the improvement of automatic disambiguation of homonymic adverbs and prepositions in the Czech National Corpus.

Key words: adverb, preposition, homonymy of parts of speech, automatic disambiguation

Úvod

Inspirací pro daný výzkum jsou práce věnované tématu homonymie v češtině a s ním související problematice automatické disambiguace homonymních tvarů a jednotek v Českém národním korpusu (např. Petkevič 2005; Jelínek – Petkevič 2011). Tento příspěvek se zaměřuje na jeden z druhů ho-

1 Tento příspěvek vznikl v rámci projektu Rozpracování lingvistických kritérií pro disambiguaci homonymních příslovcí a předložek v Českém národním korpusu řešeného na Filozofické fakultě Univerzity Karlovy v Praze v prostředků specifického vysokoškolského výzkumu na rok 2014.

monymie – slovnědruhovou homonymii příslovčí a předložek – která dosud nebyla předmětem samostatného výzkumu. Hlavním úkolem je tedy analyzovat stav a stanovit problematiku automatické disambiguace daných homonym v Českém národním korpusu, konkrétně na příkladu korpusu SYN2010, a tím vytvořit základ pro následující rozpracování lingvistických pravidel, která by mohla úspěšnost disambiguace daných jednotek v korpusu zvýšit. Vzhledem k rozsáhlosti tématu a korpusem poskytovaného jazykového materiálu byl výzkum soustředěn na homonymní příslovce a předložky s převážně místním významem, tzn. na jednotky jako například, *blízko*, *dovnitř*, *vstříc* aj.

V dosavadní odborné literatuře, kde je těmto jednotkám věnováno většinou jen několik vět až odstavců, jsou daná homonyma vyřešena následujícím způsobem: pokud se jednotka vyskytuje ve spojení se jménem (jméno následuje po předložce), jedná se o předložku; pokud se jednotka vyskytuje samostatně, jedná se o příslovce (Kroupová 1985: 11; Komárek a kol. 1986: 194; Karlík a kol. 2012: 341; Cvrček a kol. 2010: 281; Čermák 2012: 172). Manuální disambiguace homonymních příslovčí a předložek se zřetelem na výše uvedené pravidlo tedy většinou proběhne bez problémů. Přítomnost homonymie si člověk často ani neuvědomí, čehož si již všimli matematictí lingvisté (viz např. Petkevič 2005). Avšak pro automatickou disambiguaci nejsou dané jednotky tak snadným materiálem. Viz následující příklad.

Dost rychle se tak dalo zapomenout na ty probdělé noci v porodnici, kde ráno v šest, sotva co jedna z nás tří matek začala po noci neustálého kojení usínat, uklízečky rozrazily dveře, aby pokoj hlučně prošmejdily hadrem, vzápětí vbíhaly dovnitř sestřičky s teploměry, hned pak vizita a zanedlouho hromada návštěv u každé rodičky (SYN2010). V daném případě je jednotka *dovnitř* specifikována korpusem jako předložka a tvar *sestřičky* jako GenSg, tedy korpusem interpretace této části klauze vypadá jako *uklízečky vbíhaly dovnitř sestřičky*, což je gramaticky správné, avšak neodpovídá významu, který je zřejmý z kontextu: *vbíhaly dovnitř [do pokoje]*.

Navíc se vyskytují případy, ve kterých ani manuálně bez širšího kontextu není možné příslušný výskyt zjednoznačnit. Srov. Podle mého názoru jde o slepičí vejce, pokud je *uvnitř* slepice (SYN2010). Obě možné gramatické interpretace *vejce* je *uvnitř* slepice (Prep – GenSg) a je *uvnitř [vejce] slepice* (Adv – NomSg) jsou sémanticky přípustné. Výběr správné interpretace záleží na dalším kontextu: „Podle mého názoru jde o slepičí vejce, pokud je *uvnitř* slepice. Pokud by vejce snesl klokan, a z něho se vyklubal

pštros, tak by také šlo o pštroší vejce, a ne o klokaní,“ prohlásil Papineau (SYN2010). Tudíž jednotka *uvnitř* je v daném případě příslovcem. Korpus však specifikuje daný výskyt *uvnitř* jako výskyt předložky.

