Construction of Japanese Prefectural Assembly Minutes Datasets Across Three Electoral Terms: Comparative Analysis of 2011, 2015, and 2019 Four-Year Periods

Keiichi Takamaru[1], **Yuzu Uchida**[2], **Hokuto Ototake**[3], **Yasutomo Kimura**[4]
[1]Utsunomiya Kyowa University, Japan [2]Hokkai-Gakuen University, Japan
[3]Fukuoka University, Japan. [4]Otaru University of Commerce, Japan

Abstract

The presented longitudinal cross-regional corpus of Japanese prefectural assembly minutes spans 12 years (2011-2023) across three electoral terms. The corpus comprises 12,236,974 records containing 743,147,226 characters (471,496,688 tokens) of transcribed remarks from the plenary sessions of all 47 prefectural assemblies in Japan. Each dataset is organized by speaker, with assembly members linked to their electoral information, including gender, age, and electoral district. Through a comparative analysis across the three terms, we documented significant temporal changes. The proportion of members aged 25-44 decreased, whereas female representation increased. Female members use 20-30% more characters per speech than male counterparts across all age groups. The proportion of members who never speak varies from under 2% for younger females to over 10% for males aged 65+. We demonstrate the utility of the corpus through three applications: a quantitative analysis of gender and age patterns in political discourse, AI-driven computational dialectology for extracting regional linguistic features, and a web-based search and visualization system. This longitudinal cross-regional corpus provides a valuable resource for interdisciplinary research on subnational politics, computational linguistics, dialectology, and political communication in non-Western democracies. The datasets are available for research purposes upon request, with public query access provided through a web-based interface.

Keywords: Local assembly minutes, Longitudinal cross-regional corpus, Japanese political discourse, Subnational politics

1. Introduction

Local government assembly minutes are a rich and unexplored language resource, as they capture authentic political discourse at the subnational level. These documents contain valuable linguistic and socio-political information that can contribute to interdisciplinary research in multiple fields. In particular, they serve as valuable primary data for subnational political science research and public policy studies while also providing insights for dialectology and sociolinguistics. However, the systematic collection and organization of such material across multiple jurisdictions and time periods present significant challenges in terms of data standardization, metadata annotation, and longitudinal consistency.

Recent research has focused on the development of language resources based on local Japanese assembly minutes. These efforts include developing robust data collection methodologies, constructing large-scale longitudinal and cross-regional corpora that encompass prefectural and municipal assemblies, and conducting various computational linguistic analyses (Takamaru, 2013; Sakaji et al., 2019;Tsutsui et al., 2014; Kimura et al., 2016; Nagafuchi et al., 2024).

As part of this comprehensive initiative, we have been constructing datasets that provide structured metadata, thereby enabling researchers to identify "when," "where," "who," and "what" was said while maintaining controlled collection scope and temporal consistency. The Japanese Prefectural Assembly Minutes Dataset was designed using a systematic approach: the collection scope encompassed plenary sessions of all 47 prefectural assemblies across Japan, and the collection period was defined as four-year terms beginning with unified local elections[1].

In this paper, we present a longitudinal cross-regional corpus of Japanese prefectural assembly minutes spanning 12 years (2011-2023) and covering all 47 prefectures. The 2019-term dataset is the third dataset in our series and has been newly developed to enable analyses of both temporal changes and regional variations in the Japanese subnational political discourse. We provide an overview of the complete dataset through a systematic comparison of member composition and speech volume across three electoral terms and describe use cases and downstream applications demonstrating the utility of the corpus for interdisciplinary research.

The main contributions of this study are as follows: (1) construction and release of the 2019-term dataset, completing a 12-year longitudi-

[1]Among the 47 prefectural assemblies, 41 hold elections during unified local election periods. Six prefectures (Iwate, Miyagi, Fukushima, Ibaraki, Tokyo, and Okinawa) conduct elections at different times.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 6216–6223
11–16 May 2026.

