Political Science Research and Methods (2026), **14**, 1200–1216
doi:10.1017/psrm.2025.10048

ORIGINAL ARTICLE

Legislators' sentiment analysis supervised by legislators

Akitaka Matsuo[1] and Kentaro Fukumoto[2]

[1]Department of Government, University of Essex, Colchester, UK and [2]Department of Political Science, Gakushuin University, Tokyo, Japan
Corresponding author: Kentaro Fukumoto; Email: fukumoto@j.u-tokyo.ac.jp

In 2024, he joined the University of Tokyo after completing his research in 2023.

(Received 3 January 2025; revised 23 April 2025; accepted 18 June 2025; first published online 5 September 2025)

Abstract
The sentiment expressed in a legislator's speech is informative. However, extracting legislators' sentiment requires human-annotated data. Instead, we propose exploiting closing debates on a bill in Japan, where legislators in effect label their speech as either pro or con. We utilize debate speeches as the training dataset, fine-tune a pretrained model, and calculate the sentiment scores of other speeches. We show that the more senior the opposition members are, the more negative their sentiment. Additionally, we show that opposition members become more negative as the next election approaches. We also demonstrate that legislators' sentiments can be used to predict their behaviors by using the case in which government members rebelled in the historic vote of no confidence in 1993.

Keywords: Bidirectional Encoder Representations from Transformers (BERT); Japan; legislative speech; machine learning; supervised learning; text analysis

1. Introduction

Legislators' sentiment in daily legislative business carries important information for understanding their behavior. Who will agree or disagree with a bill on the table? How willingly or hesitantly do they do so? These questions are why legislative scholars have been interested in measuring legislators' sentiment (e.g. Young and Soroka, 2012; Rheault *et al.*, 2016; Proksch *et al.*, 2019; Osnabrügge, Hobolt and Rodon, 2021), although it is challenging to do so appropriately. Thanks to the recently developed technique of computational text analysis, political scientists have been increasingly utilizing legislative speech to study legislators' ideological positions (Lauderdale and Herzog, 2016; Schwarz *et al.*, 2017; Rheault and Cochrane, 2020) and topics of concern (Quinn *et al.*, 2010). Sentiment, the research object of our study, is another major area of application; scholars have measured sentiment in the European Parliament (Proksch *et al.*, 2019) and the legislatures in Canada (Rheault *et al.*, 2016), Ireland (Herzog and Benoit, 2015), Japan (Takikawa and Sakamoto, 2020), the Netherlands (Grijzenhout *et al.*, 2014), the United Kingdom (Slapin and Kirkland, 2020), and the United States (Hopkins and King, 2010).

However, sentiment detection from text is not an easy task. The method of choice is almost always either supervised learning (Herzog and Benoit, 2015) or a dictionary-based method (Proksch *et al.*, 2019; Slapin and Kirkland, 2020; Takikawa and Sakamoto, 2020), and both types require labels generated by human coders for use as training texts or as lexicons of polarity words. Nonetheless, there are three problems. First, human coding is costly. For example, the Lexicoder Sentiment Dictionary

© The Author(s), 2025. Published by Cambridge University Press on behalf of EPS Academic Ltd. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (http://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

(LSD, Young and Soroka, 2012) summarizes three dictionaries in a labor-intensive manner, and each of them is already the result of hard work by humans. Crowdsourcing (Benoit *et al.*, 2016; Haselmayer and Jenny, 2017; Rudkowsky *et al.*, 2018) may alleviate the problem to some extent but does not completely solve it. Second, the selection of labels by human coders is subjective, even when human coders are experts (Mikhaylov, Laver and Benoit, 2012). This leads to low values of the intercoder agreement that are "common for sentiment analysis tasks" (Grijzenhout *et al.*, 2014, 122). Even with the recently popularized approach of combining word vectors and seed words (Rheault *et al.*, 2016; Rice and Zorn, 2021; Watanabe, 2021; Hargrave and Blumenau, 2022), it is difficult to avoid arbitrariness in selecting seed words. Third, labeled texts might be domain specific, and their validity is not guaranteed outside the domain where the labels are created (Grimmer and Stewart, 2013; Osnabrügge, Hobolt and Rodon, 2021; Rice and Zorn, 2021). Importing general-purpose dictionaries such as LSD into political science is sometimes unsuccessful.

The core of our approach is that when training models for sentiment detection, we use labels generated by *legislators themselves* through their own statements. Specifically, we utilize debates in the Japanese national legislature. In the closing debates on a bill just before a vote is taken, legislators sometimes express their opinion toward the bill, making it clear whether they support or oppose the bill. Therefore, these debates represent the ground truth of the speakers' sentiments. Since government–opposition conflicts are the basic cleavage in the parliamentary system (Herzog and Benoit, 2015; Lauderdale and Herzog, 2016; Rudkowsky *et al.*, 2018; Proksch *et al.*, 2019; Curini *et al.*, 2020), some readers may think that we only have to refer to debate speeches made by government and opposition party leaders. However, government (opposition) party leaders do not necessarily always express positive (negative) sentiments.[1] Our approach has the same spirit as other machine learning applications where researchers exploit labels generated simultaneously with the input text, such as the product or movie rating (Pang *et al.*, 2002; Fang and Zhan, 2015; Nguyen *et al.*, 2018). Fernandes *et al.* (2019) exploit bills as prelabeled texts for topics, namely, the jurisdictions of the committees the bills are referred to.

By exploiting the debate texts as our labeled data, we can solve the above three problems. Notably, first, our approach is not costly. In effect, the legislators themselves have already labeled their debates as either positive or negative. Second, the label (i.e., pro or con regarding the corresponding bill) reflects not *our* subjective judgment but that *of the legislators themselves*, which is exactly what we aim to study. Third, domain specificity is a strength of our approach, as the domain of the labeled texts is the same as that of the unlabeled texts: legislative speeches. We contribute to the literature on the sentiment analysis of political texts by proposing a new way to construct labeled data by making the most of the constraining nature of legislative institutions (Proksch and Slapin, 2015).

Once labeled data are ready, we can train *any* supervised learning model by using the debate corpus as the training dataset and scale the remaining legislative speeches accordingly. In this article, we utilize bidirectional encoder representations from transformers (BERT, Devlin *et al.*, 2018) to train a model for classification and determine the sentiment score of each speech.[2]

We present two substantive analyses that employ our sentiment scores. The first uses our sentiment scores as a dependent variable. We demonstrate that opposition members' sentiments become negative when they are senior or the next election is approaching. The second task utilizes our sentiment scores as an independent variable. Specifically, we show that the sentiment scores help predict legislators' undisciplined behavior during the historic vote of no confidence in 1993 and party defection

[1] We thank a reviewer for suggesting these analyses. For details, see the Supplementary Materials.

[2] Widmann and Wich (2023) and Laurer *et al.* (2024) demonstrate that BERT and other transformer-based models are better at natural language processing tasks than other widely used methods are. For legislative and electoral politics applications of BERT, see Ballard *et al.* (2023) and Bestvater and Monroe (2023).

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

in the following election that led to the breakdown of the decades-long dominance of the Liberal Democratic Party (LDP).

This article proceeds as follows. The next section gives an overview of the structure of the Japanese national legislature and explains our approach. Then, in the following section, we validate our sentiment scores. The penultimate section includes two substantive analyses that use our sentiment scores. Finally, we conclude the article.

