

Liší se mluvené a psané texty ve valenci?¹

Marie Mikulová – Jan Štěpánek – Zdeňka Urešová
Ústav formální a aplikované lingvistiky, MFF UK Praha
{mikulova, stepanek, uresova}@ufal.mff.cuni.cz

Valency in spoken and written texts - is there a difference?

ABSTRACT: In this article, we introduce the Prague Dependency Treebank of Spoken Czech. The syntactic and semantic annotation of this corpus has led to the expansion of PDT-Vallex, a valency lexicon of Czech verbs, which has previously been linked only to the annotation of written texts. The expansion of the lexicon consisted of several steps: (i) new verbs were added to the lexicon; (ii) new meanings and new valency frames were added to verbs that had already been included in the lexicon; valency frames that had already been part of the lexicon were enriched with (iii) new participants (actants) and (iv) new formal realizations of participants (actants). All the above mentioned enrichments are (a) unmarked and based only on the addition of a new verb, a new meaning, a new participant (actant) or a new form, however, (b) some of them are influenced by the typical characteristics of spoken language. It would be almost impossible to find some of the verbs, some of the meanings, participants or forms in a written text. We believe that verbs in spoken language tend to exhibit different valency behavior than verbs in written texts. In this article we attempt to draw more general conclusions on the valency behavior of verbs in spoken language.

0 Úvod

Liší se mluvené a psané texty ve valenci? Mohou mít tatáž slovesa v mluveném textu jiné valenční chování, než mají v textu psaném? Valence sloves je zpravidla zkoumána na základě psaných textů. Otázku, zda se v mluveném textu dané sloveso nechová jinak, si autoři valenčních slovníků nepokládají.²

Nově budovaný korpus mluvené češtiny, který je anotovaný i na sémanticko-syntaktické úrovni, přináší nové poznatky (mimo jiné) i o valenci sloves v mluveném textu, a to velmi výrazné.

1 Pražský závislostní korpus mluvené češtiny

Pražský závislostní korpus mluvené češtiny verze 2.0 (dále PDTSC 2.0; zkratka z anglického *The Prague Dependency Treebank of Spoken Czech 2.0*) je víceúčelový korpus mluvených českých dialogů. 6 174 minut (neboli 624 380 tokenů či 73 835 vět) spontánních dialogů bylo nahráno, přepsáno a anotováno na několika vzájemně propojených rovinách. Korpus PDTSC 2.0 má podobnou strukturu jako již dříve publikovaný Pražský závislostní korpus 2.0 (Hajič et al., 2006; dále PDT 2.0). Na dvou nejnižších rovinách korpusu PDTSC 2.0 je zachován věrný přepis mluvené řeči: výstup z automatického rozpoznávače řeči a manuální přepis mluvené

¹ Příspěvek vznikl za finanční podpory projektů GA ČR P406/2010/0875 a 406/10/P193, projektu MŠMT LINDAT-Clarín LM2010013 a projektu Amalach DF12P01OVV022 podporovaného Ministerstvem kultury.

² Tradice valenčních slovníků není v rámci celosvětové lingvistiky příliš dlouhá. První valenční slovníky začaly vznikat před čtyřiceti lety, v českém prostředí až v 90. letech 20. století. Dnes pro češtinu existují valenční slovníky jak tištěné (např.: *Slovesa pro praxi* (Svozilová – Prouzová – Jirsová, 1997), *Valenční slovník českých sloves* (Lopatková – Žabokrtský – Kettnerová, 2008) – dostupný i v elektronické verzi jako *Vallex 2.5* (<http://ufal.mff.cuni.cz/vallex/2.5/>), tak elektronické (např.: *VerbaLex* (Hlaváčková - Horák, 2006), *Czech Syntactic Lexicon* (Skoumalová, 2001).