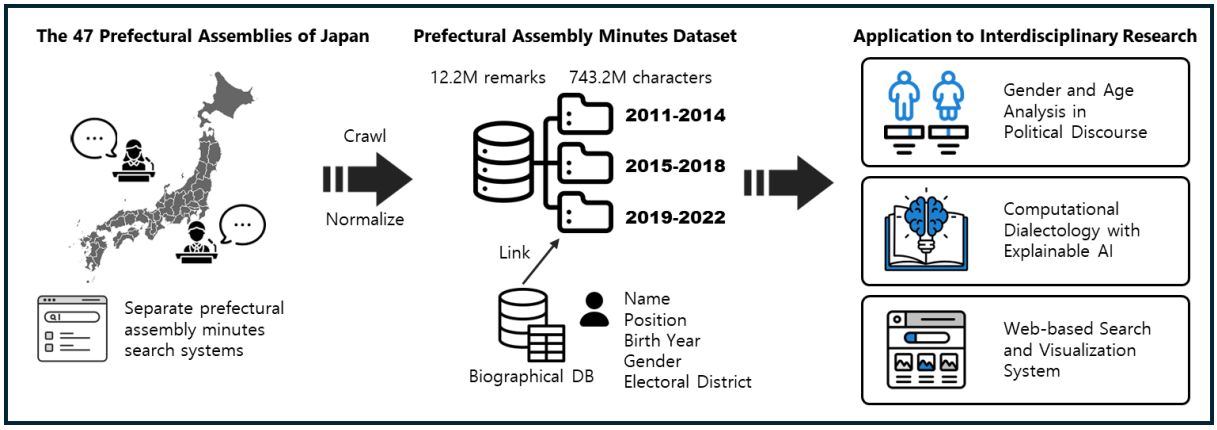

Figure 1: Overview of dataset construction

nal corpus spanning three electoral terms; (2) comprehensive comparative analyses of linguistic and participatory patterns across these terms; (3) methodological insights on consistent multi-temporal dataset construction; and (4) provision of a substantial language resource for computational linguistics, political science, and related research communities.

2. Related Work

Parliamentary and council records are extensively used as primary sources for analyzing legislative behavior and political discourse. Previous studies employed various quantitative measures to examine the different aspects of legislative activities based on the information extracted from these records.

Several studies used speech volume as an indicator of parliamentary engagement and strategic behavior. Research on the Estonian Parliament employed the word count of individual speeches in plenary transcripts as a quantitative measure of parliamentary activity and agenda visibility (Lupacheva and Mölder, 2024). Similar studies in Switzerland examined the effects of term limits on legislator behavior, showing that although some members of parliament (MPs) exhibit signs of shirking when they no longer face re-election, such effects are weaker or absent at the cantonal level (Frech et al., 2021). Research in Finland demonstrated that intraparty electoral vulnerability decreased the likelihood of MPs delivering speeches; however, interparty vulnerability did not exert a comparable effect (Poyet and Raunio, 2021).

Other studies have focused on how parties strategically use parliamentary speech for political positioning and party discipline. Studies on Chile have analyzed how parties balance and maintain party discipline by allowing legislators to cultivate constituency ties by strategically allocating speaking opportunities (Alemán et al., 2017). Research on the Austrian Nationalrat showed that parties employ parliamentary speech as a tool for competition, emphasizing their own or manifesto-salient issues, even under institutional constraints (Ivanusch, 2023). Research on the UK House of Commons investigated how government- backbench MPs use parliamentary speeches when rebelling against the party line, revealing that dissent is associated with longer, simpler, and more personal rhetoric, aiming to connect with voters (Slapin and Kirkland, 2020).

