A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics

Quantifying Implicit Political Intentions in Parliamentary Discourse

An Integrated Approach using Semi-Supervised Learning and Vector Space Information Retrieval

Niclas Frederic Sturm | 45914

Work project carried out under the supervision of:

Prof. Leid Zejnilovic

December 15, 2021

Abstract

Recent advances in Natural Language Processing and Information Retrieval have opened a new world of possibilities for the analysis of text. This study seeks to explore the possibilities of applying these techniques on political text, with a focus on quantifying intentions in parliamentary speeches and activities in Portugal. Combining vector space models and semi-supervised learning, a semantic search engine is able to extract meaningful metrics from text that help to identify political trends and quantify alignment with political issues.

Keywords: Natural Language Processing, Political Science, Semi-Supervised Learning, Information Retrieval, Discourse Analysis

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Contents

1 Introduction

Political discourse in the shape of political speeches, political campaign manifests or other contributions by political actors to the public forum is at the heart of understanding the changing and shifting political landscape over time or at a given moment. The unstructured way in which this data is often available to the public exacerbates the difficulty of analyzing text from the political domain and poses a great barrier to political transparency. Given that politics in the 21st century needs to address a plethora of political issues, it is indispensable to empower civil society, including NGOs and citizens to better understand and monitor the activities undertaken by their representatives. While in the case of Portugal, the parliament offers an *Open Data-Portal*, the functionalities appear limited, especially concerning simple information retrieval via search. There are no means of aggregation and every parliamentary activity needs to be laboriously searched for. This workproject seeks to address two issues to enable further research in this area and show the potential of State-of-the-Art Natural Language Processing solutions for discourse analysis. In terms of the type of political discourse surveyed, political *intentions* were the main focus as they are related to actions parliamentarians take as part of their duty. The implementation of this information retrieval system will introduce several components that can be reused for research and are available in an Open-Source repository *(Symmachus.jl)*: (a) A self-training labelling machine and (b) two semantic search methods for political speech and parliamentary activities. These search methods are a *speech-activity-match search*, which matches plenary speeches with parliamentary activities and a *topic search* that computes semantic alignment of a political actor with a political issue. Evaluating and demonstrating the utility of the search required a set of political issues and respective descriptions. To this end, this workproject used the Sustainable Development Goals as defined by the United Nations and their description in Portuguese. The results obtained have shown promise and demonstrate uses in research areas pertaining to Political Science, such as tracking the development of political issues over time, topic alignment for individual political actors as well as »scanning« the speeches produced by political actors for political intentions.

2 Background and Related Work

In the field of Political Science, the keyphrase »Text as Data« (Grimmer and Brandon 2013) has been gaining traction in recent years and more quantitative methods have entered the toolbox: Beyond standard corpus statistics work, such as the one by Laver, Benoit, and Garry (2003) on policy positions in the British party system, the main focus of this undertaking has been to perform *Topic Modeling*, an approach popularized by the classic but unabatedly popular implementation of the *Latent Dirichlet Allocation* (LDA) (Blei, Ng, and Jordan 2003). In the context of discourse analysis, this model has been used to create a topic model of the European Parliament's discourse (Greene and Cross 2017), mapping out a framework for the relation between mental health and Social Media discourse (Mendu et al. 2020) and analyzing a political party's internal discourse (Manucci and Amsler 2017). This approach is not free of pitfalls, however, with Brookes and McEnery calling it a »very naive model of a text« (Brookes and McEnery 2019, 5) and offering a scathing critique of the approach. The original authors of the LDA-model admit that the basic model (and thus most commonly the implementations) in the paper is based on the »bag-of-words« assumption (Blei, Ng, and Jordan 2003, 2–3). »Bags-of-words« disregard syntactic dependencies or word position. (Brookes and McEnery 2019, 7–8) criticize another disadvantage of this methodology: Since topics need to be chosen *a priori*, the associated topics can be inconclusive.

A different type of literature focuses on supervised learning. An earlier example of this method is the classification of Russian military discourse into binary categories by Stewart and Zhukov (2009), where a small hand-labelled data set, is employed to label the remainder of an unlabelled corpus. The annotated corpus is then used for quantitative analysis. Fraussen, Graham, and Halpin (2018) use a supervised learning approach to classify the prominence of political interest groups in speeches from the Australian parliament. As in the previous example, hand-labelled data was used to train an initial model that then infers labels for the whole corpus. Villegas, Mokaram, and Aletras (2021) use image data of political campaign advertisements to classify the political ideol-

ogy behind the message using a mixture of Natural Language Processing and Computer Vision. In all those cases, the problem was posed as a binary classification problem. An alternative approach leverages coding schemata for multi-class-classification, such as Bilbao-Jayo and Almeida (2018) for Spanish political campaign manifestos as well as recently Subramanian, Cohn, and Baldwin (2019) for political speech acts. Especially this last paper is in close alignment with this workproject, in that it recognizes the absence of well-defined training data in this domain (Subramanian, Cohn, and Baldwin 2019, 274) and constructs a labelling mechanism for specific modes of speech (*speech acts*).

Using speech act modelling can act as an effective relevance filter as not every speech act might be relevant to quantify political intentions. To then match relevant speech acts with parliamentary activities is a task that can be best characterized as an information retrieval problem. As parliamentary activities such as the introduction of bills, questions to the government or initiatives might not follow the discourse of a deputy literally, such search needs to be able to search beyond lexical identity and consider semantics, that is the *meaning* of a speech act. This is where classical, keyword- or token-based information retrieval systems might fail, as they focus on lexical similarity (or edit distance) alone. Using recent advances in *vector-space*-based information retrieval, this problem has become more accessible. Viewing both the search item (query) and the query result as vectors that inhabit the same vector space, this allows for an efficient representation of both the query and the retrieved items (Manning, Raghavan, and Schütze 2008, 113–114). Retrieving a fitting query result is thus turned into a Linear Algebra problem, where only vector similarity metrics need to be computed to then retrieve a matching result. Still, those similarity metrics (such as *cosine similarity*, which will be introduced later) can be computationally expensive, depending on the dimensionality of the data (ibid.). The methodology used here makes use of some shortcuts to reduce the overall computation time, such as k-Nearest-Neighbours.

One family of models that deliver State-of-the-Art-performance on these tasks are those that built

on the BERT *(Bi-directional Encoder Representations from Transformers)* (Devlin et al. 2018). BERT in turn is based on the *attention* mechanism (Vaswani et al. 2017), which allows a model to learn a context-sensitive representations of words. This is achieved using self-training, where a random proportion of an input sentence is »masked« and the model is then trained to predict the missing word (Vaswani et al. 2017, 1–2). However, as this approach is not specifically fine-tuned for multi-token situations, another layer would be needed to make this approach useful for longer documents, such as sentences or paragraphs. This is where an architecture called *sentence-transformers* (Reimers and Gurevych 2019) has shown great promise. Sentence-Transformers are trained on a Siamese network architecture, where the task is the computation of similarities between two sentence pairs. Leveraging this architecture allows for vector similarity searches that encode latent language features of sentences into a fixed-sized representation (Reimers and Gurevych 2019, 3982). An architectural overview can be found in Appendix E in Figure 8. *Sentence-Transformers* and related models are starting to become more prominent in the field, as demonstrated by the recent survey of Terechshenko et al. (2021), although the architecture is most often used for supervised learning, as in the case of Cheema, Hakimov, and Ewerth (2020) for retrieving political claims for fact-checking. Since informational retrieval methods needs to be integrated in a software system, applications of this type remain the exception. The combination of a speech-act based relevancy filter and a semantic similarity search engine for sentences and documents forms the basis for this undertaking of a domain-specific search engine for political text.

3 Data and Methods

3.1 Data Description and Processing

To better understand the interplay between political utterances in a forum like parliament and political actions taken, a special data set was constructed specifically for this purpose. It consists of two elements: (a) Speeches made before the *Assembleia da República* and (b) descriptions of parliamentary actions taken by deputies. The speech data used for this analysis was gathered from the

portal *Debates Parlamentares*, which contains historical session notes and transcripts from the 1st Portuguese Republic onward. For reason of topicality, only speeches and parliament activities from the legislature XI (starting in 2009) of the 3rd Republic to the present were considered. Since the data available at that source was not extractable through means such as an API (*Application Programming Interface*), Webscraping methods were used to parse the HTML containing the speeches, names and the date for a given parliamentary session. Parliamentary actions were extracted from the *Open Data-Portal* of the Portuguese parliament. This data source made available an `XML`-file containing the activity of deputies for a given legislative period. As such, the file needed to be parsed as well, yielding the activity date, the name of the deputy and a description and type of activity. Reasonable effort has been undertaken to clean the data in a way that seeks to minimize misrepresentations of speech, such as unexpected or systematic omission of deputies. The data used for this workproject was collected until the 29th of August 2021, meaning that speeches or parliamentary activities from after that date were not considered. The complete data set consists of 213,559 speeches and 259,748 parliamentary activities. Note that speeches can and usually do include more than one sentence. Some modifications were made to the data to accommodate the automated analysis. First, only sentences with a token length of more than 5 were considered. This was necessary, as the data contained a not insignificant share of short interjections as they are usual in a parliamentary setting. Additionally, the discourse of an individual deputy was stitched together, since interjections disrupt discourse that should be treated as a singular semantic unit. Within the discourse document thus derived, sentences containing fewer than 8 tokens within a speech string were removed. Reviewing several options for this removal procedure, a sentence length of less than 7 was not deemed meaningful enough for an automated content analysis. The individual documents, as delimited by the combination of the speaker name and the date when the speech was given, where exported in the `JSON`-format. Variables, including type and a short explanation are given in Appendix A.

3.2 Methodology

The methodology of this analysis combines a semi-supervised labelling process with a semantic search engine. As will be explained under *Labelling definitions and Operationalization*, only particular modes of speech are considered useful for the scope of this project. Figure 1 offers a high-level overview of the system, whose individual parts will be explained in subsequent sections. One of the decisive challenges was that the data set containing deputies' speeches in parliament is completely unlabelled, which calls for a method to apply labels to observations that are useful for the analysis. A self-training data set, which relies on a mixture of labelled and unlabelled data, is a potential solution to filter the data accordingly.

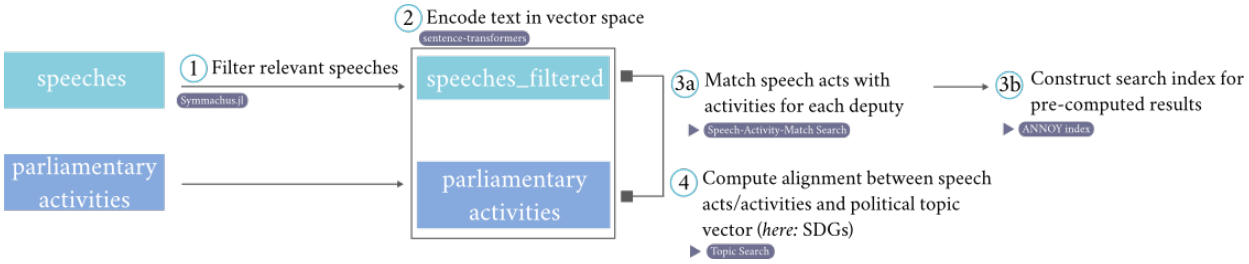


Figure 1: High-level system overview

The filtered data points are then used to query the performative data set, i.e. for each speech act, they retrieve matching parliamentary activities. Using *sentence-transformers* as the fundamental building block of the matching mechanism allows for richer results, as the search no longer relies solely on keywords. Two applications of that search engine will be discussed in section *System Demonstration and Interpretations*: (1) Speech-Activity-Match Search and (2) Topic Search. For each of the three main components – the labelling process, name and topic search – result verification steps are performed and reported in Appendix C, Appendix I and Appendix J.

