Analyses of Networks of Politicians Based on Linked Data: Case ParliamentSampo – Parliament of Finland on the Semantic Web

Henna Poikkimäki[1,2(✉)], Petri Leskinen[1,2], Minna Tamper[1,2], and Eero Hyvönen[1,2]

[1] Semantic Computing Research Group (SeCo), Aalto University, Espoo, Finland
henna.poikkimaki@aalto.fi
[2] HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Helsinki, Finland

Abstract. In parliamentary debates the speakers make reference to each other. By extracting and linking named entities from the speeches it is possible to construct reference networks and use them for analysing networks of politicians and parties and their debates. This paper presents how such a network can be constructed automatically, based on a speech corpus 2015–2022 of the Parliament of Finland, and be used as a basis for network analysis.

Keywords: Parliamentary studies · Network analysis · Linked data · Digital humanities

1 Introduction

Parliamentary speech data of plenary sessions provide a wealth of information about the state and functioning of democratic systems, political life, language, and culture [2]. Arguably, using Linked Data (LD) for representing metadata about the speeches and the politicians is useful: LD provides well-defined semantics for representing and enriching knowledge[1] aggregated from heterogeneous data sources [4,26], as well as methods for publishing the data for data-analyses and for developing applications, such as semantic portals [9].

This paper argues for using textual speech data for analyzing networks of politicians and parties. It is shown, how Named Entity Recognition (NER) and Linking (NEL) [18] can be used for extracting reference networks from parliamentary speeches, and be used for network analyses. As a case study, speeches and politicians of the Parliament of Finland are considered. We first present a data model and data service for representing reference networks of parliamentary speeches, and how such a network was created based on the ParliamentSampo data [9,15,22]. After this, a system for making network analysis is presented with examples from a Finnish dataset 2015–2019 of 65 000 speeches. Finally, contributions of the work are discussed and directions for further research outlined.

[1] https://www.w3.org/standards/semanticweb/.

S. Chiusano et al. (Eds.): ADBIS 2022, CCIS 1652, pp. 585–592, 2022.
https://doi.org/10.1007/978-3-031-15743-1_53

2 Related Works

Several parliamentary corpora have been formed from the minutes of the plenary debates, see, e.g., [14] and the CLARIN list of parliamentary corpora[2]. The related ParlaMint project[3] [6] brings together Parla-CLARIN-based national corpora. Parliamentary materials have also been transformed into the form of LD when creating the LinkedEP [26] system on the European Parliament's data, the Italian Parliament[4], the LinkedSaeima for the Latvian parliament [4], and the Finnish ParliamentSampo [9,15,22] whose data is used in this paper. The corpora have been used mostly for linguistic analyses, not for network science as suggested in our paper. For example, in [3] the content of women's parliamentary speeches in the British Parliament are analyzed, and in [1,7,10,12,20] thematic and conceptual analyses of language and the opinions were made.

Network science [21,27,29], a field revolutionized around 20 years ago, has been successful in explaining phenomena and fundamental concepts in a wide array of systems from societies to brain and cellular biology. The tools and ideas developed for this range from ideas characterising the whole network to diagnostics computed for individual nodes, such as centrality measures, node roles, and local clustering coefficients. Given the unexploited potential, there is a need to provide networks hidden in parliamentary data for researchers to study in Digital Humanities (DH) using methods of network science.

3 A Data Model and Service for Speech Networks

The ParliamentSampo data publication consists of two parts: 1) a knowledge graph (KG) of all parliamentary debate speeches in Finland from 1907 [22] and 2) a KG of the Members of Parliament (MP) and the parliament organizations [15]. In our earlier research, sociocentric and egocentric networks connecting the actors could be constructed from texts based on, e.g., mentioned names, hypertext links, genealogical relations, or similarities in characteristics such as lifetime events [5,16,24]. In this paper, the same idea is applied to parliamentary speeches that make mutual references to each other through mentioned MPs.

