

Hlásili to v rozhlase aneb vzájemné mapování lokace a aktoru v dvojjazyčném závislostním korpusu

Jana Šindlerová

Absolventka oboru Český jazyk a literatura – Anglistika – amerikanistika na FF UK v Praze, vědecká pracovnice a doktorandka Ústavu formální a aplikované lingvistiky MFF UK v Praze, korpusová lingvistka. Odborně se zabývá valencí sloves, zejména v mezijazykovém srovnání, vztahem mezi syntaxí a sémantikou, analýzou sentimentu a evaluativními konstrukcemi v češtině.

Abstrakt

V příspěvku se zabýváme jedním typem strukturní odlišnosti mezi českými a anglickými paralelními větami v dvojjazyčném korpusu PCEDT, konkrétně vzájemným mapováním valenčních doplnění aktorového typu a (nevalenčních) lokačních doplnění. Toto nestandardní mapování způsobuje v paralelních závislostních datech asymetrie dvojího typu, jednak prostý nesoulad tektogramatických značek, tzv. funktorů, a jednak konflikt korespondujícího slovníkového mapování a reálného nekorespondujícího mapování v datech. V případě sloves komunikace pak nesoulad těchto funktorů zasahuje interpretaci zdroje předávané informace, což je jedna z klíčových rolí v analýze sentimentu. Analýzou korpusových příkladů ukazujeme, do jaké míry se na rozdílech ve valenci sloves v překladově ekvivalentních větách podílejí jazykově specifické syntakticko-sémantické preference konkrétních slov na pozicích subjektu a predikátu a jak je toto mapování ovlivněno vzájemnou souhrou těchto preferencí, větné diateze a sémantického principu potlačení agentu ve větné struktuře.

Klíčová slova: valence, závislostní korpus, upozadění agentu, překladové ekvivalenty.

Keywords: valency, dependency treebank, agent backgrounding, translation equivalents.

Abstract

The paper documents one type of structural difference between Czech and English parallel sentences in the parallel Prague Czech-English Dependency Treebank, the mutual alignment of valency participants of the „Actor“ type and free modifications of the „Location“ type. This non-standard alignment causes two types of structural asymmetry in the data: the functor mismatch, i.e., the difference in participant labeling, and the conflict of the alignment stored as default in the valency lexicon and the real alignment in the data. In the case of communication verbs, the functor mismatch of the aligned participants influences the interpretation of the Source of information, which is one of the key roles to be identified

within the Sentiment Analysis. In the analysis based on corpus examples, we present our opinion that these differences originate in the interplay of various means of agent demoting, language specific preferences for filling in the position of subject and predicate of a sentence and sentence diathesis.

1. Úvod

Tento článek je inspirován výzkumem strukturních odlišností valenčního chování sloves v češtině a angličtině na základě srovnávání překladově ekvivalentních vět v dvojjazyčném korpusu. Konkrétně se budeme zabývat slovesnými dvojicemi, v nichž je na jedné straně překladu podmětové aktorové doplnění¹, které na druhé straně překladu odpovídá doplnění lokativnímu, viz (1).

- (1) (a) „The obligation is totally unwarranted,“ the statement_{ACT} said.
(b) „Dluh je zcela přemrštěný,“ stojí v prohlášení_{LOC}.

Vzájemné mapování aktoru a lokace je zajímavé zejména proto, že se jedná o mapování aktantové pozice na pozici nevalenčního volného doplnění, tedy mapování, které zasahuje sémanticky prominentní pozici a transformuje ji v pozici propozičně upozaděnou.

Pokusíme se zde ukázat, že ekvivalence lokativního a aktorového vyjádření mezi jazyky není jen jedním izolovaným typem upozadění agentu ve větné perspektivě, ale že se jedná o souhrn různých syntaktických prostředků, jimiž je agens ve větné perspektivě potlačován.

2. Analyzovaná data

Tento výzkum je založen na srovnávání konstrukcí v Pražském česko-anglickém závislostním korpusu (PCEDT) a dvojjazyčném valenčním slovníku CzEngVallex, který je jeho součástí.

PCEDT (Hajič a kol. 2012) obsahuje cca 50 000 dvojic českých a anglických vět anotovaných na rovině hloubkově syntaktické, tektogramatické struktury. Korpus vznikl na základě ručního překladu části amerického korpusu Penn Treebank, sekce Wall Street Journal.² Jednotlivé uzly tektogramatických stromů jsou prolínkovány pomocí automatické procedury, díky čemuž je možno snadno sledovat lexikální i gramatické shody a rozdíly v překladově ekvivalentních strukturách.

