

Using ParliamentSampo Linked Open Data Service and Portal for Analyzing Interruptions and Laughter in the Plenary Sessions of the Parliament of Finland

Eero Hyvönen^{1,2,*}, Petri Leskinen^{1,2,*}, Heikki Rantala^{1,2,*}, Jouni Tuominen^{1,2,*} and Annastiina Ahola^{1,*}

¹Aalto University, Department of Computer Science, Espoo, Finland

²University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG and HSSH), Helsinki, Finland

Abstract

This paper presents and demonstrates two new features implemented in the in-use semantic portal “PARLIAMENTSAMPO – Parliament of Finland on the Semantic Web”: how to search, analyze, and visualize how the Members of Parliament (MP) interrupt each others’ speeches and laugh on each other. These extensions are based on enriching the PARLIAMENTSAMPO knowledge graph (KG) by new properties extracted automatically from the manually transcribed official minutes of the plenary sessions. The analyzes and visualization can be created using the portal with a few clicks not only for individual speakers but also on the prosopographical party level and in relation to the topics and geospatial locations discussed.

Keywords

linked data, parliamentary data, digital humanities, user interface, semantic portal, SPARQL

1. Introduction

PARLIAMENTSAMPO [1] is a Finnish in-use linked open data (LOD) service¹ and semantic portal² that publishes all over million plenary speeches of the Parliament of Finland (PoF) since it was established in 1907, integrated with biographical data about the over 3000 parliamentarians who have spoken in the plenary sessions and activities of the PoF. The data consists of two knowledge graphs (KG): a Speech KG (SKG) [2] for representing the speeches and a Proposographical KG (PKG) [3] about the MPs and the parties and other organizational structures and activities of the PoF. The data and the portal have been used widely in Digital Humanities (DH) research and also by the public and the media as a means of finding out easily who has said what, when, in what context, and in what capacity.

The Speech KG has been extracted and enriched from textual transcriptions of the speeches made by the official minute-keepers of the plenary sessions. The transcriptions record not only what was spoken by whom, when, and in what session, but also if the speech was interrupted by someone or if the MPs in general reacted to the speech by shouts or laughter. Such interruptions have been of interest to the public, researchers, and the media (see, e.g., the newspaper articles listed on the project home page³). However, searching and analyzing them using the portal has not been possible, and several requests for such functionality have been submitted. To satisfy this need, this paper presents an extension to the original portal that facilitates 1) searching and analyzing interruptions of MPs and 2) moments of laughter in the PoF.

Posters, Demos, Blue Sky, and Tutorials at SEMANTICS 2026, Sep 2026, Ghent, Belgium

*Corresponding author

✉ eero.hyvonen@aalto.fi (E. Hyvönen); petri.leskinen@ext-aalto.fi (P. Leskinen); heikki.rantala@aalto.fi (H. Rantala); jouni.tuominen@helsinki.fi (J. Tuominen); annastiina.ahola@aalto.fi (A. Ahola)

🌐 <https://seco.cs.aalto.fi/u/eahyvone/> (E. Hyvönen)

🆔 0000-0003-1695-5840 (E. Hyvönen); /0000-0003-2327-6942 (P. Leskinen); 0000-0002-4716-6564 (H. Rantala); 0000-0003-4789-5676 (J. Tuominen); 0009-0008-6369-4712 (A. Ahola)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹PARLIAMENTSAMPO LOD service: <https://ldf.fi/datasets/semparl/>

²PARLIAMENTSAMPO portal: <https://parlamenttisampo.fi>

³Project home page: <https://seco.cs.aalto.fi/projects/semparl/en/>

The software PARLIAMENTSAMPO is generic and applicable to the minutes data with corresponding mark-up of parliaments in other countries or minutes of international multilateral organizations [4].