Beyond single-country studies, several large-scale corpus-construction projects have advanced comparative parliamentary research. The ParlaMint II project (Erjavec et al., 2024) constructed a comparative corpus of parliamentary debates from 29 European countries and autonomous regions, covering at least 2015-2022 and containing over one billion words, with rich metadata on 24,000 speakers, including political orientation, annotated with Universal Dependency syntactic structures and named entities. In the United States, Brown and Weber (Maxfield Brown and Weber, 2022) proposed the Council Data Project (CDP), which standardizes municipal council meeting records and provides searchable transcripts, minutes, and voting data for comparative analysis across jurisdictions. Historical analysis has also been enabled by large-scale digitization efforts. Card et al. (2022) analyzed 200,000 congressional speeches and 5,000 presidential statements from 1880 to the present, revealing the evolution of immigration discourse over more than a century.

Despite the significant contributions of these international efforts to parliamentary research, the systematic collection and analysis of subnational legislative records remain limited, particularly for non-Western democracies. Under Japan's Local

Autonomy Act, all local assemblies are legally required to record proceedings in detail and publish comprehensive minutes. This regulatory framework ensures the availability of assembly minutes across all 47 prefectures. Our study constructed a comprehensive longitudinal corpus of Japanese prefectural assembly minutes to address this research gap in subnational parliamentary studies.

3. Dataset Construction Methodology

3.1. Data Collection Process

The datasets of Japanese prefectural assembly minutes target the plenary sessions (regular and extraordinary sessions) of all 47 prefectural assemblies nationwide. The collection period of each dataset was four years, which is the electoral term for unified Japanese local elections. The 2011-term dataset contains the assembly minutes from April 2011 to March 2015. The 2015- and 2019-term datasets record the subsequent four-year periods.

The website for each prefecture has a full-text search system for the published assembly minutes. We created crawlers for each system to collect the assembly minutes. As data recording methods and display formats differ among local governments, collecting data in a format amenable to metadata annotation requires individualized handling by each local government, which partially entails manual intervention.

3.2. Data Preprocessing and Standardization

The minutes collected from all 47 prefectures were initially segmented into individual remarks based on the periods inserted by the minute preparers. Subsequently, each remark was assigned essential metadata (e.g., speaker name, meeting name, and date) and stored as a single record. To enable tracking across different terms, each assembly member was assigned a unique ID. Speaker names extracted from the minutes include titles and member numbers (not unique identifiers); therefore, a single speaker may appear under multiple notations in the speaker name field. Furthermore, discrepancies in notation, likely caused by input errors, frequently appear within the minutes. To ensure data quality and accurate speaker identification, manual verification through visual inspection was conducted after the automated extraction process to consolidate name variations and correct the discrepancies. In addition to the remarks of assembly members, meeting records document the remarks of the chairperson, governors, staff members, and expert witnesses. To analyze the characteristics of the remarks based on the differences in roles within the assembly, this corpus attached the speaker's position information as metadata to each remark. Furthermore, for assembly members, the information on date of birth, gender, and electoral district was made public during the election candidacy process. Since this attribute is useful for comparative analysis of speech content and linguistic features, we obtained the data from the election information website "Seijiyama[2]" and stored it in a table separate from the one containing the remarks, linking the information using assembly member IDs as keys. However, some assembly members had different name notations in election records and assembly minutes (e.g., hiragana vs. kanji). Therefore, manual verification through visual inspection was conducted to ensure accurate name matching and maintain data quality across datasets. As common assembly member IDs were assigned across attribute data in all the datasets, member remarks can be analyzed from various perspectives across different terms.

Term	Records	Characters	Tokens
2011	4,395,877	252,280,412	160,986,124
2015	3,827,015	244,111,644	154,275,983
2019	4,014,082	246,755,170	156,234,581
Total	12,236,974	743,147,226	471,496,688
Average	4,078,991	247,715,742	157,165,563

Table 1: Overall size of each dataset

4. Corpus Statistics and Comparative Analysis

4.1. Overall Size of Datasets

The Corpus comprises 12,236,974 records containing 743,147,226 characters (471,496,688 tokens) of transcribed remarks across the three terms. Token counts were obtained using MeCab morphological analyzer with the IPADIC dictionary. Table 1 presents the detailed statistics for each term.