řeči (se všemi jejími zvláštnostmi). Doslovný manuální přepis mluvené řeči je na další rovině nahrazen standardizovaným textem; byla zde provedena tzv. rekonstrukce standardizovaného textu z mluvené řeči.³ Automatickými nástroji je na další rovině standardizovaným větám přiřazena anotace povrchové syntaxe. Nejvyšší rovina sémanticko-syntaktické anotace byla vytvořena manuálně anotátory.⁴

1.1 Dialogy

Korpus PDTSC 2.0 obsahuje nahrávky dvojího typu. Nejprve byla využita česká část korpusu, který vznikl v rámci mezinárodního projektu Malach.⁵ Jedná se o lehce moderované dialogy s lidmi, kteří přežili holocaust. Dialogy obvykle začínají několika krátkými replikami, po kterých následují delší monology. Řeč na nahrávkách je spontánní a je velmi ovlivněna emocionálním stavem řečníka i jeho stářím (průměrný věk řečníků je 75 let). Obsahuje často šepot, kašláni, smích, pláč apod.

Druhou část korpusu tvoří dialogy, které byly nahrány v rámci projektu Companions.⁶ Tématem dialogů je konverzace nad osobní sbírkou fotografií jednoho z řečníků. Cílem projektu bylo vytvoření virtuálního společníka, který by si s lidským uživatelem povídal o jeho fotoalbu. Rozhovory byly nahrávány v nastavení „Wizard of Oz“: tazatel vystupuje při nahrávání jako virtuální postava na obrazovce počítače, ve skutečnosti je však tato postava řízena člověkem ve vedlejší místnosti a dotazovaný jakožto uživatel systému při nahrávání nevěděl, že jeho virtuální společník je ovládán člověkem.

2 Valenční slovník PDT-Vallex

Anotace valence v pražských závislostních korpusech spočívá v tom, že každému výskytu slovesa anotátor přiřadí vhodný valenční rámec z valenčního slovníku PDT-Vallex a závislá doplnění ohodnotí odpovídajícími značkami funkcí podle zvoleného rámce. Pokud náležitý rámec ve slovníku není, je slovník rozšířen (o nové sloveso, nový rámec, novou formu) tak, aby danému výskytu slovesa bylo možné přiřadit odpovídající valenční rámec.

2.1 Budování slovníku

Koncepce slovníku PDT-Vallex (Hajič et al., 2003; Hajič - Urešová, 2003; Urešová, 2006 a 2011) vychází z valenční teorie vypracované v rámci Funkčního generativního popisu (např. Sgall et al., 1986; Sgall, 1998; Panevová, 1980 a 1998). První verze slovníku PDT-Vallex vznikla při anotaci korpusu PDT 2.0. Před začátkem anotací neexistoval žádný valenční slovník a slovník vznikal paralelně se sémanticko-syntaktickou anotací vět. Také z tohoto důvodu nejde PDT-Vallex do hloubky a obsahuje téměř výlučně ta slovesa a ty jejich významy, které se vyskytly v anotovaných datech, tedy ty, jejichž valenční rámce musel anotátor znát, aby dokázal správně anotovat jednotlivá valenční doplnění a další doplnění slovesa v anotované větě.

³ Při rekonstrukci standardizovaného textu z mluvené řeči jsou z doslovného přepisu nahrávky odstraňovány zvláštnosti mluvené řeči (neplynulosti, přefeknutí, neřečové události jako zakašláni, smích apod.), proud mluvené řeči je rozčleněn do vět a výsledné věty jsou upraveny tak, aby odpovídaly standardům psaného textu. Pravidla rekonstrukce jsou popsána v anotačním manuálu (Mikulová, 2008), stručně též v článku Mikulová – Urešová (2008).

⁴ První tři roviny korpusu (automatický a manuální přepis, rekonstrukce řeči) byly již publikovány ve verzi PDTSC 1.0 (Hajič a kol., 2011). Verze PDTSC 2.0 (vydání je plánováno do konce roku 2013) bude již obsahovat i anotace na obou syntaktických rovinách.

⁵ Viz <http://malach.umiacs.umd.edu/>.

⁶ Viz <http://www.companions-project.org/>.