3.2.1 Labelling definitions and Operationalization

As a preprocessing step for the semantic search, relevant parts of political speches need to be extracted. The self-training data set uses particular units of political discourse, with the target being sentences that contain a phrase structure conforming to the linguistic idea of *deontic modality* on

a normative level or that of *proposition*. The first of these is in need of both explanation and substantiation. *Deontic Modality* is traditionally summarized as »having to do with what is morally or legally obligatory and permissible« (Charlow and Chrisman 2016, 1). For the scope of this analysis, this traditional paraphrase needs to be expanded, as parliamentary discourse is much richer than those two categories allow for. Hence, Nuyts (2016, 36) broader definition of this type of speech being an »indication of the degree of moral desirability of the state of affairs« fits well with the nature of parliamentary discourse, as various ideologies clash in debate. Given the plurality of political parties in the Portuguese parliament, it makes sense to include the notions of »performativity« and »descriptivity« (Nuyts 2016, 46–47), which indicate whether a speaker is making use of the modal category themself or reporting that of another speaker or entity. Given that deputies' discourses might reflect or reiterate a party's position, this extension is necessary. Table 1 contains an overview of some of the expressions used to label the data set in a binary fashion: `1` indicating the presence of a modality manifestation (deontic or proposition). Conversely, `0` indicates the absence of one such manifestation. Example sentences for both label types can be found in Appendix B. Importantly, the modality requirement can be fulfilled without a literal match by adhering to the broad semantic direction. Though a binary labelling process simplifies model training, it has been difficult in some cases to actually decide on the label for the manually labelled data set.

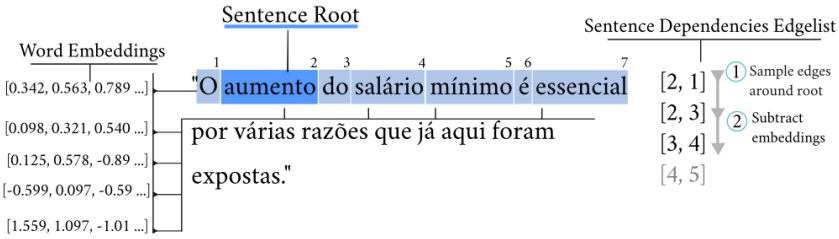
Example Expression	Modality Type	Example
É necessário . . .	Deontic	DM1
Urge/ É urgente . . .	Deontic	DM2
A nossa prioridade	Deontic	DM3
É essencial. . .	Deontic	DM4
. . . deve ser combatido/a	Deontic	DM5
Deveria . . .	Deontic	DM6
. . . importa . . .	Deontic	DM7
Estamos abertos . . .	Deontic	DM8
É preciso/Precisamos . . .	Deontic	DM9
. . . impõe-se	Deontic	DM10
Propomos. . .	Proposition	PROP1

Table 1: Expressions used for manual labelling process (positive class)

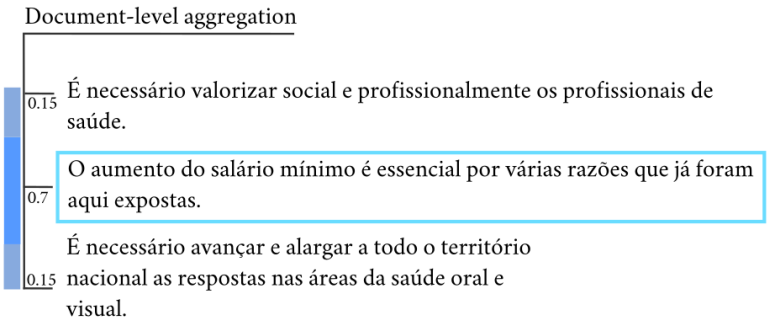
Since the original data set is unlabelled, it is difficult to discern whether the »positive class« (i.e. relevant speech acts) is the minority class or not. If this was the case, special precautions would need to be taken. In case the data was unbalanced enough, the model would incur a bias based on standard accuracy measurements (see Fernández et al. 2018, 21). The manual labelling process was thus focused on creating a balanced data set in terms of label frequency to help guide the model training process to preclude problems downstream because of imbalanced data labels. In total, 1000 instances of the data set were labelled and are available on GitHub.

3.2.2 Embedding Speech Acts: The Case for Symmachus

As raw text makes for dubious input features for a Machine Learning algorithm, a process to transform the unedited speeches into a useful representation for the task at hand is necessary. This workproject has implemented one such mechanism, named »Symmachus«. The general idea behind *Symmachus* is that of selective »attention« centered around the syntactic *root* or *head* of a sentence. According to Jurafsky and Martin (2021, 248), the root of a sentence is the word that bears the highest grammatical importance. During the process of identifying the sentence root, a dependency tree for the entire sentence is constructed. The conjecture here is that the words in the vicinity of the sentence root are those that contribute most to the overall semantic content of a sentence. Instead of considering every word in the sentence, a compression of semantic content is desirable. Implementation of this mechanism is achieved using the *Stanza* Natural Language Processing library (Qi et al. 2020) for `Python` (Van Rossum and Drake Jr 1995). The adjacency matrix of the sentence dependency tree is used to sample from the edges of the graph, i.e. the dependency relationships between words. Sampling is controlled by a scaled β-Distribution to reflect the position of the sentence root within the sentence as well as the length dimension. The sampling process generates a bi-directional context by aggregating information from tokens before and after the sampled token. Finally, for the samples drawn, *FastText* word embeddings (Bojanowski et al. 2017) are retrieved and a »semantic walk« put in place by subtracting the child node embedding from the respective head embedding (see Figure 2a). The resulting vectors are then averaged.



(a) *Symmachus* Sentence embedding procedure



(b) *Symmachus* Document embedding aggregation procedure

Figure 2: Sentence-level and document-level Word Embedding aggregation

This procedure is repeated for every sentence in the document. The second step is visualized in Figure 2b, where a given sentence's embedding derived from the previous step is aggregated with neighbouring sentences' embedding. Here, the influence of neighbours can be controlled through the context size and the weight given to the sentence in question. This two-step procedure then yields sentence embeddings, which form the feature set for the labelling machine.

3.2.3 Gradient Boosting for Label Prediction

The *Symmachus* procedure transforms the unstructured text inputs in a way that embeds them in a tabular structure. Well-suited for this is a Machine Learning method called »Gradient Boosting«. *Gradient Boosting Machines* are one of the hallmark methods in the area of supervised learning and were created by Friedman (2002). The following outline of these machines aligns with Friedman's notation. Building on the idea of *Decision Trees*, Gradient Boosting is a process whereby a *base learner* (usually a decision tree with a flat structure) is improved sequentially by furnishing it with

new learners (Friedman 2002, 367–368). Note that the following elucidation considers the case of regression, while for classification — though structurally similar — a different loss function has to be chosen to quantify how distant the predicted label is from the true label. A choice for this would be binary cross-entropy $-\sum_{i=1}^{n}\sum_{m=1}^{2} y_i \, log(f(x_i))$ (notation adapted from James et al. (2021, 410)), with n being the number of observations, i an index for the true label y and the observation x and m the number of classes.

$$f_0(\mathbf{X}_{1t}, \ldots, \mathbf{X}_{kt}) = \arg\min_{\gamma} \sum_{t-1}^{t} C(Y_t, \gamma) \tag{1}$$

The algorithm starts with an estimate, provided by the function $f_0(\mathbf{X})$. Next, the pseudo-residuals are computed by taking the derivative of the Cost function C with respect to the function f.

$$\tilde{y}_{t,m} = -\Big[\frac{\partial C(Y_t, f(X_{1t}, \ldots, X_{1k}))}{\partial f(X_{1t}, \ldots, X_{kt})}\Big]_{f(\mathbf{X})=f_{m-1}(\mathbf{X})} \tag{2}$$

The next steps is to add another learner to the chain, which is fit on the pseudo-residuals. This new learner emits a constant, for each of the *regions* of the partitioned feature space. This constant with L terminal nodes γ_{lm} of the first iteration is then appended to $f_0(\mathbf{X})$.

$$h(\mathbf{X}; \{R_{l,m}\}_1^L) = \sum_{l-1}^{L} \bar{y}_{lm} \mathbf{I}(\mathbf{X} \in R_{l,m}) \tag{3}$$

$$\gamma_{lm} = \arg\min_{\gamma} \sum_{\mathbf{X} \in R_{l,m}} C(Y_i, f_{m-1}(\mathbf{X}) + \gamma) \tag{4}$$

The speed, with which this »update« of the learned function is performed, can be controlled through a hyperparameter $0 < \eta \leq 1$. Every step after the first iteration is thus generating an increasingly specialized chain of trees to improve upon the errors of the previous step.

$$f_m(\mathbf{X}) = f_{m-1}(\mathbf{X}) + \eta \cdot \gamma_{lm} \tag{5}$$

As the implementation of the Gradient Boost Model for this workproject, the highly optimized library `XGBoost` (Chen and Guestrin 2016) has been chosen.

3.2.4 Self-training Labelling Machine

The manually labelled data set is used to train a *semi-supervised* model. This model then randomly samples from the remaining (vastly larger) unlabeled corpus until some stopping criterion is reached. For the remainder of the corpus the model infers labels. These *Pseudo-Labels* are then taken to be the »true« label of the observation. Following (Lee 2013, 3), the inferred label is the target class that has the highest predicted probability. The highest confidence label predictions are then appended to the training corpus. Repeating this cycle for several iterations, a data set is constructed, with hyperparameter validation steps during each iteration. This serves to respond to potential data drift caused by the sampling process. Instead of using a rigid, unchanging grid for the training process, only the first iteration uses a fixed set of hyperparameters. After this, a generic evolutionary-type strategy is implemented, whereby the best hyperparameter combination of each iteration is allowed to drift (»mutate«) within predefined margins to inject extra randomness into the hyperparameter search process. This broadly follows the strategy in Liu et al. (2006), however, only mutation is used for generating a new offspring grid. Once the predictive quality is deemed sufficient, the model is then »broadcast« to the entire remaining corpus; a process which can be characterized as a bulk-labeling operation. The routines of this machine were implemented using the `Julia` programming language (Bezanson et al. 2017).

Let said approach be succinctly defined by the following algorithm (based on (Zhu and Goldberg 2009, 15–16)):

Input: Labeled data $\{ (\mathbf{x_i}, y_i) \}_{i=1}^{l}$, Unlabeled data $\{ \mathbf{x_j}_{j=l+1}^{l+u} \}$

1. Set $\mathbf{L} = \{(\mathbf{x_i}, y_i)\}_{i=1}^{l}$. Set $\mathbf{U} = \{\mathbf{x_j}_{j=l+1}^{l+u}\}$
2. **Do until stop**:
    1. Train a function $f(\mathbf{L})$.
    2. Use $f(\mathbf{L})$ to predict on a sample $\mathbf{x_j}_{j+k}^{l+m}$ from $\mathbf{U}$.

3. Remove a subset **S** from **U**, then append $\{(\mathbf{x}, f(\mathbf{x})) \mid \mathbf{x} \in S\}$ to **L**.

3. Return model specifications.

4. Train a function $f(\mathbf{L})$ and apply to the remaining labels in **U**

The subset removed from the unlabelled data is deterministic, in the sense that it always keeps the same size. In each iteration, the model predictions are ordered by their predictive value, which is a scoring metric derived from the *Confusion Matrix* of the training process. Since the evaluation of political speech is a sensitive matter, it is advisable to choose a metric that makes the model pursue a cautious approach. To this end, the *F1-Score* seems a sensible choice. This metric seeks to strike a middle ground between detecting as many implicit policy intentions (through *recall*) and ascertaining that those observations classified as positive are actually positive (measured via *precision*). It is computed by:

$$\text{f1-score} = 2 \times \frac{(precision \times recall)}{(precision + recall)} \tag{6}$$

3.2.5 Vector-based Similarity Search Engine

The basic mechanism of the similarity search is shown in Figure 3. Every encoded query (i.e. sentence of a speech) of a political actor is queried against an encoded activity of that same actor. Note that speeches and legislative activities are embedded in the same latent vector space. Then, the most similar activities are retrieved by applying a k-Nearest-Neighbours model. Internally, the partitioning for the clustering model uses the *Euclidean* distance ($d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$); a metric that simply measures distance between two vectors in space. To then transform the results into a measure of similarity, the cosine similarity ($cos\_sim(x,y) = \frac{\sum_{i=1}^{n} x_i\, y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\, \sqrt{\sum_{i=1}^{n} y_i^2}}$) is computed. The model returns both the index of the activity and the similarity metric for every political actor's corpus. To facilitate access to pre-computed results for the speech-activity-match search, a search index based on ANNOY *(Approximate Nearest Neighbors Oh Yeah)* (Bernhardsson 2016) is built.