Valenční slovník CzEngVallex (Urešová a kol. 2015) byl vytvořen jako doplňkový zdroj pro anotaci PCEDT. Obsahuje cca 22 000 dvojic vzájemně ekvivalentních slovesných významů, včetně vzájemného prolínkování jednotlivých valenčních participantů v rámci. To nám umožňuje popsat rozdíly na úrovni obecných tendencí, nikoli jen náhodných jednotlivostí v syntaktických datech.

Jak PCEDT, tak CzEngVallex jsou zpracovány v rámci valenční teorie funkčního generativního popisu (Panevová 1980). V této teorii do valenčního rámce sloves vstupují tzv. aktanty, tj. sémanticky bazální argumenty větné struktury popisující základní participanty propoziční situace (aktor, patiens, adresát, efekt – výsledek děje – a origo – původ) a některá volná doplnění (lokační, časové, způsobové, příčinné upřesnění děje atd.), pokud jsou tato volná doplnění v propozici sémanticky obligatorní.

¹ Aktor je ve valenční teorii funkčního generativního popisu prvním participantem valenční struktury, obvykle nese sémantiku konatele děje.

² Česká strana paralelního korpusu tak představuje vždy jednu z nepřeberných možností překladu anglických vět. Jednostrannost korpusu vzhledem ke směru překladu představuje jisté omezení vzhledem k platnosti zde předkládaných závěrů, pro směr překladu z češtiny do angličtiny jsme však neměli k dispozici dostatek věrohodných plně anotovaných dat.

V tomto výzkumu vycházíme z hypotézy, že pro každou dvojici slovesných významů existuje jednoznačné mapování jejich valenčních participantů. Konfliktní mapování je z principu vyloučeno. Pokud se přesto v datech objeví, poukazuje na příliš nízkou míru diferenciací formálního popisného rámce.

V rámci prolínání participantů považujeme za nepříznakový stav vzájemné mapování stejných participantů (např. ACT – ACT). Přírozeně může nastat stav, kdy u dvojice významů neodpovídá počet valenčních pozic (ACT – 0), nebo si vzájemně odpovídají participanty různého druhu (ACT – LOC). Tento typ rozdílu ve větných strukturách obvykle ukazuje na jazykově inherentní rozdíly – při překladu dochází ke změnám situační perspektivy nebo k posunům v sémantické interpretaci syntaktické pozice.

Mapování aktoru (ACT) a lokace (LOC) v PCEDT zasahuje oba zmíněné typy disproporcí v prolínání významů. Buď toto mapování funguje jako prostý „functor mismatch“, tedy provázání nestejných participantů bez konfliktu ve struktuře samé, viz Obr. 1, nebo se projevuje konflikt reálného provázání různých participantů v datech s uloženým mapováním stejných participantů na stejné ve slovníku, viz Obr. 2.

Obrázek 1: Functor mismatch. Mapování participantů větné struktury pro dvojici vět:

EN: Next week's report could very well show an increase in crude inventories.

CZ: Ve zprávě z dalšího týdne by se proto mohl objevit nárůst zásob ropy.

Obrázek 2: Část tektogramatického stromu ukazující konflikt slovníkového a reálného mapování participantů. Zelené šipky ukazují nepříznakové uložené slovníkové mapování (ACT–ACT), modré šipky navržené mapování faktické (ACT–LOC).

EN: The merger agreement says the buy-out group is entitled to be repaid \$26.7 million in fees for its investment bankers [...]

CZ: Ve smlouvě o fúzi se uvádí, že vykupující skupina má nárok na úhradu 26.7 milionu dolarů za poplatky investičním bankám [...]

Mezi propozicemi, jichž se dotýká nesoulad, popř. konflikt mapování funktorů ACT a LOC, nacházíme množství sloves, která mají sémanticky evaluativní potenciál. Zejména jde o slovesa sémantické třídy komunikace a vyjadřování postojů (říci, psát, informovat, uvádět). Pozice aktoru u těchto sloves je silně spojena s pozicí „zdroje“ evaluativního významu. Dvojí možné konstruování pozice zdroje, tedy aktorové a lokativní, může přinášet problémy při automatické analýze evaluativních vyjádření v rámci analýzy sentimentu.³ Zároveň nám konflikty mapování u těchto sloves ukazují možné rozdíly vyjadřování evaluativního významu mezi jazyky, posuny v evaluativním významu při překladu, tj. otevírají důležité otázky k dalšímu zkoumání sémantické ekvivalence z hlediska translátologického.

