

# Riksdagsdebatter.se – An Interface to Explore Swedish Parliamentary Debates Since 1867

Johan Jarlbrink<sup>1</sup>, Fredrik Mohammadi Norén<sup>2</sup>, Roger Mähler<sup>1</sup>, Kajsa Palm<sup>1</sup> and Rebecka Weegar<sup>1,\*</sup>

<sup>1</sup>Umeå University

<sup>2</sup>Malmö University

## Abstract

This paper introduces version 1.0 of riksdagsdebatter.se, a freely available and user-friendly interface for searching, filtering, and exploring Swedish parliamentary speeches since 1867. The underlying corpus is based on open-access data from the Swedish Parliament, further processed and enriched within the SWERIK project, including re-OCRing the debate records, segmenting the text into individual speeches, and adding structured metadata, with ongoing improvements to data quality. The corpus currently comprises over one million annotated speeches, totaling approximately 450 million words, and offers a comprehensive record of political debate in modern Swedish history. This resource has wide-ranging applications for researchers and students in digital humanities, history, political science, and sociology, as well as for journalists, policymakers, and the general public. Version 1.0 of riksdagsdebatter.se provides four primary tools: word trends, keyword in context, n-gram analysis, and the ability to create and download filtered sub-corpora.

## Keywords

parliamentary debates, Sweden, text analysis tools, user interface, SWERIK

## 1. Introduction

The time is ripe to study Swedish parliamentary history in new and data-driven ways. The long-lived and transparent nature of the parliament has accumulated an abundance of documents, many of which have only recently been digitized. However, unlocking the full potential of these data sources requires easy-to-use tools for researchers without programming skills. This paper presents version 1.0 of a freely available and user-friendly interface – riksdagsdebatter.se – with the aim of providing tools to search, filter, and explore the Swedish parliamentary speeches since 1867. This rich dataset, documenting a vast number of political debates in Sweden, has many use cases and users, both researchers and students in the humanities and social sciences, as well as journalists, officials, politicians, and the general public. The project behind the interface, SweDeb (Swedish Parliamentary Debates), was launched in 2023 as a local infrastructure project, with initial funding from Umeå University. Humlab, the digital humanities hub at the university, runs the main development and maintenance of the interface. Users of riksdagsdebatter.se are encouraged to cite this paper as a reference in their work.

## 2. The Data Behind riksdagsdebatter.se

The annotated speech data that riksdagsdebatter.se provides access to have been produced by the research infrastructure project SWERIK (Swedish Riksdag 1867–2022: An Ecosystem of Linked Open Data, 2023–2028), funded by Riksbankens Jubileumsfond (IN22-0003) and led by Fredrik Mohammadi Norén (SWERIK, in turn, builds on work initiated in the research project WeStAc, 2019–2024, led by Pelle Snickars). The purpose of SWERIK has been, first, to build a speaker database of all Swedish members

---

*The 9th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2025), March 5–7, 2025, Tartu, Estonia.*

\*Corresponding author.

✉ johan.jarlbrink@umu.se (J. Jarlbrink); fredrik.noren@mau.se (F. Mohammadi Norén); roger.mahler@umu.se (R. Mähler); kajsa.palm@umu.se (K. Palm); rebecka.weegar@umu.se (R. Weegar)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



Digital Humanities in the Nordic and Baltic Countries Publications – ISSN: 2704-1441

of parliament (MPs) and other speakers in the parliament (such as ministers) since the establishment of the bicameral system in 1867, with metadata for, among other things, party, gender, and time in parliament. Second, SWERIK has linked speakers to relevant parliamentary documents: the private members' motions (*motioner*), the government bills (*propositioner*), the committee reports (*utskotts-betänkanden*), and – particularly relevant in the context of riksdagsdebatter.se – the parliamentary records (*riksdagsprotokoll*).