V rámci dalších anotačních projektů byl slovník dále rozšiřován. První rozšíření slovníku přinesla anotace české části Pražského česko-anglického závislostního korpusu (Hajič et al., 2011; dále PCEDT 2.0; zkratka z anglického *The Prague Czech-English Dependency Treebank 2.0*).⁷ Korpus PCEDT 2.0 obsahuje články z deníku Wall Street Journal (z roku 1989), které byly pro českou část korpusu přeloženy do češtiny. Jedná se převážně o texty s ekonomickou tematikou. PDT-Vallex byl tudíž hojně rozšířen o slovesa a významy z této oblasti (např.: *nakonfigurovat, podhodnocovat, porcovat medvěda, prát peníze, segmentovat trh, seškrtnout finanční prostředky, srovnat se s riziky*). Další velké rozšíření slovníku přinesla anotace mluveného korpusu PDTSC 2.0 (o tomto rozšíření viz dále v kap. 3).

V tab. 1 jsou jednotlivá rozšíření valenčního slovníku vyjádřena v číslech. Po anotaci korpusu PDT 2.0 obsahoval valenční slovník 5 510 slovesných hesel a 9 191 valenčních rámců. Anotace dalších korpusů, které jsou s korpusem PDT 2.0 více méně srovnatelné (korpus PCEDT 2.0 obsahuje téměř stejný počet vět, tyto věty jsou však v průměru delší; korpus, PDTSC 2.0 pak obsahuje velké množství krátkých vět s mnoha slovesy), obohatila slovník vždy o dalších cca 1 500 nových hesel a 2 500 nových valenčních rámců. (Zatím) poslední verze slovníku obsahuje přes 8 000 slovesných hesel a přes 14 000 valenčních rámců.⁸

Tabulka 1 - Rozšiřování valenčního slovníku

		PDT 2.0	PCEDT 2.0	PDTSC 2.0
Data	Počet tokenů	833 195	1 151 150	742 257
	Počet vět	49 431	49 208	73 835
	Počet výskytů sloves	88 103	118 035	123 017
	Počet přiřazených hesel	5 376	4 880	4 625
	Počet přiřazených rámců	7 674	8 285	7258
Slovník	Počet hesel ve slovníku	5 510	7 104	8 357
	Počet rámců ve slovníku	9 191	11 933	14 344
		PDT-Vallex 1.0	PDT-Vallex 2.0	PDT-Vallex 3.0

Tabulka 2 - Příklad valenčního hesla ve slovníku PDT-Vallex

⁷ Též <http://ufal.mff.cuni.cz/pcedt2.0/en/index.html>.

⁸ Čísla, která se týkají korpusu PDTSC 2.0, i čísla, která se týkají slovníku PDT-Vallex 3.0, nejsou definitivní. Korpus se kontroluje a opravují se nalezené chyby a tyto zásahy mají na uvedené údaje ještě vliv.

* věřit

ACT(.1) PAT(.3,že[.v],.c,.v) v-w7581f1 Used: 116x

věřil literatuře

v. své ženě a jejím schopnostem

v. tomu, že je to pravda

v., že ho to nezklame

ACT(.1) PAT(v-I[.4],na-I[.4]) v-w7581f2 Used: 20x

věřili v Boha

v. na strašidla

ACT(.1) ?PAT(.4,že[.v],.c,.v) ADDR(.3) v-w7581f3 Used: 18x

věřil mu všechno, co říkal

nevěřili byste nám, jak je to důležité

ACT(.1) PAT(v-I[.4]) v-w7581f4 Used: 6x

věřil v jeho schopnosti

v. v sebe (navzájem)

ACT(.1) ?PAT(o-I[.6]) EFF(.4,že[.v]) v-w7581f5 Used: 1x

činnost rady, o níž Albrightová věří, že má být rozšířena

2.2 Obsah valenčního hesla

Jednoduché valenční heslo v PDT-Vallexu je znázorněno v tab. 2. Valenční heslo ve slovníku PDT-Vallex obsahuje:

- lemma (pro jedno lemma může být a bývá ve slovníku více rámců).
- valenční rámce (jeden rámec vždy odpovídá jednomu významu uvedeného lemmatu).

Valenční rámec obsahuje:

- členy rámce označené značkami pro jejich funkce (ACT – aktor, PAT – patient, ADDR – adresát, EFF – efekt atp.), případně označené otazníkem vyjadřujícím fakultativnost členu (např. ?PAT – fakultativní člen, ACT – obligatorní člen);
- pro každý člen rámce jsou uvedeny jeho (zjištěné) povrchové realizace (formy);⁹
- ilustrativní příklad.

