Toxic Speech in the Parliament of Finland: Observing harmful language in Finland's political institutions

Ahola, Annastiina (711551)
Guevara, Gabriel (103148953)
Lindén, Wilma (886389)

Statements of teamwork and use of generative AI are included in the last appendix (Appendix I)

1 Introduction

We live in an increasingly polarized world, and the burden on authorities to prevent the dangers of division is high. Political speech is a key factor in this polarization and the way ideas are conveyed through language can have a huge impact on these political divisions. Using Lynne Tirrell's definition of toxic speech as a potentially harmful and highly reproducible type of discourse [9], it is important to reflect and analyze how prevalent this type of speech is within one of the most powerful entities in society, the government institutions. We aim our magnifier glass over the Finnish Parliament to study and understand the language they use and, more specifically, the toxicity in their political language.

ParliamentSampo [2, 6] Linked Open Data (LOD) service and semantic portal contain data on the plenary speeches of the Parliament of Finland from 1907 to the present day, totaling over a million speeches, as well as on the members of the parliament. The aforementioned speeches have been a topic of research for topic modeling [5] and sentiment analysis [8, 4], but—as far as we know—not for toxicity analysis of the contents of the speeches.

The ParliamentSampo Knowledge Graph (KG) offers rich metadata about the speeches in addition to the content itself, which makes it ideal for studying possible correlations between speech toxicity and demographic factors like the speaker's gender or party affiliation. In addition to this more basic metadata, the more recent speeches from 2015 onward have had named entities recognized and linked from the speeches with NLP methods [7], lending the KG to also be used for network-based analysis as in [3], for example. Because of these enrichments, our work will focus on the two completed electoral terms from 2015, *Electoral term 2015–2018* (Sipilä cabinet, 22.04.2015–16.04.2019) and *Electoral term 2019–2022* (Rinne and Marin cabinets, 17.04.2019–04.04.2023).

Our project aimed to answer the following research questions:

1. What are the most common types of toxicity (*toxicity*, *obscene*, *insult*, *threat*, *identity attack*, *severe toxicity*) present in the speeches?
   - Are there some time periods in which the toxicity was higher or lower than average? What could the cause be (e.g., a specific event)?
2. Does the gender of the speaker correlate with the toxicity of the speech?
3. Does the speaker's party affiliation correlate with the toxicity of the speech?
   - Does the position of the party also matter (opposition vs. government) for the parties that were in different roles in the two studied electoral terms?
4. Does the length of the speaker's experience (how long they've been a member of the parliament) correlate with the toxicity of the speech?
5. Is there any correlation between the number of interruptions and toxicity or mentions of other people and toxicity?

2 Data & methodology

2.1 Data extraction

The data was extracted from ParliamentSampo's SPARQL endpoint[1] by writing queries in the YASGUI editor[2] and downloading all the results in CSV format. To limit the number of results to a more manageable number of speeches, the restrictions for the extracted speeches were the following:

- All speeches had to be completely in Finnish[3].
- All speeches had to be either of type *varsinainen puheenvuoro* or *vastauspuheenvuoro*[4].
- All speeches had to be from the completed electoral terms of either *Electoral term 2015–2018* or *Electoral term 2019–2022*.
- All speeches had to have a linked speaker entity.

Some speeches in the original data were missing information regarding the speaker's party if the speaker had changed their party affiliation in the middle of an electoral term. For these speakers, their parliamentary group membership information was manually extracted from the Parliament of Finland's official site[5] and the metadata on the speaker's party and their role (opposition vs. government) was updated based on the date of speech and the extracted membership dates.

The text contents of the speeches also include possible interruptions or reactions from the audience (e.g., laughing) inside brackets (round, square or a mix of the two). These were cleaned out utilizing regular expressions to only get the speech content actually spoken by the marked speaker. To ensure that the speeches could be passed to the toxicity detection model discussed in the next section, the speeches were further narrowed by length to fit in the model's maximum sequence length. This meant that the final number of processed speeches was **77,868** (see Figure 1 for timeline). Some statistics on these speeches are included in Appendix F and Appendix G.

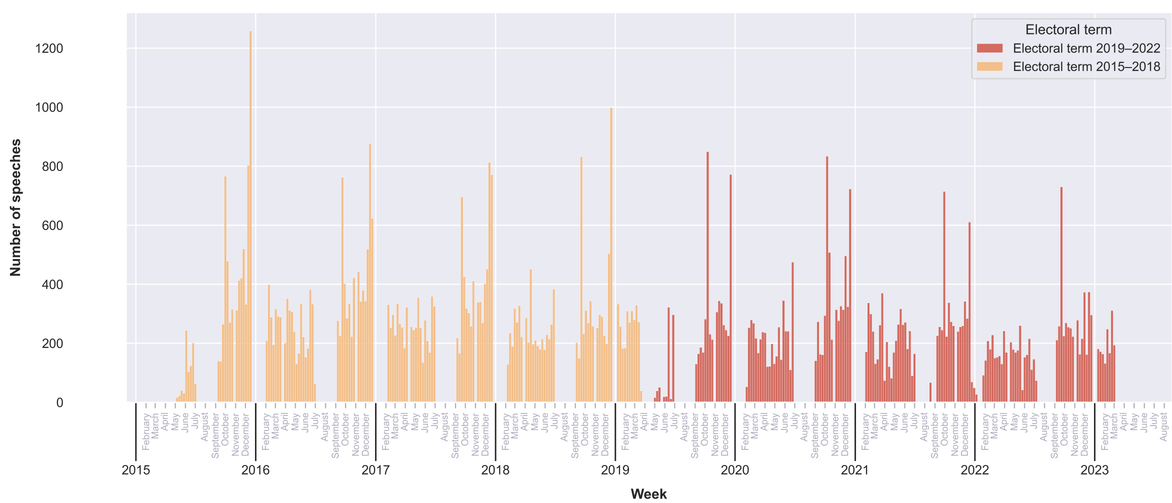

Figure 1: The weekly number of speeches on a timeline

For Research Questions 4 and 5, data on the number of electoral terms each speaker has been a part of as well as data on people referenced in speeches and interruptions of speeches were also extracted from the endpoint. All the queries used to extract the data from the SPARQL endpoint are included in Appendices A–E.

[1] https://ldf.fi/sempar1/sparql
[2] https://yasgui.triply.cc/
[3] Excluding speeches containing *both* Finnish and Swedish according to the metadata
[4] This excludes, e.g., speeches by the chairman of the parliament or speeches spoken in behalf of a group
[5] https://www.eduskunta.fi/FI/Sivut/default.aspx

2.2 Toxicity & data analysis

Toxicity calculation was done using TurkuNLP's `bert-large-finnish-cased-toxicity` model[6] for Finnish toxicity detection [1] as it was one of the two toxicity models available for Finnish on the HuggingFace platform[7] and had Finnish-specific evaluation results available. It produces toxicity scores for *toxicity*, *obscene*, *insult*, *threat*, *identity attack* and *severe toxicity* categories[8].

The text content for each of the speeches in the extracted data was passed to the model with the help of HuggingFace's `Transformers`[9] library and the results were saved to an array alongside the URI for the speech and text content. These toxicity scores were then combined with speech-related metadata and speaker-related metadata extracted from the ParliamentSampo SPARQL endpoint into a JSON file for data analysis. Data analysis on the extracted and generated data was performed using Jupyter Notebooks and Python. The primary libraries used to analyze the data were `pandas`[10] and `numpy`[11]. Data visualizations were created using `matplotlib`[12] and `seaborn`[13].

3 Results

3.1 RQ1

The most prevalent toxicity type is, as expected, toxicity (see Table 1 and Figure 2). For all speeches, the mean toxicity is roughly .003, which is more than three times larger than the mean value for the second most prevalent toxicity type (obscene). Furthermore, the maximum value for toxicity is roughly .985, which indicates a speech that is almost fully toxic, while the second largest maximum value (obscene) is roughly .191, which indicates a speech that is around one fifth obscene.

	Mean	Median	Std	Max	Min
toxicity	0.00294	0.00257	0.00794	0.98557	0.00241
obscene	0.00079	0.00075	0.00138	0.19176	0.00074
insult	0.00049	0.00049	0.00058	0.08097	0.00042
threat	0.00028	0.00027	0.00008	0.01382	0.00025
identity_attack	0.00021	0.00021	0.00015	0.02030	0.00016
severe_toxicity	0.00013	0.00014	0.00003	0.00427	0.00006

Table 1: Mean, median, standard deviation, max and min values for all toxicity types

For the five most prevalent toxicity types, the median is lower than or equal to the mean. This indicates that the distribution is somewhat skewed. Figure 3 illustrates how the 'most toxic dates'[14] tend to be toxic due to a low number of highly toxic speeches. There are exceptions, such as 12.12.2017, during which discussion topics were sexual harassment, and the government's annual report for 2016. Furthermore, Figure 4 shows a timeline of aggregate toxicity (aggregate mean toxicity per day), which shows that toxicity spikes consist of singular days that are evenly distributed throughout the timeline, rather than of longer periods, such as election season. Looking at the content of the five most toxic speeches, each contains some form of the word 'hävetä'[15]. Thus, it seems that the presence of certain toxic words has a high impact on the overall toxicity score of speeches. The `bert-large-finnish-cased-toxicity` model is also trained using online comments, which could significantly differ from transcripts of political speeches in terms of content, form, style, etc. Furthermore, transcripts do not convey the tone of voice or other forms of communication from the speaker, and thus might not fully capture the sentiment of speeches.