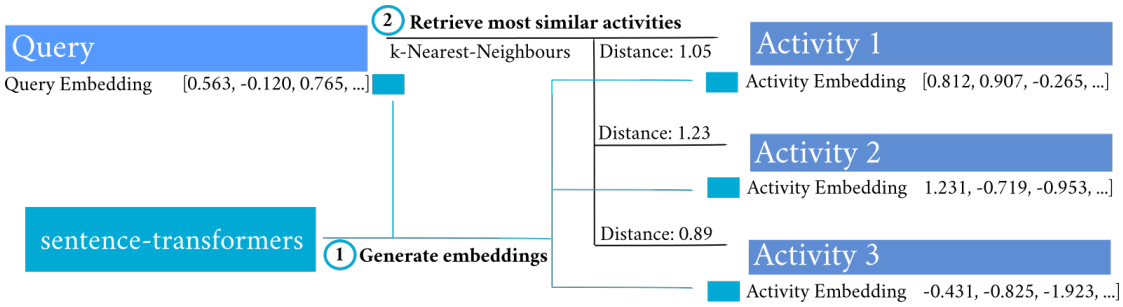


Figure 3: Basic search mechanism

As there are multiple sentence-transformer models for the Portuguese language, three of them will be evaluated here: *bert-base-portuguese-cased*, published in Souza, Nogueira, and Lotufo (2020), and *bert-portuguese-cased-nli-assin-assin-2* are specifically trained on the Portuguese language, while the other (*distiluse-base-multilingual-cased-v1*) is a multi-lingual model. With the main goal of the similarity search engine being the computation of aligned vector space, two evaluation metrics come naturally: (a) Average vector similarities over a political actor's individual corpus and (b) Average number of phrase matches over a political actor's individual corpus. The former is derived by first computing the k-Nearest-Neighbours for each speech in a political actor's individual corpus and then computing the cosine similarity for each of the k neighbours. Adding the cosine similarity step is necessary, as transformer models might not (and in this case do not) encode the same vector space. Vector spaces with higher dimensionality would naturally have higher distance measurements, which invalidates any comparison. Phrase matches are computed based on token collocations (i.e. *phrases*) of the entire corpus. More specifically, the collocation algorithm is *Normalized Pointwise Mutual Information* (Bouma 2009) in the Gensim library implementation (Rehurek and Sojk 2010). Taken together, they form the aggregate evaluation metric. Keyword matches alone quantify insufficiently the quality of a query result as the metric emphasizes literal matches only. This is why the similarity score, which captures more general, abstract similarity as well, will be weighted with $\frac{2}{3}$ and the keyword matches with $\frac{1}{3}$.

The second feature of the search engine allows for topic searches. Considering that topics are collections of sentences, a similar embedding approach can be used here, especially if the topic

descriptors are not too long. For a given set of topics, topic embeddings will be computed using domain-specific text. The output of this is a fixed-length representation of the topic in vector space. Then, this representation is broadcast to every political actor in the data set. As the distribution of topics might be very different for every political actor, only the k most similar speech sentences and activities to the topic representation will be retrieved. For these data points, two alignment metrics are computed. Figure 4 illustrates the procedure.

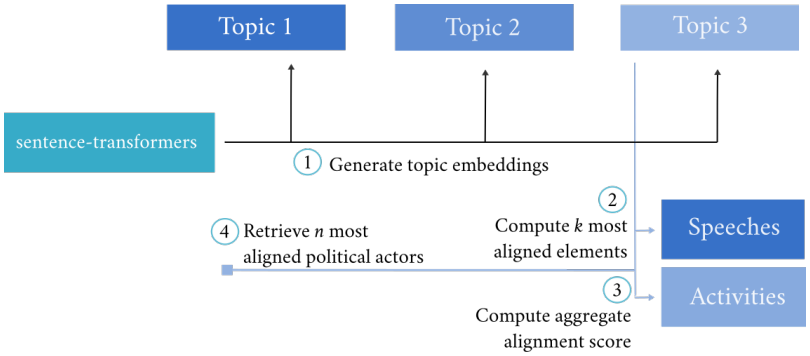


Figure 4: Topic Search mechanism

The first (internal alignment) serves to indicate whether the semantic space of relevant speeches is similar to that of the deputy's parliamentary activities. External alignment measures how aligned the activity and speech vector spaces are to the topic vector space. Finally, the n most aligned politicians are returned as the query result based on the aggregation of the two alignment metrics. Unlike the speech-activity-match search, the topic search does not evaluate phrases. For that to perform reliably, a much larger corpus would be necessary for the topics, which are mostly short paragraphs. For the evaluation of the model best suited for topic search, aggregate scores over all topic embeddings and a random sample of political actors are used.

4 Results

4.1 Labelling Machine

The model evaluation process is split into two parts: While the first concerns the self-training of the labelling process, the second concerns the evaluation of the semantic search to establish the

relationship between speech acts and activities. As the hyperparameters concerning the embedding process might be of larger importance, the corresponding parameter overview of the boosting model were relegated to Appendix D. Figure 5 shows the performance of the labelling process as described above. The F1-Score on the y-Axis is the score derived from predicting on a test set during each iteration. The proportion of training to testing data was set at 80%-20%. Overall, the labelling machine achieves high quality. As the hyperparameter search process is based on an evolutionary design and as models are re-trained in each run, the only information models receive from their parents are the »fittest« hyperparameters. Due to the computationally expensive nature of the labelling process, the experiment was only conducted once. Importantly, the figure should not mislead towards representing any accuracy-type metric. The F1-score is essentially retrieval-focused, which means that despite the high score, thousands of instances will still be incorrectly classified because of the data set's size.

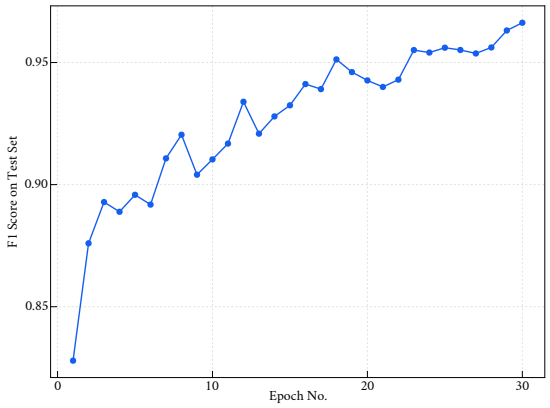
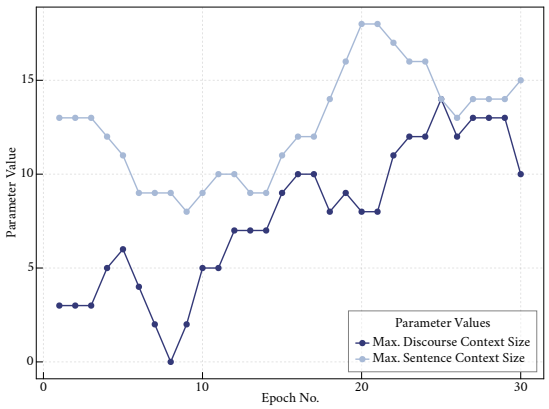
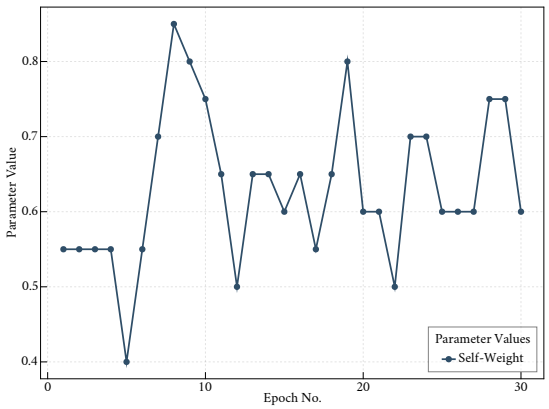


Figure 5: Performance of Labelling Process (F1) during training on test set

As for the three main hyperparameters of the *Symmachus* embedding approach, Figure 6a and Figure 6b show the evolution of the parameters throughout the labelling process. While the self-weight parameter behaves in an almost stationary way as it oscillates around the 0.6 mark, the context parameters show a clear upward tendency. Coincidentally, this suggests that in order to classify the given sentence, the model weighs the context of the sentence almost as much as the sentence itself. A random sample of inferred labels with commentary can be found in Appendix C.



(a) Changes in Context Hyperparameters



(b) Changes in Self-Weight Hyperparameter

Figure 6: Parameter Changes

The discourse context (i.e. sentences before and after the one in question) starts at around 4 sentences in both directions but evolves to 10; more than double the context. During one epoch, there is a somewhat anomalous context size of zero, which is quickly revised upward in the following epochs. A similar trend can be observed for the sentence context, though the hyperparameter development is less significant. The labelling machine starts out with a maximum of 13 tokens considered in both directions and reaches a high of 18 tokens during iteration 20. Especially the context parameters demonstrate that in order for the labelling machine to use its full capacity, the required context is relatively large, which detrimentally impacts the computational performance.

4.2 Similarity Search Engine

The functionalities of the similarity search engine were evaluated using the methodology described in the previous section. Results per model are summarized in Table 2. *A priori*, it would be reasonable to assume that language-specific models would perform better than multi-lingual ones on a single-language data set. The results support this: While the multi-lingual model achieves much higher scores on average keyword matches, the textual similarity metric is far below that of model specialized on the Portuguese language. Both Portuguese models are very close in terms of their metrics, however the `BERTimbau` model performs slightly better on keyword matches.

Avg. Keyword Matches	Avg. Similarity	Aggregate Score
`Bert-base-portuguese-cased (BERTimbau)`		
0.08970	0.72870	0.51570
`Bert-portuguese-cased-nli-assin-assin-2`		
0.07362	0.72913	0.51063
`Distiluse-base-multilingual-cased-v1`		
0.19187	0.16432	0.17351

Table 2: Speech-Activity-Match Search: Results for *sentence-transformer* models

To inspect the functional performance of the speech-activity-match search, the distribution of the number of activities and speech acts against the mean relevance of each political actor was plotted in Figure 9 in Appendix G. For reasons of visibility, the x-Axis with the number of activities and speech acts, respectively, has been log-transformed. The number of speech acts does not seem to correlate strongly with the mean relevance. This is to be expected, as the retrieval is geared towards activities. The Pearson Correlation Coefficient is 0.1643 for this relationship. On the other hand, the correlation between the number of activities and relevance is much higher, at 0.3273; a moderately strong, positive relationship. This shows that the more activities a deputy has in their corpus, the more specific — and thus relevant – results can be retrieved by the search engine, which indicates that it is working as expected.

Evaluating the topic search required query documents representing political issues. Because they circumscribe contemporary issues well and concisely, the descriptors of the *Sustainable Development Goals (SDGs)* (pt. Objetivos de Desenvolvimento Sustentável) as formulated by the United Nations were chosen. They cover relevant policy areas, such as climate change mitigation, energy policy and the combat against poverty. However, *any* topic could be chosen for usage with the

Avg. Inner Alignment	Avg. External Alignment Speech Acts	Avg. External Alignment Activities	Aggregate External Alignment
`Bert-portuguese-cased-nli-assin-assin-2`			
0.90893	0.86023	0.80347	0.83185
`Bert-base-portuguese-cased (BERTimbau)`			
0.90425	0.83800	0.78364	0.8108
`Distiluse-base-multilingual-cased-v1`			
0.24917	0.3680	0.3103	0.33918

Table 3: Topic Search: Results for *sentence-transformer* models

semantic search. Each goal was first embedded using the three models to be evaluated. Then, the topic search procedure was performed. That is, for every policy area defined by the SDGs, alignment metrics per political actors were derived. The results are summarized in Table 3. Similar to the speech-activity-match search, the language-specific models outperform the multi-lingual one for the topic search in every metric category; and quite dramatically so. The benefits of utilizing a language-specific models become quite apparent.

4.3 System Demonstration and Interpretations

This section serves to demonstrate the two search functions that have been described above as well as to illustrate potential shortcomings and fields of further research. The speech-activity-match search serves as a »scanner« to search an entire corpus of speeches for a single politician. Given a certain, relevant speech act, the activities database is queried. As initially set out as an objective, this can be used for e.g. validating politician's pledges based on stated explicit or implicit intentions.

