Shades of disdain: an analysis of disgust in Colombian congressional speech through word embeddings

Alejandro Sarria-Morales

May 21, 2024

Feedback implementation

The feedback received during in different forms during the final stretch of the development of this paper was extremely useful. In general, all the feedback was helpful in establishing a mindset in which the decision and compromises that I made early in the project needed to be questioned, reevaluated, and corrected even in cases were the feedback received did not target these elements. As such, the project benefited holistically to the exposure to foreign eyes and constructive comments.

The first, and possibly most impactful, feedback I received was during the presentation to the class, when I was questioned on the validity of using the gpt-3-small embedding model as a point of comparison to the Word2Vec model trained on the corpus of senate sessions, on which most of analysis was based. Understanding that the comparison between a contextualized embedding model (like gpt-3-small) and a static one (like Word2Vec) led me to understand that my interpretation of a part of the results and my argument that the linkage between government coalition senators and disgust related language was a phenomenon originating in the senate was incorrect, as that is not what is represented by the difference in cosine similarity scores between embedding spaces. To solve this, I removed the gpt-3-embedding model for the project, instead researching possible control models and settling on the Spanish 3B embeddings, a static Word2Vec model pretrained on various forms of political communication.

In parallel, the peer reviews were particularly useful at showing the instances were my writing needed to be polished in order to communicate the ideas that are clear in my head to someone unfamiliar with the method, the research design and setting of the project. The emphasis that both reviews on the opportunity to improve the reading flow and the connections between ideas in the methods section was very useful. I am much more confident now that the current version does a better job at translating my ideas to a new reader. Concrete suggestions, such as including examples of the use of disgust words in close proximity to senators' name did not only help to explain the methods, but are

also a tool that indirectly justifies the relevance of the project (proves the effect I am looking to measure does exists) and aids in the explanation in favor of my operationalization of the senators as a vector of their name.

Finally, the final round of feedback, which was given to first full draft of the paper, was a helpful guide in showing where I was making mistakes in my interpretations of several results and the justification for certain decisions. Key in these was the operationalization of senators as vectors of their names and last names. This has been an operationalization that was created after multiple attempts to capture the meaning given to a senator and I have struggled to put simply why I think it is the best. Seeing various comments doubting it showed the ways in which my explanation was not being sufficiently convincing, thus leading me to change my wording to better justify this decision.

Abstract

This study investigates the use of disgust-related language in official sessions of the Colombian Senate with a focus on the current legislative period following the 2022 presidential elections. The research examines whether demographic characteristics of senators — such as coalition membership, gender, and ethnicity — predict their association with disgust language. Taking a corpus of official transcripts from 20 senate sessions and employing Word2Vec to create semantic embeddings, this analysis identifies a significant semantic relationship between the use of disgust-related terms and certain groups of senators. An ordinary least squares regression model, incorporating senators' demographic characteristics, revealed that coalition affiliation and gender significantly predict the association with disgust language, while ethnicity does not. Notably, female members of the government coalition are especially targeted. These findings show evidence of systematic use of disgust language against historically marginalized groups during senate sessions. The implications of these results for the prevention of political violence, as well as the methodology contributions of this study are discussed. This research deepens the understanding of political discourse dynamics and suggests the need for mechanisms to protect democratic discourse and vulnerable political actors. Future directions for work exploring this phenomenon or adapting the method proposed to other research questions are laid out.

1 Introduction

Disgust-related language has shown to be an effective way to strategically manipulate public opinion for specific political outcomes [Kam and Estes, 2016]. Communication from political elites that is imbued with words that activate disgust sensitivities by evocating feelings of dirtiness and sickness has been shown to be influential in increasing negative attitudes towards historically oppressed outgroups [Casey, 2016, Pascal et al., 2023, Terrizzi et al., 2010] and in reduc-

ing the approval of social welfare policies [Clifford and Piston, 2017]. Moreover, a key element of its effectiveness lies in the asymmetrical power relationship between the political elites weaponizing disgust-related language and their targets. That is, a dynamic in which one group have a wider access to communication channels than the other. In this case, this would manifest as political elites having more access to communication channels than the subjects of disgust-related speech, respectively.

With the increased representation of groups that are traditionally associated with disgust in previously gatekept spaces, it remains to be seen whether the mechanisms of disgust-related political language are maintained when the power imbalance is reduced and a more direct interaction process takes place. Put simply, in contexts where both parties have a symmetric communication relationship, how is language related to disgust used by political elites?

The current senate body in Colombia is a fertile scenario to study disgust-related political communication under these specifications. The 2022 presidential elections in Colombia represented a significant shift for the traditional political landscape of the South American country. Traditionally ruled by governments ideologically situated between the right and the center-right wing, the election of Gustavo Petro and Francia Marquez as president and vice president has led to a new facet for several political phenomena [Millán, 2023]. In the context of political communication, the rise of the *Pacto Histórico* coalition to power has led to a notable change in the political discourse [Valencia Angulo, 2023, Millán, 2023]. This political coalition is characterized by its plurality, as it comprised of several political, ethnic, and sexual minorities, alongside the ideological slant to the left of the current government. This swing towards a more plural and left-leaning agenda has been widely admonished through the language used by opposition groups and traditional news media, which have made racist and classist imagery increasingly prevalent [Vásquez Padilla, 2023, Valencia Angulo, 2023]. In a country like Colombia, where political violence fueled by political discourse have to constant persecution and, in worst.case scenarios, the genocide of political outgroups [Counter, 2023, Martínez Jimenez, 2021] this trend raises concerns about a possible deterioration of political dialogue, with grave implications for critical democratic processes.

This study explores how instances of disgust-related language in the interventions made by senators in official sessions are targeted towards specific groups of senators, as predicted by historical trends of language-based political violence. Specifically, the study addresses whether coalition membership, gender, and ethnicity can be used to predict a close alignment in meaning between a senator and disgust-related language. Using word embeddings to compare this semantic relationship, I show evidence of a consistent pattern of disgust language being attached in meaning to members of the government coalition.

2 Framework

2.1 Emotional language in political communication

It would hardly be an exaggeration to argue that political elites have used emotional cues in their communication for as long as there has been a concept of political communication [Gerstlé and Nai, 2019]. Although the origins and evolution of how political speech has used words, syntactic structures and other patterns of speech related to emotion falls outside the scope of this paper, any study of of the topic would be wise to start by providing evidence not only of the existence of emotionality in political speech, but also of its intentional use by political elites.