2. Method

2.1. Structure of the Japanese legislative process

The Japanese national legislature, the Diet, has a bicameral system composed of the lower and upper chambers. Both the cabinet and groups of lawmakers can submit bills to either chamber, where cabinet bills dominate the legislative agendas. The Speaker of the chamber refers each submitted bill to a committee, which takes a vote after deliberating.

A "debate" in this article is an optional legislative procedure that comes between the end of deliberation and voting in a committee. A committee member debates as a representative of their respective party. In most cases, at most one member from each party (or each group of parties) debates. Members' debates are a means for expressing a position, usually through set speeches, rather than an exchange of opinions. For 33% of governmental bills, a debate took place in at least one chamber. The probability of a governmental bill being debated increases with its importance or the level of opposition from political parties (Fukumoto, 2000, 34).

As an official legislative procedure, such debates commonly have a specific structure. At the beginning of the debate, legislators explicitly indicate 1) that they deliver the current speech as a debate, 2) whether they are for or against the agenda, and 3) which party they represent. At the end of the debate, legislators usually wrap up their argument, repeating their position regarding the agenda, and explicitly declare that they are closing their debate.

After a vote is taken in the committee, the bill is reported back to the floor, where the chamber takes a vote for the final decision. Debates rarely take place on the floor. If the floor of the first chamber passes the bill, it is sent to the other chamber, which repeats the same procedure. If the second chamber approves the bill, it is enacted as a law.

2.2. Corpus

The corpus for this research consists of speeches included in the minutes of all standing committee meetings of both chambers during 1955–2021.[3] Note that most speeches in the legislature take place in committees rather than in the plenary. Our unit of observation is a "speech," noninterrupted lines of dialog spoken by a speaker. (In this article, the lowercase word "speaker" refers to an individual who delivers a speech and does not indicate the Speaker of a chamber.) We retain only speeches made by rank-and-file committee members and discard speeches made by committee chairs, government officials (ministers and bureaucrats), and witnesses, the latter of whom mostly come from the private sector. Our corpus contains 3, 148, 817 speeches in total.[4]

2.3. Sentiment score

2.3.1. Extracting labeled speeches

To make a labeled dataset, we first extract debate speeches. We regard a speech as a "debate speech" when the first sentence of the speech satisfies the following three conditions. 1) It contains the word "tōron" (debate) or the single Chinese character "i," which is often a part of "iken" (opinion) or "ishi"

[3]For details, see the Supplementary Materials.

[4]For preprocessing, see the Supplementary Materials.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

(intention). 2) It contains words indicating the speaker's position toward the agenda, either "sansei" (agree) or "hantai" (oppose) but not both.[5] Accordingly, we also label the debate speech's position as either pro or con. 3) It contains the word "daihyō" (represent), which indicates that the speaker is debating on the behalf of a party or parties. We read dozens of randomly chosen debate speeches defined above and confirmed that all of them are actually debates.

As we explained above, in the opening and closing of a debate speech, the speaker explicitly states his or her position and party. A supervised classifier would rely heavily on the information from stylized language such as "agree" and "disagree" or the party name of the speaker, ignoring the sentiment in the speech. That being the case, the resultant model would not be able to classify unlabeled speeches that lack those words. To prevent such a problem, we split every debate speech into three tertiles with a roughly equal number of sentences and use only the second tertile to train a supervised learning model.[6] For the same reason, we remove words with the characters "san" or "han", which can be part of "sansei" (agree) or "hantai" (oppose) in both debate and nondebate speeches.[7]

Ultimately, we find 1,928 pro-side debate speeches and 5,428 con-side debate speeches.[8] These are the "labeled speeches" in this article. We randomly sample 60% ($N = 4,413$) of the labeled speeches as the training dataset, 20% ($N = 1,471$) as the development dataset, and the remaining 20% ($N = 1,472$) as the test dataset.

2.3.2. *Training a supervised learning model*

Once we have labeled the speeches, we can train *any* supervised learning model. That said, in this article, we use a language representation model, BERT (Devlin *et al.*, 2018). One merit of BERT models is their ability to capture the context where every word is situated using its bidirectional transformer mechanism, which surpasses traditional bag-of-words approaches. Specifically, we employ a pretrained BERT model for the Japanese language, which can be fine-tuned for many downstream natural language processing tasks. Our downstream task is to classify speeches into pro- and con-side speeches.[9]

We fine-tune the model by using the second tertile of each speech in the training dataset so that it can predict the label of each speech. Then, we evaluate the model's performance in terms of the loss function of cross-entropy by using the second tertile of each speech in the development dataset. We originally planned to repeat this procedure for several epochs until the model performance stopped improving for three epochs in a row so that we could avoid overfitting. In fact, as the model metric did not improve after the first epoch, we use the model from the first epoch.

2.3.3. *Calculating the sentiment scores of speeches*

Once the fine-tuning of the BERT model is completed, we can process every speech through the model, regardless of whether the speech is labeled. In the last stage of processing, the fine-tuned BERT model produces the probability that the speech is a pro-side speech. We multiply the probability by 100 and use it as our sentiment score for the speech. If the sentiment score is closer to 100 (or 0), the speech is considered to be more positive (negative).

[5] For details, see the Supplementary Materials.

[6] Another reason is that two-thirds of the debate speeches exceed the maximum number (512) of tokens that the employed BERT model can process (for details, see the Supplementary Materials). We filter out 73 debate speeches with fewer than three sentences.

[7] In practice, we replace these words with a special `[MASK]` token to reduce their impact on our classification task while keeping other tokens' positions in the context unchanged.

[8] Among the pro-side (con-side) debate speeches, 56.3% (0.6%) are spoken by government members.

[9] For the details regarding the implementation of BERT, see the Supplementary Materials.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

Table 1. Classification performance of the BERT model using labeled speeches

Dataset					Classification performance		
Train	Dev.	Test	Unit	*N*	Accuracy	Pro-F1	Con-F1
X	X		2nd Tertile of each speech	5,884	0.937	0.872	0.958
		X	2nd Tertile of each speech	1,472	0.923	0.843	0.949
		X	Nonsplit speech	1,472	0.941	0.881	0.961

Note: The unit of observation is a labeled speech. We randomly sample 60% of the labeled speeches as the training dataset; 20% as the development dataset; and the remaining 20% goes to the test dataset. We also divide each labeled speech into three tertiles. In the third row, we use the original nonsplit labeled speeches.

3. Validation

For any text analysis, validation is a necessary step to check if the method works as expected (Grimmer and Stewart, 2013). In this section, we carry out a series of validity checks using the in- and out-sample labeled speeches, the unlabeled speeches, and human-based and dictionary-based sentiment scores.

3.1. Labeled speeches

Using labeled speeches, we evaluate the performance of the BERT model. As explained in the previous section, we divide each labeled speech into three tertiles, and for fine-tuning, we use only the second tertile of each labeled speech in the training and development datasets, which consist of in-sample labeled speeches. We explore the classification performance for out-sample labeled speeches, namely, the second tertiles of labeled speeches in the test dataset. The performance metrics are presented in Table 1, where we classify a speech as pro-side (or con-side) speech if the sentiment score is greater (less) than 50. Two F1 metrics, pro-F1 and con-F1, are the values of F1 for the pro and con classes, respectively.[10] Each row shows the metrics achieved for specific data subsets. For the second tertile of each labeled speech in the training and development datasets (in the first row), unsurprisingly, the accuracy and the two F1 scores are truly high (c.f., Devlin *et al.*, 2018). When we turn to the second tertile of each labeled speech in the test dataset (in the second row), the values of the three metrics are still high.

