

Corpus-assisted discourse analysis of family legislation changes in Slovene media

Senja Pollak (Jozef Stefan Institute, Slovenia) and Ana Marija Sobočan
(Faculty of Social Work, University of Ljubljana, Slovenia)

1 Introduction

Corpus linguistics and (critical) discourse analysis have been, with the aim of uncovering and questioning the representations of different social groups or phenomena, successfully combined in several studies (e.g. Krishnamurthy, 1996; Stubbs, 1998; Teubert, 2000; Baker, 2006; Mautner, 2009; Baker & Lavon, 2015). In our contribution we present a corpus-based discourse analysis of media coverage of the Slovene Act Amending the Marriage and Family Relations Act (hereby 'the Act'). We analyse and compare the discourses in three online and printed newspapers, *Demokracija*, *Dnevnik* and *Mladina* that were chosen based on their diverse standpoints towards the topic of interest. The selection of terms for analysis is based on terminology extraction, which by a statistical measure for comparing the terms of a domain corpus to the terms of the reference corpus, identifies the single- or multi-word terms that are characteristic for a selected corpus. In the first study (see Section 4) we analyse the concordances of the most frequent multi-word terms of the entire corpus, while in the second study (Section 5) we analyse the terms that are distinctively characteristic for each medium.

2 Corpus description

The corpus consists of 69 texts from the three online and printed media. The three are generally recognized as ideologically oriented as right-wing / conservative (*Demokracija*), centre (*Dnevnik*) and left-wing / progressive (*Mladina*). The texts were selected based on the relevance to the topic¹ and limited by the studied period spanning from the 15th of November 2015, a few weeks prior to the referendum on the Act and to the 31st of December 2015, a week after the referendum.

	<i>Mladina</i>	<i>Dnevnik</i>	<i>Demokracija</i>	All
# articles	10	27	32	69
# sentences	493	876	1836	3205
# word tokens	9,492	17,063	36,193	62,748

Table 1: Corpus statistics.

¹ Term ZZDR, the abbreviation for the Act, was used as a query in the archives of the three newspapers. The returned search results (articles were downloaded in October and November 2016) were manually filtered to confirm the relevance to the topic.)

In total 69 articles corresponded to the defined criteria (see corpus statistics in Table 1). The texts were pre-processed with the ToTrTaLe tool (Erjavec, 2011) including sentence and word tokenisation, lemmatisation and annotation with morpho-syntactic descriptions. The text was uploaded to the SketchEngine corpus linguistic tool and concordancer (Kilgarriff et al., 2014).

3 Media discourses on family legislation changes (Act Amending the Marriage and Family Relations Act)

The study focuses on media discourses related to the 'Act amending the Marriage and Family Relations Act', which aimed at changing the then current family legislation. The first attempt for changes happened in 2009, when the Ministry of work, family, social affairs and equal opportunities proposed the 'Family Code', which was supposed to substitute the Marriage and Family Relations Act originating in the 1970s (hence outdated). The Code was adopted by the Parliament, but rejected with a referendum in 2012. Already then, there was much public attention and media focus on the proposed legislative changes, and the interest of the public was mainly sparked by the legislation introducing pluralisation of families and equalising biological (parents defined by a genetic link to children) and social parenthood (parents defined by a social, care-based link to children), by equalising the rights and status of same-sex and different-sex partnerships and parental rights. In December 2014 the 'Act amending the Marriage and Family Relations Act' was filed in parliamentary procedure and was adopted, implementing a symbolic and legal equality for families of partners of same or different gender. As the Code, also the Act was rejected with a referendum (on 20th December 2015).

Other authors have already analysed the discourses related to family legislation in Slovenia, most notably Vezovnik (2015), who published an exhaustive analysis of argumentation schemes and structures related to the 'Family Code', while Kuhar (2015) published an analysis of populistic strategies of the opponents of the 'Family Code'. To the best of our knowledge, we are the first to address the context of the last referendum and the 'Act amending the Marriage and Family Relations Act' and to analyse the discourses with combined quantitative and qualitative methods (i.e. as a corpus assisted discourse study). Studying particularly the discourses on (changes of) family legislation proves to be an interesting and relevant endeavour, as it can be observed that numerous societal norms and values are reflected, reproduced or refuted in this framework.

4 Discourse analysis of top ten terms

Using the online term extraction workflow (Pollak et al., 2012), based on a comparison of noun phrases in domain and reference corpus, we extracted a list of terms ranked by their termhood score. In our case, the domain corpus is the corpus of 69 articles from the three selected media, while the reference corpus is the 100-million word corpus Kres (Logar 2012), containing Slovene texts of various genres, from daily newspapers, magazines, books (fiction, non-fiction, textbooks), web pages, and similar, with a balanced genre structure. We analysed the top 10 extracted bigrams, presented in Table 2. Since the corpus is imbalanced, the largest subcorpus

(*Demokracija*) predominantly impacts the term extraction. For example the term *theory of gender* appears nearly exclusively (72 out of 75 occurrences) in *Demokracija*.