One of many examples of this intentionality can be seen in how the use of moral-emotional language changes over the course of an election cycle in a consistent manner. Political elites use a higher degree of emotional language in their speech during primary elections in comparison to general elections [Lipsitz, 2018], which suggests a strategic use of emotionality depending on the specific of the political context. This fluctuation of intensity, frequency and type of emotionality has also been shown to depend on the scenario in which a political elite is speaking (i.e. communication in official settings is less emotionally intense than press releases and interviews) [Gerstlé and Nai, 2019]. Both of these cases show that the emotional profile of political communication is not incidental, but a premeditated choice on the part of a political elite and/or their team. It is key for any analysis of emotion in political communication to keep this in mind, as this fact is what gives validity to an assessment linking emotional language to ideological tendencies and legislative intentions.

Evidence in favor of an intentional and strategic element behind the emotional profile of elite political communication implies that there is an advantage (or the perception of an advantage, at the minimum) that comes with accurately infusing speech with emotional language, using the right combination of intensity, frequency and type of emotion. There is indeed evidence suggesting this might be case, as concrete results have shown that electoral success can be predicted by the ratio between positive and negative emotional responses primed by a candidate's press releases on social media [López-López et al., 2020]. This, however, also allows researchers to look at the emotional cues left by political elites in their communications as a way to infer intention and ideology. As an example, efforts in the field have previously used patterns of speech, and sentiment analysis in the passages of speech containing emotional words to accurately predict voting outcomes in the US Senate [Budhwar et al., 2018].

Furthermore, increasingly granular research designs, with more constrained and discreet operationalizations of emotion language have lead to interesting findings on how particular emotions are used in official contexts [Webster and Albertson, 2022, Dietrich et al., 2019]. Given that policies regarding different political groups and subgroups evoke different discreet emotions [Casey, 2016], it makes sense for any enquiry into emotionality in political speech to focus on topical and relevant emotions. Examinations of the role that different emotions play in po-

litical elite communication have been fruitful, revealing a wide variety of distinct dynamics exclusive to specific emotions [Webster and Albertson, 2022].

As a field, the study of emotional language in political communication has yielded a large and relevant amount of literature on the workings of some discreet emotions (anger and fear, in particular) while leaving questions on others for future researchers to study further. In parallel, most research on emotions in political communication has focused on emotion's influence on electoral politics and adjacent democratic processes. There is a need to expand research into other areas such as social movements and, in the case of this paper, congressional speech. There is interesting and potentially impactful research to be done in this space. For one, emotional activations against minority groups have been shown to occur when the targeted groups have been historically associated with a discreet emotion [Pascal et al., 2023]. These emotional activations have been recorded and theorized to be a causal mechanism behind non-electoral political processes, such as persecution against political minorities [Counter, 2023].

2.2 The role of disgust

Among these discreet emotions, disgust emerges as a prime candidate for explorations into how an under-researched emotion functions in elite political communication. In more general settings, *"Disgust occurs when an individual rejects contact with an object, action or event that is perceived to be impure"* (Kam and Estes, 2016, p.482). Based on what previous literature shows, this is a reaction that can be instrumentalized in political messaging via language-based activations to increase the demand for state protection from outgroups [Kam and Estes, 2016], reduce the public approval minority groups [Casey, 2016, Terrizzi et al., 2010], reduce the approval of policies benefiting them [Clifford and Piston, 2017], and trigger preexisting negative feelings against immigrants [Pascal et al., 2023]. These effects have been shown to be particularly effective when targeted at groups traditionally associated with dirtiness or sickness, and when tied to conservative beliefs and policy stances [Kam and Estes, 2016]

These effects, however, are not a comprehensive list of all the uses that disgust-related political communication may have. As an under-researched emotion in political psychology there is a wide array of political settings where the use of disgust is probably used that are yet to be researched. Even then, language-based disgust activations have been theorized to be effective at influencing political behavior based on how they may be used as a tool or a hurdle in persuasion attempts. Activating feelings of disgust has been shown to reduce the effectiveness of persuasion attempts based on reason or logic, thus *shielding* individuals from convincing contradictory beliefs [Casey, 2016]. In this sense, disgust-related language is potent in cases when partisan individuals are faced with an information processing task where the content they are being exposed to is contradictory to partisan stances [Webster and Albertson, 2022].

Moreover, a central aspect of disgust-related language in political communication - and a key element of its effectiveness - lies in the asymmetrical power relationship between the political elites weaponizing this type of messaging and

their targets. With the increased representation of groups that are traditionally associated with disgust in previously gate-kept spaces, it remains to be seen whether the mechanisms of disgust-related political language are maintained when the power imbalance is reduced and more direct interaction processes take place. In contexts where both parties have a symmetric communication relationship and similar access to communication channels, how is language related to disgust used by political elites?

2.3 NLP analysis of political speech

Although the specific question posed in the last section has not been answered by previous research, there have been similar endeavors in the field of political speech analysis that serve as a source of inspiration for the design of this project, as well as indicators of the potential validity of the methods to be applied. For instance, simple linear models have been used to accurately label senators' party based on their word usage [Bayram et al., 2019], suggesting the words used by congress members in speeches can be insightful indicators of underlying ideological elements. Moreover, patterns of speech have been used to predict votes in legislative decisions of Supreme Court Justices [Budhwar et al., 2018], underscoring how speech can be a reflection of intention even in carefully worded official utterances.

Furthermore, focusing on the validity of studying the emotional charge of political communication, previous research has been able to look at the emotional intensity of political speech to predict voting patterns [Dietrich et al., 2019]. Unfortunately, research like this, at the intersection of natural language processing, political communication and disgust, has been lacking. Nevertheless, research employing sentiment analysis methodologies to explore the online messaging of political elites shows that voting outcomes in presidential elections based on the emotional characteristics of elite communication. These studies collectively support the potential of employing text analysis techniques to study the emotional profile of political speech data, making it the next task to identify a suiting approach for the specific issue.

This project explores how disgust-related language is used in the interventions made by senators in official sessions during the current legislative period. Specifically, the study addresses whether coalition membership, gender, and ethnicity can be used to predict a close alignment in connotation between a senator and disgust-related language. To answer this question, a word embeddings approach surfaces as the prime method to be used to interrogate a corpus of senate speeches.

