

The Parlameter corpus of contemporary Slovene parliamentary proceedings

Nikola Ljubešić,* Darja Fišer,[†]* Tomaž Erjavec,* Filip Dobranič[‡]

*Department of Knowledge Technologies, Jožef Stefan Institute
Jamova cesta 39, 1000 Ljubljana
nikola.ljubesic@ijs.si, tomaz.erjavec@ijs.si

[†]Department of Translation, Faculty of Arts, University of Ljubljana
Aškerčeva cesta 2, 1000 Ljubljana
darja.fiser@ff.uni-lj.si

[‡]Danes je nov dan
Parmova ulica 20, 1000 Ljubljana
filip@danesenovdan.si

Abstract

The paper presents the Parlameter corpus of contemporary Slovene parliamentary proceedings, which currently covers the VIIth mandate of the Slovene Parliament (2014-2018). The Parlameter corpus offers rich speaker metadata (gender, age, education, party affiliation) which boost research in several digital humanities and social sciences disciplines. We analyze the linguistic production paired with the metadata from the perspective of communication and political studies. The corpus architecture allows for regular extensions of the corpus with additional Slovene data, as well as data from other parliaments, starting with Croatian and Bosnian.

1. Introduction

Parliamentary discourse is motivated by a wide range of communicative goals, from position-claiming, persuasion and negotiation to agenda-setting and opinion-building along ideological or party lines. It is characterized by role-based commitments and confrontation and the awareness of a multi-layered audience (Ilie, 2017). The unique content, structure and language of records of parliamentary debates are all factors make them an important object of study in a wide range disciplines in digital humanities and social sciences, such as political science (Van Dijk, 2010), sociology (Cheng, 2015), history (Pančur and Šorn, 2016), discourse analysis (Hirst et al., 2014), sociolinguistics (Rheault et al., 2016), and multilinguality (Bayley, 2004).

Despite the fact that parliamentary discourse has become an increasingly important research topic in various fields of digital humanities and social sciences in the past 50 years (Chester and Bowring, 1962; Franklin and Norton, 1993), it has only recently started to acquire a truly interdisciplinary scope (Bayley, 2004). Recent developments enable cross-fertilization of linguistic studies with other disciplines and in-depth exploration of institutional uses of language, interpersonal behavior patterns, interplay between language-shaped facts, and reality-prompted language ritualization and change (Ihalainen et al., 2016).

With an increasingly decisive role of parliaments and their rapidly changing relations with the public, mass media, executive branch and international organizations, further empirical research and development of integrative analytical tools is necessary in order to achieve a better understanding of parliamentary discourse as well as its wider societal impact, in particular with studies that represent diverse parts of society (women, minorities, marginalized groups) and cross-cultural studies (Hughes et al., 2013).

2. Parliamentary corpora

The most distinguishing characteristic of records of parliamentary debates is that they are essentially transcriptions of spoken language produced in controlled and regulated circumstances. For this reason, they are rich in invaluable (sociodemographic) meta-data. They are also easily available under various Freedom of Information Acts set in place to enable informed participation by the public and to improve effective functioning of democratic systems, making the datasets even more valuable for researchers with heterogeneous backgrounds.

This has motivated a number of national as well as international initiatives (for an overview, see Fišer and Lenardic (2018)) to compile, process and analyze parliamentary corpora. They are available for most countries within the CLARIN ERIC research infrastructure for language resources and technology, with the UK's Hansard Corpus being the largest (1.6 billion tokens) and spanning the longest time period (1803-2005) while corpora from other countries are significantly smaller (most comprise between 10 and 100 million tokens) and cover significantly shorter periods (mostly from the 1970s onwards).