Je evidentní, že komplikace v disambiguaci homonymních příslovcí a předložek v korpusu vznikají na pozadí rozsáhlé homonymie pádových tvarů, což úspěšnost interpretace ještě více ohrožuje. Nejsložitějšími jsou případy, ve kterých správná disambiguace probíhá spíše na sémantické úrovni, jež je pro počítač zatím jen těžko dosažitelná. Příspěvek však bude zaměřen především na ty korpusové chyby, které by bylo možné podle našich předpokladů odstranit pomocí pečlivě vypracovaných pravidel.

Následující kapitoly obsahují seznam homonymních příslovcí a předložek s převážně místním významem, analýzu aktuálního stavu jejich disambiguace v korpusu (na příkladu korpusu SYN2010) a stanovený souhrn problematických míst způsobujících chyby v automatické disambiguaci zkoumaných homonym.

Seznam homonymních příslovcí a předložek s místním významem

Seznam zkoumaných homonym byl zpracován excerpcí z ilustračních soupisů uvedených v odborné literatuře – ve studiích věnovaných analýze předložek, příslovcí resp. slovních druhů obecně a také v příslušných kapitolách českých mluvnic (Kroupová 1985; Komárek a kol. 1986; Karlík a kol. 2012; Cvrček a kol. 2010; Čermák 2012). Tento pracovní seznam v současné době obsahuje 46 jednotek (včetně fonetických variant), které se v korpusu vyskytují a které jsou specifikovány podle slovního druhu (příslovce nebo předložka). V níže uvedeném seznamu jsou tyto jednotky uspořádány sestupně podle frekvence svých výskytů v korpusu SYN2010:

Vedle, daleko, skrze, uprostřed, uvnitř, dovnitř, okolo, blízko, naproti, nedaleko, poblíž, blíž, blíže, vstříc, skrz, napříč, vpředu, zevnitř, navrch, opodál, doprostřed, vně, nablízku, vevnitř, zkraje, zespodu, svrchu, vprostřed, vespod, zevně, zespada, odspodu, naspodu, navrchu, nablízko, zboku, dospodu, zprostředka, zespod, doprostředka, dospod, poblíže, vprostředku, zprostřed, vespodu, naspod.

Distribuce homonymních příslovcí a předložek s místním významem v korpusu SYN2010: korpusové statistiky a jejich manuální ověření

Podle údajů korpusu SYN2010 lze zkoumané homonymní jednotky utřídit do následujících skupin: 1) větší část dokladů je specifikovaná jako doklady předložky (*vedle, uprostřed, okolo, naproti, nedaleko, poblíž, poblíže, napříč,*

skrz, doprostřed, vně, vprostřed, zprostřed); 2) větší část dokladů je specifikovaná jako doklady příslovce (*dovnitř, blíž, zevnitř, vevnitř*); 3) počet dokladů specifikovaných jako doklady předložky a počet dokladů specifikovaných jako doklady příslovce je skoro stejný (*uvnitř, blízko, vstříc, zkraje, zprostředka, doprostředka, vprostředku*); 4) všechny doklady jsou specifikované jako doklady příslovce (*daleko, blíže, vpředu, navrch, opodál, nablízku, zespodu, svrchu, vespod, zevně, zespoda, odspodu, naspodu, navrchu, nablízku, zboku, dospodu, zespod, dospod, vespodu, naspod*); 5) všechny doklady jsou specifikované jako doklady předložky (*skrze*).

Manuální analýza těchto dokladů specifikovaných korpusem jako příslovce resp. předložky však ukázala, že mezi výskyty jednotek označovaných korpusem jako příslovce jsou často zařazeny výskyty, ve kterých jsou tyto jednotky evidentně předložkami a naopak. V některých případech je procento chyby značně vysoké (viz tabulky níže). Chyby byly objeveny během frontální manuální analýzy konkordancí a frekvenčních i kolokačních seznamů zkoumaných jednotek.

