

Semantic categories of the lemmata *Arab* and *Muslim* in British English newspapers: A corpus-based discourse analysis

Dalea Awada

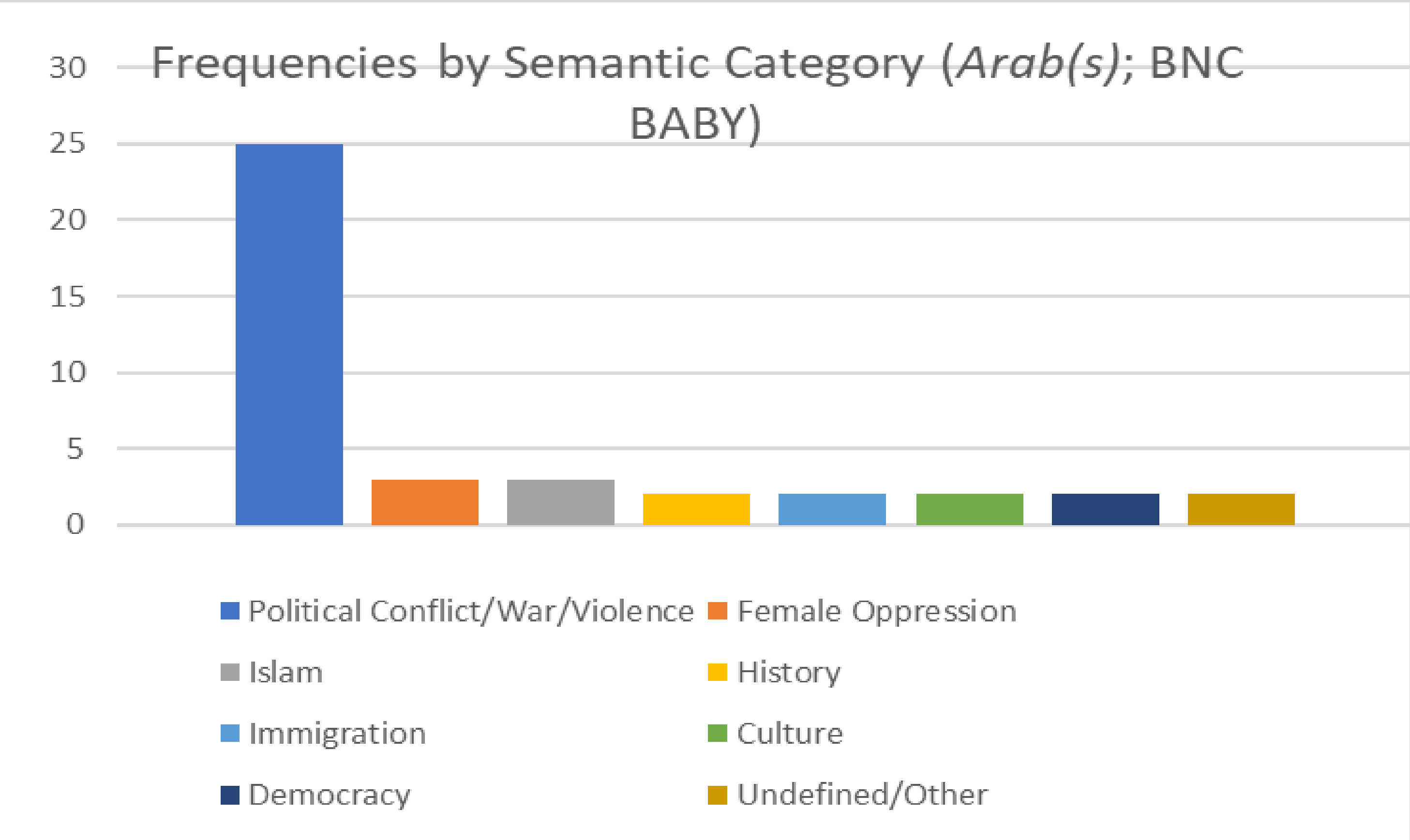
Introduction

Critical Discourse Analysis (CDA) examines linguistic phenomena by connecting them to questions of ideology and power, viewing discourses as a “form of social practise” (Fairclough & Wodac 1997 through Cook 2011). It is, however, also a rather broad “study of everything” (Cook 2011) that has to actively be applied to a category like linguistic analysis. Corpus linguistics enable the discipline of Applied Linguistics Discourse Analysis to add a quantitative dimension to its research, serving as an important tool. This study aims to investigate the usage of the lemmata *Arab* and *Muslim* in British English newspapers as available through the BNC. It does so by categorizing their uses semantically, sorting these semantic categories then in regard to their frequency, to reveal their cognitive associations in British English newspapers, and hence, in social discourse.

Data and methodology

It is still an open question how many samples would have to be (manually) analysed to be representative of some sort. The query [word=“Arabs?”]::match.text_derived_type=“newspaper” finds 268 matches; [word=“Muslims?”] as many as 478 matches. A manual categorization of these matches as a single person seems rather unrealistic; **the following results therefore stem from the BNC-Baby that found 36 matches just to demonstrate the project.** Other open questions and problems remain. The data of the BNC is 26 years old, making it possibly outdated for such a socially relevant study. Additionally, the restriction to *Arab(s)* and *Muslim(s)* as opposed to, for example *Islam* has to be clarified and possible orthographic variations like *Moslem(s)* considered. A chi-squared test will have to be applied to the variables to evaluate their actual relevance. The semantic categorization should eventually be used to create a comprehensive list of the different semantic categories for the lemmata in British English newspapers, as well as answering whether *Arab* and *Muslim* are being used as clearly distinctive entities.

Results



Conclusion

-> mostly contexts of political conflict, war, violence, BUT: newspapers inform about political activities around the world -
> comparison to eg. *Irish* in newspaper corpus interesting to determine whether Racial/Islamophobic agenda at play (Northern Ireland conflict during same time as Arab-Israeli conflict that is being referred to in most cases)
-> sample too small to conclude anything
-> a comparison of *Arab(s)* and/or *Muslim(s)* to *Northern Irish*, *Irish*, *Catholic*, *Protestant* might be more meaningful to determine whether political conflict in the Southwest Asian/Northern African region during the time of the compilation of the corpus was especially overrepresented in British media discourse

References

Guy Cook (2011). *Discourse Analysis*. The Routledge Handbook of Applied Linguistics. Edt: James Simpson. Taylor & Francis, 431 – 444.
Bandar Al-Hejin (2015). *Covering Muslim women: Semantic macrostructures in BBC News*. Discourse & Communication 2015, Vol. 9(1), 19 – 46.