The Slovene parliamentary corpus SlovParl 2.0 (Pančur, 2016) contains minutes of the Assembly of the Republic of Slovenia for the legislative period 1990-1992 when Slovenia became an independent country. The corpus comprises over 200 sessions, almost 60,000 speeches and 11 million words. It contains extensive meta-data about the speakers, a typology of sessions and structural and editorial annotations and is uniformly encoded to the Text Encoding Initiative (TEI) Guidelines, a de-facto standard for encoding and annotating textual data in Digital Humanities. It is available under the CC-BY licence in the CLARIN.SI repository of language resources and via the CLARIN.SI concordancers

(Pančur et al., 2017). SloParl is thus an exemplary corpus but contains material from a quite limited, and not very recent, time period. This makes the corpus of limited use for the rich body of research on recent parliamentary activities.

Contemporary Slovenian parliamentary debates are monitored by the analytical tool *Parlamer*¹ which makes use of linguistic as well as non-linguistic data, such as MPs' attendance and voting results. While this is a wonderful tool for journalists and citizen scientists and gives valuable insight into contemporary parliamentary data, the material is confined to the functionalities of the tool and as such cannot be freely manipulated by scholars according to their specific research needs.

The goal of the research presented in this paper was to convert the *Parlamer* database into a corpus and to enrich the linguistic data with the session and speaker metadata. Section 3. gives the basic information on the corpus structure and size, Section 4. presents the analysis of the corpus according to the speaker metadata (gender, education, age, party affiliation), and Section 5. gives some conclusions and directions for further research. While the focus of the paper is the parliamentary language material which we analyze with standard corpus and natural language processing approaches, the aim of the analysis is to inform media and political studies.

3. Corpus compilation

The corpus covers the full VIIth mandate of the Slovene Parliament, ranging from August 1st 2014 to March 19th 2018. Currently the central entities in the corpus are the sessions and speeches given by the members of parliament and other speakers as this is the most interesting content to researchers in the areas of linguistics, communication and political science.

The data model currently used for encoding the corpus is presented in Figure 1. Parliamentary sessions are equipped with the mandate they belong to and the name and date of the session. Sessions are further broken down into individual speeches which, in addition to the content of the contribution, is annotated with speaker name and other speaker information (date of birth, gender, education and education level, party affiliation). The rich speaker data are available for the members of parliament and members of the government but not for all other speakers in the parliament (e.g., field experts, representatives of governmental agencies, non-governmental organizations or civil initiatives). This is why the analyses in Section 4. are performed based on the instances for which the metadata is available in the corpus.

Some basic statistics regarding the corpus are given in Table 1. The transcripts were also processed with the standard linguistic annotation pipeline for Slovene consisting of *reldi-tokeniser*, which segments the text string into tokens and sentences, and *reldi-tagger*, which adds morphosyntactic descriptions (MSDs) and lemmas to the word tokens (Ljubešić et al., 2016).

A basic analysis of the morphosyntactic annotations of the corpus in form of the most significant differences in

```
[
  {
    mandate,
    session_name,
    date,
    speeches:
      [
        {
          id
          content,
          speaker:
            {
              id
              birth_date
              education
              education_level
              gender
              name
              party
            }
        },
        ...
      ]
    },
    ...
  ]
```

Figure 1: The current JSON data model of the corpus

Level	Count
Sessions	362
Days	514
Speeches	218,398
Speakers	1,984
Sentences	3,070,314
Words	61,039,385
Tokens	70,874,201

Table 1: Basic statistics of the corpus

frequency of MSD tags between the KRES balanced corpus of Slovene and the *Parlamer* corpus are given in Table 2.

The results show that the parliamentary speeches, as expected, contain more present tense word forms (*Vm.r[1-3][sp]*), especially in the first person singular or plural (*Vm.r1[sp]*), demonstrative pronouns (*Pd-.**), the first person singular personal pronoun (*Pp1-sn*), the first person auxiliary verbs (*Va-.1.-n*) and adverbs (*Rgp*) compared to general Slovene.