Figure 2 shows the total number of characters spoken in each prefecture in the 2019-term dataset, with the smallest number of total characters spoken at the Yamagata Prefectural Assembly (2,141,262 characters). The largest number was at the Okinawa Prefectural Assembly (10,859,757 characters), followed by the Tottori Prefectural Assembly (10,298,648 characters). This substantial variation in the total number of characters spoken can be attributed to factors such as whether the

[2]https://seijiyama.jp/

Term	Gender	25–44	45–64	65+	N/A	Total
2011	M	532	1,646	497	16	2,691
2011	F	49	170	32	4	255
2015	M	447	1,421	629	1	2,498
2015	F	56	185	40	0	281
2019	M	377	1,385	649	44	2,455
2019	F	61	199	57	12	329
2019	N/A	0	0	0	1	1

Table 2: Number of assembly members by gender and age group for each electoral term

Term	Gender	25–44	45–64	65+
2011	M	43,583	44,626	31,299
2011	F	44,455	59,969	59,366
2015	M	45,814	45,378	30,469
2015	F	42,886	55,950	66,057
2019	M	46,921	45,448	32,262
2019	F	46,824	52,745	59,147

Table 3: Average number of characters per speech by gender and age group

assemblies were predominantly ceremonial or featured active deliberations.

4.2. Trends in the Number of Assembly Members

Table 2 shows the number of serving assembly members in each term. According to the Ministry of Internal Affairs and Communications[3], the total number of authorized seats in the 47 prefectural assemblies was 2,735 in 2012 and 2,676 in 2023. In line with the reduction in the number of authorized seats, the total number of assembly members shows a declining trend. However, for the six assemblies that hold elections outside the unified local election period, the dataset included members both before and after re-election. Therefore, the total number of members in the dataset exceeds the total number of authorized seats.

Comparing the age groups of members in each term (Figure 3), the proportion of those aged 25-44 decreases with each term, whereas the proportion of older members aged 65 years and above increases. There were 690 newcomers or former members who did not serve in the 2015 term but were elected in the 2019 term; of these, 188 (27.2%) were aged 25-44. Although younger members were added as newcomers, the influence of continuously re-elected members is significant, and the overall age composition is gradually increasing.

[3]https://www.soumu.go.jp/index.html

Furthermore, in comparing the gender ratio in each term (Figure 4), the proportion of women is observed to increase each term in all age groups. However, even in the 2019 term, which had the highest proportion of female assembly members, the proportion of women remained at only 12%.

4.3. Amount of Speech by Assembly Members

Table 3 lists the average number of characters spoken by gender and age group in each term. The overall average number of characters spoken was 43,153 characters in the 2011 term, 43,031 characters in the 2015 term, and 43,177 characters in the 2019 term. Considering the number of characters spoken by attribute, a slightly increasing trend can be observed in the average number of characters spoken by those aged 25-44. In addition, the following trends were observed across all terms:

1. For male assembly members, the number of characters spoken decreases as age group increases
2. For female assembly members, the number of characters spoken increases as age group increases
3. In all age groups, female assembly members speak more characters than male assembly members.

Table 4 shows the number and proportion of non-speaking assembly members (in parentheses) by gender and age group for each term. Non-speaking assembly members refer to those who served (even partially) during the relevant period but had zero characters spoken in plenary sessions during that period. The overall non-speaking rate was approximately 5% (5.2% in the 2011 term, 4.4% in the 2015 term, and 4.5% in the 2019 term). Regarding non-speaking assembly members by attribute, the following trends were observed across all terms.

1. Male assembly members have a higher non-speaking rate compared to female assembly members
2. For both men and women, the non-speaking rate increases as age group increases
3. In particular, more than 10% of male assembly members aged 65 and over are non-speaking in every term.

