



Univerzita Karlova v Praze  
Filozofická fakulta  
Ústav anglistiky a amerikanistiky  
Náměstí J. Palacha 2, 116 38 Praha 1, Tel. 02-21619341, Fax 02-21619382

**Ondřej Tichý, Digitization of Old and Middle English Dictionaries.** Diplomová práce.

UK FF, Praha, Ústav anglistiky a amerikanistiky, květen 2007

Posudek diplomové práce

Diplomová práce Ondřeje Tichého se v mnoha směrech vymyká běžné lingvistické diplomové práci, co se týče zaměření, hloubky zpracování a konkrétního výsledku. Zabývá se přehledově teoretickou a praktickou lexikografií staré a střední angličtiny, zabývá se komputerovou lexikografií, a to nikoli na obecné rovině, ale na rovině konkrétní aplikace (vytváření slovníku) se všemi příslušnými technickými aspekty a konečně předkládá funkční elektronický slovník, digitalizovanou podobu anglosaského slovníku Bosworthe-Tollera. Fakt, že realizovaný projekt digitalizace slovníku získal grant Vzdělávací nadace Jana Husa a že projekt navazujícího výzkumu je v takové podobě, že mohl být předložen Grantové agentuře Univerzity Karlovy, mluví sám za sebe.

V přehledu (str. 9-22) papírových slovníků staré a střední angličtiny komentuje autor 34 (!) staroanglických pramenů zabývajících se staroanglickým lexikem a dalších 13 pramenů pro středoanglickou slovní zásobu. U každého pramene je uveden bibliografický údaj a stručná, ale zato zasvěcená charakteristika. Obdobným způsobem je zpracován přehled (str. 29-33) existujících projektů elektronických slovníků staré a střední angličtiny (Bosworth, Bright, McSparran, Mayhew, Roberts, Baker, Healey et al. a dále Johnson, Slade). Přehled uzavírá srovnání komentovaných slovníků podle standardního modelu: makrostruktura (heslář, záhlaví, problémy s jejich kanonickou formou a tříděním), mikrostruktura (součásti hesel) a typografická stránka. Zde bych poznamenal, že u Figure 1 (str. 43) – instrukce k alfanumerickým symbolů (1-11) a (a-e) – nejsou příliš jasné: např. proč číslování od 1, když se 1 v tabulce neobjevuje („to save space“). Zajímavý je pokus (str. 44-48) o kvalitativní zhodnocení obsahu stávajících elektronických slovníků (zejm. ve srovnání s MED, McSparran, Middle English Compendium, a DOE, Healey et al., Dictionary of Old English).

Přehled slovníků je východiskem pro obecnější zhodnocení specifika historických slovníků angličtiny a pro formulování vhodných parametrů pro příští elektronické slovníky a digitalizační projekty v této oblasti (zdroje a metodika digitalizace). Velmi vhodné a moderní je zdůraznění uživatelského aspektu, tzn. koncipování elektronických aplikací s ohledem na předpokládané uživatele. Tímto zhodnocením se autor dostává k anglosaskému slovníku Bosworthe a Tollera, resp. k procesu jeho scannování, který je velmi podrobně popsán ve všech hlavních fázích. Dlužno říci, že tyto technické, popisné oddíly, ač čtenářsky nepřilíživě, jsou velmi užitečné z hlediska jakékoli další digitalizace. Zatím poslední fází digitalizace Bosworthe-Tollera je vytvoření aplikace, tzn. funkční elektronické podoby slovníku, jejímž prostřednictvím Ondřej Tichý poprvé zpřístupnil tento slovník širokému okruhu koncových uživatelů. Závěrečný oddíl nastiňuje dva další úkoly, resp. fáze výzkumu, které by měly v rámci digitalizace Bosworthe a Tollera následovat, sémantické označování dat (semantic tagging) a vytvoření programu, který by dokázala vygenerovat morfologické charakteristiky zvoleného staroanglického slova (morphological analyzer).

**Připomínky, komentář:** K většině částí slovníku nemám připomínky. Přehledové a srovnávací oddíly jsou psány velmi poučeně a kompetentně. Je zřejmé, že autor je dobře orientován v literatuře i v lexikografické teorii a praxi. V oblasti technického zpracování pak znalosti diplomanda natolik převyšují znalosti posuzovatele, že připomínky mohou být nanejvýš metodologického rázu. Jistě by bylo možné některé věci doplnit, uvažovat o alternativních řešeních, ale šlo by spíše o osobní názor a preference. Nicméně se domnívám, že oddíl 6.4 (Application Development, st. 72) mohl být pojat obširněji. Vzhledem k tomu, že valná část toho, co mu předchází, slouží do značné míry jako platforma a zdůvodnění pro vznik aplikace, mohl být tento oddíl koncipován jako samostatná kapitola (zahrnující i příslušní pasáže z kap. 4) uceleně popisující vznik aplikace, problémy s tím spojené i smysl a funkce aplikace. Koneckonců na rozdíl od předchozích kapitol představuje práce na aplikaci autorův originální vklad a tedy nejcennější část diplomové práce. Rozhodně doporučuji v odevzdávané verzi diplomové práce nahradit CD se slovníkem diskem, které obsahuje pouze text práce: jednak to vyžaduje pokyn děkana, jednak je zbytečné vystavovat vypůjčovatele pokušení. Další desiderata, která mne při čtení práce napadla:

- doplnit zkratky užívané v textu (e.g. MED, DOE), většinou se kladou za obsah
- možná by neškodilo, kdyby byl text na konci doplněn stručným glosářem technických výrazů z počítačové oblasti (zpracování jazyka); nedá se předpokládat, že zejména starší generace lingvistů je příliš zběhlá v pojmech jako XML ap.