In order to extract such reference networks, the original RDF speech graph was enriched with Natural Language Processing (NLP) methods, such as named entity recognition (NER) and linking (NEL) [23]. The extraction was done using the upgraded Nelli tool [25] by querying the textual speeches from the ParliamentSampo SPARQL endpoint[5]. The speeches were first cleaned from the interruptions and then lemmatized using the Turku Neural Parser[6] [11]. The FinBERT-NER model [28] was used on lemmatized texts for NER. After NER,

[2] https://www.clarin.eu/resource-families/parliamentary-corpora.
[3] https://www.clarin.eu/content/parlamint-towards-comparable-parliamentary-corpora.
[4] http://data.camera.it.
[5] See the datamodel for speeches [22], MPs [15], and their enrichments respectively.
[6] http://turkunlp.org/Turku-neural-parser-pipeline.

the mentioned people, places, groups, organizations, and their related information were then linked internally to the ParliamentSampo knowledge graph of MPs using the ARPA tool [17] on the extracted named entities. For a broader data enrichment, linkings to external data sources were created, including the Kanto[7] vocabulary for Finnish actors provided by the National Library, the YSO Places ontology[8], PNR[9] gazetteer of Finnish place names by the National Survey, and the Semantic Finlex[10] [19] legal KG of the Ministry of Justice.

The NEL results were transformed into RDF as instances of the class NamedEntity described in Table 1. Namespace *provo* refers to the PROV-O ontology[11]. The instances link the resources behind the mentions to speakers.

Table 1. Metadata schema for the class for NamedEntity.

Element URI	C	Range	Meaning of the value
:surfaceForm	1	xsd:string	Original surface form in text
:count	1	xsd:integer	Number of entity mentions in a speech
:category	1	xsd:string	Type of the named entity
skos:relatedMatch	0..*	rdfs:Resource	Links to ontologies for named entities
provo:wasAssociatedWith	1..*	:NamedEntityMethod, provo:SoftwareAgent	Provenance information about the method used to extract the named entity

The accuracy of the NER was estimated for 100 randomly selected mentions of people, places, organizations, and expressions of time. The precision was 97%, recall 77%, and F1-score 86%. Based on our initial evaluation, 88% of the speeches contained a named entity of which 30% contained a person name. From this evaluation set, roughly 20% contained only person names and for them the F1-score was 100%. The linking of person names was more tricky. Currently, the person names were linked internally to ParliamentSampo successfully only if a person was mentioned using the full name. The results for linking people were calculated for 50 randomly selected speeches (containing 105 person mentions) where precision was 95%, recall 80%, and F1-score 87%. The family name references were not linked to MPs and it remains as a future work to connect the family names and the full names properly.

4 Analyzing Networks of Politicians

This section gives examples of analysing networks of MPs using the Python package NetworkX [8]. Two different reference networks were constructed based on speeches given during the electoral term 2015–2019 and so far linked person

[7] https://finto.fi/finaf/en/.
[8] https://finto.fi/yso-paikat/en/?clang=en.
[9] http://www.ldf.fi/dataset/pnr.
[10] https://data.finlex.fi.
[11] https://www.w3.org/TR/prov-o/.

names. Speeches that do not contain any linked person names were excluded from the analysis. In addition, administrative speeches of the Speaker of the Parliament of Finland were not taken into account. Analyses of reference networks can, e.g., reveal MPs who are most active in parliamentary debates and help to recognize possible disputes between MPs or parties.

The first reference network has MPs as nodes and the other one parties. In the former case, links point from the speaker to the mentioned person, and the weight of the link corresponds to the total number of speeches with at least one mention. The network has in total 209 MPs that have been mentioned or have mentioned someone. The total number of mentions to other MPs extracted from the speeches is 2108. Mentions of people who were not MPs or ministers at the chosen electoral term were filtered out of the result set.

To study and visualize the network, hub and authority values were calculated using the HITS algorithm [13]. Ten MPs with highest authority values and ten nodes with highest hub values are shown in Fig. 1. From the MPs with the highest authority values, Juha Sipilä, Timo Soini, and Petteri Orpo were ministers and leaders of their parties during the 2015–2019 term. During the same years, Antti Rinne, Ville Niinistö, and his successor Touko Aalto from the opposition served also as leaders of their parties. Three MPs, Touko Aalto, Pia Viitanen, and Ben Zyskowicz, are both top hubs that make references as well as top authorities often mentioned by other MPs. None of the MPs with highest hub values were ministers.