Výsledky této analýzy jsou znázorněny v následujících tabulkách. Ve sloupcích je uvedena frekvenční distribuce homonymních příslovcí a předložek v korpusu SYN2010 a počet chybně disambiguovaných výskytů nalezených při manuální analýze a také procento těchto výskytů na celkovém počtu výskytů dané konkrétní jednotky, značkové buď jako příslovce anebo jako předložka. Pro výše uvedené skupiny jednotek jsou sestaveny samostatné tabulky. Tučně jsou zvýrazněna frekventovanější homonyma, která jsou v korpusu disambiguována se značným procentem chyby.

Z výše uvedených tabulek vyplývá zajímavá zákonitost: pokud má homonymní jednotka vysokou frekvenci ve funkci předložky (Skupina 1), je procento chyby v automatické disambiguaci této jednotky jako předložky velmi malé (většinou nepřesahuje 1 %), avšak procento chyby v disambiguaci této jednotky jako příslovce je výrazně větší (převážně kolem 5 % a více), tj. frekvence předložkových výskytů tohoto homonyma je fakticky ještě vyšší, než uvádí korpus. Obdobně (jen opačně) je na tom Skupina 2, ve které mají jednotky vysokou frekvenci jako příslovce: procento chyby v automatické disambiguaci těchto jednotek jako příslovce je nepatrné (nepřesahuje ani 0,5 %), avšak procento chyby v disambiguaci těchto jednotek jako předložky je v této skupině velmi vysoké (srov. *zevnitř* – 30 %; *dovnitř* – 40,2 %; *vevnitř* – až 94 %), což znovu dokazuje, že jedna z funkcí (v tomto případě adverbialní) je pro dané homonymní jednotky výrazně

převládající. Co se týče Skupiny 3, relace mezi procentem chyby a frekvencí předložkových a adverbálních výskytů jednotek z této skupiny je většinou vyrovnaná, stejně jako samotná frekvenční distribuce těchto homonym (zhruba půl na půl), a tudíž je vyrovnaná i jejich předložková a adverbální funkce.

Skupina 1. Větší část dokladů homonymní jednotky je specifikovaná v korpusu SYN2010 jako doklady předložky.

Homonymní jednotka	Předložka (podle automat. disambiguace)			Příslovce (podle automat. disambiguace)		
	Počet výskytů	Z toho výskytů adverbii (podle manuální disambiguace)		Počet výskytů	Z toho výskytů předložek (podle manuální disambiguace)	
		Počet	%		Počet	%
<i>vedle</i>	24007	131	0,5	2298	126	5,5
<i>uprostřed</i>	8490	22	0,25	1963	94	4,8
<i>nedaleko</i>	3572	28	0,8	1356	234	17
<i>naproti</i>	4047	23	0,6	1227	58	4,7
<i>okolo</i>	6440	82	1,3	865	10	1,2
<i>poblíž</i>	3399	17	0,5	723	57	8
<i>skrz</i>	1872	0	0	312	26	8
<i>napříč</i>	1577	9	0,6	260	14	5,4
<i>vně</i>	493	55	11	175	4	2,3
<i>doprostřed</i>	605	0	0	161	10	6
<i>vprostřed</i>	129	3	2,3	20	1	5
<i>poblíže</i>	48	0	0	9	2	22,2
<i>zprostřed</i>	7	0	0	1	0	0

Skupina 2. Větší část dokladů homonymní jednotky je specifikovaná v korpusu SYN2010 jako doklady příslovce.

Homonymní jednotka	Příslovce (podle automat. disambiguace)			Předložka (podle automat. disambiguace)		
	Počet výskytů	Z toho výskytů předložek (podle manuální disambiguace)		Počet výskytů	Z toho výskytů adverbii (podle manuální disambiguace)	
		Počet	%		Počet	%
<i>dovnitř</i>	8089	6	0,07	580	233	40,2
<i>blíž</i>	3528	13	0,4	234	17	7,3
<i>zevnitř</i>	1148	1	0,09	170	51	30
<i>vevnitř</i>	389	0	0	18	17	94

Skupina 3. Počet dokladů specifikovaných jako doklady předložky a počet dokladů specifikovaných jako doklady příslovce je skoro stejný.

