

Posudek habilitační práce Kláry Osolobě: *Morfologie českého slovesa a tvoření deverbativ jako problém strojové analýzy češtiny*. Spisy Masarykovy univerzity v Brně číslo 401, Brno 2011

Sám problém je naznačen už na nápadité obálce knižního vydání monografie. Pravidelnost formy (strojová) se rozchází s jazykovou praxí. (Je sice *trpět - trpělivý - trpělivost*, ale jen *snášet - snášenlivý - snášenlivost*). Klasickou lingvistiku zajímá, zda se vložené *-n-* objevilo analogií s deverbativy od sloves s *-n-* v základu (kořeni), jako *hanět - hanlivý - hanlivost* a nejspíš tento jev označí jako perintegraci. Strojovou lingvistiku pak zajímá jaké pravidlo vymyslet, aby podobné základy byly počítačovým programem rozpoznány a tvar byl vytvořen správně. Klára Osolobě zaměřila svou snahu na morfologicky nejsložitější slovní druh, na sloveso a jeho slovotvorné deriváty. Klade si za cíl poznat možnosti automatické analýzy deverbálních derivací, formálně popsat realizované případy derivace vybraných typů deverbativ a vytvořit taková formální pravidla a takovou metodologii popisu, že bude použitelná i pro jiná automatická zpracování přirozeného jazyka. Teoretickými a praktickými východisky jsou Dokulilova slovotvorná teorie aplikovaná v Mluvnici češtiny a v dílčích statích různých autorů, morfemická segmentace slov Slavičkové a Šišky, strojově zpracovaný SSSJČ a korpusy SYN a SYN2000. Pro alternace formantů s pohybným *-e-* vychází z prací Scheerových a Zikové. Pro výklady sporných případů se obrací k literatuře z etymologie a historie jazyka.

Úvodní kapitoly jsou věnovány představení automatické analýzy přirozeného jazyka na Ústavu českého jazyka na FFMU a dosavadní práci na české formální morfologii a jejím testování na strojovém slovníku češtiny, na nichž se autorka významným způsobem podílela. Její předchozí práce, strojový morfologický slovník, umožňuje ke sto sedmdesáti tisícům kmenů přiřadit pravidlo, kterým lze generovat slovní tvar a slovní druh. Z uvedeného slovníku vychází další nástroj, a to automatický morfologický analyzátor *ajka*. I zainteresovanému laikovi, jakým jsem ve strojové analýze jazyka já, je z hlediska formální analýzy zřejmá potřeba co nejpodrobnější segmentace slov a slovních tvarů. Teoretický přínos do oblasti klasické lingvistiky se pak neodehrává ani tak v teorii morfematiky, i když není vyloučen, ale spíše v doplnění jevů, které zůstaly dosud buď nepovšimnuty, nebo jen na okraji dosavadních popisů. Z hlediska strojové analýzy je jedno, zda budou při stanovení formálních pravidel překračovány hranice klasicky vymezených morfů či hranice morfologie a lexikologie. Plošný pohled na soubor morfů/segmentů (včetně konektů) a jejich grafických alternací a podob poskytuje na soustavu slovesných tvarů nový pohled, v mnohém odlišný od tradičních popisů. Výjimky a okrajové typy jsou na téže úrovni jako zástupci reprezentativních typů a podtypů, protože cílem je nenechat jediné sloveso s jeho běžně i neběžně realizovanými tvary stranou. Výhodou postupu je, že použití pravidla ukáže, zda je zohledněn opravdu každý tvar a každé sloveso ze stanovené skupiny, a také, zda jde o skupiny uzavřené, či otevřené.