On the other hand, the parliamentary proceedings contain significantly fewer proper names (*Np.**), numbers (*Md.*), verb participle forms (*Vm.p-.**), personal pronoun dative cases (*Pp3..d.**) and general and possessive adjective forms (*A.p.**) than general Slovene.

4. Corpus analysis

This section presents a brief analysis of the corpus content given four main variables: gender, education, age and political affiliation of the speakers. In each session we disregard the speeches given by the most frequently occurring speaker because it is evident that this speaker was in charge of leading the session and, as a consequence, their content

¹<https://parlamer.si>

KRES	Parlameter
Npmsn	Vmpr1s
Mdc	Pd-nsn
Ncmsn	Vmpr1p
Vmep-sm	Pp1-sn
Npfsn	Pd-nsa
Vmep-sf	Va-r1p-n
Pp3msd-y	Vmpr1p-n
Va-r3d-n	Va-f1p-n
Npmsl	Rgp
Vmpp-sm	Pd-nsg
Pp3fsd-y	Pr-nsn
Agpmsny	Vmbr3s
Npmsg	Vmbr2p
Vmem2s	Pd-msa
Ncmisi	Pd-fsa
Vmpp-sf	Pd-msl
Px—d-y	Va-r2p-n
Mdo	Pi-msa
Aspfsn	Va-f1s-n
Aspmsnn	Pd-fsn

Table 2: Most significant differences in morphosyntactic categories used in the KRES balanced corpus of Slovene and the Parlameter corpus

would skew the distributions we are interested in. We also disregard all speeches of speakers for whom we do not have the necessary metadata.

4.1. Gender

The basic statistics regarding the number of speakers per gender and their linguistic production are given in Table 3. In total, the gender information is available for 139 speakers. 85 or 61% of those are male while 54 or 39% are female. Male speakers delivered three quarters of the speeches while female speakers only one quarter. However, on average, the speeches given by female speakers were 20% longer than those by male speakers.

	Male	Female
# of speakers	85	54
% of speakers	61	39
# of speeches	88,896	31,072
% of speeches	74	26
Avg # of words per speech	355	424

Table 3: Basic statistics regarding the gender

A keyword analysis, based on the Log Likelihood score given the male and female subcorpus, is presented in Figure 3. Among the top 100 keywords from the female speeches, apart from the general (17%) and administrative (32%) vocabulary, the most prominent topics are health (17%), social issues (13%), family (8%), and environmental (8%) issues, followed by education (2%) and finance (2%). In terms of word types, by far the most prevalent are nouns (62%) and adjectives (25%). Among the top 100 keywords from the

male speeches, there is much more general (51%) and administrative (30%) vocabulary, which mostly pertains to the meta-discussions of the procedures in parliamentary sessions, followed by proper names (11%). Specific topics are few and far between: transportation (6%), technology (1%) and finance (1%).

This analysis is very general as keywords were classified out of context and in cases of polysemous keywords, only the most predominant sense was considered, but still gives a valuable insight into the contributions by male and female speakers. That the nature and style of male speeches is quite different from the female ones can also be seen from the analysis of the types of words ranked as the most specific for male speeches. While nouns are the most frequent category here as well (42%), much more of those are used to address or refer to other people, e.g., *gospod*, *kolega*, *poslanec*, *predsednik*, and proper nouns, i.e., names of colleague MPs, ministers, parties and companies. Keywords from male speeches contain many more verbs (15%), adverbs, pronouns and particles, indicating a much more discursive and debating style than female speeches.

4.2. Level of education

In this section we present the basic statistics regarding the number of speakers per each education level and their linguistic production in Table 4. The codes for levels of education are the following:

- 5: secondary school
- 6/1: higher education degree
- 6/2: university bachelor degree
- 7: university master degree
- 8/1: scientific master degree
- 8/2 scientific doctorate degree

The statistics show that most of the members of parliament hold the old university or the new Bologna master degree (7), with a similar number of members holding the new Bologna bachelor degrees (6/1 and 6/2) and the old scientific masters or PhD (8/1 and 8/2) degrees.