However, prefectural assemblies notably have a larger number of members than municipal assemblies; therefore, due to time constraints, not all

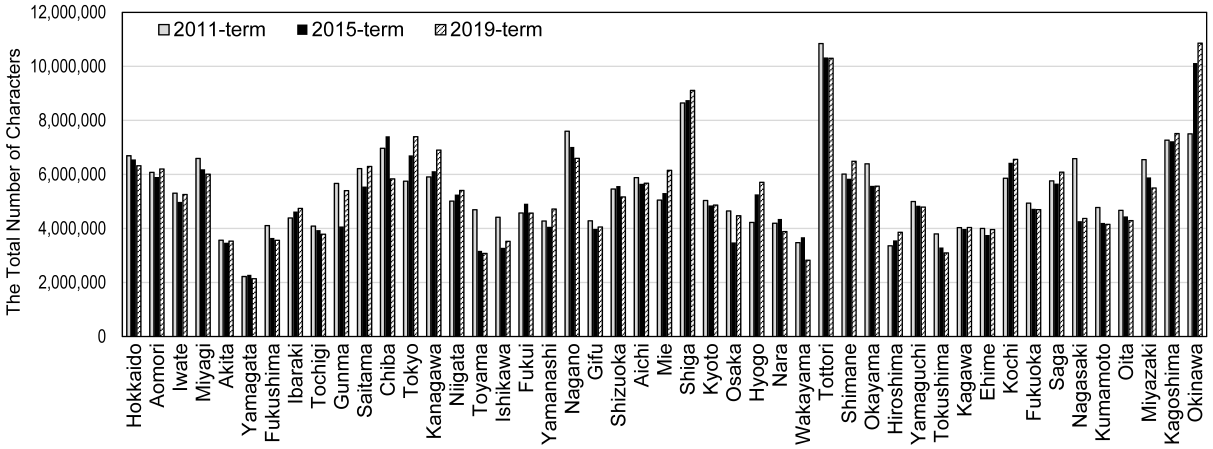

Figure 2: Total number of characters by prefecture in 2019-term

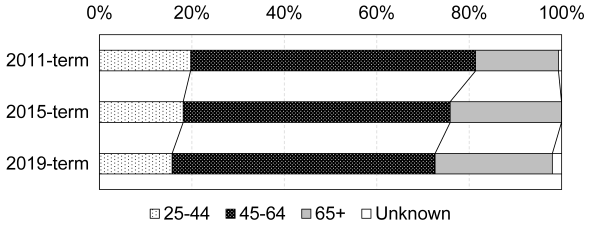

Figure 3: Age composition by term

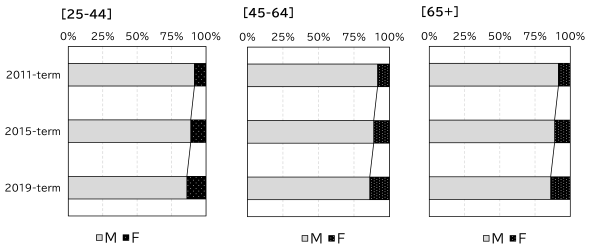

Figure 4: Gender ratio by age group across terms

Term	Gender	25–44	45–64	65+
2011	M	9(1.7%)	76(4.6%)	62(12.5%)
2011	F	1(2.0%)	2(1.2%)	2(6.3%)
2015	M	9(2.0%)	38(2.7%)	71(11.3%)
2015	F	0(0.0%)	2(1.1%)	1(2.5%)
2019	M	7(1.9%)	36(2.6%)	76(11.7%)
2019	F	0(0.0%)	3(1.5%)	3(5.3%)

Table 4: Non-speaking members (%)

	2011-Term	2015-Term
Total members	887	684
Non-speaking	96 (10.8%)	40 (5.8%)

Table 5: Non-speaking rate of retired and defeated members

members have sufficient opportunities to speak in plenary sessions. Additionally, members serving as assembly chairpersons rarely speak as ordinary members during their tenure, not withstanding those who become assembly members midway through a term via by-election and those who resign midway through a term. For these reasons, it should be noted that non-speaking members do not necessarily represent those who intentionally choose not to speak.