[6] https://huggingface.co/TurkuNLP/bert-large-finnish-cased-toxicity
[7] https://huggingface.co/
[8] Definitions follow Perspective API's annotation guidelines: https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages?language=en_US
[9] https://huggingface.co/docs/transformers/en/index
[10] https://pandas.pydata.org/
[11] https://numpy.org/
[12] https://matplotlib.org/
[13] https://seaborn.pydata.org/
[14] The dates with the highest average aggregate toxicity scores
[15] '[h]ävetkää', '[h]ävettää', 'hävetä', '[h]ävytöntä'

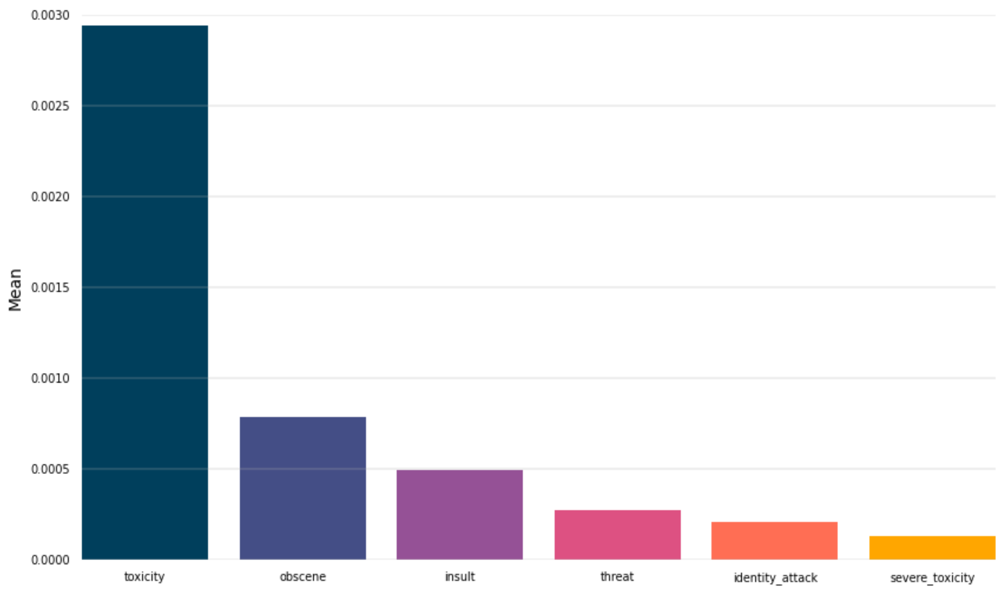

Figure 2: Bar plot showing mean values across all toxicity types

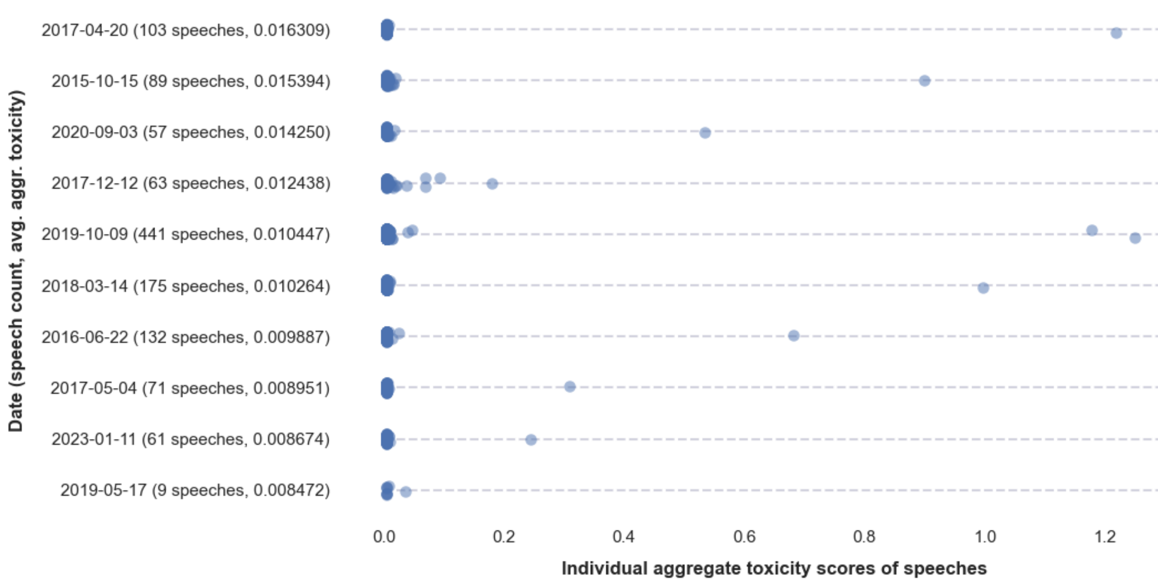

Figure 3: The distribution of aggregate toxicity scores for speeches from the 10 'most toxic dates'

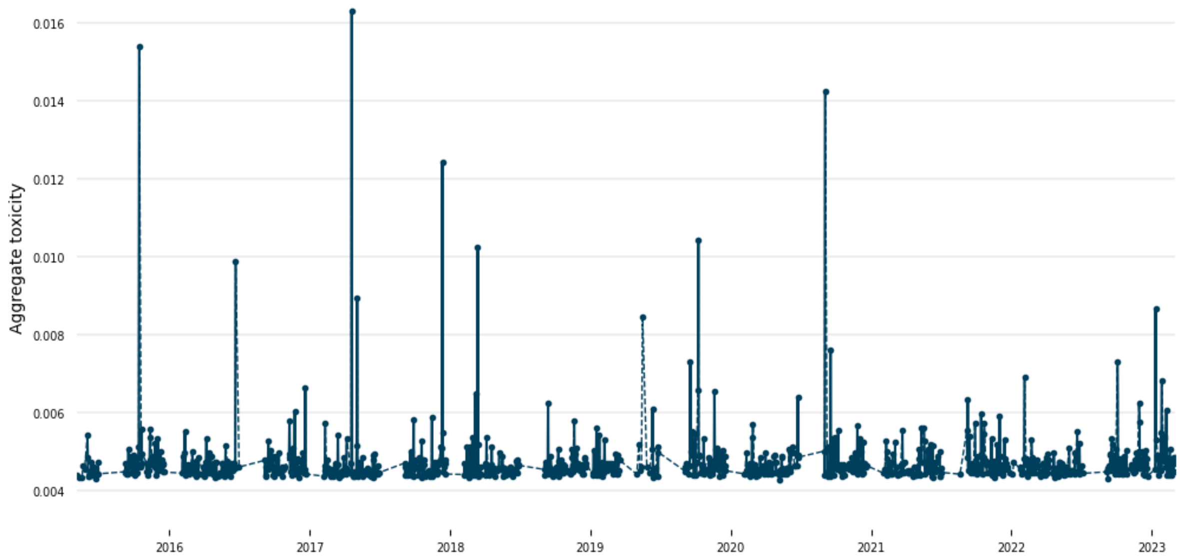

Figure 4: Timeline of aggregate toxicity (aggregate mean toxicity per day)

3.2 RQ2

Mean values are similar although slightly higher for males across all toxicity types (see Table 2 and Figure 5). Figure 6 illustrates the toxicity distribution for females and males. Males are more distributed across toxicity types, whereas females are either non-toxic or highly toxic. This might be due to randomness and sample size, or due to some difference in behavior. It is, however, worth noting that for females the total number of speeches is 33,772, whereas for males it is 44,096. Thus, statistical estimates for females may be less stable than for males and results should be interpreted with this in mind.

	Mean	**Median**	**Std**	**Max**	**Min**
F toxicity	0.00288	0.00257	0.00771	0.98557	0.00241
M toxicity	0.00300	0.00257	0.00811	0.90363	0.00241
F obscene	0.00078	0.00075	0.00126	0.19176	0.00074
M obscene	0.00079	0.00075	0.00146	0.17870	0.00074
F insult	0.00049	0.00049	0.00060	0.08097	0.00042
M insult	0.00049	0.00049	0.00056	0.06890	0.00042
F threat	0.00027	0.00027	0.00009	0.01382	0.00025
M threat	0.00028	0.00027	0.00007	0.00908	0.00025
F identity_attack	0.00021	0.00021	0.00015	0.02030	0.00016
M identity_attack	0.00021	0.00021	0.00015	0.01759	0.00016
F severe_toxicity	0.00013	0.00014	0.00003	0.00427	0.00006
M severe_toxicity	0.00013	0.00014	0.00002	0.00244	0.00006

Table 2: Mean, median, standard deviation, max and min values for females (F) and males (M)