Regarding the number of speeches given, the distribution roughly follows the distribution of speakers, with the least educated speakers speaking less frequently. These speakers, however, hold the longest speeches, which is an exception as overall the length of speeches tends to grow with the level of education.

4.3. Age

We organize the speakers' age by the decade in which they were born. The basic statistics regarding the number of speakers per each age group and their linguistic production in Table 5. The results of the analysis show that the most represented group are speakers born in the 1960s who were in their 40s and 50s in the mandate covered by the corpus. The most active group (roughly estimated as the difference between the percentage of speakers and the percentage of speeches given) are the youngest and the oldest members

	5	6-1	6-2	7	8-1	8-2
# of speakers	12	4	24	30	17	14
% of speakers	14	5	28	34	20	16
# of speeches	9,850	2,158	25,469	38,693	14,784	16,335
% of speeches	9	2	24	36	14	15
Avg # of words per speech	616	388	430	483	526	504

Table 4: Basic statistics regarding the education level

	1940s	1950s	1960s	1970s	1980s
# of speakers	8	34	44	36	13
% of speakers	6	25	33	27	10
# of speeches	14,804	22,187	32,421	26,372	22,298
% of speeches	13	19	27	22	19
Avg # of words per speech	210	502	590	509	407

Table 5: Basic statistics regarding age (decade of birth)

of parliament, giving roughly twice the amount of speeches than their representation is.

Interestingly, the average length of the speeches given follows roughly the distribution of the number of speakers in each age group, with the members born in the 60s holding the longest speeches, while the shortest speeches, more than half in length, are given by the oldest members. The youngest members also hold significantly shorter speeches than the three central age groups.

4.4. Political orientation

Our final analysis considers the activity and linguistic production of members given their party affiliation. The results for the six parties with the highest number of active members of parliament are given in Table 6.

The results show that all parties expect SMC give more speeches than their member number would suggest. The most active are SD and Levica, both left-wing, with SD one of the ruling parties and Levica in the opposition in this composition of the parliament. They account for twice the amount of speeches than their member count.

Regarding the length of the speeches, the speeches of the opposition parties are much longer than those by the ruling parties. The longest speeches are given by the right-wing SDS party, followed by another right-wing party, the NSI. The average length of their speeches are 4 times longer than those of DeSUS and SD who give the shortest speeches.

The top 100 keywords from the speeches given by members of the six most prominent parties are displayed in Figure 2. The biggest ruling party SMC's keywords clearly reflect their position and role in the parliament, which is to propose and pass legislation as well as take care of the procedural activities in sessions. Their keywords are very neutral and impersonal, highly procedural, administrative and legislative.

The keywords from their main opposition SDS, on the other hand, are much more discursive, critical and emotional. The most prominent topics in SDS speeches seem to be the judiciary branch, health care and migrants.

The member of the coalition, the DeSUS party, mostly dealt with the social welfare system and health care topic-wise but also made a lot of procedural comments and interacted with and referred to a lot of relevant actors by name or position.

The keywords of the third coalition party SD are almost exclusively related to procedural activities and references to other individuals.

The keywords of the opposition right-wing party NSI show big differences with their closest party SDS. They are very program-driven, mainly tackling economic issues. Interestingly, NSI is the only party with explicit references to religion.

Finally, the keywords of the left-wing opposition party Levica clearly show the core values and goals of this party, which are social rights and equality. Interestingly, the style of the keywords of Levica range from colloquial (e.g., *blazno, bajta*) to sophisticated (e.g., *nemara, ubesedovati*), thereby differing quite a lot from the rest of the parties in the parliament.