Table 4 contains two »ideal« examples, where the search retrieves not only relevant results, but also

Speech Act (Query)	**Most relevant result**
Presidente, e Deputados, cerca de 40 000 animais são retirados das ruas e recolhidos em centros de recolha oficial (CRO) de animais, sejam eles municipais, sejam eles intermunicipais.	Avaliação da aplicação da Lei nº 69/2014, de 29 de agosto, sobre a criminalização de maus tratos a animais, proteção aos animais e alargamento dos direitos das associações zoófilas.
Name: Mariana Silva (PEV) — Result Relevance: 0.75159	
No combate à desinformação, é igualmente essencial continuar a reconhecer a existência de uma imprensa livre e independente, incómoda para os poderosos e exigente no escrutínio de quem decide.	Pela não atribuição de subsídios públicos aos órgãos de comunicação social, mantendo a Imprensa independente e como contrapoder ao Estado.
Name: João Cotrim de Figueiredo (IL) — Result Relevance: 0.76864	

Table 4: Information Retrieval (Speech-Activity-Match Search) Demonstration I

results that are not strictly lexically, but semantically similar to the speech act. One issue identified with the speech-activity-match search »scanner« approach is that there might be either no relevant activities for a deputy, or the speech act, upon which the scan is based, bears an inexact label. In this case, the search engine still retrieves results based on vector space similarity, but those result might prove to be practically irrelevant to a potential user. A diverse range of examples can be found in Appendix I. To dampen the impact of this, a sensitivity metric was added to the engine that returns results for a query only, if the relevance (i.e. the cosine similarity) of the best result candidate lies above a certain threshold. For the opposite case, where the speech act itself does not actually belong to the initially defined speech act category, the relevance filter needs to be refined.

The semantic search enables the computation of aggregate alignment scores for specific topics. As such, Table 5 shows the most aligned (i.e. sorted by the mean of speech act alignment and activity alignment) political actors for SDGs numbers 1 to 7. Results for SDGs 8-17 are shown in Appendix H in Table 10. The tabulations make sense ideologically, as deputies of parties from the

SDG No.	Most aligned political actors
1 – Erradicar a pobreza	Sandra Cunha (BE) - 0.8611; José Manuel Pureza (BE) - 0.8595 Isabel Pires (BE) - 0.8566; Carlos Matias (BE) - 0.8566
2 – Erradicar a fome	Paula Santos (PCP) - 0.8676; José Manuel Pureza (BE) - 0.8648 Carlos Matias (BE) - 0.8640; Bruno Dias (PCP) - 0.8639
3 – Saúde de Qualidade	José Manuel Pureza - 0.8694; Ricardo Vicente (BE) - 0.8691 André Silva (PAN) - 0.8664; José Moura Soeiro - 0.8639
4 – Educação de Qualidade	José Manuel Pureza (BE) - 0.8400; Rita Rato (PCP) - 0.8374 André Silva (PAN) - 0.8370; Moises Ferreira (BE) - 0.8348
5 - Igualdade de género	Sandra Cunha (BE) - 0.8528; José Manuel Pureza (BE) - 0.8515 Pedro Filipe Soares - 0.8501; Catarina Martins (BE) - 0.8486
6 - Água potável e Saneamento	José Manuel Pureza (BE) - 0.8442; Fabíola Cardoso (BE) - 0.8438 Ricardo Vicente (BE) - 0.8427; Carlos Matias (BE) - 0.8393
7 - Energias renováveis e accesíveis	Luís Leite Ramos (PSD) - 0.8553; João Dias (PCP) - 0.8536 Carlos Matias (BE) - 0.8509; Fabíola Cardoso - 0.8505

Table 5: Topic Search: Most aligned political actors per SDG I

left-wing of the spectrum, such as PCP (*Partido Communista Portuguesa*), BE (*Bloco de Esquerda*) and PAN (*Pessoas-Animais-Natureza*) are by virtue of their ideology more aligned with topics such as gender equality than centrist or right-wing parties. However, there are a few anomalies, such as SDG No. 13 (Ação Climática), where just one of the four most aligned political actors is from the left and none from an »ecological« party. A manual result verification of randomly sampled political actors is performed in Appendix J.

Table 6 represents a global summary of mean alignment with SDG descriptors since 2009. The Sustainable Development Goals have only existed since 2015, the topics they encode have certainly not come into existence from that date only. It is apparent that parties of the opposition, especially those on the left wing of the political spectrum are more aligned with SDGs on a global basis, whereas

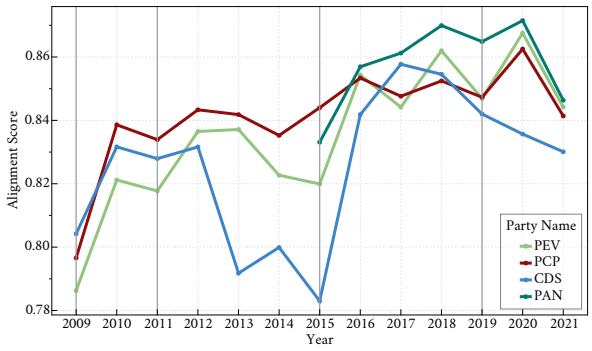
Party Name	SDG Aggregate Alignment Score	Speech Act Alignment Score	Activity Alignment Score
PAN (*Pessoas-Animais-Natureza*)	0.8577	0.8909	0.8245
PCP (*Partido Comunista Portuguesa*)	0.8414	0.8641	0.8187
PEV (*Partido Ecologista »Os Verdes«*)	0.8354	0.8735	0.7973
BE (*Bloco de Esquerda*)	0.8404	0.8510	0.8299
CDS (*Centro Democrático e Social*)	0.8255	0.8487	0.8024
IL* (*Iniciativa Liberal*)	0.8148	0.8645	0.7651
Chega*	0.8111	0.8381	0.7842
Independent Deputies*	0.7958	0.8419	0.7497
PSD** (*Partido Social-Democrata*)	0.7828	0.8218	0.7437
PS** (*Partido Socialista*)	0.7762	0.8286	0.7238

Note: * These parties have only one (*IL* and *Chega*) deputy and two delegates (*Independents*), respectively. Results are therefore most likely dubious.
Note: ** Governing parties between 2009 and 2021. Their scores are most likely underestimated by the model, as the activities of the government likely do not appear in the individual deputies' records who belong to that party. This would have to be accounted for separately.

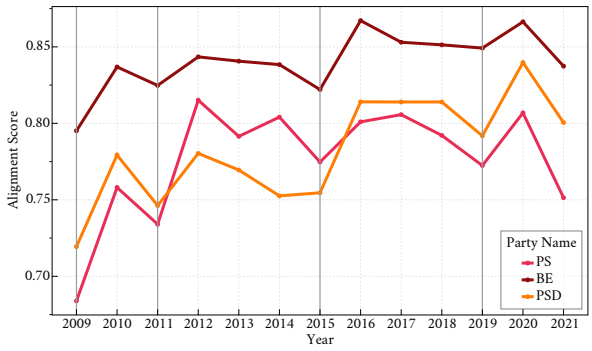
Table 6: Topic Search: Global Alignment Scores

the »governing parties« of PS and PSD have the lowest global scores. This last result is most likely tied to the structure of the data, as explained in the table notes of Table 6. Another peculiarity is the observation that for every party, the speech acts are always more aligned with the SDG topic vectors than their activities. Whether this is an indication of the infamous »discrepancy« between political actor's actions and their speeches, requires a case investigation that is out of scope of this workproject. Even though these findings might be tempting, caveats should be considered: Scores are not adjusted, neither is delegation size in parliament (i.e. a larger party might have a more »dissonant« voice because of a higher number of deputies) or information about whether the party was in power at the time. Another issue seems to be a compression of the similarity measurement space: The entire political party spectrum spans a range of 0.08, which could be a symptom of a lack of fine-tuning on a domain-specific text corpus and would need to be re-calibrated.

A final demonstration concerns the global alignment with SDG descriptors over time, summarized by party in Figure 7. Parties with only one delegate were excluded from the plots. Similar to the per-topic alignment, left-wing as well as »ecological« parties in Portugal seem to be much



(a) Party SDG alignment over time I



(b) Party SDG alignment over time II

Note: Grey vertical lines correspond to electoral cycles

Figure 7: Party alignment with SDG descriptors over time

more aligned with SDG topic descriptors, a trend that has manifested itself seemingly from the 2011 elections onward. For brief periods between 2016 and 2017, conservative parties temporarily scored higher than left-wing parties. As with the global alignment score, the »big-tent« parties of PSD and PS have the lowest scores of all parties. Except for the case of CDS after the election in 2019, every party's alignment score jumps after an electoral cycle and shows an overall upward trend. The information retrieval methods implemented seems to value specificity over generality, which is in line with the search results presented: Smaller, more focused parties appear at the top of the results.

5 Discussion

5.1 Theoretical and Practical Contributions

The objectives of this workproject have been two-fold: One has been to demonstrate that advances in Natural Language Processing have given social scientists a new way to look at political text by being able to explicitly compute alignment of political actors with well-described political issues as well as with their own activities in parliament; a technique that goes well beyond traditional corpus statistics. Not only this offer potential time-savings, as the system reduces the need to crawl political text manually, but also does it create new metrics like topic or speech-action alignment to

quantify and evaluate the performance of political actors. The next step from there was to create prototype of one such system that could be added to the methodological toolbox. Transfers of this methodology to other genres of political text, such as party manifestos or even multi-modal analysis of video data, are plausible. Combining the alignment approach with search in a potential public system serves to improve the citizenry's awareness of their elected leaders actions and hands them a powerful tool to strength accountability, as they can freely explore a political actor's record. Ideally, they can make a more informed electoral choice based on the computation of the search tool. This has already precedent in other countries, such as in Brazil, where the site *politicos.org.br* (manually) »ranks« politicians according to their actions in Congress.

5.2 Limitations

Significant limitations of the system remain: Since political text is a delicate data type, an enthusiastic, yet too uncritical usage of the models and algorithms presented, is inadvisable and further investigations have to be conducted, especially regarding (ideological?) bias in Natural-Language-Processing-based information retrieval methods (such as in the excellent meta study by Weidinger et al. (2021)). Furthermore, it would be important to fine-tune the models used to the kind of text that they are used for, that is political text. Moreover, some models (such as *BERTimbau*) are trained on *Brazilian* Portuguese, which might exhibit different linguistic subtleties than those considered here for European Portuguese. The preprocessing step could benefit from new labelling methods such as Data Programming (Ratner et al. 2016) or learning from possibly noisy labels (Jiang et al. 2018). The search also suffers from domain knowledge agnosticism: In its current state, the system has no knowledge of whether a political actor is a deputy with or without additional offices or even a party leader, which might influence the weight of their speeches and activities. To enable the proper assessment of the alignment of governing parties, the government's legislative track record needs to be included in the data set. Following these improvement proposals would result in findings that both more robust and comprehensive. Overall, the methods trialled form a solid foundation upon which to build future investigations.

Bibliography

Baldi, Pierre and Chauvin, Ives (1993). "Neural Networks for Fingerprint Recognition". *Neural Computation* 5, pp. 402–418.

Bernhardsson, Erik (2016). *Annoy (Approximate Nearest Neighbors Oh Yeah)*. Available at: `https://github.com/spotify/annoy`.

Bezanson, Jeff et al. (2017). "Julia: A fresh approach to numerical computing". *SIAM Review* 59, pp. 65–98.

Bilbao-Jayo, Aritz and Almeida, Aitor (2018). "Political discourse classification in social networks using context sensitive convolutional neural networks". Association for Computational Linguistics.

Blei, David M., Ng, Andrew Y., and Jordan, Michael I. (2003). "Latent Dirichlet Allocation". *Journal of Machine Learning Research* 3, pp. 993–1022.

Bojanowski, Piotr et al. (2017). "Enriching Word Vectors with Subword Information". *Transactions of the Association for Computational Linguistics* 5, pp. 135–146.

Bouma, Gerlof (2009). "Normalized (pointwise) mutual information in collocation extraction." *From Form to Meaning. Processing Texts Automatically. Proceedings of the Biennial GSCL Conference*. Ed. by Christian Chiarcos, Richard Eckhart de Castilho, and Manfred Stede. Tübingen: Gunter Narr Verlag, pp. 31–40.

Bromley, Jane et al. (1993). "Signature Verification using a SiameseTime Delay Neural Network". *NIPS'93: Proceedings of the 6th International Conference on Neural Information Processing Systems*. NIPS, pp. 737–744.

Brookes, Gavin and McEnery, Tony (2019). "The Utility of Topic Modeling for Discourse Studies: A critical evaluation". *Discourse Studies* 21, pp. 3–21.

Charlow, Nate and Chrisman, Matthew, eds. (2016). *Deontic Modality*. 1st ed. Oxford: Oxford University Press.

Cheema, Gullal S., Hakimov, Sherzod, and Ewerth, Ralph (July 2020). "Check_square at CheckThat! 2020: Claim Detection in Social Media via Fusion of Transformer and Syntactic Features". arXiv: `2007.10534 [cs.CL]`.