2. Related Work

Parliaments and cultural heritage organizations in different countries have created parliamentary speech corpora and digital parliamentary datasets of both historical and contemporary parliaments [5, 6]. The goal of this work has been to improve the findability, accessibility, interoperability, and re-usability of these key documents of democratic societies for the public, researchers, and other users. The digitization has also allowed researchers to engage in novel and interdisciplinary research using the new parliamentary data. Semantic web technologies have also been applied for linking and enriching the parliamentary data with other datasets. In the project Linked Data of the European Parliament (LinkedEP), the debates of the European Parliament and the political affiliation information were connected as linked data into other datasets, such as DBpedia and the EuroVoc thesaurus [7]. Other examples of linked data parliament initiatives are the LinkedSaeima for the Latvian parliament [8], the Italian Parliament data⁴, and the historical Imperial Diet of Regensburg of 1576 project [9].

Several linguistic and social science studies have been conducted using minutes of the PoF. In [10] the history of Nordic right of public access to nature was studied. Andrushchenko et al. [11] have used their grammatically structured corpus for selected digital humanities research cases. Simola [12] has explored the differences in political speech between parties in the long term (1907–2018), and Makkonen and Loukasmäki [13] have used topic modeling to study the plenary debates of PoF in 1999–2014. Encoding and analysis of interruptions and laughter in parliamentary datasets have previously been discussed, for example, in the context of the Austrian parliament [4], but without actual applications.

3. Technical Solution

Interruptions, including moments of laughter, were marked up in parentheses in the content text of the plenary minutes, or in newer minutes by embedding custom XML markup in the minutes. The interruptions of general laughter were marked with the word “laughter”⁵. During the data speech transformation [2] the interruptions were found from the minutes and represented as instances of the class `: Interruption` that were linked to the speech instances by the property `: isInterruptedBy`. The class `: Interruption` has three properties explained in Table 1.

Table 1
RDF schema properties for the class `: Interruption`.

Element URI	Cardinality	Range	Meaning of the value
<code>: content</code>	1	<code>rdfs: Literal</code>	Content of the interruption
<code>: interrupter</code>	0..1	<code>rdfs: Literal</code>	Source of the interruption
<code>: speaker</code>	0..1	<code>bioc: Person</code>	Interrupter URI, if interrupter was mentioned

Interruptions can be analyzed in different ways based on the data, and this possibility has already been used in some Google Colab-based analyzes [1]. However, the portal did not initially offer the possibility to search and analyze different kinds of interruptions. The interruptions with laughter can be found from the content of the speeches, for example, by using regular expressions. However, to be able to create a facet for laughter in speeches, it was practical to add a new property `<http://ldf.fi/schema/parliamentsampo-portal/hasInterruptionType>` for each speech in the speech KG data. Its value tells whether the speech has or does not have interruptions or laughter. In this way, the SKG speech graph can be conveniently used to implement faceted search using SPARQL queries

⁴<http://data.camera.it>

⁵The word used is usually the Finnish word for laughter “*naurua*” or the Swedish word “*skratt*” in Swedish speeches. The older minutes sometimes use the word “*hilpeyttä*” which is Finnish for hilarity.

