PNAS RESEARCH ARTICLE | CLASSIFICATION

Analysis of the Evolution of Alliances and Polarization in the French National Assembly through MPs' Votes and Speeches (14th–17th Legislatures)

Rémi Berthelot (s254144)[a,1], Antoine Bois-Berlioz (s252363)[a], and Pierre Hollebèque (s254136)[a]

This manuscript was compiled on December 12, 2025

France is currently facing a political crisis characterized by legislative gridlock and the disruption of the traditional left-right cleavage following the rise of En Marche and political extremes. Although qualitative assessments exist, a quantitative framework to track the structural evolution of alliances and polarization leading to this stalemate is lacking. Here we analyze the National Assembly from the 14th to the 17th legislature using a network approach based on official roll-call votes. We construct graphs where edges represent the percentage of shared votes and apply Louvain community detection to identify empirical alliances over time, complemented by an adapted TF-IDF text analysis to determine their linguistic priorities. Our results quantify the fragility of historic parties favoring a central bloc, followed by the rise of extremes and a structural distancing between groups that explains the current absence of a majority. This study offers a scalable framework to observe long-term legislative trends.

Political Polarization, French National Assembly, Network Analysis, Community Detection, Natural Language Processing (NLP).

Significance

Political polarization is a central concern for France and other democracies. Beyond general perceptions, this study investigates the practical evolution of legislative behavior in the French National Assembly from the 14th to the 17th legislature. By analyzing co-voting patterns and speech topics, we observe the formation of alliances and the subjects prioritized by different communities. This approach allows us to verify empirically whether there is a structural shift in political organization and an observable polarization in both legislative groupings and parliamentary discourse.

France, like other Western democracies, has recently been grappling with significant political crisis. This instability is characterized by government fragility and legislative gridlock, resulting from an inability to build consensus on common programs due to deep partisan disagreements. Although the challenges of compromise are not new, they appear to have intensified in recent years and months. This phenomenon is largely driven by structural shifts in the political landscape over time. Indeed, the emergence of new parties—most notably the creation of Emmanuel Macron's *En Marche*—has disrupted the traditional left-right cleavage associated with historic parties. This reconfiguration is also evident in the balance of power: the traditional parties (left: *Le Parti Socialiste*; and right: *Les Républicains*) no longer hold a majority in the National Assembly.

In this shifting context, partisan priorities and alliances have evolved, revealing a form of polarization that has been addressed qualitatively. In this study, we propose a quantitative framework to analyze these phenomena. We examine the evolution of alliances from the 14th to the 17th legislature using roll-call votes, applying network analysis. The aim is to detect communities of deputies with similar voting patterns and compare them to formally declared alliances. To achieve this, we employ clustering algorithms, specifically the Louvain method, on a graph where nodes represent deputies and edges denote the percentage of shared votes. Furthermore, modularity calculations are performed to compare the detected communities with official party affiliations. This approach allows us to measure polarization by tracking the proportion of shared votes between individual deputies, parties, and alliances over time. Finally, to understand the underlying rationale for these community formations, we analyzed parliamentary debates using Natural Language Processing (NLP), based on TF-IDF weighting, to identify recurring themes and shared vocabulary within these communities.

Author affiliations: [a]02805 Social Graphs and Interactions, MSc Student, Technical University of Denmark, DK-2800 Kongens - Lyngby, Denmark

---

All members contributed to the report's writing.
Antoine Bois-Berlioz was in charge of the individual vote analysis.*
Pierre Hollebèque was is charge of the cluster analysis.*
Rémi Berthelot was in charge of the textual analysis.*
Antoine Bois-Berlioz primarily contributed to data preprocessing and collection. **The other members** also carried out part of the preprocessing.
* 'In charge' means that this member was the final reporter for this section. However, all members cross-validated with each other and actively contributed to this report.
Data, Materials, and Software Availability: The report project is presented on this github project: github.com. The data has been scraped from the official `French National Assembly Data Website` also using the official `API Endpoint`.

[1] **Contact**: s254144@dtu.dk

Results

1. Data

1.1. Processing. To achieve this, we used open data from the `National Assembly`. We focused on the 14th, 15th, 16th, and 17th legislatures, as these are the only ones fully digitized. To better observe trends over time, each legislature is analyzed separately. The study covers the period from 2012 to the present, typically characterized by five-year mandates. The 16th legislature marks an exception, as the presidential dissolution of the Assembly divided the standard term into two separate legislatures (the 16th and 17th).

