Alan Rosendo

Partidos brasileiros refletem a opinião dos senadores?

Brasil

2023

Alan Rosendo

Partidos brasileiros refletem a opinião dos senadores?

Dissertação apresentada ao Curso de Mestrado Acadêmico em Economia, Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Economia

Universidade de Brasília - UnB
Faculdade de Administração Contabilidade e Economia - FACE
Departamento de Economia - ECO
Programa de Pós-Graduação

Orientador: Prof. Daniel Oliveira Cajueiro, Dr.

Brasil
2023

Alan Rosendo
Partidos brasileiros refletem a opinião dos senadores?/ Alan Rosendo. – Brasil, 2023-
72p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Daniel Oliveira Cajueiro, Dr.

Dissertação (Mestrado) – Universidade de Brasília - UnB
Faculdade de Administração Contabilidade e Economia - FACE
Departamento de Economia - ECO
Programa de Pós-Graduação, 2023.

1. Party cohesion. 2. Text clustering. 3. Machine learning. II. Universidade de Brasília. III. Faculdade de Administração, Contabilidade e Economia - FACE. IV. Departamento de Economia IV. Partidos brasileiros refletem a opinião dos senadores?

Alan Rosendo

Partidos brasileiros refletem a opinião dos senadores?

Dissertação apresentada ao Curso de Mestrado Acadêmico em Economia, Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Economia

Trabalho aprovado. Brasil, 03 de março de 2023:

Prof. Daniel Oliveira Cajueiro, Dr.
Orientador

Bernardo Mueller
Convidado 1

Maisa Melo
Convidado 2

Brasil
2023

Resumo

Este trabalho tem como objetivo verificar se senadores de um mesmo partido possuem pontos de vista ideológicos semelhantes. Medimos os posicionamentos ideológicos dos senadores usando seus discursos e técnicas de processamento de linguagem natural. Construímos nosso método com base em três etapas. Primeiro, limpamos os discursos usando técnicas de pré-processamento de linguagem natural. Posteriormente, dividimos os senadores em grupos de acordo com a semelhança de seus discursos. Por último, comparamos a composição dos clusters formados endogenamente com a composição dos partidos. Nosso conjunto de dados de discursos vai da 51ª à 55ª legislatura do Senado brasileiro. Descobrimos que senadores do mesmo partido tendem a ter discursos semelhantes. Isso também ocorre entre partidos de mesma ideologia. Além disso, caracterizamos cada cluster com suas palavras mais relevantes. Este tipo de caracterização permite identificar a posição da parte como *esquerda*, *centro* ou *direita*.

Palavras-chave: Clusters, Senadores, Partidos, Ideologia, Processamento de Linguagem Natural, Discursos, Senado Federal, Coesão Partidária.

Abstract

This work aims to verify whether senators from the same party have similar ideological points of view. We measure the senators' ideological points using their speechs and natural language processing techniques. We build our method based on three steps. First, we clean the speechs using nature language pre-processing techniques. Second, we split the senators into clusters according to the similarity of their speeches. Third, we compare the composition of the endogenously formed clusters with the composition of the parties. Our dataset of speechs come from the 51th to the 55th Brazilian senate legislature. We find that senators from the same party tend to have similar speeches. This also occurs between parties of the same ideology. Furthermore, we characterize each cluster with its most relevant words. This kind of characterization allows the identification of the position of the party as *left*, *center* or *right*.

Keywords: Clusters, Senators, Parties, Ideology, Natural Language Processing, Speeches, Federal Senate, Party cohesion.

List of Figures

List of Tables

List of abbreviations and acronyms

ABCP	Associação Brasileira de Ciência Política
AI	Artificial Intelligence
CADE	Conselho Administrativo de Defesa Econômica
CBF	Confederação Brasileira de Futebol
CPI	Comissão Parlamentar de Inquérito
CPMF	Contribuição Provisória sobre Movimentação Financeira
DC	Democracia Cristã
DEM	Democratas
FMI	Fundo Monetário Internacional
FUNDEF	Fundo de Manutenção e Desenvolvimento do Ensino Fundamental
GDP	Gross Domestic Product
ICMS	Imposto sobre Circulação de Mercadorias e Serviços
INCRA	Instituto Nacional de Colonização e Reforma Agrária
MDB	Movimento Democrático Brasileiro
NLP	Natural Language Processing
PAC	Programa de Aceleração do Crescimento
PCdoB	Partido Comunista do Brasil
PDT	Partido Democrático Trabalhista
PEC	Proposta de Emenda Constitucional
PFL	Partido da Frente Liberal
PIB	Produto Interno Bruto
PL	Partido Liberal
PMB	Partido da Mulher Brasileira

PMDB	Partido do Movimento Democrático Brasileiro
PMN	Partido da Mobilização Nacional
PMR	Partido Municipalista Renovador
PP	Partido Progressistas
PPB	Partido Progressista Brasileiro
PPL	Partido Pátria Livre
PPS	Partido Popular Socialista
PR	Partido da República
PRB	Partido Republicano Brasileiro
PROS	Partido Republicano da Ordem Social
PRTB	Partido Renovador Trabalhista Brasileiro
PSB	Partido Socialista Brasileiro
PSC	Partido Social Cristão
PSD	Partido Social Democrático
PSDB	Partido da Social Democracia Brasileira
PSL	Partido Social Liberal
PSOL	Partido Socialismo e Liberdade
PT	Partido dos Trabalhadores
PTC	Partido Trabalhista Cristão
PTRB	Partido Renovador Trabalhista Brasileiro
PV	Partido Verde
SUDENE	Superintendência do Desenvolvimento do Nordeste

Contents

1 Introduction

The Brazilian electoral rules, with proportional legislative votes and open lists, generated a single party system, extremely fragmented, with many parties and little identification of the electorate's party. Brazil currently has 31 political parties [1]. Several analysts argue that Brazilian parties are not one of the main dimensions through which the political system should be analyzed and understood (Figueredo et al., 2022). The Brazilian framework is one of low permanent programmatic (Bolognesi et al., 2022).

In this work we propose an unsupervised machine learning approach to investigate whether we can associate the division of politicians into parties with the ideology of their members. Our analysis is based on a public dataset that includes all speeches given by senators from 1999 to 2018. Thus, our measure of ideology is the information content available in the speeches of the senators. After cleaning and organizing the dataset, we represent the speeches of the senators using a nature language model and we use this representation to divide the senators in clusters. In the end, we compare the data-driven formed clusters with the recent party classification proposed by Bolognesi et al. (2022). We find that senators from the same party tend to have similar speeches and senators from parties of the same ideological spectrum also have similar speeches. Thus, according to our results, the division of politicians into parties is related to the ideology of their members. Furthermore, we also provide a measure of party cohesion that is completely data-driven. Our measure suggests that the parties that act in the Brazilian Senate are highly cohesive.

Our work relates to the study of the cohesion of political parties. In particular, our work specifically connects with the study of ideological cohesion that we may understand as a general agreement among the members of a party about certain ideological standpoints. A relevant part of the literature analyzes roll call voting behavior in order to measure party cohesion (Poole; Rosenthal, 2001; Poole, 2005; Hix et al., 2007). However, Jahn and Oberst (2012) calls the attention that roll call analysis is probably a better indicator for party discipline than party cohesion because this analysis generates timevariant results for individual parties in a systematic manner. This conclusion is only valid provided that legislative voting is open and not secret. An interesting tour of force is to analyze ideological cohesion using textual data from intraparties debates (Giannetti; Laver, 2008). Although this technique is able to measure party cohesion individually for each party, it is not able to measure party cohesion in a comparative perspective. Furthermore, party disagreement is stronger at intra-party congresses than in the parliament (Jahn; Oberst, 2012). Different from these works, our work neither uses roll call voting data nor uses intraparty data.

[1] Available at: <https://www.tse.jus.br/partidos/partidos-registrados-no-tse/registrados-no-tse>

It is worth mentioning that when studying political parties, it is common in the literature to distinguish between party cohesion and party discipline (Hazan, 2006). In the definition of classical literature, cohesion is "the extent to which, in a given situation, group members can be observed to work together for the group's goal in one and the same way" and discipline “refers either to a special type of cohesion achieved by enforcing obedience or to a system of sanctions by which such enforced cohesion is attained" (Özbudun, 1970).

Our work also depends on the ideological party classification proposed by Bolognesi et al. (2022) that is based on the experts' answers of questionnaires. It is worth mentioning that there are other ways to classify parties based on their political spectrum such as using the party program (Franzmann; Kaiser, 2006; Tarouco; Madeira, 2013) and the behavior of parliamentarians (Scheeffer et al., 2016). Furthermore, the literature also classifies political parties according to the objective they pursue into three types: vote-seeking, office-seeking and policy-seeking (Wolinetz, 2002). Vote-seeking parties are the ones that seek to maximize the votes received. They have political positions and flexible alliances that are changed with the objective of maximizing votes. Party organization intensifies during the election period and reduces in the periods between elections. Parties are “teams of men" seeking to maximize their electoral support for the purpose of controlling government (Strom, 1990). Office-seeking parties maximize participation in government and political alliances. Party activities are restricted to election periods. These parties seek to maximize, not their votes, but their control over political office (Strom, 1990). Policy-seeking parties prioritize political issues, occupy the state to achieve the implementation of their issues. They are always mobilized, even in non-election periods, and seek to convince voters of their issues.

In Brazil, there is an intense debate about the discipline, cohesion and, in general, the role of the parties. Neiva (2011b) calls our attention to the fact that there are two clear strands in the literature. The first strand that he call “pessimists" defends that the Brazilian party system is almost chaotic and formed by fragile, ideological inconsistent and nondisciplinary parties (Mainwaring; Liñán, 1997). On the other hand, the other strand that he calls “optimists" asserts that the Brazilian party system is disciplined consistent and predictable (Neiva, 2011a; Santos, 2002). Our work provides empirical evidence in favor of the optimists.

We organize our work as follows. Section 3 describes the procedures we adopt to tune and estimate the models. We detail the data set we use in Section 2 and present the results in Section 4. Section 5 summarizes and concludes the work.

2 Data

Our dataset contains 76734 senators' speeches given in the Brazilian Senate, from 1999 to 2018. In order to build our dataset, we use public xml data available at the webpage of the Federal Senate of Brazil. [1] Besides the speeches, our data also includes the date of the speech, the name of the senator who gave the speech, and the party of the senator at the date of the speech. Table 1 summarizes our data.

Since the objective of our work is to study the ideological cohesion of parties, we consider a senator who has belonged to different parties as different senators. This makes sense because the action of party switching may suggest that either the ideology of the senator may have changed or the ideology of the party may have changed. Furthermore, we are using the speeches of senators as proxies for the parties ideologies. So, the important unity here is the party and not the senator. Table 2 shows that the number of partisan shifts per legislature is not negligible.

We compare the ideological content of the parties per legislature. There are two important reasons for that: First, there are different parties in different legislatures; Second, we can consider the effect of the executive in the choices of the parties. Therefore, for each legislature, we identify each senator by the tuple [*senator*, *party*]. Thus, if a senator X leaves party A and registers as party B, we assume that senator $[X, A]$ is different from $[X, B]$.