Homonymní jednotka	Příslovce (podle automat. disambiguace)			Předložka (podle automat. disambiguace)		
	Počet výskytů	Z toho výskytů předložek (podle manuální disambiguace)		Počet výskytů	Z toho výskytů adverbii (podle manuální disambiguace)	
		Počet	%		Počet	%
<i>uvnitř</i>	5115	76	1,5	5318	293	5,5
<i>blízko</i>	3596	99	2,8	2595	30	1,2
<i>vstříc</i>	1270	23	1,8	992	0	0
<i>zkraje</i>	145	6	4,1	194	10	5,2
<i>zprostředka</i>	10	0	0	10	1	10
<i>doprostředka</i>	8	0	0	10	0	0
<i>vprostředku</i>	4	0	0	4	0	0

Skupina 4. Všechny doklady homonymních jednotek jsou specifikovány v korpusu SYN2010 jako doklady příslovce.

Do Skupiny 4 patří jednotky, které nemají v korpusovém slovníku morfologické analýzy předložkovou interpretaci (např. kvůli zastaralosti nebo nízkému výskytu). Nalezené výskyty těchto jednotek v předložkové funkci jsou ve většině případů spíše výjimkou, což dokazuje výše uvedená tabulka. Avšak pro některé jednotky, jako např. *blíže*, *navrch*, *nablízku* (viz výše), je podíl předložkových výskytů chybně disambiguovaných jako příslovce dosti značný. Například podle korpusových statistik je jednou z nejfrekventovanějších kolokací „příslovce“ *navrch* slovní tvar *hlavy* (85 výskytů v pozici 1P, 83 z těchto výskytů mají tag GenSg). V tomto případě se jedná především o frazém (*mít*) *oči navrch hlavy*. Mimo to se *navrch* jako předložka vyskytuje v korpusu také v přímém místním významu (ještě 10 výskytů, viz tabulka), přičemž v pravé pozici (první a resp. také druhé) má většinou správně rozpoznáný tvar 2. pádu. Srov. *Navrch hotového koktejlu dejte drčený led a vše zakápněte brusinkovým džusem* (SYN2010). Doplnění slovníku morfologické analýzy by mělo tyto chyby odstranit. Stejně jsou však vyžadována doprovodná disambiguační pravidla, která by zabránila možnému množení chyb v automatické disambiguaci, což se dá vzhledem k nové homonymii očekávat.

Ojedinělou jednotkou v přehledu korpusové distribuce zkoumaných homonym je jednotka *skrze* (knižní varianta jednotky *skrz*), která je v korpusu jednoznačně interpretována jako předložka s frekvencí 2420 výskytů (SYN2010). Srov. *skrz* má v korpusu obě interpretace, přičemž větší část dokladů této jednotky je specifikována jako doklady předložky (viz Skupina 1). Tato neshoda je také podmíněna obsahem slovníku morfologické analýzy. Avšak ze všech výskytů *skrze* jako předložky byl skutečně nalezen pouze jeden adverbialní výskyt, který v daném případě vzniká jako zřejmá elipsa, obdobně jako výskyt předložky *před* ve stejném kontextu. *Žádá, aby jeho platnost byla prohlášena před farářem (před, nikoli skrze!), resp. před dvěma svědky.* (SYN2010). To znovu dokazuje, že se toto homonymum vyskytuje převážně jako předložka, což jsme již rozebírali výše.

Homonymní jednotka	Příslovce (podle automatické disambiguace)		Homonymní jednotka		Příslovce (podle automatické disambiguace)		
	Počet výskytů	Z toho výskytů předložek (podle manuální disambiguace)	Počet výskytů	Z toho výskytů předložek (podle manuální disambiguace)	Počet výskytů	Z toho výskytů předložek (podle manuální disambiguace)	
		Počet					%
<i>daleko</i>	17812	8	0,04	<i>odspodu</i>	102	1	0,98
<i>bliže</i>	2370	239	10,1	<i>naspodu</i>	87	7	8
<i>vpředu</i>	1430	4	0,3	<i>navrchu</i>	83	12	14,5
<i>navrch</i>	1179	95	8,1	<i>nablízko</i>	22	0	0
<i>opodál</i>	925	14	1,5	<i>zboku</i>	22	1	4,5
<i>nablízku</i>	642	59	9,2	<i>dospodu</i>	21	3	14,3
<i>zespodu</i>	243	3	1,2	<i>zespod</i>	19	5	26,3
<i>svrchu</i>	194	0	0	<i>dospod</i>	16	2	12,5
<i>vespod</i>	140	6	4,3	<i>vespodu</i>	6	0	0
<i>zevně</i>	116	1	0,9	<i>naspod</i>	2	1	50
<i>zespoda</i>	114	1	0,88				