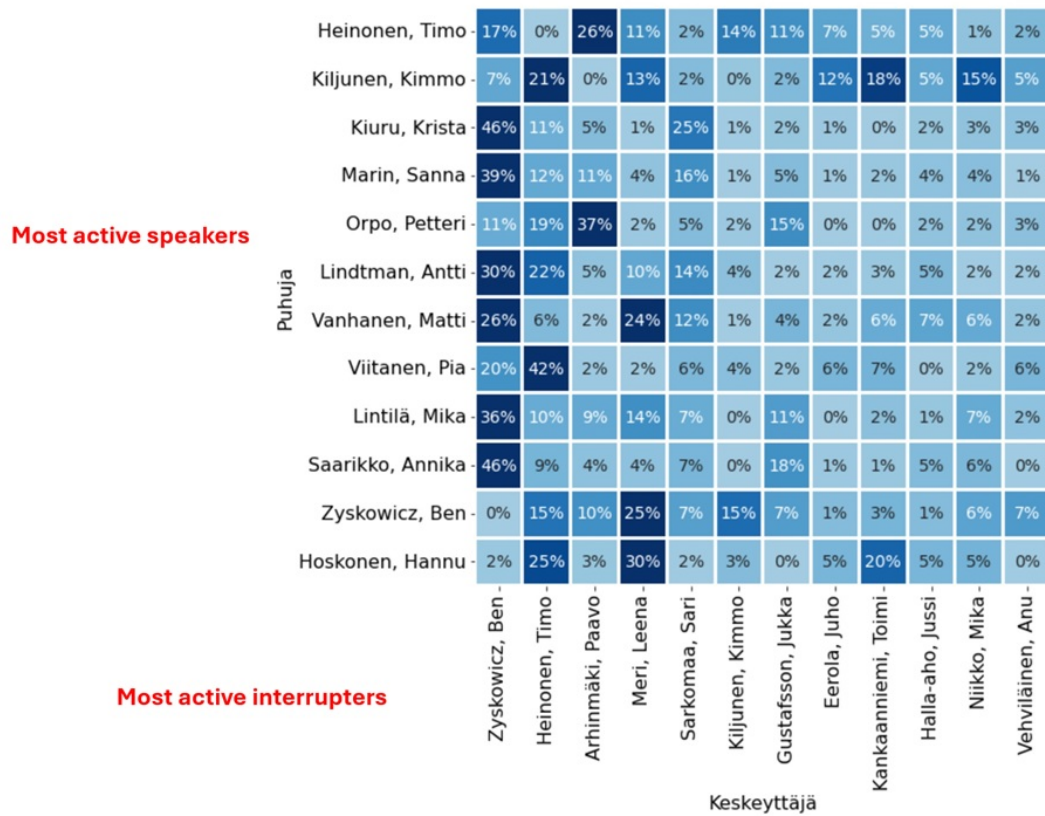


Figure 1: Table of correlations that indicate how the most active speakers (y-axis) were interrupted by most active interrupters (x-axis) during the election season 2029–2022.

created by the user interface based on Sampo-UI [14, 15]. The properties were created using a SPARQL CONSTRUCT query with a regular expression.

Sampo-UI aims to create a kind of generic, domain-agnostic “standard” user interface (UI) logic model for LOD applications, based on the basic concepts of classes, their properties, and instances in a KG. The framework can therefore be applied easily to search, browse, and visualize virtually any KG available in a SPARQL endpoint. In contrast to traditional programming libraries, Sampo-UI is based on a declarative, configuration-based approach: a new UI is created by taking an existing Sampo-UI interface as a starting point, by pointing to a new SPARQL endpoint, by configuring SPARQL queries to match the underlying data model used in the endpoint, and by selecting and modifying configurations for the needed visualization components.

The landing page of a Sampo portal⁶ based on the Sampo-UI framework⁷ contains a series of *application perspective* windows that allow one to search, browse, and analyze the underlying KG from different perspectives, based on the classes of the KG. A perspective for a class contains a faceted search engine whose facets are based on the properties of the class; by making selections on the facets, a corresponding subset of individuals of the class is retrieved as the search result, and hit counts on the facet categories are updated to guide the search. The result set can then be analyzed on different tabs, e.g., on a map or timeline or using charts. At any point, an individual can be chosen for a closer look at its *instance page* that provides comprehensive linked data and visualizations about the individual on separate tabs.

4. Analyzing Interruptions and Laughter

Interruptions can be analyzed using either scripting on top of the SPARQL endpoint [16] or by the portal. For example, the matrix visualization of Fig. 1 was created using Google Colab to analyze the interruptions of speeches of the PoF in 2019–2022. The y-axis lists the most active speakers and x-axis the MPs that have interrupted their speeches. For example, of the interrupted speeches of MP Annika Saarikko (Centre Party), the Minister of Finance, 46% are due to MP Ben Zyskowitz, representing the National Coalition Party in opposition, and 18% to MP Jukka Gustafsson representing the social democratic party SDP in the government, indicating possibly different opinions inside the government.