Table 1 – Summary statistics

Legislature	Period of the legislature [2]	N° speeches	N° parties
51st	1999-2003	7433	12
52nd	2003-2007	18314	15
53rd	2007-2011	17122	16
54th	2011-2015	17346	21
55th	2015-2019	12147	30

We use senators' speeches and not deputies' speeches for the following reasons: each senator speaks more often and they have more political experience.

[1] We may find the entire data in <https://www12.senado.leg.br/dados-abertos/conjuntos?portal=Legislativo&grupo=plenario>.
[2] The last year of legislature not has speechs.

Table 2 – Partisan shifts

Legislature	N° shifts
51st	27
52nd	25
53rd	36
54th	21
55th	40

3 Methods

In this section, we present the methods we use to investigate whether there is a high connection between senators' ideology and their parties. As a measure of senators' ideology we use the content of their speeches. We then use a data-driven approach to group those senators who have similar speeches into groups and then compare these groups to the parties they actually belong to. We may summarize our approach in four steps: The first step is the data pre-processing step where we clean up the speeches and associate them with each senator. In the second step, we represent the speeches of the senators using a natural language model. In the third step, we use an unsupervised machine learning algorithm to divide senators into "homogeneous" groups according to the ideology of their speeches. In the last step, we compare these groups with the actual parties the senators belong and political inclination of the senators.

In Section 3.1, we show how we organize and clean the data. In Section 3.2 we present the vector space model that is the language model we use to represent the senators' speeches. In section 3, we present the k-means algorithm, which is the algorithm we use to group senators with similar ideologies based on the content of their speeches. Finally, in Section 3.4, we review the method we use as reference to identify the ideology of the senators' political parties.

3.1 Data pre-processing

For each legislature, we create a document for each tuple [*senator*, *party*]. In each document, we include all the speeches of that senator delivered while he belonged to that party. To make it simple, from now on, we assume that always we refer to the senators' data, we are referring to the tuple [*senator*, *party*] data.

Since we use these documents of speeches as proxy for the ideology of the senators, we clean these documents in order to keep only relevant information. Thus, we remove from each document the following tokens:

1. Verbs[1];
2. Words that do not provide information about the content of the speeches:
    a) Names of all senators;
    b) Party abbreviations;

[1] We use the Spacy library part of speech tag to identify the verbs. The documentation is available at: <https://spacy.io/api/doc>

c) State names and abbreviations;

d) Stopwords[2];

e) List of frequent words in speech with little relation to the topic. We may find the complete list of words in A. An example of a word that belongs to this list is “judgment"[3]

3.2 Vector space model

In order to represent the document of speeches associated with each senator, we use the vector space model.

To make a precise description of the vector space language model, we borrow some ideas from the field of natural language processing. Let the term w_i be a word or group of consecutive words identified by the unique index i. N_v is the number of distinct terms. d_i is the set of all compiled speeches of senator i.

The vocabulary $\mathcal{V} = \{w_1, \ldots, w_i, \ldots, w_{N_V}\}$ is the set of all distinct terms (present in all documents) and $I_V = \{1, \ldots, N_V\}$ is the set of all term indexes. The documents $d_j = [w_{i_1}, \ldots, w_{i_k}, \ldots, w_{i_{L_j}}]$ consist of a list of L_j non-unique consecutive terms ($1 \leq k \leq L_j$ and $i_k \in I_V$), while $\mathcal{V}^{d_j}$ is the vocabulary that appears in the document d_j. Finally, $\mathcal{D} = \{d_1, \ldots, d_j, \ldots, d_{N_S}\}$ is the set of all documents that contain the senators' speeches.

The term-document matrix $\mathbf{M}$ is a $N_V \times N_S$ matrix that establishes a relation between a term and a document with senators' speeches. In a document-term matrix, rows correspond to terms and columns correspond to the documents $d_k \in \mathcal{D}$:

$$\begin{array}{c} \\ w_1 \\ w_2 \\ \vdots \\ w_{N_V} \end{array} \begin{array}{c} \begin{array}{cccc} d_1 & d_2 & & d_{N_S} \end{array} \\ \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1N_C} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2N_C} \\ \vdots & \vdots & \cdots & \vdots \\ \omega_{N_V1} & \omega_{N_V2} & \cdots & \omega_{N_VN_S} \end{bmatrix} \end{array} . \quad , \tag{3.1}$$

We build the weight $\omega_{i,j}$ considering one factor related to the term frequency, the other related to the document frequency, and the last one related to a normalization:

$$\widetilde{\omega}_{i,j} = \begin{cases} f_{\text{tf}}(\text{tf}_{i,j}) \times f_{\text{idf}}(\text{df}_i) & \text{if } \text{tf}_{i,j} > 0 \\ 0 & \text{if } \text{tf}_{i,j} = 0 \end{cases} \tag{3.2}$$

$$\omega_{i,j} = \frac{\widetilde{\omega}_{i,j}}{\text{norm}_j} \tag{3.3}$$

[2] We use the list of stopwords provided by the spacy and nltk libraries.
[3] From the Portuguese word “Acórdão".

where $f_{\text{tf}}(\text{tf}_{i,j})$ is the weight associated with the frequency $tf_{i,j}$ of term i in document j (i.e., the number of times a term i arises in document j), $f_{\text{idf}}(\text{df}_i)$ is the weight associated with the document frequency df_i of term i (i.e., the number of documents that term i arises), and norm_j is a document length normalization factor to compensate undesired effects of long documents. There are several possible choices for these weights and we may find a collection of them in Baeza-Yates and Ribeiro-Neto (2008) and Manning et al. (2008). The intuition behind this model is that a term is important in a document if it arises several times in this document (i.e., high token frequency) and it does not arises many times in all other documents (i.e. low document frequency).

In our work, we use these weights as provided by the Python Scikit-Learn library (Pedregosa et al., 2011) [4]. This library computes the weight $f_{\text{tf}}(\text{tf}_{i,j})$ for term i and document j as

$$f_{\text{tf}}(\text{tf}_{i,j}) = \text{tf}_{i,j},$$

the weight $f_{\text{idf}}(\text{df}_i)$ for term i as

$$f_{\text{idf}}(\text{df}_i) = \log \frac{1 + N_S}{1 + \text{df}_i} + 1,$$

and the normalization factor for document j as

$$\text{norm}_j(\tilde{\omega}) = \sqrt{\sum_{i}^{N_V} \tilde{\omega}_{i,j}^2}.$$

Thus, after applying the vector space model to the collection of documents $\mathcal{D}$, we are able to represent each document $d_j \in \mathcal{D}$ by a vector, where each coordinate of this vector is a weight that gives the importance of a term i in document j. We may call this vector as tf-idf vector.

3.3 Cluster

In order to split the senators in clusters based on the content of their speeches, we use the K-means model because this is the bench-marking for similar cases.

K-means clustering is a method that aims at partitioning N_S observations into $K\,(\leq N_S)$ clusters, assuming each observation belongs to the cluster with the nearest mean (MacQueen, 1967; Lloyd, 1982; Gnanadesikan, 2011). In order to build the set of clusters C, we have to run the algorithm due to Lloyds presented in Figure 1.

[4] The documentation is available at <https://scikit-learn.org/stable/modules/feature_extraction.html#text-feature-extraction>

procedure *Lloyds*
Choose K points to the initial clusters
while *C changes* ***do***
for all $x \in X$ ***do***
Find cluster C_k with center c_k that is the closest to x (using the distance)
Add x to C_k
end for
for all Cluster k **do**
Recalculate c_k as the average of all the members of C_k
end for
end while
end procedure

Figure 1 – The Lloyds algorithm.

procedure $K - means + +$
Pick a center evenly and randomly from among the data points.
For number of centers $< K$ **do**
For each point x on the data, calculate $D(x)$, the distance between x and the nearest center that has already been chosen.
Choose a new point at random as a new center, using a weighted probability distribution where a point x is chosen with probability proportional to $D(x)^2$.
end for
end procedure

Figure 2 – The K-means++ initialization method.

In order to choose the initial centroids, we use k-means++ initialization. The initialization method works as follows:

Our computational implementation of the K-means algorithm uses the sklearn library [5]. We use k-means++ initialization presented in Figure 2 [6].

Note that the K-means algorithm depends on the definition of the number of clusters. In order to define the number of clusters K, we use the Calinski-Harabasz index. The Calinski-Harabasz (CH) index evaluates the degree of dispersion between clusters comparing inter-cluster with intra-cluster distances (Wang; Xu, 2019):

$$CH(K) = \frac{B(K)(N-K)}{W(K)(K-1)},$$

where $B(K) = \sum_{k=1}^{K} n_k||c_k - \bar{x}||$, $W(K) = \sum_{k=1}^{K}\sum_{x_j \in C_k} ||x_j - c_k||$, K is the number of clusters, n_k is the number of points in cluster k, c_k is the center of cluster k, $\bar{x}$ is the average of all points, $B(K)$ is the intra-cluster distance, $W(K)$ is the intercluster distance,

[5] The documentation is available at <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>.

[6] In particular, we use the "greedy K-means++". It differs from the vanilla K-means++ by making several trials at each sampling step and choosing the best centroid among them (Pedregosa et al., 2011).

and N is the number of samples. The higher the value of the Calinski-Harabasz index, the better is the division of the data provided by the clustering algorithm.

In our work, we apply the K-means clustering method to the documents with senators speeches represented by their tf-idf vectors.

In order to find the best cluster representation of our set of documents with senators' sppeches, we run a grid of simulations varying parameters associated with the generation of the tf-idf vectors and the number K of clusters. For each run, we evaluate the Calinski-Harabasz index. For each legislation, we choose the combination of parameters with the highest value of Calinski-Harabasz index. We consider the following parameters:

1. Freq plan: Lowest frequency that a term appears in all the speeches of the same senator;
2. Min doc: Lowest frequency that a term appears in speeches by different senators;
3. Max doc: Highest percentage of speeches a term appears;
4. Number of clusters: The amount of clusters;

Table 3 summarizes the values we use to run these simulations.

Table 3 – Parameters of computer simulations

Variable	Values
Freq plan	3 - 5 - 10 - 15 - 20
Min doc	2 - 3 - 4
Max doc	0.4 - 0.5 - 0.6 - 0.7
Numbers of clusters	2 - 3 - 4 - 5 - 6 - 7 - 8 - 9

We may use the percentage of senators of a given party that arise in the cluster with the largest concentration of them per legislature as a measure of party cohesion. As this type of analysis is not usual, we defined the grid of values of the computational simulations.

3.4 Party ideological classification

In our work, we use the work of Bolognesi et al. (2022) to classify the ideology behind the Brazilian political parties. Bolognesi et al. (2022) build this classification based on a survey with political scientists. They invited political scientists from the ABCP to respond, through a web-based platform, about how they classify the then thirty-five[7] Brazilian political parties on the left-right axis on a spatial scale from zero to ten. They estimate the political spectrum of the parties according to the Table 4.