The parliamentary records document all speeches delivered in parliament, along with accounts of decisions and activities that take place during each parliamentary session (*riksmöte*). There are three types of sub-records: the body record (*stomprotokollet*), the debate record (*debattprotokollet*), and the preliminary record (*snabbprotokollet*). The body record includes information such as attendance lists, agendas, voting outcomes, and written questions submitted by MPs to ministers. The debate record comprises transcriptions of spoken contributions in the chamber (Nygren 1985). These transcriptions, however, undergo editorial processing, meaning they do not offer verbatim representations of the debates (Harvard 2011). Typically, the body and debate records are merged in chronological order to produce the official printed parliamentary record. The preliminary records, which serve as early drafts of the finalized versions, are excluded from the SWERIK corpus.

SWERIK has annotated two preliminary records: speech text and non-speech text, the latter divided into speaker introductions and other textual elements. Speaker introductions typically include the speaker's name and may also contain their title and party affiliation: "Herr statsministern PALME" (Mr Prime Minister [Olof] Palme), or "EWA LARSSON (mp)" (Ewa Larsson, Green Party). The information about the person in the speaker introduction is used to map the right speaker to their corresponding speeches through SWERIK's speaker database (Yrjänäinen et al. 2024). In total, SWERIK's version 1.3 of the speech corpus contains more than one million annotated speeches and a database of more than six thousand unique speakers.<sup>1</sup> The annotated speeches and the system of links between speeches and speakers constitute the main data structure behind riksdagsdebatter.se.

## 2.1. The Structure and Tools of riksdagsdebatter.se

Thanks to various mass digitization projects, the debate records of many national parliaments are now available in digital formats. This development has, in turn, spurred the creation of various user interfaces for exploring parliamentary speech data. When developing riksdagsdebatter.se, the SweDeb project has drawn inspiration from several state-of-the-art platforms, including those for the British Parliament ([hansard.parliament.uk](https://hansard.parliament.uk)), the Finnish Riksdag ([parlamenttisampo.fi](https://parlamenttisampo.fi)), the German Bundestag ([opendiscourse.de](https://opendiscourse.de)), and the Canadian Parliament ([lipad.ca](https://lipad.ca)).

The interface riksdagsdebatter.se consists of two main parts, a back-end API (application programming interface) implemented in Python ([python.org](https://python.org)), and a front-end implemented using Vue3 ([vuejs.org](https://vuejs.org)). Furthermore, SweDeb is a project based on principles of open science and its source code is openly available on GitHub.<sup>2</sup> Since the main user groups of riksdagsdebatter.se are based in Sweden, the language of the interface is Swedish. Although version 1.0 of riksdagsdebatter.se is limited to tools for exploring Swedish parliamentary debates, the SWERIK data structure allows for the future integration of additional MP-linked document types, including motions, government bills, and committee reports, into the interface. The principles behind riksdagsdebatter.se might also be used as a basis to develop similar interfaces for exploring other countries' parliamentary data.

Besides three introductory pages (landing page, about the project, and frequently asked questions), riksdagsdebatter.se consists of four tool pages: word trends, keyword in context (KWIC), n-gram analysis, and a tool to create sub-corpora of the speech data. Each tool allows the user to filter speeches based on year, political party represented in the parliament, gender, chamber, and individual speakers. A brief description of each interface tool is found below. Both the KWIC tool and the n-gram analysis tool are based on the corpus analysis tool *Corpus Workbench* (Evert and Hardie 2011).

---

<sup>1</sup>All code, data, and version history of the SWERIK project are open access and can be found here: <https://github.com/swerik-project> (accessed 23 January 2025).

<sup>2</sup><https://github.com/humlab-swedeb> (accessed 23 January 2025).

**Word trends** enables the user to search for one or several terms and display how the frequency of these terms has changed over time. The tool displays word trend results as a trendline graph and as a table. There is an option to show the total frequencies of the word trends and normalized frequencies. The tool also provides a list of all annotated speeches containing the search terms, presented as a table with information about the record, the year, and the speaker (including name, gender, and party). If the user clicks on a speech in the list, the full annotated speech text is displayed. In each speech, there are four links. The first directs the user to the Wikidata page of the speaker (SWERIK's speaker database is built through Wikidata). The second is a link to the PDF image of the debate record page where the specific speech can be found. The third is an option to download the speech. The fourth link gives the user the opportunity to file an error report regarding the speech data (e.g., concerning OCR, metadata, and segmentation) and directs the user to SWERIK's GitHub environment. The tool also provides an option to download the graph, the table, and the selected speeches.