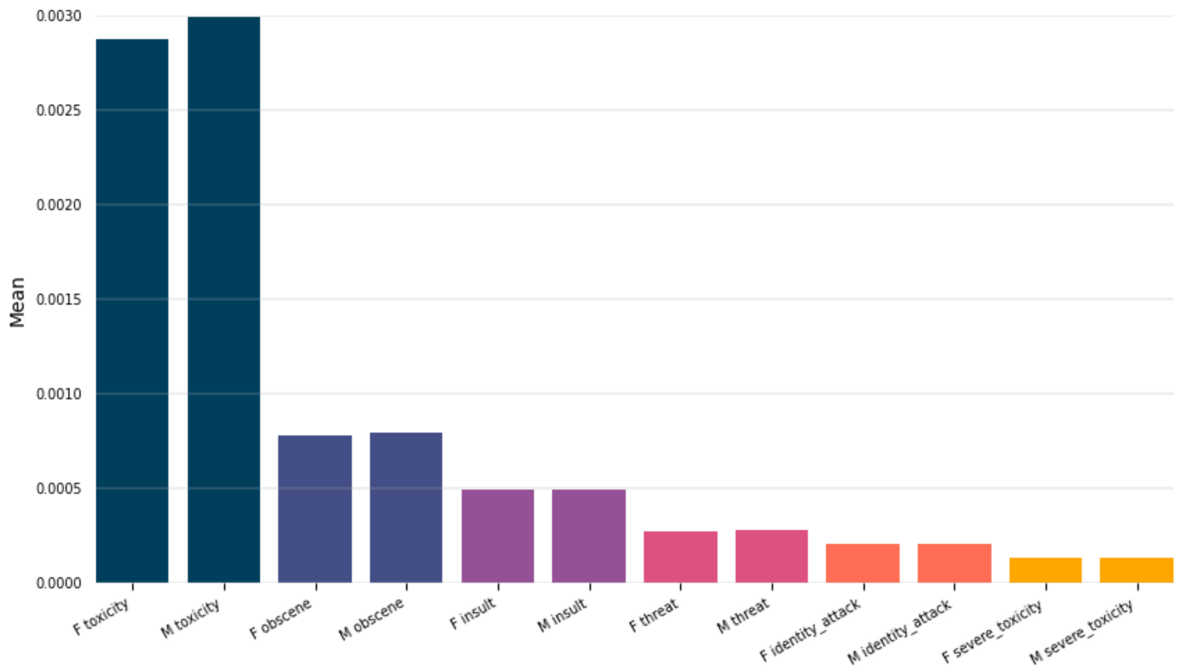

Figure 5: Bar plot showing mean values across all toxicity types for females (F) and males (M)

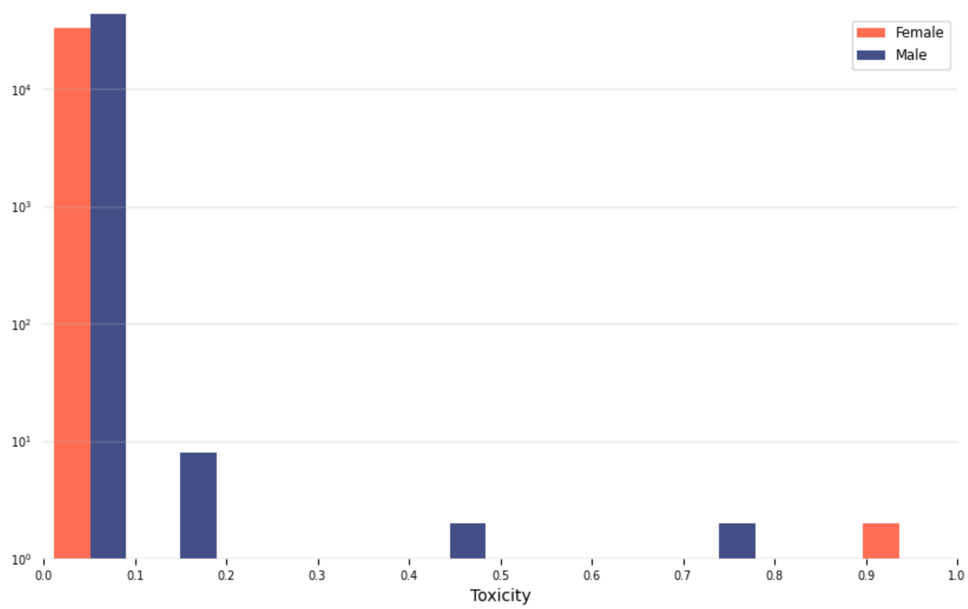

Figure 6: Histogram showing toxicity distribution for females and males

3.3 RQ3

A majority of the 10 most toxic speakers in terms of average aggregate toxicity have been affiliated with the Finns Party (see Table 3 and Figure 7). Overall, only Power Belongs to the People scores higher—very significantly in this case when compared to differences in toxicity for the rest of the parties—than the Finns Party in mean toxicity (see Table 4), its only member in the parliament being Ano Turtiainen—also the overall most toxic speaker on average—after being removed from the Finns party.

Speaker name	**Gender**	**All party affiliations**[16]	**2015–2018**	**2019–2022**	**Term count**[17]	**Avg. aggr. toxicity**
1. Turtiainen, Ano (1967–)	M	PS, VKK	-	X	1	0.010792
2. Tynkkynen, Sebastian (1989–)	M	PS	-	X	1	0.010657
3. Mäkynen, Jukka (1961–)	M	PS	-	X	1	0.009923
4. Guzenina, Maria (1969–)	F	SDP	X	X	4	0.009743
5. Biaudet, Eva (1961–)	F	RKP	X	X	6	0.009721
6. Raassina, Sari (1963–)	F	Kok.	X	-	1	0.008555
7. Hakkarainen, Teuvo (1960–)	M	PS	X	X	3	0.007818
8. Peltokangas, Mauri (1966–)	M	PS	-	X	1	0.007782
9. Laakso, Sheikki (1968–)	M	KaL, PS	-	X	1	0.007654
10. Packalén, Tom (1969–)	M	PS	X	X	3	0.007579

Table 3: The 10 most toxic speakers in terms of average aggregate toxicity

	Mean	**Median**	**Std**	**Max**	**Min**	**Count**
Seven Star Movement	0.00258	0.00252	0.00016	0.00313	0.00244	31
Christian Democrats	0.00265	0.00255	0.00051	0.01233	0.00241	3849
PERSONAL PARLIAMENTARY GROUP	0.00269	0.00254	0.00039	0.00412	0.00246	20
Centre Party	0.00271	0.00255	0.00338	0.41065	0.00241	15406
National Coalition Party	0.00283	0.00256	0.00465	0.51291	0.00241	13680
Green League	0.00284	0.00256	0.00355	0.24670	0.00241	6135
Left Alliance	0.00290	0.00258	0.00162	0.05134	0.00242	5763
Social Democratic Party of Finland	0.00292	0.00256	0.00989	0.94626	0.00241	15566
Finnish Reform Movement	0.00309	0.00257	0.01310	0.73963	0.00242	3318
Movement Now	0.00333	0.00266	0.00188	0.01557	0.00247	175
Swedish People's Party of Finland	0.00334	0.00256	0.02236	0.98557	0.00242	1963
Finns Party	0.00342	0.00262	0.01020	0.90363	0.00241	11797
Power Belongs to the People	0.00686	0.00276	0.01568	0.12143	0.00245	145

Table 4: Mean, median, standard deviation, max, min and count values for toxicity in parties

Every party that were in opposition and government during the two electoral terms have a higher mean toxicity during the opposition term (see Table 5). However, all parties also contributed more speeches during the opposition term. Thus, this might be a reason for the higher toxicity during opposition. Opposition parties might also criticize the government more heavily as a political strategy to gain more votes during the next election. Figure 8 shows the toxicity timeline for parties in opposition during the Sipilä cabinet and government during the Rinne and Marin cabinets. Figure 9 shows the toxicity timeline for parties in government during the first cabinet and in opposition during the second. Notable here is the increase in toxicity in the Finns Party after the party split in 2017 and the Finns Party moved into the opposition while the newly founded Finnish Reform Movement took their role in the government.

[16]The list includes *all* party affiliations from the KG; see Appendix H for a list of parties and the abbreviations used for them
[17]The number of electoral terms the speaker has been a part of by the end of the 2019–2022 electoral term

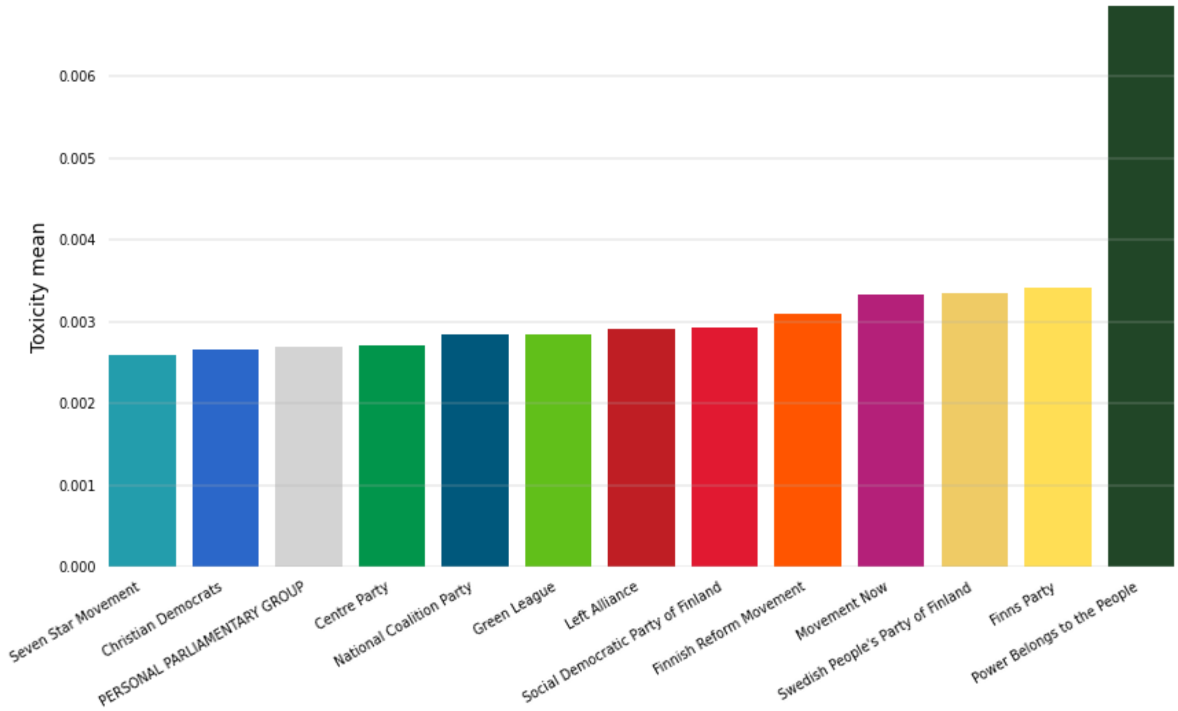

Figure 7: Bar plot showing mean toxicity values for all parties

	Govt Mean	Opp Mean	Govt Count	Opp Count
Green League	0.00274	0.00291	2498	3615
National Coalition Party	0.00280	0.00287	6663	6974
Social Democratic Party of Finland	0.00284	0.00298	6362	9171
Left Alliance	0.00285	0.00294	2092	3665
Swedish People's Party of Finland	0.00290	0.00365	811	1151
Finns Party	0.00314	0.00358	4054	7400