For each term, we downloaded a ZIP file containing the voting records. The file structure differed for the 14th legislature, which used a single JSON file for all votes. The later legislatures used separate files for each vote. We also extracted deputy information, including their political groups. Since deputies sometimes change groups during a term, we only recorded their most recent affiliation.

We processed this data to create two final files for each legislature. The first file records how each deputy voted: For, Against, or Abstention. The second file contains deputy details and a complete record of their speeches. To gather the speeches, we linked every vote to its specific session and fetched the transcripts via the `API endpoint`. We then used pattern recognition to extract every speech and comment made by each deputy.

	14th Leg.	15th Leg.	16th Leg.	17th Leg.
Groups	9	13	11	12
Deputies	642	689	643	646
Votes	1354	4417	4106	3404

Table 1. Summary of the dataset extracted for each legislature, detailing the number of political groups, the total count of unique deputies involved, and the total volume of recorded votes processed.

1.2. Quality. In Table 1, the 14th and 15th legislatures show fewer recorded votes because many decisions were made by a show of hands, which creates no digital record. Public ballots are only recorded when requested by a quarter of the assembly. The opposition often uses this request to slow down debates. Consequently, the 16th and 17th legislatures contain three to four times more data, even though the terms were shorter due to the dissolution of the 16th assembly in 2024.

Although the National Assembly is strictly limited to 577 seats, the dataset records a higher number of unique deputies for each legislature. This discrepancy arises because substitute members frequently replace elected deputies during the term. These replacements typically occur when a deputy is appointed to the government, passes away, or resigns due to incompatible mandates. Consequently, a single seat is often occupied by multiple individuals sequentially, increasing the total count of deputies in the data.

To validate data quality, we analyzed histograms showing the proximity of deputies within their own parties and relative to other parties, as displayed in Figure 1. We observed that internal party cohesion increased over time, while proximity between different parties decreased. This trend peaked during the 16th legislature, where the relative majority enforced strict voting discipline and

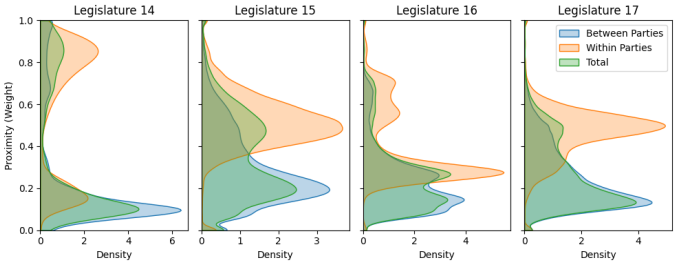

Fig. 1. Density Distribution of Voting Proximities in the National Assembly. Kernel density estimation plots illustrating the voting similarity between deputies across four legislatures. The vertical axis represents the proximity weight ranging from 0 (divergent votes) to 1 (identical votes). The distributions are split into three categories: **(Orange)** Within Parties, showing internal group cohesion; **(Blue)** Between Parties, showing alignment between different political groups; **(Green)** The total distribution of all voting relationships.

divided the assembly into distinct blocs. In contrast, the 14th and 15th legislatures featured a stable majority, which allowed for closer voting ties between different groups. The 17th legislature maintains high internal cohesion and clear group separation, though polarization has slightly decreased compared to the previous term. Finally, we must note that some deputies are listed as "Non-inscrits" because they lack the sufficient numbers to form a parliamentary group, a status that directly impacts their voting patterns and positioning in the analysis.

2. Graph creation and analysis

2.1. Graph creation. To construct the network, the graph was initialized by creating a unique node for each deputy. Voting records were then analyzed to quantify the agreement between pairs. Specifically, instances where two deputies simultaneously voted "For" were counted. To normalize this metric, the total number of votes where both deputies were present was tracked, regardless of their decision. The edge weight between two nodes was calculated as the ratio of shared "For" votes to the total number of common votes. Edges were added to the graph based on these weights, and the giant connected component was finally extracted to retain only the largest continuous network of deputies.