Eventually, we will scale every speech without dividing it. Thus, in the third row of Table 1, we report the three metrics for the nonsplit labeled speeches in the test dataset, which are better than their second tertile counterparts (second row) and even than those of the second tertile of each labeled speech in the training and development datasets (first row). This improved performance is probably due to the increased amount of information from the texts due to not splitting the labeled speeches. In the Supplementary Materials, we demonstrate that the model also classifies the first and third titles successfully, and the performance of the model does not change over time. These results demonstrate that our BERT model works quite well for out-sample labeled speeches.

3.2. Unlabeled speeches

We now use the unlabeled speeches for validation ($N = 3,141,461$). Since most agendas are proposed by the government and most unlabeled speeches are technically questions posed by legislators to the government, scholars expect government members' sentiments to be more positive than opposition members' sentiments (Herzog and Benoit, 2015; Lauderdale and Herzog, 2016; Rudkowsky *et al.*, 2018; Proksch *et al.*, 2019; Curini *et al.*, 2020). In our corpus, 11.9% of speeches are from government parties, and 88.1% are from the opposition.[11] Figure 1 illustrates the densities of sentiment scores of government (solid line) and opposition members (dotted line). As expected, the sentiment

[10]For the details regarding the F1 score, see the Supplementary Materials.

[11]The definition of the government party in this article is a political party with at least one member appointed as a cabinet minister or a parliamentary secretary, and other parties are opposition parties. For details, see the Supplementary Materials.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

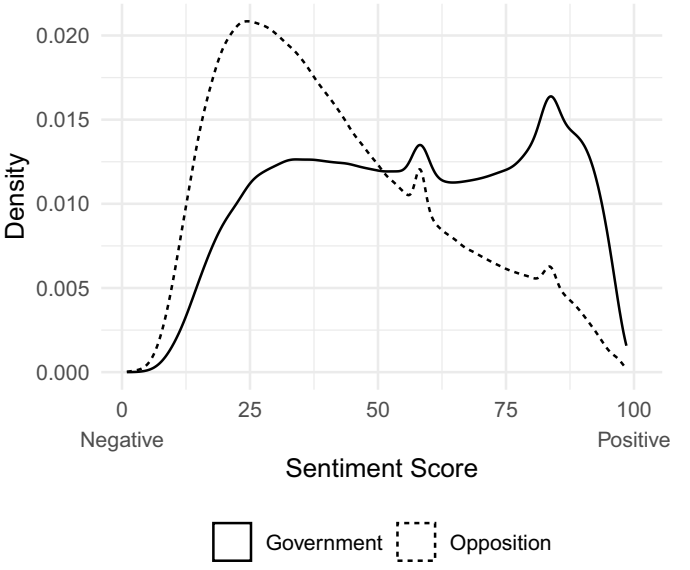


Figure 1. Sentiment score distribution by government status.
Note: The densities of sentiment scores of government and opposition members are displayed as solid and dotted lines, respectively.

scores of government members tend to be higher, while those of opposition members tend to be lower.[12] This figure also reveals a little-known substantive fact: even government members sometimes (or quite often) make negative speeches.[13]

We also expect that when a party switches from the government to the opposition (the opposition to the government), the sentiment scores of its members' speeches decrease (increase). To visualize this change, we present the time series of the annual average sentiment scores of the two largest parties (Figure 2). For each year and party, we calculate the average sentiment scores of the speeches made in the session in which the original national budget was passed.[14] As the LDP was in the government for most of the study period (except for 1994 and 2010–2012, which are shaded in the figure), we contrast the LDP (black dots, solid line) with its major opponent (gray dots, dashed line), which is the largest opposition party when the LDP is in the government and the largest party in the government when the LDP is in the opposition.[15] It is clear that the sentiment score of a party is higher when the party is a government party than when it is an opposition party. This pattern implies that our sentiment score captures not an ideal point but rather a sentiment.

[12]There are two noticeable bumps around the sentiment scores of 0.6 and 0.85. The former is a mass of six-character speech ("Owarimasu" in Japanese), which means "I stop (speaking)." The latter is the group of short speeches meaning "Thank you" (e.g., "Arigatō gozai mashita" in Japanese). It is common for either type of speech to come at the end of a series of speeches by a speaker.

[13]For details, see the Supplementary Materials.

[14]In principle, this type of session begins in December of the previous year (before the fiscal year of 1991) or January (after 1992) and ends in May or June unless it is extended or the lower chamber is dissolved. Before the fiscal year of 1991, we include speeches in December with the corresponding fiscal year.

[15]Specifically, the LDP's major opponents were the Japan Socialist Party (JSP) before 1994, the New Frontier Party during 1995–1997, the Democratic Party of Japan ("Minshu Tō" in Japanese) from 1998 to 2016 (until March 26), the Democratic Party ("Minshin Tō" in Japanese) in 2016 (from March 27) and 2017, and the Constitutional Democratic Party from 2018 on. In 1994, we use speeches by the JSP members until April 27 because the JSP did not join the Tsutomu Hata Cabinet, which started the next day.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

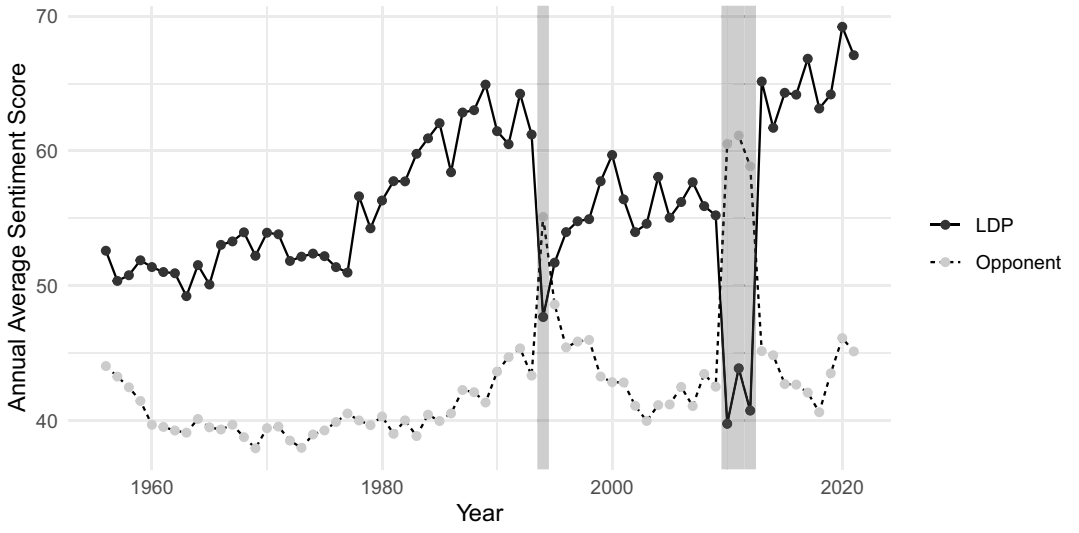


Figure 2. Annual average sentiment scores for the two largest parties.
Note: For the LDP (black dots, solid line) and its major opponent (gray dots, dashed line), we present the annual average sentiment scores of speeches made in the session in which the original national budget was passed.