Table 5 presents the number and rate of non-speaking members among those who did not serve in the subsequent term (i.e., those who were defeated or retired). Retiring or defeated members showed significantly higher non-speaking rates than the overall membership in both the 2011-term ($\chi^2 = 83.19$, $p < 0.001$) and 2015-term ($\chi^2 = 4.86$, $p < 0.05$).

Table 6 presents the number and rate of non-speaking members among those who did not serve in the previous term (i.e., newcomers and former members who were re-elected). Conversely, newcomers and returning members demonstrated significantly lower non-speaking rates than the overall membership in both the 2015-term ($\chi^2 = 13.55$, $p < 0.001$) and 2019-term ($\chi^2 = 37.64$, $p < 0.001$).

The column labeled "Returned" indicates members who were not serving in the 2015-term but were elected in the 2019-term and had previously served in the 2011-term–that is, members who returned after an absence. The non-speaking rates for both members not serving in the previous term and those who returned are very low compared with the overall non-speaking rate. This type of data analysis, spanning three consecutive terms, is only possible by making construction of a multi-term dataset an ongoing process.

	2015-Term	2019-Term	Returned
Total Numbers	720	689	44
Non-speaking	14 (1.9%)	2 (0.3%)	1 (2.3%)

Table 6: Non-speaking rate of new and returning members

5. Use Cases and Downstream Applications

To illustrate the practical applications of our datasets, we present three examples from our own research: two use cases demonstrating their utility in political science and dialectology, and a web-based search and visualization system we developed to facilitate data access and analysis.

5.1. Gender and Age Analysis in Political Discourse

Uchida et al. (2019) employed this corpus to visualize gender- and age-related characteristics in the utterances of prefectural assembly members across Japan. Using statistical methods such as log-likelihood ratios, this study extracted distinctive terms and revealed differences by gender and age. For instance, female legislators frequently used terms related to "health and welfare," "childbirth and childcare," and "education and research," whereas older male legislators frequently used terms related to "agriculture, forestry, and fisheries." Such analyses demonstrate how the corpus enables quantitative investigation of words in local assemblies and political communication patterns.

5.2. Computational Dialectology with Explainable AI

Linguists can conduct synchronic studies on regional variations by comparing minutes across nations. Because speakers are local residents and typically natives of their regions, these records provide valuable data for observing regional linguistic variation, making assembly minutes an important resource for dialectology and sociolinguistic research.

Takamaru and Ototake (2023) demonstrated the potential of AI technology in automatically extracting dialectal features from assembly minutes. This approach combines machine-learning classifiers that predict the location of statements with eXplainable AI (XAI) techniques that identify location-specific linguistic cues. These methods enable the discovery of region-specific phenomena that are difficult to detect using traditional manual surveys, thereby demonstrating new possibilities for computational dialectology.

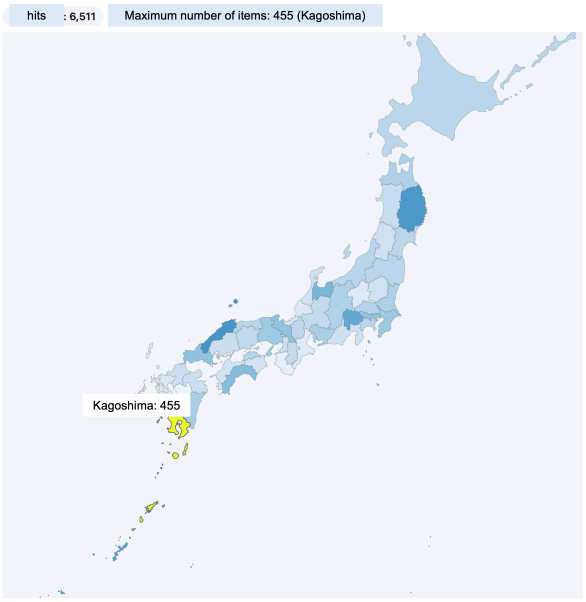

Figure 5: Example of map search (childcare support)

5.3. Web-based Search and Visualization System

Ototake et al. (2018) developed a web-based system Giimiru to search and visualize local assembly minutes across multiple Japanese municipalities.