[7] The number of the policital parties is variable in time

Table 4 – Party ideological classification scale. (Bolognesi et al., 2022)

Ideology	Values
Extreme left	0 - 1.50
Left	1.51 - 3.00
Center left	3.01 - 4.49
Center	4.50 - 5.50
Center right	5.51 - 7.00
Right	7.01 - 8.50
Extreme right	8.51 - 10.00

In the period studied, while some parties changed their name, others merged. We always use the party classification resulting from these modifications. We consider the following party modifications according to Table 5:

Table 5 – Party modifications

Old name	New name
PPS	Cidadania
PFL	DEM
PR	PL
PMDB	MDB
PMR and PRB	Republicanos
Pode	Podemos
PP and PPB	Progressistas

We use the party ideological classification provided by Bolognesi et al. (2022) and presented in Table 6:

Table 6 – Party ideological classification. (Bolognesi et al., 2022)

Party	Average	Ideology
PSTU	0.51	Extreme left
PCO	0.61	Extreme left
PCB	0.91	Extreme left
PSOL	1.28	Extreme left
PCdoB	1.92	Left
PT	2.97	Left
PDT	3.92	Center left
PSB	4.05	Center left
Rede	4.77	Center
PPS	4.92	Center
PV	5.29	Center
PTB	6.10	Center right
Avante	6.32	Center right
SDD	6.50	Center right
PMN	6.88	Center right
PMB	6.90	Center right
PHS	6.96	Center right
MDB	7.01	Right
PSD	7.09	Right
PSDB	7.11	Right
Podemos	7.24	Right
PPL	7.27	Right
PRTB	7.45	Right
Pros	7.47	Right
PRP	7.59	Right
PRB	7.78	Right
PR	7.78	Right
PTC	7.86	Right
DC	8.11	Right
PSL	8.11	Right
Novo	8.13	Right
Progressistas	8.20	Right
PSC	8.33	Right
Patriota	8.55	Extreme right
DEM	8.57	Extreme right

To be able to compare the political spectrum of the clusters formed by the senators' speeches. In our work, we split the parties into only three groups, according to Table 7.

Table 7 – Unified ideological positioning.

Ideology	Unified group
Extreme left	Left
Left	Left
Center left	Center
Center	Center
Center right	Center
Right	Right
Extreme right	Right

4 Results

Based on Table 3, we run the grid of simulations and obtain the value of the Calinski-Harabasz index for each set of parameters. For each legislature, Table 8 shows only the set of parameters associated with the highest Calinski-Harabasz indexes. Using this set of parameters, we create the TF-IDF vectors associated with each senator in each legislature and build the clusters of senators. To make easy the understanding of the content of the clusters, for each legislature, we create a word cloud for each cluster. For each legislature, our main results are the percentage of senators of each party that belongs to each cluster and the ideological composition of each cluster.

Table 8 – Parameters in each legislature

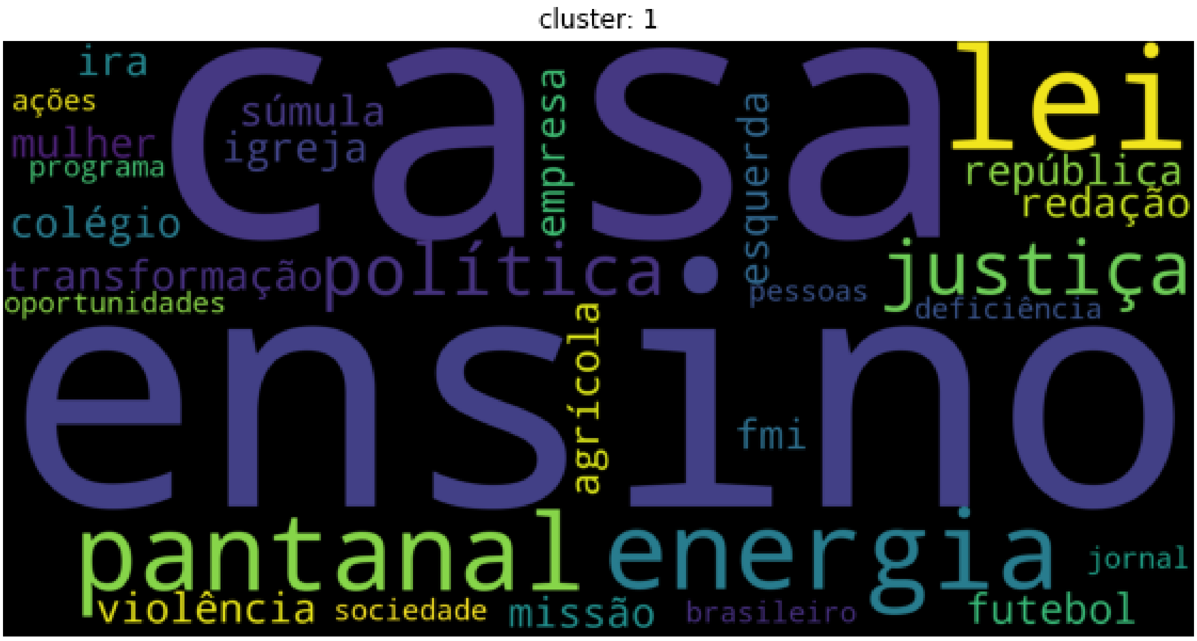
Leg	Freq plan	Min doc	Max doc	N° of clusters	Index of Calinski-Harabasz
51	20	2	0.7	3	5.921
52	10	3	0.5	2	7.594
53	5	2	0.6	2	7.893
54	15	3	0.6	3	6.609
55	3	4	0.6	2	7.964

4.1 51st legislature

According to the Calinski-Harabasz index, there are three clusters in the 51st legislature. Figures 3, 4 and 5 present the most representative words of each cluster. While Table 9 presents the absolute presence and percentage of senators from each party according to the clusters. Table 10 shows the ideological classification of parties according to Tables 6 and 7, with % Ideo is the percentage of senator of this ideology in this cluster and %Clus is the percentage of this senators in this cluster. Tables 20, 21 and 22 in Appendix B present the most important words and their translations in respectively Clusters 0, 1 and 2 of the 51st legislature.

Figure 3 – Word cloud, cluster 0, 51st legislature

Figure 4 – Word cloud, cluster 1, 51st legislature



In cluster 0, according to Figure 3, the most representative words are related to parliamentary activity in a generic way, with no direct relation to a theme of a specific ideology, such as: house (casa) it is common for senators to refer to the Senate as "this

[1] Percentage of senators parties that are in this cluster
[2] Percentage of senators of this ideology in this cluster
[3] Percentage of senators in cluster of this ideology

Figure 5 – Word cloud, cluster 2, 51st legislature

Table 9 – Division of parties between clusters in the 51st legislature

Party	Cluster label	Senators	Percentage [1]
PDT	0	6	100.00
PL	0	1	100.00
PSB	0	4	100.00
PT	0	9	100.00
PV	0	1	100.00
PPS	0	3	75.00
PSDB	0	14	60.87
PFL	0	17	60.71
PTB	0	3	60.00
PMDB	0	21	51.22
PPB	0	3	42.86
No party	0	2	20.00
No party	1	8	80.00
PPB	1	3	42.86
PTB	1	2	40.00
PMDB	1	15	36.59
PFL	1	9	32.14
PSDB	1	6	26.09
PPS	2	1	25.00
PPB	2	1	14.29
PSDB	2	3	13.04
PMDB	2	5	12.20
PFL	2	2	7.14

Table 10 – Ideological composition of parties in each cluster, in the 51st legislature

Cluster	0			1			2		
Ideology	Sen	% Ideo [2]	% Clus [3]	Sen	% Ideo	% Clus	Sen	% Ideo	% Clus
Left	9	100.00	11.54	0	0.00	0.00	0	0.00	0.00
Center	13	86.67	16.67	2	13.33	5.71	0	0.00	00.00
Right	56	62.92	71.79	33	33.00	94.29	11	11.00	100.00

house"; Brazilian (brasileira); policy (política). With less weight than these, there are many words related to terms frequently used by the left: Amazon (Amazônia); public (publico); work (trabalho); justice (justiça); education (edufcação). At the time, it is worth mentioning that we had a right-wing government[4] that had a broad parliamentary base, which is why left parties tended to act more in the center.

Most parliamentarians are concentrated in cluster 0, including all leftist parties. This indicates a unity in relation to left parties and a difference in relation to center and right parties. This indicates a difference in the discourse of the parties, according to the ideology in the 51st legislature.

In the case of cluster 1, as shown in 4, some terms more related to the actions of right-wing parties stand out, such as: IMF (FMI) (International Monetary Fund); church (igreja); opportunities (oportunidades); company (empresa); violence (violência); transformation (transformação); agricultural (agrícola). The presence of the term energy (energia) may be related to the energy crisis that occurred in Brazil in the period, including a period of electricity rationing (Bardelin, 2004). The presence of the term football (futebol) must be related to the CPI created in the period to investigate the CBF (Azevedo; Rebelo, 2001).

In cluster 2 (Figure 5), there are only a few parliamentarians, the vast majority on the right. The presence of several terms related to the performance of the government stands out, such as: FUNDEF [5]; resources (recursos); North East (nordeste); roads (estradas); development (desenvolvimento); agriculture (agricultura); Sudene; corruption (corrupção). It is noteworthy that the word with the greatest weight in this group is Indian (índio), which is generally not related to the activities of right parliamentarians. However, during this period there was the trial of the murderers of the Galdino Indian (Piubelli, 2012), a crime that had great national repercussion and must have influenced the speech of the senators.

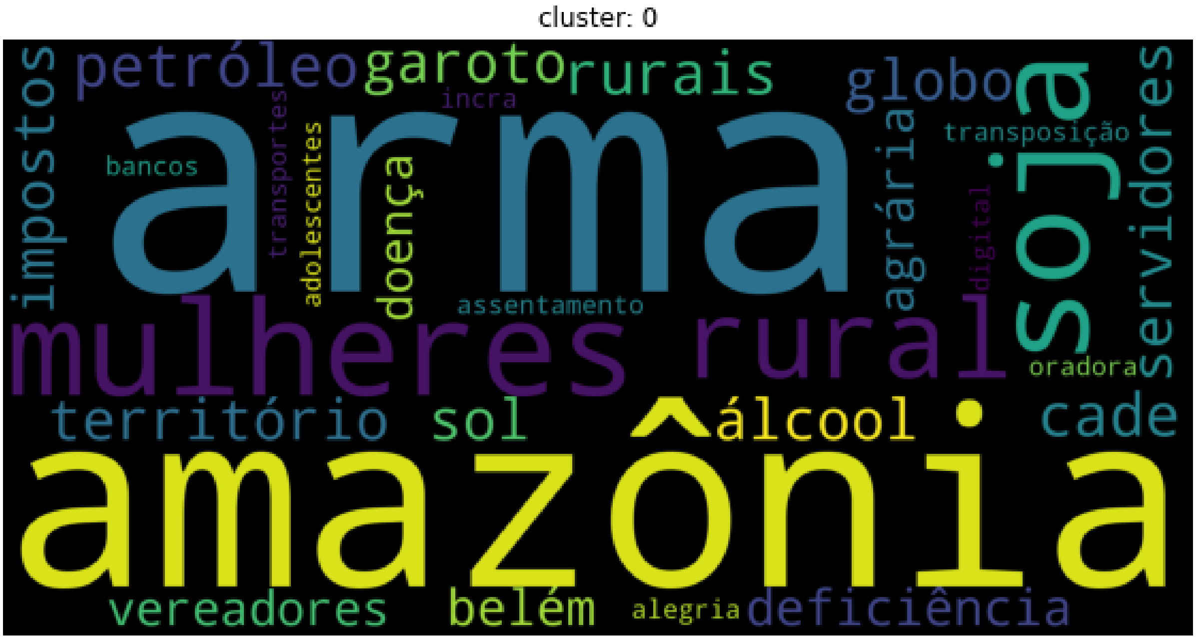
[4] It is worth mentioning that we are using here the (Bolognesi et al., 2022)'s classification that considers the PSDB a right wing party. Although this government run a liberal economy, there is no indication this government implemented any kind of conservative policy. Furthermore, many social policies that started in this government were amplified in the subsequent left-wing governments.

[5] Implemented by Law No. 9424/1996 and available at: <http://www.planalto.gov.br/ccivil_03/leis/L9424compiled.htm>

4.2 52nd legislature

According to the Calinski-Harabasz index, there are only two clusters in the 52nd legislature. Figures 6 and 7 present the most representative words of each cluster. While Table 11 presents the absolute presence and percentage of senators from each party according to the clusters, Table 12 shows the ideological classification of parties according to Tables 6 and 7. Tables 23 and 24 in Appendix B present the most important words and their translations in respectively Clusters 0 and 1 of the 52nd legislature.

Figure 6 – Word cloud, cluster 0, 52nd legislature

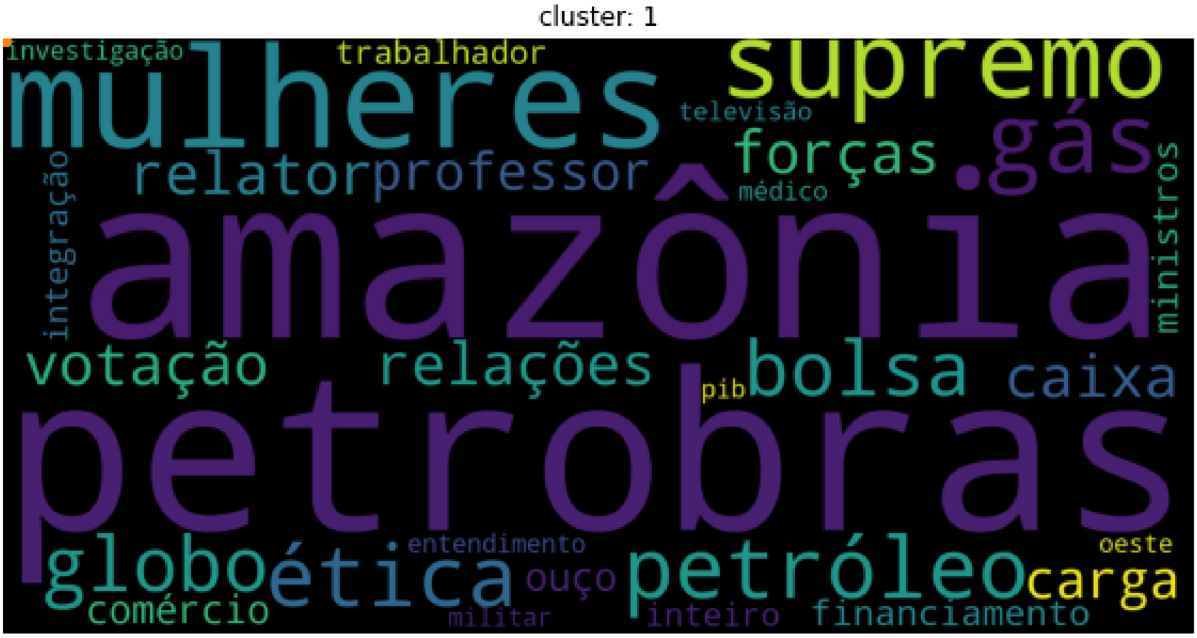


In the 52nd legislature, the president of Brazil is Lula, from the Partido dos Trabalhadores (Workers' Party), a left-wing party. The left-wing parties are concentrated almost entirely in cluster 1, the centre parties are equally divided between the clusters, while the right-wing parties are a little more concentrated in cluster 0. In this legislature, we observed a great distinction in the speeches between right and left parties.

In cluster 0, according to Figure 6, the word gun (arma) is the most prominent in this group, it is closely related to the right-wing agenda and in the period there was approval of the disarmament statute [9]. The following terms are also highlighted: transport (transportes), rural (rurais), transposition (transposição), agrarian (agrária), territory

[6] Percentage of senators parties that are in this cluster
[7] Percentage of senators of this ideology in this cluster
[8] Percentage of senators in cluster of this ideology
[9] Law No. 10826/2003, available at <https://www.planalto.gov.br/ccivil_03/leis/2003/L10.826compiled.htm>

Figure 7 – Word cloud, cluster 1, 52nd legislature



(território) and taxes (impostos), which are themes usually present in the agendas of right parties.

In cluster 1, as shown in Figure 7, we highlight the words Petrobras and Caixa, which are the largest state-owned companies that received the largest state investments in the period. In addition to the words that are historically related to leftist parties: worker (trabalhador) and teacher (professor).

Both clusters have the word amazônia (Amazon) highlighted. During this period, there was a major change in Brazilian environmental policy, with the appointment of Minister Marina Silva (Oliveira, 2016). There was also the murder of the American Dorothy Stang due to land disputes (Lisboa; Branco, 2022). This highlighted the environmental issue.

Table 11 – Division of parties between clusters in the 52nd legislature

Party	Cluster	Senators	Percentage [6]
PC DO B	0	6	100.00
PMR	0	1	100.00
PP	0	1	100.00
PRB	0	2	100.00
No party	0	5	100.00
PTB	0	5	71.43
PPS	0	2	66.67
PSB	0	3	60.00
PMDB	0	23	58.97
PSOL	0	1	50.00
PSDB	0	10	47.62
PL	0	2	40.00
PFL	0	8	38.10
PT	0	3	20.00
PDT	0	1	16.67
PDT	1	5	83.33
PT	1	12	80.00
PFL	1	13	61.90
PL	1	3	60.00
PSDB	1	11	52.38
PSOL	1	1	50.00
PMDB	1	16	41.03
PSB	1	2	40.00
PPS	1	1	33.33
PTB	1	2	28.57

Table 12 – Ideological composition of parties in each cluster, in the 52nd legislature

Cluster	0			1		
Ideology	Sen	% Ideo [7]	% Clus [8]	Sen	% Ideo	% Clus
Left	5	27.78	8.20	13	72.22	20.00
Center	9	50.00	14.75	9	50.00	13.85
Right	47	52.22	77.05	43	47.78	66.15

4.3 53rd legislature

According to the Calinski-Harabasz index, there are also only two clusters in the 53rd legislature. Figures 8 and 9 present the most representative words of each cluster. While Table 13 presents the absolute presence and percentage of senators from each party according to the clusters, Table 14 shows the ideological classification of parties according to Tables 6 and 7. Tables 25 and 26 in Appendix B present the most important words and their translations in respectively Clusters 0 and 1 of the 53rd legislature.

Figure 8 – Word cloud, cluster 0, 53rd legislature

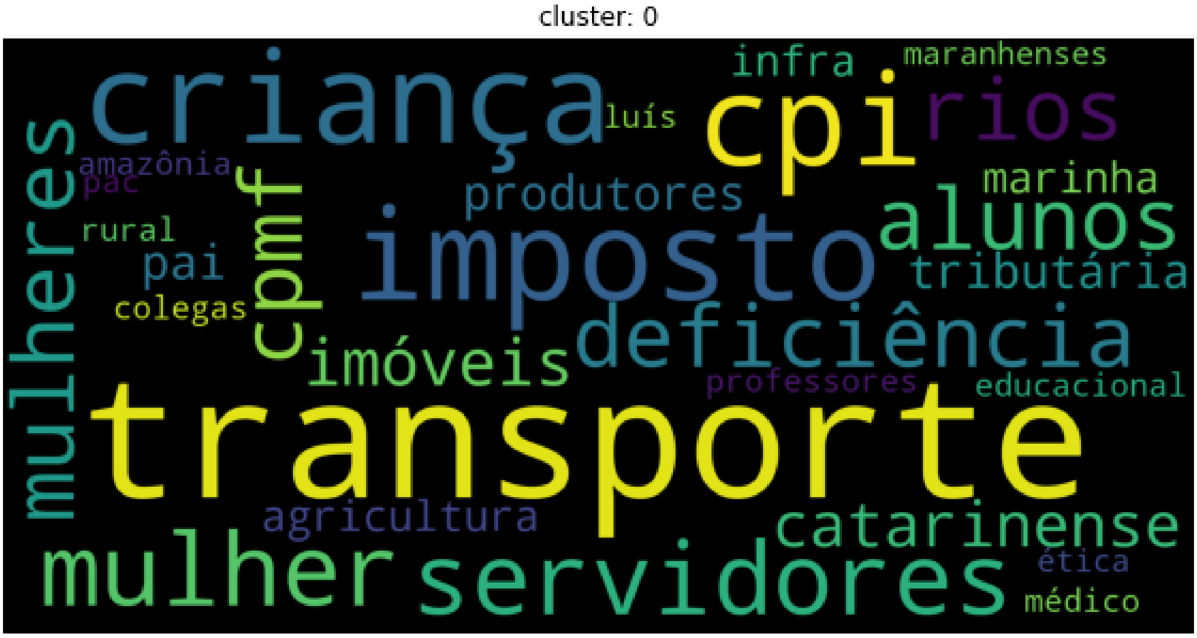
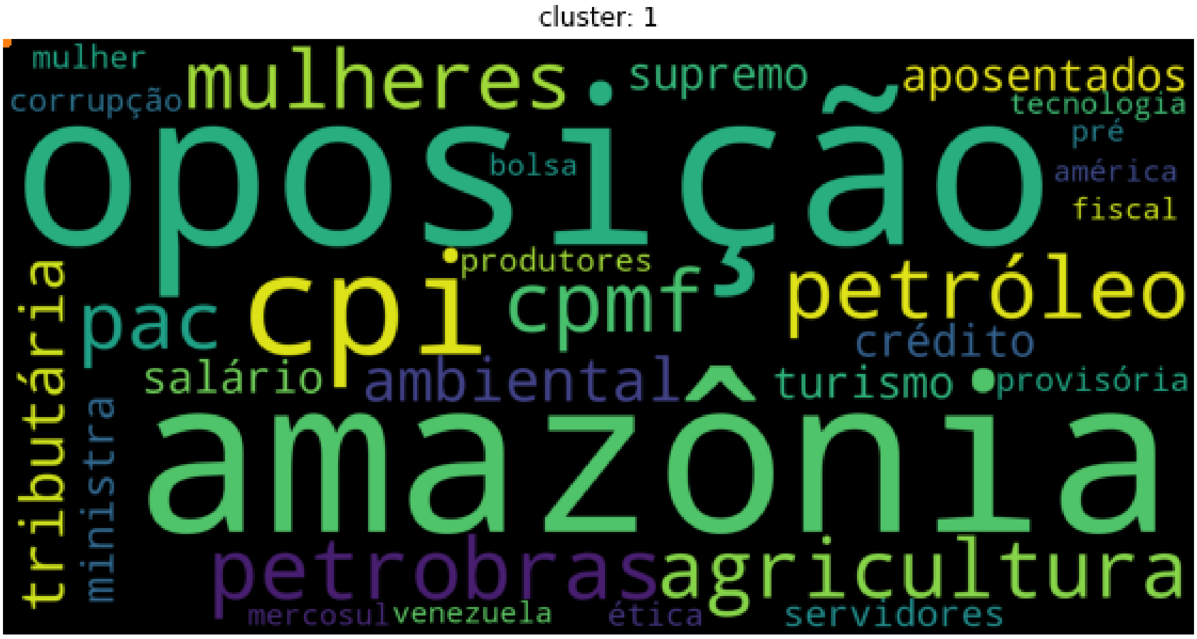