5. Conclusion

In this paper we presented the Parlameter corpus of contemporary Slovene parliamentary proceedings. We analyzed the linguistic production of the speakers according to the speaker metadata. We have shown that while male speakers take the floor much more often than their female colleagues, females make longer contributions. Female speakers mostly address the topics of social, health, family and environmental issues, while male speakers do not cover specific topics, but differentiate in using more verbs, adverbs, pronouns and particles, indicating a more discursive and debating style. In terms of education level, speakers with PhDs too deliver more but shorter speeches. Older speakers (those born in the 1950s) rarely speak but their speeches are the longest. When comparing speeches according to party lines, they are evenly distributed according to party representation in the parliament, most likely due to parliamentary bylaws. The average speech lengths of the ruling parties SMC, DeSUS and SD are the same whereas

nekako tudi predlog tikati odbor praviti ter moderen člen glede smc navzoč sprememba glasovati predlagan poudariti zveza unija član določen vsebina pravzaprav smer podlaga amandma omenjen zakon zakonodajno izpostaviti obravnava članica lesen omeniti postopek kar vidik prehajati center vseeno morda umetniški vsekakor odstavek obravnavati pristojnost namen okvir les sprejet evropski cilj smisel zelo prostor ureditev izpostavljen pojasniti naveden matičen nadomeščati obrazložitev raven republika določati predstavnica pripomba poslovnik vloženi sklep skupina obveščati področje pomemben 2017 urejati točka sklad javnofinančen razprava tale telo poslanka faza predlagatelj podati dopolnjen praven nek uporaba malce nadalje odločanje aktivnost vezan turizem delovanje sodelovanje predložiti beps tako

biti demokrati govorniki pogledati tisti slovenski vedeti reči nič kaj ali sodnik stranka koalicija zdaj narediti noben komisija predsednik povedati potem ministrica sodišče človekov nekdo mandatno takrat soden dati vrhoven mandat napisati vaš kršiti problem danes gledati ukc kakšen niti preiskovalen stvar pač jaz ustaven ampak janez pisati tam korupcija očitno minister koliko janša iti kandidat zadeva migrant cerar dobiti zgoditi tožilec enkrat predlagati vlada prej denar ker pravnik ilegalen senat sedeti opornica nikoli pravosoden policija sodnica zakaj dejati priti samo žilen klemenčič zdajle sodstvo zločin verjetno mark največkrat kpk davkoplačevalec kjer kako volilen človek vprašati spraševati državljanka pravosodje kangler

lep izvoliti desus skupina poslanski hvala predlog stališče predstaviti pokojnina mag seveda gospod upokojenec gospa beseda zakon zdravstven pokojninski poslanka franc ter invalid dopolnitev socialen novela pripraviti predstavitev anja prehajati kultura marija delo področje matičen zdravstvo horvat naj torej prekinjati jožef matej predlagatelj irena pozdrav obravnava zavarovanje namreč tašner žan upokojen majcen miha star žnidar usklajevanje dajati uroš marko zdravje dostojen zaključevati celarc dušan blagajna žibert telo zavod kolar regres andrej dneven milojka nekateri podpredsednik spoštovan mahnič prikl javen kordiš sprememba tanko marijan bojan kulturen pristojen dimic tomaž pojbič branko zdravko janko čuš tedaj proceduralno vatovec dediščina predložiti podkrajšek bah

izvoliti demokrat beseda socialen gospod mag gospa poslanski skupina lep hvala želeti izčistiti pomemben dajati rep-lika predlagatelj razprava rast obravnava zagotovo predstavitev tudi razpravljati zaključevati okvir hip postopkovno hainz pravzaprav anja primož pripraviti franc jožef horvat dopolnitev znotraj prehajati matej imeti izjemno kolikor predstaviti ter stališče sicer bah deti žibert vendarle bistveno položaj potrebno jože gospodarski tomaž marko janko marija branko predlog banka prekinjati zakon ključni zahteven muršič poskušati izobraževanje obdobje zvonko godec zame ugotavljati zato kočevski ferluga možnost mogoč vselej andrej potek kordiš prijava holding mlakar srečevati jan dimic tanko iva lisec veber slediti peter skozi kriza podkrajšek bizjak