For operationalization purposes, instances of a senator being semantically linked to disgust will be thought of as proximity between a senator's name and an instance of disgust related speech during an intervention. The most basic case of such an instance would be a sentence in the form of "senator name is disgust related word." This hypothetical syntactic structure represents the simplest instance of a senator being semantically tied to disgust during the session. As seen in the previous example, this task fundamentally requires a

distributional interpretation of the words in the corpus.

The concrete puzzle at hand, figuring out which words from the group "senator" are closely linked to words from the group "disgust" suits word embeddings perfectly, as this method is specifically designed to represent words distributionally [Grimmer et al., 2022]. Furthermore, as a strategy for selecting and representing text for analysis and interpretation, word embedding methodologies have been successfully implemented in the past to identify cultural bias in the shape of class connotations [Kozlowski et al., 2019], with word embeddings applied to political communication being able to output valid and intuitive results by human standards [Rodriguez and Spirling, 2022].

2.4 The Colombian political context and potential targets of disgusts

The particularities of Colombian politics make for a suiting political context to study this dimension of disgust-related political communication. Compared to other countries in the region, Colombian politics have had a higher incidence of ethnic minority political elites [Van Cott, 2005] than other countries in the region. This is a result of various policies designed to increase plurality among the legislative branch. As an example, ethnic minority groups have been given special seats in the senate and the chamber of representatives since 1991, making their presence in the high chambers of the legislative body a constant. Similar policies for farmer, victims of the armed conflict, and other political minority groups have been implemented in the years since. As a consequence, many political, social and ethnic tensions are played out and recorded in congressional sessions.

As an additional point of divergence between the settings where research on emotional language and NLP in political communication gets done, the strength of traditional political parties in Latin American has decreased significantly during the 21st century [Lupu, 2016]. As a result, the amount and relevance of inferences that can be done based on partisanship (both at the political elite and citizen level) has been in decline. Language-based disgust activations have not been studied in a political contexts such as these, creating an initial uncertainty regarding whether emotional cues related to disgust are still effective when not supported by a series of partisan beliefs and sense of identity.

Even then, historical and political dynamics in Colombian politics, coupled with previous research on disgust in political communication, may serve as indicators of which groups are likely to be referred to using language related to disgust political elite messaging. Given the close tie that disgust related language has with political exclusion and political violence [Casey, 2016, Pascal et al., 2023]. Thus, a look at which groups have been victims of these dynamics might clue where disgust language tends to be directed. Three groups were identified in the literature as likely marks: Political minorities [Counter, 2023, Martínez Jimenez, 2021], ethnic minorities [López-López et al., 2020], and women [Martínez Jimenez, 2021]. These cases were selected as in all them, language-based political violence was shown to be a consistent predecessor to political

exclusion, political violence, and, in some cases, genocide. Importantly, there is a non-trivial amount of overlap between the three groups, leading to political dynamics and relationships to political violence specific to the intersections of two or more of these identities.

With this historical backdrop in mind, the current state of Colombian politics serves as a fertile scenario to study how disgust language is used by political elites in contexts where access to communication between them and potential targets of disgust-related language is symmetrical. The 2022 presidential elections represented a significant shift for the traditional political landscape of the South American country. Traditionally ruled by governments ideologically situated between the right wing and the center-right, the election of Gustavo Petro and Francia Marquez as president and vice president has led to a new facet for several political phenomena [Millán, 2023]. In the context of political communication, the rise of the *Pacto Histórico* coalition to power has led to a notable change in the political discourse. This political coalition is characterized by its plurality, as it is comprised of several political, ethnic, and sexual minorities, alongside the ideological slant to the left of the current government. This swing towards a more plural and left-leaning agenda has been widely admonished through the language used by opposition groups and traditional news media, which have made racist and classist imagery increasingly prevalent [López-López et al., 2020, Valencia Angulo, 2023, Vásquez Padilla, 2023].

In a country like Colombia, with a worrying track record of political violence fueled by political discourse against several of the groups represented in the government coalition [Counter, 2023, Martínez Jimenez, 2021], this trend raises concerns about a possible deterioration of political dialogue, with grave implications for critical democratic processes. As with an unfortunate number of research domains, the role of disgust in political communication remains a vastly understudied field in Colombian and Latin American politics. In this sense, a computational approach to disgust related speech targeting previously identified groups would be a relevant contribution to the literature, as it would serve as a starting point for future research on the under-explored intersection between Latin American politics, political communication, and natural language processing. An additional point of relevance for this type of research then pops up, as doing computational political science research on understudied contexts with institutional configurations different from the ones most frequently studied can yield compelling findings that may lead to future research in comparative political science, computational methods, and many of their sub-fields.

Building on the intersection of disgust-related in political communication, NLP approaches to political speech, and the current political conditions of Colombia, this study aims to explore how disgust is used in elite political communication scenarios where these elites and the targets of disgust have a similar access to communication channels, such as in congressional sessions. Additionally, this paper aims to examine Colombia's unique political landscape to understand the impact of disgust-related language on diverse socio-political groups, providing insights into potential targets and consequences of such discourse as well as serving as a demonstration of the relevance of applying NLP

techniques to study political communication in the region. Through a comprehensive synthesis of existing research and empirical analysis, this study study seeks to contribute to the growing body of literature on emotion, language, and politics, all with a computational twist. By deepening the understanding of the complex interplay between emotion and political discourse, this research hopes to provide valuable insights for future academic and policy work.

3 Data

The foundation of the analysis, the text from the senate sessions, was constructed from official records obtained from the Colombian Senate's website. Using a scraper, I downloaded the official records from 20 sessions from the current legislative period (July 2022 to February 2024, when the scrapping was done). The senate sessions records contain a detailed account of the topics discussed within each session, the text of all of the relevant documents, voting and attendance records, and a word-by-word transcript of the debate in the session. For this study, only the latter were looked at. The transcripts contain, verbatim, every intervention done by senators during the debate portion of sessions. Crucially, these documents have no metadata and no formal labelling.