Problematická místa způsobující chyby v automatické disambiguaci homonymních příslovcí a předložek

Zároveň se sondou distribuce homonymních příslovcí a předložek v korpusu SYN2010 byly analyzovány chyby v automatické disambiguaci těchto jednotek. Analýza ukázala, že se korpus mýlí vždy v určitých místech, tzn. vyskytující se disambiguační chyby je možné typizovat a systematizovat.

Výše již bylo uvedeno, že chybná automatická disambiguace může být způsobena nedostatky slovníku morfologické analýzy, jako v případě jednotek ze Skupiny 4, které jinou než adverbialní interpretaci v korpusu nemají. Tato okolnost však ve srovnání s chybami vznikajícími na základě koexistence obou interpretací ve slovníku ještě není tak nebezpečná. Analýza těchto chyb ukázala, že komplikace v automatické disambiguaci homonymních příslovcí a předložek jsou způsobeny nejen samotnými zkoumanými homonymy, ale často jejich nejbližšími kolokáty (především v pozici 1P). Disambiguace homonymní jednotky, která se může vyskytovat buď jako předložka, anebo jako příslovce, přímo koreluje s disambiguací jejich pravých kolokátů, a přitom pokud jde o předložkový výskyt, bude tato korelace obousměrná. Pro správnou disambiguaci předložky je rozhodující správná disambiguace jejího pravého kolokátu a naopak. Tuto situaci ještě více komplikuje možná adverbialní funkce dané homonymní jednotky. „S adverbii je vzhledem k volnému slovosledu v češtině obzvláštní potíž: mohou být ve větě prakticky kdekoli a nemají zřetelnou syntaktickou spojitelnost“ (Petkevič 2005: 98), kterou stejná jednotka ve funkci předložky samozřejmě má. Srov. následující příklady. *Pak si přitáhla blíž sklenici s narciskami a chvíli se na ně dívala* (SYN2010). Jednotka *blíž* je v daném případě disambiguována korpusem jako předložka a slovní tvar *sklenici* v souladu s tím jako DatSg. *Malé ohně jim plály u nohou, ani ne tak kvůli teplu, ale aby svým světlem přilákaly blíž stříbřité ryby* (SYN2010). Znovu *blíž* jako předložka, a tedy slovní spojení *stříbřité ryby* jako GenSg. Je evidentní, že obtížnost automatické disambiguace homonymní jednotky *blíž* za prvé ještě více narůstá kvůli rekčním vlastnostem této jednotky jako předložky – pojí se se dvěma pády (Gen, Dat) – a za druhé je přímo způsobena homonymií pádových tvarů jmen. Avšak rozsáhlá morfologická homonymie, na kterou je čeština velice bohatá, není jediným faktorem disambiguačních chyb.

Typické jsou chyby způsobené zejména následujícími problematickými kolokáty zkoumaných homonym (především jde o pozici 1P):

(1) Propria cizího původu (jednoslovná a víceslovná), která nejsou ve slovníku automatické morfologické analýzy (nejčastěji jsou tagována jako X@). Např. mezi výskyty jednotky *blízko* disambiguovanými v korpusu jako výskyty příslovce jsme našli celkem 84 výskytů *blízko* v kontextu cizích proprií (1P = X@), přičemž všechny dané výskyty jsou výskyty ve funkci předložky. Srov. *První zločin se stal blízko Epinettes, na druhém konci Paříže* (SYN2010);

(2) Přídavná jména odvozená od přechodníků. Příklady výskytů chybně specifikovaných jako výskyty předložky: *Poslouchám nedaleko hrající jazzovou kapelu* (SYN2010); *Jinde posedává pár kovbojů a nedaleko pasoucí se koně čekají na divokou jízdu* (SYN2010); *Svěřil se vedle sedící ženě, že mu nedávno zemřela manželka...* (SYN2010);