Table 5: Mean and count values for toxicity in parties that were in opposition and government during the two electoral terms

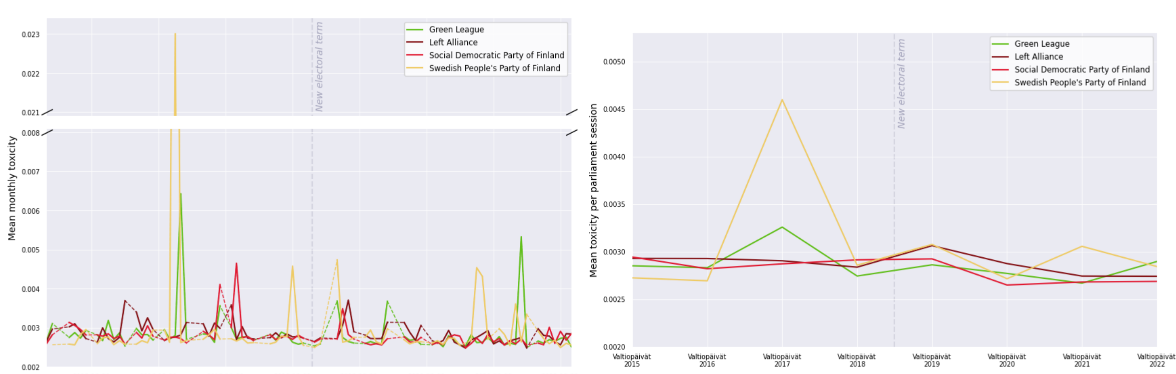

Figure 8: Timeline of mean toxicity grouped by month (left) and by parliamentary session (right) of parties first in opposition, then government.

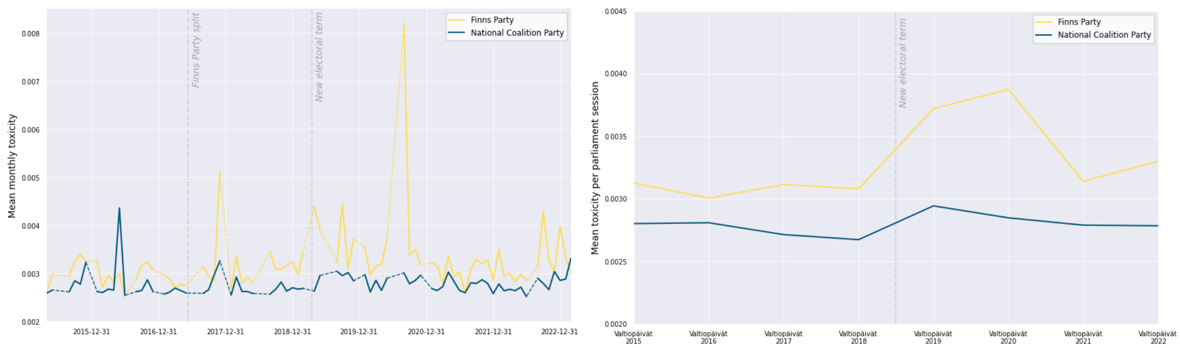

Figure 9: Timeline of mean toxicity grouped by month (left) and by parliamentary session (right) of parties first in government, then opposition.

3.4 RQ4

Initial analysis of the mean toxicity across each toxicity category for different numbers of terms, as presented in Figure 10, does not reveal a clear difference in toxicity levels related to term count. The variation in toxicity based on the number of terms appears minimal, and no discernible trend in toxicity values is evident in this initial representation.

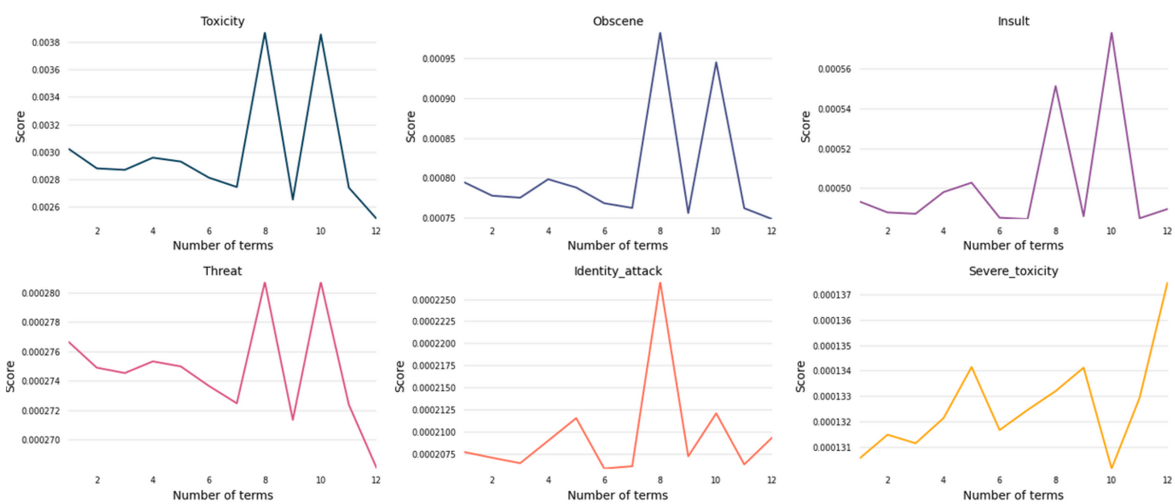

Figure 10: Small multiples of different toxicity types (mean) compared to number of terms.

However, a more defined pattern emerges when normalizing toxicity categories by the number of entries within each count of terms, as shown in Figure 11. The data indicates that toxicity remains stable throughout the term count with a slight decrease in general toxicity. However, we must address the spikes in the eighth term and the tenth term, which occur very likely due to a low sample of data. It is crucial to acknowledge that the significantly lower number of members of parliament in higher term counts may skew these data and affect the robustness of the observed trend (see Table 6).

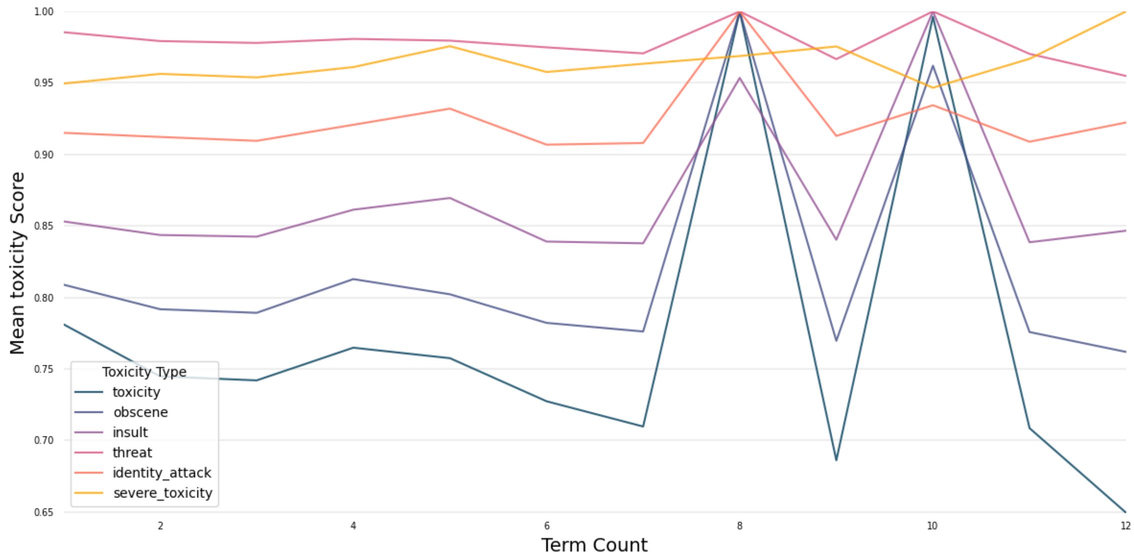

Figure 11: Line plot of different toxicity types (mean) divided by number of parliament members by term compared to number of terms normalized.