3.3. Human-coding

Another way to check whether our sentiment score is a valid measure of legislators' sentiment is to examine whether it is consistent with human judgment. To obtain human-coded data, we created 12 sets of 65 pairs of speeches. Each pair was assessed by 12 graduate students, who determined which speech was more positive. The Bradley and Terry method (Bradley and Terry, 1952) was applied to estimate the latent sentiment for each speech. Then, by exploiting the human-based sentiment score as a ground truth, we explored whether, and how well, the legislator-supervised sentiment score aligned with the human-based sentiment score.

Below is the process for creating the 65 pairs of speeches.[16] We randomly chose 10 nondebate speeches for every ten-point interval of rounded sentiment scores between 10 and 90. Since few nondebate speeches have scores below 0.5 or above 99.5, we randomly chose 10 speeches with scores of 0 (or 100) from nondebate speeches with scores between 1 and 2 (or between 98 and 99). We also randomly chose 10 second-tertiles of pro-side debate speeches and 10 second-tertiles of con-side debate speeches. Thus, we obtained $10 \times 11 + 10 \times 2 = 130$ speeches. For each human coder, we generated randomly matched 65 pairs of speeches. The order of 65 pairs was also randomized.[17]

We estimated the sentiment score of each of the 130 speeches included in the 65 pairs by applying the Bradley–Terry model with no explanatory variables. This model assumes that there is a latent score attached to each speech and that for each pair of speeches, a coder is more likely to choose the speech with a higher score. We refer to the latent score for each speech assigned by the model as the human-based sentiment score; below, we refer to our sentiment score as the legislator-supervised sentiment score.

Figure 3 shows the relationship between the human-based sentiment score and the legislator-supervised sentiment score. In the left panel, the box plots display the distribution of the human-based sentiment scores for a group of the 10 nondebate speeches corresponding to each tenth rounded value (0, 10, …, 100) of legislator-supervised sentiment scores, arranged from left to right. Overall,

[16]We sampled 130 speeches from speeches containing 500 or fewer tokens.
[17]In addition, we created 10 pairs of speeches for validation and attention checks. For the details regarding human coding, see the Supplementary Materials.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

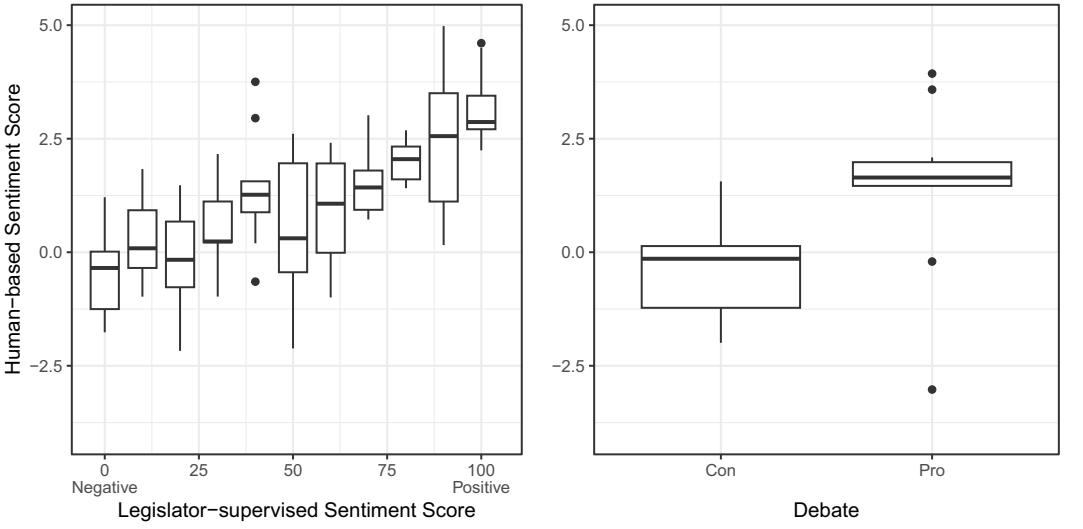


Figure 3. Box plots of human-based sentiment scores.
Note: In the left panel, each box plot represents the distribution of the human-based sentiment scores for a group of the 10 nondebate speeches for each tenth rounded value (0, 10, …, 100) of legislator-supervised sentiment score, arranged from left to right. In the right panel, the box plots represent the distribution of the human-based sentiment scores for the 10 pro-side and the 10 con-side debate speeches in the right and left columns, respectively.

the human-based sentiment scores are positively related to the legislator-supervised sentiment scores (correlation coefficient of 0.646).[18]

The right panel of the figure provides the box plots of the human-based sentiment scores for the 10 pro-side and the 10 con-side debate speeches in the right and left columns, respectively. The human-based sentiment scores of the pro-side debate speeches tend to be positive, while those of the con-side debate speeches tend to be negative.

In the Supplementary Materials, we also demonstrate that the legislator-supervised sentiment score explains the judgements of the human coders very well. Overall, human coding has shown that the legislator-supervised sentiment score successfully represents legislators' sentiment.

3.4. Dictionary-based sentiment score

As a final validity check, we compare the legislator-supervised sentiment scores with those calculated from existing dictionaries. As we noted before, the performance of dictionaries in text analysis is domain-specific and may underperform in some domains of application. We examine the case of parliamentary speech.

The landscape of sentiment dictionaries in Japanese is very different from that in English, where there are a number of existing reputable dictionaries. There is no standard dictionary of this kind in Japanese. Nevertheless, we employ the dictionary by Takamura *et al.* 2005 because it is relatively popular.[19] We calculate the dictionary-based sentiment score for each speech as an average of the scores of words and phrases matched to the dictionary, weighted by the inversed document frequency.[20]

We compare three sentiment scores we so far have obtained: legislator-supervised, human-based, and dictionary-based. To evaluate their relationship, we calculate the correlation coefficients between

[18]We discuss the value of the correlation coefficient in the Supplementary Materials.
[19]The dictionary data is available at http://www.lr.pi.titech.ac.jp/~takamura/pndic_en.html.
[20]We use the `textstat_valence` function of the `quanteda.sentiment` package for this calculation (Benoit *et al.*, 2018).

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

them. When focusing on the correlation coefficients with the human-based sentiment score, the legislator-supervised sentiment score has a higher value (0.646) than the dictionary-based sentiment score does (0.195). This finding suggests that the legislator-supervised sentiment score extracts the sentiment of parliamentary speeches more successfully than the dictionary-based sentiment score does.

In the Supplementary Materials, we also show that a legislator-supervised sentiment score model outperforms a dictionary-based sentiment score model in terms of classification task performance using human-coded data.

4. Analyses

Having established the validity of our sentiment score in the previous section, we now present two substantive analyses using this score.

4.1. Determinants of sentiment

In this first analysis, we explore how electoral and legislative factors affect legislators' sentiments.

4.1.1. Electoral system

Until the 1993 election, the lower chamber used the single nontransferable voting (SNTV) system, where the district magnitude M is three to five in most districts, each citizen has just one vote, votes cannot be transferred between candidates, and the seats are simply awarded to the top M vote-getters in each district. After the 1996 election, the chamber adopted a two-tier system. The lower tier consists of single-member districts (SMD, $M = 1$), while the upper tier is composed of closed-list proportional representation (PR) districts ($6 \leq M \leq 33$). The upper chamber employs another two-tier system. The lower tier employs the SNTV system ($1 \leq M \leq 6$). The upper tier consists of the national at-large district that used SNTV ($M = 50$) until 1980, closed-list PR ($M = 50$) from 1983 to 1998, and open list PR ($48 \leq M \leq 50$) after 2001.