2.3 Zachycení nestandardních jevů ve valenčním slovníku

Anotace mluveného korpusu PDTSC 2.0 si vyžádala nové úpravy v zápisu valenčních hesel. Pro označení různého stupně nestandardních jevů byl do zápisu valenčních hesel zaveden znak procenta (%). Tento znak lze použít v následujících kontextech:

- za lemmatem, kde značí nestandardní lemma. Jeden znak % se používá pro nespisovná, expresivní či jinak „podivná“ lemmata (např. *čumět %*). Dva znaky % mají vulgární lemmata (např. *chlastat %%*).
- za celým rámcem. Zde % označuje nestandardní slovesný rámec, nějaký méně obvyklý význam daného slovesného lemmatu (např. *bruslit: ACT(.1) PAT(v+6) % Bruslil jsem v chemii.*). Dva znaky % používáme pro vulgární slovesné významy (např. *držet: ACT(.1) DPHR (hubu) %% Drž hubu!*).
- za značkou pro funkci valenčního členu, kde označuje nestandardní valenční člen, který se v daném významu obvykle neužívá, a působí proto nepatříčně (např. *dobýt: ACT (.1) PAT(.4) ?ORIG%(od+2) Angličané dobyli Palestinu od Turků.*).

⁹ Uvádí se tzv. kanonické formy (srov. Urešová, 2011).

- za formou, kde označuje nestandardní formální realizaci daného valenčního členu, která se obvykle neužívá a v psaném textu by působila nepatřičně, stylisticky neobratně (např. *dráždit: ACT(.1) ADDR(.4) ?PAT(k+3;na+4%) Dráždí mě to na kašel.*).

Různé kontexty užití znaku % lze v rámci zápisu valenčního hesla kombinovat. Pomocí znaku % u lemmatu a zároveň u valenčního rámce zachycujeme případy, kdy u nespisovné podoby (např. *píct %*) jinak běžného spisovného slovesa (*péci*) označuje jeden z valenčních rámců příznakový význam (např. *ACT(.1) PAT(s+7) % Mohl bych píct s jinou.*), zatímco další valenční rámce označují nepříznakové významy (např. *ACT(.1) PAT(.4) Budeme píct koláče.*).

3 Obohacování valenčního slovníku o mluvené jevy

Některá rozšíření valenčního slovníku PDT-Vallex při anotaci mluvených textů korpusu PDTSC 2.0 byla nepochybně daná specifikem mluvených projevů. V psaném textu by se takové sloveso, význam slovesa, valenční člen, forma s velkou pravděpodobností nevyskytly. V kap. 2.3 jsme popsali konvence, kterými jsou tyto příznakové jevy ve slovníku zachyceny. Tab. 3 ukazuje, kolikrát a v jakých kontextech byl výše popsaný znak % využit. Jednotlivé typy příznakového rozšíření slovníku (příznaková lemmata, rámce, valenční členy a formy) popisujeme v následujících kapitolách podrobněji.

	%	%%
Příznakové lemma	151	8
Příznakové rámec	188	12
Příznaková forma	51	0
Příznakový valenční člen	6	0

Tabulka 3 - Příznakové jevy v PDT-Vallexu 3.0

3.1 Příznaková slovesná lemmata

Již jsme konstatovali, že při sémanticko-syntaktické anotaci korpusu PDTSC přibylo do valenčního slovníku cca 1 500 nových valenčních hesel. Tento přírůstek byl způsoben zejména novým tematickým zaměřením anotovaných textů (viz kap. 1.1); do slovníku přibyly hesla z oblasti běžného (rodinného) života jako *háčkovat, houbařit, koledovat, pošťuchovat se, přebalit dítě, přivdat se, sáňkovat, zavařovat*, ale též i hesla z autentických výpovědí pamětníků holocaustu jako *proválčit, vybombardovat, odvlíknout, přežít, srocovat se*.