Term	Toxicity	Severe Toxicity	Obscene	Threat	Insult	Identity Attack	# Members
1	0.7811	0.9494	0.8088	0.9854	0.8531	0.9150	144
2	0.7447	0.9562	0.7917	0.9792	0.8436	0.9121	108
3	0.7419	0.9538	0.7892	0.9779	0.8424	0.9094	72
4	0.7648	0.9610	0.8127	0.9807	0.8613	0.9206	34
5	0.7574	0.9756	0.8021	0.9795	0.8694	0.9319	19
6	0.7273	0.9576	0.7821	0.9748	0.8390	0.9067	18
7	0.7096	0.9633	0.7761	0.9706	0.8378	0.9079	9
8	1.0000	0.9687	1.0000	0.9999	0.9534	1.0000	3
9	0.6860	0.9754	0.7696	0.9666	0.8402	0.9128	2
10	0.9962	0.9465	0.9619	1.0000	1.0000	0.9343	3
11	0.7085	0.9668	0.7757	0.9702	0.8385	0.9088	5
12	0.6493	1.0000	0.7618	0.9548	0.8466	0.9222	1

Table 6: Normalized toxicity values by number of terms

3.5 RQ5

The observed difference in toxicity values between interrupted and uninterrupted speeches is minimal, as seen in Figure 12. Although certain toxicity metrics, such as overall toxicity, exhibit slightly higher values in interrupted speeches, this variation is negligible. Furthermore, the unequal distribution and the resulting skewed representation of the data, as illustrated in Figure 13, preclude drawing firm conclusions regarding the toxicity of more frequently interrupted speeches. Consequently, the available data and analysis do not indicate a substantial difference in toxicity between speeches that were interrupted and those that were not.

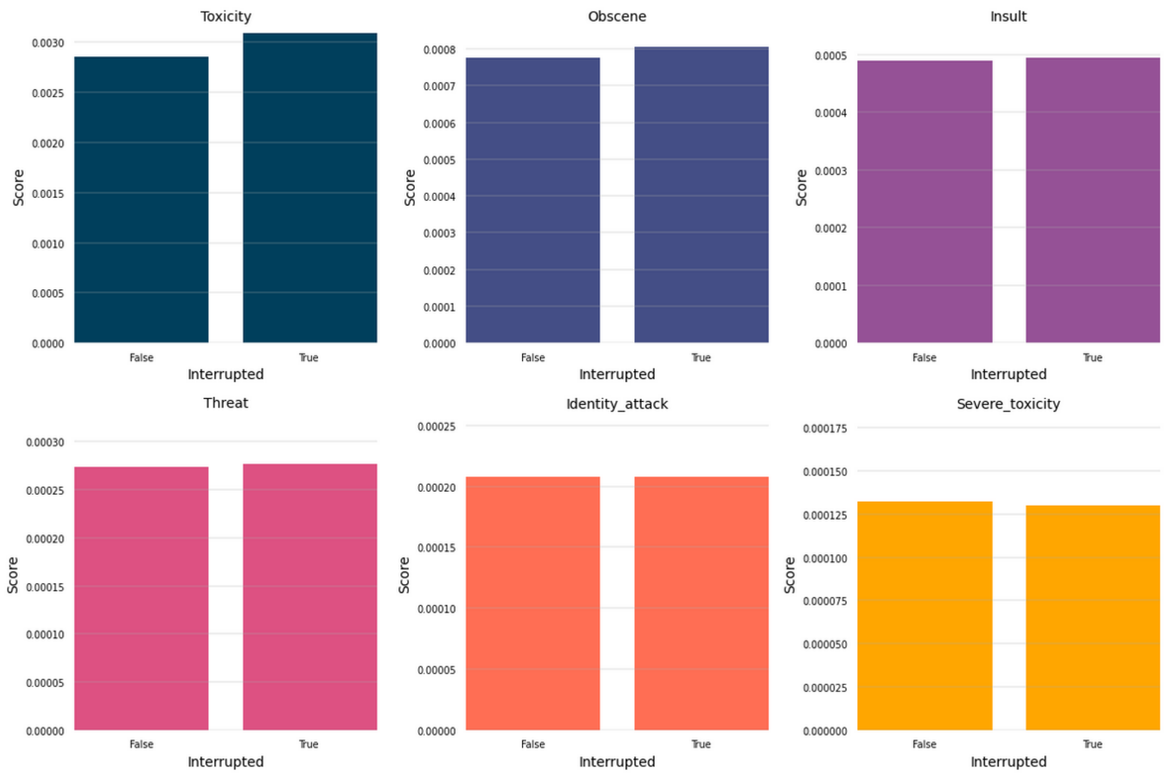

Figure 12: Bar chart comparing toxicity values between speeches that were interrupted and speeches that were not interrupted

A similar pattern emerges in the analysis of references. As shown in Figure 14, the differences in toxicity values are minimal. The unequal distribution of the data also renders the interpretation of potential visualizations, such as that in Figure 15, highly complex. Therefore, it can be concluded that there is no substantial difference in toxicity between speeches that included references to other parliament members and those that did not.

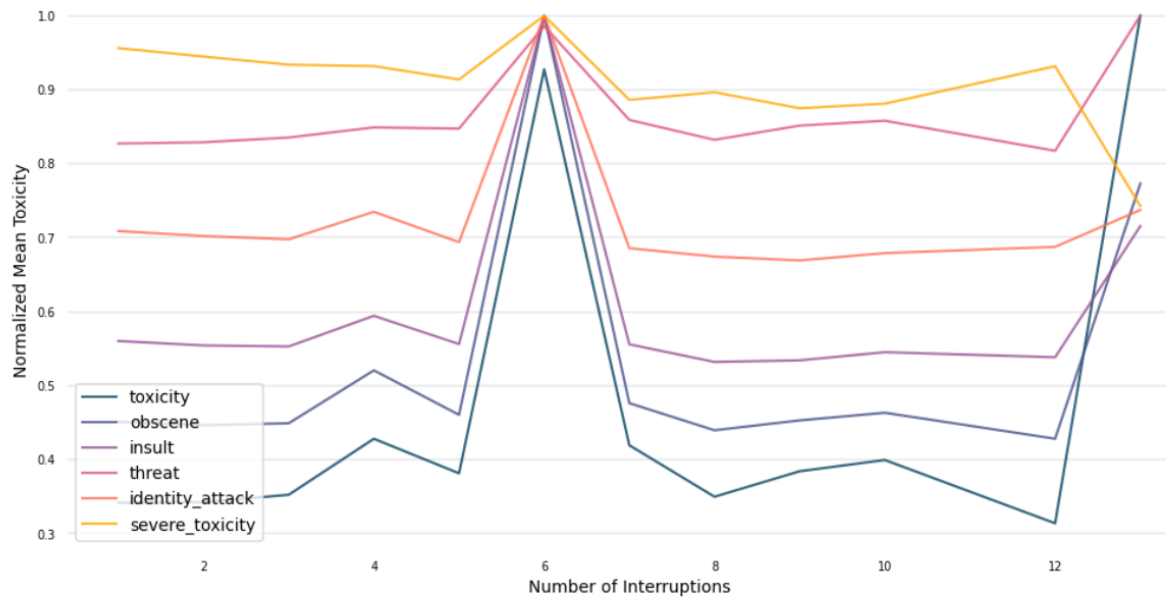

Figure 13: Line chart comparing different types of toxicity in speeches by number of interruptions

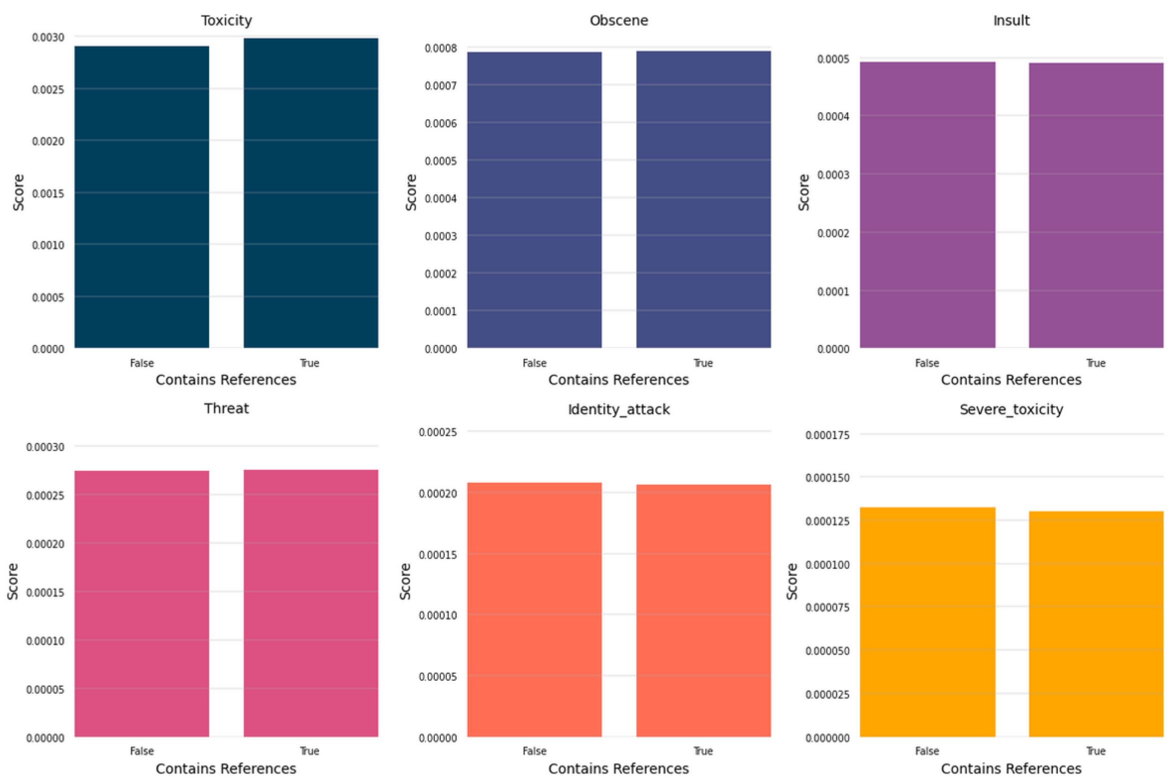

Figure 14: Bar chart comparing toxicity values between speeches that container references and speeches that did not contain references