I converted the pdf files to raw text files using the pdfminer version 20191125 python library. Then, I processed each document into clean text via a series of regular expressions. Once the text was preprocessed, I divided each file into normalized and tokenized sentences using the spacy version 3.7.2 library for NLP and the *es_core_news_sm* model for text analysis in Spanish [Honnibal and Montani, 2017]. Despite the efficiency of the automated cleaning process, certain nuances, and inconsistencies inherent in the data required meticulous manual intervention. This supplementary cleaning ensured the removal of anomalies that automated methods could overlook, such as incorrectly formatted text, misidentified sections, and non-textual elements that could skew the analysis. The resulting corpus consisted of 2,606,433 instances of 5263 unique lemmatized tokens.

Next, I implemented another scrapping protocol on a different section of the official senate webpage to gather the official list of 104 senators currently serving. Initially, the scrapping returned a list of senators with their party affiliation. Their information about coalition (government, opposition or independent), gender (male or female) and ethnicity (part of an ethnic minority group or not) was hand labelled based on public records, public statements, and personal webpages for each senator. The process lead to the construction of a database of 104 senators, each labelled with its respective gender (male = 74, female = 30), coalition (government = 29, independent = 40, opposition = 35), and ethnicity (ethnic minority member = 5).

Finally, the list of disgust related words was created by referencing the Spanish Emotional Lexicon (SEL) [Díaz Rangel et al., 2014], a validated dictionary of emotional words in spanish were each word is associated to an emotion. To augment this dictionary and tailor it to the context of political communication,

a few-shot classification task was implemented using the OpenAI GPT - 4 API to generate an expanded list of words. The resulting words were validated by spanish speakers with domain knowledge of political communication in Colombia and added to the list of emotional words. Further validation is given to the list of disgust related language by their close proximity to each other in the embedding space (as seen on Figure 1).

4 Methods

To examine the semantic relationships between senators and disgust related language, I created two sets of embeddings to represent each senator and each disgust word, with a vector mapping them in a multidimensional semantic space. Following a distributional hypothesis for language, which proposes that *"we can learn something about the semantic meaning of word on the basis of the words that appear frequently in close proximity in a small context window around the word"* (Grimmer et al., 2022, p.79), word embeddings convert words in the corpus into unique tokens which are represented by vectors projected in a multidimensional space. The values inside these vectors represent the probability that a word is used near the word in question. As such, operation done with these vectors allow for representations of meaning in the semantic space.

The first set of embeddings was created using a Word2Vec model [Řehůřek and Sojka, 2010] trained on the processed text from the senate sessions. This resulting model contains each of the unique tokens as a vector representing its meaning in the corpus of senate sessions. Following the findings of previous studies on the effectiveness of word embeddings on congressional speech [Rodriguez and Spirling, 2022] the model parameters were set to a context window of 6 words and vectors of 100 dimensions. The Spanish 3B words Word2Vec Embeddings pretrained model [Bilbao-Jayo and Almeida, 2018] was used as a second embedding strategy. Using this model, trained on a larger corpus of political text in Spanish, a list of vectors related to every senator's name and every disgust word in the corpus was obtained. The reason for doing two sets of word embeddings lies in the training data for each. The first Word2Vec model is trained on the corpus data. As such, the semantic space created represents the unique set of meanings and connotations given to words during senate speeches. On the other hand, the Spanish 3B embeddings models are trained on a significantly larger and less specific corpus, leading to vectors representing the meaning of senator names and disgust words in a meaning space well beyond the one created by congressional speech during the period. In other words, the senate Word2Vec model represents the "local" meaning of the relevant strings, while the Spanish 3B embeddings represents their "global" meaning.

In order to answer the questions pertinent to this study, the vectors representing a list of disgust related words (more details in the data section) and the senators were used as the units of analysis. Going back to the operationalization of a senator being semantically linked to disgust in the text produced by the senate sessions, the vector representing each senators' name was calculated by

taking the average of the vectors for the names and last names for each senator. For example, *Aída Yolanda Abella Esquivel*, a senator of the government coalition, was represented by the average vector for the *"Aída"*, *"Yolanda"*, *"Abella"*, and *"Esquivel"* vectors. Under the congressional speech rules that govern the Colombian senate, senators can be refered to by any of their names of last names, which leads to instances such as the following, in which a disgust word (decay) is used in close proximity to last name of a senator: *"(...)the rhythm of the* ***decay*** *brought over by Senator* ***Abella****, has been fast(...)"* or this one, where the first name of Senator Ariel Ávila is used: *"This result is impossible Senator* ***Ariel****, and you are introducing a* ***sickening*** *incentive"*. These two examples represent the many ways in which a senator's name may be used around disgust related speech, validating the need to include this variety in each senator's representing vector.

The average vector of all of the disgust words in the corpus was also calculated to be used as a reference for later analysis. Figure 1 provides a visual representation of the estimated location of these vectors (with the senator name vectors being color coded by coalition membership) in the semantic spaces created by both models. In these dimensionally reduced spaces, the disgust words are closer to the senator vectors in the Word2Vec model than in the Spanish 3B embeddings, with the latter's vectors being more clustered overall.

Using the resulting vectors as inputs, each senator was assigned their corresponding name vectors, one for each of the embedding models. The cosine similarity between these and the average disgust vector was calculated and assigned to each senator as their disgust cosine similarity score. Cosine similarity is chosen as a measure for indicating the semantic distance between two word embedding vectors as capturing the angle between the vectors in the embedding space effectively reflects how close the senators name and the disgusts words are related in meaning, regardless of their magnitude. This approach has been used in previous research comparing the meaning of cultural and political constructs (similar to the symbols behind senators names and disgust related language) [Kozlowski et al., 2019, Rodriguez and Spirling, 2022] to produce valid and interpretable results.

Finally, this data was then used for statistical analysis. First, a two-tailed t-test was be done to determine if the differences in disgust cosine similarity scores from the two sets of embeddings are different, indicating a difference in meaning between the senate sessions semantic space and the larger corpus of the Spanish 3B embeddings training data. Next, OLS regressions were run, taking the disgust cosine similarity scores as the dependent variables with coalition, gender, ethnicity, and relevant interaction terms between the three of them as independent variables. This was done with the objective of determining whether certain groups of senators have a closer association in meaning to disgust related language than others.

