

Vyjádření školitele k disertační práci

RNDr. Jaroslavy Hlaváčové

Formalizace systému české morfologie a ohledem na zpracování českých textů

Autorka si ve své disertační práci vytkla nesnadný úkol: navrhnout – jak sama uvádí – systém, který jednotně, konzistentně a co nejúplněji popíše všechny morfologické jevy, které jsou potřebné při práci s českým jazykovým korpusem. Tento systém má popsat morfologii češtiny **z hlediska automatického zpracování češtiny**: to jest nejde jen o popis morfologických faktů češtiny samých, nýbrž zejména o popis co nejvhodnější pro počítačovou implementaci. Na jedné straně je tu tedy snaha zachytit celý systém morfologie češtiny (jazyka z hlediska morfologického nesmírně spleťového) teoreticky vhodným způsobem, na straně druhé pak formálně přesně tento systém popsat se zřetelem k automatickému zpracování češtiny. Autorka vychází jednak ze závěrů schůzky skupiny, která se na podzim roku 2005 snažila koncipovat implementovatelný systém morfologie češtiny na národní bázi (konkrétně sjednotit přístup pražský a brněnský), jednak jeho závěry (jež nebyly nikdy prakticky realizovány) tvůrčím způsobem dotahuje.

Ve dvou hlavních částech se zabývá systémem kategorií a hodnot popisujících morfologii současné češtiny včetně lemmatizace (první část práce) a novým způsobem zápisu flektivních a derivačních vzorů (druhá část práce). Základním východiskem v první části je pro ni tzv. Zlaté pravidlo morfologie, tj. jednojednoznačný vztah slovního tvaru a jeho morfologického popisu včetně lemmatu. Podle tohoto pravidla není možné, aby dva různé slovní tvary byly popsány jediným způsobem v podobě dvojice <lemma,morfologická_charakteristika>

Na druhou stranu může mít jeden slovní tvar více morfologických popisů uvedeného typu jen v případě skutečné homonymie tvaru. Ukazuje se, že v dosavadním systému značkování češtiny, přinejmenším pražském, pravidlo někdy neplatí, zejména s ohledem na *varianty*. Varianty jsou obtížné pole, na něž se autorka neohroženě pouští, přičemž zavádí pojem *vícenásobného lemmatu* pro popis variantních lemmat různých typů (obecněčeské, případně nářeční varianty, pravopisné varianty, flektivní a jiné varianty) a *jejich kombinací* a pojem *mutace* (vlastně varianty) tzv. *flektivního* a *globálního* typu. V jistém smyslu autorka vnáší i systém do dosavadního, místy ne úplně konzistentního popisu slovních druhů. Mimo varianty se přitom zaměřuje na obtížné případy lemmatizace, zejména

- na tzv. *složeniny* (složenina je tvar tvořený podtvary/morfy náležejícími dvěma či více lemmatům, nikoli kompozitum v tradičním lingvistickém smyslu). Má tu na mysli slova typu: *jehos*, *očs*, *več*, *zaň*; v práci jsou pojednány všechny možné složeniny rozdělené podle slovních druhů a způsoby jejich vyhledávání v korpusech podle autorčiny koncepce
- na dosud nevhodně zachycené *zvláštnosti* u jednotlivých slovních druhů (záporná lemmata u různých slovních druhů, stupňování typu *sebe-* adjektiv a adverbii, duál atp.).

Podrobně popisuje hodnoty jednotlivých slovních druhů a morfologických kategorií právě se zřetelem k obtížným případům a ve snaze vždy dodržet Zlaté pravidlo morfologie a v budoucnu umožnit uživatelům korpusů značkových podle její koncepce náležitě vyhledávat obtížné případy. Tuto první část pokládám za jádro celé práce.

Ve druhé části práce je podrobně zpracován systém parametrizovatelných flektivních a derivačních vzorů. Zajímavá je pasáž o tzv. postfixovém guesseru.

Práce je psána pěkným slohem, je dobře čitelná a přehledná; nekonzistence a překlepy se v ní hledají jen velmi těžko.

Konstatuji, že jsem s výsledky doktorandčina studia (ač trvalo dost dlouho) a s úrovní její disertační práce spokojen. Doktorandka odvedla velký kus práce v oblasti formálního popisu české morfologie a lze se jen nadít toho, že výsledky její práce se také projeví v praxi při novém morfologickém značkování a lemmatizaci českých textů, zejména korpusových. Mám za to, že prokázala své schopnosti vědecky pracovat, a to ve spleťité oblasti morfologie češtiny a jako školitel doporučuji práci k obhajobě.

V Praze dne 24. 3. 2009



doc. RNDr. Vladimír Petkevič, CSc.
školitel