3. Lokativní subjekt jako typ upozadění agentu ve větné perspektivě

Část jevů, které zde popisujeme, byla pro češtinu popsána jako případ tzv. lokativního subjektu. Lokativní subjekt jako případ systémové, lexikalizované elipsy je zmíněn v Mluvnici češtiny 3 (Daneš a kol. 1987: 44). Věty s lokativním subjektem se zde řadí mezi prostředky deagentizace a jsou označeny jako tzv. „konstrukce s transponovaným užitím formy 3. osoby plurálu.“ Neřadí se sem již ale konstrukce s jinou než primární diatezí, např. věty typu „O tom se hovořilo na schůzi.“

³ Analýza sentimentu (Sentiment Analysis, Opinion Mining, postojová analýza) jsou metody automatické extrakce vyjádření subjektivních postojů a emocí z textu. Slouží jako užitečný nástroj zjišťování veřejného mínění v oblasti politiky, médií i trhu (Pang a Lee, 2008, s. 1–135), (Veselovská, Hajič a Šindlerová, 2012).

Společně tyto dvě konstrukce řeší Panevová (1979) v článku o všeobecném konateli. Uvádí, že obě vyjadřují deagentnost a jsou vzájemně synonymní. Možnou souvislost konatele s lokativní frází však nezmiňuje.

Konstrukcemi s transponovaným užitím formy 3. osoby plurálu se zabývají i Grepl a Karlík (1983), ale ani zde není specifikována souvislost anonymizovaného či generalizovaného agentu přímo s lokativní frází.

Oproti tomu Šmilauer (1969) mluví o možnosti vyjádření podmětu příslovečným výrazem v lexikalizovaných elipsách označujících např. instituce, ačkoli tuto tezi dále nerozvádí.

Viták (2015) v bakalářské práci o vyjadřování všeobecného konatele v češtině a angličtině (na překladových datech) lokativní frází ve spojení s 3. os. pl. aktivního slovesa označuje jako specifikační vzhledem ke konateli.

Malá a Brůhová (2016) v dosud nevydané studii spojují lokativní subjekty se sémantickým vzorcem „uvádění na scénu“ a na základě paralelního korpusu vymezují některé typické sémantické třídy, které se s tímto vzorcem spojují.

Shrme-li situaci, existuje v odborném diskurzu nejistota ohledně pozice lokativní fráze v konstrukcích s potlačenou pozicí agentu. Někteří autoři ji přímo spojují s pozicí subjektu, jiní hovoří o lexikalizované elipse, někteří jí přisuzují dějové specifikační roli, ale od subjektové (resp. aktorové) pozice ji oddělují.

4. Mapování pozic lokace a aktoru v závislostních datech

Při hledání propozičních dvojic s nestandardním mapováním ACT–LOC jsme vyhledávali dvojice vět, v nichž jsou tato doplnění vzájemně prolínkována a zároveň jsou dětmi slovesných uzlů. V paralelních syntakticky anotovaných datech jsme zjistili nápadnou asymetrii mezi angličtinou a češtinou. Pro rozložení ACT_{EN}-LOC_{CZ} jsme našli 357 relevantních výskytů v datech. Naprostá většina těchto výskytů byla zcela systémová, paralelní věty si sémanticky odpovídaly. Oproti tomu pro rozložení ACT_{CZ}-LOC_{EN} jsme našli pouze 173 výskytů. Jednalo se ve většině případů o nestandardní konstrukce – volný překlad (2), žargon (3), frazémy (4), syntakticky specifické konstrukce (5), apod.

(2) (a) There's a pretty good program inside all the noise_{LOC} of ``Mancuso FBI."'

(b) Na to, kolik humbuku_{ACT} kolem seriálu "Mancuso FBI" panuje, je to opravdu dobrý pořad.

(3) (a) The total short interest in Nasdaq stocks as of mid-October was 237.1 million shares, up_{LOC} from 223.7 million in September.

(b) Celkový počet akcií prodaných nakrátko byl u akcií Nasdaq v polovině října 237.7 milionu akcií, což je oproti září zvýšení_{ACT} z 223.7 milionu akcií.

