

Bc. Alena Kvítková

Textual patterns and idiolect: a corpus-assisted study of individual textual profiles of Hillary Clinton and Donald Trump. Textové vzorce a idiolekt: korpusová studie individuálních textových profilů Hillary Clintonové a Donalda Trumpa. Diplomová práce. ÚAJD FF UK. 2018.

Posudek vedoucí práce

Předložená diplomová práce porovnává jazyk vystoupení dvou kandidátů na prezidenta Spojených států amerických, Hillary Clintonové a Donalda Trumpa, na pozadí projevů ostatních kandidátů v předvolebních debatách v roce 2016. Využívá metody korpusové lingvistiky: zkoumá klíčová slova, jejich kolokace a klíčové shluky slov. Na základě detailní analýzy se autorce podařilo odhalit slova a slovní spojení, která jsou příznačná pro každého ze sledovaných kandidátů, nebo jichž kandidáti užívají odlišným způsobem.

Teoreticky vychází práce z přístupu označovaného jako CADS (Corpus-assisted discourse analysis), který kombinuje analýzu diskurzu a metody korpusové lingvistiky. Tímto přístupem jsou dány metody analýzy – autorka postupuje od kvantitativních dat (práce čerpá materiál z korpusů předvolebních debat, které autorka sestavila) k jejich pečlivému kvalitativnímu popisu, který se opírá o znalost kontextu, jazykového i situačního. Práce je přehledně strukturovaná. Postupně jsou popsány charakteristické rysy promluv obou kandidátů, každá tato část obsahuje shrnutí poznatků. Poznátky o vyjadřování obou kandidátů jsou pak porovnány a v závěru autorka hodnotí také vhodnost použitých metod.

Použité zdroje jsou dobře dokumentovány v příloze práce. Pro přehlednost by bylo vhodné v tabulce 1 na str. 26 uvést také, kolik různých mluvčích obsahují referenční korpusy.

Analýza je provedena velmi pečlivě, autorka dospěla k originálním, zajímavým a dobře podloženým výsledkům. Moje připomínky směřují proto převážně k teoretickým východiskům práce a přesnosti vymezení jednotlivých pojmů a metodologických postupů.

- Otázkou je, do jaké míry lze vystoupení kandidátů v předvolebních debatách pokládat za nepřipravená („unprepared debates“, str. 9, 26). V té souvislosti bych byla opatrnější ve formulacích, které popisuje zjištěné rozdíly mezi promluvami Clintonové a Trumpa jako rozdíly v idiolektu. Jak vyplývá z definice idiolektu (str. 24), zkoumaný materiál zachycuje jen jistou výseč – obvyklý způsob vyjadřování obou kandidátů v politických debatách v roce 2016. Nezohledňuje další parametry idiolektu („age, gender, social class, personal traits“). Přesnější popis toho, co práce sleduje, je uveden na str. 26: „a... authentic picture of Clinton’s and Trump’s speech patterns in a political speech situation“. Je, myslím, třeba také rozlišovat mezi tím, jak mluvčí formuluje své postoje, a tím, co si myslí: klíčová slova neukazují, že „Clinton thinks things through“ (str. 55), ale spíše to, jak Clintonová věci prezentuje, jaká slova volí.