4.1.2. Data

The unit of observation is a speech. We use both labeled and unlabeled speeches. The dependent variable, *Sentiment Score*, is our legislator-supervised sentiment score. Below, we introduce four independent variables, referring to prior research (Herzog and Benoit, 2015; Proksch *et al.*, 2019).[21]

To begin, for members of the lower chamber, we define the variable *Seniority* as the number of terms they have served. The term of members of the upper chamber is six years, which is approximately twice as long as the average years (2.8 years) for which members of the lower chamber actually served during the study period given terminations due to early elections.[22] Thus, for members of the upper chamber, we calculate Seniority by doubling the number of terms they served, and we subtract one from this number when they are still in the first half of their current term (c.f., Sato and Matsuzaki, 1986, 366). Senior legislators deploy sentiment in their speeches more efficiently and strategically than junior legislators (Osnabrügge, Hobolt and Rodon, 2021). Therefore, the coefficient of Seniority is expected to be positive for the government and negative for the opposition.

Second, we define the variable *Electoral Time* as the number of years that have passed between the day of the speech and the election day when the legislator was most recently (re-)elected.

[21] For the data sources, see the Supplementary Materials. We exclude 81, 740 (2.6%) speeches whose speakers' names cannot be matched to those in the data sources of the independent variables or whose speakers were elected to fill vacancies.

[22] The term of the lower chamber members is four years.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

Our prediction is that as the next election approaches, government and opposition members increase the contrast with their competitors by expressing positive and negative sentiments in their speeches, respectively (Proksch *et al.*, 2019; Crabtree *et al.*, 2020; Ishima, 2024, 122). Accordingly, as Electoral Time increases, sentiment scores become higher for the government and lower for the opposition.

Third, we define the variable *Legislative Time* as the number of months that have passed between the day of the speech and the beginning of the current session. Usually, a few sessions are held every year. We expect that toward the end of each session, government and opposition members compromise on bills and thus tone down the sentiment in their speeches. Therefore, the higher the Legislative Time, the lower and higher the sentiment scores that government and opposition members tend to have, respectively.

Finally, we use the variable *Electoral Strength*, the number of votes received by a legislator in the previous election divided by the Droop quota of his or her district (Cox and Rosenbluth, 1995. See also Reed and Scheiner, 2003; Desposato and Scheiner, 2008, and Saito, 2009).[23] This is missing for legislators elected from closed-list PR districts. We expect that the sentiment scores of electorally vulnerable government (opposition) members are high (low) because such legislators try to appeal to the electorate by speaking in a charged manner. Thus, the coefficient of Electoral Strength will be negative for the government and positive for the opposition.

4.1.3. Results

As explained in the previous subsubsection, we expect that the independent variables affect sentiment in a different way depending on the type of legislators. Accordingly, we divide the whole dataset by government status and chamber (and electoral system) and analyze each data subset separately.

Figure 4 presents the coefficient plots.[24] Each row of panels corresponds to an independent variable. The left and right columns of panels correspond to the cases of government and opposition members, respectively. In each panel, the three black dots indicate the lower chamber, while the three white dots indicate the upper chamber. In the first analysis for each chamber (Analysis 1 for Lower Chamber and Analysis 4 for Upper Chamber), indicated in circles, we regress the sentiment score on Seniority, Electoral Time, and Legislative Time as well as legislator fixed effects but not on Electoral Strength so that we can analyze legislators from all types of electoral systems, including PR legislators. In the second and third analyses (triangle and square) for each chamber, we include Electoral Strength as an additional independent variable. In the case of the lower chamber, Analyses 2 and 3 use legislators elected from SNTV and SMD, respectively. In the case of the upper chamber, Analyses 5 and 6 use legislators elected from the lower tier and the pre-1980 upper tier under the nationwide SNTV, respectively.[25] The dots show the point estimates of the coefficients with the bars for the 95% confidence intervals calculated from the standard errors clustered by legislators.

For opposition members, Seniority has significantly negative effects, as anticipated, while for government members, the effects are insignificant. All coefficients of Electoral Time are negative and significant for opposition members, as expected. However, contrary to our expectations, they are mostly insignificant for government members. In Analyses 1 and 3 of the lower chamber, the effects of Legislative Time are significantly negative for government members and significantly positive

[23]The Droop quota is the minimum number of votes to guarantee a win, which is obtained by dividing the total number of valid votes cast in the district j (V_j) by the district magnitude plus one ($V_j/(M_j+1)$). Electoral Strength is comparable across districts with different magnitudes.

[24]For the exact values of the coefficient estimates and the standard errors, see the Supplementary Materials.

[25]We do not analyze the post-1980 upper tier under PR systems, as Electoral Strength cannot be defined in a comparable manner.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