(4) (a) The rivals currently are locked in a costly bidding contest_{LOC} for Hollywood film rights.

(b) Protivníci jsou momentálně zablokováni nákladnou nabídkovou soutěží_{LOC} o práva k hollywoodským filmům.

(5) (a) She was in her most radiant, expressive voice_{LOC}.

(b) Její oslňující, výrazný hlas_{ACT} zazněl v plné kráse.

Pouze malý počet paralelních vět představoval případy podobného druhu jako u opačného směru překladu (6).

(6) (a) The legislation also requires broad new disclosures of the race, sex and income level of borrowers, but that information won't be gathered in new studies_{LOC} for several months at least.

(b) Legislativa také vyžaduje rozsáhlé nové údaje o rase, pohlaví a úrovni příjmů dlužníků, ale nové analýzy_{ACT} tyto informace nebudou získávat přinejmenším po několik měsíců.

V analýze rozdělíme případy vzájemného mapování ACT_{EN} a LOC_{CZ} v závislostních datech do dvou skupin. První skupinu tvoří případy, kdy je takové mapování pro daná slovesa základní, tj. nejčastější nebo jediné možné, a proto uloženo jako výchozí mapování ve slovníku CzEngVallex. V datech nedochází ke konfliktům, ale pouze k tzv. „nesouladu funktorů“ (functor mismatch), tj. prolínování různých funktorů bez dalších vlivů na strukturu věty. Druhou skupinu tvoří slovesa, jejichž základním a ve slovníku uloženým mapováním je $ACT-LOC$, resp. souladné mapování syntaktických funktorů (jež odpovídají sémantické interpretaci agentu), přičemž v závislostních datech dochází na jedné straně k upozadění agentu, sekundárnímu vyjádření pomocí lokativní fráze, a tudíž ke konfliktu reálného prolínování v syntaktických stromech a prolínování navrženého překladovým valenčním slovníkem.

4.1 Nesoulad funktorů

V CzEngVallexu se vyskytuje 55 dvojic slovesných rámců (významů) s uloženým mapováním české lokace na anglický aktor. Jedná se vesměs o překlady pomocí sémanticky konverzivního slovesa. Tato sémantická konverze je často založena na kontrastu dynamičnosti (ACT) a státnosti (LOC) slovesa, viz výše př. (1), popř. kauzativitě a absenci kauzace u slovesa (7).

(7) (a) The machine_{ACT} can *run* software written for other Mips computers, the company said.
(b) Na počítači_{LOC} mohou *běžet* programy napsané pro ostatní počítače Mips, uvedla společnost.

Konverzivní slovesa jsou při překladu volena z různých důvodů. Buď v češtině neexistuje vhodné kauzativní sloveso, které by bylo možno spojit se jménem v pozici subjektu (8), neexistuje syntakticky adekvátní překlad, nebo je dynamická/kauzativní varianta (*postrádá*) vnímána jako stylově neadekvátní (9).

(8) (a) The plant has a hairy stem that_{ACT} *produces* flowers and diminutive seeds.
(b) Rostlina má chlupatý stonek, na němž_{LOC} *rostou* květy a maličká semena.

(9) (a) Half the cells_{ACT} taken from lung cancer tissue they tested *are missing* this gene.
(b) U poloviny_{LOC} buněk odebraných z tkáně rakoviny prsu, které testovali, tento gen *chybí*.

Mezi případy, kdy je mapování $ACT-LOC$ uloženo jako základní již ve slovníku, se ale někdy jedná o dvojici významů, která se v datech objevila unikátně, a proto bylo do CzEngVallexu uloženo prolínování aktoru s lokativní frází, přestože sloveso teoreticky umožňuje agentní výraz v pozici subjektu (10). Pokud by se alternativní mapování $ACT-ACT$ v datech objevilo, jednalo by se již o konfliktní struktury.

(10) (a) The proposal_{ACT} *reiterates* the U.S. desire to scrap or reduce a host of trade-distorting subsidies on farm products.
(b) V návrhu_{LOC} *je zdůrazněn* požadavek Spojených států na zrušení nebo omezení vysokého počtu dotací na zemědělské produkty, které narušují obchodování.