	14th Leg.	15th Leg.	16th Leg.	17th Leg.
Nodes	635	647	605	603
Edges	198,403	204,378	181,929	179,915
Avg. Degree	228.62	226.71	158.22	163.06
Density	0.986	0.978	0.996	0.991
Avg. Clustering Coef.	0.303	0.327	0.235	0.247

Table 2. Topological properties of the deputy networks across the four legislatures. The table presents the size of the graph (Nodes, Edges), connectivity metrics (Degree, Density), and community structure indicators (Clustering Coefficient).

As shown in Table 2, the networks are highly dense, meaning nearly every deputy has shared a "For" vote with every other colleague. In this analysis, the weighted degree represents a deputy's global level of agreement with the entire assembly. The data reveals a significant drop in the average degree and clustering coefficient over time, which indicates increasing political polarization. The weighted degree distribution in Figure 2 for the 14th and 15th legislatures is distinctly bimodal, reflecting an organization structured around two coherent blocks:

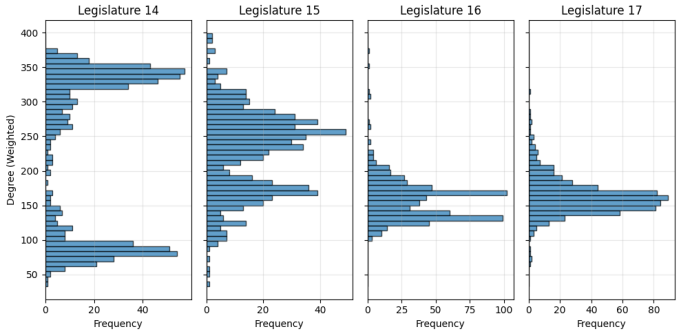

Fig. 2. Distribution of Weighted Degrees across Legislatures. Histograms displaying the frequency of voting agreement scores among deputies. The vertical axis represents the weighted degree (global consensus level), while the horizontal axis indicates the number of deputies holding that score. The visualization highlights a structural shift from a bimodal distribution (two distinct blocks) in the 14th legislature to a single, lower-value cluster in the 16th and 17th legislatures.

a disciplined majority and a homogeneous opposition. Starting with the 16th legislature, this bimodality weakens due to the emergence of a relative majority and fluctuating coalitions. This dynamic leads to an unimodal distribution in the 17th legislature, signaling increased fragmentation and the decline of traditional parliamentary blocks in favor of a dispersed system less oriented around two stable poles.

2.2. Communities. Once the graph is constructed, we use it to attempt to identify alliances and dynamics within the assembly. The aim is to construct communities in order to compare them with the alliances announced between parties during the same period, but also to determine if parties are solidly embedded within a single community or if certain parties are divided across several.

To achieve this, we employ a Louvain clustering approach on every legislature. Thus, we obtain the communities shown in Figure 3. This figure displays both the distribution of deputies, colored according to their political group, and the community to which they are attached.

Furthermore, Figure 4 illustrates the proportions in which the parties are distributed among the different communities across the legislatures. It is important to note that the same community number can correspond to very different groups between two legislatures.

	14th Leg.	15th Leg.	16th Leg.	17th Leg.
Within Community				
Com. 0	0.79	0.54	0.31	0.48
Com. 1	0.31	0.33	0.62	0.37
Com. 2	0.19		0.54	0.46
Between Communities				
Com. 0 - 1	0.29	0.20	0.13	0.17
Com. 0 - 2	0.13		0.19	0.14
Com. 1 - 2	0.08		0.29	0.26
Global Metric				
Louvain modularity	0.11	1.12	0.21	0.19

Table 3. Average percentage of votes shared within and between communities across legislatures, and global Louvain modularity score. Figures 3 and 4 show which parties represent these communities in each legislature. The digit used for one is not consistent.

We can observe that during the 14th legislature (**A** in Figure 3 & 4), under President François Hollande, two communities predominate in terms of membership numbers and generally represent the right-wing and left-wing parties. Far-right parties are very sparse, to the point that their members are too few to be recognized as a formal group in the assembly, and are therefore classified as Non-Inscrits (NI). However, another community exists containing part of the GDR party, considered the furthest to the left at the time. Although they are few in number, their voting behavior diverges enough from the rest for the clustering algorithm to identify them as a distinct political community, as it can be seen in Table 3.

The 15th legislature (**B** in Figure 3 & 4) was formed following elections held shortly after Emmanuel Macron took office as President. His party, LREM, came largely in the lead, forming a strong majority in the National Assembly. A remarkable phenomenon can then be observed: the majority and its close allies act as a single bloc (**Com.0** in Figure 3-**B**). This is particularly evident in the Table 3, which shows that more than half of the 'For' votes are shared within this community. In contrast, the only other community represents the entirety of the opposition forces (both right and left) whose votes prove to be much more dispersed (Table 3).

During the 16th legislature (**C** in Figure 3 & 4), the configuration changes. Firstly, we see an increase in the number of RN members of parliament, a phenomenon that will persist throughout the next legislature. They are numerous enough to form a community entirely on their own, supplemented by a few right-wing deputies and non-affiliated members (**Com.2** in Figure 3-**C**). Another interesting insight is that the historic right-wing party, LR, previously a member of the opposition, moves closer to the presidential majority's community (**Com.0** in Figure 3-**C**) as shown in Table 3. The left finds itself within its own community (**Com.1** in Figure 3-**C**), which includes a majority of deputies from the *La France Insoumise* group—one of the most marked to the left—whose number of seats has grown to the detriment of the socialists.

Finally, regarding the latest legislature (**D** in Figure 3 & 4), the general trends continue. RN members and a segment of the right are aligned. This is primarily due to a split on the right, where a faction formed the new party UDR and officially allied with the RN. Critically, even 40% of those who didn't join UDR still exhibit this community alignment (**Com.2** in Figure 3-**D**) based on their voting patterns. The central government bloc is still constituted by its historic allies as well as the remainder of the right (**Com.1** in Figure 3-**D**). It is actually closer to the far-right community than the left-wing one (Table 3 **1-2**). Finally, the left-wing parties remain in a configuration similar to the previous legislature (**Com.0** in Figure 3-**D**). In addition to this, whereas in previous legislatures a community would appear to have a majority in terms of the number of representatives, the 17th legislature offers a configuration without a significant lead for the president's group.

In conclusion, over the course of these last four legislatures, we have witnessed the collapse of the classic two-bloc model (left and right), notably due to the arrival of a centrist actor led by Emmanuel Macron. With a large majority, this initially placed the rest of the political spectrum in a weak opposition position. However, since the 15th legislature (**B** in Figure 3 & 4), the number of the most left-wing members (such as LFI) and especially

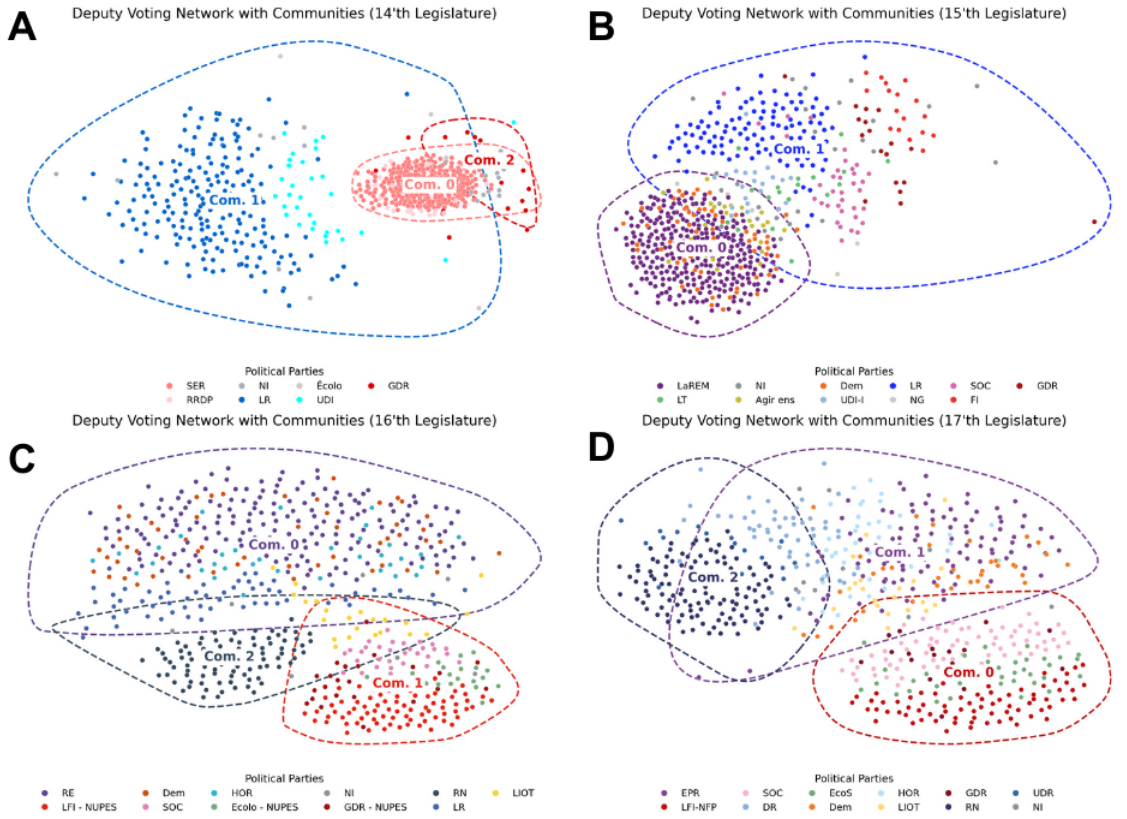

Fig. 3. Network Analysis of Political Alignments in the French National Assembly. Graph of deputies across four legislatures. Individual nodes represent deputies, colored according to their political party. The zones indicate communities detected using the Louvain method, taking the color of the dominant party within each cluster. The periods covered are: (**A**) 14th Legislature (June 2012 – June 2017); (**B**) 15th Legislature (June 2017 – June 2022); (**C**) 16th Legislature (June 2022 – June 2024); (**D**) 17th Legislature (July 2024 – Present).

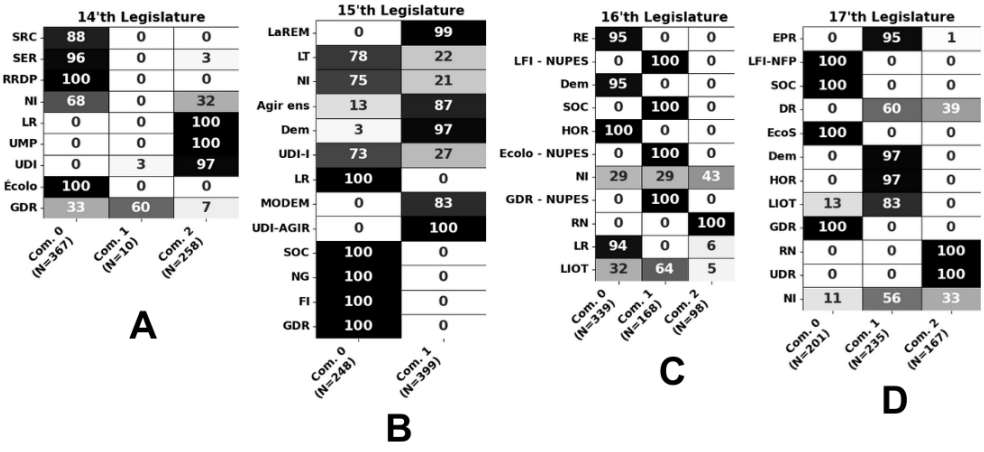

Fig. 4. Composition of Louvain Communities by Political Party. This matrix displays the distribution of party members across the detected communities. Each row represents a political party, and each column corresponds to a community identified by the Louvain method. The values indicate the percentage of each party's membership belonging to a specific community. The data covers four periods: **(A)** 14th Legislature (June 2012 – June 2017); **(B)** 15th Legislature (June 2017 – June 2022); **(C)** 16th Legislature (June 2022 – June 2024); **(D)** 17th Legislature (July 2024 – Present).

on the far right (RN and UDR) has increased significantly. The historic left has partly rallied its votes to LFI, while the right has split between the government party and the RN, transforming the latter from a party too small to have a group in the assembly into the leader of one of the three most important communities, with the most coordinated voting behavior. The traditional parties have therefore given way to the central bloc and the communities represented by more extreme parties, without giving a clear majority to any of the three blocs.

3. Written analysis over the last legislature

Explaination of the process. To complete, a written analysis of the last legislature is proposed using the previous calculated Louvain communities. This textual examination is made over every recorded parliamentary act of speaking

Fig. 5. Textual Analysis of Political Speeches in the French National Assembly. Word cloud representing the most relevant terms from the different parliamentary communities during the 17th Legislature (July 2024–Present). Word size reflects the relative frequency or importance of each term within its respective group.

and aims to observe the polarization of the speeches, and assess the fracture suggested between the different parties.

To achieve this, we propose a modified version of the term frequency–inverse document frequency measurement (1). This approach combines the original TF-IDF -which highlights words that are specific to a community- and a class-based variant, as proposed in the machine learning literature for text classification like with BERT model (2). The goal is to identify words that are common across an entire community while filtering out those used exclusively by individual members and words common across all communities. The calculation is presented in *Methods* and the communities are named by their dominant political orientation. With this method, singular words repeated redundantly by an individual are excluded to emphasize terms that are common to a community.

Presentation of the results. This analysis of three French political wordclouds presented in Figure 5 reveals distinct, and often adversarial, thematic priorities for each bloc.

The ***Left Alliance*** centers its rhetoric on strong social and economic critique, directly attacking perceived structural power and elites. Major terms include *Bolloré*, referring to Vincent Bolloré, a right-leaning billionaire who controls a vast media empire (including CNews and Europe 1), whose influence the left views as antidemocratic. Other large terms refers to the patriarchy or the monarchy directly framing the opposition.

The ***Center-Right Party*** features relatively few major words, indicating a less centralized, perhaps more fractured or moderate public discourse. This aligns with its position as a central, presidential party that attempts to navigate global cooperation, reflected in terms like *réussisse* (succeed) and *encourageants* (encouraging), balanced by *frustrant* (frustrating) and *inquiète* (worries).

The ***Far-Right Party*** is overwhelmingly preoccupied with themes of islam, identity, and security. The term *charia* (Islam Law) dominates, while the presence of *Lola* directly references a murder of a young girl, which became a national flashpoint for security and immigration discussions. The language is framed as an attack on the opposing extreme, explicitly naming the *ultagauche* (ultra-left) alongside culturally sensitive terms like *étouffantes* (suffocating).

Discussion

Thanks to this study, we have first shown, based on the votes of members of the French National Assembly, that political parties have been diverging for 10 years now into two distinct groups: a left-wing group and a second center-right group. This divergence is characterized by a decrease in votes shared between established communities (Table 3), while internal fragmentation is also apparent. Finally, we supported this divergence between parties by using speeches made by all members of parliament (Figure 5) during the current legislative term and showed that the specific concerns of political groups are diverging and becoming unique to their position.

This project combines a graph study and NLP and is unique compared to existing projects that only trace dialogues from an NLP perspective without studying common votes (3–5). However, the method of linking between representatives is open to debate. Indeed, the proportion of positive votes in common is not necessarily indicative of a link. One option would be to weight the differences bidirectionally or to count the overall similar votes.

Materials and Methods

The score for a term t in a community c is calculated as:

$$S(t,c) = tf(t,c) \cdot idf(t,C) \cdot cdf(t,c)$$

where $tf(t,c)$ is the term frequency of term t within community c, computed with double normalization K: $tf(t,c) = K + (1-K) \cdot \frac{f_{t,c}}{\max_{\{t' \in c\}} f_{t',c}}$ with $K = 0.25$.

The term $idf(t,C)$ is the inverse document frequency of term t across all communities C: $idf(t,C) = \log\left(\frac{N}{n_t+1}\right)$ where N is the total number of communities, and n_t is the number of communities containing term t.

Finally, the added term $cdf(t,c)$ is the community document frequency, representing how frequent term t is among members of community c: $cdf(t,c) = \frac{n_{t,c}}{N_c}$ where $n_{t,c}$ is the number of members in community c using term t, and N_c is the total number of members in community c.

1. Wikipedia, tf–idf — Wikipedia (https://en.wikipedia.org/wiki/Tfidf) (2025).
2. M Grootendorst, Class-based tf-idf (c-tf-idf) (https://maartengr.github.io/BERTopic/getting_started/ctfidf/ctfidf.html) (2025).
3. L Datan, Datan : Outil d'analyse des votes et comportements des députés (https://datan.fr) (2020).
4. C de Galembert, O Rozenberg, C Vigour, eds., *Faire parler le Parlement : Méthodes et enjeux de l'analyse des débats parlementaires pour les sciences sociales*, Droit et Société. (LGDJ (Librairie Générale de Droit et de Jurisprudence), Paris), (2014).
5. V Rennard, G Shang, D Grari, J Hunter, M Vazirgiannis, Fredsum: A dialogue summarization corpus for french political debates (https://arxiv.org/abs/2312.04843) (2023).