Chen, Tianqi and Guestrin, Carlos (2016). "XGBoost: A Scalable Tree Boosting System". *Proceedings of the 22nd ACM SIGKDD InternationalConference on Knowledge Discovery and Data Mining*. Ed. by ACM. ACM, pp. 785–794.

Devlin, Jacob et al. (Oct. 2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". arXiv: `1810.04805 [cs.CL]`.

Fernández, Alberto et al. (2018). *Learning From Imbalanced Datasets*. 1st ed. Cham: Springer Nature.

Fraussen, Bert, Graham, Timothy, and Halpin, Darren R. (2018). "Assessing the prominence of interest groups in parliament: a supervised machine learning approach". *The Journal of Legislative Studies* 24, pp. 450–474.

Friedman, Jerome H. (2002). "Stochastic Gradient Boosting". *Computational Statistics & Data Analysis* 38, pp. 367–378.

Greene, Derek and Cross, James P. (Jan. 2017). "Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach". *Society for Political Methodology* 25, pp. 77–94.

Grimmer, Justin and Brandon, Stewart M. (2013). "Text As Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts". *Political Analysis*, pp. 1–31.

James, Gareth et al. (2021). *An Introduction to Statistical Learning. With Applications in R*. 2nd ed. New York: Springer.

Jiang, Lu et al. (2018). "MentorNet: Learning Data-Driven Curriculumfor Very Deep Neural Networks on Corrupted Labels". *Proceedings of the 35th International Conference on Machine Learning*. Vol. 35. Stockholm: ICML.

Laver, Michael, Benoit, Kenneth, and Garry, John (May 2003). "Extracting Policy Positions from Political Texts Using Words as Data". *American Political Science Review* 97, pp. 311–331.

Lee, Dong-Hyun (2013). "Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks". *ICML 2013 Workshop : Challenges in Representation Learning*. ICML.

Liu, Ruiming et al. (2006). "Optimizing the Hyper-parameters for SVM by Combining Evolution Strategies with a Grid Search". *Intelligent Control and Automation: International Conference*

on Intelligent Computing, ICIC 2006 Kunming, China, August 16–19, 2006. Ed. by De-Shuang Huang, Kang Li, and George William Irwin. Berlin, Heidelberg: Springer, pp. 712–721.

Manning, Christopher D., Raghavan, Prabkhakar, and Schütze, Hinrich (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Manucci, Luca and Amsler, Michi (2017). "Where the wind blows: Five Star Movement's populism, direct democracy and ideological flexibility". *Italian Political Science Review* 48, pp. 109–132.

Mendu, Sajana et al. (2020). "A Framework for Understanding the Relationship between Social Media Discourse and Mental Health". *Proceedings of the ACM on Human-Computer Interactions* 4.

Nuyts, Jan (2016). "Analyses of the Modal Meanings". *The Oxford Handbook of Modality and Mood*. Ed. by Jan Nuyts and Johan Van Der Auwera. Oxford: Oxford University Press, pp. 31–49.

Qi, Peng et al. (Mar. 2020). "Stanza: A Python Natural Language Processing Toolkit for Many Human Languages". arXiv: `2003.07082 [cs.CL]`.

Ratner, Alexander et al. (2016). "Data Programming: Creating Large Training Sets, Quickly". *30th Conference on Neural Information Processing Systems (NIPS 2016)*. NeurIPS. Barcelona.

Rehurek, Radim and Sojk, Petr (2010). "Software Framework for Topic Modelling with Large Corpora". *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta, Malta: ELRA, pp. 45–50.

Reimers, Nils and Gurevych, Iryna (Nov. 2019). "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks". *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, pp. 3982–3992.

Souza, Fábio, Nogueira, Roberto, and Lotufo, Roberto (2020). "Brazilian Conference on Intelligent System (BRACIS 2020)". Ed. by Ricardo Cerri and Ronaldo C. Prati. Intelligent Systems. Cham: Springer. Chap. BERTimbau: Pretrained BERT Models for Brazilian Portuguese, pp. 403–417.

Stewart, Brandon M. and Zhukov, Juri M. (2009). "Use of force and civil – military relations in Russia:an automated content analysis". *Small Wars & Insurgencies* 20, pp. 319–343.

Subramanian, Shivashankar, Cohn, Trevor, and Baldwin, Timothy (2019). "Target Based Speech Act Classification in Political Campaign Text". *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics, pp. 273–282.

Terechshenko, Zhanna et al. (Oct. 2021). "A Comparison of Methods in Political Science Text Classification: Transfer Learning Language Models for Politics". *Available at SSRN*. DOI: htt p://dx.doi.org/10.2139/ssrn.3724644.

Van Rossum, Guido and Drake Jr, Fred L (1995). *Python tutorial*. Tech. rep. Amsterdam: Centrum voor Wiskunde en Informatica.

Vaswani, Ashish et al. (2017). "Attention is all you need". *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*. Vol. 31. NeurIPS, pp. 6000–6010.

Villegas, Danae Sánchez, Mokaram, Saeid, and Aletras, Nikolaos (May 2021). "Analyzing Online Political Advertisements". arXiv: 2105.04047 [cs.CL].

Weidinger, Laura et al. (Dec. 2021). "Ethical and social risks of harm from Language Models". arXiv: 2112.04359 [cs.CL].

Zhu, Xiaojin and Goldberg, Andrew B. (2009). *Introduction to Semi-Supervised Learning*. San Rafael: Morgan & Claypool.

A Data Dictionary

Variable Name	Variable Type	Description
doc_uuid	`String`	Alphanumeric string to identify the document
lookup	`Array[Integer, String]`	String index/string pair to identify a word
dependency_graph	`Array[Integer, Integer]`	Integer/Integer pair to denote head/child node in the dependency graph
doc_length	`Integer`	Length of the sentence
sentence_root	`Integer`	String index position of the sentence root
sentence_literal	`String`	Raw text of the sentence
sentence_id	`Integer`	Index identification of a sentence in the document
actor_name	`String`	Name of the speaker
discourse_time	`Date[yyyy-mm-dd]`	Date when the speech was given

Table 7: Speeches – Data Dictionary

Variable Name	Variable Type	Description
time	`Date/UNIX-Time`	Time when activity was registered in the parliament
text	`String`	Summary of the parliamentary activity.
key	`String`	Type of parliamentary activity
		"rel": Edition of initiative or petitions
		"ini": Legislative initiative
		"req": Questions to the government
		"actP": Votes in the plenary
name	`String`	Name of the deputy

Table 8: Parliamentary Activities – Data Dictionary

B Labelling examples

As explained in the section *Labelling definitions and Operationalization*, specific examples representing linguistic modalities were used to manually label a small corpus of data. This section seeks to give some more space to these examples. Note that [...] indicates an ellipsis, indicating that the expression has been truncated for illustrative purposes. The following list contains sentences labelled with `1`, the list below that names sentences labelled with `0`, bearing the name **NONCONF** (Non-Conforming).

1. **DM1** (Label: 1): É necessário avançar de forma a que, nos cuidados paliativos, se consiga uma referenciação mais célere, com particular atenção à implementação das equipas comunitárias de suporte em cuidados paliativos, para que estas atendam os doentes no domicílio. – João Dias (PCP).

2. **DM2** (Label: 1): É urgente a proibição de despedimentos em empresas com lucros, tomar medidas de proteção do emprego e dos direitos do trabalho [...]. – Catarina Martins (BE).

3. **DM3** (Label: 1): Efetivamente, a nossa prioridade é garantir maior justiça fiscal, fazer os incentivos certos em matéria fiscal e não entrar num leilão de redução indiferenciada dos impostos para tudo e para todos. – António Costa (PS).

4. **DM4** (Label: 1): É essencial tomar medidas que permitam a resolução dos défices operacionais crónicos e das dívidas financeiras crescentes e insustentáveis quer do transporte ferroviário quer do transporte rodoviário público. – Afonso Oliveira (PSD).

5. **DM5** (Label: 1): O uso indevido de trabalho independente deve ser combatido nos devidos termos da lei, porque é um imperativo ético e de solidariedade [...] – Maria das Merces Soares (PSD).

6. **DM6** (Label: 1): A regeneração e a defesa da biodiversidade deveriam ser uma componente e uma aposta [...] – Inês de Sousa Real (PAN).

7. **DM7** (Label: 1): Entendemos que importa concretizar medidas que garantam a efetiva e plena inclusão das crianças e dos jovens com necessidades especiais em todo o ensino obrigatório [. . .] – Diana Ferreira (PCP).

8. **DM8** (Label: 1): Neste sentido, estamos abertos a um amplo debate que inclua as autarquias locais, que promova a implementação de mais praias concessionadas, com rigorosa fiscalização do cumprimento das competências definidas no contrato de concessão. – Rui Silva (PSD).

9. **DM9** (Label: 1): É preciso tomar medidas e é preciso garantir, de uma vez por todas, o ensino artístico, a sua valorização e a valorização do Conservatório e da sua comunidade educativa. – Rita Rato (PCP).

10. **DM10** (Label: 1) : Impõe-se assegurar uma eficaz regulação da concorrência, combatendo os abusos de posição dominante e de dependência económica, apostando no mercado interno do qual dependem grande parte das PME. – Bruno Dias (PCP).

11. **PROP1** (Label: 1): Por tal, propomos a extensão da contribuição a todas as embalagens plásticas, secundárias e terciárias. – André Silva (PAN).

1. **NONCONF** (Label: 0): Aí se gritava: «quanto mais calados, mais roubados. - Bernadino Soares (PCP).

2. **NONCONF** (Label: 0): Para esse debate, o CDS não está disponível e eu acho que ninguém, aqui, deveria estar disponível. - Cecília Meireles (CDS).

3. **NONCONF** (Label: 0): Para se ter uma noção, em 2013 e em 2014 foram adotadas do estrangeiro 10 crianças e 27 foram adotadas para o estrangeiro. - Pedro Mota Soares (CDS).

4. **NONCONF** (Label: 0): Em plena crise económica, com o recuo do negócio do crédito, os bancos carregaram nas comissões para assegurarem os seus níveis mínimos de rentabilidade. - Paulino Ascenção (BE).

5. **NONCONF**: (Label: 0): Não pode ser de outra maneira, porquanto o Partido Socialista está em aliança política permanente com o PSD, para o PEC 1, para o PEC 2 e para os demais PEC que aí virão. - Luís Fazenda (BE).

6. **NONCONF** (Label: 0): O nosso compromisso é não perder mais tempo, um tempo que é precioso para os portugueses. - Carlos César (PS).

7. **NONCONF** (Label: 0): A Grécia é, agora, exatamente o País que tem, em termos de dívida e em termos de rating, algo pelo qual — dizem-nos — este Governo batalha. - João Galamba (PS).

8. **NONCONF** (Label: 0): Então, porque continua o Governo a manter a sua obsessão com projectos para os quais, objectivamente, não existe financiamento? - Miguel Frasquilho (PSD).

C Labelling Process: Result Transparency

To verify a sample of results from the self-labelling process in a way that goes beyond the reported *F1-Score*, the following list contains 10 randomly drawn results from the labelling process with meta data. Due to the probabilistic output of the Gradient Boosting Machine (for which the rounding threshold was one hyperparameter), the label probability is included alongside the rounded label. The overview contains the sentence text, the name of the speaker, date as well as probabilistic label and the rounded (final) label as computed by the labelling machine. Judgements on label correctness are based on the author's judgements.

1. »Que esta tragédia constitua, em si, uma força para conseguirem atingir aquele objectivo e, ao mesmo tempo, que constitua também um reforço da energia, que o povo polaco sempre demonstrou a todos os seus parceiros internacionais, no sentido de ser capaz de superar todas as dificuldades, toda a dor, toda a tragédia para investirem naquele trabalho que, verdadeiramente, importa: o de desenvolvimento em termos de civilização, tarefa, esta, que nunca acaba e que, porventura, estes episódios nos ajudam a reconhecer como a primeira de todas elas.« - **Maria de Belém Roseira** (2010-04-15): 0.9540 ($\rightarrow$ 1). *Incorrect label*.

    * Commentary: While the attainment of an otherwise »objectivo« that is alluded to in this sentence is not explicitly specified, it does not refer to the speaker herself but an external entity (the Polish people). The label probability that the model has generated is quite high, but the sentence itself does not conform to the idea of modality as elaborated under *Labelling definitions and Operationalization*.