This system addresses the challenge of assembly minutes published in diverse formats across local governments, making cross-municipal analysis difficult. The system provides five main functionalities: (1) keyword search with context expansion; (2) semantic similarity search using embedding-based methods; (3) geographic visualization of search results; (4) time-series analysis; and (5) cross-tabulation of multiple variables. Figure 5 shows an example of the map visualization feature. These capabilities support diverse applications, including sociolinguistic research on dialectal variation, econometric analysis using panel data, and tracking of policy discussions.

6. Data Availability

The Japanese Prefectural Assembly Minutes Datasets will be made available for research purposes upon publication of this paper. Although the original assembly minutes are public records, the value-added datasets include standardized formats, comprehensive metadata, unique speaker IDs, and linked attributes.

Researchers can access datasets through

1. Direct data access under a research use

agreement (contact information will be provided upon publication)

2. Public web-based query interface for exploratory analysis and visualization

The datasets are distributed in standard formats (CSV and JSON) with comprehensive documentation. Sample data and detailed documentation will be made available at the time of publication.

7. Conclusion

In this paper, we describe the construction of the Japanese Prefectural Assembly Minutes Datasets across three electoral terms and provide an overview of their characteristics by comparing the 2011-term, 2015-term, and 2019-term datasets with respect to the number of speakers, speech character counts, and non-speaking assembly members. We also present use cases demonstrating how these datasets support research in political discourse analysis, dialectology, and data accessibility through web-based tools.

By constructing these datasets spanning three terms, we demonstrate the capability of conducting longitudinal comparative analyses that would not be possible with a single-term dataset. Our research group has continuously built prefectural assembly minutes datasets covering a 12-year period across the 2011, 2015, and 2019 terms. Although the dataset volume exceeds what can be manually observed in its entirety, human validation was conducted at critical stages to ensure data quality, particularly for metadata annotation and name disambiguation (assigning member IDs).

The availability of this longitudinal dataset opens new possibilities for computational political science research, enabling studies on political discourse evolution, policy agenda changes, and democratic representation patterns at the subnational level. We believe that this resource will contribute significantly to both computational linguistics and political sciences. Moving forward, we plan to incorporate each Prefectural Assembly Minutes Dataset into the local assembly minutes search system "Giimiru v2" (Ototake et al., 2025) and continue refining the data while promoting its interdisciplinary use.

8. Acknowledgements

This work was supported in part by JSPS KAKENHI Grant Numbers 25K04045 and 24K15196, and by JST RISTEX Grant Number JPMJRS25L2.

9. Bibliographical References

Eduardo Alemán, Margarita M. Ramírez, and Jonathan B. Slapin. 2017. Party strategies, constituency links, and legislative speech. *Legislative Studies Quarterly*, 42(4):637–664.

Dallas Card, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. 2022. Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, 119(31):e2120510119.

Elena Frech, Niels D. Goet, and Simon Hug. 2021. Shirking and slacking in parliament. *Legislative Studies Quarterly*, 46(2):493–514.

Christoph Ivanusch. 2023. Issue competition in parliamentary speeches? a computer-based content analysis of legislative debates in the austrian nationalrat. *Legislative Studies Quarterly*, 49(1):203–232.

Tatiana Lupacheva and Martin Mölder. 2024. A place to speak and be heard? Parliamentary speech and media attention in Estonia, 2011–2019. *Legislative Studies Quarterly*, 49(4):905–924.

Corentin Poyet and Tapio Raunio. 2021. Reconsidering the electoral connection of speeches: The impact of electoral vulnerability on legislative speechmaking in a preferential voting system. *Legislative Studies Quarterly*, 46(4):1087–1114. First published online 27 September 2020.