nov slovenija krščanski naš digitalen evropski občina podjetnik jaz kolegica drag najbrž dober pomurski unija evro ven-darle kohezijski morda država projekt davčen želeti tisoč pomurje leto denar vipavski gotovo družina mlad regija demokrat obžalovati gospodarstvo lastnik stvar reforma bančen družinski zemljišče kmetijski komunalno posloven investitor kmet politika razumeti komunalen donacija penzija zgraditi lizbonski božičnica kolega zasedanje zunanji vladen obrtnik poko-jninski praktično demografija enostavno parlament kibernetiki konkurenčnost poudarjati župan vesel program regresen odpadek asistenca parcela operativen blagajna agenda podonavski gozd kapica bog rodnost vplačevati vodovod strukturen ikt okrog bolezen zdravljenje zgrajen gradben fantastičen šola graditi plečnikov bolniški članica piten lanski proračunski

levica združen nek navsezadnje skratka malo kapital delavec delavski penez privatizacija nekako bistvo revščina hoteti desnica resno socialist gor pogosto odkrito reven dol rad nato bolj sporazum podjetje tuliti minimalen prečenjeb debata kap-italizem politika žica resen predsedujoč dobiček stoletje konoplja stanovanjski neoliberalen onkraj bogat maribor koper resoren pogovarjati brati zaposlen plača socialen prečiti cel begunec sočasno četrt nehati lobi korporacija bajta firma prekaren ameriški izhajajoč privaten žival evro deregulacija profit skoraj neenakost kot rezilen družben stanovanje pre-bivalec pol zgodba ampak citat čeprav ips izvršilen lekarna blazno logika namesto težiti ubesedovati nemara človek tukaj neki trenutno živeti soupravljanje minimum vračljivost niti

Figure 2: Most prominent terms in speeches given by members of six most prominent parties (SMC, SDS, DeSUS, SD, NSI, Levica)

otrok zdravstven javen ukrep zavod pravica zdravnik starš tudi socialen zdravstvo čakalen ukc program področje družina varstvo leto sprememba družinski ter človekov zdravje bolnišnica opornica izobraževanje nasilje žilen res meniti pomoč peticija ministrstvo torej sredstvo delo prav pacient višina kakovost izvajanje šola dodatek center novela doba priprava pozdravljen letošnji oskrba žival romski storitev bolnik oseba krma dejavnost varstven enak preživnina dobavitelj otroški posamezen naročilo medicinski ženska ekološki zavarovanje vendar ureditev potreba zdravilo odhodek podneben mleko mark predlagan zakon potreben živilo obvezen gensko zaposlovanje dolgotrajen cilj duševen vsekakor sicer dostopnost oziroma podati brezposeln odrasel medical izguba transfer rastlina jamstvo denaren 2016

nek imeti reči gospod tisti mag zbor hoteti gledati naprej mandatno malo jaz dneven tir red predsednik tam noben ali zgodba navzoč poslanec gospa vlada levica volilen ura resnica iti zadeva tanko seveda točka navsezadnje resen beseda glasovati državen dalje najbrž zdaj videti prehajati luka kolega kakšen biti stvar postopkoven moj kandidat franc početi priti združen matej relativno nekaj postopkovno koper obrazložitev vsaj praviti banka tak seja proti jože razumeti sds predstavitev kaj trček verjetno digitalen poslanski gor stališče minister infrastruktura misliti preprosto nekdo ime resno zaključevati opozicija uber janša prekinjen nekako minuta sklep dol promet železnica tonin ker glas

Figure 3: Most prominent terms in female and male speeches

	SMC	SDS	DeSUS	SD	NSI	Levica
# of speakers	43	21	13	9	6	6
% of speakers	44	21	13	9	6	6
# of speeches	20656	23876	17340	17367	8788	10753
% of speeches	21	24	18	18	9	11
Avg # of words per speech	366	522	151	152	462	427

Table 6: Basic statistics regarding political orientation

the opposition parties the speeches of SDS, NSI and Levica are more than twice longer.