2. »Por isso, Presidente, e Deputados, é que Eusébio se foi afirmando, pela sua forma de ser e por aquilo que tinha sido também a sua carreira desportiva, como uma das marcas de Portugal e, por isso mesmo, como diz o voto, como um «embaixador de Portugal, não só por chegar a todos os cantos do mundo, mas porque aquilo que chegava a todos os cantos do mundo era a de uma pessoa boa, a de um homem íntegro, um homem que gostava da sua Pátria e que

inspirava a sua Pátria.« - **Luís Montenegro** (2014-01-10): 0.8638 ($\rightarrow$ 1). *Incorrect label*.

   * Commentary: This laudatory statement for a deceased football player has a high label probability, yet does not quite fit the deontic label that is has, as the statement does not relate to anything explicitly political.

3. »As suas características, nomeadamente o seu preço acessível, tornaram o plástico um material usado em larga escala para os mais variados fins.« - **António Topa** (2019-04-11): 0.5664 ($\rightarrow$ 1). *Correct label*.

   * Commentary: While this statement does not contain an explicit modality marker as in the examples, the speaker does convey a certain message about a political issue, namely the advantage of using plastics over other materials.

4. »Agora, de certeza que não é um programa, como todos os outros, que deva existir para derrapar.« - **Paulo Macedo** (2012-10-31): 0.003 ($\rightarrow$ 0). *Correct label*.

   * Commentary: The sentence is classified as not containing the required modality, which makes sense in this case. There are no specific political keywords or statements, to which the labelling machine could refer in terms of semantic vector space.

5. »Tal situação só ocorreu devido ao reconhecimento por parte de todos os grupos parlamentares de que o setor da economia social, pelo relevante contributo que presta à sociedade portuguesa, nas mais variadas áreas, é merecedor de todo o nosso empenho e traduz a capacidade de congregar vontades em torno de uma lei em que todos se revejam.« - **Maria das Mercês Soares** (2013-03-20): 0.9999 ($\rightarrow$ 1). *Correct label*.

   * Commentary: The algorithm in this case is almost certain that this is the required modality and one is tempted to concur: The sentence contains references to a political issue as well as a statement of support by the speaker.

6. »Esta decisão cabe, e só pode caber, às pessoas que infelizmente se encontram nesta situação.« - **Bebiana Cunha** (2020-02-20): 0.2793 ($\rightarrow$ 0). *Correct label*.

- Commentary: This statement does not contain enough information to be classified as positive, even though it could have been based on the reference to »esta decisão«.

7. »Obviamente, isso tem consequências no conteúdo concreto daquilo que estamos a tratar.« - **João Pinho de Almeida** (2020-06-03): 0.4775 (→ 0). *Correct label*.

    - Commentary: This sentence is – like its predecessor – to short to be given the positive label. Given that not only the sentence in question alone, but also the context is considered, the algorithm was close to classifying this as positive.

8. »No mês de Fevereiro, em sete dias, apenas quatro dias cumprem o horário de funcionamento previsto, até às 24 horas; nos outros dias, não funciona, não presta o serviço, encerrando, em muitos casos, às 17 horas.« - **Maria das Mercês Soares** (2011-02-24): 0.8131 (→ 1). *Correct label*.

    - Commentary: This is another case of an implicit modality. Here, the speaker is stating her disapproval of a certain state of affairs with the implicit message that this should be changed.

9. »Vamos às questões mais substantivas: o dia de hoje, o início de diálogo com a troica.« - **José Junqueiro** (2013-09-16): 0.1400 (→ 0). *Correct label*.

    - Commentary: This statement does contain a political marker highly relevant in Portugal for a number of years: The »Troica« (European Central Bank, International Monetary Fund and European Commission). However, the statement in itself is quite neutral and does not offer a political leaning by the speaker, which is the most likely reason why it classified as negative.

10. »Infelizmente, tal não tem acontecido entre nós, neste como noutros domínios.« - **Luís Capoulas** (2011-01-13): 2.379e-5 (→ 0). *Correct label*.

- Commentary: This statement (and in extension its neighbours) do not contain the relevant keywords or indications of deontic modality, making the algorithm's classification reasonable.

Summary Statement: Out of 10 randomly sampled speech acts, the model has correctly classified 8. The two incorrectly labelled statements were false positive, where the model assumed class `1` when in fact they should have been labelled as `0`. This translates to an F1-Score of 0.75, which contrasts with an epoch average of 0.927 during the self-labelling process.

D Symmachus Model: Boosting Hyperparameters

The following table summarizes the final hyperparameters of the Gradient Boosting Machine used for the Symmachus document embeddings. η refers to the learning rate, that is how large the influence of the new learners should be. Max. Depth refers to the maximum depth of the decision trees used in the learning process. The number of rounds refers to the number of training epochs. The Truth Threshold refers to the threshold beyond which an observation is measured as belonging to the positive class.

Parameter Name	Parameter Value
η (Eta)	0.4
Max. Depth	4
Number of rounds	125
Truth Threshold	0.6

Table 9: Gradient Boosting Machine – Hyperparameters

E SentenceTransformer Architecture

The following figure shows a high-level overview of the *SentenceTransformer* architecture. Fundamental for the understanding of the architecture are the so-called »Siamese« Neural Networks that were introduced independently by Baldi and Chauvin (1993) and Bromley et al. (1993). For the original implementation, the name »Siamese« was based on the fact that the weights of the two layer stacks were constrained to be the same. The architecture uses the output of a `BERT`-Model to derive representations for sentences. Appendix E shows the case where similarity between embeddings is computed using the cosine similarity. Another choice is possible, which would be useful for potentially fine-tuning the model. This choice would be to use triplets. As per the original paper, triplets use an *anchor sentence* to compute the loss for two sentences, tuning the weights so that the sentence more similar to the anchor has a higher value on the similarity metric (e.g. cosine similarity or euclidean distance) than the more dissimilar sentence. This can be expressed by the following formula from the original paper (Reimers and Gurevych 2019, 3984):

$$max(\|s_a - s_p\| - \|s_a - s_n\| + \varepsilon, 0) \tag{7}$$

Here, s_a refers to the anchor sentence, s_p the positive, i.e. more similar sentence, and s_n the negative or more dissimilar sentence. The ε is to assure that the positive sentence's embedding is closer to the anchor sentence than that of the negative sentence.

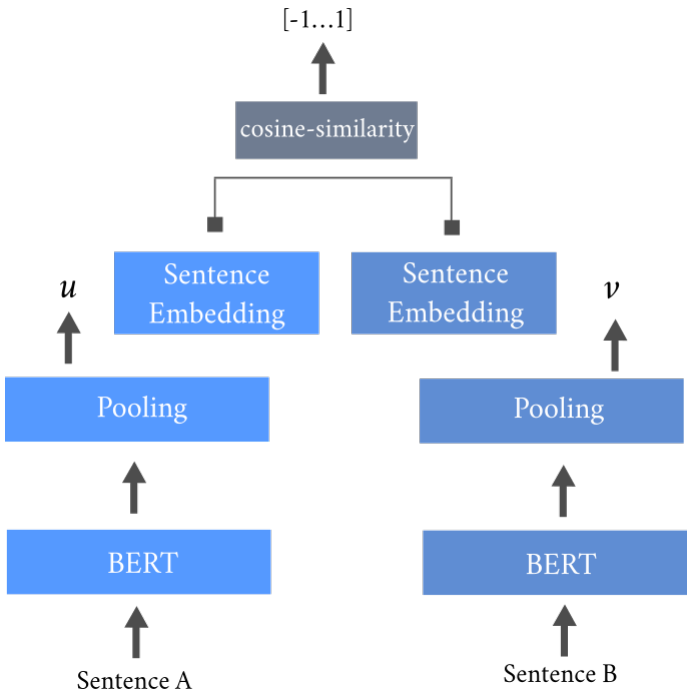


Figure 8: *SentenceTransformer* Architectural Overview. Adapted from: Reimers and Gurevych (2019, 3984), Figure 2.

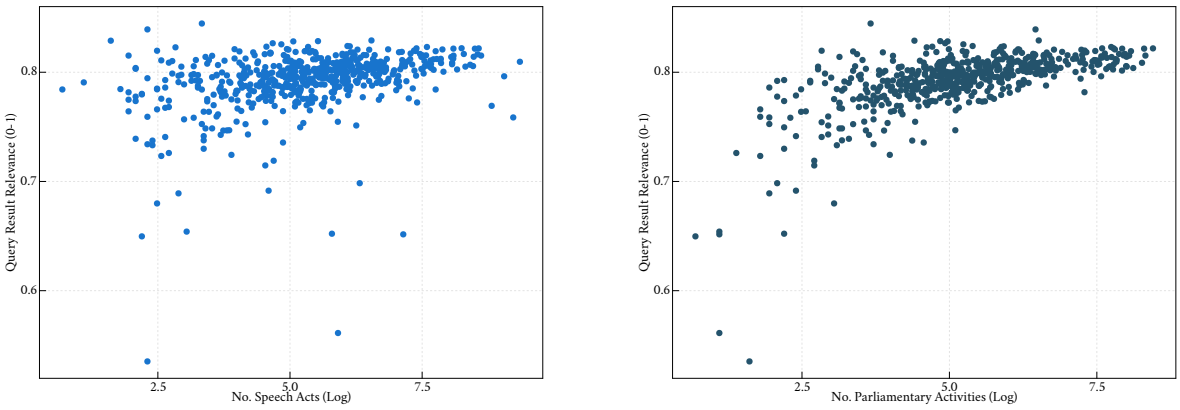
F Sentence-Transformer Model References

Filho, Ricardo (2021). Bert-portuguese-cased-nli-assin-assin-2. Available at: https://huggingface.co/ricardo-filho/bert-portuguese-cased-nli-assin-assin-2. Last accessed: 4.11.2021.

UKP Lab (2021). Distiluse-base-multilingual-cased-v1. Available at: https://huggingface.co/sentence-transformers/distiluse-base-multilingual-cased-v1. Last accessed: 4.11.2021.

NeuralMind Inteligência Artificial (2021). Bert-base-portuguese-cased. Available at: https://huggingface.co/neuralmind/bert-base-portuguese-cased/tree/main. Last accessed: 4.11.2021

G Relevance of Speech Acts and Activities by Corpus Length



(a) Relevance of Speech Acts

(b) Relevance of Activities

Figure 9: Relevance of Speech Acts and Activities by Corpus Length

H SDG Vector Space Alignment Per Topic

SDG No.	Most aligned political actors
8 – Trabalho digno e cresimento económico	José Manuel Pureza (BE) - 0.8631; José Moura Soeiro - 0.8624 Paula Santos (PCP) - 0.8619; Bruno Dias (PCP) - 0.8609
9 – Indústria, inovação e infraestruturas	Bruno Dias (PCP) - 0.8518; José Manuel Pureza (BE) - 0.8504 Paula Santos (PCP) - 0.8500; Carlos Matias (BE) - 0.8457
10 – Reduzir as desigualdades	Bruno Dias (PCP) - 0.9428; Paula Santos (PCP) - 0.9413 Ana Rita Bessa (CDS) - 0.9394; António Filipe (PCP) - 0.9390
11 - Cidades e comunidades sustentáveis	João Oliveira (PCP) - 0.9254; Paula Santos (PCP) - 0.9243 Pedro Mota Soares (CDS) - 0.9237; Bruno Dias (PCP) - 0.9234
12 - Produção e consumo sustentáveis	Bruno Coimbra (PSD) - 0.9224; Carlos Matias (BE) - 0.9214 Luís Leite Ramos (PSD) - 0.9184; Renato Sampaio (PS) - 0.9179
13 - Ação climática	Luís Leite Ramos (PSD) - 0.9343; Carlos Matias (BE) - 0.9324 Ana Rita Bessa (CDS) - 0.9320; Pedro Mota Soares (CDS) - 0.9313
14 - Proteger a vida marinha	Carlos Matias (BE) - 0.9358; Sandra Cunha (BE) - 0.9352 José Manuel Pureza (BE) - 0.9333; Ana Rita Bessa (CDS) - 0.9333
15 - Proteger a vida terrestre	Ricardo Vicente (BE) - 0.9220; Paula Santos (PCP) - 0.9220 Bruno Dias (PCP) - 0.9217; Andrße Silva (PAN) - 0.9208
16 - Paz, justiça e instituições efficazes	António Filipe (PCP) - 0.9378; Paula Santos (PCP) - 0.9372 Rita Rato (PCP) - 0.9368; Bruno Dias (PCP) - 0.9362
17 - Parcerias para a implementação dos objetivos	Ana Rita Bessa (CDS) - 0.9187; Paula Santos (PCP) - 0.9176 Carlos Matias (BE) - 0.9163; Bruno Dias (PCP) - 0.9156

Table 10: Topic Search: Most aligned political actors per SDG II

I Speech-Activity-Match Search: Result Transparency

The implementation of the speech-activity-match search incorporates some of the fundamental algorithmic steps of the system developed in this workproject, i.e. using vector space representations of natural language to retrieve similar observations using similarity metrics *(Cosine, Euclidean)* and algorithms *(ANN, KNN)*. To validate the results produced, a manual mirroring of the algorithmic steps that lead to the retrieval of matching parliamentary activities for a given speech act will be conducted. This means taking a speech act and combing the activity archive on *parlamento.pt* for a given deputy. Speech acts, for which no manual best result could be retrieved, are marked. If the parliamentary activity retrieved by the model cannot be improved upon, that activity is considered the best result for the manual inspection as well. Hence, a sample of 5 political actors with 3 speech acts each will be drawn. For each political actor's sample a commentary will be offered below. Results of this process can be found in Table 11 to Table 17. The political actors in question are **Carla Tavares** (PS), **Adão Silva** (PSD), **Sandra Cunha** (BE), **Pedro do Carmo** (PS), **Andreia Neto** (PSD).