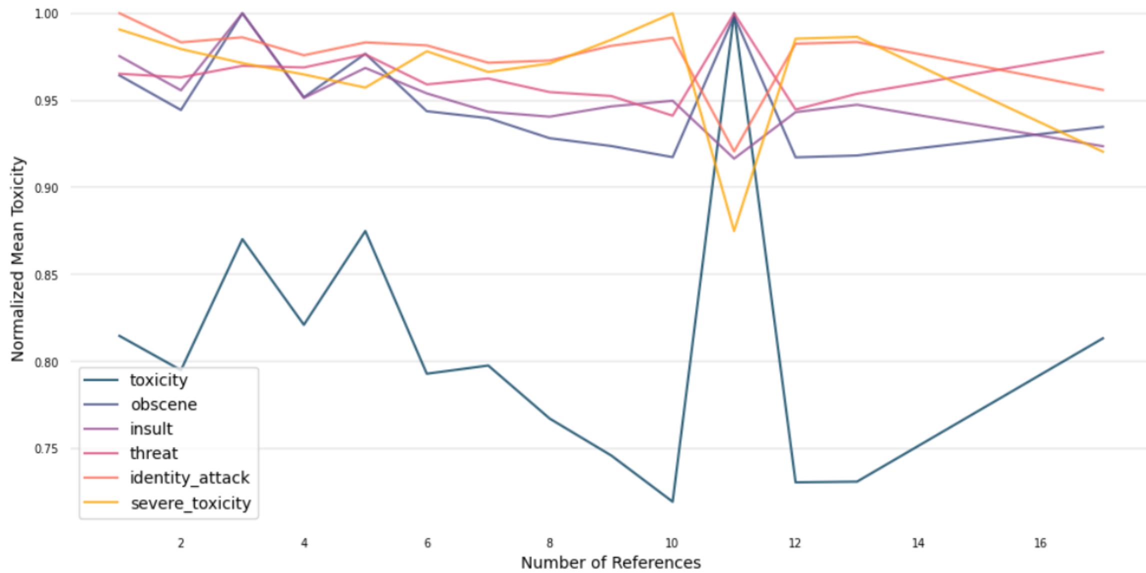

Figure 15: Line chart comparing different types of toxicity in speeches by number of references

4 Discussion & conclusions

Our aim was to study the prevalence of toxic speech and potential demographic factors that affect it in the speeches of the Parliament of Finland. The model we used for the toxicity analysis is trained on online comments as opposed to spoken speeches dealt with here. A written text possibly expresses tone in a more explicit manner in contrast to speeches where the actual tone of the voice already serves as a strong indicator of intended tone—something the model cannot account for. However, the results still present some interesting insights into how toxicity correlates with demographic factors.

From a temporal perspective, toxicity spikes seem to be brief, related, e.g., to the specific topic at hand, or even caused by a singular outlier speech that is flagged as extremely toxic by the model. The baseline toxicity level seems to stay relatively constant throughout the two examined electoral terms instead of being affected by things like a nearing election season, where one might expect parties to try to stand out from others.

For speaker demographics, male and female speakers on average have similar toxicity scores with male speakers being slightly more toxic in their speeches. Grouping speeches by the party affiliation of the speaker, on the other hand, leads to more variation in the mean toxicity scores with one party—Power Belongs to the People—leading significantly. This could indicate that the general party atmosphere and possible expectations of conduct (e.g., whether a party is seen as more professional or compassionate vs. unprofessional or combative) could possibly affect the speaker's toxicity or them becoming a member of that party in the first place. An interesting future avenue of research could be to look at whether switching parties seems to affect the mean toxicity values for speakers (e.g., Wille Rydman switching from the National Coalition Party to the Finns Party).

The number of terms does not seem to be strongly correlated with levels of toxicity. The unequal distribution of data and the low number of members in higher terms make it difficult to draw conclusions. However, some slight differences, such as a reduction in general toxicity, can be observed.

For future research, focusing on individual, long-standing members of the parliament could give better ideas about how a person's experience affects their toxicity. The analysis carried out in this project was limited to just two electoral terms, but instead the focus could be on *all* the speeches from a selected subset of people, who have large electoral term counts. For example, Ben Zyskowicz has a long career as a member of the parliament (10+ electoral terms) and is one of the most vocal members (and number one interrupter of speeches). It would be interesting to see whether the average toxicity level of his speeches has changed throughout his career in any way.

Speeches that were interrupted are generally slightly more toxic. However, this difference is negligible and does not add meaningful information to this investigation. The same phenomenon occurs when studying the toxicity of speeches that contained references to other parliament members.

The data has the contents of the interruptions as well and could provide some insights into, e.g., toxicity directed across parliament lines (opposition vs. government). However, this would require some cleaning and filtering of the data to remove interruptions without actual spoken content (e.g., *chairman knocks*) and their content could still be difficult to automatically score on a toxicity scale due to shortness.

Having analyzed the parameters of toxic speeches in the Parliament of Finland, we have a better understanding of the usage of hateful political discourse in this institution. Toxic speeches do exist in the Finnish parliament and might vary depending on timing, such as before elections or political party. We hope that the findings of this paper can be useful for future studies in Finnish political speech and possibly future research related to the effects of toxic speech.

References

[1] Eskelinen, A., Silvala, L., Ginter, F., Pyysalo, S., and Laippala, V. Toxicity detection in Finnish using machine translation. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)* (Tórshavn, Faroe Islands, May 2023), University of Tartu Library, pp. 685–697.

[2] Hyvönen, E., Sinikallio, L., Leskinen, P., Drobac, S., Leal, R., Mela, M. L., Tuominen, J., Poikkimäki, H., and Rantala, H. Publishing and using parliamentary Linked Data on the Semantic Web: ParliamentSampo system for Parliament of Finland. *Semantic Web 16*, 1 (2025), SW–243683.

[3] Poikkimäki, H., Leskinen, P., Tamper, M., and Hyvönen, E. Analyses of networks of politicians based on Linked Data: Case ParliamentSampo - Parliament of Finland on the Semantic Web. In *New Trends in Database and Information Systems* (August 2022), Springer International Publishing, pp. 585–592.

[4] Ristilä, A. Hot topics – Sentiment analysis on the plenary sessions of the Finnish parliament 1970-2020. In *Digital Humanities in Nordic and Baltic Countries (DHNB 2025). Book of Abstracts.* (March 2025). Poster paper.

[5] Ristilä, A., and and, K. E. Observing political and societal changes in Finnish parliamentary speech data, 1980–2010, with topic modelling. *Parliaments, Estates and Representation 43*, 2 (2023), 149–176.

[6] Sinikallio, L., Leskinen, P., Drobac, S., Tamper, M., Leal, R., Tuominen, J., Rantala, H., La Mela, M., Koho, M., and Hyvönen, E. ParliamentSampo Knowledge Graph. Zenodo, 2025. https://doi.org/10.5281/zenodo.7636419.

[7] Tamper, M., Leal, R., Sinikallio, L., Leskinen, P., Tuominen, J., and Hyvönen, E. Extracting knowledge from parliamentary debates for studying political culture and language. In *Proceedings of the 1st International Workshop on Knowledge Graph Generation From Text and the 1st International Workshop on Modular Knowledge co-located with 19th Extended Semantic Conference (ESWC 2022)* (May 2022), S. Tiwari, N. Mihindukulasooriya, F. Osborne, D. Kontokostas, J. D'Souza, and M. Kejriwal, Eds., vol. 3184, CEUR WS, pp. 70–79.

[8] Tarkka, O., Koljonen, J., Korhonen, M., Laine, J., Martiskainen, K., Elo, K., and Laippala, V. Automated emotion annotation of Finnish parliamentary speeches using GPT-4. In *Proceedings of the IV Workshop on Creating, Analysing, and Increasing Accessibility of Parliamentary Corpora (ParlaCLARIN) @ LREC-COLING 2024* (Torino, Italia, May 2024), D. Fiser, M. Eskevich, and D. Bordon, Eds., ELRA and ICCL, pp. 70–76.

[9] Tirrell, L. Toxic speech: Inoculations and antidotes. *The Southern Journal of Philosophy 56* (2018), 116–144.

