



Leško Marek, The individual textual profile: a corpus based study of idiolect /

Individuálna textový profil: korpusově založený výzkum idiolektu

Magisterská diplomová práce, Ústav anglického jazyka a didaktiky, FF UK Praha, září 2012

Posudek oponenta

Hlavním cílem práce Marka Leška je vymezit charakteristiky idiolektu konkrétní osoby (prezidenta Baracka Obamy), tedy stanovit její individuální textový profil, pomocí korpusové metody, tedy především analýzy klíčových slov (keyword analysis), na základě oficiálních televizních debat v rámci předvolební prezidentské kampaně. Základem analýzy je předpoklad, že klíčová slova poskytují zásadní informaci o obsahu a stylu textu.

Studie má logickou strukturu, nejprve seznamuje čtenáře s důvody, proč zkoumat idiolektu a proč korpusově. V druhé kapitole vymezuje stručně cíle, v třetí pak velmi systematicky metodologii práci. Její popis zároveň osvětluje teoretické pozadí (které obvykle předchází a tvoří samostatnou část). Probírá klíčová slova a související pojem, klasifikaci a identifikaci klíčových slov a další metody, které analýzu klíčových slov doplňují. Ve 4. kapitole označené Summary (autorovo oblíbené slovo, patrně součást jeho idiolektu) of the Research Data představuje oba korpusy, Obamův a referenční, a jejich kvantitativní parametry. Kupodivu přehled zjištěných klíčových slov (Obama corpus), který by sem podle mého názoru patřil, je uveden až v následující kapitole 5, nazvané The analysis of Keywords in the Obama corpus. Závěrečná 6. kapitola je trochu překvapivě označena jako Summary of Findings. Překvapivě ze dvou důvodů. Za prvé vnáší určitý nový pohled na data (kategorizace rysů Obamova idiolektu, s. 68), tzn. nejde o prosté shrnutí toho, což už bylo řečeno, a spadá tedy spíše do rozboru a diskuse výsledků. Za druhé stojí na místě, kde bychom očekávali Conclusions, ale „conclusions“ (rekapitulující smysl a zjištění celé studie) to v pravém slova smyslu nejsou. Práce v důsledku toho působí neukončeně.

Studie je v každém případě zajímavá a poučná. Myslím, že první část jejího názvu, individuální textový profil, je přesnější než jeho druhá část, výzkum idiolektu. Mám za to, že určovat idiolekt na základě 37 klíčových slov je trochu málo a že jde spíše o metodologickou sondu toho, jak by bylo možno postupovat. Výzkum idiolektu by vyžadoval analýzu jazyka dané osoby v různých typech textů a v různých situacích, tedy na podstatně větším materiálu. To vzhledem ke stanovenému rozsahu diplomové práce nelze očekávat, takto jde z větší části spíše o obsahově-stylistickou analýzu určitého typu textu. Např. tematické indikátory (s. 68) mohou sice „represent a part of [Obama's] profile“, ale může to být aspekt profilu záměrně naordinovaný jeho poradci pro tato konkrétní televizní vystoupení. To ovšem hodnotu práce nesnižuje.

Při čtení jsem narazil na několik věcí, které stojí za komentář. Z části 1.2 vyplývá, že autor nerozlišuje mezi „corpus-based“ a „corpus-driven“. V literatuře je ale dnes poměrně rozšířené povědomí, v čem se oba přístupy liší. Srov. heslo v A Glossary of Computer Linguistics (Baker, Hardie, McEnery, 2006), s kterým autor ostatně pracuje: **corpus-based** Tognini-Bonelli (2001) makes a useful distinction between corpus-based and corpus-driven investigations. The former uses a **corpus** as a source of examples to check researcher intuition or to examine the **frequency** and/or plausibility of the language contained within a smaller **data** set. The researcher does not question pre-existing traditional descriptive units and categories. A corpus-driven analysis is a more inductive process: the corpus itself is the data and the patterns in it are noted as a way of expressing regularities (and exceptions) in language. A corpus-driven analysis tends to only use minimal theoretical presuppositions about grammatical structure.



V části 3.1 v rámci popisu keywords by stálo za to zmínit i rozdíl činěný mezi *keywords* a *key words* (Scott a Tribble užívají v názvu své publikace dvouslovnou formu). Alespoň Jukka Tyrkko (s. 80, Bondi-Scott, *Keyness in Texts*) tvrdí, že je „conceptual difference between *keywords* and *key words*, the former meaning lexical items possessing statistical keyness and the latter items perceived by human readers as key“.

V textu se objevují dobré nápady, např. zavedení dvou podkategorií lexikálních slov (s. 12, 28) nebo myšlenka pracovat místo s colligation s clusters (s. 21). Nicméně jaká jsou kritéria pro rozlišení tématických a stylových indikátorů – proč je např. slovo *advisers* stylovým a nikoli obsahovým indikátorem?

K textu bych měl tyto formální výtky: závorka s bibliografickým údajem (Mollin 2009: 371) se nedává za větu, ale do věty, k níž se vztahu (viz s. 9 a jinde); postrádám seznam zkratk na začátku, řešení poznámkou pod čarou (s. 27) není příliš dobré (co zkratka p.m.w., předpokládám, že jde o „per million words“?); bibliografie a odkazový aparát mohly být zpracovány pečlivěji (viz dílčí připomínky).

Dílčí připomínky:

- s. 10: Coniam (2004) neuvedeno v bibliografii, ač zmiňováno v textu
- s. 12: poznámka 8 - nejasné, proč není glosář standardně uváděn pod jménem autorů, chápu, že Glossary je kratší, ale stačilo by Baker et al.
- s. 19: *lexems*, i.e. canonical forms – toto není šťastná formulace (podobně v Glossary, „The base form of a word.“), lexém není kanonická nebo základní forma, ta pouze lexém reprezentuje
- s. 48: označení „functional style indicator“ pro gramatická slova je bez bližšího vysvětlení trochu matoucí v kontextu domácího termínu „funkční styly“
- s. 22: Hyland (2008) chybí v bibliografii
- s. 67: začátek první věty potřebuje upravit
- s. 80: Baker, Hardie, McEnery (2006) – publikace mimo abecedy, neúplný název: A Glossary
- s. 80: Leed, D. (2008) neúplná bibliografická informace – neúplný název článku (*Corpora and discourse analysis: new ways of doing old things*), chybí editoři publikace (Bhattia, Flowerdew, Jones) a chybí stránky (86-99)
- s. 81: publikace Bondi-Scott měla být uvedena v bibliografii i jako samostatná položka, navíc způsob, jakým je uvedena pod Scott (2010) je nestandardní

Hodnocení: Práce je, až na uvedené nedostatky, dobře zvládnutá po strukturní, obsahové i formální stránce. Autor se v problematice zjevně spolehlivě orientuje a dokáže si poradit s úkolem, který vyžaduje nejen znalosti, ale i schopnost je kreativně uplatnit. Také po jazykové stránce je práce pěkná a je psána úsporným jazykem s minimem chyb.

Závěr: Práci považuji za zdařilou, autor při plnění zadání prokázal znalosti i samostatnost. Škoda, že některé věci nebyly zcela dotaženy do konce. Práci doporučuji k obhajobě a předběžně navrhuji hodnocení „velmi dobrý“ až „výborný“ v závislosti na obhajobě.

V Praze dne 3.9. 2012

prof. PhDr. Aleš Klégr