- V teoretické části by autorka měla být opatrnější v generalizacích. Nemyslím si např., že příčinou popularity korpusové lingvistiky v devadesátých letech je „the widespread popularity of computers“ (str. 13).
- Terminologii je třeba používat přesně:
 - autorka zaměnila termíny inter- a intra-textual (str. 16);
 - zdá se, že autorka nerozlišuje dostatečně mezi frekvencí a statistickou signifikancí („Corpus linguistics searches for what is statistically considered a frequent pattern and regards such a pattern as significant for a further analysis.“); také ve vymezení klíčových slov hrají roli statistické metody (log-likelihood, chi-square) a ne jen frekvence (str. 19); log-likelihood a chi-square nejsou metody, které zjišťují frekvenci slov v korpusu (str. 21);
 - „concordance“ (str. 16) je způsob zobrazení výsledků, nástroj je concordancer, pro odhalení kolokací existují jiné nástroje než řazení konkordančních řádků;
 - clusters/lexical bundles nejsou totéž jako n-grams (str. 18), vymezení „clusters“ se mísí s popisem technických možností použitého softwaru, popis forem „clusters“ je zjednodušující a neúplný, chybí odkaz na zdroj formální klasifikace, chybí zmínka o funkční klasifikaci „lexical bundles“;
 - velikost a složení referenčního korpusu jsou často diskutované otázky – v práci chybí citace k tvrzení, že „it is customary to choose a reference corpus of at least the same size as the target one“;
 - v jaké vzdálenosti od analyzovaného slova (collocation span) se vyhledávaly kolokace (např. v tabulce 5 na str. 35 je vzdálenost 2 slova vlevo a vpravo od zkoumaného slova, platí to obecně?); proč obsahují kolokace víceslovné jednotky i jednotlivá slova (*of the United States, the*)? Je obecně stanoven frekvenční limit pro zahrnutí kolokace do analýzy? Např. *strategy* se objevuje jako kolokát slova *comprehensive* dvakrát, „and thus cannot be considered a collocate“ (str. 34) – to je problematické hlavně v souvislosti se zvolenou metodou identifikace kolokací, MI-score, protože tato metoda obvykle identifikuje jako kolokace spíše méně frekventovaná slova s omezenou kolokabilitou. Proč byla zvolena metoda MI-score?
 - Jak se rozhodovalo, ve kterých případech („in certain cases...“, str. 28) budou prováděna srovnání s referenčním korpusem?
 - Co je „sufficient frequency pro keywords“? – rozdíly ve frekvenci klíčových slov s nejvyššími hodnotami LL jsou značné (1638 – 36 u Trumpa).

Analyzováno je 50 klíčových slov s nejvyšší hodnotou „keyness“ z každého subkorpusu (Clintonové a Trumpa). Autorka se přiklání ke klasifikaci klíčových slov na vlastní jména, „aboutness words“ a indikátory stylu, lexikální a gramatické. Dobře zároveň ukazuje, že zařazení do jednotlivých kategorií může být problematické (proč není např. *voted* (str. 39) „aboutness word“?). Přesvědčivě ukazuje autorka frazeologii spojenou s *people - millions and millions of people, people that I deal with, that I talk to* (str. 62) – je vidět, jak *people* funguje jako lexikální indikátor stylu, podobně *way* v *by the way*.

Autorka prozkoumala podrobně kolokace sledovaných slov, u některých sledovala, v jakých shlcích se vyskytují. Analýza je pečlivá a odhaluje jinak špatně postřehnutelné rozdíly. Pěkný je např. popis „grammatical indicators of style“ u Clintonové a interpretace užívání zájmena *we* (str. 43-44) – zajímavé je pak porovnání s užitím téhož zájmena v Trumpových prosloveh (ve fixních spojeních *we have no idea, we have no choice* (str. 72)). U analýzy klíčových slov není vždy třeba analyzovat jednotlivě všechna identifikovaná slova, je možné je např. sdružovat do větších celků nebo neprovádět vždy všechny kroky – např. u předložek se jeví n-gramy jako vhodnější než kolokace.

Je třeba dát pozor na interpretaci, která nevyplývá přímo z výsledků. Zdá se např., že z uvedených dokladů (str. 37) nevyplývá přímo, že Clintonová je „a sympathetic and supportive person“.

„Clusters“ by bylo vhodné sdružít do „patterns“ – dobře je to vidět v tabulce 22 na str. 81 – *let me just tell you*. Lépe je pak vidět jejich kontaktoá a textově-organizační funkce (v souladu s dalšími jazykovými prostředky v projevech obou kandidátů, např. *you know, OK*).

Zajímavé je Trumpovo užívání zájmena *me* (str. 73), i když *I* mezi klíčovými slovy není. Je *me* kolokát slova *nobody*? (str. 87)

Teoretická část zmiňuje jako jednu z metod zkoumání textového profilu type/token ratio. Byla tato metoda použita?

Práce má požadovanou formální a jazykovou úroveň, jazykových chyb je minimum (*comprehensive vs. comprehensible*, str. 86).

Použitá literatura je rozsáhlá a pro zkoumané téma relevantní. (Jen titul Hunston, S. (2000) *Corpus Linguistics*. Birmingham: University of Birmingham nemohu v bibliografii S. Hunston najít.)

Závěr

Předložená práce splňuje požadavky kladené na diplomové práce. Zatímco v teoretické a metodologické části lze najít některé nepřesnosti a nedostatky, analýza a interpretace shromážděných dat je pečlivá a přesvědčivá. Navrhuji proto předběžně hodnocení velmi dobře až výborně.

V Praze 1. 9. 2018

doc. Markéta Malá, Ph.D.