4.2 Konfliktní struktury

Ke konfliktům dochází v případech, kdy překladově ekvivalentní slovesné významy nejsou primárně sémanticky konverzivní. Obě slovesa potenciálně otvírají pozici sémantického agentu, a přitom na obou stranách překladu dochází k tomu, že tento agens nenaplní syntaktickou pozici aktoru, ale je nějakým způsobem ve struktuře potlačen. Na anglické straně překladu se do pozice subjektu dostává entita, která nemá primárně interpretaci konatele, ale je s konatelem v určitém metonymickém vztahu (11). Jde o výrazy, které implicitně dovolují lokativní interpretaci: názvy institucí (11a), geografické entity, státy (11b), média, dokumenty (11c), apod.

(11) (a) The House action yesterday came as the Senate_{ACT} *remained* mired in difficulties over a \$17.25 billion measure covering the budgets for the State, Commerce, and Justice departments in fiscal 1990.

(a') Ke včerejšímu jednání sněmovny došlo, jelikož v senátu_{LOC} *přetrvávaly* problémy ohledně ustanovení o částce 17.25 miliardy dolarů, které se týká rozpočtů na fiskální rok 1990 pro ministerstvo zahraničí, obchodu a spravedlnosti.

(b) After numerous occurrences of questionable teacher help to students, Texas_{ACT} *is revising* its security practices.

(b') Po častém výskytu případů diskutabilní pomoci učitele studentům *přehodnocují* v Texasu_{LOC} svůj systém zabezpečení testů.

(c) The form_{ACT} *asks* for such details as the client's name, Social Security number, passport number and details about the services provided for the payment.

(c') Ve formuláři_{LOC} *jsou* k vyplnění takové položky, jako je jméno klienta, číslo sociálního zabezpečení, číslo pasu a detailní popis služeb poskytovaných za danou platbu.

Neagentní a potenciálně lokativní charakter takového doplnění dokládá i obvyklá dvojitá možnost anaforického odkázání (12), respektive možnost konstruovat překlad obojím způsobem – pomocí subjektu i pomocí místního určení (13).

(12) (a) The industry, which_{ACT} currently *employs* about 25,000 people in London, has shed about 2,500 jobs over the past two years.

(b) V tomto sektoru, který_{ACT} v současné době *zaměstnává* 25 000 lidí v Londýně, bylo během posledních dvou let zrušeno na 2 500 pracovních míst.

(c) V tomto sektoru, kde_{LOC} *je* v současné době *zaměstnáno* [...]

(13) (a) The report_{ACT} *says* that „the restructuring will take place over a two-year period and will principally involve the transfer and termination of employees in our U.S. operations.“

(b) Zpráva_{ACT} *uvádí*, že „restrukturalizace proběhne během dvou let a bude zahrnovat zejména přesun a ukončení práce zaměstnanců v amerických provozech.“

(c) Ve zprávě_{LOC} *se uvádí*, že [...]

Všechny výskyty uvedeného překladového vzorce mají společný charakter. V anglické větě se objevuje sloveso v primární diatezi, které sémanticky otevírá pozici agentu, ale na subjektové pozici má entitu, kterou nelze interpretovat agentivně. Na české straně potom nalézáme různé kompenzační strategie u predikátů. Slovesa mohou zůstat významově blízká, ale mění se větná perspektiva, diateze slovesa či jeho reflexivita. V (14a) je na české straně pasivní

slovesná forma, (14b) je dokladem užití reflexivního tvaru slovesa,⁴ (14c) můžeme považovat za příklad rezultativního užití konstrukce slovesa „být“ s participiem.

(14) (a) Two Minneapolis shareholder suits_{ACT} in the past month *have accused* top officers of making „various untrue statements.“

(a') Minulý měsíc *bylo* ve dvou žalobách_{LOC} ze strany akcionářů z Minneapolisu nejvyšší vedení *obviněno* z toho, že činí „různá nepravdivá prohlášení.“

(b) The exchange_{ACT} *had forecast* that the contract would begin trading by December.

(b') Na burze_{LOC} *se předpovídalo*, že smlouva se začne obchodovat od prosince.

(c) The statement_{ACT} *added*, „Someone has regrettably chosen to selectively summarize portions of highly classified correspondence between the two branches of government.“

(c') V prohlášení_{LOC} *je dále uvedeno*: „Někdo se bohužel rozhodl selektivně shrnout některé části přísně tajné korespondence mezi těmito dvěma vládními institucemi.“

Nakládání s výše uvedenými konstrukcemi nám ukazuje, že formální stránka predikátu a lokačního doplnění zastupujícího konatele spolu neoddělitelně souvisí. Jak transformace subjektu v adverbiale, tak změna slovesné diateze patří mezi strategie potlačování agentu ve větě struktuře a vzájemně se při generování výsledné jazykové struktury podmiňují.