Do slovníku však přibyly i lemmata, která by se v psaných textech spíše neobjevila. Jedná se o lemmata hovorová, expresivní, familiární, hanlivá, ale i zhrubělá a vulgární. Většina těchto příznakových lemmat je doložena i v našich českých výkladových slovnících; např.: *baštit, ani nedutat, drandit na lyžích, sbuchnout z lanovky, vydyndat z někoho zálohu, pakovat ho ke kraji, pendelovat za prací* a další.¹⁰ Část příznakových lemmat pak není v dostupných výkladových slovnících (natož ve slovnících valenčních) zaznamenána vůbec, taková jsou například slovesa: *doštráchat se, sdrkotat, splanýrovat, odlepkovávat, zbandat* či *zbankat* (př. (1) až (6)).

Některá nová příznaková lemmata ve slovnících sice najdeme, ale s jiným významem, než jaký přinesla anotovaná data. V SSČ je například obsaženo sloveso *flinkat se* ve významu „flákat se, nic nedělat“ (*flinkat se po ulicích*). V datech se však toto sloveso objevilo ve významu „flinkat něco“ (př. (7)). Ledabylá činnost ve škole nebo v práci je vůbec častým

¹⁰ Lemmata jsme vyhledávali ve Slovníku spisovné češtiny pro školu a veřejnost (2003), ve Slovníku spisovného jazyka českého (1989) a v Příručním slovníku jazyka českého (1935–1957).

tématem nahrávek, a slovník tak byl obohacen o celou škálu výrazů pro vyjádření tohoto významu (např. *flákat se, flinkat, haprovat, ulejšvat se, poflakovat se*; př. (7) až (11)).

- (1) *Doštráchal se k porodní babičce.*
- (2) *Sdrkotal jsem ze schodů.*
- (3) *Splanýrovali celou zahradu. Materiál se musel splanýrovat.*
- (4) *Tam se odlepkovávají jikry.*
- (5) *(Jezdili jste často?) Ne, to nás zbankal náš spolužák.*
- (6) *Zbandala mě na němčinu.*
- (7) *V tercii jsem školu začal flinkat.*
- (8) *Začal flákat školu.*
- (9) *Haprovala jsem ve všech předmětech.*
- (10) *Trošku jsem se ulejšvala.*
- (11) *Poflakujeme se jako letní vlaštovky.*

Zvláštní skupinu nových příznakových lemmat tvoří lemmata utvořená nějakým slovotvorným procesem (nejčastěji prefixací, často několikanásobnou) z lemmat nepříznakových. Zřejmě snaha vyjádřit se co nejpřesněji a zároveň co nejstručněji vede mluvčí k takovým novotvarům, jako jsou slovesa v příkladech (12) až (19). Na to, že některé tyto prefixace společně s připojením zvrtné částice *se/si* je možné utvořit od všech nedokonavých (avšak ne iterativních) sloves vždy se stejnou změnou významu, upozornila Hlaváčová – Nedolužko (2012). Předpona *na-* (př. (12)) vyjadřuje význam „hodně něco dělat“, předpona *po-* (př. (13)) vyjadřuje význam „dělat něco v klidu, většinou příjemně“ a předpona *za-* (př. (14)) nese význam „dělat něco po delší dobu a užít si to“.

V datech se vyskytlo také několik neobvyklých novotvarů s předponou *do-* (př. (15)) a *při-* (př. (16) a (17)). Častým slovotvorným procesem je také transpozice (př. (18)), obvykle ve spojení s prefixací (př. (19)).

- (12) *Manžel toho moc nenaskautoval, protože komunisté to zatrhli.*
- (13) *Udělám jim to ke svačině, aby si pojedli.*
- (14) *Zajdu si tam zaplavčikovat dvakrát nebo třikrát do měsíce.*
- (15) *Čekala jsem, až se dopřipraví.*
- (16) *Museli jsme drobet přiopravit střechu.*
- (17) *Autobusák připozastaví, když vidí zájemce o jízdu.*
- (18) *Dnes už se bojím padat, takže jen běžkařím.*
- (19) *Vnučka se pobabila.*

Přesné stylistické posouzení těchto slov ponecháváme lexikografům, je však nesporné, že tato slovesa jsou příznaková a přinášejí do valenčního slovníku jevy (v tomto případě celá slovníková hesla), na které bychom při zpracování pouze psaných textů nejspíše nenarazili.