5 Results

5.1 Embedding spaces

Figure 1 provides a visual representation of the estimated location of the vectors for the disgust words and senators names (color coded by coalition membership) in the semantic spaces created by both models. An initial look at these dimensionally reduced spaces appears to indicate that the disgust words are closer to the senator vectors in the Senate Word2Vec model than in the Spanish 3B embeddings, with the latter's vectors being more clustered overall.

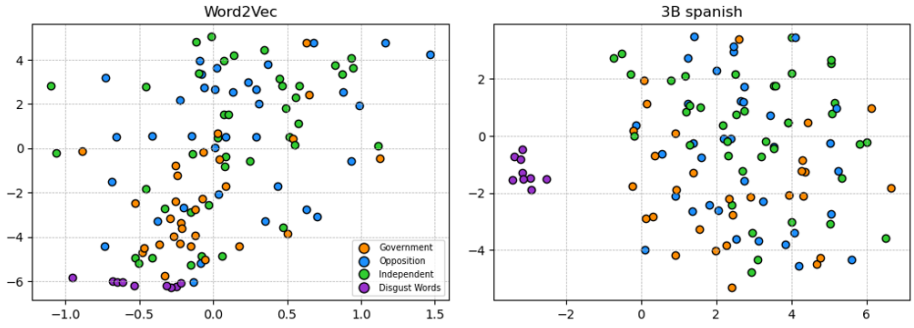

Figure 1: t-SNE dimension reduction scatterplots for senator and disgust vectors

A fair deal of construct validity is given to two both models by looking at the clustering of the disgust word vectors in Figure 1. Out of all of the relevant vectors in the model, these are the ones whose meanings were known before hand and, additionally, we would expect them to be close in semantic distance. Their clustering in the t-SNE plot shows that the vectors are correctly representing meaning in ways that make sense. Furthermore, these result gives validity to the use of a dictionary method to capture instances of disgust in the senate text. Taking these factors into account, the following statistical analysis is strengthened based on the validity of the embedding models which support it.

5.2 Disgust cosine similarity scores

An initial look at both sets of disgusts cosine similarity scores suggests little correlation between the two for each senator. A Pearson's Correlation test ($r(206) = -0.08$, $p > 0,05$) confirms this assessment, suggesting that the relationship between disgust and a senator in the senate sessions is not correlated in any significant way to its meaning in the Spanish 3B training data. Figure 2 shows the distribution of the disgust similarity scores for senators grouped by coalition, gender, and ethnicity for both embedding models. A noticeable difference in the measures is seen, as scores calculated from the Senate Word2Vec model are larger (mean = 0.599, std. dev. = 0.068) than the ones calculated

from the Spanish 3B Embeddings (mean = 0.301, std. dev. = 0.04). A two-tailed t-test comparing this difference in means resulted in a significant value ($t(206) = 43.2$, $p < 0.001$), leading to the conclusion that the semantic association between senators and the disgust words in the corpus is different between the semantic spaces generated by senate sessions speech and the large corpus used by Spanish 3B embeddings.

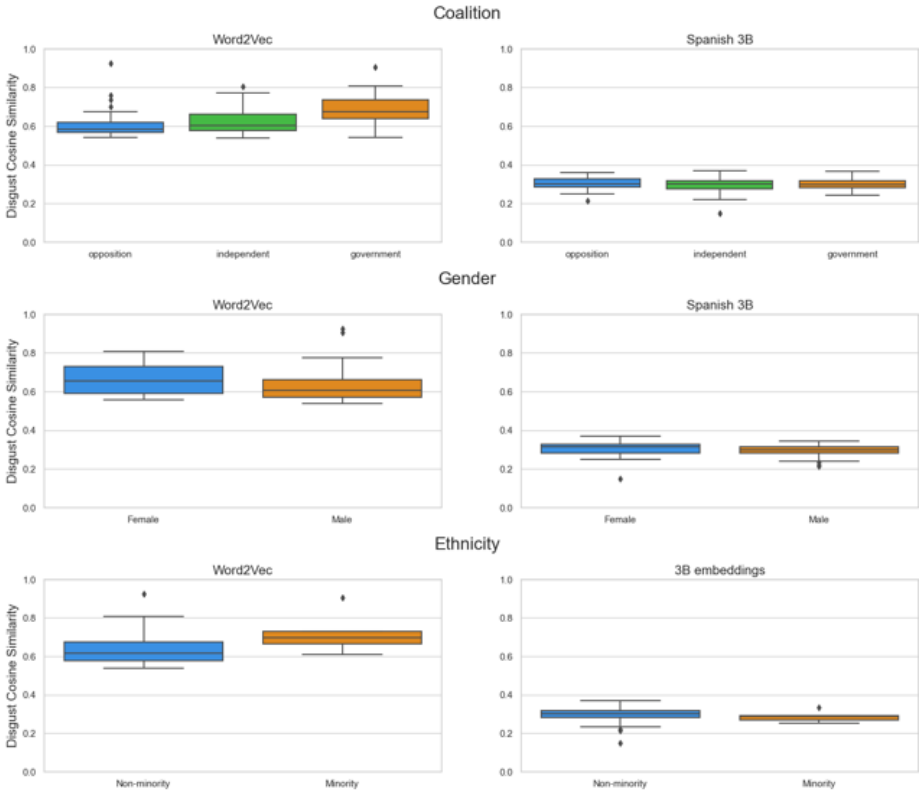

Figure 2: Distribution of disgust cosine similarity scores

5.3 Regression models

Two sets of OLS regression models, one for each word embedding model, were implemented to test whether coalition membership, gender and ethnic identity can predict variance in senators' disgust cosine similarity scores. Figure 3 shows the Q-Q plots for the best fitting regression model for each embedding space. The linear distribution of the residuals in the models serve as evidence of a linear relationship, thus justifying the choice of model.

