

Oponentský posudek bakalářské práce Daniely Bodanské

Homonymie pasivních participií v současné češtině

Bakalářská práce Daniely Bodanské se zaměřuje na náhodnou slovnědruhovou homonymii pasivních participií a jiných slovních druhů než sloves a adjektiv (např. *brána, bit, minuta, buzen*). V teoretické části práce přináší popis vlastností pasivních participií ve vybraných českých mluvnicích a popis procesu značkování českých jazykových korpusů. Autorka poté pro svůj výzkum vybírá pasivní participia, u nichž v korpusu *SYN2010* dochází či může docházet k chybnému značkování, a na základě analýzy jednotlivých příkladů se snaží identifikovat pravidla, jejichž zařazení do značkovacího procesu by zvýšilo úspěšnost slovnědruhového značkování slovních tvarů. Cílem práce je tak přispět dílčím způsobem ke kvalitě značkování českých jazykových korpusů.

Tento cíl je v úvodu práce dobře představen a autorka zároveň přehledně popisuje strukturu celého textu. Na druhou stranu jsem v úvodní kapitole postrádala jednak vymezení slovnědruhové homonymie (včetně upřesnění, že se v Petkevičově klasifikaci jedná o homonymii náhodnou, což se čtenář dozvídá až na str. 23), jednak alespoň několik příkladů pasivních participií, aby bylo zjevné, jakými typy se bude autorka v práci zabývat.

První kapitola teoretické části věnovaná dosavadnímu popisu pasivních participií v české lingvistice parafrázuje informace z chronologicky seřazených mluvnických či encyklopedických děl. Struktura této části je přehledná, nicméně z několika důvodů se jedná o nejproblematictější část celého textu. Za prvé je zřejmé, že se jedná o výběr z mluvnic a syntetických příruček počínaje Trávníčkovou *Mluvnici spisovné češtiny* z roku 1951. Není přitom jasné, proč jsou některé mluvnice zařazeny a některé (například *Čeština – řeč a jazyk* nebo *Mluvnice současné češtiny*) vynechány. Lze si proto například klást následující otázky: Co z informací, jež přináší Trávníčkova mluvnice, nenajdeme v jiných pracích? V čem spočívá jeho přínos pro pochopení pasivních participií, když zrovna jeho terminologie a kategorizace slovních druhů je odlišná od dnešního odborného úzu? Proč na druhou stranu chybí perspektiva výše zmíněných mluvnic? Je to proto, že nepřinášejí žádnou další relevantní informaci nad rámec zjištění uvedených v zařazených příručkách, nebo proto, že se je autorka rozhodla z nějakého nespécifikovaného důvodu nezohlednit? Proč se v obecném popisu pasivních participií neberou v potaz odborné studie? Souhrnně řečeno, domnívám se, že tato kapitola práce dobře dokládá, že tam, kde autor nepodává úplný přehled vývoje v přístupu k určité lingvistické problematice, není chronologický postup nejvhodnější. Mnohem adekvátnější je kompilační postup, kde autor vybírá z relevantních publikací důležité informace a řadí je logicky podle postupného zpřesňujícího popisu daného jevu, nikoliv podle data vzniku dané publikace, odkud informace čerpá.

Výhodnost nechronologické, logické výstavby přesvědčivě dokládá druhá kapitola teoretické části, v níž se autorka zabývá principy značkování korpusů a postupy, jimiž lze dosáhnout co největší úspěšnosti v automatickém značkování korpusových dat.

V praktické části zaměřené na chyby ve značkování, identifikaci jejich příčin a návrh zpřesňujících pravidel se postupuje podle jednotlivých slovních druhů, s nimiž jsou pasivní participia homonymní. Pozornou analýzou kontextu autorka upozorňuje na zajímavé souvislosti v chybně označovaných případech a zároveň implicitně přispívá k poznání různých aspektů organizace české věty. Ukazuje mimo jiné, jak jedna chyba ve značkování může plodit chyby další (viz například časové adverbium *ráno* chybně označené jako nominativ singuláru substantiva v neutru znemožňuje určit slovní tvar *buzen* jako participium pasiva, neboť s tvarem *ráno* coby adeptem na větný subjekt neshoduje v morfoloických kategoriích). Přesně identifikovat a popsat, proč vznikla chyba ve značkování, a

zároveň předložit návrh jiného řešení je úkol náročný a autorka si v něm počíná rozvážně a její návrhy prokazují schopnost syntaktické analýzy a citlivost pro jazykový kontext. Během obhajoby bych autorku prosila o upřesnění následujících několika bodů:

1. Na straně 27 autorka řeší disambiguaci složité konstrukce *Máte za sebou v posledních týdnech testování ve formuli GP2, která je brána pro piloty jako příprava pro přechod do formule 1*, přičemž konstatuje, že slovosled *kteřá je pro piloty brána jako* by zejména z důvodu spojky *jako* bezprostředně následující za přičestím byla patrně snazší pro automatické značkování. Varianta první je, na rozdíl od té druhé, až do spojky dvojnásobná nejen pro disambiguační program, ale i pro rodilého mluvčího. Prosím autorku, aby rozdíl v dvojnásobnosti obou struktur u obhajoby explicitněji rozvedla.

2. Na straně 29 autorka navrhuje chápat spojení *budit + ráno* jako ustálenou kolokaci, a tedy začlenit ji do anotačního systému jako kolokační pravidlo, a to na základě údaje, že v SYN2010 se toto spojení vyskytuje ve sto případech. Prosím autorku, aby doplnila informaci, jaký je celkový výskyt tohoto slovesa v korpusu, bez tohoto údaje je daná informace nerelevantní.

3. Tvar *děleno* autorka v souladu se značkováním použitým pro SYN2010 hodnotí jako homonymní mezi participiem a spojkou, přičemž upozorňuje, že kategorizace tohoto výrazu v jeho matematickém významu coby spojky není vždy akceptováno. Autorka akceptuje slovnědruhovému řazení přítomné v korpusu SYN, není však jasné, na základě jakých argumentů. Prosím autorku, aby předložila argumenty pro tuto kategorizaci.

Hlavním výstupem předložené bakalářské práce jsou lingvisticky formulované návrhy nových pravidel, jež by po začlenění do značkovacího programu zvýšila přesnost značkování, tj. zvýšila četnost správného určení pasivních participií, resp. s nimi homonymních jiných slovních druhů. V tomto kontextu jsem v práci poněkud postrádala autorčino zhodnocení toho, která z navržených nových pravidel jsou implementačně snadná a která obtížnější. Tj. pravděpodobně přidání kolokační dvojice nebo frazému do frazeologického komponentu je snazším úkonem než požadovat po značkovacím programu, aby zohlednil širší kontext celé klauze. Prosím autorku, aby se při obhajobě pokusila zhodnotit pravidla navržená v práci i po této stránce.

Dále bych ráda, aby autorka u obhajoby podrobněji vysvětlila, jakým způsobem jsou do disambiguační gramatiky začleněna valenční pravidla, tj. informace o valenčním poli jednotlivých lexikálních jednotek, resp. schematizovaných vzorců (typů predikátů). Zdá se to, vzhledem ke zkoumané problematice, být důležitý aspekt, přesto se o něm v teoretické části mnoho nedozvídáme (až na jeden typ negramatických jevů v Olivově klasifikaci na str. 19). Více informací o této složce by také zlepšilo představu čtenáře o tom, jak mohou vypadat nová pravidla spjatá s valencí, jež v praktické části práce autorka místy navrhuje (například valenční rámec v jednom významu pasivního participia *zatažen + do čeho*).

Po stylové a pravopisné stránce je bakalářská práce dobrým příkladem kvalitního odborného textu. Autorčin styl je úsporný a srozumitelný, po formulační stránce lze upozornit jen na drobnosti (například se nedoporučuje začínat první větu celé kapitoly vedlejší větou, viz *Aby bylo možné zabývat se...* na str. 8, dále viz dvojnásobnou strukturu *Takto se předmět v pasivu stává podmětem u sloves přechodných* na str. 11, kterou je možné vyjasnit změnou slovosledu). Po terminologické stránce není příliš uspokojivý anglický abstrakt práce (*homonymic* je adjektivum, správně je *a homonym*, synchronní korpusy se překládají jako *synchronic corpora*, nikoliv *synchronized*).

Jako celek lze předloženou bakalářskou práci hodnotit pozitivně. Kolegyně Bodanská řeší výzkumnou otázku, která je dobře vymezena mimo jiné s ohledem na očekávaný rozsah bakalářské práce. Řešení

výzkumné otázky je uspokojivé a podané mimořádně čtivým a koherentním způsobem. Z hlediska celku se mé připomínky zmíněné výše jeví jako dílčí, a proto práci ráda doporučuji k obhajobě a navrhuji hodnotit ji známkou výborně.

V Praze dne 9. 6. 2015

Mgr. Eva Lehečková, Ph.D.