3.2 Příznakové valenční rámce

Příznakové valenční rámce zachycují významy, které se daným slovesným lemmatem obvykle nevyjadřují, a v primárně psaných textech bychom pro daný význam volili spíše jiný výraz. Neobvyklý je například význam „být v úzkých, neumět“ u slovesa *bruslit* (př. (20)). Přenesený význam „být v úzkých, neumět“ známe spíše z jiného sportu, který se ovšem

provokuje v nezamrzlé vodě. Dalším příkladem příznakového významu je význam „zapomenout“ u slovesa *spolknout* (př. (21)), u kterého SSČ vedle základního významu „vpravit do jícnu“ uvádí jako obvyklé jen přenesené významy „nevyslovit“ (*spolknul jsem odpověď*) a „snést, strpět“ (*musíš leccos spolknout*). Srov. i další př. (22) až (26). Významy užitých sloves jsou naznačeny v závorkách za příklady.¹¹

- (20) *Bruslil jsem v chemii.* („být v úzkých, neumět“)
- (21) *Spolkl jsem, jak se jmenuje ten klub.* („zapomenout“)
- (22) *Počítač se kousnul.* („přestat fungovat“)
- (23) *Konzumuju cestování.* („užívat si“)
- (24) *Všechno za nás zatáhne.* („zaplatit“)
- (25) *Dokopala jsem ho k tomu, že se alespoň vyučil.* („donutit“)
- (26) *Už to prasklo, že se scházíme.* („vyjít najevo“)

Příznakovost rámce může také být dána „jen“ nestandardním uspořádáním jeho valenčních členů včetně neobvyklých forem těchto členů. Většinou jde o neobratné, kostrbaté vyjádření ze strany mluvčího, které však je natolik srozumitelné, že při standardizaci mluvených textů (viz kap. 1) zůstalo zachováno. Příkladem je užití dvou akuzativů u slovesa *cvičit* (př. (27)). Nestandardní akuzativní forma se několikrát vyskytla také u sloves s významem pohybu (př. (28) až (31)) a k těmto slovesům vedle významu „pohybování se někde/odněkud/někam“ přidává ještě další významové odstíny spojené se slovesy, od kterých se sem zřejmě přenesla.

- (27) *Kněz nás děti cvičil všelijaké básničky a písničky.*
- (28) *To se stalo, když jsem přijel poslední jízdu.* („dojet, zvládnout“)
- (29) *Musel jsem celou cestu odjet sám.* („řídít“)
- (30) *Jezdili jsme expresní rychlíky do Prahy.* („obsluhovat, dělat strojvedoucího“)
- (31) *Jezdívávali jsme Německo, Maďarsko, Polsko.* („navštěvovat“)

Další skupinu příznakových valenčních rámců tvoří případy, v nichž je ke slovesu neobratně připojen nový valenční člen, který se však s daným slovesem obvykle nepojí. Vyjádřením dalšího valenčního členu je pozměněn původní význam slovesa, ve valenčním slovníku je proto tento jev zachycen prostřednictvím nového rámce (na rozdíl od přidání pouze nového valenčního členu do stávajícího rámce; k tomu viz dále v kap. 3.3). Sloveso *zedničit* se například objevilo v korpusu s příznakovým rámcem obsahujícím dva valenční členy: aktor a patient v akuzativu (př. (32)).

- (32) *Zedničili jsme dům.*

3.3 Příznakové členy valenčního rámce

Pouze několik ojedinělých případů příznakových valenčních členů jsme do slovníku zanesli jako příznakový valenční člen v jinak nepříznakovém valenčním rámci. Například u slovesa *domluvit se* ve významu „dokončit mluvený projev“ obvykle objevuje jen aktor (např. *Nech mě domluvit.*), v mluveném korpusu se však vyskytly i dva případy, kdy byl u tohoto slovesa vyjádřen ještě valenční člen v akuzativu vyjadřující téma mluveného projevu (př. (33)). U slovesných významů jako *dát si něco do pořádku* nebo *dobýt území* obvykle očekáváme pouze dva valenční členy. V psaném textu by se v těchto užitích nejspíše neobjevilo vyjádření

¹¹ Žádný ze zde uvedených významů není zaznamenán ve SSČ.

stavu věci před tím, než se bude dávat do pořádku (př. (34)), ani vyjádření nepřítele, kterého je třeba při dobývání území porazit (př. (35)).

