



**Universität Regensburg**

Universität Regensburg · D-93040 Regensburg

Ústav českého jazyka FF MU  
Prof. PhDr. Petr Karlík, CSc.  
Arna Nováka 1  
CZ-660 88 Brno  
Tschechische Republik

PHILOSOPHISCHE FAKULTÄT III  
SPRACH-, LITERATUR- UND KULTURWISSENSCHAFTEN  
Bohemicum / Institut für Slavistik

**Prof. Dr. Marek Nekula**  
Telefon +49 941 943-3526  
Telefax +49 941 943-1861  
Universitätsstraße 31  
D-93053 Regensburg

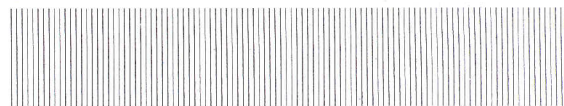
marek.nekula@sprachlit.uni-regensburg.de  
www.bohemicum.de

14. února, 2012

### **Posudek habilitační práce**

**Klára Osolsobě: *Morfologie českého slovesa a tvoření deverbativ jako problém strojové analýzy češtiny*. Brno: Masarykova univerzita 2011. 220 s.**

Monografie Kláry Osolsobě navazuje na vlastní práce autorky k automatické analýze přirozeného jazyka, které vznikly v rámci úspěšně obhájených grantů Grantové agentury České republiky. Ve své monografii se přitom opírá nejen o své dílčí studie, ale také o softwarové a jiné nástroje (analyzátor *ajka*, tagované korpusy...), které vznikly mj. i na bázi jejích výstupů, zvláště na bázi jejího strojového slovníku češtiny se 170.000 kmeny z roku 1996. O výsledky její práce se ostatně opírají i v praxi obecně rozšířené lingvistické aplikace (jazykové korektory, korpusy...). Zatímco ve své disertaci se autorka zaměřila na českou formální morfologii (tvarosloví), její habilitační práce se vyrovnává s derivační morfologií (slovotvorba). Autorka si v ní klade tyto klíčové otázky: Jsou dosavadní popisy derivační morfologie úplné a natolik exaktní, že je lze bez dalšího formalizovat, a pokud ne, jak tato formalizace vypadá? V čem s ohledem na stávající korpusy a nástroje spočívají limity této formalizace, resp. formální slovotvorné analýzy a automatického značkování? Případné limity je přitom podle autorky možno uchopit přes údaje o přegenerování, případně podgenerování dat (viz kap. 3). Jde tedy o práci, v níž vedle formalizace derivačních, přesněji substitučních pravidel hraje ústřední roli empirie a experiment. Ten spočívá v testování formálních substitučních pravidel na korpusech dat, čímž se ověřuje, zda je zvolený formální popis efektivní, a zda vhodně formalizuje, resp. zpřesňuje tradiční morfologický popis. Dle výsledku testu se formální popis dle potřeby dále zpřesňuje. Jedním z využívaných korpusů je přitom i již zmíněný rozsáhlý strojový slovník češtiny, jehož je Klára Osolsobě autorkou.



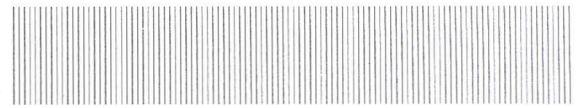
Už autorčin strojový slovník češtiny obsahuje vedle tvarotvorných vzorů i definice některých pravidelných derivací. Předkládaná habilitační práce pak v oblasti derivační morfologie zkoumá možnosti a meze automatické analýzy na příkladu slovesa, které nepochybně představuje morfologicky nejkompexnější slovní druh. To je také důvod, proč se exemplární analýza v monografii soustředí na formální popis derivace sloves a deverbativ, resp. na vybrané pravidelné typy verbálních derivací v češtině. V této souvislosti je třeba podotknout, že toto omezení je vědomé, protože v práci jde o odpovědi na obecnější otázky (viz výše). Součástí práce je ovšem v 9. kapitole i *komentovaná ukázka* derivačního slovníku deverbativ, který je jako *celek* uložen a zpřístupněn na serveru FF MU Brno a který je třeba považovat za svého druhu součást předkládané monografie, a to nikoli jen za součást ve smyslu materiálové přílohy, ale za svébytný *komplexní výstup*, jemuž předcházela formální analýza deverbativ, jejíž správnost strojový slovník svou funkčností dokládá. Tím roste i váha předkládané monografie, která vysvětluje architekturu a exemplárně dokládá funkčnost této široce využívané aplikace, potažmo i funkčnost formálního popisu derivační morfologie.