Figure 9 – Word cloud, cluster 1, 53rd legislature



The left-wing parties are concentrated almost entirely in cluster 1, the center parties are more in cluster 1, but there is a reasonable amount in cluster 0. The right-wing parties are divided almost equally between cluster 0 and 1. We again observe a difference in

[10] Percentage of senators parties that are in this cluster
[11] Percentage of senators of this ideology in this cluster
[12] Percentage of senators in cluster of this ideology

Table 13 – Division of parties between clusters in the 53rd legislature

Party	Cluster	Senators	Percentage [10]
PV	0	1	100.00
No party	0	3	100.00
PFL	0	13	76.47
DEM	0	18	60.00
PR	0	3	60.00
PDT	0	4	50.00
PSC	0	1	50.00
PTB	0	5	45.45
PMDB	0	10	38.46
PRB	0	1	33.33
PSB	0	1	33.33
PSDB	0	6	31.58
PT	0	3	18.75
PC DO B	1	1	100.00
PP	1	1	100.00
PSOL	1	1	100.00
PT	1	1	81.25
PSDB	1	13	68.42
PRB	1	2	66.67
PSB	1	2	66.67
PMDB	1	16	61.54
PTB	1	6	54.55
PDT	1	4	50.00
PSC	1	1	50.00
DEM	1	12	40.00
PR	1	2	40.00
PFL	1	4	23.53

Table 14 – Ideological composition of parties in each cluster, in the 53rd legislature

Cluster	0			1		
Ideology	Sen	% Ideo [11]	% Clus [12]	Sen	% Ideo	% Clus
Left	3	16.67	4.62	15	83.33	19.23
Center	10	45.45	15.38	12	54.55	15.38
Right	52	50.49	80.00	51	49.51	65.38

speeches between the right-wing and left-wing parties, although in previous legislatures this difference is larger.

Lula was re-elected president during this period. We observed that the division between clusters remained very similar to that of the previous legislature, indicating a strong influence of the president in the speeches of parliamentarians.

According to figure 8, the most influential words in cluster 0 are: child (criança), producers (produtores), transport (transporte), agriculture (agricultura) and servants (servidores). We observed the presence of many words in common, probably due to the influence of the president. We highlight the CPMF, a tax whose collection ended in 2007 (Cagnin; Freitas, 2015). The PAC theme also appears in both clusters and refers to the investment program implemented by the federal government.

In cluster 1, as shown in figure 9, we highlight the presence of the words: Amazon (amazônia) and environmental (ambiental). It is interesting how the Amazon theme has great weight in the speeches of leftist parties. In this cluster, the word Venezuela also appears, which despite having a small presence, was not present in the speeches of previous legislatures.

4.4 54th legislature

According to the Calinski-Harabasz index, there are three clusters in the 54th legislature. Figures 10, 11 and 12 present the most representative words of each cluster. While Table 15 presents the absolute presence and percentage of senators from each party according to the clusters, Table 16 shows the ideological classification of parties according to Tables 6 and 7. Tables 27, 28 and 29 in Appendix B present the most important words and their translations in respectively Clusters 0, 1 and 2 of the 54th legislature.

Figure 10 – Word cloud, cluster 0, 54th legislature

Figure 11 – Word cloud, cluster 1, 54th legislature

Figure 12 – Word cloud, cluster 2, 54th legislature

We have decided to keep the division of 3 clusters to be consistent with the methodology, but only one senator arises in cluster 0. Because this, the figure 10 has little variation in themes and words.

[13] Percentage of senators parties that are in this cluster
[14] Percentage of senators of this ideology in this cluster
[15] Percentage of senators in cluster of this ideology

Table 15 – Division of parties between clusters in the 54th legislature

Party	Cluster	Senators	Percentage [13]
PMDB	0	1	3.70
PC DO B	1	4	100.00
PRB	1	2	100.00
PSB	1	4	100.00
PSC	1	1	100.00
PSOL	1	2	100.00
PV	1	1	100.00
PT	1	14	93.33
PMDB	1	17	62.96
PP	1	5	62.50
PSDB	1	9	56.25
PTB	1	4	44.44
PDT	1	3	42.86
PR	1	3	30.00
DEM	1	2	11.11
MDB	2	3	100.00
PMN	2	1	100.00
PPL	2	1	100.00
PPS	2	1	100.00
PROS	2	1	100.00
PSD	2	3	100.00
No party	2	2	100.00
DEM	2	16	88.89
PR	2	7	70.00
PDT	2	4	57.14
PTB	2	5	55.56
PSDB	2	7	43.75
PP	2	3	37.50
PMDB	2	9	33.33
PT	2	1	6.67

Table 16 – Ideological composition of parties in each cluster, in the 54th legislature

Cluster	0			1			2		
Ideology	Sen	% Ideo [14]	% Clus [15]	Sen	% Ideo	% Clus	Sen	% Ideo	% Clus
Left	0	0.00%	0.00%	20	95.24%	28.57%	1	4.76%	4.41%
Center	0	0.00%	0.00%	11	52.38%	15.71%	10	47.62%	44.12%
Right	1	2.50%	100.00%	39	43.33%	55.71%	50	55.56%	51.47%

In the 54th legislature, leftist parties were concentrated in cluster 1, only one leftist senator had his speeches classified in cluster 2 and right parties were more concentrated in cluster 2 with an important amount of senators (43.33%) in cluster 1.

In the 54th legislature, the president was Dilma Roussef, from the Partido dos Trabalhadores (workers' party). We have a composition of the clusters similar to the two previous ones. In cluster 1 (Figure 11), we have several words related to the economy, probably due to the economic crisis that occurred in Brazil in this period (Barbosa, 2017). Among these words, we highlight: company (empresa); investments (investimentos); income (renda); economical (econômica); opportunities (oportunidade); construction (obras); growth (crescimento); economic (econômico); market (mercado); flat (plano); stock (ações).

In cluster 2 (Figure12), we have several political themes and the economic theme does not prevail. We highlight: water (água); constructions (obra); North East (nordeste); cash (dinheiro); teaching (ensino); violence (violência); cpi; crime; drugs (drogas). These are different themes and there is no predominance of any specific theme.

4.5 55th legislature

According to the Calinski-Harabasz index, there are again two clusters in the 55th legislature. Figures 13 and 14 present the most representative words of each cluster. While Table 17 presents the absolute presence and percentage of senators from each party according to the clusters, Table 18 shows the ideological classification of parties according to Tables 6 and 7. Tables 30 and 31 in Appendix B present the most important words and their translations in respectively Clusters 0, 1 and 2 of the 55th legislature.

Figure 13 – Word cloud, cluster 0, 55th legislature

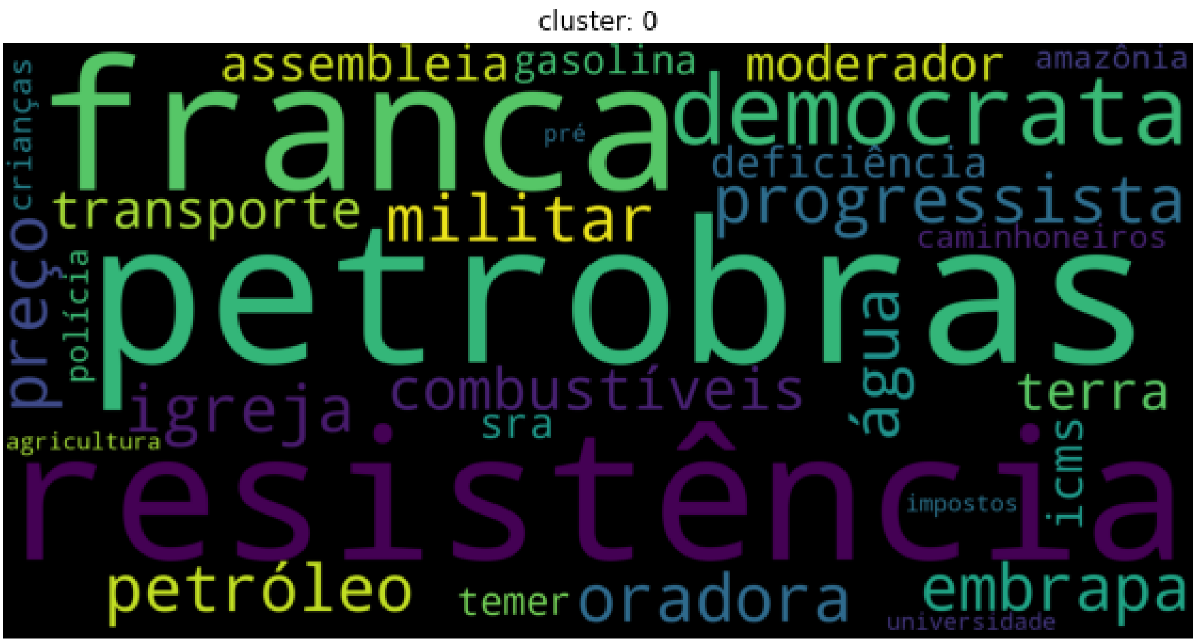
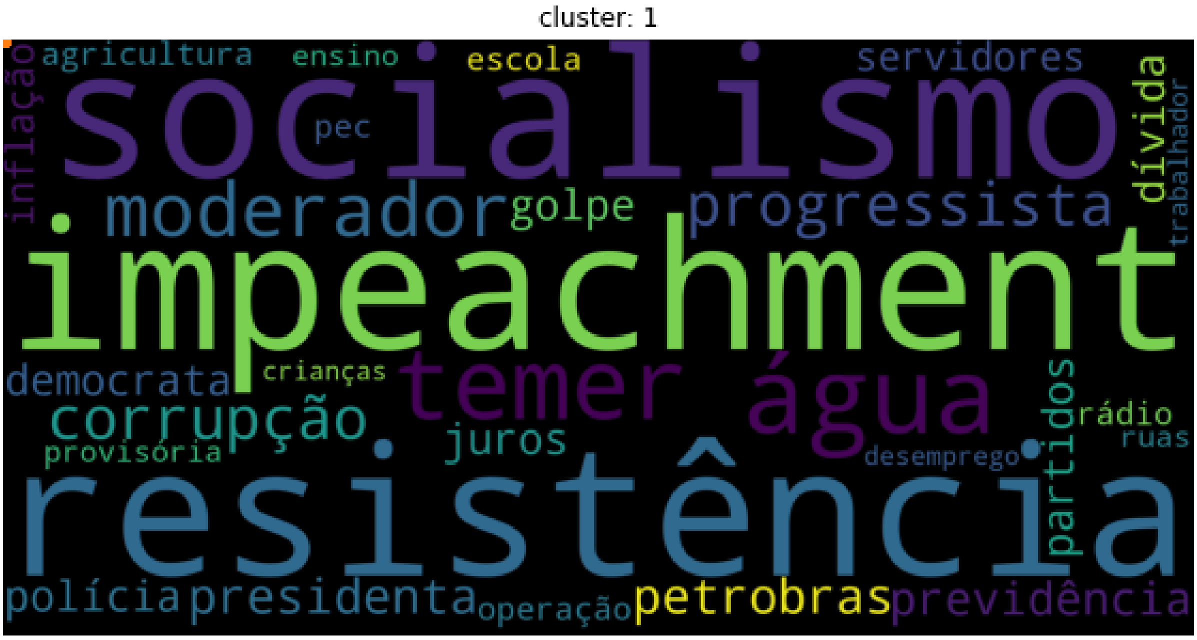


Figure 14 – Word cloud, cluster 1, 55th legislature



We have found that the left-wing parties are concentrated almost entirely in cluster 1, with only 1 left-wing senator in cluster 0. The center parties are more in cluster 1, but there is a reasonable amount in cluster 0. The right parties are equally divided between cluster 0 and 1.