- hlouběji objasnit smysl a způsoby využití sémantického značkování u Bosworth-Tollerova slovníku (něco je naznačeno na str. 78, ale možná podrobnější představa o využití by mohla pomoci i při samotném koncipování systému tagů)
- morphological analyzer – nepovede snaha o systematické doplnění lemmat a flektivních tvarů u všech staroanglických slov k riziku, že bude v některých případech nutno vytvářet nedoložené tvary (samozřejmě analogií s příslušným paradigmatem), u nichž není záruka, že odpovídají historické jazykové realitě? (Odpověď je částečně antcipována na str. 81)

Pokud mám nějaké připomínky, týkají se jazykové stránky. Přestože se úroveň angličtiny diplomanda po mém soudu výrazně zlepšila a dá se říci, že je obecně velice dobrá, přesto – zvláště v kontrastu s úrovní obsahové stránky – „it still leaves something to be desired“. Týká se to především užívání členů, ale i lexikálních a gramatických drobností, které působí rušivě v textu, který je jinak formálně velmi pěkně zpracován. Některé příklady:

- 9/2 – lest (= aby ne) > unless
- 12/1 – the author professes ... his later abandonment of the principle > admits/confesses to his subsequent
- 22/3 – some possible fabricated > possibly
- 28/2 – followed an etymological order > were ordered etymologically/arranged in etymological order
- 33 note 23 – all the world's writing systems > all of the world's
- 34/1 – whatever progress has ... scholarship made > whatever progress... scholarship has made
- 35/1 – predetermines whether the dictionary be semasiological > will be
- 35/2 – compile as exhaustive wordlist as possible > a wordlist
  - we will see > can see
- 36/1 – lest the user be unable to locate ??? enabling the user to locate
- 50 – independent from > of
- 54/2 – may have fewer troubles finding > difficulties
- 59 – out of question > out of the question
- 60/3 – ponder the contribution > assess?
  - only a small space for > only little room/space for
- 77 – microstructural analysis > microstructure analysis
- 85/2 – in a greater detail > in greater detail (srov. much/more detail)
- 92 (oddíl 10.3) – visually the entry ... allow you ... to copy/paste them > paste
- 93/ř. 2 – hinting “ENTER” > hitting ?

Ad aplikace/slovník: Obecně považuji zpracování slovníku (formu i funkce, které nabízí) za velmi šťastné a sympatické a myslím, že bude obrovským přínosem pro mediavelistickou obec. V krátkém čase, který jsem měl k jeho otestování, jsem se dopracoval k těmto postřehům:

- dvě zásadní připomínka (jev častý u elektronických slovníků, offline i online): (1) chybí „front matter“, tzn. úvod autorů, jejich popis slovníku a hesel, především ale soupis zkratk a značek, kterých je v textu požehnaně a většinou nerozumím. To zahrnuje i soupis a zkratky excerpovaných textů; (2) součástí aplikace by měla být i instrukce, jak slovník používat !
- trochu postrádám možnost „scrolling“ u reprodukováných stránek slovníku (image), konsektivní prohlížení stránek je jinak možné pouze procházením za sebou jdoucích záhlaví v hesláři
- chybí instrukce o kopírování z textu hesel (funkci „copy“ pod File na liště nenajdeme), předpokládá se (možná příliš optimisticky), že každý uživatel bude znát příkazy Ctrl-C a Ctrl-V.
- trochu mám problémy s wildcard search, pro wordlist tato funkce, zdá se, není zavedena; ve fulltext search vedl konkrétní dotaz (e?gel) k 5 odpovědím (po dlouhém hledání, zakončeném hláškou „Out of memory“), z nichž ani jedna nebyla kupodivu „engel“, ale např. heáfod-hrægel (!?).
- škoda, že aplikace neuvádí „number of hits“, což by bylo leckdy užitečné
- v textu slovníkových hesel jsem narazil na značky, které vypadají jako mark-up language signs a patrně by tam neměly být (konkrétně jde o heslo ge-wunian): @bt\_d0460, @bt\_b0472

**Závěr:** Předložená práce je specifická v tom, že nutí uplatňovat jakýsi „double standard“. Na jedné straně jde o magisterskou diplomovou práci, na druhé straně její konkrétní výsledek je funkční, de facto profesionální aplikace, který svádí k porovnání s komerčními produkty i odbornými vědeckými výstupy. Oba „standards“ splňuje a lze říci, že úroveň zpracování, množstvím investované práce a hodnotou výstupu pro obor je srovnatelná s rigorózními pracemi (ostatně výzkumný projekt, který nastiňuje, odpovídá zadání pro doktorskou práci). Z těchto důvodů doporučuji práci k obhajobě a předběžně ji hodnotím známkou výborně.