

Posudek disertační práce Ondřeje Tichého *Nástroj na tvaroslovnou analýzu staré angličtiny / Morphological Analyser of Old English*

Disertační práce Ondřeje Tichého se zabývá tématem, které je v současné fázi vývoje historicky orientované lingvistiky (ne jen anglistiky) mimořádně živé: poté, co historický výzkum jazyka dostal před více než dvaceti lety první významný impuls ze strany korpusové lingvistiky ve formě možnosti operativně zkoumat starší vývojové fáze jazyků na nebývalé velkých datových souborech, tj. po etapě charakterizované především samotnou výstavbou historických a diachronních korpusů, přichází období orientované směrem, jímž je zaměřena i předložená disertace – směrem ke zvýšení výzkumného potenciálu historických dat prostřednictvím dodané lingvistické informace, především lemmatizace a morfologického značkování.

Autor – stejně jako každý, kdo se touto problematikou v současné době zabývá – má výhodu v tom, že se při své práci může opřít o zkušenosti získané při lemmatizaci a morfologickém značkování korpusů současných jazyků, v jejichž případě má dodávání lingvistických informací k textovým datům už poměrně dlouhou tradici. V souladu s tím Ondřej Tichý pracuje s širokým okruhem literatury a internetových informací nejen z oboru lingvistiky (především historické a korpusové), ale i z oblasti informatiky. Tyto zdroje autor přijímá jako cennou inspiraci, současně k nim však přistupuje s náležitou kritičností a formuje tak na jejich základě vlastní metodu. Podobně využívá i částečné inspirace, kterou představují některé dílčí projekty, které byly v minulosti realizovány přímo se zaměřením na starou angličtinu a jejichž přehled podává na s. 14–17.

Základní koncepce autorova morfologického analyzátoru vychází z analýzy pravopisné stránky staroanglických textů a staroanglického tvarosloví, kam se koncentruje většina problémů, s nimiž se musí vypořádat každé počítačové zpracování (respektive vůbec každý rozbor či popis) staré angličtiny. Ondřej Tichý podává stručný zasvěcený přehled obou oblastí a identifikuje hlavní charakteristiky, které musí koncepce analyzátoru reflektovat (jde zejména o rozsáhlou intreflexi a variabilitu i rozkolísanost pravopisu). S přehledem a systematickostí řeší autor i konkrétní problémy, s nimiž je tvorba analyzátoru spojena v oblasti dostupných dat pro slovníkovou databázi. Z naznačených autorových analýz, ale nevyhnutelně i ze skutečností, jako je neexistence vhodně anotovaného staroanglického korpusu nebo neexistence úplného popisu morfematiky staré angličtiny, vyplývá výsledná koncepce analyzátoru. Jeho principem se nestává morfematická analýza (na rozdíl např. od analyzátoru novočeských textů, vůči němuž autor svůj přístup vymezuje), nýbrž porovnávání interpretovaného slovního tvaru s položkami databáze vygenerovaných slovních tvarů.

Autorův postup považuji za plně zdůvodněný a jeho výklad za jasný a přesvědčivý. K předložené disertaci nemám žádné podstatné připomínky, uvádím jen jedno formální upozornění: Autor zřejmě užívá výrazu *analyser* v užším (obvyklém) i v podstatně širším smyslu (s. 51; v tomto smyslu jsou součástí „analyser’s operation“ nejen sama analýza konkrétních slovních tvarů, ale i příprava vstupních dat a vytváření slovníku). Toto obecnější užití by podle mého názoru bylo vhodné nahradit jiným výrazem.

V rámci rozpravy nad předloženou disertací bych uvítal, kdyby se autor vyjádřil k následujícím tématům:

- Jak velké je riziko přegenerování při vytváření paradigmat pro databázi staroanglických slovních tvarů a jejich podob? Počítá se v budoucnu s průběžnou validací (a případnou redukcí) generovaných tvarů a podob na základě textů?

- Pedagogická motivace, resp. zřetel k pedagogickému využití nástroje na morfologickou analýzu, uváděný v oddílu 1.2. (Aims and Motivations), tvoří významnou součást autorova projektu. V kterých rysech pedagogický aspekt nejvýrazněji ovlivnil (popřípadě na základě současných úvah ještě v budoucnu ovlivní) výslednou podobu aplikace?
- Variabilita a rozkolísanost písařského úzu patří k nezávažnějším problémům rekonstrukce fonologických podob starších fází jazyků – a výsledně také k nezávažnějším problémům morfologické analýzy i výstavby historických a diachronních korpusů. Existují podle autorova názoru reálné možnosti, jak tento problém, v případě staré angličtiny mimořádně komplikovaný, minimalizovat pomocí standardizované transkripce (popřípadě transkripce propojené s transliterací a/nebo s digitalizovanými staršími texty)?

Předložená disertace Ondřeje Tichého je zpracována s velkým přehledem i nadhledem jak v oblasti lingvistické problematiky související se zvoleným tématem, tak v oblasti informatiky, a navíc vyniká schopností přesně a výstižně formulovat i značně komplikované problémy. Jako celek nejen představuje metodologický přínos k počítačovému zpracování staré angličtiny, ale je cílena i ke konkrétní aplikaci, která bezpochyby najde uplatnění v pedagogické praxi i při výzkumu nejstarších doložených vývojových fází anglického jazyka a umožní podstatné rozšíření informačního potenciálu historických anglických korpusů. Jak je zřejmé, jde podle mého názoru o práci, která rozhodně splňuje požadavky kladené na doktorské disertace, a doporučuji ji tedy bez výhrad jako vhodný podklad k obhajobě i k udělení hodnosti Ph.D.

V Praze 7. května 2014

prof. dr. Karel Kučera, CSc.,
Ústav Českého národního korpusu
Filozofické fakulty UK v Praze