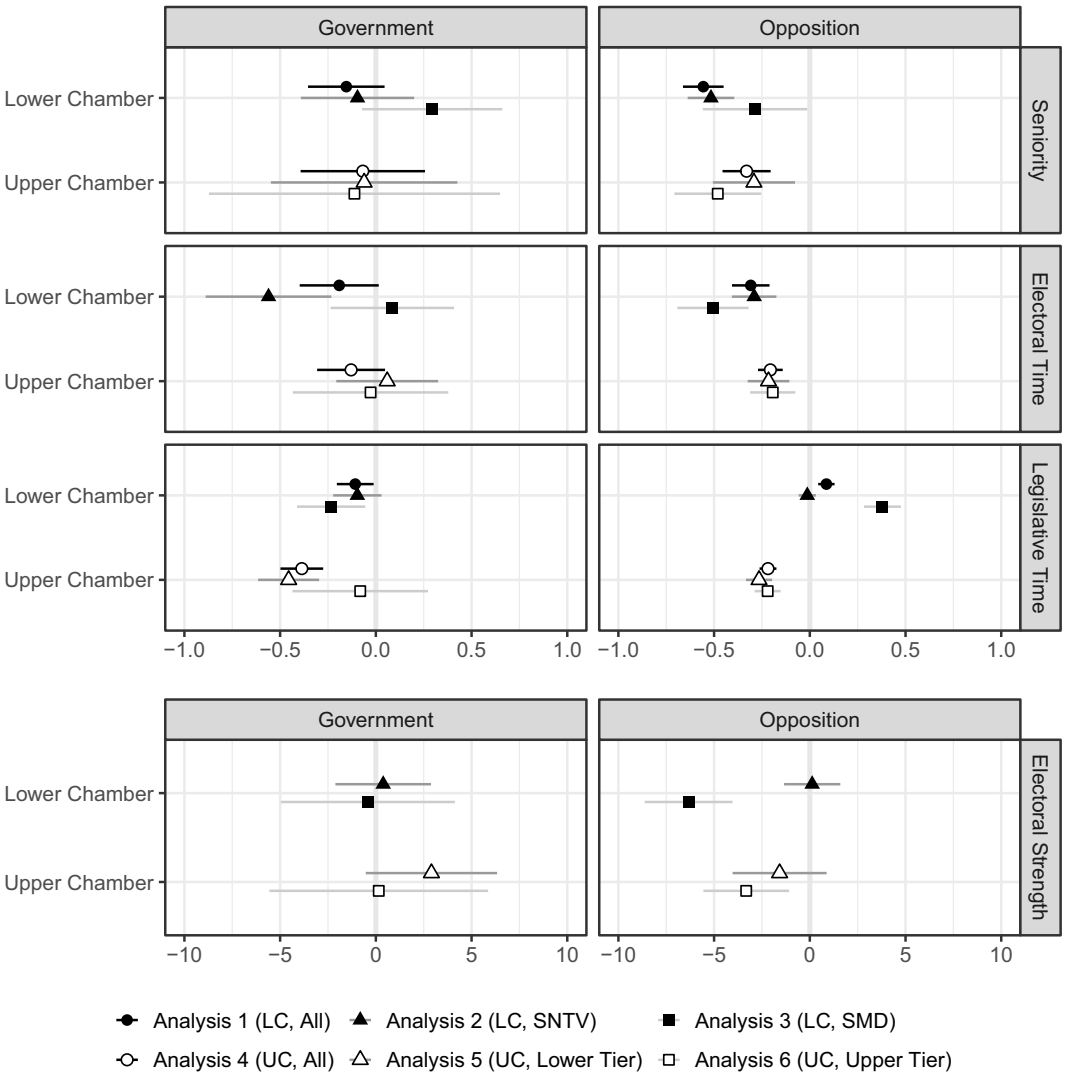


Figure 4. Summary of regression results: coefficient estimates.
Note: The unit of observation is a speech. We regress Sentiment Score on the three or four independent variables as well as legislator fixed effects. Each row of panels is for an independent variable. The left (right) panel deals with government (opposition) members' speeches. Horizontal bars around point estimates are the 95% confidence intervals derived from standard errors clustered by legislators. Three analyses are conducted for each chamber. For the lower chamber (LC, top three rows in each panel), Analysis 1 uses all speeches. In Analyses 2 and 3, we analyze speeches by legislators elected from SNTV and SMD, respectively. For the upper chamber (UC, bottom three rows in each panel), Analysis 4 uses all speeches. In Analyses 5 and 6, we analyze speeches by legislators elected from the lower tier and the pre-1980 upper tier under the nationwide SNTV, respectively.

for opposition members, which is consistent with our argument. In the upper chamber, the coefficients of Legislative Time remain (mostly significantly) negative for government members, although surprisingly, they turn significantly negative for opposition members. Electoral Strength has insignificant coefficients for government members and (sometimes significantly) negative coefficients for opposition members, which we did not expect. In summary, it is fair to conclude robustly that opposition members tend to express negative sentiments when they are senior or the next election is approaching.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

4.2. The rebellion and defection of LDP members in 1993

In the second analysis, we explore whether the sentiment score predicts legislators' behavior at the historic 1993 vote of no confidence in the Ki'ichi Miyazawa Cabinet and the following split-up of the LDP. In principle, parliamentary democracies rarely experience rebellion (Kam, 2009). Japan is an extreme case: during the LDP government periods, the party has experienced as few as four instances of rebellions where some members voted against the party discipline.[26] Among them, only the vote of no confidence in 1993 resulted in a change of the government parties. This is why we focus on this case. If our sentiment score truly measures the heterogeneity of sentiments even among government members, it should be helpful for predicting who would rebel and/or defect.[27.]

4.2.1. Background

In 1993, the LDP was divided over "political reform."[28] The main issue was whether it would change the SNTV electoral system of the lower chamber. In the 126th session of the legislature (from January 22 to June 18, 1993), the pro-political reform group of LDP members (called "reformers") managed to submit political reform bills to the lower chamber. Opposition members also proposed their own political reform bills. The cabinet did not submit any bills on this matter. Both sets of bills were referred to the Special Committee for Research on Political Reform (hereafter, the Reform Committee). There, advocates of the reform argued for the bills, while the anti-reform group of LDP members resisted them.

To the disappointment of the reformers, however, the leadership of the LDP decided not to pass the bills toward the end of the session. Opposition parties submitted a motion of no confidence in the Miyazawa Cabinet. As the LDP had a majority in the lower chamber, it would have defeated such a motion under normal circumstances. This time, however, the motion was passed because a number of the reformers, mostly members of the Hata faction, rebelled and supported the motion. Prime Minister Miyazawa then dissolved the lower chamber. As many expected, 44 members from the Hata faction defected from the LDP and formed a new party, the Japan Renewal Party ("Shinseito"). However, it came as a surprise when an additional 10 lawmakers, most of whom did not rebel in the vote of no confidence, also left the LDP and launched the New Party Harbinger ("Shinto Sakigake," hereafter the NPH). After the general election, these two new parties and six other ex-opposition parties formed a coalition government that ousted the LDP from power for the first time in almost four decades.

4.2.2. Data

The unit of observation is an LDP member who spoke at the Reform Committee in the 126th session ($N = 39$). We have two dependent variables. One is a dummy variable of whether the LDP member supported the motion of no confidence in the Miyazawa Cabinet on June 18, 1993 (*Rebel*). There were 25 nay votes, 8 yes votes, and 6 abstentions, the last of which we regarded as missing values. The other dependent variable is a dummy variable of whether the legislator defected from the LDP in the general election on July 18, 1993 (*Defect*). Since every member ran for reelection, there are no missing values. Note that these two dependent variables are slightly different; all eight rebels defected, five of six absentees did not defect from the LDP, and four members did not rebel but did defect (to the NPH).

The main independent variable, *Sentiment Score*, is the average of our legislator-supervised sentiment scores of the LDP member's speeches at 18 meetings of the Reform Committee (January 22 to

[26]They were the votes of no confidence in 1980 and 1993, the nomination of the Prime Minister in 1979, and the postal reform bills in 2005.

[27.]Slapin and Kirkland (2020) show that a few linguistic characteristics of MP speeches are predictive of rebellion in the House of Commons of the United Kingdom.

[28]For details, refer to Cox and Rosenbluth (1995); Kato (1998); Reed and Scheiner (2003), and Saito (2009).

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

Table 2. Regression of LDP members' rebellion and defection in 1993

	Dependent variable:	
	Rebel	Defect
Intercept	−17.771*	−10.116*
	(7.693)	(4.765)
Legislator-supervised Sentiment Score	0.117*	0.060
	(0.051)	(0.034)
Seniority	−0.164	−0.174
	(0.252)	(0.198)
Electoral Strength	12.842	8.007
	(7.021)	(4. 665)

Note: $^{*}p < 0.05$. The unit of observation is an LDP member on the Reform Committee in 1993. We regress Rebel or Defect on the three independent variables. We report standard errors in parentheses.

May 25, 1993) in the 126th session. Note that the sentiment score represents the sentiment toward the political reform bills, not toward the government, because all of the agendas in the Reform Committee were the political reform bills.[29] Since the leadership of the LDP killed the bills at the last minute, we expect that the higher Sentiment Score is, the more likely the LDP member is to have rebelled or defected. Note also that there were no debate speeches because no vote over the bills was taken at the Reform Committee.

The control variables are Seniority and Electoral Strength (in the 1990 general election), as defined in the previous subsection, which prior research also refers to as determinants of the same dependent variables (Cox and Rosenbluth, 1995; Kato, 1998; Reed and Scheiner, 2003; Desposato and Scheiner, 2008; Saito, 2009).