**KWIC** allows the user to search for a term or phrase and see the contexts in which these occur. The size of the context is decided by the user through a word window of up to 20 words to the left or to the right of the search target. In contrast to the word trends tool, it is possible to lemmatise selected terms or phrases, which means that the search result includes all occurrences of a term with the same grammatical root. The search result is presented as a list in which the search target and its left and right contexts are displayed. As with the word trends tool, it is possible to click on a speech to be able to read its full text, with the same functionality as previously described, including downloading the KWIC contexts.

**N-gram** is similar to KWIC and facilitates a way to search for a term or a phrase and see their immediate contexts. An n-gram is a sequence of words where the n represents the number of words included in the sequence. In the tool, it is possible to search for n-grams of up to 20 words. The n-gram result is presented as a descending list of the most common word contexts (n-grams) containing the search target. In the result, it is possible to click on a specific n-gram and get a list of all the speeches containing that specific n-gram. For this tool, it is also possible to click on a speech, read it in full, and download the selected n-grams as well as the speeches.

**The tool for creating sub-corpora of speeches** enables users to download a customized subset of speeches using filters for year, political party represented in the parliament, gender, and individual speakers. Based on the chosen filters, the result is presented in a list of speeches, with the same options as the speech lists in the other tools. The user is also able to download the filtered sub-corpus of speeches.

### 3. Concluding Remarks and Future Directions

This paper presents version 1.0 of riksdagsdebatter.se, a user-friendly interface with the ambition to provide tools to explore the Swedish parliamentary speeches since 1867. While riksdagsdebatter.se enables users to explore word trends, KWIC, n-grams, as well as a tool to create a sub-corpus of the speech data, the possibilities to develop the interface are many. For example, other tools can be integrated, such as collocate analysis, topic modelling, and word embeddings. Moreover, based on the output and data updates from the SWERIK project, other Swedish parliamentary document categories can be added to and explored through riksdagsdebatter.se with the same linking structure to connect speakers to their respective relevant documents, such as motions, government bills, and committee reports.

**Acknowledgements:** Creating version 1.0 of riksdagsdebatter.se has been supported by Umeå University (FS 1.3.2-2339-22), Humlab at Umeå University, and Huminfra ([huminfra.se](http://huminfra.se)). An extra word of thanks to Marita Nilsson who was initially part of the project team. The authors would also like to thank the members who have been part of the reference group for valuable support: Hanna Bäck (Lund University), Josefin Hägglund (Malmö University), Simon Lindgren (Umeå University), Måns Magnusson (Uppsala University), Claes Ohlsson (Linnaeus University), Leif-Jöran Olsson (Gothenburg University), and Linda Sandström (Umeå University).

## References

- Evert, Stefan, and Andrew Hardie. 2011. “Twenty-first century corpus workbench: updating a query architecture for the new millennium.” In *Proceedings of the Corpus Linguistics 2011 Conference*, 1–21. Birmingham: University of Birmingham.
- Harvard, Jonas. 2011. “Riksdagsprotokollen som medium.” In *Dolt i offentligheten: Nya perspektiv på traditionellt källmaterial*, edited by Staffan Förhammar, Jonas Harvard, and Dag Lindström, 25–42. Lund: Sekel Bokförlag.
- Nygren, Rolf. 1985. “Det svenska riksdagstrycket.” In *Handbok i Nordiskt parlamentstryck*, edited by Rolf Nygren, 109–137. Stockholm: Sveriges Riksdag.
- Yrjänäinen, Väinö, Fredrik Mohammadi Norén, Robert Borges, Johan Jarlbrink, Lotta Åberg Brorsson, Anders P Olsson, Pelle Snickars, and Måns Magnusson. 2024. “The Swedish parliament corpus 1867–2022.” In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 16100–16112. <https://aclanthology.org/2024.lrec-main.1400>.