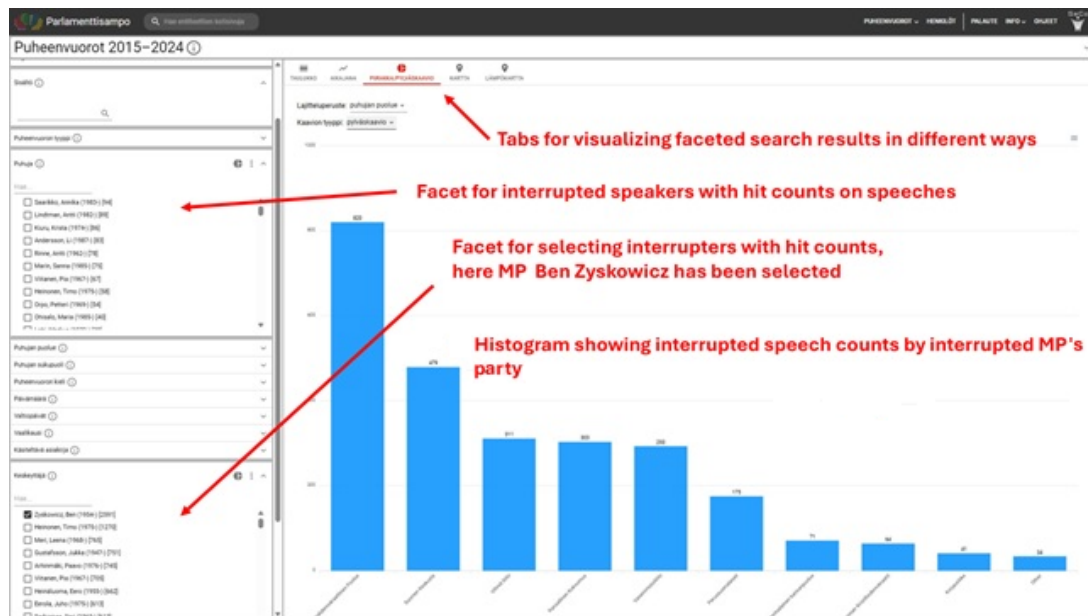


Figure 2: The 2591 interruptions 2015–2024 of Ben Zyskowitz analyzed by interrupted MP and on a party level.

In Fig. 2, the faceted search integrated with data analytic tools in the portal is used. Here the period 2015–2024 is considered, Ben Zyskowitz is selected from the facet Interrupters, and the search result (i.e, interrupted speeches) is visualized using the histogram chart tab, but in this case on a party level.

Speeches that provoked laughter can be analyzed in a similar way. In Fig. 3 the period 2015–2024 is considered during which the speech data were available born digital and of best quality. For doing the analyzes, the speech application perspective for searching and analyzing speeches was extended with a new facet “Laughter for the speech” (“Naurua ja hilpeyttä puheelle” in Finnish) with binary values: “Laughter” and “No laughter”. By selecting Laughter, it can be seen that out of the total of 178 879 speeches in 2015–2024, 1985 (1.1%) speeches were laughed at. Of the total speeches made by male speakers, 1.6% were laughed at, compared with 0.6% of speeches made by female speakers, so, proportionally, men seem to make over 2.5 times as many funny speeches as women. There are also substantial differences on a party level: 1.3% of the speeches given by MPs of the Centre Party were laughed at while only 0.5% of the speeches by members of the obviously more serious Greens party were laughed at.