In the future we plan to enrich the corpus with additional session records of other parliamentary seatings but also with additional metadata available through the Parameter system, such as voting data and accepted legislation, which are also valuable for addressing a number of research questions in various research communities. In parallel, we also plan to develop comparable corpora from other parliaments, starting with Croatian and Bosnian.

Acknowledgements

The work presented in this paper was partially funded by the Slovenian Research Agency within the national basic research project “Resources, methods and tools for the understanding, identification and classification of various forms of socially unacceptable discourse in the information society” (J7-8280, 2017-2020).

6. References

- Paul Bayley. 2004. *Cross-cultural perspectives on parliamentary discourse*, volume 10. John Benjamins Publishing.
- Jennifer E Cheng. 2015. Islamophobia, muslimophobia or racism? parliamentary discourses on Islam and Muslims in debates on the minaret ban in Switzerland. *Discourse & Society*, 26(5):562–586.
- Daniel Norman Chester and Nona Bowring. 1962. *Questions in parliament*. Clarendon Press.
- Darja Fišer and Jakob Lenardic. 2018. Parliamentary corpora in the CLARIN infrastructure. In *Selected papers from the CLARIN Annual Conference 2017, Budapest, 18–20 September 2017*, number 147, pages 75–85. Linköping University Electronic Press.
- Mark N Franklin and Philip Norton. 1993. *Parliamentary Questions: For the Study of Parliament Group*. Oxford University Press, USA.
- Graeme Hirst, Vanessa Wei Feng, Christopher Cochrane, and Nona Naderi. 2014. Argumentation, ideology, and issue framing in parliamentary discourse. In *ArgNLP*.
- Lorna M Hughes, Paul S Ell, Gareth AG Knight, and Milena Dobrova. 2013. Assessing and measuring impact of a digital collection in the humanities: An analysis of the sphere (stormont parliamentary hansards: Embedded in research and education) project. *Digital Scholarship in the Humanities*, 30(2):183–198.
- Pasi Ihalainen, Cornelia Ilie, and Kari Palonen. 2016. *Parliament and Parliamentarism: A Comparative History of a European Concept*. Berghahn Books.
- Cornelia Ilie. 2017. Parliamentary debates. *The Routledge Handbook of Language and Politics*.
- Nikola Ljubešić, Tomaž Erjavec, Darja Fišer, Tanja Samardžić, Maja Milicević, Filip Klubička, and Filip Petkovski. 2016. Easily accessible language technologies for Slovene, Croatian and Serbian. In *Proceedings of the Conference on Language Technologies and Digital Humanities*, pages 120–124.
- Andrej Pančur and Mojca Šorn. 2016. Smart big data: Use of slovenian parliamentary papers in digital history. *Prispevki za novejšo zgodovino/Contributions to Contemporary History*, 56(3):130–146.
- Andrej Pančur, Mojca Šorn, and Tomaž Erjavec. 2017. *Slovenian parliamentary corpus SlovParl 2.0*. Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1167>.
- Andrej Pančur. 2016. Označevanje zbirke zapisnikov sej slovenskega parlamenta s smernicami TEI (Encoding the Slovenian Parliament Session Minutes in Line with the TEI Guidelines). In *Proceedings of the Conference on Language Technologies and Digital Humanities*, pages 142–48.
- Ludovic Rheault, Kaspar Beelen, Christopher Cochrane, and Graeme Hirst. 2016. Measuring emotion in parliamentary debates with automated textual analysis. *PLoS one*, 11(12):e0168843.
- Teun A Van Dijk. 2010. Political identities in parliamentary debates. *European Parliaments under Scrutiny. Discourse strategies and interaction practices*, pages 29–56.