In the model using the Senate Word2Vec disgust cosine similarity score as the dependent variable, initial analyses yielded a statistically significant model

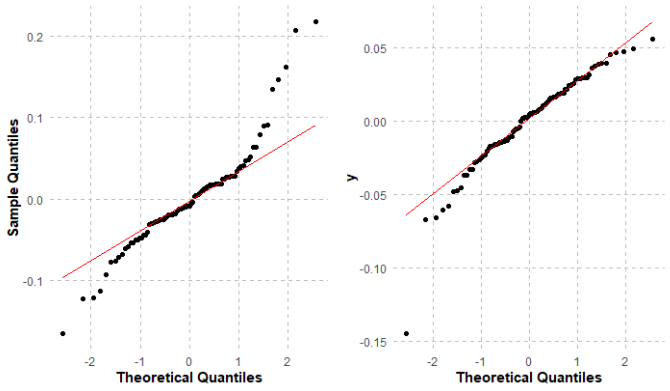

Figure 3: Q-Q plots for regression models using cosine similarity measures from Senate Word2Vec (left) and Spanish 3B Embeddings (right)

($R^2 = 0.127$, $F(4, 94) = 4.6$, $p < 0.01$) revealing that being a part of the independent coalition ($\beta = -0.05$, $p < 0.01$), opposition coalition ($\beta = -0.05$, $p < 0.01$), and gender ($\beta = 0.03$, $p < 0.05$) significantly predict the association with disgust language. This result suggests that female senators and senators from the government coalition tend to be closer in semantic distance to disgust than male senators and senators from the independent and opposition coalition. Senators belonging to an ethnic minority showed no significant variance in their disgust cosine similarity scores.

Next, a series of more complex models were trained and were tested to improve the fit of the model. These contained relevant interactions terms between variables, which were chosen based on the literature on historical predictors of political violence in the country [Herrera and Pertuz Bedoya, 2015, Valencia Angulo, 2023]. Of these, the introduction of an interaction term between gender and coalition was the most effective at doing so. The creation of this term led to a better model ($R^2 = 0.204$), improving the percent of variance explained by being a part of the independent ($\beta = -0.1$, $p < 0.001$) or opposition ($\beta = -0.1$, $p < 0.001$) while identifying being a female member of the government coalition as an additional explainer of variance in the model ($\beta = 0.09$, $p < 0.01$), showing a statistically significant closeness in meaning between their associated vectors and the vectors for words related to disgust. Interestingly, in this model, gender was not a significant predictor of variance in disgust cosine similarity scores by itself, suggesting that the effect that gender has on the semantic distance to disgust-related language becomes significant when tied to other, probably more salient, categories, as is the case with coalition membership in this model.

In contrast, various versions of OLS regression models using the Spanish 3B embedding disgust cosine similarity measure as the dependent variable were implemented. Even thought all of the configurations of independent variables and interactions terms tried with the previous model were applied, no statistically

significant results were obtained. As such, this method outputs evidence that the differences in semantic distance by coalition that are evidenced in the space create by senate speech are not present in the semantic space created by the larger corpus of the Spanish 3B Embeddings training data.

6 Discussion

The influential role that emotional language plays in every level of political communication cannot be overstated. The number of use cases for imbuing political speech with words, syntactic structures, and non-textual elements of communication (such as pitch and/or rhythm) related to emotion is potentially endless. Given that there is evidence showing that the use of this type language is both intentional and strategic [Lipsitz, 2018, Gerstlé and Nai, 2019], the opportunity to develop research on understudied emotions and political context emerges not only as a fruitful academic endeavor, but also an urgent question for any democratic society, for which answers and solutions are a dire need.

It is in this context that this study looks to understand how disgust-related language is used by political elites during official sessions in the Colombian Senate. Previous literature on disgust in political communication has shown its use as a tool for political violence against groups that have been historically associated with disgust [Casey, 2016, Terrizzi et al., 2010, Clifford and Piston, 2017, Pascal et al., 2023]. However, the literature on this emotion is limited when compared to other discreet emotions, such as anger and anxiety [Webster and Albertson, 2022]. As such, there is academic relevance into painting a clear picture of how disgust operates in different situations and settings.

In this paper I interrogated whether the dynamics of disgust-related speech in elite political communication are maintained when access these elites and the disgust-relevant targets have a symmetrical access to communication channels. Due to its political and ethnic characteristic, the current Colombian senate was chosen as a setting for this endeavor. This choice came with the additional advantage of studying disgust in an under-researched political context in the field of emotion in political communication.

The semantic distance between senators and disgust-related language was calculated using two sets of word embedding algorithms and analyzed via a OLS regression, showing that, in the meaning space created by the speech in senate sessions, senators from the government coalition are significant closer in meaning to disgust-related language than those from the independent and opposition coalitions. Within this group, female members of the government coalition were more highly associated to disgust than their male counterparts. Initially, these results show that disgust-related language in political communication persist even in scenarios, such as senate sessions, where access to communication channels is symmetrical.

Digging deeper, these findings suggest that in the current Colombian political context, members of the government coalition could represent one or more outgroups associated with disgust. Historically, political minorities in Colom-

bia have been targets of language-based political violence leading to extreme political exclusion [Martínez Jimenez, 2021] and, in some cases, the genocide of these groups [Counter, 2023]. Furthermore, the current government coalition, which is formed by, among others, some of these political minority groups has already been a target of language-based political violence since its inception as a political front [Valencia Angulo, 2023, Millán, 2023]. This, coupled with the negative emotional reactions that left-leaning governments tend to generate in Latin American public opinion [López-López et al., 2020], confirm that the current government coalition in Colombia is a disgust relevant target. This is a novel finding for the literature on the role of disgust in political communication, as political minorities had not been previously identified as disgust relevant groups. The particular cultural, political and institutional conditions that lead to this semantic tie between disgust and political minorities would be an intriguing question for compared political psychologists advancing this line of research.

Moreover, the fact that women have also been historical targets of political violence in the country [Herrera and Pertuz Bedoya, 2015], coupled with the closer semantic distance between female senators from the government coalition in comparison to male senators from this group, could indicate that certain identitarian categories become disgust relevant when coupled with other, more salient identities. More research into the interplay of disgust-relevant labels in relation to disgust-related political language would certainly be an interesting avenue for future research.

The non-significance of ethnicity in the results of the models is surprising, as ethnic minorities are part of the groups that been shown to be frequent subject of disgust related language in political communication [Pascal et al., 2023] and have been a tragically frequent target of political violence in Colombia [Martínez Jimenez, 2021]. The small number of ethnic minority senators in the sample (5 out of 104) may have something to do with this result. Making it hard to categorically affirm that ethnic minority senators are *not* frequent targets of disgust related language. As it stands, future work studying the specific link between ethnicity and disgust related political communication may be useful at reinterpreting the results of this project.