1. **Carla Tavares** (PS)

    I. The first speech act deals with the issue of demographic decline. While the model retrieves an activity that relates to housing prices, this not quite fit the broad tone of the speech act, although both demographic decline and housing prices can be connected in the sphere of »social issues«. A speech act relating to reproductive health might have been more fitting.

    II. Here, the speech act refers to workers' rights. The most relevant result retrieved by the model fits quite well and there was no reason to substitute the model result with a different result from manual inspection.

    III. For this speech act, it has proven impossible to retrieve a fitting speech act using the statement alone. The statement was uttered in the context of family policy during a

debate. The model's most relevant result does not refer to the correct political issue.

2. **Adão Silva** (PSD)

   I. This speech act concerns the pension system and the erosion of public trust in it. Here, an activity about the contributions regulation seems the best-matching activity. However, the model retrieves an activity that peripherally touches on »social« issues but is not specific enough.

   II. This speech act seems to have an anomalous model result, as it does not retrieve a parliamentary activity but something different entirely. Faulty data is the most likely culprit. Through manual search, a fitting activity about supporting the labour market has been retrieved.

   III. For the speech referring to the SNS *Serviço Nacional de Saúde)*, a base law about health seems a fitting activity. However, the model seems to have retrieved a result that broadly refers to *segurança social*, the social security system, but not specifically the health system.

3. **Sandra Cunha** (BE)

   I. This speech act broadly covers environmental issues, particularly concerning the protection of forests. Here, a very specific parliamentary activity can be found. The search result retrieved by the model can be classified as inhabiting a similar idea space (protection against pollution), yet the specificity is somewhat lacking, since the recommendation does not relate to forestry, but pollution at an air field.

   II. This speech act does not contain sufficient information to retrieve a fitting activity. Based on the mention of *mulheres*, an activity relating to women's rights could be possible. The model retrieves an activity relating to care homes, which might or might not be sensible. Without a context it is difficult to make a decision.

III. Similar to the previous speech act, this one lacks specificity, so some degree of guesswork is involved. The model's most relevant result seems completely off and irrelevant. Again, an activity relating to women's right and protection (relating to pandemic-related restrictions) could be fitting.

4. **Pedro do Carmo** (PS)

I. The speech act seeks to extol the valour of the Portuguese agricultural system. Because there is mention of certain *necessidades*, the activity retrieved by manual inspection contains a data collection system about the food supply. The model seems to have gone in a different direction and considers some sort compensatory payment the most relevant result. This would certainly fit the laudatory tone of the speech act.

II. In the case of this speech act, the manual inspection and the model results concur. Referenced in the speech act are pesticides and how they can be deployed as well as some provisions for education about their usage. The matching activity retrieved by the model contains those exact provisions and a manual inspection of the rest of the activity corpus has not yielded a better result.

III. The geographic locations mentioned in the speech act have made it trivial to search for a fitting speech act by hand. However, the model too has retrieved that same activity. For this speech act in particular, the geographic mention alone would probably be sufficient for a relevant speech act, but the collocation of *oportunidades* in the speech act and *desenvolviment integrado and participado* in the activity are quite remarkable.

5. **Andreia Neto** (PSD)

I. Through the mention of *direitos, liberdades e garantias* this speech act is firmly located in the juridical sphere. The model retrieves a law proposal to introduce a study area on the children's rights convention, which is certainly a plausible choice. Given the context of international affairs, the drive to find similar solutions *(procurar soluções*

semelhantes) suggests a different spin, which is why the most relevant result by manual inspections differs slightly.

II. For this relatively abstract speech act, an oddly specific activity about the profession of a »night guard« can be found. The model retrieves an activity that is not quite relevant in the (admittedly) abstract context. One might construe a connection through the idea of public security.

III. This speech act does contain a specific policy proposal, only a mention of a declaration of intent on criminal policy *(Lei-Quadro da Política Criminal)*. One might find a law on the reorganization of the judicial system fitting here. The model, on the other hand, has retrieved an activity that is in the same idea sphere of criminal law and justice, which generally matches.

Summary Statement: Considering the results of the 15 sampled speech acts from 5 political actors, the model retrieved reasonably relevant results in 5-6 of 15 cases. It is important to note that for one political actor the model retrieved relevant results for all sampled speech acts, while for two none were relevant.

Speech Act (Query)	Most relevant result (Manual)	Most relevant Result (Model)
Segundo o Ageing Report de 2018, publicado pela Comissão Europeia, Portugal deverá perder 23% da sua população até 2070, descendo dos 10,3 milhões de pessoas em 2016 para cerca de 8 milhões de pessoas em 2070.	Regime de proteção na pré conceção, na procriação medicamente assistida, na gravidez, no parto, no nascimento e no puerpério. (872/XIII)	Recomenda ao Governo a criação de limites máximos à renda apoiada em função da taxa de esforço para impedir aumentos exponenciais da renda apoiada nos bairros sociais geridos pelo IHRU.
Além disso, verificou-se ainda a redução do número de trabalhadores com horários de trabalho incompletos e com trabalho a tempo parcial, o que foi compensado pelo aumento do número de trabalhadores com horários e salários completos, o que tem um manifesto impacto positivo ao nível dos rendimentos e da diminuição dos índices de pobreza dos trabalhadores.	Repõe o valor do trabalho suplementar e o descanso compensatório, aprofundando a recuperação de rendimentos e contribuindo para a criação de emprego (15.ª alteração ao Código do Trabalho aprovado pela Lei n.º 7/2009, de 12 de fevereiro) (553/XIII)	Repõe o valor do trabalho suplementar e o descanso compensatório, aprofundando a recuperação de rendimentos e contribuindo para a criação de emprego (15.ª alteração ao Código do Trabalho aprovado pela Lei n.º 7/2009, de 12 de fevereiro)
Presidente, antes de mais, quero cumprimentar e felicitar o PCP pela escolha do tema para as declarações políticas de hoje.	—	Solicita a alteração da Lei n.º 7/2007, de 5 de Fevereiro, que criou o cartão de cidadão e rege a sua emissão e utilização, no sentido de serem aditados ao circuito integrado do cartão (chip) elementos de identificação adicionais e de ser criado um cartão »braçadeira eletrónica« para pessoas em situação vulnerável.

Table 11: Carla Tavares – Speech-Activity-Match Search Verification

Speech Act (Query)	**Most relevant result (Manual)**	**Most relevant Result (Model)**
E a verdade, e Deputados, é que não há nada mais corrosivo num sistema de segurança social do que a falta de confiança dos cidadãos contribuintes face ao destino dado às suas contribuições e aos seus impostos.	Alteração da Lei n.º 110/2009, de 16 de Setembro, Código dos Regimes Contributivos do Sistema Previdencial da Segurança Social. (44/XI)	Recomenda ao Governo uma avaliação da aplicação do Decreto-Lei nº 29/2001, de 3 de Fevereiro (que estabelece o sistema de quotas de emprego para pessoas com deficiência, com um grau de incapacidade igual ou superior a 60% nos serviços e organismos da administração central e local)
O que está agora a gerar emprego são as alterações ao Código do Trabalho!	Institui medidas transitórias e excepcionais de promoção do emprego. (528/XI)	*O segundo semestre está a arrancar e ainda há 13 604 estudantes à espera de saber se terão bolsa este ano letivo, di-lo a comunicação social. É por causa da redução do horário de trabalho para as 35 horas?*

Table 12: Adão Silva – Speech-Activity-Match Search Verification I

Speech Act (Query)	**Most relevant result (Manual)**	**Most relevant Result (Model)**
No momento em que o Serviço Nacional de Saúde está praticamente esgotado na sua capacidade para responder à pandemia e no momento, bem comprovado, em que não é capaz de responder às exigências correntes — de acessibilidade às consultas, às cirurgias — a minha pergunta é esta: porque não fazer esta cooperação, esta correlação entre o setor público, o Serviço Nacional de Saúde, e o setor social e o setor.	Lei de Bases da Saúde (1065/XIII)	Recomenda ao governo que assegure que a reflexão e ponderação sobre a possibilidade de integração da caixa de previdência dos advogados e dos solicitadores (CPAS) na segurança social, a ser equacionada pelo governo, seja necessariamente feita em estreita articulação com a CPAS, a ordem dos advogados e a ordem dos solicitadores e agentes de execução.

Table 13: Adão Silva – Speech-Activity-Match Search Verification II

Speech Act (Query)	Most relevant result (Manual)	Most relevant Result (Model)
É o fruto de anos de desinvestimento nesta área, de uma aposta num ordenamento da floresta que visa o lucro fácil e rápido em detrimento da proteção da floresta, do desenvolvimento sustentável e da proteção das populações.	Pela proteção do património cultural face aos riscos das actividades de produção agrícola e florestal (999/XIV)	Recomenda ao Governo que desenvolva todos os esforços diplomáticos para garantir o fim da poluição e a descontaminação dos solos e aquíferos contaminados por derrames de hidrocarbonetos na base aérea das Lajes
Os projetos do Bloco de Esquerda estão, neste momento, em sede de especialidade, à espera dos contributos e do apoio das iniciativas dos outros grupos parlamentares para que possamos responder a este problema e, assim, podermos proteger estas mulheres de uma vez por todas.	Recomenda ao Governo medidas para a não exclusão de mulheres dos procedimentos de procriação medicamente assistida por atrasos devidos à pandemia de Covid-19 (1019/XIV)	Recomenda a elaboração de um estudo e de um manual de boas práticas para os lares de idosos, o reforço da fiscalização por parte da Segurança Social a estas instituições e o reforço das respostas públicas ao nível dos cuidados continuados e do apoio domiciliário a idosos
O que queremos, com a declaração política que hoje aqui trouxemos, é tentar perceber se os Deputados e as Deputadas estão na disposição também de fazer alguma coisa e de proteger estas mulheres, de uma vez por todas.	Garante o acompanhamento da mulher grávida na assistência à gravidez e em todas as fases do parto mesmo durante a pandemia de Covid-19 (636/XIV)	Recomenda ao Governo que não transfira mais verbas para o Fundo de Resolução com vista à injeção de capital no Novo Banco até que a auditoria às suas contas esteja concluída