Část práce věnovaná nástroji *Deriv* určenému pro testování úspěšnosti automatické slovotvorné analýzy je informací o základních softwarových rysech nástroje, o úskalích tzv. podgenerování a přegenerování, o potřebě postupů, které přispějí k vyloučení nesprávně generovaných jednotek atd. Formulace tzv. substitučního pravidla vychází ze slovníkových údajů a z intuice, pokud jde o významový vztah k základu, a z přesně formulované substituce části jednoho řetězce znaků jiným řetězcem znaků (obojí s přesně definovatelnou posloupností) a toto pravidlo vede ke vzniku dvojice základové slovo a odvozené slovo. Autorka použila základní ideu uvedeného softwaru a dotvořila a zpřesnila jeho pravidla na zvolených deverbativech. Ukazuje se nejen propracovanost *Deriv* na jeho různých úrovních,

ale i možnost získaná data interaktivně sledovat. *Deriv* je propojen s internetovým prohlížečem (*DebDict*), který umožňuje prohlížet materiál pocházející ze zpracovaných výkladových slovníků a dále je propojen s korpusem SYN2000 a dalšími korpusy. Úspěšnost automatické analýzy je dána porovnáním počtu možných derivátů a počtu správných derivací vzniklých aplikací vytvořených formálních pravidel. Zevrubné představení použitých nástrojů, jejich charakteristiky a způsoby použití představují nutný úvod k vlastní problematice práce.

Při poměrně značné homonymii základů i derivačních prostředků vzniká řada problémů. Interaktivní hodnocení chyb je nejen prostředkem k vyloučení nepravých dvojic, ale v rukou tak zkušené autorky i podnětem k hledání cest, jak vznik sporných či přegenerovaných dvojic eliminovat.

Poznámky k jednotlivým typům deverbativ:

Kromě přegenerování i podgenerování řešila autorka také problém, jímž je značný počet variant morfů kořenových i kmenových. Vedle obecných pravidel generujících všechny zkoumané slovotvorné dvojice všech tříd jsou vytvořena pravidla generující dvojice se speciálními či méně početnými alternacemi. K jedinému (nouzovému) příkladu na alternaci -*ě/i prdič* můžeme nabídnout ženskou podobu k *vonět – vonička*.

K deverbativům od l-ového tvaru slovesa v roli příjmení dodejme na okraj, že existují četné případy, kdy se proprium od původní podoby slovesného základu odlišilo: *Otáhal, Zapletal*. Nabízí se otázka, zda neexistují i příjmení pocházející z nestandardních minulých přítčestí, tedy takových, kde -*l* chybí: *nes, spad, moh, bouch*. (Jako příklady zřejmě existují jen složeniny, a to lidový posměšný název *rytíř Skozyspad*, či příjmení *Vozembouch*.) Za vhodné doplnění poskytnuté autorkou většině klasických popisů považujeme ženské a střední podoby proprií typu *Úlehla, Vosáhlo* i uvedená kompozita.

Autorka zaznamenala rozdíl v alternaci kvantity vokálu u slovesa *lhát – lhal – lhaní*, který platí pro prefigovaná slovesa (*zalhat*), nikoli pro kompozita jako *vhrdlohát*. Tento tvar lze považovat za kompozitum jen kvůli staročeské podobě a psaní dohromady v 19. stol. (u B. Němcové). Jako kompozitum však existuje v nářeční už deetymologizované podobě *vrdlouhat*. Zajímavá by byla - zřejmě jen okazionální - forma *velelhát* a *velelhář*, což by svědčilo o neprefixální povaze *vele*.

Pohled na prefigovaná slovesa a jejich tvary z hlediska vokalizace prefixu *ode-, pode- přede-* a vysledování pravidel jejich výskytu u slovesných tříd a tvarů nabízí doplnit ještě útvary s *obe-*.

K už zmíněným rozdílům mezi klasickou a strojovou slovotvorbou připojme další. Klasická derivatologie se snaží dopátrat primárního tvaru pomocí exkursů do historie, slavistiky i dialektologie, a nenajde-li jednoznačné řešení, zůstane u konstatování dvojí možné motivace. Strojová analýza musí jednoznačně určit odvozovací základ i tam, kde je podle klasického výkladu situace ve skutečnosti složitá nebo nejasná: např. při konkurenci slovesa a dějového jména .

K deverbálním adjektivům.