5. Conclusions

This paper proposed novel cases for DH research of parliamentary debates: analyzing interruptions and laughter in parliamentary speeches. As a demonstration, new extensions implemented in the in-use

⁶Sampo portal series: <https://seco.cs.aalto.fi/applications/sampo/>

⁷Code on GitHub <https://github.com/SemanticComputing/sampo-ui>; project homepage <https://seco.cs.aalto.fi/tools/sampo-ui/>

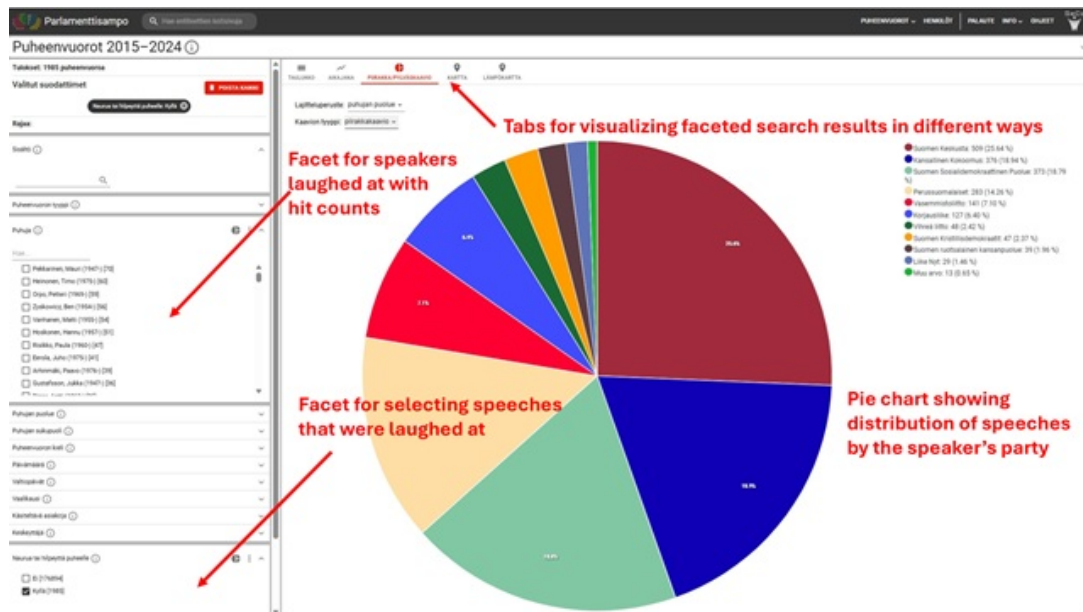


Figure 3: Laughter speeches in the PoF selected and visualized on the pie chart tab on a party level.

semantic portal PARLIAMENTSAMPO were presented.

Acknowledgments

Support of the Finnish Digital Humanities research infrastructure initiative FIN-CLARIAH/DARIAH-FI⁸ funded by the Research Council of Finland and NextConnectionEU is acknowledged. Computational resources of CSC – IT Center for Science have been used.

Declaration on Generative AI

During the preparation of this work, the author(s) used Writefull in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] E. Hyvönen, L. Sinikallio, P. Leskinen, S. Drobac, R. Leal, M. La Mela, J. Tuominen, H. Poikkimäki, H. Rantala, Publishing and using parliamentary linked data on the semantic web: ParliamentSampo system for Parliament of Finland, *Semantic Web* (2025) 1–23. doi:10.3233/SW-243683.
- [2] L. Sinikallio, S. Drobac, M. Tamper, R. Leal, M. Koho, J. Tuominen, M. L. Mela, E. Hyvönen, 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, Schloss Dagstuhl- Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, 2021, pp. 1–17. URL: <https://drops.dagstuhl.de/opus/volltexte/2021/14544/pdf/OASlcs-LDK-2021-8.pdf>.
- [3] P. Leskinen, E. Hyvönen, J. Tuominen, Members of Parliament in Finland knowledge graph and its linked open data service, in: *Further with Knowledge Graphs. Proceedings of the 17th International Conference on Semantic Systems*, 6–9 September 2021, Amsterdam, The Netherlands, IOS Press, 2021, pp. 255–269. doi:10.3233/SSW210049.