(33) *To jsme domluvili zemědělství. To jsme domluvili všechno o zemědělství.*

(34) *Dal si do pořádku z vraku auto.*

(35) *Angličané dobyli Palestinu od Turků.*

3.4 Příznakové formální realizace

Poslední skupinu rozšiřování valenčního slovníku o příznakové charakteristiky získané při anotaci mluvených textů představují neobvyklé formální realizace valenčních členů. Jde o případy, kdy valenční člen je vyjádřen formou, která se obvykle neužívá, v psaném textu by působila nepatřičně, stylisticky nevhodně. Při anotaci mluvených textů jsme se setkali s celou řadou neobvyklých předložkových vazeb (př. (36) až (40)). Zpravidla jde o nějakou kontaminaci, o zkřížení vazeb náležejících k různým (významově, případně i zvukově podobným) slovesům.

(36) *Mám trošku chrapot a dráždí mě to na kašel.*

(37) *Mám k té fotografii hovořit?*

(38) *Mám jednu fotku, která k tomu navazuje.*

(39) *Nepocítovali jsme to za dřinu.*

(40) *Vždycky jsem strašně tíhla na ségřiny děti.*

V mluvených projevech se také objevily zjednodušené formální realizace. Ve větách byl použit pouze „jednoduchý“ akuzativ jména namísto „složitější“ předložkové vazby (př. (41) až (44)). Jinou zjednodušenou formální realizací, která se v mluvených projevech objevuje vedle „jednoduchého“ akuzativu, je infinitiv slovesa (př. (45)).

(41) *Mám pokračovat celou moji anamnézu?*

(42) *Co bych ještě vzpomínala?*

(43) *Po cestě jsem jí zkoušela násobilku.*

(44) *Pamatujete se nějakou píseň?*

(45) *Babička mě zaučovala různé věci. Zaučoval jsem ho jezdit na elektrikách.*

Jako příznaková forma se také jeví některé realizace přímého předmětu pomocí závislé klauze připojené podřadící spojkou *že* jako v př. (46) až (50). Právě tato realizace valenčního členu byla nejčastější příznakovou formou, o kterou jsme rozšiřovali seznamy formálních realizací.

(46) *Požádali jsme si, že bychom chtěli do ciziny.*

(47) *V Pramenu se dostávalo, že jim zaplatili rekreaci.*

(48) *Začali připravovat, že se to rozdělí.*

(49) *Prožili jsme, že třeba byly nálety v Rokycanech.*

(50) *Zželelo se nám, že jsme Petra nevzali s sebou.*

Závislá klauze připojená spojkou *že* je také formou, kterou se v mluvené řeči do již plně obsazeného valenčního rámce slovesa dostává další informace, další valenční člen. Například u slovesa *věřit* byly v PDT-Vallexu před anotací mluvených textů valenční rámce, které jsou uvedeny v tab. 2. Kvůli větě v př. (51) byl druhý valenční rámec rozšířen o příznakový valenční člen vyjadřovaný závislou klauzou se spojkou *že*.

(51) *Věřila na Mikuláše, že za ní skutečně přišel.*

4 Závěr

Anotace valence sloves v Pražském závislostním korpusu mluvené češtiny ukázala, že slovesa v mluveném textu mohou mít jiné valenční chování, než mají tatáž slovesa v textu psaném. Skladba valenčních členů u konkrétního slovesného významu může být v mluvené řeči jiná – bohatší, než v textu psaném. Seznam forem pro vyjadřování valenčních členů, zdá se, je v mluvené řeči v mnohém jednodušší (složitě předložkové vazby jsou nahrazovány jednoduššími bezpředložkovými pády). Slovesa jinak nepříznaková nabývají v mluveném textu nových (metaforických) významů. Valenční slovník byl při anotaci navíc obohacen o nová příznaková lemmata (o slovesa hovorová, expresivní, slangová, o celou řadu různých novotvarů).