A SPARQL Query: Contents of speeches

```
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX semparl: <http://ldf.fi/schema/semparl/>
PREFIX semparl-portal: <http://ldf.fi/schema/parliamentsampo-portal/>

SELECT DISTINCT ?speech ?content
WHERE {
  ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/fin> .
  FILTER NOT EXISTS { ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/swe> . }

  ?speech semparl:content ?content .

  ?speech semparl:speechType ?type .
  VALUES ?type {
    <http://ldf.fi/semparl/speechtypes/Puheenvuoro>
    <http://ldf.fi/semparl/speechtypes/Vastauspuheenvuoro>
  }

  ?speech semparl-portal:facet_electoral_term ?term .
  VALUES ?term {
    <http://ldf.fi/semparl/times/electoral-terms/e_2015-04-22-2019-04-16>
    <http://ldf.fi/semparl/times/electoral-terms/e_2019-04-17-2023-04-04>
  }

  ?speech semparl:speaker ?speaker .
}
```

B SPARQL Query: Metadata on speeches and speakers

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX semparl: <http://ldf.fi/schema/semparl/>
PREFIX semparl-portal: <http://ldf.fi/schema/parliamentsampo-portal/>

SELECT DISTINCT ?speech ?type_label ?term_label
?speaker ?gender_label ?party_label ?role_label ?date
WHERE {
  ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/fin> .
  FILTER NOT EXISTS { ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/swe> . }
  ?speech semparl:content ?content .
  ?speech semparl:speechType ?type .
  VALUES ?type {
    <http://ldf.fi/semparl/speechtypes/Puheenvuoro>
    <http://ldf.fi/semparl/speechtypes/Vastauspuheenvuoro>
  }
  ?type skos:prefLabel ?type_label .
  ?speech semparl-portal:facet_electoral_term ?term .
  VALUES ?term {
    <http://ldf.fi/semparl/times/electoral-terms/e_2015-04-22-2019-04-16>
    <http://ldf.fi/semparl/times/electoral-terms/e_2019-04-17-2023-04-04>
  }
  ?term skos:prefLabel ?term_label .
  ?speech semparl:speaker ?speaker .

  OPTIONAL {
    ?speech semparl-portal:facet_gender ?gender .
    ?gender skos:prefLabel ?gender_label .
    FILTER(LANG(?gender_label) = 'en')
  }

  OPTIONAL {
    ?speech semparl:party ?party .
    ?party skos:prefLabel ?party_label .
    FILTER(LANG(?party_label) = 'en')
  }

  OPTIONAL {
    ?speech semparl:parliamentaryRole ?role .
    ?role skos:prefLabel ?role_label .
    FILTER(LANG(?role_label) = 'en')
  }

  ?speech dct:date ?date .
}
```

C SPARQL Query: Number of electoral terms

```
PREFIX crm: <http://erlangen-crm.org/current/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX semparl: <http://ldf.fi/schema/semparl/>
PREFIX semparl-portal: <http://ldf.fi/schema/parliamentsampo-portal/>
PREFIX bioc: <http://ldf.fi/schema/bioc/>

SELECT ?term_label ?speaker
(COUNT(DISTINCT ?electoral_term) as ?term_count)
WHERE {
  ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/fin> .
  FILTER NOT EXISTS { ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/swe> . }
  ?speech semparl:content ?content .
  ?speech semparl:speechType ?type .
  VALUES ?type {
    <http://ldf.fi/semparl/speechtypes/Puheenvuoro>
    <http://ldf.fi/semparl/speechtypes/Vastauspuheenvuoro>
  }
  ?type skos:prefLabel ?type_label .
  ?speech semparl-portal:facet_electoral_term ?term .
  VALUES ?term {
    <http://ldf.fi/semparl/times/electoral-terms/e_2015-04-22-2019-04-16>
    <http://ldf.fi/semparl/times/electoral-terms/e_2019-04-17-2023-04-04>
  }
  ?term skos:prefLabel ?term_label .
  ?speech semparl:speaker ?speaker .
  ?speech dct:date ?date .

  ?speaker bioc:bearer_of/crm:P11i_participated_in ?membership_event .
  ?membership_event a semparl:ParliamentaryGroupMembership .
  ?membership_event crm:P10_falls_within ?electoral_term .
  ?electoral_term crm:P81a_begin_of_the_begin ?electoral_term_start .
  FILTER(?electoral_term_start <= ?date)
}
GROUP BY ?term_label ?speaker
```

D SPARQL Query: Referenced people

```
PREFIX crm: <http://erlangen-crm.org/current/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX semparl: <http://ldf.fi/schema/semparl/>
PREFIX semparl-portal: <http://ldf.fi/schema/parliamentsampo-portal/>
PREFIX semparl-linguistics: <http://ldf.fi/schema/semparl/linguistics/>
PREFIX bioc: <http://ldf.fi/schema/bioc/>

SELECT DISTINCT ?speech
?referenced_person ?reference_count ?referenced_person_gender ?referenced_person_party
(GROUP_CONCAT(DISTINCT ?party_membership ; separator=' ; ')
as ?referenced_person_party_memberships)
WHERE {
  ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/fin> .
  FILTER NOT EXISTS { ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/swe> . }
  ?speech semparl:content ?content .
  ?speech semparl:speechType ?type .
  VALUES ?type {
    <http://ldf.fi/semparl/speechtypes/Puheenvuoro>
    <http://ldf.fi/semparl/speechtypes/Vastauspuheenvuoro>
  }
  ?type skos:prefLabel ?type_label .
  ?speech semparl-portal:facet_electoral_term ?term .
  VALUES ?term {
    <http://ldf.fi/semparl/times/electoral-terms/e_2015-04-22-2019-04-16>
    <http://ldf.fi/semparl/times/electoral-terms/e_2019-04-17-2023-04-04>
  }
  ?term skos:prefLabel ?term_label .
  ?speech semparl:speaker ?speaker .
  ?speech dct:date ?date .

  ?speech semparl-linguistics:referenceToPerson ?reference .
  ?reference semparl-linguistics:count ?reference_count .
  ?reference skos:relatedMatch ?referenced_person .
  OPTIONAL {
    ?referenced_person bioc:has_gender/skos:prefLabel ?referenced_person_gender .
    FILTER(LANG(?referenced_person_gender) = 'en')
  }

  OPTIONAL {
    ?referenced_person bioc:bearer_of/crm:P11i_participated_in ?group_event .
    ?group_event a semparl:ParliamentaryGroupMembership .
    ?group_event semparl:organization/rdfs:subClassOf ?organization .
    ?organization semparl:party/skos:prefLabel ?referenced_person_party .
    FILTER(LANG(?referenced_person_party) = 'en')

    ?group_event crm:P4_has_time-span ?timespan .
    ?timespan crm:P81a_begin_of_the_begin ?timespan_start .
    FILTER(?timespan_start <= ?date)

    OPTIONAL {
      ?timespan crm:P82b_end_of_the_end ?timespan_end .
      FILTER(?date <= ?timespan_end)
```

```
    }
  }

  OPTIONAL {
    ?referenced_person semparl:party/skos:prefLabel ?party_membership .
    FILTER(LANG(?party_membership) = 'en')
  }
}
GROUP BY ?speech ?referenced_person ?reference_count
?referenced_person_gender ?referenced_person_party
```

E SPARQL Query: Interruptions

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX semparl: <http://ldf.fi/schema/semparl/>
PREFIX semparl-portal: <http://ldf.fi/schema/parliamentsampo-portal/>
PREFIX bioc: <http://ldf.fi/schema/bioc/>

SELECT DISTINCT ?speech ?interruption ?interrupter ?chairman_interruption
?interruption_speaker ?interruption_gender ?interruption_party ?interruption_content
WHERE {
  ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/fin> .
  FILTER NOT EXISTS { ?speech dct:language <http://id.loc.gov/vocabulary/iso639-2/swe> . }
  ?speech semparl:content ?content .
  ?speech semparl:speechType ?type .
  VALUES ?type {
    <http://ldf.fi/semparl/speechtypes/Puheenvuoro>
    <http://ldf.fi/semparl/speechtypes/Vastauspuheenvuoro>
  }
  ?type skos:prefLabel ?type_label .
  ?speech semparl-portal:facet_electoral_term ?term .
  VALUES ?term {
    <http://ldf.fi/semparl/times/electoral-terms/e_2015-04-22-2019-04-16>
    <http://ldf.fi/semparl/times/electoral-terms/e_2019-04-17-2023-04-04>
  }
  ?term skos:prefLabel ?term_label .
  ?speech semparl:speaker ?speaker .
  ?speech dct:date ?date .

  ?speech semparl:isInterruptedBy ?interruption .
  OPTIONAL { ?interruption semparl:interrupter ?interrupter . }
  OPTIONAL {
    ?interruption semparl:speaker ?interruption_speaker .
    OPTIONAL {
      ?interruption_speaker bioc:has_gender/skos:prefLabel ?interruption_gender .
      FILTER(LANG(?interruption_gender) = 'en')
    }
  }
  OPTIONAL {
   ?interruption semparl:party/skos:prefLabel ?interruption_party .
    FILTER(LANG(?interruption_party) = 'en')
  }
  ?interruption semparl:content ?interruption_content .
  ?interruption semparl:chairmanInterruption ?chairman_interruption .
}
```

F Statistics: Speeches

Table 7 lists the number of speeches for the two included speech types. The number of speeches for the two included electoral terms are listed in Table 8. These statistics are visualized in pie chart format in Figure 16.