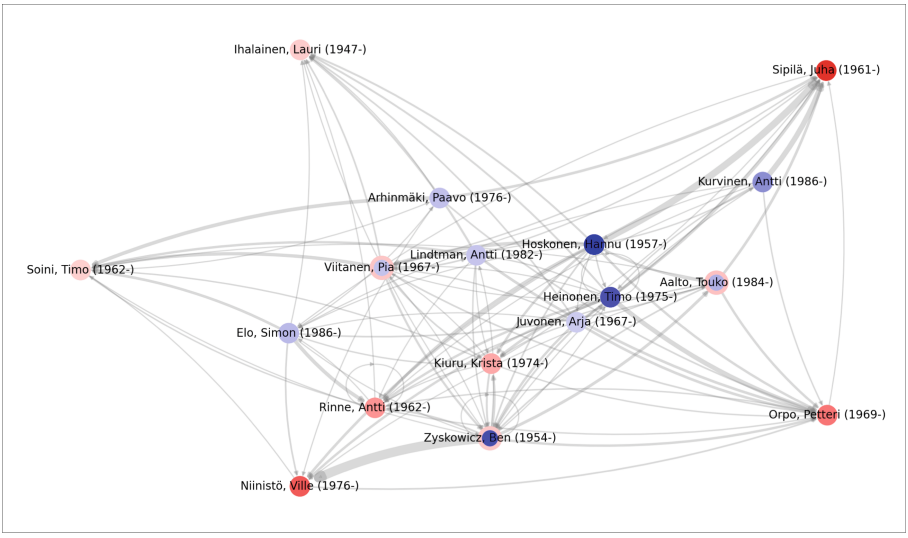

Fig. 1. Ten MPs with highest hub and authority values based on the HITS algorithm. The darker red, the larger authority value, and the darker blue, the larger hub value. (Color figure online)

The second reference network uses parties as nodes instead of MPs. Mentions between MPs were obtained in a similar manner as for the first reference network. MPs were then grouped into 11 parties by their memberships. The Sankey diagram in the Fig. 2 depicts how MPs in government parties refer to MPs in other parties and Fig. 3 the other way around. For example, MPs in the Green League refer mostly to MPs in the Centre Party and National Coalition Party, probably due to disputes related to, e.g., environmental issues. The Sankey diagrams were rendered using the Python module pySankey[12].

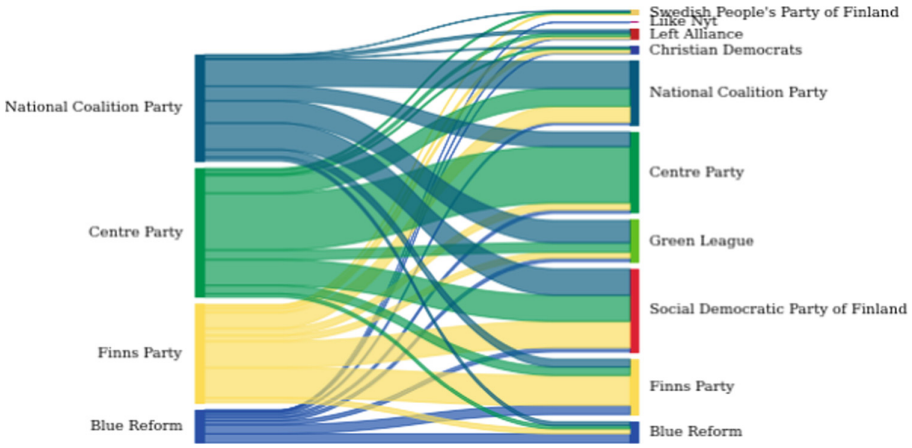

Fig. 2. The mentions from government parties (on the left) to all parties (on the right) with nonnormalized weights.

Analyses like these can be used to point out phenomena of potential interest in parliamentary discussion, but in order to really interpret the results traditional close reading is needed. The faceted search engines and other tools of the ParliamentSampo portal make it easy to filter out, e.g., those speeches of an individual MP or all members of a party that mention particular MPs [9].

5 Discussion

This paper presented the idea of creating reference networks of MPs based on references in their speeches. As a case study, reference networks based on 65 000 parliamentary speeches were created using methods of NER and NEL, and resulting examples of network analysis were presented. Our first experiments suggest that network analysis can be used to detect possibly interesting phenomena in the discussions but interpreting the results require close reading the related texts. In order to support both distant and close reading tasks, the ParliamentSampo system under development aims to integrate seamlessly semantic faceted search

[12] https://github.com/anazalea/pySankey.