Table 14: Sandra Cunha – Speech-Activity-Match Search Verification

Speech Act (Query)	Most relevant result (Manual)	Most relevant Result (Model)
Uma vez mais, o interior, a agricultura, a pecuária e outros setores produtivos responderam e corresponderam às necessidades dos portugueses, que voltaram a dizer «presente.	Desenvolvimento um sistema de recolha de dados, relativos aos preços e ao mercado, da cadeia de abastecimento alimentar (2208/XIII)	Recomenda ao Governo a adoção de medidas de minimização dos prejuízos verificados no sector da fruticultura e em produções agrícolas, face às condições atmosféricas extremas ocorridas a 31 de maio no Centro e Norte do País.
Os aplicadores de produtos fitofarmacêuticos, nomeadamente agricultores, podem, assim, dirigir-se às organizações de agricultores e outros representantes do setor e também aos serviços regionais do Ministério da Agricultura, Florestas e Desenvolvimento Rural, no sentido de obterem mais esclarecimentos, bem como de se inscreverem nas ações de formação disponíveis e obterem o respetivo certificado, que lhes permitirá a continuação do exercício da sua atividade dentro da legalidade.	Altera os prazos e critérios para a formação de aplicador de produtos fitofarmacêuticos - Primeira alteração à Lei n.º 26/2013, de 11 de abril que regula as atividades de distribuição, venda e aplicação de produtos fitofarmacêuticos para uso profissional e de adjuvantes de produtos fitofarmacêuticos e define os procedimentos de monitorização à utilização dos produtos fitofarmacêuticos, transpondo a Diretiva n.º 2009/128/CE, do Parlamento Europeu e do Conselho, de 21 de outubro, que estabelece um quadro de ação a nível comunitário para uma utilização sustentável dos pesticidas, e revogando a Lei n.º 10/93, de 6 de abril, e o Decreto -Lei n.º 173/2005, de 21 de outubro (67/XIII)	Altera os prazos e critérios para a formação de aplicador de produtos fitofarmacêuticos - Primeira alteração à Lei n.º 26/2013, de 11 de abril que regula as atividades de distribuição, venda e aplicação de produtos fitofarmacêuticos para uso profissional e de adjuvantes de produtos fitofarmacêuticos e define os procedimentos de monitorização à utilização dos produtos fitofarmacêuticos, transpondo a Diretiva n.º 2009/128/CE, do Parlamento Europeu e do Conselho, de 21 de outubro, que estabelece um quadro de ação a nível comunitário para uma utilização sustentável dos pesticidas, e revogando a Lei n.º 10/93, de 6 de abril, e o Decreto -Lei n.º 173/2005, de 21 de outubro

Table 15: Pedro do Carmo – Speech-Activity-Match Search Verification I

Speech Act (Query)	**Most relevant result (Manual)**	**Most relevant Result (Model)**
O Sudoeste Alentejano e a Costa Vicentina precisam é de respostas que digam «sim às populações, que apresentem oportunidades, que invistam nas pessoas e no território.	Recomenda ao Governo que retome o Plano de Ordenamento do Parque Natural do Sudoeste Alentejano e Costa Vicentina como instrumento de desenvolvimento integrado e participado (835/XIV)	Recomenda ao Governo que retome o Plano de Ordenamento do Parque Natural do Sudoeste Alentejano e Costa Vicentina como instrumento de desenvolvimento integrado e participado

Table 16: Pedro do Carmo – Speech-Activity-Match Search Verification II

Speech Act (Query)	Most relevant result (Manual)	Most relevant Result (Model)
Sendo a Europa um espaço comum e globalizado, os problemas, necessariamente, tendem a procurar soluções semelhantes, tanto mais quanto estamos a falar de direitos, liberdades e garantias.	Recomenda ao Governo a atribuição ao Provedor de Justiça da função de coordenar e monitorizar a aplicação da Convenção sobre os Direitos da Criança em Portugal (1807/XIII)	3ª alteração à Lei n.º 2/2008, de 14 de janeiro (regula o ingresso nas magistraturas, a formação de magistrados e a natureza, estrutura e funcionamento do centro de estudos judiciários), incorporando uma área de estudo que incida sobre a Convenção sobre os Direitos da Criança
e Deputados, a atividade de segurança privada tem, nos termos do respetivo quadro legal, uma função subsidiária e complementar da atividade das forças e serviços de segurança pública do Estado.	Estabelece o regime jurídico da atividade de guarda-noturno (775/XII)	3ª alteração à Lei n.º 2/2008, de 14 de janeiro (regula o ingresso nas magistraturas, a formação de magistrados e a natureza, estrutura e funcionamento do centro de estudos judiciários), incorporando uma área de estudo que incida sobre a Convenção sobre os Direitos da Criança
Ora, Sr.ª Ministra, este compromisso obriga a respeitar escrupulosamente o disposto na Lei-Quadro da Política Criminal.	Altera a Lei da Organização do Sistema Judiciário (145/XIII)	Solicita que a Assembleia da República requeira ao Tribunal Constitucional a declaração de inconstitucionalidade de normas do Estatuto dos Militares da Guarda Nacional Republicana ou que tome medidas legislativas para repor o regime vigente antes das alterações decorrentes do Decreto-Lei n.º 159/2005, de 20 de Setembro.

Table 17: Andreia Neto – Speech-Activity-Match Search Verification

J Topic Alignment: Result Transparency

This section is dedicated to a case investigation of five randomly sampled political actors from the results of the topic search. The politicians in question are:

1. **Manuela Ferreira Leite** (PSD)
2. **Alexandra Tavares de Moura** (PS)
3. **Francisco Rocha** (PS)
4. **Carlos Alberto Gonçalves** (PSD)
5. **Ofélia Ramos** (PSD)

Based on their deputy profile from *parlamento.pt* and the SDG descriptor they are aligned with the most (based on the results of the topic search) the models performance will be assessed. The methodology is based on an expert review, augmented with corpus statistics (e.g. frequency of words). Corpus statistics are considered separately for the most aligned speech acts and the rest of the corpus. First, the deputy's parliamentary activity profile is manually evaluated regarding the model result, i.e. whether the model is correct in identifying an SDG as the most representative for that deputy. The second step consists in repeating the evaluation for speech acts that are obtained from the automatic labelling process. This is achieved through the interpretation of corpus statistics on speech acts to identify the SDG with which the political actor would be aligned the most.

1. **Manuela Ferreira Leite** (PSD): This political actor is most aligned with SDG No. 16 – *Paz, Justiça e Instituições efficazes* with a mean external alignment (including speeches and activities in parliament) of 0.776. Her speeches are much more aligned than her activities (speeches: 0.907, activities: 0.644). Here, it is important to note that she was a deputy before the legislature that is considered here (11^{th} legislature). Based on the content of SDG No. 16, one would expect activities relating to combating corruption, the rule of law, an

informed public as well as good institutions. Due to her being an economist, most activities relate to economics in government, such as improving financial literacy (405/XI) as well as a law to support the government on budget matters (295/XI/1). Due to the strong focus on stabilizing the state budget (80-COFXI; 76-COF-XI) and participation in the budgetary commission, the model alignment score regarding activities makes sense as it does not quite fit the overall significance of SDG No. 16. Common phrases relate to a crisis situation (*crise, endividamento externo, viabilizar orçamento*) and how that crisis situation relates to the public (*opinião pública, transparência, verdade, crise confiança, confiança governo*). The most aligned speech acts refer mainly to issues pertaining to the state budget *(orçamento)* and the macroeconomic situation *(agravar, relançar economia, endividamento páis)*. There are verbs such as *salvaguar[dar], conseguir crescimento*. One very strong example in the most aligned speech acts concerns the rule of law, most others concern the aforementioned issues. In the sense that the rule of law (*envenenando a sociedade com a convicção de que há protegidos e perseguidos*) and a stable state budget that is not quite dependent on external debt concerns all institutions of the state, SDG No. 16 with its item of *»Desenvolver instituições eficazes, responsáveis e transparentes«* fits well, even though it seems to over-emphasize economic issues.

2. **Alexandra Tavares de Moura** (PS): This political actor is most aligned with SDG No. 10 - *Reduzir as Desigualdades* with a mean external alignment of 0.923. Speech acts are more aligned with the SDG descriptor than activities (speeches: 0.944, activities: 0.901). The vector space, in which speeches and activities live, is quite aligned in this case. Her background Clinical Psychology and her activity as director of a Nursing School, gives rise to most of her parliamentary activities being in the area of health (e.g. 405/XIV, 403/XIV, 990/XIV - Referring to the career prospects of nurses). However, this is not reflected in the topic alignment scores, as SDG No. 10 is the one with the highest alignment score. Looking at the speech acts, we find references to *administração pública, carreira, valorização carreira* and *função públic[a]*, mostly in the context of the health service. This is strikingly similar to rest

of the corpus, which explains the high alignment between those two vector spaces (0.945). Overall, the speech acts and activities retrieved in the context of SDG No. 10 do not quite fit the descriptors of that SDG. A better fit would be SDG No.3 – *Sáude de Qualidade*. Indeed, SDG No. 3 is only the 9th most aligned SDG for the totality of her corpus.

3. **Francisco Rocha** (PS): This political actor is most aligned with SDG No. 10 - *Reduzir as Desigualdades* with a mean external alignment of 0.926. His speech acts are more aligned than activities (speeches: 0.934, activities: 0.918). In terms of parliamentary activities, these stretch over legislative periods XIII and XIV. Law proposals in legislature XIII relate to the protection of minors (1239/XIII, 1190/XIII) or protection against sexual abuse (1155/XIII). Votes in the plenary thematize issues relating to forms of discriminatory practices, such as recognition of LGBTQI+ rights (Voto/850/XIII), racial discrimination (Voto/775/XIII, Voto/584/XIII). In legislature XIV, we find similar votes on LGBTQI+ rights (Voto/580/XIV, Voto/51/XIV). There are also votes on women's rights and protection against domestic violence (Voto/75/XIV, Voto/120/XIV, Voto/199/XIV, 747/XIV). Based on the activity alone, SDG No. 10 would fit well. However, the most aligned speech acts extracted by the model do not refer to any of these issues and are more concerned with issues relating to forestry and agriculture (e.g. *florestal, sustentabilidade, rural, importância floresta*). The rest of the corpus concurs and the most frequent phrases are *florestal, rural, agricultura, setor florestal*. In this case, the model scoring appears misaligned and another SDG clearly would have been the better choice.

4. **Carlos Alberto Gonçalves** (PSD): This political actor is most aligned with SDG No. 16 – *Paz, Justiça e Instituições efficazes* with a mean external alignment of 0.916. His speeches are more aligned than activities (speeches: 0.924, activities: 0.907). His case is a special one as he is a special deputy for the »district« of Europe, i.e. representing all Portuguese communities living abroad in Europe. In this, role the strengthening of ties with those communities motivates several law proposals, such as 295/XI about strengthening Portuguese culture and language abroad or 392/XII about granting Portuguese citizenship to foreign-born grandchil-

dren of Portuguese citizens. Another proposal refers to combating corruption (875/XIV). Large parts of votes are also dedicated to areas of foreign policy, e.g. the Portuguese reaction to developments in Afghanistan (2712/XIV/2), the condemnation of the use of poison gas in Syria (Voto/520/XIII) and the condemnation of the terror attack on Charlie Hebdo (Voto/242/XII). As for the most aligned speech acts, they refer – among others – to political prisoners in Cuba and the political situation in Guinea. Most other speech acts refer to the Portugal's international relations or the situation of Portuguese communities abroad. Indeed, among the most common phrases in the most aligned speech acts are *apelar, representação, união europeia*. This contrasts with common phrases in the rest of the corpus, which is in accordance with issues treated in the activities, such as *communidade portuguesa, emigração, estrangeiro, ensino português estrangeiro*. Based on the activities and speech act phrases, it makes sense to group him in SDG No. 16, as it is the goal that most embodies foreign policy with respect to human rights and the rule of law.

5. **Ofélia Ramos** (PSD): This political actor is most aligned with SDG No. 10 - *Reduzir as Desigualdades* with a mean external alignment of 0.915. Her speeches are narrowly more aligned than activities (speeches: 0.917, activities: 0.913). As legislature XIV is the only one so far, this is the one considered here for speech acts and parliamentary activities. Several activities relate to workers' rights, such as expanding the grief period for workers whose child has died (1018/XIV) or a law regarding remote work (812/XIV). An interest can also be found in social matters, as evidenced by questions on a missing payment in family subsidies (183/AC/XIV/2) and early retirement for people with disabilities (233/AC/XIV/2). The speech corpus contains common phrases such as *segurança social, lar ilegal, apoio, social, promover emprego* and *salário mínimo*. Fitting SDGs might be either SDG No. 8 - *Trabalho Digno e Crescimento Económico* or SDG No. 10, in the latter case especially concerning the clause *adotar políticas, especialmente ao nível fiscal, salarial e de proteção social [. . .]*. The phrases in the most aligned speech acts mirror this: They relate to social subsidies *(subsídio)*, the promotion of employment *(promover emprego)* and the phrase "workers" *(trabalhadores)*.

Considering the evidence, the algorithm computed well the most aligned SDG, as the manual inspection reaches similar conclusions.

Summary Statement: For the five political actors considered during this verification process, the model has reasonably selected the SDG with which they are aligned with the most in 3 out of 5 cases.