

An Exploratory Content Analysis of the Annual Presidential Speeches of Peru, 1956 - 2019 *

Carla Solis Uehara
csuehara@uchicago.edu

Luis Eduardo San Martin
lsanmartin@uchicago.edu

March 25, 2020

Abstract

We use content analysis on a corpora containing the annual presidential speeches of Peru. We restrict our sample to the range of years from 1956 to 2019. The main hypothesis we try to test is whether the results obtained go in line with the analysis and work that political scientists have drawn from Peru. For this purpose, we rely on four exploratory methods: count of words and phrases, network analysis, clustering and word embeddings. We find that our results correlate to the findings of political scientists about Peru during this period. These results set a promising path in investigating the discovered trends further.

Keywords: *presidential speeches; exploratory content analysis; politics of Peru.*

1 Introduction

Peru has gone a long way to consolidate itself as a democratic nation, which has been the rule for the last 20 years. Before that, a tumultuous path was the dominant. Proof of it is the variety in governance directions it has had, even until today: It returned to be a democracy in the late 1950's, after going through a period of militarized governance, to become a dictatorship from 1969 to 1980. After that, even though there was another comeback to democracy, from 1980 to 1990, with Fernando Belaunde and Alan Garcia; Alberto Fujimori's dictatorship ruled for 10 years (1990-2000). The last 20 years has meant a period of sound economic growth and a slow consolidation of the population's trust in the prevailing democratic system.

All of the major plans that these leaders had for Peru and all what they accomplished during the previous year are usually reflected in the annual presidential speech (Sánchez 2002c). Generally speaking, the language these corpora uses is positive and enthusiastic about all what happened during the last year, and filled with hope about what will come ahead. Presidents usually center the discourse in the people and may often reach populist tones. Money and investment amounts are usually a big subject as well.

*This is the Research Appendix document of our final project for the course SOCI 40133 - Computational Content Analysis at the University of Chicago. The project open repository is: <https://github.com/ccsuehara/Peruvian-presidents-speeches>

This work uses content analysis tools to obtain social and political insights from these speeches and contrast them with the main ideas that political scientists have theorized about these presidents. We analyze the period 1956-2019, encompassing 64 speeches, 11 presidents who gave a presidential speech¹, and 13 different administrations where the president gave a speech. Table 1 shows the period of each president and administration in this time span.

Table 1: Peruvian Presidents and Administrations:1956-2019

Administration	President	Period
Prado	Manuel Prado	1956-1961
Lindley	Nicolas Lindley	1962
Belaunde (1)	Fernando Belaunde Terry	1963-1968
Velasco	Juan Velasco Alvarado	1968-1975
Morales Bermudez	Francisco Morales Bermudez	1975-1980
Belaunde (2)	Fernando Belaunde Terry	1980-1985
Garcia (1)	Alan Garcia Perez	1985-1990
Fujimori (1)	Alberto Fujimori Fujimori	1990-1995
Fujimori (2)	Alberto Fujimori Fujimori	1995-2000
Toledo	Alejandro Toledo Manrique	2001-2006
Garcia (2)	Alan Garcia Perez	2006-2011
Humala	Ollanta Humala Tasso	2011-2016
Kuczynski / Vizcarra	Pedro Pablo Kuczynski	2016-2018
Kuczynski / Vizcarra	Martin Vizcarra Cornejo	2018-Present

The main importance of this work relies on its novelty: political scientists who study Peru usually use comparative and (non-computational) qualitative methods in their works. The present explorations will open a path for the use of computational content analysis techniques in Peru's political analysis.

The rest of this document is distributed as follows: section 2 describes how we got our corpora and the data preparation processes we conducted, section 3 describes the method-

1. We only consider presidents who gave a presidential speech in its classical annual venue and date: the Congress of Peru on July 28 of each year

ologies our analysis used, section 4 presents the results of each methodology, and section 5 concludes.

2 Data

We used data from the website of the Congress of Peru, where every presidential speech for the period of analysis is published in the form of a PDF file. To get the data, we built a script to download every speech in our period of analysis, transform them into TXT files, and save the result in separate files for each speech.

Then, the data preparation process followed these steps:

1. Removal of unnecessary characters remaining from the PDF to TXT transformation.

We used the Python property of strings `replace` and the Regular Expressions Python library for this.

2. Word tokenization, followed by word normalization. We used the `word_tokenization` function from the `lucem_illud.2020` package for the word tokenization part, and the word normalization model of `spaCy` (in Spanish) for word normalization.

3. Sentence tokenization, followed by word tokenization within sentences, and word normalization. We used the `sent_tokenize` function of the `nltk` package for sentence tokenizing. The word tokenization and normalization used the same functions described in 2.

For the sake of convenience, the result was stored in a data frame with columns containing the final and intermediate results of 1, 2 and 3, along with the raw TXT and PDF data and some metadata about the speeches, like the year they were delivered and the presidential administrations they belonged to. We have 64 observations in total.

3 Methodology

Our methodology follows an exploratory approach for content analysis applied to our corpora, the Annual Presidential Speeches of Peru from 1956 to 2019. Specifically, we apply a number of computational content analysis techniques to extract historical and political insights from the speeches.

The techniques we apply are the following:

1. Counting Words and Phrases
2. Networks
3. Clustering
4. Word embeddings

We justify this exploratory approach in two reasons: the novelty of the application of these techniques in political science research with presidential information for the Peruvian context, and the novelty in the use of these corpora.

4 Results

4.1 Counting Words and Phrases

This section will take the normalized words and treat them as if they were variables. Specifically, it will analyze the statistical properties of the most frequent ones. First, we make a recount of the number of words used by each of the presidents. Then, we will review the most important words, plotted through a word cloud, and draw some conclusions on the words that tend to dominate the word space. Next, we will graph the most popular words, in an approximation of what are the central topics of each administration. Finally, we will use measures of distance to see how "far" the language each presidents use are located to each other.

For this section, we rely on the use of the `altair` and `matplotlib` libraries for visualizing our results, as well as the `wordcloud` dependency from `nltk` for plotting our word cloud.

Figure 1 shows the distribution of the word count each president uses by year. Among the presidents with the longest speeches are Manuel Prado (11,000 words on average, mainly due to the fact that in 1957 he gave a speech of 212 pages, which was practically a report of all political and economic activities that happened during his first year of government), Belaunde (5,300 words on average) and Alan Garcia (5,200 words on average). The presidents who use less words are, besides Lindley (445 words), Kuczynski (1,918 words

on average) and Fujimori (3,300 words on average). Coincidentally, these last two presidents didn't get to finish their presidency periods, due to same reason: They were involved in widely known acts of corruption.

Figure 1: Normalized words per speech

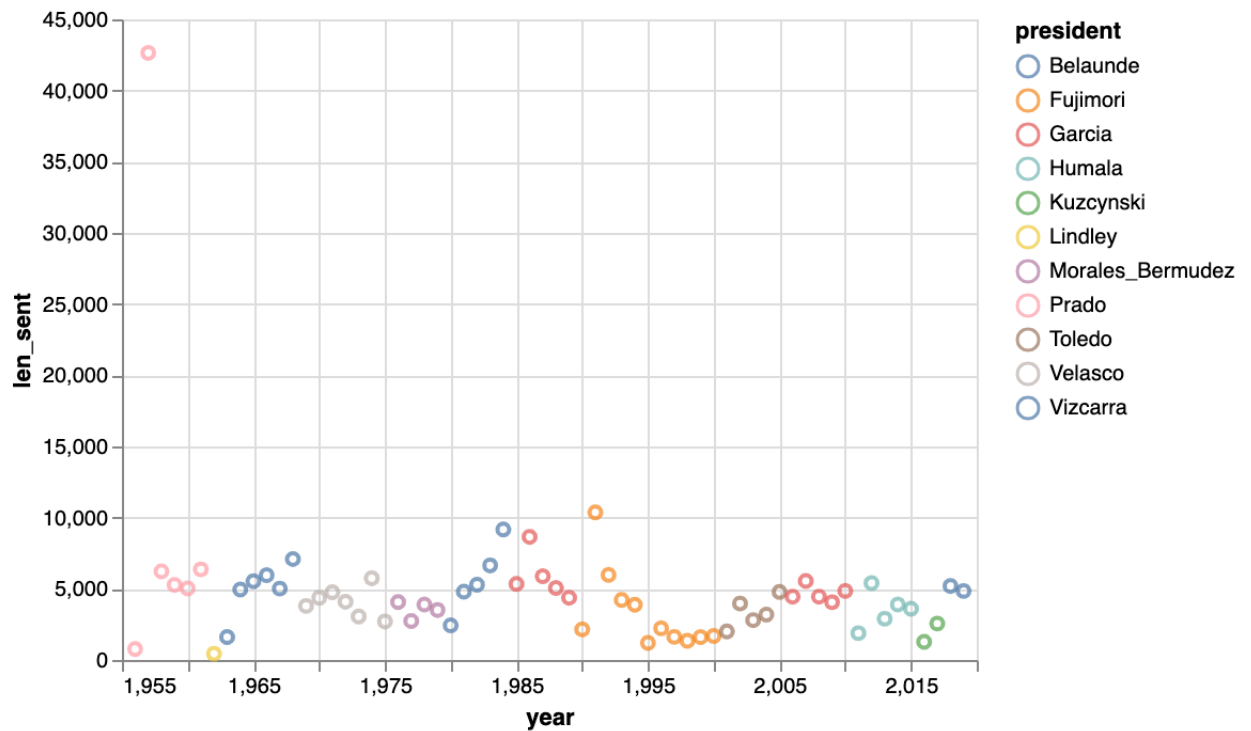


Figure 2 shows the word cloud that stresses what are the major topics that these presidents have in common. The main concern that is central to each of these leaders is the country itself (*país* and *Perú* are consistently ranked amongst the first for each of these distributions), also, the time span of these speeches, which are annual (*año*). Govern and politic are also important terms. Many verbs that refer to *building-out* dominate the landscape: program, develop, allow, form, start and achieve, are present.

investment(*inversion*). President Kuczynski follows suit with the emphasis put in new projects (*proyectar, proyecto*). Finally, we get to president Vizcarra. Vizcarra was not elected, he was Kuczynski's vice president and took office after he resigned to avoid being impeached by the Congress because of acts of corruption. This might be why corruption (*corrupcion*) is a key word in his speeches.

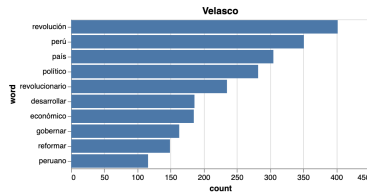


Figure 3: Juan Velasco

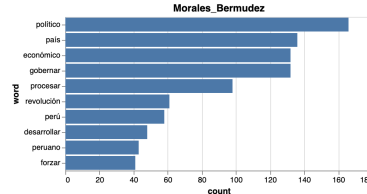


Figure 4: F. Morales B.

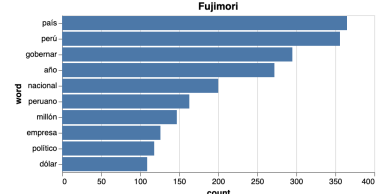


Figure 5: Alberto Fujimori

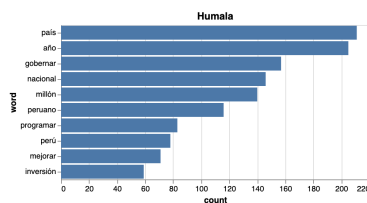


Figure 6: Ollanta Humala

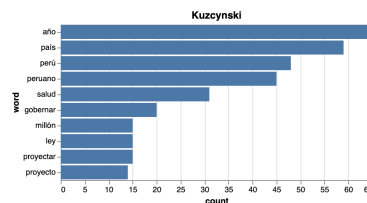


Figure 7: Pedro P. Kuczynski

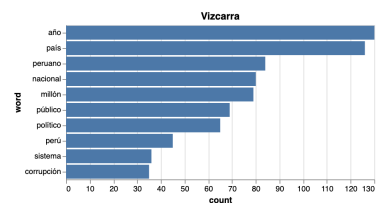


Figure 8: Martin Vizcarra

Next, we analyze how "close" each speech is from each other. For this purpose, we use the Kullback-Leibler divergence measure, and we project it into two dimensions. We restrict the analysis to the years that came after the military coup, in 1968. The following graph shows that the KL measure gives us a grouping of speeches, based on the vector-space that was formed based on the words. In the bottom center there is a cluster of them that belonged to the decades of the 1960's and the 1970's, which coincide with the military dictatorships. Also, we see at the center another different group that go from the 1980's to the 1990's, which correspond to the democratically elected leaders, but still had different priorities in mind with respect to the third group, which are the presidents that come from 2001 and beyond.

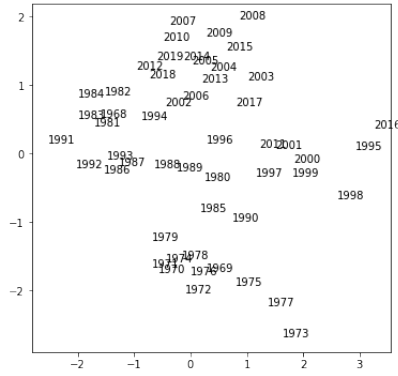


Figure 9: Multidimensional scaling of the KL

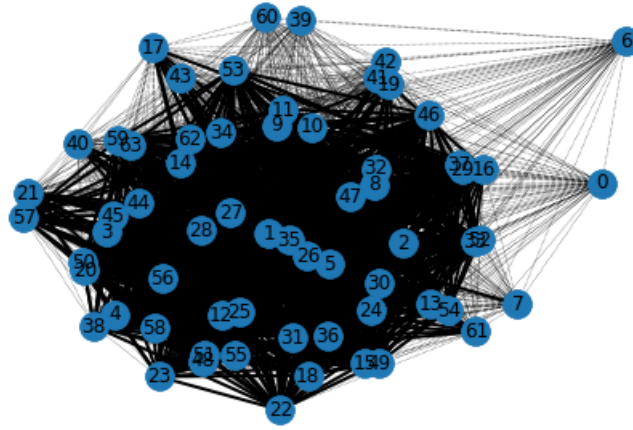
4.2 Networks

In this section we will take a look at how the language used by the presidents locate them in an network spectrum. We will get to appreciate how intertwined their speeches are and how relate to each other.

At the center of the network, node 1 reflects the 212 pages long that President Prado gave in 1957. It resembles more of a report that gave account, in high detail, of all the actions the government undertook, described in a very technical way. Very next to it, node 18, that corresponds to Velasco's penultimate year, is, surprisingly, at the center. The most distant ones are node 6, or the 1962, delivered by General Lindley, which was an atypical year.

On the periphery, nodes 0, 34, 39 and 60, coincidentally, correspond to the first year of mandate of or 1956, was the first year of presidency for Manuel Prado, Alberto Fujimori (two periods) and Pedro P. Kuczynski. This analysis helps us reinforce the idea that first years of mandate are usually atypical from the rest of the period.

Figure 10: Network representation of speeches



4.3 Clustering

With clustering we explore the associations and patterns present in the relative word frequency of each speech. Specifically, we would like to assess if a cluster analysis can predict speeches that come from the same president or administration, and if speeches within the same cluster signal similarities in political discourses or closeness in the left-right political spectrum. Though this section mainly focuses on the clustering results, it also consists of an analysis of the two principal components of the TF-IDF projection of each speech from our corpora.

For this end, we will conduct a K-means cluster analysis of the speeches using the TF-IDF word projection of the normalized words of each speech. We will generate from 3 to 10 clusters and evaluate which one fits better with the corpora we have. We will also complement this analysis with a Silhouette score calculation for each different clustering. For the TF-IDF word projection, clusters generation and Silhouette analysis, we relied on the `TfidfVectorizer`, `KMeans`, `silhouette_score` and `silhouette_samples` functions of the `sklearn` package.

To visualize the result, we will also generate the first two components of the Principal Component Analysis (PCA) of the TF-IDF matrix for each speech. Then we will do a scatter plot of each speech and its position according to these two principal components to see how are the clusters distributed. This result is shown in Figure 11 and Figure 12. We will use the `PCA` function of the `sklearn` package for this part.

Table 2: Average Silhouette Score result for each clustering classification

Number of Clusters	Avg. Silhouette Score
3	0.103
4	0.094
5	0.089
6	0.091
7	0.072
8	0.078
9	0.079
10	0.088

The Silhouette Analysis results show that the cluster classification with the highest score is 3, among all the number of clusters generated. Importantly, the average score seems to decrease as the number of clusters used increase. Thus, we will interpret our clustering results using a classification of three clusters. Furthermore, the scatter plots of Figure 11 and Figure 12 show that there is a clear agglomeration of speeches in three big groups after we get the first two principal components of their TF-IDF projections.

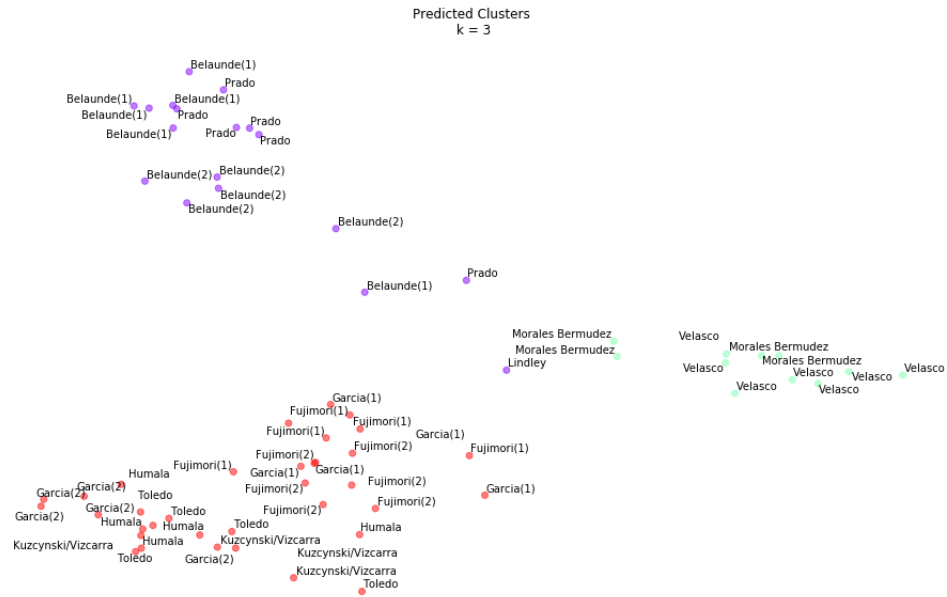


Figure 11: Clustering classification projected in the PCA space of the two principal components, labeled by administration

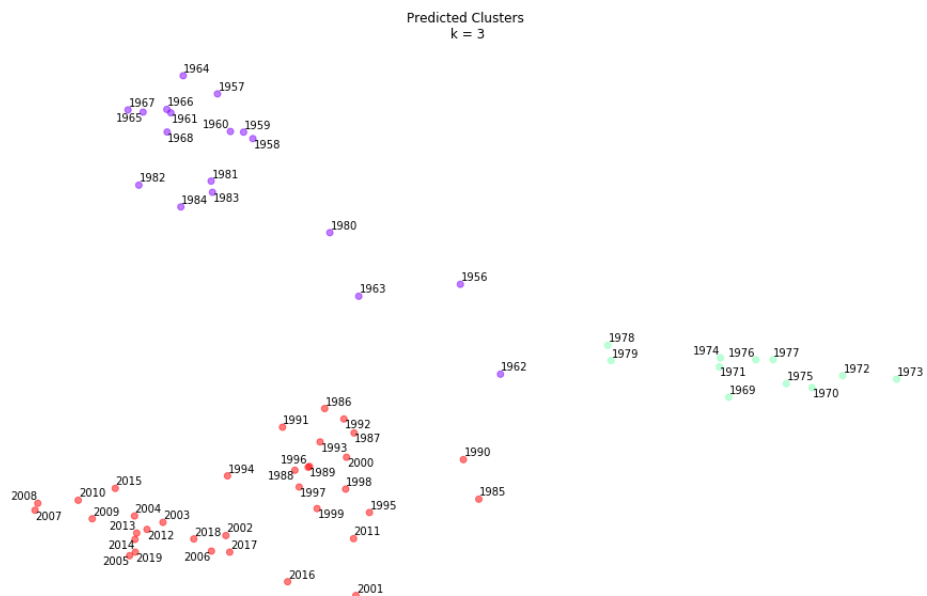


Figure 12: Clustering classification projected in the PCA space of the two principal components, labeled by year

In this kind of analysis the clusters should reveal high-level patterns in the relative frequency of words across speeches. The strongest trend we observe in these results is that cluster one (in purple) contains every democratically elected government until 1984 in our period of analysis, except for the government of president Lindley, which was atypical because he came to power through a military coup and stayed in office only for a year. Then, the second cluster (in cyan) contains every government that came to power through a military coup except for Lindley's administration, but we also observe that in our PCA plot Lindley's speech comes very close to the speeches of this cluster. Finally, the third cluster (in red) encloses every administration elected from 1985 onward.

This comes as a surprising result if we consider that the two administrations of cluster 2 (Morales Bermudez and Velasco), despite of being two military governments who took power by force, were very different left-wing (Velasco) and right-wing (Morales Bermudez) dictatorships. In fact, both administrations are usually characterized as antagonistic of each other: while Velasco explicitly tried to establish a "revolution" to take the role of the Peruvian economic elites in the country's development, Morales Bermudez's coup ousted Velasco from power and tried to modify and mitigate the reforms carried out by the "revolution" (Jaquette and Lowenthal 1986). Interestingly, this result also brings into light another important fact of the Peruvian military regimes of the decade of 1970: in a rhetoric level, Morales Bermudez's administration claimed his commitment with the on-going so-called revolution and the continuity of the policies started by Velasco (Jaquette and Lowenthal 1986). Our results, indeed, are very much in line with this.

Another interesting pattern that we observe in this result is the subtle differences in the clustering results and the PCA decompositions of the speeches belonging to the same president but to different administrations. This is the case of Belaunde (first term: 1963-1968, second term: 1980-1985), Garcia (first term: 1985-1990, second term: 2006-2011), and Fujimori (first term: 1990-1995, second term: 1995-2000), where it should be noted that Fujimori was the only president whose two administrations took place consecutively. In every one of these cases, the speeches of the first and second terms of each president fall within the same cluster. A closer look, though, reveals that the speeches from the first and second term of Fujimori have closer PCA projections than the speeches of the first and second terms of Belaunde, and that these have a closer PCA projection than the first and

second term speeches of Garcia. The first finding is explained by considering that Fujimori's two consecutive terms were very much a continuity of each other, where he first broke the constitutional order by closing the Congress and then legitimated this decision with the relative economic and social success of his entire administration (Sánchez 2002a), while Belaunde saw his second term (1980-1985) as the chance to continue with the policies of his first term (1963-1968), which ended abruptly when Velasco deposed him from office (Sánchez 2002b). The finding about Garcia's two terms deserves a deeper exploration of reasons. The speeches from his two terms, in fact, seem to show very distinct PCA decompositions between them: the speeches from Garcia's first term are close to Fujimori's and the speeches from the second term are close to Toledo (2001-2006) and Humala (2011-2016). This goes in line with a well-acknowledged fact by political scientists studying Peru's political landscape: Garcia's election platform and political discourse during his two presidential terms were remarkably different (Duárez Mendoza 2018). In the first term, he was elected on a platform of vindicating the long-deprived rights of the impoverished masses of the urban slums and the rural countryside (Cotler 1985), while in his second term his political discourse changed radically towards the hegemony of neoliberal economic policies (Duárez Mendoza 2018).

4.4 Word embeddings

We used word embeddings to explore the semantic spaces projected by each speech and the social and political insights derived from this result. Specifically, we projected the semantic space of each speech using document embeddings and focused in which speeches show a more unique meaning, providing hypotheses for the possible reasons why some speeches show innovative or similar contents.

We started this process by using the `Doc2Vec` function of the `gensim` package to train a model containing the semantic space of each speech projected in one hundred "meanings". To get similitude metrics for the speeches, we calculated the cosine difference of the projection of every pair of speeches using the `cosine_similarity` function of the `sklearn` package. This result is shown in Figure 13.

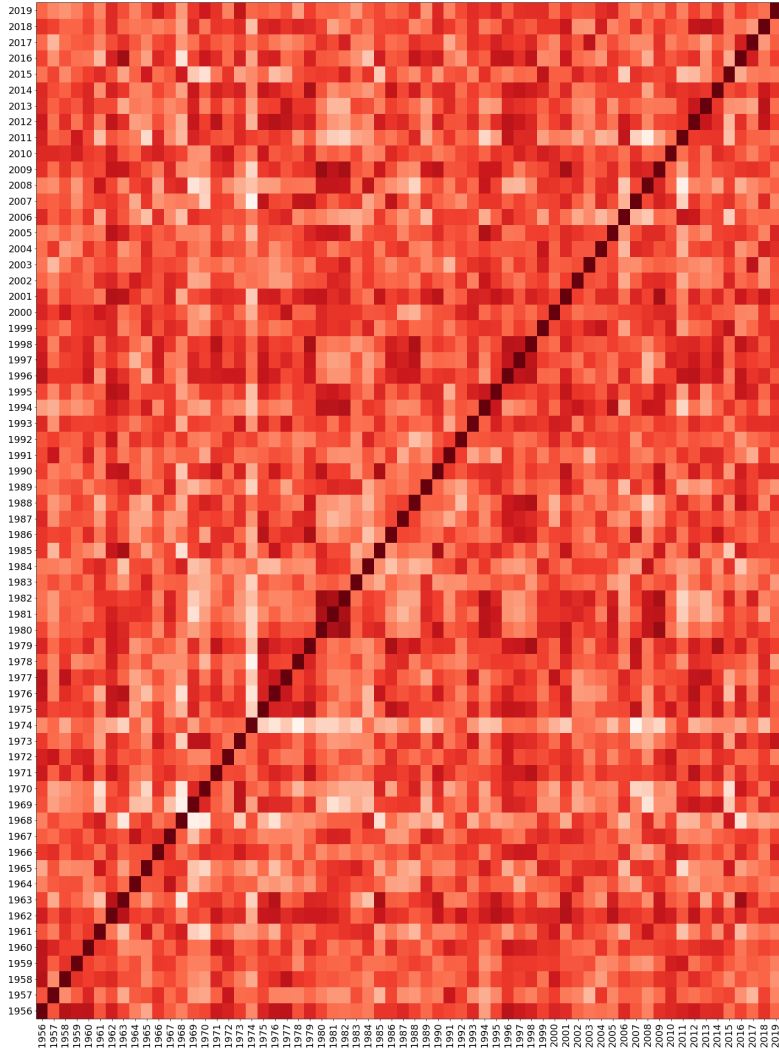


Figure 13: Pairwise cosine distance of the Doc2Vec semantic spaces of each speech. A darker color signals higher similarity between two speeches.

The matrix shown in Figure 13 reveals some interesting patterns in the distance between the semantic space of speeches. First, one salient feature are the uniqueness of some speeches, displayed by a large number of light squares in a single row or column. This is specially notable for the 1974 speech, that looks clearly distant to several speeches after it. The same happens to a lesser degree with the speeches of years 1969, 2008 and 2011.

Another important characteristic are the small clusters of similitude in speeches from

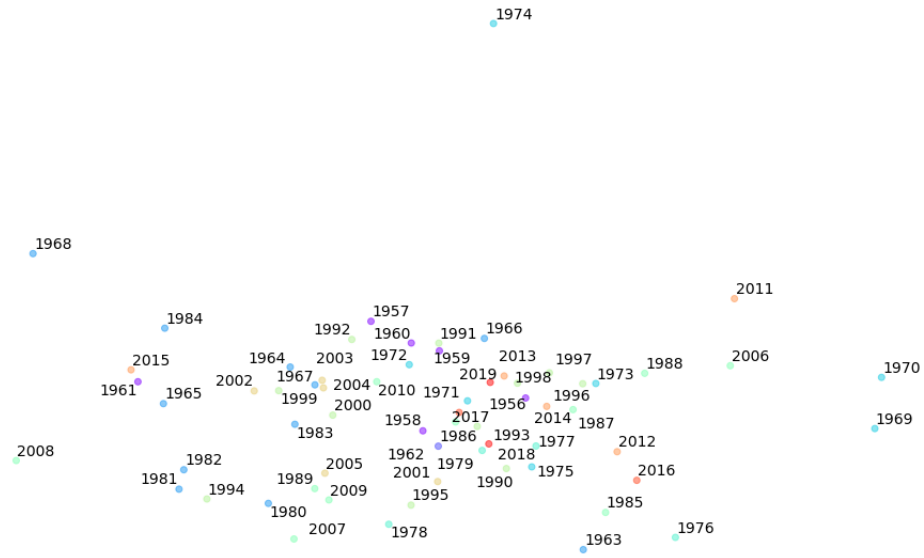


Figure 15: Doc2Vec semantic projection of the speeches projected in their two principal components, labeled by year. The same dot color indicates the same president.

The results of Figures 14 and 15 bring more clarity about the speeches with a unique content. Similar to what Figure 13 displays, the speech of 1974 stands out of the crowd, along with (to a lesser extent) the speeches of 1968, 1969, 1970, 2008 and 2011. Remarkably, half of these speeches correspond to president Velasco's administration (1969, 1970 and 1974); a result that, once again, captures the uniqueness of this regime in Peru. 1969 and 1970 were the first two years of Velasco's administration and also the first two years after Peru's land reform (*reforma agraria*, in Spanish), a major political milestone that transformed class relations and ushered more equitable land distributions in the country (Eckstein 2015). Looking at the actual content of 1974's speech, we see that its prominence might correspond to the fact that it included the government plan of the *Gobierno Revolucionario de las Fuerzas Armadas* (Revolutionary Government of the Armed Forces, the title the regime gave to herself); and Velasco actually did read the whole plan in front of the Congress when he delivered this speech.

The salience of the speeches of 2008 and 2011 is another interesting result. An explanation for the 2008 speech's singularity arises after looking at its content and the global land-

scape when it was delivered. It corresponds to the third year of president Garcia's second administration (2006-2011), and took place amidst the beginning of the Great Recession. The content reveals, since its initial pages, that it focused on the challenging year Garcia's government had ahead instead of emphasizing the achievements and future projects in the optimistic tone the presidential speeches usually use. The speech of 2011, on the other hand, belongs to president Humala's administration (2011-2016) inaugural year. When delivered, Humala established himself apart of the economic policies of his predecessors and centered his language in the social programs and re-distributive policies his administration planned to implement.

5 Conclusions

All our findings are in line with the theories of political scientists about Peru. Our results especially highlight the use of the presidential speeches as a way of presenting the administration's annual results (Sánchez 2002c); and the uniqueness of Velasco's regime, "The Peruvian Experiment", among every other presidential term (McClintock and Lowenthal 2015).

These results set a promising path to political scientists studying Peru to systematize and provide a more solid ground to their theories. Much of the field's work about the country is based in case studies and comparative politics, and there is tremendous ground to expand these analyses using computational techniques. The authors humbly hope this work can shine a light into the exploitation of computational content analysis tools that could complement the current methods applied by political scientists studying this country; and that this effort could lead to a more profound understanding of the uniquely complex and fascinating world of politics in Peru.

References

- Bielschowsky, Ricardo. 1985. "Evolucion de las ideas de la CEPAL (Spanish) [Evolution of the ideas of CEPAL]." *Revista de la CEPAL* (Santiago de Chile) Nro. Extraordinario.
- Contreras, Carlos, and Marcos Cueto. 2013. *Historia del Perú Contemporáneo*. Lima: Instituto de Estudios Peruanos.
- Cotler, Julio. 1985. "¿Nuevos Horizontes para el Perú?. (Spanish) [New Horizons for Peru?]." *Revista Nueva Sociedad* (La Rioja) 80.
- Duárez Mendoza, Jorge Luis. 2018. "Ser Aprista en Tiempos Neoliberales. Un análisis discursivo del segundo gobierno de Alan García Pérez (2006-2011). (Spanish) [To Be Aprista in Neoliberal Times. A Discourse Analysis of Alan García Pérez's Second Term (2006-2011)]." *Revista Temas y Debates* (Rosario) 36.
- Eckstein, Susan. 2015. "Revolution and Redistribution in Latin America." Chap. 11 in *The Peruvian Experiment Reconsidered*. Edited by Cynthia McClintock and Abraham F. Lowenthal, 369–408. Princeton: Princeton University Press.
- Jaquette, Jane S., and Abraham F. Lowenthal. 1986. *El experimento peruano en retrospectiva*. (Spanish) [*The Peruvian Experiment in Restrospective*]. Documento de Trabajo 19. Instituto de Estudios Peruanos.
- McClintock, Cynthia, and Abraham F. Lowenthal. 2015. *The Peruvian Experiment Reconsidered*. Princeton: Princeton University Press.
- Sánchez, Juan Martín. 2002a. "De la fortuna democrática a la fama autoritaria: 1990-1999. (Spanish) [From democratic fortune to authoritarian fame: 1990-1999]." Chap. 8 in *Perú 28 de Julio: discurso y acción política el día de Fiestas Patrias, 1969-1999. (Spanish) [Peru July 28: speech and political action on the National Day, 1969-1999]*. 201–226. Sevilla: Instituto de Investigaciones Dr. José María Luis Mora / Consejo Superior de Investigaciones Científicas.

- Sánchez, Juan Martín. 2002b. "El cruel Perú prometido: 1983. (Spanish) [The cruel promised Peru:1983]." Chap. 6 in *Perú 28 de Julio: discurso y acción política el día de Fiestas Patrias, 1969-1999. (Spanish) [Peru July 28: speech and political action on the National Day, 1969-1999]*. 155–176. Sevilla: Instituto de Investigaciones Dr. José María Luis Mora / Consejo Superior de Investigaciones Científicas.
- . 2002c. *Perú 28 de Julio: discurso y acción política el día de Fiestas Patrias, 1969-1999. (Spanish) [Peru July 28: speech and political action on the National Day, 1969-1999]*. Sevilla: Instituto de Investigaciones Dr. José María Luis Mora / Consejo Superior de Investigaciones Científicas.

6 Appendix: Brief political history of Peru

We want to motivate the reader and provide them with a succinct context about the political history this paper covers. In this section we detail the political history of Peru from 1956 to the year 2000. Our main motivation is that for the first time during the elections of 1955, the right to vote was universalized for adult women. This section is based on Contreras and Cueto 2013. From 1956 to 1961, president Prado, who was a representative of the oligarchic order that ruled Peru, instituted a liberal model where he reduced the presence of state in economy. There were profound social changes during this time, like the great migration from the country to the urban sites, that weren't reflected in thoughtful political changes with this president.

The "moderate reformism" followed its course with Fernando Belaunde's government, from 1963 to 1968, with a highly educated leader who wanted to unify the country, mainly, the Peruvian amazon. He wanted to reform the property rights of land, but the oligarchies and latifundistas wouldn't let him. Moreover, a scandal of great magnitudes that happened with regards to an agreement between the International Petroleum Company (IPC) and the Peruvian government, that disfavored the country, the constant backlash of the congress, and the economic crisis of 1967, obliged him to resign.

After this, the military coup of 1968, led by Juan Velasco, changed the structure of the political and economic order in Peru. The main development framework of this era was the industrialization by the substitution of imports (ISI), so many of the productive indus-

tries were expropriated and controlled by the State, the word Peru was on fashion for the mining, fishing, maritime transportation, telecommunications , electricity and air flights (PetroPeru, PescaPeru, EntelPeru, ElectroPeru, for example). Most importantly, the social property, where all workers shared ownership of the firm, was instaurated. One of the most significant changes was the change in property of the productive land: 9.1 million hectares out of the 27 million that exist in Peru, were expropriated and given to the workers, in an attempt to bring equality to land ownership, and to give the land to who works it. Morales Bermudez represented a second phase of the military government who disagreed with many of the radical measures that were took by Velasco. The economic crises that were in part caused by the harsh expropriations from Velasco's time, didn't give much space to undertake government actions.

Finally, 1980 represents a return to democracy with Fernando Belaunde, the same man who was forced to resign before the military coup. His government was oriented towards the reduction of the State that was imposed by the coup, the austerity in fiscal expenditure, and the reduction of the State in the economy. The main difficulty he had to face was the El Nino phenomenon, which left Peru with great economic losses.

1985 was the inaugural year for President Alan Garcia. His administration was characterized by a heavy control of the Economy by the State: price controls, selective devaluation of the foreign currency and freezing the exchange rates. He also stopped paying the foreign debt. One of his biggest mistakes was the attempt to nationalize the banks and the creation of a parallel dollar, which contributed to an accumulated hiperinflation of 2 million %.

Alberto Fujimori's democratically elected dictatorship lasted 10 years. He undermined many democratic institutions that took long to build, corrupting them in its core. Also, he impoverished citizen participation in politics by weakening syndicates. Nowadays, he is in prison because of the crimes he committed while being president.

Alejandro Toledo's administration was a new comeback to democracy, and a bet towards decentralization. His administration established the institutional framework for regional governments. Before that, it would have seemed that Peru was Lima. Also, the healthy economic environment, propitiated by high prices of commodities, favored Peru in having a bigger economic power. Similar paths were followed by the second adminis-

tration of Alan Garcia.

The decade of the 2010's is still a matter of discussion among historians and political scientists, and will be broaden in future iterations of this work.