Bigram rank by termhood value	Term	Term translation	Frequency
1	zakonska zveza	marriage	238
2	novela zakona	amending act	137
3	družinsko razmerje	family relationship	96
4	teorija spola	theory of gender	75
5	družinski zakonik	family code	58
6	človekova pravica	human rights	72
7	istospolna skupnost	same-sex partnership	40
8	istospolni par	same-sex couple	38
9	novela zzzdr	act amending the marriage and family relations act	32
10	ustavno sodišče	constitutional court	53

Table 2: 10 top-ranked bigram terms.

We analysed all corpus examples that contained the selected terms and interpreted them from the perspective of critical discourse analysis. We identified six discursive frames:

- family as 'natural' and heterosexual relationship as the norm
- theory of gender
- dissolution of 'Family' and values, a threat to the social order
- hegemony of the ruling elites, manipulations and state of mind
- human rights, women's rights and children's rights
- civil rights, nation building and nationalism

The analysis shows that some terms (e.g., '*theory of gender*') and discourses significantly mark only the texts in one of the media (*Demokracija*) and that certain terms (e.g. *human rights*) are used in conflicting discourses (i.e. both proponents and opponents of marriage equality are portrayed as human rights violators). Significantly, the homophobic, racist, and nationalist discourses of opponents of family legislation changes are producing moral and sexual panics (moral panics specifically related to issues of sexuality and reproduction).

5 Discourse analysis of contrasting terms

As in the selection presented in Section 4, we used the term extraction workflow (Pollak et al., 2012), but here we performed the term extraction for each subcorpus (medium) separately. Since we were interested in contrasting terms, characteristic for each of the subcorpora, we compared the lists of 150 top-ranked terms of each subcorpus and kept for analysis only the terms (n-gram noun phrases) that were listed in only in one of the three term lists. We grouped the contrasting terms in different semantic fields:

- partnership relationships
- legislation and referendum
- social semantic fields: gender, children, upbringing, nationality, human rights, values
- stakeholders / agents
- meta-discourse

We compared the differences in framing the referendum of three media. Different terms of the three media are characteristic for each field. For example, the differentiating terms selected from *Demokracija* under stakeholders category contain family members (husband, father, grandparent, mother, grandchild, wife, mom) that are presented in heteronormative matrices of an essentialist conception of family.

6 Conclusions

The corpus-driven analysis of the discourses of the three media has revealed the discourses and differences in reporting in three online and printed media reporting on the family legislation changes and the referendum. Standing out are the discourses propagating homophobia, islamophobia, anti-feminism, etc. that were the dominating discourses especially in the conservative medium *Demokracija*, and contributed to creating an atmosphere of danger and fear. By using term extraction, we limited our analysis on characteristic nouns (or noun phrases with a noun as a head word), but in future work we will extend the analysis to other parts-of-speech.

References

- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, P. & Lavon, E. (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse and Communication*, 9(2), 401–422.
- Erjavec, T. (2011). Automatic linguistic annotation of historical language: ToTrTaLe and XIX century Slovene. In K. Zervanou & P. Lendvai (Eds.), *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (pp. 33–38). Stroudsburg, PA: ACL.
- Kilgarriff, A., Baisa, V., Bušta, J. Jakubíček, M. Kovař, V., Michelfeit, J., ...Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36.

- Krishnamurthy, R. (1996). Ethnic, Racial and tribal: The language of racism? In R. Caldas-Coulthard & M. Coulthard (Eds.), *Texts and Practices: Readings in Critical Discourse Analysis* (pp. 129–149). London: Routledge.
- Kuhar, R. (2015). Konec je sveta, kakršnega poznamo: Populistične strategije nasprotnikov Družinskega zakonika. *Časopis za kritiko znanosti*, 43(260), 118–132.
- Mautner, G. (2009). Checks and balances: How corpus linguistics can contribute to CDA. In R. Wodak & M. Meyer (Eds.), *Methods of Critical Discourse Analysis* (pp. 122–143). London: Sage.
- Logar Berginc, N., Grčar, M., Brakus, M., Erjavec, T., Arhar Holdt Š. & Krek, S. (2012): *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in cckRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina; Faculty of Social Sciences.
- Pollak, S., Vavpetič, A. Kranjc, J. Lavrač, N. & Vintar, Š. (2012). NLP workflow for online definition extraction from English and Slovene text corpora. In J. Jancsary (Ed.), *Proceedings of the 11th Conference on Natural Language Processing - KONVENS* (pp. 53–60). Wien: ÖGAI.
- Stubbs, M. (1998). German loanwords and cultural stereotypes. *English Today*, 14(1), 19–26.
- Teubert, W. (2000). A province of a federal superstate, ruled by an unelected bureaucracy. Keywords of the Euro-sceptic discourse in Britain. In A. Musolff, C. Good, P. Points & R. Wittlinger (Eds.), *Attitudes Towards Europe: Language in the Unification Process* (pp. 45–88). Aldershot: Ashgate.
- Vezovnik, A. (2015). Slovenski predlog Družinskega zakonika in njegovi politični argumenti. *Teorija in praksa* 52(6), 1191-1211.