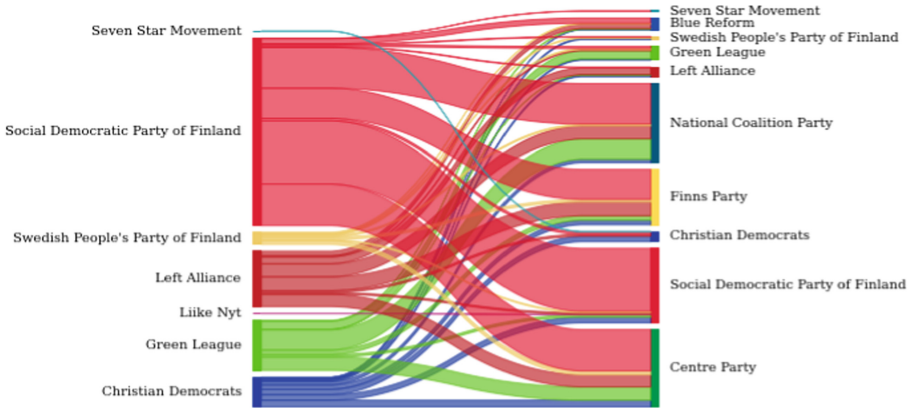

Fig. 3. The mentions from opposition parties (on the left) to all parties (on the right) with nonnormalized weights.

and browsing facilities with data-analytic tools. On one hand, faceted search can be used to filter out subsets of speeches to be studied with methods of network analysis, and on the other hand, faceted search can be used for finding speeches when interpreting analyses made using external tools on top of the SPARQL endpoint, like the ones presented in Sect. 4.

Acknowledgements. Our work is funded by the Academy of Finland and is related to the EU project InTaVia (https://intavia.eu) and the EU COST action Nexus Linguarum (https://nexuslinguarum.eu). Computing resources of the CSC – IT Center for Science were used.

References

1. Beelen, K., et al.: Digitization of the Canadian parliamentary debates. In: Säily, T., Nurmi, A., Palander-Collin, M., Auer, A. (eds.) Exploring Future Paths for Historical Sociolinguistics, pp. 83–107. John Benjamins, Amsterdam (2017). https://doi.org/10.1017/S0008423916001165
2. Benoît, C., Rozenberg, O. (eds.): Handbook of Parliamentary Studies: Interdisciplinary Approaches to Legislatures. Edward Elgar Publishing (2020). https://doi.org/10.4337/9781789906516
3. Blaxill, L., Beelen, K.: A feminized language of democracy? The representation of women at Westminster since 1945. Twentieth Century Br. Hist. **27**(3), 412–449 (2016). https://doi.org/10.1093/tcbh/hww028
4. Bojārs, U., Daris, R., Lavrinovičs, U., Paikens, P.: LinkedSaeima: a linked open dataset of Latvia's parliamentary debates. In: Acosta, M., Cudré-Mauroux, P., Maleshkova, M., Pellegrini, T., Sack, H., Sure-Vetter, Y. (eds.) SEMANTiCS 2019. LNCS, vol. 11702, pp. 50–56. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33220-4_4
5. Elson, D.K., McKeown, K., Dames, N.J.: Extracting social networks from literary fiction. aclweb.org https://www.aclweb.org/anthology/P10-1015.pdf