The notion that disgust-related language in political communication is not merely a reflection of societal prejudices but a deliberate strategy employed by opposition and political elites is underscored in light of the non-significant Spanish 3B Embeddings model results, which show no deferential relationship between senators from different coalitions. The distinction in the significance of the regression models coming from different embedding models indicates that the phenomenon under investigation is inherently linked to the specific context of senate sessions, where the senate Word2Vec model was trained, rather than being an extrapolation of external semantic relationships. This delineation is critical, as it emphasizes the specificity of disgust language as a tool wielded by political actors, aligning with the broader literature on political violence in Colombia [Herrera and Pertuz Bedoya, 2015, Martínez Jimenez, 2021, Millán, 2023].

More urgently worrying is the implication that historical dynamics of vi-

olence against political minorities and women are currently being reproduced and fueled in the senate. Going back to the dynamics of disgust in political communication, this emotion has been shown be physiologically associated with dirtiness and sickness, activating demands of protection tied to cleansing [Kam and Estes, 2016]. This consideration, tied with the intentional and strategic nature of emotional language in political communication [Lipsitz, 2018, Gerstlé and Nai, 2019] serve as troubling signs of a potential but tragically familiar strategy against political minorities in the country: othering and extermination [Counter, 2023]. This insight is particularly relevant against the backdrop of Colombia's tumultuous history with political violence, highlighting an urgent need for mechanisms that uphold the integrity of democratic discourse and protect those in vulnerable political positions

Despite the grave implications of the findings, the successful application of word embeddings to analyze political speech marks a significant methodological contribution to the field. Echoing the methodologies of previous studies that scrutinized political behavior [Bayram et al., 2019, Budhwar et al., 2018, Dietrich et al., 2019] and semantic connotations [Kozlowski et al., 2019, Rodriguez and Spirling, 2022] using text as data, the approach used in this project demonstrates the efficacy of word embeddings as a computationally inexpensive and insightful tool for dissecting political attitudes and emotional language in congressional speech. Additional elements of the methodological design of this study, such as the use of validated emotional language dictionaries to label congressional speech, show promise based on their success on this paper. An easily adaptable and relatively user-friendly workflow for analyzing the emotional characteristics of congressional speech in Latin American contexts can be derived from the methods described in this paper and may serve to accelerate research and facilitate collaborations in the intersecting areas of natural language processing, political communication, and Latin American politics.

6.1 Limitations

The method used in this project does have some limitations that are important to keep track of when analyzing the results and extrapolating the findings to real world scenarios. First, the calculation of the senator vector in the Word2Vec model (the centroid taken from their names and last names) could have a stronger theoretical standing. Alternative methods (only using last names) failed to capture vectors for a significant group of senators and yielded less interpretable results in the models ($R^2 = 0.003$, with no significant coefficients in the best model), so the original operationalization was maintained. Replications of this project using alternative strategies to represent senators in the semantic space are welcomed and could potentially be beneficial for the validity of the results presented here and their interpretation.

An important caveat to this interpretation lies in nature of word embeddings and methods using cosine similarity to compare meaning. The results show that, in the senate speech corpus, senators from the government coalition are significantly closer in meaning to disgust than other senators. This however

is not a causal claim and the result may well be mediated by an element not observed in the design of the study. Still, this consideration does not obscure the findings or the historical precedence that supports it but it does highlight the need for future exploration that increase the nuance of this findings and the academic understanding of the role of disgust related language in political communication.

Additionally, taking only the senators as the subjects of the study leads to having a small number of observations for the analysis. This heavily restricts statistical analysis and the types of modelling that can be done. This problem could be solved in future explorations of this topic by adding other members of congress such as members of the chamber of representatives or senators from previous periods to produce more robust statistical findings.

6.2 Future directions

This project adds a new element to previously developed workflows for disgust language analysis by switching the focus from speakers to subjects of text, being able to draw conclusions about senators' attitudes towards each other. A word embeddings approach based on this methodology is particularly useful at detecting the meaning that is given to senators and the vast array of symbols they represent as elected officials. A more granular definition of disgust, dividing it into elements relating to dirtiness and sickness, its essential components [Kam and Estes, 2016], could be used to validate and gain a deeper insight into the results of this project. Moreover, replacing disgust with other constructs such as class, sexuality, and masculinity (among others) could lead to interesting results as to how meaning around those topics is constructed and assigned to senators in congressional sessions.

The model itself could be more complex to detect facets of the social game that current methodology is not able to capture. Particularly interesting in this case would be time. Adding a temporal element to the analysis may allow for the information collected via these methods to answer an additional set of questions. Inquiries into differences in semantic relationships between groups of senators and disgust related language across difference legislative and presidential periods, for one, would be particularly relevant. Apart from time, expanding the type of congressional sessions from text is scrapped, and keeping track of it, could yield interesting information about differences in semantic relationship in different institutional bodies, which would in turn result in information about the unique dynamics of, for example, the senate vs. the chamber of representatives.

Finally, a causal model using the measures created for this project would be extremely good fit for the overall aims of this research agenda. The disgust similarity score of each senator could be used as predictor, mediator, and confounder in a wide variety of cases such as evaluating the impact of disgust related language on the public perception and electoral success of senators, assessing its role in shaping policy debates and legislative outcomes, or understanding the dynamics of coalition formation and maintenance within the Senate, among

many others. These approaches, all based on the workflow designed for this project, could realistically lead to deeper insights into the mechanisms of political discourse and power dynamics within the Colombian Senate and other institutional bodies.

7 Conclusion

Disgust related language has been shown to frequently target historically oppressed groups, reducing the approval of policies trying to break this cycle. By using word embeddings to measure the semantic distance between senators and disgust related language in congressional speech, this study evidences how weaponizations of disgust are still present even when communications asymmetries between political elites and targeted groups are reduced. Additionally, by switching the analytic focus from speakers to targets of disgust related speech, and situating this endeavor in the Colombian context, political minorities where identified as a disgust relevant group, and effect which may be increased when associated with other disgust relevant categories.

Furthermore, the differences in semantic distances between the two embedding models suggests that instances of disgust related language in the Colombian senate are a phenomenon specific to the senate, which carries worrying implications for the health of political plurality and the safety of political minority leaders in the country.