At the beginning of the 55th legislature, the president was Dilma Roussef, from the Partido dos Trabalhadores (Workers' Party). At the end of 2015 there was an impeachment process that culminated in the removal of Dilma and Michel Temer assumed the presidency of Brazil.

Most senators were classified in cluster 1 (significant words as shown in Figure14). Many words are related to impeachment, such as: impeachment; Temer; democrat (democrata); resistance (resistência); president (presidenta); coup (golpe); moderator (moderador). We emphasize that virtually all left-wing senators are in this cluster and the impeachment process has a great influence in the discourse of the entire legislature.

In cluster 0, according to Figure13, impeachment was not the dominant theme, having a lot to do with the Temer government, such as: truck drivers (caminhoneiros), Petrobras, transport (transporte), price (preço), resistance (resistência), fuels (combustíveis), gasoline (gasolina), icms. Probably due to the truckers' strike that greatly affected Brazil (Lourenço, 2018). In this cluster there is only one left-wing senator. Despite being Dilma's deputy, Temer's support base is predominantly formed by center and right-wing senators,

[16] Percentage of senators parties that are in this cluster
[17] Percentage of senators of this ideology in this cluster
[18] Percentage of senators in cluster of this ideology

who formed this cluster.

Table 17 – Division of parties between clusters in the 55th legislature

Party	Cluster	Senators	Percentage [16]
CIDADANIA	0	2	100.00
DC	0	2	100.00
PL	0	1	100.00
PODEMOS	1	4	100.00
PRTB	0	1	100.00
PSL	0	1	100.00
PTRB	0	1	100.00
MDB	0	18	94.74
DEM	0	4	66.67
PODE	0	4	66.67
No party	0	5	62.50
PDT	0	6	50.00
PV	0	1	50.00
PSD	0	3	37.50
PMDB	0	9	33.33
PP	0	3	33.33
PR	0	2	33.33
PSC	0	1	33.33
PTB	0	2	28.57
PSB	0	2	25.00
PSDB	0	4	22.22
PT	0	1	7.14
PC DO B	1	1	100.00
PMB	1	1	100.00
PPS	1	2	100.00
PRB	1	3	100.00
PROS	1	1	100.00
PSOL	1	1	100.00
PTC	1	1	100.00
REDE	1	1	100.00
PT	1	13	92.86
PSDB	1	14	77.78
PSB	1	6	75.00
PTB	1	5	71.43
PMDB	1	18	66.67
PP	1	6	66.67
PR	1	4	66.67
PSC	1	2	66.67
PSD	1	5	62.50
PDT	1	6	50.00
PV	1	1	50.00
No party	1	3	37.50
DEM	1	2	33.33
PODE	1	2	33.33
MDB	1	1	5.26

Table 18 – Ideological composition of parties in each cluster, in the 55th legislature

Cluster	0			1		
Ideology	Sen	% Ideo [17]	% Clus [18]	Sen	% Ideo	% Clus
Left	1	6.25	1.54	15	93.75	16.30
Center	10	35.71	15.38	18	64.29	19.57
Right	54	47.79	83.08	59	52.21	64.13

4.6 Party cohesion measure

In Section 3, we mention that we may use the percentage of senators of a given party that belongs to the cluster with the greater concentration of them as a measure of party cohesion. Table 19 uses the results of Tables 9, 11, 13, 15 and 17 to present our measure of party cohesion for each party.

Table 19 – Party cohesion. We do not consider parties with less than 3 senators in the legislature.

Party	% 51st	% 52nd	% 53rd	% 54th	% 55th
DEM/PFL	60.71	61.90	76.47	88.89	66.67
MDB/PMDB	51.22	58.97	61.54	62.96	66.67
PC do B	-	-	-	100.00	-
PDT	100.00	83.33	50.00	57.14	50.00
PL/PR	-	60.00	60.00	70.00	66.67
PMR/PRB	-	-	66.67	-	100.00
PODE/PODEMOS	-	-	-	-	66.67
PP/PPB	42.86	-	-	62.50	66.67
PSB	-	-	66.67	100.00	75.00
PSC	-	-	-	-	66.67
PSD	-	-	-	100.00	62.50
PSDB	60.87	52.38	68.42	56.25	77.78
PT	100.00	80.00	81.25	93.33	92.86
PTB	60.00	71.43	54.55	55.56	71.43
NO PARTY	80.00	100.00	100.00	-	62.50

According to table 19, we conclude that the most cohesive party, in the studied period, is the Partido dos Trabalhadores (workers' party). In general, the parties show cohesion. An interesting finding is the great cohesion among non-party senators.

5 Summary and conclusion

Our work has proposed a data-driven based approach to measure party ideological cohesion. We investigate a dataset formed by senators' speeches from 1999 to 2018. We conclude that senators from the same party tend to use similar speeches. In the analyzed period, the speeches of left-wing senators are more similar than the speeches of center and right-wing parties. We may justify these findings by two important points. First, most of the analyzed period was governed by the left-wing party (the PT governed Brazil from 2002 to 2016) and may have influenced center and right-wing parties. Second, there are fewer left-wing senators. This fact makes it easier for speeches to be more similar.

Bibliography

Azevedo, C.; Rebelo, A. A corrupção no futebol brasileiro. *Revista Motrivivência*, p. 15–45, 2001. Cited on page 32.

Baeza-Yates, R.; Ribeiro-Neto, B. *Modern Information Retrieval*. 2nd. ed. USA: Addison-Wesley Publishing Company, 2008. ISBN 9780321416919. Cited on page 23.

Barbosa, F. d. H. A crise econômica de 2014/2017. *Estudos avançados*, SciELO Brasil, v. 31, p. 51–60, 2017. Cited on page 42.

Bardelin, C. E. A. *Os efeitos do racionamento de energia elétrica ocorrido no Brasil em 2001 e 2002 com ênfase no consumo de energia elétrica.* Tese (Doutorado) — Universidade de São Paulo, 2004. Cited on page 32.

Bolognesi, B.; Ribeiro, E.; Codato, A. Uma nova classificação ideológica dos partidos políticos brasileiros. *Dados*, SciELO Brasil, v. 66, 2022. Cited 7 times on pages 11, 17, 18, 25, 26, 27, and 32.

Cagnin, R. F.; Freitas, M. C. P. de. Tributação das transações financeiras: a experiência brasileira com o iof e a cpmf. *Análise Econômica*, v. 33, n. 63, 2015. Cited on page 38.

Figueredo, F. C. d.; Mueller, B.; Cajueiro, D. O. A natural language measure of ideology in the brazilian senate. *Revista Brasileira de Ciência Política*, SciELO Brasil, 2022. Cited on page 17.

Franzmann, S.; Kaiser, A. Locating political parties in policy space: A reanalysis of party manifesto data. *Party politics*, Sage Publications London, Thousand Oaks, New Delhi, v. 12, n. 2, p. 163–188, 2006. Cited on page 18.

Giannetti, D.; Laver, M. Party cohesion, party discipline and party factions in italy. In: *Intra-party politics and coalition governments.* : Routledge, 2008. p. 162–184. Cited on page 17.

Gnanadesikan, R. *Methods for statistical data analysis of multivariate observations.* : John Wiley & Sons, 2011. Cited on page 23.

Hazan, R. *Cohesion and Discipline in Legislatures: Political Parties, Party Leadership, Parliamentary Committees and Governance.* Routledge, 2006. (Cohesion and Discipline in Legislatures: Political Parties, Party Leadership, Parliamentary Committees and Governance). ISBN 9780415360142. Available on: <https://books.google.com.br/books?id=_lUg7dvXowwC>. Cited on page 18.

Hix, S.; Noury, A. G.; Roland, G. *Democratic politics in the European Parliament.* : Cambridge University Press, 2007. Cited on page 17.

Jahn, D.; Oberst, C. Ideological party cohesion in macro-comparative politics: The nordic social democratic parties from a comparative perspective. *Scandinavian Political Studies*, v. 35, n. 3, p. 222–245, 2012. Available on: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9477.2012.00287.x>. Cited on page 17.

Lisboa, M. G. de O.; Branco, R. C. C. Dorothy stang e o protagonismo feminino na luta pela terra e território de vida paraense. *Revista GeoAmazônia*, v. 9, n. 18, p. 57–82, 2022. Cited on page 34.

Lloyd, S. Least squares quantization in pcm. *IEEE transactions on information theory*, IEEE, v. 28, n. 2, p. 129–137, 1982. Cited on page 23.

Lourenço, G. M. O efeito da greve dos caminhoneiros. *Vitrine da Conjuntura, Curitiba*, v. 11, n. 6, 2018. Cited on page 44.

MacQueen, J. Classification and analysis of multivariate observations. In: University of California Los Angeles LA USA. *5th Berkeley Symp. Math. Statist. Probability*. 1967. p. 281–297. Cited on page 23.

Mainwaring, S.; Liñán, A. P. Party discipline in the brazilian constitutional congress. *Legislative studies quarterly*, JSTOR, p. 453–483, 1997. Cited on page 18.

Manning, C. D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008. ISBN 0521865719, 9780521865715. Cited on page 23.

Neiva, P. R. P. Coesão e disciplina partidária no senado federal. *Dados*, SciELO Brasil, v. 54, p. 289–318, 2011. Cited on page 18.

Neiva, P. R. P. Disciplina partidária e apoio ao governo no bicameralismo brasileiro. *Revista de Sociologia e Política*, SciELO Brasil, v. 19, p. 183–196, 2011. Cited on page 18.

Oliveira, M. S. d. Movimento para as instituições: ambientalistas, partidos políticos e a liderança de marina silva. 2016. Cited on page 34.

Özbudun, E. *Party Cohesion in Western Democracies: A Causal Analysis*. Sage Publications, 1970. (Comparative politics series, Nº 6). ISBN 9780803901063. Available on: <https://books.google.com.br/books?id=DU2JSAAACAAJ>. Cited on page 18.

Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Cited 2 times on pages 23 and 24.

Piubelli, R. Memórias e imagens em torno do índio pataxó hãhãhãe galdino jesus dos santos (1997 a 2012). 2012. Cited on page 32.

Poole, K. T. *Spatial models of parliamentary voting*. : Cambridge University Press, 2005. Cited on page 17.

Poole, K. T.; Rosenthal, H. D-nominate after 10 years: A comparative update to congress: A political-economic history of roll-call voting. *Legislative Studies Quarterly*, JSTOR, p. 5–29, 2001. Cited on page 17.