4.2.3. Results

We employ logistic regression models. Table 2 displays estimates of the coefficients and the standard errors in brackets. In the left column, we report the results where the dependent variable is Rebel. The coefficient of Sentiment Score is significantly positive, as expected. That is, the more positive an LDP member's speech is toward the political reform bills, the more likely the member is to have defied the party in the vote of no confidence. The coefficients for Seniority and Electoral Strength are not significant. In the left panel of Figure 5, the line shows the predicted probabilities of rebellion by Sentiment Score, with Seniority and Electoral Strength at their respective median values. The tick marks at the top and bottom represent Sentiment Scores of rebelling and nonrebelling legislators, respectively. The finding that sentiments are predictive of rebellions is not trivial in light of the previous research. For instance, Slapin and Kirkland (2020, 165) report that "[t]here is little clear pattern between the presence of negative and positive sentiment and rebellion."

A similar pattern arises in the analysis of defection from the LDP as the dependent variable (right column of Table 2 and right panel of Figure 5). The effect of Sentiment Score is positive, which is consistent with our argument, while it comes near reaching a conventional level of significance ($p = 0.079$). The effects of Seniority and Electoral Strength are not significant.

Overall, our findings imply that observers should be able to predict who is likely to rebel or defect if they listen to legislators' speeches carefully.

[29] Prior research had difficulty finding a good measure of members' preference for the political reform and substituted the signature to a statement on the political reform on December 18, 1992, as a compromise (Reed and Scheiner, 2003; Desposato and Scheiner, 2008; Saito, 2009).

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

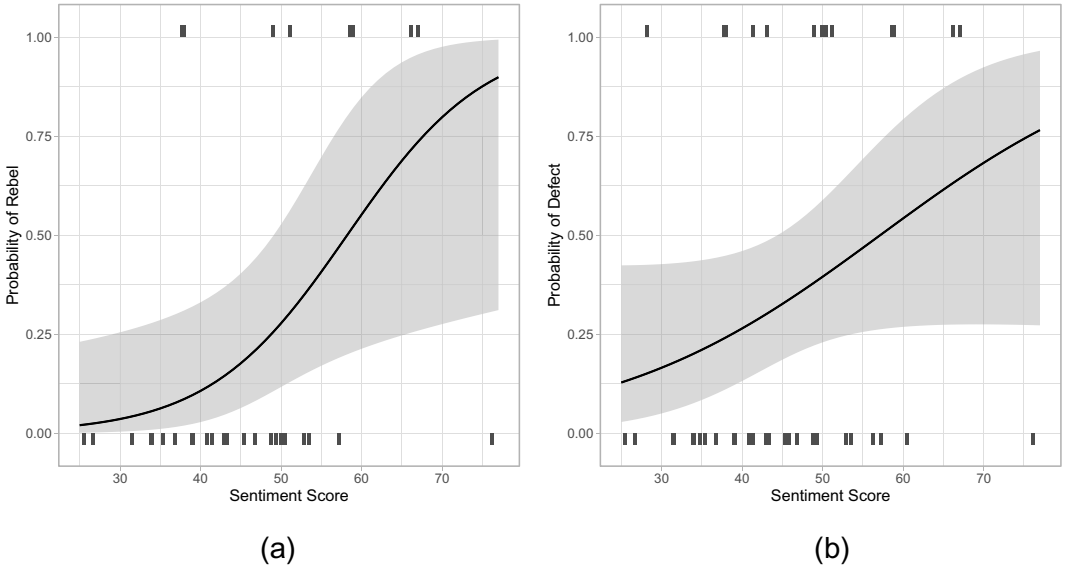


Figure 5. Predicted probabilities of LDP members' rebellion and defection in 1993 by Sentiment Score. (a) Rebel and (b) Defect.

Note: The lines show the predicted probabilities of rebellion (left) and defection (right) on the basis of Sentiment Score, with Seniority and Electoral Strength at their average values. The gray-shaded areas around the lines represent the 95% confidence intervals for the predictions. The black tick marks at the top and bottom represent the sentiment scores of rebelling (or defecting) and nonrebelling (or nondefecting) legislators, respectively.

5. Conclusion

The general message of this article in a broader context is simple: for scholars of text analysis who conduct supervised learning and thus need labeled texts, it is advisable to explore and exploit, if any, training texts that have already been labeled (as either positive or negative in the case of analysis of sentiment polarity) by those who generate the texts (Pang *et al.*, 2002; Fang and Zhan, 2015; Nguyen *et al.*, 2018; Fernandes *et al.*, 2019). We showcase a good example of the Japanese legislature, where the highly structured debates generate such prelabeled texts available for scientific data analysis.

In our substantive analyses, we show that the sentiment score is useful both in studying the nature of party politics over time and in explaining legislators' behavior in a historic case. In the first analysis, it turns out that the seniority of legislators affects sentiment polarization by making senior opposition legislators more negative in their speeches; in addition, as the next election approaches, opposition members express more negative sentiments. In the second analysis, government members who expressed positive sentiment toward the political reform bills tended to vote for the motion of no confidence in the cabinet that killed the bills.

The core idea of our process for extracting sentiments from prelabeled debate speeches is applicable to legislative corpora in countries other than Japan. As Proksch and Slapin (2015, 1) observed, "in all parliaments, members debate bills before they vote on them." For instance, in the U.S. House of Representatives, after a committee reports a bill to the floor, one hour of general debate on the bill follows on the floor. This time is divided evenly between the majority and minority parties, their lawmakers deliver prepared speeches, and the party leaders are informed of "member *sentiment* and the mood of the House" toward the bill (Oleszek *et al.*, 2020, 203–205, 211–213, 217, emphasis added). Other examples are corpora consisting of debates that have taken place in the European Parliament (Benoit *et al.*, 2016) and the Swiss National Council (Schwarz *et al.*, 2017). For these corpora, scholars may be able to exploit the associated debates as automatically labeled texts, as we do for the Japanese

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

Diet. Or at least, hopefully, political scientists can look for patterns in speech where, under a small set of assumptions, labels can be reliably extracted.

Some readers may question whether labeled and unlabeled speeches are comparable. Their comparability is the basis for the effectiveness of our classifier when applied to unlabeled speeches. A few of our responses are presented below. First, we detect sentiments that are expressed in the same way as sentiments presented by legislators in closing debates, which are politically salient. The types of sentiments expressed in these speeches are precisely what we (and we believe many scholars) are interested in. Second, the Supplementary Materials show that the saliency of debate speeches does not affect the performance of our classifier. Third, most unlabeled speeches are *technically* questions, although they are mostly legislators' arguments made under the name of "questions" akin to labeled speeches (i.e., debate speeches). That said, our classifier may be less effective when applied to unlabeled speeches that are less salient or segments of unlabeled speeches that are purely procedual questions.[30]

In summary, we hope that our article contributes to the literature on sentiment analysis and enhances the understanding of legislators' behaviors, particularly under parliamentary systems.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psrm.2025.10048. To obtain replication material for this article, please visit https://doi.org/10.7910/DVN/6K5E3Y.

Acknowledgements. We would like to thank Diego Escobari, Etienne Gagnon, Masataka Harada, Musashi Hinck, Silvia Kim, Da-chi Liao, Go Murakami, Marlene Mußotter, Ludovic Rheault, Jonathan Slapin, and Arthur Spirling for their helpful feedback. We appreciate American Journal Experts for providing their proofreading service. We are entirely responsible for the scientific content of the article, and the article adheres to the journal's authorship policy.

Funding statement. This work was supported by the Japan Society for the Promotion of Science [grant numbers KAKENHI JP16K13340 and JP19K21683]. The financial sponsor played no role.

Competing interests. The authors declare none.