Předpokladem automatické slovtvorné analýzy je přirozeně formální popis, tj. popis zaměřený na formu a formalizaci jejího popisu. Autorka ve své práci vychází z grafické formy slova či slovního tvaru (s. 15), vztah základového a odvozeného slova chápe ve formálním smyslu jako vztah „záměny“ (s. 17). Ve 4. kapitole navíc usouvztažňuje formální a derivační morfologii, přičemž až na blíže nedefinovaný termín „konekt“ zůstává kompatibilní s tradiční terminologií Dokulilovy školy. S ohledem na segmentaci jednotek ale samozřejmě reflektuje novější formální popis české morfologie formální i derivační.

Za klíčovou považují 5. kapitolu autorčiny monografie, kde na rozdíl od předchozích obecných kapitol vstupuje do detailního popisu pravidel variantnosti lexikálních a tvarotvorných kmenů jako tvarotvorných a slovtvorných spoluformantů. Rozlišuje v ní alternaci kmenotvorné přípony (KmV) včetně alternace souhlásky, která je součástí kmenotvorné přípony (kt), alternaci samohlásky/diftongu v základu (KoV), alternaci finální, resp. iniciální souhlásky kořene (kf, ki), alternaci samohlásek v prefixu (PV) a alternaci na švu prefixu a základu (P0e). Detailněji si všímá jejich kombinací při jejich materializaci v korpusech. Na tomto pozadí se pak snaží dosavadní popisy v gramatikách přeformulovat do jednoznačných pravidel, která v 7. kapitole dále formalizuje pro automatickou analýzu a která testuje na elektronických korpusech.

V 6. kapitole autorka vysvětluje využití softwarového nástroje Deriv pro prozkoumání možností a mezí automatické slovtvorné analýzy, přesněji pro testování „pokrytí formálních pravidel popisujících vybrané případy derivací v češtině“ (s. 45), které slouží „optimalizaci navržených funkcí“. Nejde v něm jen o funkce vyhledávání, ale i o funkce prohlížení, třídění, editování a ukládání, jichž využívá i již zmíněná aplikace popsaná v 9. kapitole. Autorka se v 6. kapitole zabývá především otázkou vyhledávání token prostřednictvím řetězců tagů

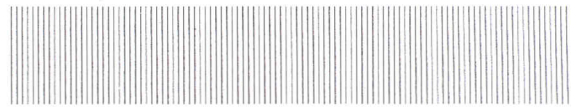




v kombinaci s hláskovými proměnnými a řeší otázku přegenerování a podgenerování a vysvětluje možnosti vyladění dotazu v případě přegenerování. Upozorňuje také na důležitost typologie chyb „při automatickém generování slovotvorných vztahů (...) pro zkvalitnění ručního zpracování automaticky generovaných dat“ (s. 50). Tyto chyby jsou především důsledkem homonymie na nejruznějších úrovních.

7. kapitola pak představuje konkrétní podobu zpracování na počátku formulovaných úkolů, tedy formalizaci deverbativních jmen tvořených od slovesného tvaru, kmene a základu prostřednictvím substitučních pravidel. Tato pravidla jsou podle slovotvorných typů kvantitativně testována s ohledem na svou úspěšnost při vyhledání, tak aby nedošlo k přegenerování, resp. podgenerování. Úspěšnost navržených formalizovaných substitučních pravidel je v této i následující kapitole detailně diskutována jak s ohledem na možnosti jejich zpřesnění, tak s ohledem na míru jejich úspěšnosti v závislosti na parametrech jako připojení sufixu na tvar, kmen a základ i parametrů jako homonymie a alternace, přičemž autorka dochází k závěru, že u přegenerovanosti obecně hraje větší roli strukturovanost sufixu než jeho homonymie, čímž ovšem nepodceňuje vliv homonymie na přegenerování. Konstatuje dále, že přegenerování je nižší tam, kde dochází k hláskovým alternacím, a že deriváty od kořene/základu mají vyšší míru přegenerovanosti než deriváty od kmene. Tím pojmenovává parametry, které omezují efektivnost automatické analýzy jazyka, jejíž limity v detailu ukázala při aplikaci jednotlivých substitučních pravidel. V diskuzi konkrétně ukázala i to, v čem spočívají možnosti automatické analýzy jazyka. Ty je – obecně řečeno – možno vidět v možnosti (opakovaného) experimentálního zaostření substitučních pravidel na základě výsledků jejich testování.