Závěr

Na počátku jsme v rámci primární hypotézy řekli, že konflikty v syntakticky anotovaných paralelních datech jsou obvykle důsledkem malé diferenciovanosti teoretického popisného rámce. V tomto případě jde zejména o výraznou sémantickou podspecifikovanost pozice aktoru. Tektogramatická interpretace v rámci FGP klade velký důraz na morfologickou formu daného doplnění. Je-li dané doplnění v subjektu a nominativní, pak je převážně volena interpretace aktorová, zatímco interpretace lokativní je fakticky vyloučena. Oproti tomu lokativní forma fakticky vylučuje interpretaci aktorovou, a to i přesto, že v české lingvistické tradici je v některých teoriích lokativní subjekt připouštěn. Angličtina tak na rozdíl od češtiny snadno zaplňuje pozici aktoru sémanticky neagentními výrazy, pokud je agens nespécifický, generalizovaný. To je důsledek nutnosti zaplnit formální subjektovou pozici ve struktuře. V případě lokativních subjektů se tak děje na bázi metonymie. Čeština naopak snadno ponechává subjektovou pozici formálně prázdnou a nespécifičnost agentu vyjadřuje pomocí škály příznakových morfologických prostředků. Lokativní fráze ze sémantického hlediska zaplňuje uvolněnou pozici propozičně upozaděného aktantu, a to nejen pozici aktorovou. V striktně sémantickém pojetí bychom tak mohli uvažovat nejen o lokativním konateli, ale např. i o lokativním původu (ORIG): „Dozvěděl jsem se o tom *na schůzi*.“ nebo lokativním adresátu (ADDR): „Oznámím to dnes *ve škole*.“

Poděkování

Tento článek vznikl za částečné podpory Grantové agentury České republiky v rámci projektu č. GA15-06894S K lingvistické struktuře evaluativního významu v češtině. This work has been using language resources developed and/or stored and/or distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth, and Sports of the Czech Republic (project LM2015071).

⁴ Vzhledem ke komplexnosti problematiky zařazení reflexivních slovesných forem k aktivní či pasivní interpretaci se zde nebudeme zabývat konkrétním dělením a podržíme se generalizované terminologie uváděné v Mluvnici češtiny 3.

Reference

- Daneš, F. – Hlavsa, Z. – Grepl, M. (1987): *Mluvnice češtiny 3, Skladba*. Praha: Academia.
- Grepl, M. – Karlík, P. (1983): *Gramatické prostředky hierarchizace sémantické struktury věty*. Brno: UJEP.
- Hajič, J. a kol. (2012): *Prague Czech-English Dependency Treebank 2.0*. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University in Prague, <http://hdl.handle.net/11858/00-097C-0000-0015-8DAF-4>.
- Malá, M. – Brůhová, G. (2016) Tracking presentative semantic patterns through parallel corpora, in *KOLT 2016: Abstrakty*. http://ucnk.korpus.cz/kolt2016/KOLT2016_abstrakty.pdf
- Panevová, J. (1980): *Formy a funkce ve stavbě české věty*. Praha: Academia.
- Panevová, J. (1979): Všeobecný konatel a jeho vztah k mluvčímu, in *Otázky slovanské syntaxe*. IV/1, Sborník sympozia Aktualizační (pragmatické) složky výpovědi v slovanských jazycích. Brno: Univerzita J. E. Purkyně, s. 101-106.
- Pang, B. – Lee, L. (2008): *Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval*.
- Šmilauer, V. (1969): *Novočeská skladba*. Praha: Státní pedagogické nakladatelství.
- Urešová, Z. a kol. (2015): *CzEngVallex*. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University in Prague, <http://hdl.handle.net/11234/1-1512>.
- Veselovská, K. – Hajič jr., J. – Šindlerová, J. (2012): Creating annotated resources for polarity classification in Czech, in *Proceedings of KONVENS*, s. 296-304.
- Viták, F. (2015): *Vyjádření obecného konatele v češtině a angličtině*. Bakalářská práce. Plzeň: ZČU.