(3) Přídavná jména a přídavná jména, která se vyskytují v postpozici k příslušnému podstatnému jménu a která jsou většinou částí přísudku. Příklady chyb (jednotka je disambiguována jako předložka): *Dveře byly samozřejmě zevnitř zamčené nebo zajištěné* (SYN2010); *Celá scenerie se nejlépe pozoruje z terasy pavilonu, který však není zevnitř přístupný* (SYN2010); *Některé rostliny mají lácny vně lysé, jiné krátce, hustě ochlupené* (SYN2010);

(4) Další příslovce. Příklady výskytů chybně disambiguovaných jako výskyty předložky: *Pro oba nebylo uvnitř dost místa* (SYN2010); *Každou minutu bylo uvnitř víc lidí* (SYN2010); *Zvenčí i v interiéru doslova překvapuje robustností a uvnitř pak také prostorností* (SYN2010); *Toyota Auris je uvnitř designově zajímavé auto* (SYN2010).

Pokud je v případě (1) hlavním důvodem obtížnosti automatické disambiguace nedostatek tvarů ve slovníku (v daném případě cizích proprií), tak v případech (2), (3), (4) je evidentním důvodem chyb absence působení příslušných disambiguačních pravidel, která by například nepřipouštěla disambiguovat výskyt jednotky jako předložku, když tato jednotka evidentně nemá syntaktickou spojitelnost s tvarem (nebo tvary) po ní následujícím, jako například v dokladu: *Koruna zvonkovitá, 10–22 mm dlouhá, světle modrofialová, vně lysá, uvnitř chlupatá* (SYN2010) aj. (viz výše).

Závěr

V příspěvku byla probrána distribuce homonymních příslovcí a předložek s převážně místním významem v korpusu SYN2010 a předloženy výsledky

manuálního ověření úspěšností automatické disambiguace těchto jednotek v daném korpusu. Bylo ukázáno, že problematika disambiguace daných slovnědruhových homonym v korpusu vzniká na pozadí rozsáhlé morfologické homonymie v češtině. Vyšlo však najevo, že chyby v automatické disambiguaci, které byly v příspěvku systematizovány podle problematických kolokátů zkoumaných homonymních jednotek, vznikají také z technických důvodů, a to jednak z důvodu nedostatku tvarů ve slovníku automatické morfologické analýzy a jednak z důvodu nedostatku v morfologickém analyzátoru disambiguačních pravidel, která by mohla mnohé z uvedených chyb odstranit. Jako další etapu výzkumu vidíme rozpracování souhrnu pozitivních a negativních lingvistických pravidel se zřetelem na lexikální, morfologické a syntaktické okolí homonymních příslovčí a předložek, který by bylo možné následně uplatnit k zlepšení automatické disambiguace daných jednotek v korpusu.

Bibliografie:

- Cvrček, V. a kol. (2010): *Mluvnice současné češtiny 1. Jak se píše a jak se mluví*. Praha: Karolinum.
- Čermák, F. (2012): *Morfematika a slootovorba češtiny*. Praha: Nakladatelství Lidové noviny.
- Jelínek, T. – Petkevič, V. (2011): Systém jazykového značkování současné psané češtiny, in: Petkevič, V. – Rosen, A. (eds.), *Korpusová lingvistika Praha 2011, sv. 3: Gramatika a značkování korpusů*. Praha: Nakladatelství Lidové noviny / Ústav českého národního korpusu, s. 154–170.
- Karlík, P. – Někula, M. – Rusínová, Z. (eds.) (2012): *Příruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.
- Komárek, M. – Kořenský, J. – Petr, J. – Veselková, J. (eds.) (1986): *Mluvnice češtiny II. Tvarosloví*. Praha: Academia.
- Kroupová, L. (1985): *Sekundární předložky v současné spisovné češtině*. Praha: Ústav pro jazyk český ČSAV.
- Petkevič, V. (2005): Spletitá krása homonymie češtiny a jak se s ní vypořádat, in: Panevová, J. – Hladká, B. (eds.), *Aleg(r)ace pro Evu*. Sborník u příležitosti 70. narozenin prof. Evy Hajičové. Praha: [S.l.], s. 95-100.
- Český národní korpus – SYN2010. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z WWW: <<http://www.korpus.cz>>.