V obecném posudku určeném pro Vědeckou radu Filozofické fakulty Masarykovy univerzity je obtížné detailně diskutovat formální popisy a dílčí řešení. Ta jsou až na výjimky přesvědčivá, validovaná mj. i funkčností zmíněných korpusových a softwarových aplikací. Přesto si autor posudku klade otázku, zda by při dalším vydání habilitační práce nebylo s ohledem na širší lingvistickou obec dobré vsunout explicitní definice pojmů kmen, kořen, konekt aj. a nespoléhat se na jejich implicitní porozumění mj. i na základě tabulky na s. 18-19, jejíž příklady pro běžného otvírají otázky, které pro úspornost výkladu zůstávají nezodpovězeny (tj. proč je u *sběratel*, *hnětač*... jako tvarotvorný formant vedena Ø, zatímco u *soudce*, *rytec*... nikoli, proč je u *opilec* jako tvarová koncovka vedena Ø, zatímco u *zarytec* nikoli). Jen z příkladů nemůže být běžnému čtenáři jasné, v čem spočívá podstata konektu a v čem se liší od kmentotvorné přípony, přičemž užití termínu konekt se, jak autorka poznamenává, liší od užití termínu v tradiční slovotvorbě. Na s. 24 si autor posudku klade otázku, zda je vhodné modelovat výklad na příkladu *tisknout/tisknul/tisknut/tisknutý*..., když vedle toho existují varianty *tisknout/tiskl/tištěn/tištěný*..., nebo zda by nepřispělo k lepší orientaci uvádět na straně 24-25 příklady nejen u pozorování 2, ale tak u pozorování 1 a 3 (příklady lze ovšem dohledat na dalších stranách, resp. v poznámkách na dalších stranách)



apod. Chybí také vysvětlení různých interpretací v poznámce 30 na s. 28, snad by bylo dobré ověřit umístění „ale“ na 2. ř. zdola na s. 32 a v poznámce 63 na s. 39 adekvátnost příkladu *důlkovat* apod. Větší problém vidím jen v tom, že se v práci – s ohledem na korpusové zdroje – bazíruje na psané formě, jež má nicméně potenciálně fonemickou realizaci, takže se klade otázka segmentace např. u typů *trp-ě-n* nebo tam, kde se bije fonemické realizace a umístění  $\emptyset$ , jak je tomu např. na s. 32 u *vypušt- $\emptyset$ -ěn*, což je dobře vidět i při zformulování pravidel na s. 36-37, kde se v pravidlech objevuje zjednodušený fonologický zápis, zatímco v příkladech se užívá grafematický zápis. Jak vidět, jde ale v těchto případech jen o detaily, které lze snadno opravit, resp. s ohledem na obecného čtenáře v dalším vydání podle potřeby zformulovat empatičtěji.

Celkově lze říci, že autorka v teoretické části předložila velmi hutně psanou práci, kterou docení zvláště ten, kdo se zabývá formálním popisem jazyka a má zkušenost s jeho automatickou morfologickou analýzou. Práce se cele soustředí na dílčí, nutno ale třeba podotknout klíčový a nejobtížnější výsek derivační morfologie a sleduje v této souvislosti obecnější – klíčové – otázky matematické lingvistiky. V žádném případě tak nejde o práci fragmentární. Je jí navíc třeba vidět v širším kontextu autorčiných studií, monografií a aplikací, které zahrnují jak formalizovanou formální morfologii, tak formalizovaný lexikon, tedy jádrové lingvistické disciplíny. Práce je navíc kontextualizována v širším kontextu automatické analýzy jazyka, jejíž výsledky autorka s přehledem shrnuje. Širší kontextualizace práce v rámci matematické lingvistiky chybí, autorka nicméně odkazuje na svou studii v kolektivní monografii *Kapitoly z dějin české jazykovědné bohemistiky (2007)*, v níž je tato kontextualizace provedena. Je také v této souvislosti třeba podotknout, že úsporná a hutná diskuze teoretických a metodických východisek představuje v matematické lingvistice běžný postup. Daleko větší důraz je v ní kladen na empirii a experimentální řešení formulovaného úkolu, které se testuje, dokladuje a validuje, resp. dle potřeby dále zpřesňuje. Metody matematické lingvistiky a automatické analýzy nelze proto bez dalšího měřit lingvistickými metodami jiných metodologických škol. Ty autorka nicméně samozřejmě bere v úvahu ať už směrem k tradiční lingvistice, nebo směrem k teoretické lingvistice jako v případě *templatic morphology*. Reflektuje totiž a kriticky diskutuje jejich poznatky, řešení i terminologii. V tomto smyslu jednoznačně prokazuje, že je nejen profilovanou specialistkou, jejíž řešení jsou východiskem obecně užívaných aplikací, ale také lingvistkou, která se účastní širšího lingvistického diskurzu.

Práce tak z mého pohledu splňuje požadavky standardně kladené na habilitační práci a doporučuji její přijetí.

(Marek Nekula)