6. Erjavec, T., et al.: The ParlaMint corpora of parliamentary proceedings. Lang. Resour. Eval. (2022). https://doi.org/10.1007/s10579-021-09574-0
7. Guldi, J.: Parliament's debates about infrastructure: An exercise in using dynamic topic models to synthesize historical change. Technol. Cult. **60**(1), 1–33 (2019). https://doi.org/10.1353/tech.2019.0000
8. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using networkx. In: Varoquaux, G., Vaught, T., Millman, J. (eds.) Proceedings of the 7th Python in Science Conference, Pasadena, CA, USA, pp. 11–15 (2008)
9. Hyvönen, E., et al.: Finnish parliament on the semantic web: using ParliamentSampo data service and semantic portal for studying political culture and language. In: DIPADA, Digital Parliamentary data in Action. CEUR Workshop Proceedings, vol. 3133 (2022). https://ceur-ws.org/Vol-3133/paper05.pdf
10. Ihalainen, P., Sahala, A.: Evolving conceptualisations of internationalism in the UK parliament: collocation analyses from the league to Brexit. In: Fridlund, M., Oiva, M., Paju, P. (eds.) Digital histories: Emergent Approaches Within the New Digital History, pp. 199–219. Helsinki University Press (2020). https://doi.org/10.33134/HUP-5-12
11. Kanerva, J., Ginter, F., Miekka, N., Leino, A., Salakoski, T.: Turku neural parser pipeline: an end-to-end system for the CoNLL 2018 shared task. In: Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 31 October–1 November 2018, Brussels, Belgium, pp. 133–142. Association for Computational Linguistics, Brussels, Belgium (2018)
12. Kettunen, K., La Mela, M.: Semantic tagging and the Nordic tradition of everyman's rights. Digit. Sch. Humanit. (2021). https://doi.org/10.1093/llc/fqab052
13. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. Assoc. Comput. Mach. **46**, 604–632 (1999). https://doi.org/10.1145/324133.324140
14. Lapponi, E., Søyland, M.G., Velldal, E., Oepen, S.: The Talk of Norway: a richly annotated corpus of the Norwegian parliament, 1998–2016. Lang. Resour. Eval. **52**(3), 873–893 (2018). https://doi.org/10.1007/s10579-018-9411-5
15. Leskinen, P., Hyvönen, E., Tuominen, J.: Members of Parliament in Finland knowledge graph and its linked open data service. In: Graphs. Proceedings of the 17th International Conference on Semantic Systems, 6–9 September 2021, Amsterdam, The Netherlands, pp. 255–269 (2021). https://doi.org/10.3233/SSW210049, https://ebooks.iospress.nl/volumearticle/57420
16. Leskinen, P., Rantala, H., Hyvönen, E.: Analyzing the lives of Finnish academic people 1640–1899 in Nordic and Baltic countries: AcademySampo data service and portal. In: 6th Digital Humanities in Nordic and Baltic Countries Conference, Long Paper. CEUR Workshop Proceedings (March 2022, forth-coming)
17. Mäkelä, E.: Combining a REST lexical analysis web service with SPARQL for mashup semantic annotation from text. In: Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., Tordai, A. (eds.) ESWC 2014. LNCS, vol. 8798, pp. 424–428. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11955-7_60
18. Martinez-Rodriguez, J.L., Hogan, A., Lopez-Arevalo, I.: Information extraction meets the semantic web: a survey. Semant. Web Interoperability Usability Applicability **11**(2), 255–335 (2020)
19. Oksanen, A., Tuominen, J., Mäkelä, E., Tamper, M., Hietanen, A., Hyvönen, E.: Semantic Finlex: transforming, publishing, and using Finnish legislation and case law as linked open data on the web. In: Knowledge of the Law in the Big Data Age. Frontiers in Artificial Intelligence and Applications, vol. 317, pp. 212–228. IOS Press (2019)

20. Quinn, K., Monroe, B., Colaresi, M., Crespin, M.H., Radev, D.R.: How to analyze political attention with minimal assumptions and costs. Am. J. Polit. Sci. **54**, 209–228 (2010). https://doi.org/10.1111/j.1540-5907.2009.00427.x
21. Saramäki, J., Moro, E.: From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. Eur. Phys. J. B **88**(6), 1–10 (2015). https://doi.org/10.1140/epjb/e2015-60106-6
22. Sinikallio, L., et al.: Plenary debates of the parliament of Finland as linked open data and in Parla-CLARIN markup. In: 3rd Conference on Language, Data and Knowledge, LDK 2021, pp. 1–17. Schloss Dagstuhl- Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing, August 2021. https://drops.dagstuhl.de/opus/volltexte/2021/14544/pdf/OASIcs-LDK-2021-8.pdf
23. Tamper, M., Leal, R., Sinikallio, L., Leskinen, P., Tuominen, J., Hyvönen, E.: Extracting knowledge from parliamentary debates for studying political culture and language, May 2022, forth-coming
24. Tamper, M., Leskinen, P., Hyvönen, E., Valjus, R., Keravuori, K.: Analyzing biography collection historiographically as linked data: case national biography of Finland. Semant. Web Interoperability Usability Applicability (2021, accepted). https://seco.cs.aalto.fi/publications/2021/tamper-et-al-bs-2021.pdf
25. Tamper, M., Oksanen, A., Tuominen, J., Hietanen, A., Hyvönen, E.: Automatic annotation service APPI: named entity linking in legal domain. In: Harth, A., et al. (eds.) ESWC 2020. LNCS, vol. 12124, pp. 208–213. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62327-2_36
26. Abercrombie, G., Batista-Navarro, R.: Sentiment and position-taking analysis of parliamentary debates: a systematic literature review. J. Comput. Soc. Sci. **3**(1), 245–270 (2020). https://doi.org/10.1007/s42001-019-00060-w
27. Vespignani, A.: Twenty years of network science. nature **558**(7711), 528–529 (2018). https://doi.org/10.1038/d41586-018-05444-y
28. Virtanen, A., Kanerva, J., Ilo, R., Luoma, J., Luotolahti, J., Salakoski, T., Ginter, F., Pyysalo, S.: Multilingual is not enough: BERT for Finnish (2019)
29. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. nature **393**, 440–442 (1998). https://doi.org/10.1038/30918