REFERENCE

- HAIJČ, J. et al. (2011): *Prague Czech-English Dependency Treebank 2.0*. Data/software. Prague: Institute of Formal and Applied Linguistics; <http://ufal.mff.cuni.cz/pcedt2.0/>.
- HAIJČ, J. et al. (2011): *Prague Database of Spoken Czech*. Data/software. Prague: Institute of Formal and Applied Linguistics; <http://ufal.mff.cuni.cz/pdtsc1.0/>.
- HAIJČ, J. et al. (2006): *Prague Dependency Treebank 2.0*. Data/software. Philadelphia: Linguistic Data Consortium.
- HAIJČ, J. et al. (2003): PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In: *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*. Vaxjo: Vaxjo University Press, 57-68.
- HAIJČ, J. - UREŠOVÁ, Z. (2003): Linguistic Annotation: from Links to Cross-Layer Lexicons. In: *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*. Vaxjo: Vaxjo University Press, 69-80.
- HLAVÁČKOVÁ, D. – HORÁK, A. (2006): VerbaLex - New Comprehensive Lexicon of Verb Valencies for Czech. In: *Computer Treatment of Slavic and East European Languages*. Bratislava: Slovenský národný korpus, 107-115.
- HLAVÁČOVÁ, J. – NEDOLUŽKO, A. (2012): Příklad pravidelných slovotvorných vzorců v automatickém zpracování češtiny a ruštiny. In: *ITAT*. Bratislava: Přírodovedecká fakulta, Univerzita P. J. Šafárika v Košiciach, 53-56.
- LOPATKOVÁ, M. – ŽABOKRTSKÝ, Z. – KETTNEROVÁ, V. (2008): *Valenční slovník českých sloves*. Praha: Karolinum.
- MIKULOVÁ, M. (2008): *Rekonstrukce standardizovaného textu z mluvené řeči v Pražském závislostním korpusu mluvené češtiny. Manuál pro anotátory*, Technická zpráva č. TR-2008-38, Praha: Ústav formální a aplikované lingvistiky MFF UK.
- MIKULOVÁ, M. - UREŠOVÁ, Z. (2008): Rekonstrukce standardizovaného textu z mluvené řeči. In: *Čeština v mluveném korpusu*. Praha: Lidové noviny, 167-176.
- PANEVOVÁ, J. (1980): *Formy a funkce ve stavbě české věty*. Praha: Academia.
- PANEVOVÁ, J. (1998): Ještě k teorii valence. *Slovo a slovesnost*, 59, 1-14.
- Příruční slovník jazyka českého (1935–1957)*. Praha: Ústav pro jazyk český ČSAV; <http://bara.ujc.cas.cz/psjc/>.

SGALL, P. et al. (1986): *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Praha: Academia.

SGALL, P. (1998): Teorie valence a její formální zpracování. *Slovo a slovesnost*, 59, 15-29.

SKOUMALOVÁ, H. (2001): *Czech Syntactic Lexicon*. PhD thesis, Prague.

Slovník spisovné češtiny pro školu a veřejnost (2003). Praha: Academia, 3.vydání.

Slovník spisovného jazyka českého (1989). Praha: Academia; dostupný i z WWW: <http://ssjc.ujc.cas.cz/search.php>

SVOZILOVÁ, N. – PROUZOVÁ, H. – JIRSOVÁ, A. (1997): *Slovesa pro praxi. Valenční slovník nejčastějších českých sloves*. Praha: Academia.

UREŠOVÁ, Z. (2006): Verbal Valency in the Prague Dependency Treebank from the Annotator's Viewpoint. In: *Insight into Slovak and Czech Corpus Linguistics*, Bratislava: Veda, 93-112.

UREŠOVÁ, Z. (2011): *Valenční slovník Pražského závislostního korpusu (PDT-Vallex)*. Praha: Ústav formální a aplikované lingvistiky MFF UK.