⁸FIN-CLARIAH/DARIAH-FI: <https://seco.cs.aalto.fi/projects/fin-clariah/>

- [4] T. Wissik, Encoding interruptions in parliamentary data: From applause to interjections and laughter, *Journal of the Text Encoding Initiative* (2021). doi:10.4000/jtei.4214.
- [5] M. La Mela, F. Norén, E. Hyvönen (Eds.), *Digital Parliamentary Data in Action (DiPaDA 2022): Introduction*, volume 3133, CEUR WS, 2022. URL: <http://ceur-ws.org/Vol-3133/paper00.pdf>.
- [6] D. Fišer, M. Eskevich, J. Lenardič, F. de Jong (Eds.), *Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, 2022. URL: <https://aclanthology.org/2022.parlaclarin-1.0>.
- [7] A. Van Aggelen, L. Hollink, M. Kemman, M. Kleppe, H. Beunders, The debates of the European Parliament as Linked Open Data, *Semantic Web* 8 (2017) 271–281. doi:10.1007/s42001-019-00060-w.
- [8] U. Bojārs, R. Dargis, U. Lavrinovičs, P. Paikens, LinkedSaeima: A linked open dataset of Latvia’s parliamentary debates, in: *Semantic Systems. The Power of AI and Knowledge Graphs. SEMANTICS 2019*, Springer–Verlag, 2019, pp. 50–56. doi:10.1007/978-3-030-33220-4_4.
- [9] R. Bleier, F. Zeilinger, G. Vogeler, From early modern deliberation to the semantic web: Annotating communications in the records of the Imperial Diet of 1576, in: M. La Mela, F. Norén, E. Hyvönen (Eds.), *Proceedings of the Digital Parliamentary Data in Action (DiPaDA 2022) Workshop co-located with 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNb 2022)*, volume 3133, CEUR WS, 2022, pp. 86–100. URL: <http://ceur-ws.org/Vol-3133/paper06.pdf>.
- [10] K. Kettunen, M. La Mela, Semantic tagging and the nordic tradition of everyman’s rights, *Digital Scholarship in the Humanities* 37 (2021). doi:10.1093/llc/fqab052.
- [11] M. Andrushchenko, K. Sandberg, R. Turunen, J. Marjanen, M. Hatavara, J. Kurunmäki, T. Nummenmaa, M. Hyvärinen, K. Teräs, J. Peltonen, J. Nummenmaa, Using parsed and annotated corpora to analyze parliamentarians’ talk in Finland, *Journal of the Association for Information Science and Technology* 185 (2021) 1–15. doi:10.1002/asi.24500.
- [12] S. Simola, A century of partisanship in Finnish political speech, 2020. URL: <https://sites.google.com/site/sallasimolaecon/home/research>.
- [13] K. Makkonen, P. Loukasmäki, Eduskunnan täysistunnon puheenaiheet 1999–2014: Miten käsitellä LDA-aihemalleja?, *Politiikka* 61 (2019) 127–159.
- [14] E. Ikkala, E. Hyvönen, H. Rantala, M. Koho, Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces, *Semantic Web* 13 (2022) 69–84. doi:10.3233/SW-210428.
- [15] H. Rantala, A. Ahola, E. Ikkala, E. Hyvönen, How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework, in: *VOILA! 2023 Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs 2023*, CEUR Workshop Proceedings, Vol. 3508, 2023. URL: <https://ceur-ws.org/Vol-3508/paper3.pdf>.
- [16] E. Hyvönen, L. Sinikallio, P. Leskinen, M. La Mela, J. Tuominen, K. Elo, S. Drobac, M. Koho, E. Ikkala, M. Tamper, R. Leal, J. Kesäniemi, Finnish parliament on the semantic web: Using ParliamentSampo data service and semantic portal for studying political culture and language, *Digital Humanities in the Nordic and Baltic Countries Publications: Digital Parliamentary data in Action (DIPADA 2022)*, Workshop at the 6th Digital Humanities in Nordic and Baltic Countries Conference 4 (2022). doi:10.5617/dhnbpub.11261.