Santos, F. Partidos e comissões no presidencialismo de coalizão. *Dados*, SciELO Brasil, v. 45, p. 237–264, 2002. Cited on page 18.

Scheeffer, F. et al. Ideologia e comportamento parlamentar na câmara dos deputados: faz sentido ainda falar em esquerda e direita? 2016. Cited on page 18.

Strom, K. A behavioral theory of competitive political parties. *American journal of political science*, JSTOR, p. 565–598, 1990. Cited on page 18.

Tarouco, G. d. S.; Madeira, R. M. Partidos, programas e o debate sobre esquerda e direita no brasil. *Revista de Sociologia e politica*, SciELO Brasil, v. 21, p. 149–165, 2013. Cited on page 18.

Wang, X.; Xu, Y. An improved index for clustering validation based on silhouette index and calinski-harabasz index. In: IOP Publishing. *IOP Conference Series: Materials Science and Engineering*. 2019. v. 569, n. 5, p. 052024. Cited on page 24.

Wolinetz, S. B. Beyond the catch-all party: approaches to the study of parties and party organization in contemporary democracies. *Political parties: Old concepts and new challenges*, Oxford University Press Oxford, p. 136–165, 2002. Cited on page 18.

Appendix

APPENDIX A – list of removed words

The removed words are all lowercase, as all speeches were treated that way. This is necessary for word recognition. The list is composed as follows: stopwords (from the NLTK library and Spacy), names of senators from the period, names of Brazilian states and their acronyms and manual addition of words that do not bring meaning related to the senator's speech.

a, à, abençoe, abraço, abreu, ac, ação, acerca, acima, acir, acórdão, acordo, acre, ada, adelmir, ademais, ademir, adeus, adir, aécio, aelton, afonso, agenda, agora, agradeço, agripino, aí, ainda, airton, ajuste, al, alagoas, alberto, albino, albuquerque, alcântara, alckmin, alcolumbre, além, alemães, alencar, alessandro, alfredo, algo, alguém, algum, alguma, algumas, alguns, ali, aliás, almeida, aloizio, aloysio, althoff, altura, alvaro, álvaro, alves, am, amapá, amaral, amauri, amazonas, amazonense, ambas, ambos, amélia, amigo, amin, amir, amorim, ana, anais, anastasia, andrade, ângela, angelo, anibal, aníbal, antero, antes, antonio, antônio, ao, aonde, aos, ap, apanhamento, aparte, apenas, apoia, apoio, apontar, após, aquela, aquelas, aquele, aqueles, aqui, aquilo, aracaju, araújo, área, áreas, argello, arlindo, armando, arns, arolde, arruda, art, arthur, artigo, artur, as, às, aspas, assim, assis, assunto, assuntos, ataídes, atalaia, até, atenção, atos, atrás, através, atual, augusto, aureliano, autor, autoria, autoridades, azeredo, aziz, b, ba, bahia, baixo, bala, barbalho, barbaridade, barreto, barros, base, bastante, batalha, batista, bauer, beber, bel, belini, bello, bem, benedito, benício, berger, bernardo, bessa, bezerra, bilhões, bittar, blairo, bloco, blumenau, boa, boaventura, bolsonaro, bom, borges, bornhausen, botelho, br, braga, brasileiras, brasileiros, brasília, braziliense, breve, brito, brizola, buarque, bulhões, cá, cabral, cada, cafeteira, caiado, calheiros, calixto, câmara, camargo, camata, cameli, caminho, campainha, campos, candidato, cândido, capiberibe, capital, cardoso, carepa, carlos, carmo, caruaru, carvalho, casagrande, casildo, caso, casos, cássio, cassol, castro, catarina, catorze, cavalcanti, ce, ceara, ceará, cedo, cento, central, cerca, certamente, certeza, césar, charles, chaves, chico, ciarlini, cícero, cid, cidadania, cidinho, cima, cinco, cintra, ciro, citada, claro, claudino, cleide, clésio, clodoaldo, clovis, código, coelho, coisa, coisas, collor, colombo, com, comissão, como, companheiro, comprida, comprido, condições, confederações, conforme, confúcio, congresso, conhecida, conhecido, consciência, conselho, conste, contarato, conto, contra, contudo, convivência, coronel, correia, corrente, costa, couto, criação, cristovam, crivella, cuiabá, cuja, cujo, cunha, cury, custa, cyro, da, dá, dalirio, daniella, dantas, dão, daquela, daquele, dar, dário, das, davi, davim, de, debaixo, debate, deca, década, decisão, defeso, dela, delas, delcídio, dele, deles, dem, demais, demarchi, demóstenes, dentro, depois, deputado, deputados, des, desafios, desde, dessa, desse, desses, desta, destaca, deste, deve, devem, deverá, dez, dezanove, dezasseis, dezassete, dezoito, df, diante, dias, dilma, diniz, dirceu, direita, direta, diretamente, dirigente, disso, distrito, diz, dizem, dizer, djalma, do, documento, documentos, dois, dom, donizeti, dornelles, dos, douglas, doze, dr, duarte, duas, duciomar, dunga, duque, durante, durval, dutra, dúvida, e, é, edição, edison, editorial, eduardo, efetivamente, efraim, ela, elas, elber, ele, eles, eliane, elias, elifas, eliseu, eliza, eliziane, elmano, em, embora, emenda, emília, eminente, encaminhamento, enquanto, então, entre, entrevista, epitácio, época, era, eram, éramos, ernandes, es, és, escórcio, especial, especialmente, esperidião, espírito, essa, essas, esse, esses, esta, está, estado, estados, estadual, estamos, estão, estar, estará, estas, estás, estava, estavam, estávamos, este, esteja, estejam, estejamos, estes, estevão, esteve, estive, estivemos, estiver, estivera, estiveram, estivéramos, estiverem, estivermos, estivesse, estivessem, estivéssemos, estiveste, estivestes, estou, estrutura, eu, euclydes, eunício, eurípedes, eventual, evidentemente, ex, exa, exame, exatamente, excelência, exemplo, exemplos, expedito, experiências, expressão, fabiano, fácil, faço, fagundes, falta, fará, farias, fátima, fato, fatos, faustino, fávaro, favor, favoráveis, faz, fazeis, fazem, fazemos, fazer, fazes, fazia, fecury, federal, fernandes, fernando, ferraço, ferreira, férrer, fez, figueiró, filho, fim, final, finalidade, fiquene, fiscalização, flávio, fleury, flexa, fogaça, foi, folha, fomos, fonseca, for, fora, foram, fôramos, forem, forma, formos, fortes, fosse, fossem, fôssemos, foste, fostes, francelino, francisco, franco, freire, freitas, fui, fulano, função, fundamental, furlan, gabardo, gabrilli, gama, garibaldi, gentil, geovani, geral, geraldo, gerson, gilberto, gilvam, giordano, girão, givago, gladson, gleisi, go, goellner, goiano, goiás, gomes, gondim, goulart, governador, governadora, governadores, grande, grandes, grave, grazziotin, grosso, grupo, guerra, guimarães, gurgacz, há, haja, hajam, hajamos, hão, hartung, havemos, hei, heinze, helena, hélio, heloísa, henrique, heráclito, hoffmann, hora, houve, houvemos, houver, houvera, houverá, houveram, houvéramos, houverão, houverei, houverem, houveremos, houveria, houveriam, houveríamos, houvermos, houvesse, houvessem, houvéssemos, hugo, humberto, ideli, ilustre, importância, importante, importantes, inácio, inciso, inclusive, infralula, iniciar, inicio, integrante, interessante, interesse, interno, ione, ir, irá, irajá, iris, íris, isso, isto, itamar, ivo, ivonete, izalci, já, jader, jamais, janeiro, jaques, jarbas, jayme, jean, jefferson, jereissati, jesus, joão, joaquim, jonas, jorge, jorginho, josé, jucá, juíza, júlia, júlio, junho, júnior, junto, justamente, juvêncio, kajuru, kaká, kátia, km, kubitschek, lá, lado, lamentavelmente, lando, lasier, lauro, leila, leite, leomar, leonel, lhe, lhes, líder, líderes, lídice, ligado, lima, lindberg, lindbergh, lira, listas, lobão, local, logo, longe, longo, lopes, loyola, lucas, lucena, lúcia, lúcio, lúdio, lugar, luis, luiz, lula, luz, luzia, ma, macapá, machado, maciel, magalhães, maggi, magno, maguito, maia, mailza, maio, maior, maioria, maiorias, mais, major, mal, maldaner, malta, manaus, maneira, mão, mãos, mara, maranhao, maranhão, maranhense, marcelo, marcio, marco, março, marconi, marcos, maria, marina, marinho, marinor, mário, marisa, marluce, marta, martins, mas, mata, matéria, matérias, mato, matos, matusalém, mauro, máximo, mdb, me, mecias, medeiros, medida, medidas, médio, meio, melhor, mello, melo, mendes, menor, menos, mercadante, mês, mesa, meses, mesmo, mesquita, mestrinho, meu, meurer, meus, mg, michel, mil, milhões, mim, minas gerais, minha, minhas, ministério, ministro, minutos, miranda, missões, modelo, moka, momento, monteiro, morais, moreira, moro, mossoró, mota, motta, moura, mozarildo, ms, mt, muita, muitas, muito, muitos, muniz, mw, na, nabor, nada, não, napoleão, naquela, naquele, nas, nascimento, nelsinho, nem, nenhum, nenhuma, nery, nessa, nesse, nesta, neste, neto, neuto, neves, ney, nezinho, nilda, ninguém, níura, nível, no, nº, nobre, nobres, nogueira, noite, nome, norte, nos, nós, nossa, nossas, nosso, nossos, nova, novamente, novas, nove, novo, novos, num, numa, número, números, nunca, nunes, nuns, o, obrigada, obrigado, obviamente, óbvio, octávio, ocupo, odacir, ofício, oitava, oitavo, oito, olimpio, oliveira, olivir, omar, onde, ontem, onze, ora, orador, oriovisto, ornelas, os, osmar, osvaldo, otavio, otto, ou, outra, outras, outro, outros, pa, pacheco, paes, paim, palavra, palavras, palma, palmas, palocci, papaléo, papel, para, pará, parabéns, paraíba, paraná, parece, parga, parlamentar, parlamentares, parte, partido, partir, pastor, pastore, patrícia, patriota, patrocínio, paul, paulo, pavan, pb, pc, pcdob, pdt, pe, pedro, pegar, pela, pelas, pelo, pelos, pereira, peres, péres, perillo, período, pernambuco, perrella, perto, pessoa, petecão, petrônio, pfl, pi, piauí, piauiense, pimentel, pinheiro, pinto, pior, piselo, piva, pl, plenário, plínio, pmb, pmdb, pmn, pmr, pode, pôde, podem, podemos, poder, poderá, podia, põe, põem, pois, políticos, pontes, ponto, pontos, população, por, porém, porquanto, porque, porquê, portanto, portela, portinho, porto, posição, possibilidade, possível, possivelmente, posso, pouca, pouco, povo, pp, ppb, ppl, pps, pr, praia, prates, prb, preciso, prefeitos, presidência, primeira, primeiro, principalmente, prisco, problemas, processo, programas, projeto, pronunciamento, própria, próprio, pros, próxima, próximo, prp, prtb, psb, psc, psd, psdb, psl, psol, pt, ptb, ptc, puderam, pv, quais, qual, qualquer, quando, quanto, quarta, quarto, quase, quatro, que, quê, quem, quer, querem, querido, quero, questão, quieta, quieto, quinta, quintanilha, quinto, quinze, raimundo, ramez, randolfe, raras, raupp, razão, realmente, recado, rede, reditario, regime, regimento, regina, reginaldo, registro, rêgo, reguffe, relação, renan, renato, renilde, renildo, repito, republicanos, requeiro, requerimento, requião, resende, revista, rezende, ribamar, ribeiro, ricardo, rio, rita, rj, rn, ro, roberto, rocha, rodolpho, rodrigo, rodrigues, rogério, rollemberg, romário, romero, romeu, ronaldo, rondônia, roraima, roriz, rosado, rosalba, rose, roseana, roussef, rr, rs, ruben, rudson, russo, sabe, saber, saboya, sadi, salgado, salvador, salvatti, sampaio, sandoval, sandra, santa, santana, santiago, santo, santoro, santos, sao, são, sarney, sartori, saturnino, sc, se, sebastião, secretaria, seguinte, segunda, segundo, sei, seis, seja, sejam, sejamos, selma, sem, semana, sempre, senadora, senadoras, sendo, senhor, sentido, ser, será, serão, serei, seremos, sergio, sérgio, sergipano, sergipe, seria, seriam, seríamos, seridó, serra, serrano, serviço, serys, sessão, sete, setembro, sétima, sétimo, setor, seu, seus, sexta, sexto, sf, sibá, sido, silva, silveira, sim, simon, simone, siqueira, sistema, situação, slhessarenko, só, soares, sob, sobre, sobretudo, sobrinho, sodré, sois, solicito, somente, somos, soraya, sou, sousa, souto, souza, sp, spartido, srª, sras, srªs, styvenson, sua, suas, suassuna, subsecretaria, sul, suplente, suplicy, tabela, tais, tal, talvez, também, tanta, tanto, tão, taques, taquigrafia, taquigráfico, tarde, tasso, tavola, te, tebet, telmário, tem, tém, têm, tema, temos, tempo, tendes, tenha, tenham, tenhamos, tenho, tenório, tens, tentar, tentaram, tente, tentei, teoria, teotonio, teotônio, ter, terá, terão, terceira, terceiro, terei, teremos, teresina, teria, teriam, teríamos, termos, teu, teus, teve, thelma, thronicke, tião, tinha, tinham, tínhamos, tipo, títulos, tive, tivemos, tiver, tivera, tiveram, tivéramos, tiverem, tivermos, tivesse, tivessem, tivéssemos, tiveste, tivestes, to, tocantinense, tocantins, toda, todas, todo, todos, toledo, tomás, torres, tourinho, trad, três, treze, tribuna, tu, tua, tuas, tudo, tuma, ubirajara, último, últimos, um, uma, umas, unidos, uns, us, usa, usar, uso, vai, vais, val, valadares, valdir, valentim, valério, valmir, valor, valter, vamos, vanderlan, vanessa, vânia, vão, vários, vasco, vasconcelos, veja, vem, vêm, veneziano, vens, ver, vez, vezes, viana, vice, vicente, vicentinho, vieira, vilela, vinda, vindo, vinte, virgílio, virginio, vital, você, vocês, vos, vós, vossa, vossas, vosso, vossos, voto, vou, wagner, waldeck, waldemir, waldomiro, walter, wellington, weverton, wilder, wilson, wirlande, yanai, zambiasi, zenaide, zequinha, zero, zeze, zona.