Speech type	Number of speeches
Vastauspuheenvuoro	44,095
Varsinainen puheenvuoro	33,773

Table 7: Speech type statistics

Electoral term	Number of speeches
Electoral term 2015–2018	43,605
Electoral term 2019–2022	34,263

Table 8: Electoral term statistics

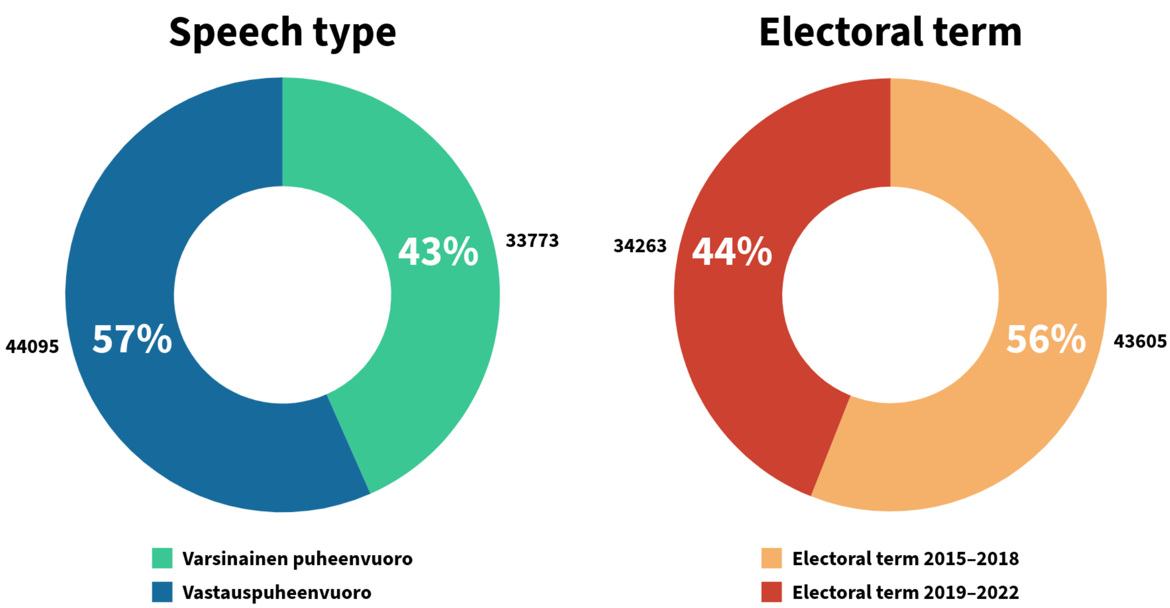

Figure 16: Pie chart visualization of speech type (left) and electoral term (right) statistics

G Statistics: Speakers

Table 9 lists the number of speeches by the gender of the speaker. The number of speeches by the role of the speaker (e.g., if they are a part of the government or opposition based on their affiliation at the time of giving the speech) is listed in Table 10. Table 11 lists the number of speeches by the speaker's party at the time of giving the speech. The statistics on speaker's gender and role are visualized in pie chart format in Figure 17 and party affiliations in Figure 18.

Speaker's gender	**Number of speeches**
Male	44,096
Female	33,772

Table 9: Statistics on the speaker's gender

Speaker's role	**Number of speeches**
Government	41,285
Opposition	36,563
Government official	20

Table 10: Statistics on the speaker's role based on party affiliation

Speaker's party	**Number of speeches**
Social Democratic Party of Finland	15,566
Centre Party	15,406
National Coalition Party	13,680
Finns Party	11,797
Green League	6,135
Left Alliance	5,763
Christian Democrats	3,849
Finnish Reform Movement	3,318
Swedish People's Party of Finland	1,963
Movement Now	175
Power Belongs to the People	145
Seven Star Movement	31
government official	20
personal parliamentary group	20

Table 11: Statistics on the speaker's party affiliation

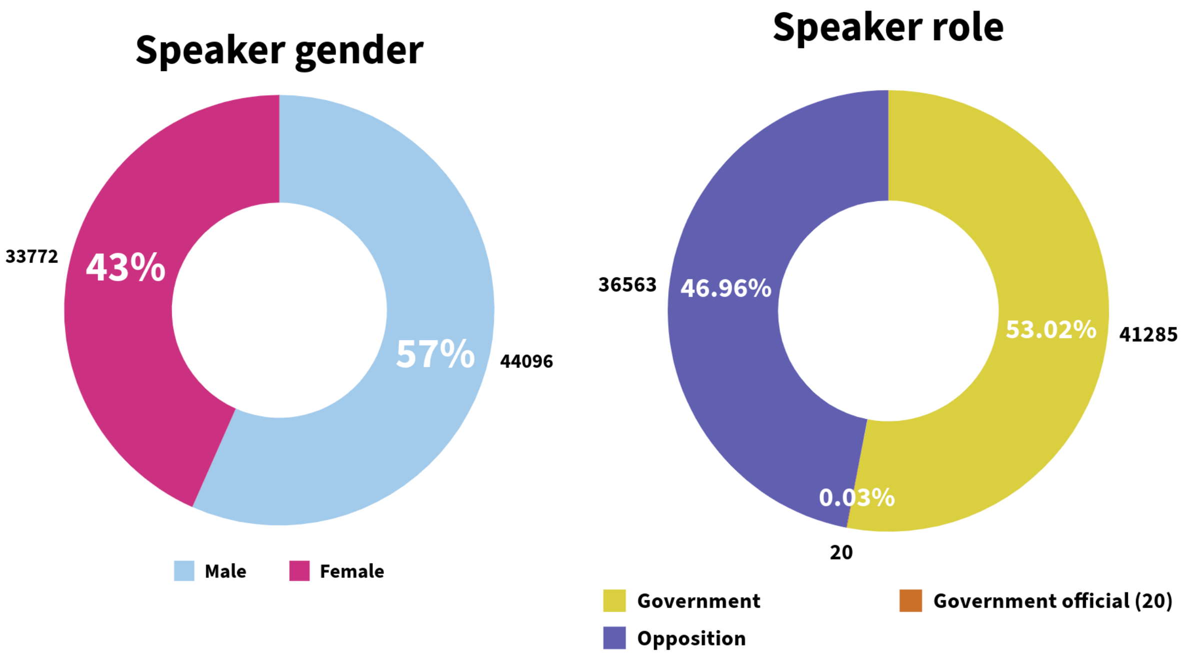

Figure 17: Pie chart visualization of speaker gender (left) and role (right) statistics

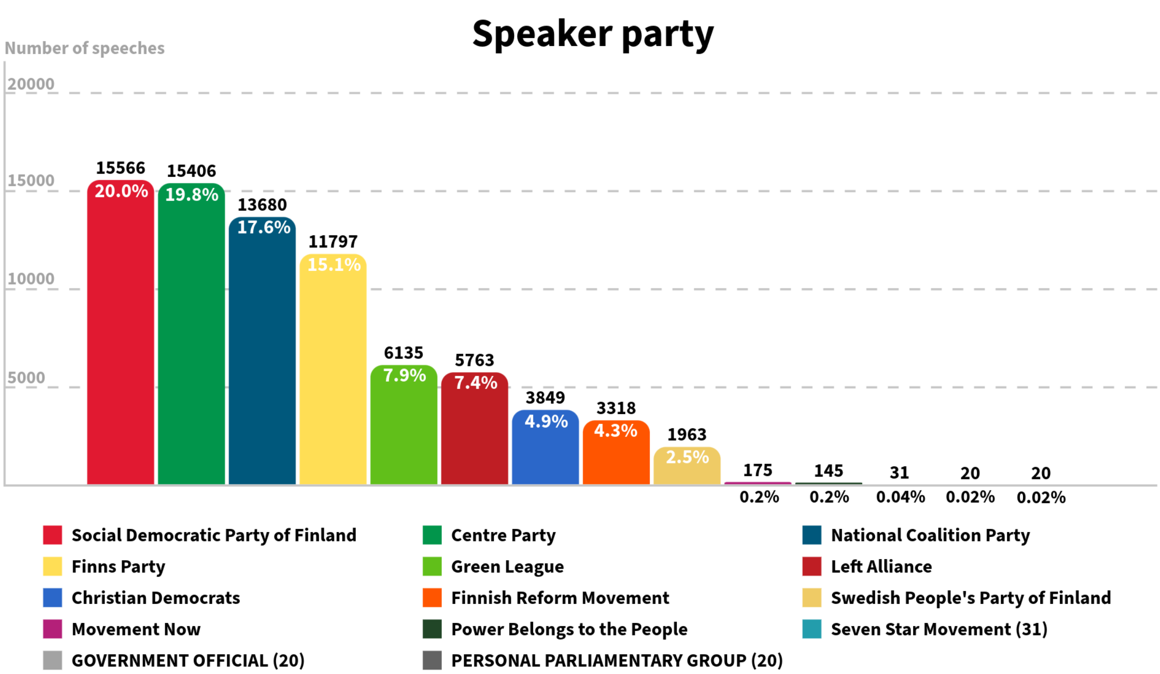

Figure 18: Bar chart visualization of speaker party affiliation statistics

H Party abbreviations

English name	Finnish name	Abbreviation
Social Democratic Party of Finland	Suomen Sosiaalidemokraattinen Puolue	SDP
Centre Party	Suomen Keskusta	Kesk.
National Coalition Party	Kansallinen Kokoomus	Kok.
Finns Party	Perussuomalaiset	PS
Green League	Vihreä liitto	Vihr.
Left Alliance	Vasemmistoliitto	Vas.
Christian Democrats	Suomen Kristillisdemokraatit	KD
Finnish Reform Movement	Korjausliike	KL
Swedish People’s Party of Finland	Suomen ruotsalainen kansanpuolue	RKP
Movement Now	Liike Nyt	Liik
Power Belongs to the People	Valta kuuluu kansalle	VKK
Seven Star Movement	Seitsemän tähden liike	TL
Kansalaisliitto	Kansalaisliitto	KaL

Table 12: Abbreviations used for parties

I Teamwork statement & use of generative AI

I.1 Statement of teamwork

The division of work for all authors is described in the table below:

Name	Tasks
Ahola, Annastiina	- Data extraction and cleaning - Running toxicity analysis
Guevara, Gabriel	- Analysis and visualizations for RQ4-5
Lindén, Wilma	- Analysis and visualizations for RQ1-3

I.2 Use of generative AI

Overleaf's Writefull AI tool was used for spellchecking and checking grammar and style. In addition, generative AI was used to generate code snippets to help with generating some visualizations. The content produced was reviewed and edited by the authors.