Although worrying, having obtained valid and intuitively sound results works in favor of using word embeddings to analyze political speech. The method employed in this study offers a powerful and computationally inexpensive tool for detecting and measuring political attitudes, providing a workflow for analyzing ways in which senators assign meaning to each other. Beyond their academic value, the insights gleaned from this study hold profound implications for policymakers, civil society actors, and anyone invested in fostering a more inclusive, respectful, and violence-free political dialogue in Colombia and potentially in similar contexts globally.

8 Data Availability

The datasets, coding notebooks and replication materials for this paper are found in the project's GitHub Repository.

References

[Bayram et al., 2019] Bayram, U., Pestian, J., Santel, D., and Minai, A. A. (2019). What's in a Word? Detecting Partisan Affiliation from Word Use in Congressional Speeches. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

[Bilbao-Jayo and Almeida, 2018] Bilbao-Jayo, A. and Almeida, A. (2018). Automatic political discourse analysis with multi-scale convolutional neural networks and contextual data. *International Journal of Distributed Sensor Networks*, 14(11):1550147718811827. Publisher: SAGE Publications.

[Budhwar et al., 2018] Budhwar, A., Kuboi, T., Dekhtyar, A., and Khosmood, F. (2018). Predicting the Vote Using Legislative Speech. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, pages 1–10, Delft The Netherlands. ACM.

[Casey, 2016] Casey, L. (2016). *The Politics of Disgust: Public Opinion Toward LGBTQ People & Policies*. Thesis, University of Michigan, Ann Arbor.

[Clifford and Piston, 2017] Clifford, S. and Piston, S. (2017). Explaining Public Support for Counterproductive Homelessness Policy: The Role of Disgust. *Political Behavior*, 39(2):503–525.

[Counter, 2023] Counter, M. (2023). Genocide in "Common Parlance": Grief and the Political Persecution of Colombia's Patriotic Union Party. *Journal of Genocide Research*, 0(0):1–20.

[Dietrich et al., 2019] Dietrich, B. J., Enos, R. D., and Sen, M. (2019). Emotional Arousal Predicts Voting on the U.S. Supreme Court. *Political Analysis*, 27(2):237–243. Publisher: Cambridge University Press.

[Díaz Rangel et al., 2014] Díaz Rangel, I., Sidorov, G., and Suárez Guerra, S. (2014). Creación y evaluación de un diccionario marcado con emociones y ponderado para el español. *Onomázein Revista de lingüística, filología y traducción*, 29:31–46.

[Gerstlé and Nai, 2019] Gerstlé, J. and Nai, A. (2019). Negativity, emotionality and populist rhetoric in election campaigns worldwide, and their effects on media attention and electoral success. *European Journal of Communication*, 34(4):410–444. Publisher: SAGE Publications Ltd.

[Grimmer et al., 2022] Grimmer, J., Roberts, M. E., and Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

[Herrera and Pertuz Bedoya, 2015] Herrera, M. C. and Pertuz Bedoya, C. (2015). Narrativas femeninas del conflicto armado y la violencia política en Colombia: contar para rehacerse. *Revista de Estudios Sociales*, (53):150–162. Number: 53 Publisher: Universidad de los Andes.

[Honnibal and Montani, 2017] Honnibal, M. and Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

[Kam and Estes, 2016] Kam, C. D. and Estes, B. A. (2016). Disgust Sensitivity and Public Demand for Protection. *The Journal of Politics*, 78(2):481–496. Publisher: The University of Chicago Press.

[Kozlowski et al., 2019] Kozlowski, A. C., Taddy, M., and Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, 84(5):905–949. Publisher: SAGE Publications Inc.

[Lipsitz, 2018] Lipsitz, K. (2018). Playing with Emotions: The Effect of Moral Appeals in Elite Rhetoric. *Political Behavior*, 40(1):57–78.

[Lupu, 2016] Lupu, N. (2016). *Party brands in crisis: Partisanship, brand dilution, and the breakdown of political parties in Latin America*. Cambridge University Press.

[López-López et al., 2020] López-López, P. C., Oñate, P., and Rocha, (2020). Social media mining, debate and feelings: digital public opinion's reaction in five presidential elections in Latin America. *Cluster Computing*, 23(3):1875–1886.

[Martínez Jimenez, 2021] Martínez Jimenez, K. (2021). Genocidal practices and crimes against humanity in Colombia: the case of the Unión Patriótica (Patriotic Union) and social leaders in the post-agreement. *Revista de la Facultad de Derecho y Ciencias Políticas*, 51(135):489–516.

[Millán, 2023] Millán, J. (2023). Colombia: una nueva gramática del poder. *Nueva Sociedad*, (303):98–105. Num Pages: 98-105 Place: Caracas, Argentina Publisher: Fundacion Foro Nueva Sociedad.

[Pascal et al., 2023] Pascal, E., Holman, A. C., and Miluț, F. M. (2023). Emotional relevance and prejudice: testing the differentiated effect of incidental disgust on prejudice towards ethnic minorities. *Frontiers in psychology*, 14:1177263.

[Rodriguez and Spirling, 2022] Rodriguez, P. L. and Spirling, A. (2022). Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research. *The Journal of Politics*, 84(1):101–115.

[Terrizzi et al., 2010] Terrizzi, J. A., Shook, N. J., and Ventis, W. L. (2010). Disgust: A predictor of social conservatism and prejudicial attitudes toward homosexuals. *Personality and Individual Differences*, 49(6):587–592.

[Valencia Angulo, 2023] Valencia Angulo, L. E. (2023). ¡Negro(a) hijueputa! In *Debates y reflexiones contemporáneas sobre el racismo en el suroccidente de Colombia*, pages 93–150. Universidad Icesi, 1 edition.

[Van Cott, 2005] Van Cott, D. L. (2005). Building inclusive democracies: Indigenous peoples and ethnic minorities in Latin America. *Democratization*, 12(5):820–837. Publisher: Routledge _eprint: https://doi.org/10.1080/13510340500322215.

[Vásquez Padilla, 2023] Vásquez Padilla, D. H. (2023). Chasing racists, protecting racism: revisiting anti-discrimination law in Colombia. *Latin American and Caribbean Ethnic Studies*, 0(0):1–20.

[Webster and Albertson, 2022] Webster, S. W. and Albertson, B. (2022). Emotion and Politics: Noncognitive Psychological Biases in Public Opinion. *Annual Review of Political Science*, 25(1):401–418.

[Řehůřek and Sojka, 2010] Řehůřek, R. and Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA.