Evolution of Political Discourse: Analysing Thematic Trends in Spanish Investiture Speeches (1979-2023)

Sergi Margalef Comallonga

Abstract: *This research explores the thematic evolution of investiture speeches in Spain from 1979 to 2023, employing quantitative text analysis and topic modelling. By analysing speeches delivered during important political moments, the research identifies significant shifts in thematic priorities, reflecting evolving political dynamics. Results reveal a notable increase in discussions related to governance and institutions in recent years, indicative of changing policy focus. However, findings regarding economic themes, the Constitution, and climate change diverge from initial hypotheses. Despite some limitations, this study contributes to understanding Spain's political landscape and offers insights into the evolution of key themes in investiture speeches over time.*

1. Introduction

Over the past five decades, Spain has experienced significant political, social and economic changes after nearly four decades of Francoist dictatorship. The so-called Transition to democracy in 1978 marked an important turning point in Spain's recent history and, since then, the country has evolved and faced different challenges. From the consolidation of the democratic system and the attempted coup in 1981 by General Tejero, to the war against the terrorist organization ETA, the 2008 global financial crisis, and the rise of the Catalan independence movement. Political leader's speeches can provide a wealth of information about the constantly changing political discourse and policy agenda, especially investiture speeches to become the next president of the country. That is the reason why this research seeks to answer the following question:

- What are the key themes and how have they evolved in investiture speeches in Spain over time?

Understanding the thematic evolution of investiture speeches can provide insight into the shifting political dynamics in Spain and the issues that have shaped the country's democratic trajectory. By identifying and analysing the dominant themes present in the investiture speeches I intend to answer the research question and also compare the evolution over.

The dependent variable in this study is the identified themes from the investiture speeches, these themes represent the evolution of the political discourse and indicate salient issues and priorities addressed by political leaders. The dataset created to answer the question covers all the investiture speeches from 1979 to 2023. To analyse the thematic evolution of investiture speeches I use topic modelling with the R package keyATM (Eshima et al., 2024), this tool can uncover trends and patterns that may not be apparent through traditional qualitative analysis alone. To validate the topic modelling a dictionary analysis has been conducted.

2. Literature Review and Theory

The vast quantity of textual data that political figures produce nowadays is enormous and keeping track of all of it seems like an impossible task (Han & Lim, 2021), that is why quantitative text analysis can be a powerful tool for navigating this sea of information. However, there are political speeches that gain more attention because of their relevance, such as inaugural addresses and the State of the Union in the United States, as well as investiture speeches in Spain. Previous research has delved into the significance of Presidential Inaugural speeches in the United States (Huen, 2011), acknowledging their symbolic nature and their role in introducing key themes and policies (Biria & Mohammadi, 2012; Light, 2014).

The investiture speech in Spain occurs following an election when a candidate, named by the King, presents their candidacy to assume the role of President of the Spanish Government. This speech seeks the confidence of the Congress of Deputies, as outlined in Article 99 of the Spanish Constitution. The candidate addresses the chamber, and indirectly all the citizens, to present their program for the next legislature. Should the candidate fail to secure majority support from the Congress, they have the option to address the chamber once more, 48 hours later, for a duration of 10 minutes, per Article 171.5 of the Standing Orders of the Congress. In this second vote, they need a simple majority.

Several studies have explored investiture speeches in Spain, from works that focus on one concrete speech and the subsequent debate (Fernandez-Ulloa & López Ruiz, 2021) to others that include a larger corpus. For example, the work by Quevedo-Redondo (2023) includes 16 speeches spanning from 1979 to 2022 and explores the existence of rhetoric of equality, women's issues, ethics of care, and the pragmatic-empathic communicative pattern traits. In contrast, the present study extends its scope and focuses on a broader number of topics to identify the key themes and their evolution over time. Fernández Castrillo (2016) adopts a more general approach to identifying themes and analyses if these are similar to the most important issues that concern Spanish citizens based on surveys conducted by the Spanish Centre for Sociological Research. One underlying finding from this study is that the economy is prevalent in all speeches and that the themes match citizens' most important issues However, this analysis is qualitative, and the application of quantitative text analysis and topic modelling could offer deeper insights into Spanish investiture speeches. A more quantitative approach is used by Álvarez Esteban (2011) analysing 10 investiture speeches from 1979 to 2008, focusing on statistical analyses of lexical patterns.

The objective of this research is to contribute to the existing literature and offer new insight into investiture speeches in Spain. Central to this objective is the use of quantitative text analysis and topic models, offering a nuanced understanding of speech content that traditional qualitative methods may overlook. Another relevant fact is that I use the most up-to-date corpus, which includes all the investiture speeches. Additionally, the previous research corpus did not include speeches from candidates when they addressed the chamber for the second time. The inclusion of second-vote speeches is motivated by the hypothesis that candidates, constrained by time, strategically tailor their discourse to emphasise aspects related to institutional dynamics, political actors, and governance, distinct from the broader thematic range covered in initial addresses.

Building upon Fernández Castrillo's (2016) observations regarding the prominence of economic discourse, I hypothesise that the economy will be prevalent in all speeches, but more in times of crisis, such as 2008 and the more recent caused by Covid-19 and the war in Ukraine. Two other topics that I will focus on are the Constitution and Climate Change. Due to the transition to democracy, I theorise that the Constitution is a topic relevant during the first years, diminishing over time before resurging in relevance with the rise of the Catalan independence movement.

Similarly, Climate Change is anticipated to emerge as a progressively prominent topic, reflective of growing societal concerns and environmental imperatives.

Furthermore, ideological orientations of governing parties—centre-left (Socialist Party) and centre-right (Popular Party and Union of the Democratic Centre)—are posited as influential determinants shaping speech content. A comparative analysis expects centre-left parties to prioritize social welfare themes, contrasting with centre-right parties' emphasis on economic matters.

The expectations that I have are the following:

> H1: Second-vote speeches prioritise institutional, political, and governance themes.
>
> H2: Economic discourse intensifies during economic crises.
>
> H3: The Constitution is an important topic during the first years of democracy and gains importance again with the Catalan independence movement.
>
> H4: Climate Change gains prominence over time.
>
> H5: Centre-left parties focus on social welfare topics more than centre-right parties that talk more about the economy.

In addition to these central hypotheses, I must consider external events, which may not be fully captured by the topic models. Furthermore, while quantitative text analysis offers valuable insights, it simplifies political discourse's complexity and may overlook contextual factors. Additionally, the proposed linear evolution of thematic priorities may oversimplify political dynamics.

3. Methodology

3.1. Data

The corpus used for this research consists of an original dataset comprising all investiture speeches in Spain, from the first one, the 30th of March 1979, until the most recent, the 15th of November 2023. During this period, 26 investiture speeches have been given. In total, without any pre-processing the number of tokens accends to 224.354. Table 1 presents a breakdown of the speeches and token lengths categorised by the three political parties that have governed.

Table 1: Descriptive Statistics

		Length in tokens		
	Number of speeches	Minimum	Mean	Maximum
Socialist Party	14	1110	9592.857	18205
Popular Party	9	1404	7448.556	12760
Union of the Democratic Centre	3	489	7672.333	13180
Total	26	489	8629	18205

The collection of the speeches was carried out meticulously through the Open Data website of the Congress of Deputies. Despite being a time-consuming task, this manual collection process was essential to guaranteeing the dataset's standardization and suitability for analysis. Moreover, it ensured the inclusion of all investiture speeches, enhancing the completeness and reliability of the data.

3.2. Analysis

3.2.1.   *Preprocessing*

The data was cleaned and pre-processed using the quanteda package (Benoit et al., 2018), this involved removing stopwords, symbols and infrequent terms. Additionally, different multi-word expressions, such as *European Union* and party names, were compounded into a single token for a more coherent analysis. As the topic modelling process was fine-tuned, words that did not add any value to the analysis were removed.

3.2.2.   *Topic Modelling*

To analyse the thematic trends in Spanish investiture speeches with topic models I have used the R package keyATM (Eshima et al., 2024). This tool offers distinct advantages over alternatives like stm (Roberts et al., 2019), primarily by building on quanteda and boasting user-friendly functionality and comprehensive documentation. Initially, I intended to run a dynamic topic model because it takes into account time trends; however, the results and interpretability of the base model were better than the dynamic model.

Another advantage of Keyword-Assisted Topic Models is that you can provide a small number of keywords that improve the performance (Eshima et al., 2024). The final number of topics has been the result of trail-and-error. At the beginning, I started with 20 topics, the same number of issues that the Spanish Centre for Sociological Research identifies. Later I reduced the number to 12 keyword topics and 3 non-keywords. The results that test the hypothesis of the topic model contain the 12 identified topics.

3.2.3.   *Validation: Dictionary Method*

A dictionary comprising the 12 identified topics was created. The main purpose of the dictionary method was to validate the topic model. The words chosen to be included in the dictionary are the same used for the keywords in the topic model. This approach allowed to cross-reference the identified topics with the predefined set of keywords. The list of the dictionary can be found in Appendix A.

4. Results

Table 2 shows the 12 labelled topics identified with the keyATM topic model. These have been identified thanks to the keywords that were generated based on the most important issues according to Spanish Center for Sociological Research. In total, they identified 20 but some of them overlapped and it was summarised. I tested between 3 and 8 unlabelled topics to see if I could come up with extra topics, however, none yielded satisfactory results for finding new themes. Appendix B offers a comprehensive table detailing the 12 identified topics, their respective percentages, and the most frequent terms associated with each.

Table 2: Labelled topics

1. Economy	2. Employment	3. Welfare
4. Health	5. International	6. Climate
7. Constitution	8. Territorial	9. Public
10. Institutions	11. Government	12. Rights

With the results of this model, I have plotted their frequency over time and the results can be seen in Figure 1. Although it is difficult to extract conclusions and test our hypotheses from this first plot, one noticeable observation is the increase in the proportion of discussions related to "Institutions" and "Government" during the past decade. This phenomenon could be attributed to the fact that since 2016 of the 13 speeches only 3 have ended in successful investitures, resulting in a change of policy focus to more governance and representative rhetoric.
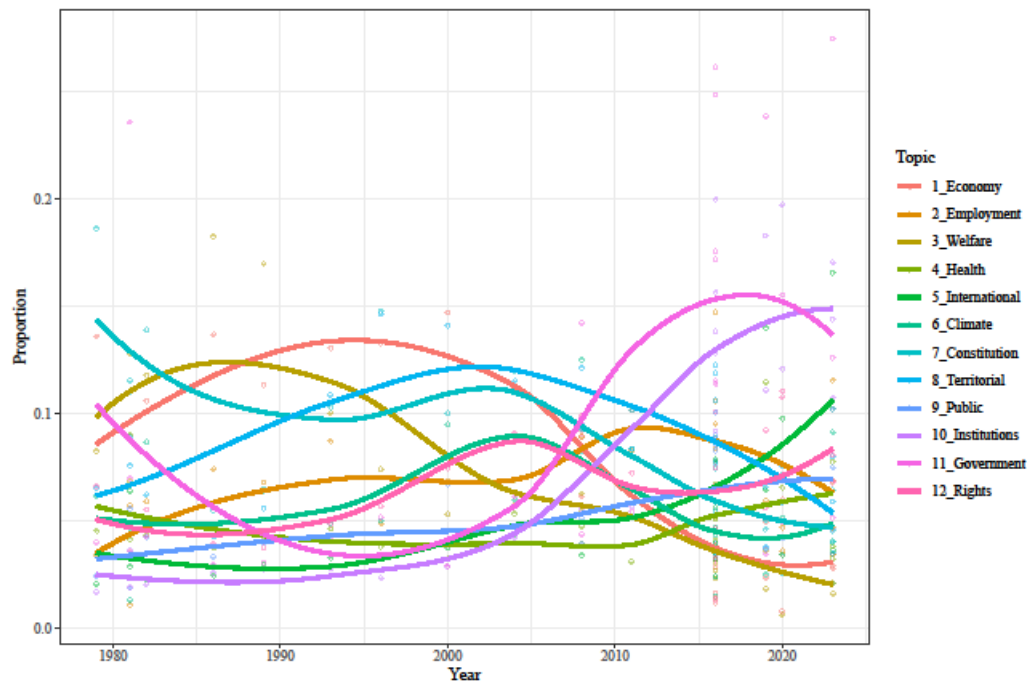

Figure 1: Plot showing topics evolution over time

To test the first hypothesis we can look at the mean proportion of institutions and government topics for first-vote speech and second-vote. With these results, in Table 3, we can see that the mean of these topics is higher in second-vote speeches as I have theorised. To be sure of the results I have also run a linear regression model, with the topics as the dependent variable and the type of vote as the independent variable.

Table 3: Mean proportion of institutions and government topics for vote

	Mean Institutions	Mean Government
First Vote	6.48	6.45
Second Vote	14.38	21.68

The results are in Table 4, the coefficients and the p-values confirm that second-vote speeches prioritise institution and government themes.

Table 4: Institutions and government regression models

	Inst.	Gov.
Intercept	0.06^{***}	0.06^{***}
	(0.01)	(0.01)
Second Vote	0.08^{**}	0.15^{***}
	(0.02)	(0.02)
R^2	0.37	0.74
Adj. R^2	0.34	0.73
Num. obs.	26	26

$^{***}p < 0.001;\ ^{**}p < 0.01;\ ^{*}p < 0.05$

To test the second hypothesis about economic discourse I have plotted the results of economy and employment topics, the result is Figure 2. Surprisingly, the topic proportion of Economy is relatively high before 2000, and then it decreases. Employment has a different evolution with an increase during the 2008 economic crisis. In consequence, the second hypothesis cannot be validated due to the fact that the results are rather inconclusive.
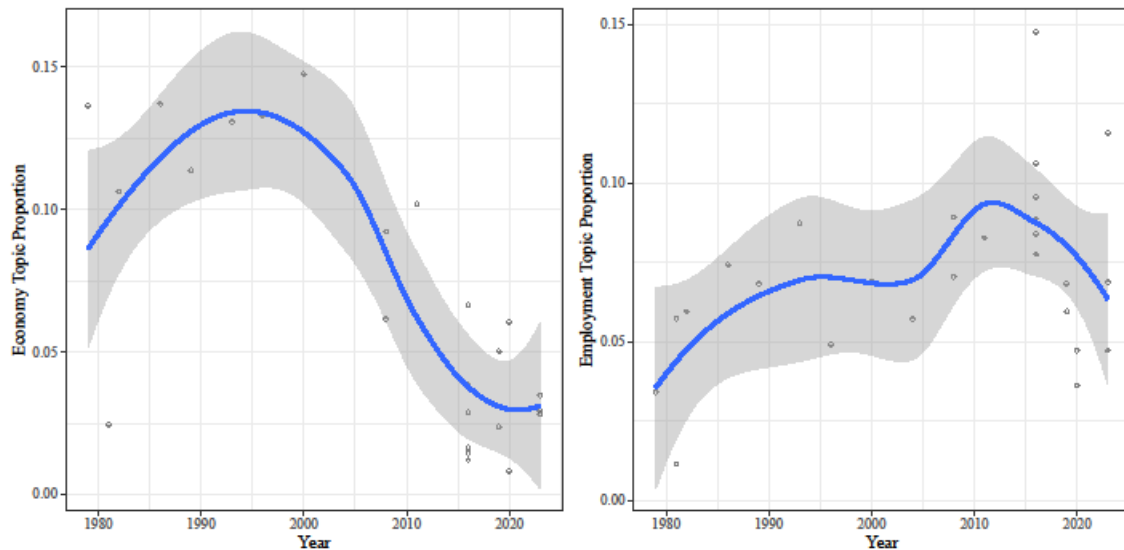

Figure 2: Plot showing economic topics evolution

The evolution of the topic Constitution can be seen in Figure 3. The hypothesis related to this topic said that the proportion would be high in the first years and with the rise of the Catalan independence movement. The plot shows that during the first years, there is a high topic proportion of the Constitution, but we do not see an increase in the topic after 2012 when the Catalan independence movement gained popularity.
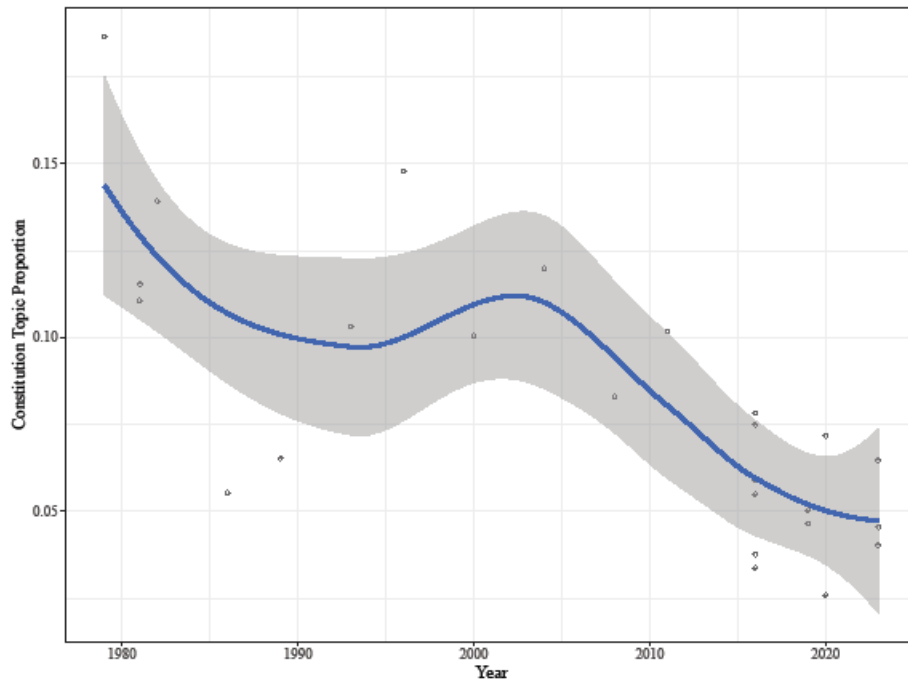
6

Figure 3: Plot showing Constitution topic evolution

Climate change has been a growing concern during the last decades, becoming a worldwide theme set in the policy agenda. That is the reason why I theorised that it would gain proportion of investiture speeches over time. From the results in Figure 4, we can see an increase in the topic from 2000 to 2010 but this increase has not been sustained after 2010. One reason could be that a lot of second-vote speeches have happened since then, and as we have seen the focus has shifted to governance and institutions topics, leaving less proportion to other topics such as Climate. With these results, we cannot confirm the fourth hypothesis.
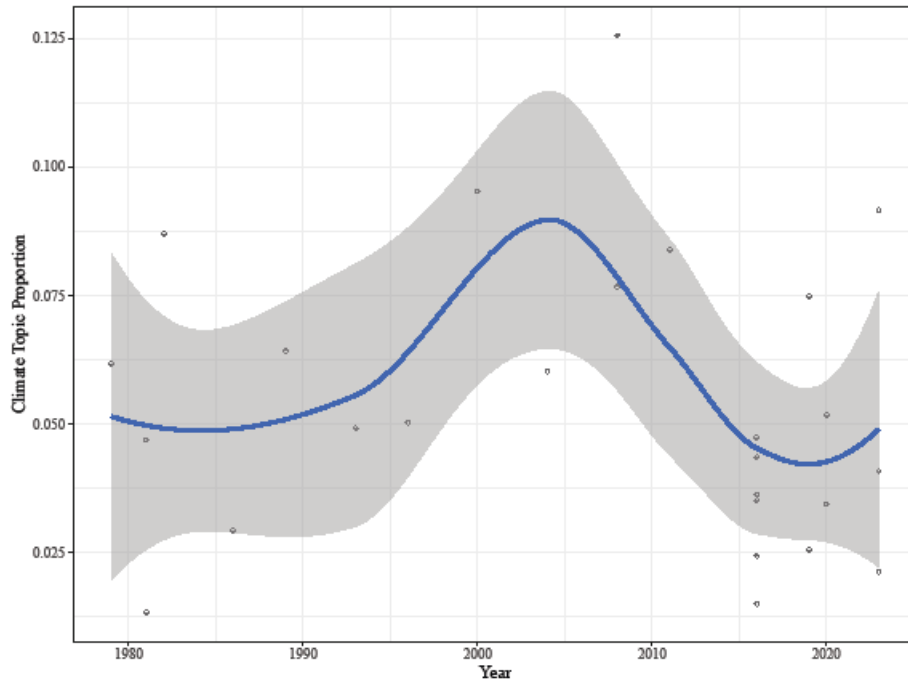

Figure 4: Plot showing Climate topic evolution

The last hypothesis to test was that the centre-left party would focus more on social welfare themes (Welfare and Rights), whereas the centre-right parties would focus on economic topics (Economy and Employment). In Appendix C a table can be found with the mean for each topic according to ideology. Table 5 shows the regression models that test this hypothesis. From the coefficients and the p-values only Rights are statistically significant, with the direction I expected. However, the general results cannot confirm the hypothesis.

Table 5: Economic and social welfare regression models

	Econ.	Emp.	Welfare	Rights
Intercept	0.07^{***}	0.07^{***}	0.07^{***}	0.08^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
Centre Right	0.00	0.01	-0.01	-0.03^{**}
	(0.02)	(0.01)	(0.02)	(0.01)
R^2	0.00	0.04	0.01	0.27
Adj. R^2	-0.04	0.00	-0.03	0.24
Num. obs.	26	26	26	26

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

To validate the results of the topic modelling I have done a dictionary analysis with the same topics to see if the frequent words identified by the topic model corresponded with the same ones as the dictionary. In Figure 5 and Figure 6, we can see the results of the evolution of each topic over time. Except for Welfare, International and Public the shape of the graph is similar among both methods.
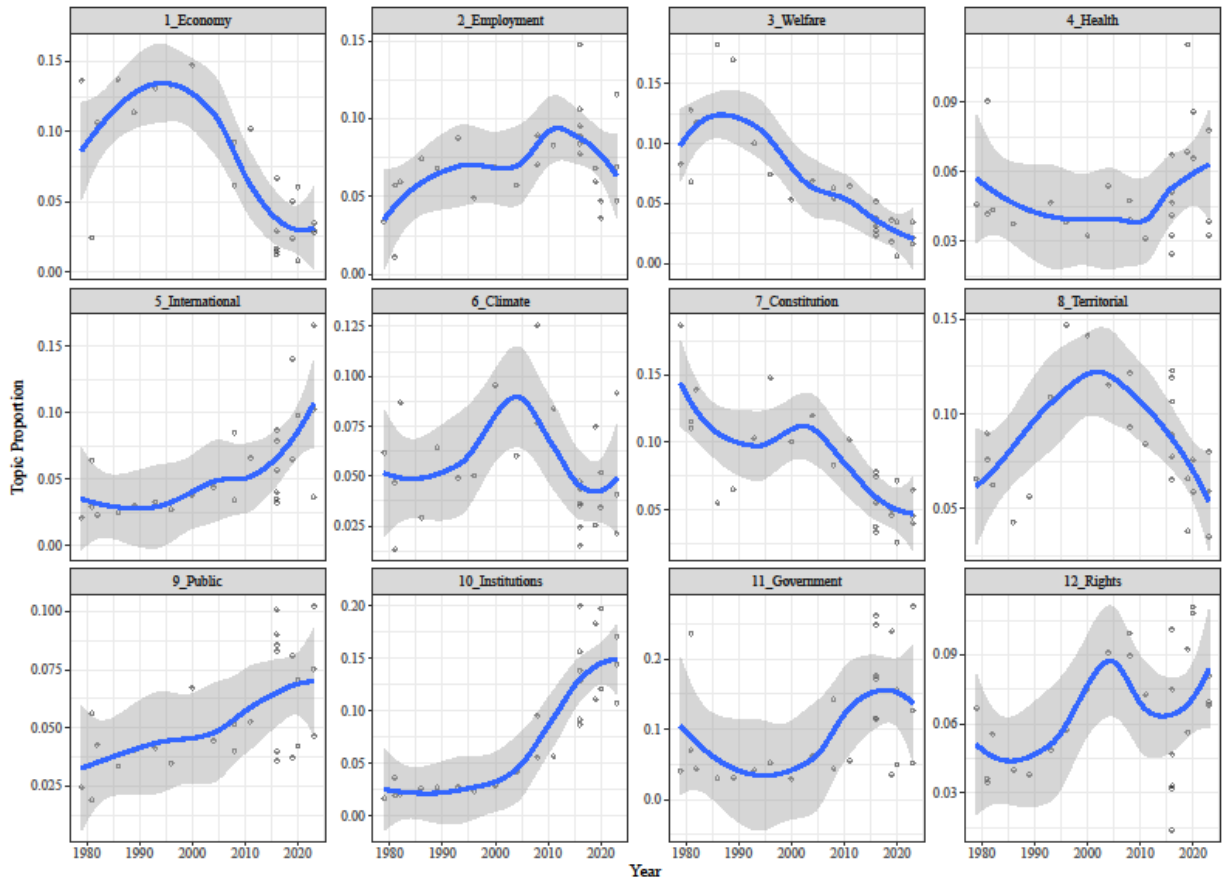

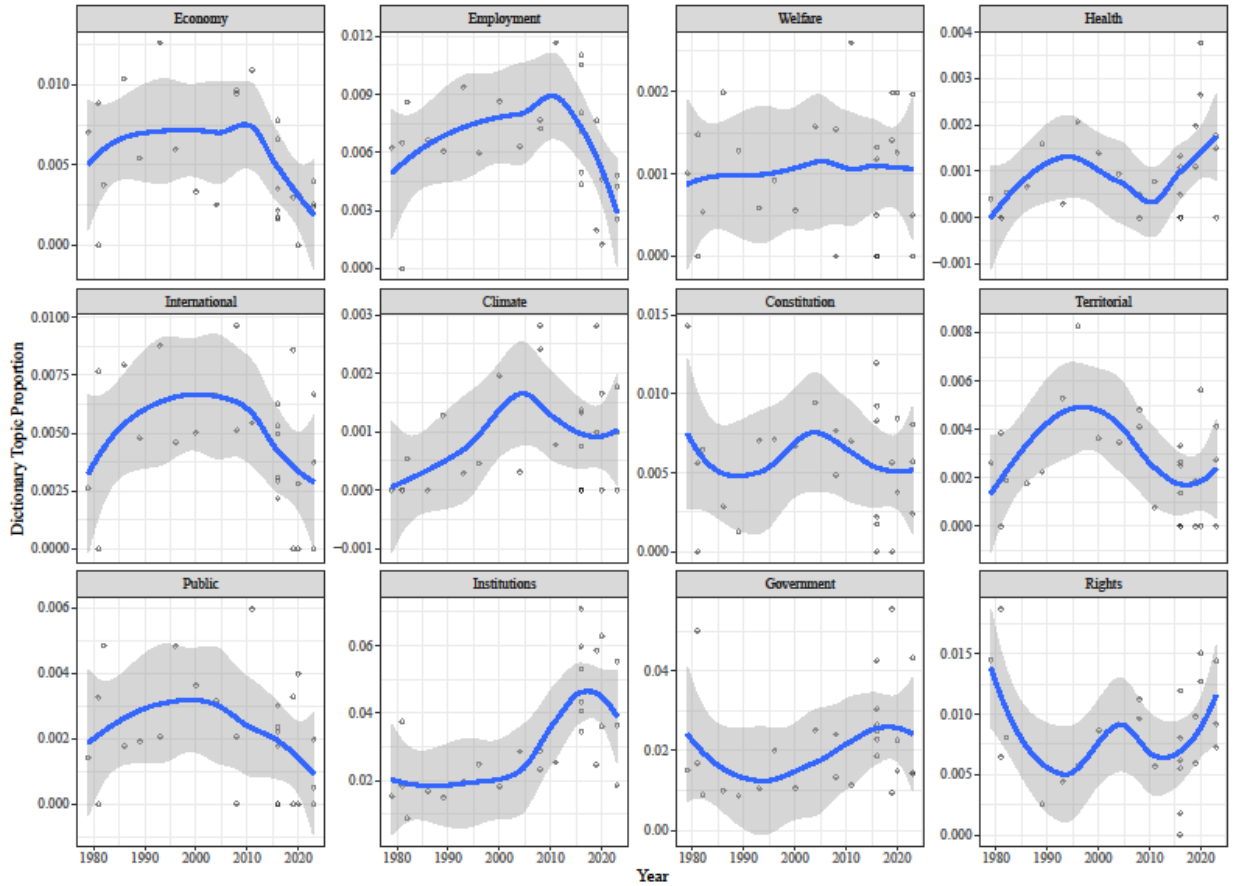Figure 5: Plot showing topic model evolution

Figure 6: Plot showing dictionary evolution

After reviewing the validation results and before preceding the final section, I have some thoughts on this research and the methodologies employed. The inspiration for creating the corpus came from a similar dataset found in the quanteda package (Benoit et al., 2018), which focuses on the inaugural speeches of US presidents. Admittedly, a problem with my corpus is its relatively modest size, which raises questions about the suitability of topic modelling. Nevertheless, the attempt to apply innovative and cutting-edge methodologies serves as a valuable exercise, offering insights into both the strengths and limitations of such approaches.

5. Conclusion

The primary focus of this study was to investigate the evolution of key themes in investiture speeches in Spain over time. This research aimed to answer the question: "What are the key themes and how have they evolved in investiture speeches in Spain over time?" by examining the thematic content of investiture speeches spanning from 1979 to 2023.

The analysis revealed several significant findings regarding the thematic trends in investiture speeches. Firstly, there was a notable increase in discussions related to "Institutions" and "Government" during the past decade, which could be attributed to the recent challenges faced by political leaders in securing majority support for investitures. This shift underscores the evolving policy focus towards governance and representative rhetoric in response to changing political dynamics, also confirming the first hypothesis.

Regarding the other hypothesis, they cannot be confirmed. I did not find the economy to be a prevalent theme and the increase during economic crisis. The evolution of "Constitution" and "Climate" was not the one I had hypothesised, the first one not increasing the proportion with the Catalan independent movement and the second not increasing constantly over time. The analysis exploring ideological differences between centre-left and centre-right parties could not yield statistically significant results.

The validation of the topic modelling through a dictionary analysis provided further insights into the consistency and reliability of the identified topics. Despite some discrepancies and probably not the most suitable methodology. However, this study contributes to the existing literature by offering a comprehensive analysis of the thematic evolution in Spanish investiture speeches and with the creation of this corpus different methods and questions can be further researched.

References

Álvarez Esteban, R. (2011). El discurso de investidura como instrumento de comunicación política en España (1979-2008). *aDResearch: Revista Internacional de Investigación en Comunicación, 4*, 28-48. https://doi.org/10.7263/ADRESIC-004-06

Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A (2018). "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software*, 3(30), 774. doi:10.21105/joss.00774, https://quanteda.io.

Biria, R., & Mohammadi, A. (2012). The socio pragmatic functions of inaugural speech: A critical discourse analysis approach. *Journal of Pragmatics, 44(10),* 1290–1302. https://doi.org/10.1016/j.pragma.2012.05.013

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55(4)*, 77–84. https://doi.org/10.1145/2133806.2133826

Congress of Deputies. (1982). Standing Orders of the Congress of Deputies. Retrieved from https://www.boe.es/buscar/act.php?id=BOE-A-1982-5196

Congress of Deputies. (2024). *Open Data website of the Congress of Deputies*. Retrieved from https://www.congreso.es/en/datos-abiertos

Constitution of Spain. (1978). Retrieved from https://www.boe.es/buscar/act.php?id=BOE-A-1978-31229

Eshima, S., Imai, K., & Sasaki, T. (2024). Keyword-Assisted Topic Models. *American Journal of Political Science*, *68(2)*, 730–750. https://doi.org/10.1111/ajps.12779

Fernández Castrillo, E. (2016). El discurso político en España. Un análisis comparado de los discursos de investidura (1979-2011). *Universidad de Valladolid. Facultad de Filosofía y Letras*. Retrieved from http://uvadoc.uva.es/handle/10324/21700

Fernandez-Ulloa, T., & López Ruiz, M. D. C. (2021). La argumentación en los discursos de Pedro Sánchez y Pablo Casado en el debate de investidura (España, 4 de enero de 2020). *Círculo de Lingüística Aplicada a la Comunicación,* 86, 201–214. https://doi.org/10.5209/clac.70608

H. Wickham. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Han, H., & Lim, S. (2021). Inaugural Speech Classification with Named Entities and Key Phrases. *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 110–113. https://doi.org/10.1109/BigComp51126.2021.00029

Huen, K. L. B. (2011). Text analysis with a systemic functional approach: Analyzing US presidential inaugural address (Outstanding Academic Papers by Students (OAPS)). Retrieved from City University of Hong Kong, CityU Institutional Repository.

Leifeld P (2013). "texreg: Conversion of Statistical Model Output in R to LaTeX and HTML Tables." *Journal of Statistical Software*, 55(8), 1–24. https://doi.org/10.18637/jss.v055.i08.

Light, R. (2014). From Words to Networks and Back: Digital Text, Computational Social Science, and the Case of Presidential Inaugural Addresses. *Social Currents, 1(2),* 111–129. https://doi.org/10.1177/2329496514524543

Pedersen T (2024). *patchwork: The Composer of Plots*. R package version 1.2.0.9000, https://github.com/thomasp85/patchwork, https://patchwork.data-imaginist.com.

Quevedo Redondo, R. (2023). La feminización de la retórica en discursos de investidura: El caso español. *Revista Prisma Social,* (40), 213–237. Retrieved from https://revistaprismasocial.es/article/view/4874

Roberts, M.E., Stewart. B.M., Tingley, D. (2019). Stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, 91(2), 1-40. https://doi.org/10.18637/jss.v091.i02

Spanish Center for Sociological Research. (2024). https://www.cis.es

Watanabe, K., & Zhou, Y. (2022). Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches. *Social Science Computer Review, 40(2)*, 346–366. https://doi.org/10.1177/0894439320907027

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation.* R package version 1.1.4, https://github.com/tidyverse/dplyr, https://dplyr.tidyverse.org.

Wickham H, Vaughan D, Girlich M (2024*). tidyr: Tidy Messy Data*. R package version 1.3.1, https://github.com/tidyverse/tidyr, https://tidyr.tidyverse.org.

AI Tools	
GitHub Copilot	During the process of coding, Copilot was activated and it helped with predicting the code and being more efficient.
ChatGPT	It helped to improve the figures and when I needed help coding. For example, to display the topics in the order I wanted.
Grammarly	It was useful to check grammar mistakes and improvement in some structure suggestions.
QuillBot	It helped with checking grammar and spelling mistakes.

Appendix A

Keywords	
Economy	crisis económica, crecimiento económica, economía, política económica, económico
Employment	empleo, paro, trabajo, desempleo
Welfare	estado bienestar, salario mínimo, servicios públicos, seguridad social
Health	sanidad, salud, vacunas, covid, sanitario
International	unión europea, política exterior, internacional, europa
Climate	cambio climático, medio ambiente, sostenibilidad, contaminación
Constitution	constitución, constitucional, ley, esatdo
Territorial	comunidades autónomas, financiación autonómica, territorial, autonomía
Public	administraciones públicas, sector público, poderes públicos, público, servicio público
Institutions	confianza cámara, parlamento, congreso, gobierno, cámara, señorías
Government	presidente, gobierno, ministro, poder
Rights	derechos, libertad, igualdad, justicia, oportunidad

Appendix B

Topic Label	Percentage	Most Frequent Words
Economy	6.94	sociedad, desarrollo, seguridad, económica, acción
Employment	7.12	ciudadanos, empleo, políticas, economía, trabajo
Welfare	6.17	países, sólo, esfuerzo, proceso, crecimiento
Health	5.14	ley, lugar, plan, lucha, impulsar
International	5.93	país, europa, pacto, millones, mujeres
Climate	5.33	legislatura, sistema, mundo, educación, cuatro
Constitution	8.10	constitución, reforma, medidas, sociales, objetivo
Territorial	8.45	gobierno, quiero, futuro, mejor, nuevo
Public	5.52	todas, cada, debemos, partido popular, público
Institutions	8.91	señorías, cámara, personas, gobierno, mayoría
Government	11.13	gobierno, acuerdo, bien, señor, diputados
Rights	6.50	social, derechos, diálogo, igualdad, justicia

13

Appendix C

	Mean Economy	Mean Employment	Mean Welfare	Mean Rights
Centre Right	6.94	7.71	5.58	5.04
Centre Left	6.93	6.62	6.67	7.76