References

Ballard AO, DeTamble R, Dorsey S, Heseltine M and Johnson M (2023) Dynamics of polarizing rhetoric in congressional tweets. *Legislative Studies Quarterly* **48**, 105–144.

Benoit K, Conway D, Lauderdale BE, Laver M and Mikhaylov S (2016) Crowd-sourced text analysis: Reproducible and agile production of political data. *American Political Science Review* **110**, 278–295.

Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S and Matsuo A (2018) Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software* **3**, 774.

Bestvater SE and Monroe BL (2023) Sentiment is not stance: Target-aware opinion classification for political text analysis. *Political Analysis* **31**, 235–256.

Bradley RA and Terry ME (1952) Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39**, 324–345.

Cox GW and Rosenbluth FM (1995) Anatomy of a split: The liberal democrats of Japan. *Electoral Studies* **14**, 355–376.

Crabtree C, Golder M, Gschwend T and Indridason IH (2020) It is not only what you say, it is also how you say it: The strategic use of campaign sentiment. *Journal of Politics* **82**, 1044–1060.

Curini L, Hino A and Osaki A (2020) The intensity of government–opposition divide as measured through legislative speeches and what we can learn from it: Analyses of Japanese parliamentary debates, 1953–2013. *Government and Opposition* **55**, 184–201.

Desposato S and Scheiner E (2008) Governmental centralization and party affiliation: Legislator strategies in Brazil and Japan. *American Political Science Review* **102**, 509–524.

Devlin J, Chang M-W, Lee K and Toutanova K (2018) BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* **1**:4171–4186.

Fang X and Zhan J (2015) Sentiment analysis using product review data. *Journal of Big Data* **2**: Online.

Fernandes JM, Goplerud M and Won M (2019) Legislative bellwethers: The role of committee membership in parliamentary debate. *Legislative Studies Quarterly* **44**, 307–343.

[30]For details, see the Supplementary Materials.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

Fukumoto K (2000) *Nihon no Kokkai Seiji: Zen Seifu Rippō no Bunseki [Politics in the Japanese Diet: A Statistical Analysis of Postwar Government Legislation]*. Tokyo: University of Tokyo Press.

Grijzenhout S, Marx M and Jijkoun V (2014) Sentiment analysis in parliamentary proceedings. In Kaal B, Maks I & van Elfrinkhof A (eds.) *From Text to Political Positions: Text Analysis across Disciplines* Amsterdam: John Benjamins Publishing Company 117–134.

Grimmer J and Stewart BM (2013) Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* **21**, 267–297.

Hargrave L and Blumenau J (2022) No longer conforming to stereotypes? Gender, political style and parliamentary debate in the UK. *British Journal of Political Science* **52**, 1584–1601.

Haselmayer M and Jenny M (2017) Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding. *Quality and Quantity* **51**, 2623–2646.

Herzog A and Benoit K (2015) The most unkindest cuts: Speaker selection and expressed government dissent during economic crisis. *Journal of Politics* **77**, 1157–1175.

Hopkins DJ and King G (2010) A method of automated nonparametric content analysis for social science. *American Journal of Political Science* **54**, 229–247.

Ishima H (2024) Talking like opposition parties? Electoral proximity and language styles employed by coalition partners in a mixed member majoritarian system. *Legislative Studies Quarterly* **49**, 721–740.

Kam C (2009) *Party Discipline and Parliamentary Politics.* Cambridge: Cambridge University Press.

Kato J (1998) When the party breaks up: Exit and voice among Japanese legislators. *American Political Science Review* **92**, 857–870.

Lauderdale BE and Herzog A (2016) Measuring political positions from legislative speech. *Political Analysis* **24**, 374–394.

Laurer M, van Atteveldt W, Casas A and Welbers K (2024) Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and BERT-NLI. *Political Analysis* **32**, 84–100.

Laver M, Slava M and Benoit KR (2012) Coder reliability and misclassification in the human coding of party manifestos. *Political Analysis* **20**, 78–91.

Nguyen H, Veluchamy A, Diop M and Iqbal R (2018) Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches. *SMU Data Science Review* **1**, Online.

Oleszek WJ, Oleszek MJ, Rybicki E and Heniff B (2020) *Congressional Procedures and the Policy Process.* (11 ed.) Thousand Oaks: Sage, CQ Press.

Osnabrügge M, Hobolt SB and Rodon T (2021) Playing to the gallery: Emotive rhetoric in parliaments. *American Political Science Review* **115**, 885–899.

Pang B, Lee L and Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing* 79–86.

Proksch S-O, Lowe W, Wäckerle J and Soroka S (2019) Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. *Legislative Studies Quarterly* **44**, 97–131.

Proksch S-O and Slapin JB (2015) *The Politics of Parliamentary Debate.* Cambridge: Cambridge University Press.

Quinn KM, Monroe BL, Colaresi M and Radev DR (2010) How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science* **54**, 209–228.

Reed SR and Scheiner E (2003) Electoral incentives and policy preferences: Mixed motives behind party defections in Japan. *British Journal of Political Science* **33**, 469–490.

Rheault L, Beelen K, Cochrane C and Hirst G (2016) Measuring emotion in parliamentary debates with automated textual analysis. *PLoS One* **11** e0168843.

Rheault L and Cochrane C (2020) Word embeddings for the analysis of ideological placement in parliamentary corpora. *Political Analysis* **28**, 112–133.

Rice DR and Zorn C (2021) Corpus-based dictionaries for sentiment analysis of specialized vocabularies. *Political Science Research and Methods* **9**, 20–35.

Rudkowsky E, Haselmayer M, Wastian M, Jenny M, Emrich Štefan and Sedlmair M (2018) More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures* **12**, 140–157.

Saito J (2009) Infrastructure as the magnet of power: Explaining why Japanese legislators left and returned to the LDP. *Journal of East Asian Studies* **9**, 467–493.

Sato S and Matsuzaki T (1986) *Jimintō Seiken [Liberal Democratic Party Government]*. Tokyo: Chūō Kōron Sha.

Schwarz D, Traber D and Benoit K (2017) Estimating intra-party preferences: Comparing speeches to votes. *Political Science Research and Methods* **5**, 379–396.

Slapin JB and Kirkland JH (2020) The sound of rebellion: Voting dissent and legislative speech in the UK House of Commons. *Legislative Studies Quarterly* **45**, 153–176.

Takamura H, Inui T and Okumura M (2005) Extracting semantic orientations of words using spin model. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)* 133–140.

Takikawa H and Sakamoto T (2020) The moral-emotional foundations of political discourse: A comparative analysis of the speech records of the U.S. and the Japanese legislatures. *Quality and Quantity* **54**, 547–566.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

Watanabe K (2021) Latent semantic scaling: A semisupervised text analysis technique for new domains and languages. *Communication Methods and Measures* **15**, 81–102.

Widmann T and Wich M (2023) Creating and comparing dictionary, word embedding, and transformer-based models to measure discrete emotions in German political text. *Political Analysis* **31**, 626–641.

Young L and Soroka S (2012) Affective news: The automated coding of sentiment in political texts. *Political Communication* **29**, 205–231.

https://doi.org/10.1017/psrm.2025.10048 Published online by Cambridge University Press

Cite this article: Matsuo A and Fukumoto K (2026) Legislators' sentiment analysis supervised by legislators. *Political Science Research and Methods* **14**, 1200–1216. https://doi.org/10.1017/psrm.2025.10048