

Metodické pokyny k výzkumu slootvorby

Metodické pokyny k výzkumu slootvorby

I. OBECNĚ

Východiskem výzkumu slootvorby zaměřeného na její popis nejprve v Kapitolách z české gramatiky (výzkum bude probíhat v rámci projektu GA ČR č. 405/06/1057 v letech 2006-2008) a poté v prvním svazku dvousvazkové velké korpusové Gramatiky současné češtiny (GSČ I. – Slovo a slovní skupina) budou korpusové nálezy slootvorných derivátů s jednotlivými slootvornými sufixy a jejich variantami (např. –tel, -č, -ačka).

Pro substantivní deriváty budeme pracovat se seznamem sufixů a jejich variant v TŠČ II. Pro ostatní slovní druhy nám bude východiskem Šmilauerovo Novočeské tvoření slov.

Protože máme již vyhledáno, zčásti zpracováno (editováno) a v databázích MaM uloženo značné množství nálezů slootvorných derivátů z korpusu SYN2000, bude žádoucí, abychom primárně i nadále pracovali s tímto korpusem.

K práci s korpusem SYN2005 a s internetovými vyhledávači Google, Morfeo a Jyxo budeme přistupovat za těchto okolností:

- (a) Dokladů derivátů s daným sufixem bude málo (méně než deset dokladů ze tří různých zdrojů) nebo nebude žádný;
- (b) Dokladů derivátů s daným sufixem bude sice poměrně dost (několik desítek), ale budou pocházet z jediného zdroje či z několika málo příznakových zdrojů (např. jen poezie, jen autoři, o nichž je známo, že experimentují s jazykem, jen odborná literatura atd.);
- (c) V SYN2000 nenalezneme doklady derivátů, které jsou uvedeny jako běžné v TŠČ II či v SSJČ anebo které bude za takové pokládat badatel či které bude badatel znát z vlastního úzu;
- (d) Se zadanou prací na bázi korpusu SYN2000 budeme hotovi a budeme chtít ověřovat produktivitu (neproduktivitu) daného derivačního typu na (a) širší materiálové bázi; (b) na novějším jazykovém materiálu

II. METODIKA PRÁCE S KORPUSEM

Principiálně půjde vždy o to, vytěžit z korpusu VŠECHNY DERIVÁTY S URČITÝM SUFIXEM, KTERÉ JSOU V NĚM OBSAŽENY, či alespoň CO NEJVÍCE Z NICH. Pokud bychom se spoléhali na vyhledávání podle lemat a tagů, hrozí nebezpečí, že: (a) Údaje o frekvenci budou více či méně (někdy bohužel velmi) nepravdivé; (b) Značný počet – zejména jedno- a málodokladových derivátů v korpusu nenalezneme.

Proto navrhuji při vyhledávání derivátů, zejména derivátů se specifickým sufixem, u nichž lze předpokládat poměrně malou produktivitu a frekvenci, POSTUPOVAT NÁSLEDUJÍCÍMI METODAMI:

A. Základní metoda vyhledávání

Základní metodou budiž vyhledávání pomocí dotazu typu:

```
[(lemma="*.ouš") & (tag="N.*")]
```

Nález bude obsahovat slova jako 'bělouš, teplouš, ...'.

Postup při editaci nálezu budiž tento:

1. Zadáme příkaz provést frekvenční distribuci podle lematu.
2. Seznam derivátů si uložíme do svého wordovského adresáře.
3. Seznam prohlédneme a vymažeme z něj balast (např. vlastní jména, slova bez jasně motivačně-fundačního vztahu, slova nerelevantní formálně-sémantické struktury apod.
4. Pročištěný frekvenční seznam si zkopírujeme a abecedně ho utřídíme. Abecedně seřazený seznam si rovněž uložíme do svého wordovského adresáře.

B. Doplnková metoda vyhledávání č. 1:

Zadáme sekvenci sufixálních písmen, jednak bez koncovky, jednak s jednotlivými koncovkami, a to buď postupně nebo jedním komplikovaným dotazem (je-li to proveditelné) a tag pro nerozpoznané tvary, tj. např.:

```
[(word="*.ouš") & (tag="X@.*")]  
[(word="*.ouše") & (tag="X@.*")] atd.
```

Následuje editační postup jak popsáno výše.

C. Doplnková metoda vyhledávání č. 2

Vyhledáváme pomocí sekvence sufixálních znaků se zástupnými vyhledávacími znaky pro přední a zadní část slova, např.

.ouš.

Výsledkem tohoto příkladového dotazu je ovšem 70 703 dokladů. S tak velkým množstvím obvykle nebudeme pracovat. Zadáme tedy dotaz se zástupnými znaky *. pouze pro začátek slova:

*.ouš

Výsledkem tohoto dotazu je 1883 dokladů, které už můžeme během poměrně krátké chvíle prohlédnout: Zatrhneme si v menu ukázání tagu a pak budeme hledat především doklady označené X@ a doklady špatně otagované. Z nich vybereme relevantní deriváty a zbytek vymažeme. Deriváty s uvedením tagu a zdroje si uložíme a připojíme je k editovanému nálezu téhož sufixu. Výsledek celého postupu viz v Editovaných nálezech pod sufixem -ouš v kategorii Nominace - Jména nositelů vlastnosti.

Je třeba vždy dostatečně popsat způsob hledání a způsob editace! (To dosud v mnohých úložkách - i mých! - chybí!)

III. SÉMANTICKO-STYLOVÁ a STRUKTURNÍ KLASIFIKACE SLOVOTVORNÝCH DERIVÁTŮ

Poté co jsme výše popsanými postupy získali frekvenční a abecední seznam (doufejme) všech či alespoň většiny derivátů s daným sufixem (variantou sufixu), provedeme sémanticko-stylovou analýzu těchto derivátů.

A. KLASIFIKACE VÝZNAMOVÁ

U derivátů s určitým sufixem budeme rozlišovat SÉMANTICKÉ KATEGORIE.
Např. u DERIVÁTŮ NA -(A)ČKA odlišíme

zahajovačka (děj) – zasedačka (místnost) – vrtačka (nástroj) – žehnačka (osoba)

B. KLASIFIKACE STYLOVÁ:

Budeme rozlišovat:

(a) Deriváty LEXIKALIZOVANÉ od NELEXIKALIZOVANÝCH (event. různé stupně lexikalizace – podle údajů slovníků, podle frekvence a podle vlastní řečové zkušenosti (introspekce); např. zabíječka – zdviháčky;

- (b) Deriváty STANDARDNÍ – HYPERSTANDARDNÍ (knižní, resp. formální) – SUBSTANDARDNÍ;
např. hokejista – policista – kopálista;
- (c) Deriváty EXPRESIVNÍ od NEEXPRESIVNÍCH, např. žadatel – chlastatel (event. rozlišíme různou míru a druh expresivity, např. 'vulgární', 'žertovné'; patetické apod.
- (d) Deriváty z hlediska žánrově stylového nespecifikované od takto specifikovaných; např. zabíračka – zdvihačka (sport. slang)

C. KLASIFIKACE STRUKTURNÍ

1. Rozlišíme deriváty od různých slovních druhů, zejména DEVERBATIVA, DESUBSTANTIVA a DEADJEKTIVA.
2. Rozlišíme všechny rozdílné formální a/nebo formálně-významové vztahy mezi slovem základovým a derivátem. Např. 'kutil' - od sloves s infinitivním kmenem na -i- (kutil) x 'čumil' (výjimečně odvozeno od slovesa s kmenotvorným -e- (čumět).