U adjektiv na -*cí* by bylo podle našeho soudu vhodné nevytvářet dvojici *stonat – stůňoucí*, ale nechat *stonat* pod *dělat – dělající*, tedy *stonající*, jinými slovy přihlédnout k faktu, zda

existuje přechodník *stůňouc*, když je brán jako základový tvar pro vznik adjektiva. Jde-li jen o ilustraci typu alternace, pak by byl vhodný příklad méně sporný.

Adjektiva na *-vý* představují svou rozmanitostí variant přípony záležitost spíše slovníkovou než formalizovatelnou. Proto je nutné přesné zachycení alternací ocenit. Při prioritě fundace samozřejmě dochází tu i jinde k oslabení motivace, jako např. *znát - značný*, podobně u řady dalších dvojic ovlivněných historickým vývojem.

Plně podporuji názor autorky, který vyslovila při poznávání složitých vztahů mezi infinitivy a adjektivy na *-ný* stejně jako mezi deverbativními substantivy a infinitivy různého vidu a adjektivy od nich odvozenými. Ten názor praví, že ani větší množství formálních pravidel není s to vystihnout tyto vztahy bezpodmínečně a že nelze jejich povahu vystihnout ani ručním dořešením prostě proto, že vztahy mezi členy slovotvorného hnízda (jak píše autorka) nejsou jednoznačně a přesně dány. Vyjadřujeme se, jak víme, s jistou vágností, stačí nám někdy přibližná motivace a univerzálnější slovotvorný prostředek, jímž sufix *-ný* nepochybně je, a zpřesnění pak zůstává rolí kontextu, což derivační formální pravidlo zohlednit zatím nemůže. O mimořádných schopnostech autorky svědčí nejen technická propracovanost výstavby formálních pravidel, ale i schopnost změnit jisté části postupu, vyžaduje-li to povaha materiálu, jako je tomu u deverbálních adjektiv na *-ný*. Zde se stýká bohatá struktura slovesných základů a tvarů s univerzálností slovotvorného adjektivního prostředku. Autorka pojímá tuto kapitolu jako soubor průzkumných řešení a návrh k diskusi, což svědčí o její odpovědnosti a kritickém vhledu do složité problematiky. Autorčina řešení případů přegenerování exkursy do zatímních řešení a do etymologie, stejně jako vytváření menších skupin derivačních dvojic na základě společných rysů substitučních pravidel jsou přesvědčivá a využitelná nejen pro autory klasických popisů.

Závěrem můžeme konstatovat: Práce představuje velmi ucelené a přesně formulované odborné pojednání o velkém komplexu českého slovesa a jeho derivátů substantivních a adjektivních. Snaha vytvořit platná pravidla derivace s co nejmenšími chybami a přegenerováními ospravedlňuje i postupy jdoucí proti některým detailům v teorii odvozování, která je vedena významovou motivací, vztahem pojmové báze a příznaku, konstrukcí slovotvorné kategorie a typu atd. Podobně je tomu ve vztahu k morfematicce: u segmentací potřebných pro vytvoření pravidel je vedlejší například to, o jaké typy morfů jde, jde-li o perintegraci atd. Díky vypracované metodologii je stavba jednotlivých kapitol přehledná, výklady a postupy se opakují v přesném sledu. Množství materiálu především ze slovníku morfologického analyzátoru *ajka* (založeném na hesláři SSJČ) vede k možnostem objasnit okrajové i neokrajové jevy označované za výjimky nebo nejasné příklady, doplnit opomenutá řešení, vede i k opravám a vyjasněním vztahu mezi homonymními případy (srov. např. adjektiva na *-čí* a činitelská jména). Opravují se tak některá tvrzení z klasické slovotvorby, dochází k zpřesnění vztahů fundačních i fundačně motivačních. Specialisté na strojovou analýzu češtiny ocení výsledný derivační slovník deverbativ popsaných typů a informaci o přístupu k němu na webu i zpracování umožňující interaktivní vstupy do slovníků v DebDict a morfologického analyzátoru *ajka*. Závěrem ráda konstatuji, že není nejmenších pochyb, že **habilitační práce splňuje požadavky standardně kladené na úroveň habilitačních prací v oboru český jazyk.**

Brno 9. února 2012


Prof. PhDr. Zdenka Rusínová, CSc.