Hiroki Sakaji, Yasutomo Kimura, Izumi Kiyoshi, and Hiroyasu Matsushima. 2019. Extraction of volitional utterances from Japanese local political corpus. *The 2nd International Workshop on Cross-disciplinary Data Exchange and Collaboration*, pages 24–29.

Jonathan B. Slapin and Justin H. Kirkland. 2020. The sound of rebellion: Voting dissent and legislative speech in the UK house of commons. *Legislative Studies Quarterly*, 45(2):153–183.

Keiichi Takamaru. 2013. What topics are discussed in a local assembly?: A text-mining analysis of the minutes of Utsunomiya City Assembly (in Japanese). *The annual report of urban and regional economic studies*, 13:162–176.

Keiichi Takamaru and Hokuto Ototake. 2023. An exploratory survey of dialects in minutes of local assemblies: A dialectological study using machine learning and xai (in Japanese). *Studies in Dialects*, 9:27–51.

Yuzu Uchida, Keiichi Takamaru, Hokuto Ototake, and Yasutomo Kimura. 2019. Visualization of activity of local assembly members based on log-likelihood ratio and political vocabulary ratio (in Japanese). *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 31(2):662–671.

10. Language Resource References

Tomaž Erjavec, Matyáš Kopp, Nikola Ljubešić, Taja Kuzman, Paul Rayson, Petya Osenova, Maciej Ogrodniczuk, Çağrı Çöltekin, Danijel Koržinek, Katja Meden, Jure Skubic, Peter Rupnik, Tommaso Agnoloni, José Aires, Starkaður Barkarson, Roberto Bartolini, Núria Bel, María Calzada Pérez, Roberts Darundefinedis, Sascha Diwersy, Maria Gavriilidou, Ruben van Heusden, Mikel Iruskieta, Neeme Kahusk, Anna Kryvenko, Noémi Ligeti-Nagy, Carmen Magariños, Martin Mölder, Costanza Navarretta, Kiril Simov, Lars Magne Tungland, Jouni Tuominen, John Vidler, Adina Ioana Vladu, Tanja Wissik, Väinö Yrjänäinen, and Darja Fišer. 2024. ParlaMint II: advancing comparable parliamentary corpora across Europe. *Lang. Resour. Eval.*, 59(3):2071–2102.

Yasutomo Kimura, Keiichi Takamaru, Takuma Tanaka, Akio Kobayashi, Hiroki Sakaji, Yuzu Uchida, Hokuto Ototake, and Shigeru Masuyama. 2016. Creating Japanese political corpus from local assembly minutes of 47 prefectures. In *Proceedings of the 12th Workshop on Asian Language Resources*, pages 78–85.

Eva Maxfield Brown and Nicholas Weber. 2022. Councils in action: Automating the curation of municipal governance data for research. *Proceedings of the Association for Information Science and Technology*, 59(1):23–31.

Keiyu Nagafuchi, Yasutomo Kimura, Kazuma Kadowaki, and Kenji Araki. 2024. Building a large corpus and pre-trained language models from national and local assembly minutes (in Japanese). *Journal of Natural Language Processing*, 31(2):707–732.

Hokuto Ototake, Hiroki Sakaji, Keiichi Takamaru, Akio Kobayashi, Yuzu Uchida, and Yasutomo Kimura. 2018. Web-based system for Japanese local political documents. *International Journal of Web Information Systems*, 14(3):357–371.

Hokuto Ototake, Keiichi Takamaru, Yuzu Uchida, and Yasutomo Kimura. 2025. Overview of the local assembly minutes search system Giimiru v2 (in Japanese). In *Proceedings of the 31st Annual Meeting of the Association for Natural Language Processing*.

Takashi Tsutsui, Takuya Gaman, Takashi Oshiro, Kohei Sugawara, Takahiro Nagai, Hideyuki Shibuki, Yasutomo Kimura, and Tatsunori Mori. 2014. Construction of regional assembly minutes corpus and a proposal for annotation scheme for implementing political information system. *Journal of Natural Language Processing*, 21(2):125–155.