APPENDIX B – list of words in clusters

The words of the clusters and their translations are in the following tables 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31:

Table 20 – Wordcloud of the cluster 0, legislature 51

Word in portuguese	Word in english
casa	house
política	policy
brasileira	brazilian
recursos	resources
ano	year
mundo	world
desenvolvimento	development
trabalho	work
vida	life
sociedade	society
lei	law
república	republic
pessoas	people
social	social
saúde	health
pública	public
público	public
países	country
discurso	speech
região	region
economia	economy
respeito	respect
mercado	market
justiça	justice
educação	education
banco	bank
amazônia	amazônia
programa	program

Table 21 – Wordcloud of the cluster 1, legislature 51

Word in portuguese	Word in english
casa	house
ensino	teaching
lei	law
energia	energy
pantanal	pantanal
justiça	justice
política	policy
transformação	transformation
ira	ire
missão	mission
fmi	imf
república	republic
esquerda	left
súmula	docket
mulher	womam
agrícola	agricultural
colégio	school
igreja	church
futebol	soccer
empresa	firm
redação	redaction
violência	violence
oportunidades	opportunities
jornal	newspaper
sociedade	society
pessoas	people
brasileiro	brazilian
deficiência	deficiency
programa	program
ações	actions

Table 22 – Wordcloud of the cluster 2, legislature 51

Word in portuguese	Word in english
região	region
nordeste	nordeste
estradas	roads
amazônia	amazônia
desenvolvimento	development
sudene	sudene
livros	books
parnaíba	parnaíba
recursos	resources
petróleo	petroleum
corrupção	corruption
regiões	regions
recife	recife
regional	regional
campina	meadow
índios	indians
índio	indian
ano	year
educação	education
figueiredo	figueiredo
política	policy
ética	ethic
fundef	fundef
água	water
capanema	capanema
queiroz	queiroz
cidade	city
padre	priest
agricultura	agriculture
vida	life

Table 23 – Wordcloud of the cluster 0, legislature 52

Word in portuguese	Word in english
amazônia	amazônia
mulheres	women
soja	soy
arma	weapon
rural	rural
petróleo	petroleum
cade	cade
rurais	rural
território	territory
garoto	boy
álcool	alcohol
globo	globo
servidores	civilservant
sol	sun
impostos	taxes
deficiência	deficiency
agrária	agrarian
belém	belém
vereadores	city councilman
doença	illness
transposição	transposition
assentamento	settlement
digital	digital
incra	incra
oradora	speaker
bancos	banks
armas	weapons
adolescentes	teenagers
transportes	transport
alegria	happiness

Table 24 – Wordcloud of the cluster 1, legislature 52

Word in portuguese	Word in english
amazônia	amazônia
petrobras	petrobras
mulheres	women
supremo	supreme court
gás	gas
petróleo	petroleum
ética	ethic
globo	globo
bolsa	stock exchange
relações	relations
votação	vote
relator	referendary
forças	forces
caixa	caixa
professor	teacher
carga	load
comércio	business
ouço	hear
financiamento	financing
ministros	ministers
trabalhador	worker
integração	integration
inteiro	whole
investigação	investigation
oeste	west
pib	gdp
televisão	television
militar	military
médico	doctor
entendimento	understandings

Table 25 – Wordcloud of the cluster 0, legislature 53

Word in portuguese	Word in english
cpi	cpi
imposto	tax
criança	child
servidores	civil servants
mulher	woman
deficiência	deficiency
cpmf	cpmf
mulheres	women
alunos	students
rios	rivers
imóveis	properties
catarinense	catarinense
produtores	producers
pai	father
tributária	tributary
agricultura	agriculture
transporte	transport
infra	infrastructure
marinha	navy
professores	teachers
transportes	transports
amazônia	amazônia
pac	pac
colegas	mate
educacional	educational
rural	rural
médico	medic
maranhenses	maranhenses
luís	luís
ética	ethic

Table 26 – Wordcloud of the cluster 1, legislature 53

Word in portuguese	Word in english
amazônia	amazônia
oposição	oppsotition
cpi	cpi
petrobras	petrobras
petróleo	petroleum
cpmf	cpmf
agricultura	agriculture
mulheres	women
pac	pac
ambiental	environmental
tributária	tributary
aposentados	retirees
supremo	supreme court
crédito	credit
turismo	tourism
salário	wage
ministra	minister
servidores	civil servants
provisória	provisional
produtores	producers
ética	ethic
corrupção	corruption
mulher	woman
tecnologia	technology
américa	america
venezuela	venezuela
bolsa	stock exchange
mercosul	mercosul
fiscal	fiscal
pré	pre

Table 27 – Wordcloud of the cluster 0, legislature 54

Word in portuguese	Word in english
doenças	diseases
culpa	fault
campo	field
município	county
empresarial	bussiness
empreendimentos	enterprises
empate	draw
empenho	effort
empreendedor	entrepreneur
empreendedores	entrepreneurs
empreendedorismo	entrepreneurship
empreendimento	enterprise
empregadores	employers
empregador	employer
empresariado	bussiness community
emprego	employment
empregos	jobs
empreteira	contractor
empreteiras	contractors
empresa	company
empregados	employees
útil	useful
emoção	emotion
emendas	amendments
embaixador	ambassador
embaixo	below
embargo	embargo
embargos	embargoes
embate	clash
embraer	embraer

Table 28 – Wordcloud of the cluster 1, legislature 54

Word in portuguese	Word in english
presidenta	chairwoman
oportunidade	opportunity
mulheres	women
investimentos	investments
empresas	companies
energia	energy
crescimento	growth
proposta	proposal
direitos	rights
participação	participation
segurança	safety
tribunal	court
ações	actions
renda	income
sociais	social
políticas	policies
nordeste	northeast
ecônomica	economic
mercado	market
defesa	defense
plano	plan
públicos	public
obras	construction
trabalhadores	workers
empresa	company
passado	past
violência	violence
internacional	international
públicas	public
econômico	economic

Table 29 – Wordcloud of the cluster 2, legislature 54

Word in portuguese	Word in english
nordeste	northeast
dinheiro	money
água	water
presidenta	chairwoman
ensino	teaching
mulheres	women
força	force
produtores	producers
violência	violence
mulher	woman
homem	man
prefeito	mayor
ética	ethic
obras	constructions
inflação	inflation
obra	construction
minoria	minority
drogas	drugs
energia	energy
penal	criminal
cultura	culture
turismo	tourism
polícia	cop
transporte	transport
horário	horary
proposta	proposal
amazônia	amazônia
cpi	cpi
crime	crime
município	county

Table 30 – Wordcloud of the cluster 0, legislature 55

Word in portuguese	Word in english
petrobras	petrobras
resistência	resistance
franca	free zone
democrata	democrat
progressista	progressive
preço	price
oradora	speaker
petróleo	petroleum
militar	military
água	water
embrapa	embrapa
igreja	church
combustíveis	fuels
transporte	transport
terra	land
assembleia	assembly
icms	icms
moderador	moderator
deficiência	deficiency
sra	lady
gasolina	gasoline
temer	temer
caminhoneiros	truckers
amazônia	amazon
polícia	police
crianças	children
agricultura	agriculture
impostos	taxes
pré	pre
universidade	university

Table 31 – Wordcloud of the cluster 1, legislature 55

Word in portuguese	Word in english
resistência	resistance
impeachment	impeachment
socialismo	socialism
água	water
temer	temer
moderador	moderator
progressista	progressive
corrupção	corruption
petrobras	petrobras
presidenta	president
previdência	welfare
democrata	democrat
juros	fees
polícia	police
partidos	parties
golpe	coup
dívida	debt
servidores	public servants
inflação	inflation
agricultura	agriculture
pec	pec
operação	operation
rádio	radio
escola	school
provisória	provisional
ensino	teaching
ruas	streets
desemprego	unemployment
trabalhador	worker
crianças	children