

# Can a troubled political era be detected by machine learning methods? An application on the End of Year speeches of the Italian Presidents

Franco M. T. Gatti<sup>1</sup>, Arjuna Tuzzi<sup>2</sup>

<sup>1</sup>Università degli Studi di Padova – franco.gatti.1@phd.unipd.it

<sup>2</sup>Università degli Studi di Padova – arjuna.tuzzi@unipd.it

## Abstract

President Oscar Luigi Scalfaro took office in 1992, when the Italian political and judicial climate was highly turbulent. As his speeches may be considered a good source of information for grasping the transition between the so called "First" and the "Second" Republic, this study aims at analysing the 71 End of Year speeches pronounced by the 11 Presidents of the Italian Republic in order to check whether we are able to detect a clear-cut fracture within the corpus by means of supervised machine learning methods applied to texts. For this purpose, our corpus has been arranged in three different configurations to assess the performance of different classification tasks within the training set (cross-validation procedure) and to observe in which one of the three arrangements the algorithms perform better. For the final attribution task we chose the configuration that showed the highest accuracy and we split the set of speeches in a training set and a test set to observe how the algorithms assign the text chunks of the test set to the two classes (First or Second Republic). A comparison between the performance of the two different algorithms used in this work, i.e. Support Vector Machine (SVM) and Random Forest (RF) has been discussed throughout the whole study. With reference to previous studies, Scalfaro's speeches confirmed their role as transition messages and the maximum precision was reached by both algorithms when they were discarded by the two classes.

**Keywords:** machine learning, Support Vector Machine, Random Forest, presidential speeches, End of Year speeches

## Riassunto

Il Presidente Oscar Luigi Scalfaro ha iniziato il suo mandato nel 1992, quando il clima politico e giudiziario italiano era fortemente turbolento. Dal momento che i suoi discorsi possono essere considerati una buona fonte informativa per cogliere la transizione tra le cosiddette "Prima" e "Seconda" Repubblica, questo lavoro mira a studiare i 71 messaggi di fine anno pronunciati dagli 11 Presidenti della Repubblica Italiana per stabilire se è possibile rilevare una frattura netta all'interno del corpus, con l'ausilio di metodi di *supervised machine learning* applicati ai testi. Per raggiungere questo obiettivo, il corpus è stato preparato in tre configurazioni diverse al fine di saggiare la performance di diversi processi di classificazione all'interno del training set (procedura di *cross-validation*) e di osservare in quale delle tre configurazioni gli algoritmi raggiungono la migliore performance. Per l'ultima procedura di attribuzione abbiamo scelto la configurazione che ha mostrato la precisione più alta abbiamo suddiviso l'insieme dei discorsi in un training set e un test set, al fine di osservare come i frammenti del test set vengono assegnati dagli algoritmi alle due classi (Prima o Seconda Repubblica). Un confronto tra le performance dei due diversi algoritmi utilizzati in questo lavoro, *Support Vector Machine* (SVM) e *Random Forest* (RF), è stato discusso nel corso della ricerca. I discorsi di Scalfaro hanno confermato il loro ruolo di messaggi di transizione già evidenziato in lavori precedenti e la massima precisione è stata raggiunta da entrambi gli algoritmi quando sono stati esclusi dalle due classi.

**Parole chiave:** machine learning, Support Vector Machine, Random Forest, discorso presidenziale, discorsi di fine anno

## 1. Introduction

In 1992, a few months before Oscar Luigi Scalfaro took office as the 9th President of the Italian Republic, the climate within Italian society grew troubled. Some judiciary investigations uncovered a thick network of bribe affairs between almost all the main political parties and many influential representatives of the business and financial world. Furthermore, the assassination of magistrate Giovanni Falcone - who was conducting some investigations on the links between Cosa Nostra (a well-known mafia-type Italian criminal organization) and some political leaders of the country - took place on May 23, 1992, precisely on the same days when the sessions for the election of the President of the Republic were taking place.

This tense period was exploited by the new President Scalfaro as a chance to strengthen one of the key functions of the Head of State office: a role of *super partes* guarantor. In order to keep the balance in a troubled country he took the opportunity to deliver speeches that were outdistanced from those of his predecessors and that appear quite different from those of his successors (Cortelazzo 2007); even in some of those institutional ceremonies well known for characterizing the role of the President with a certain rituality. This is the case of the End of Year speeches that represent a traditional and relevant social event in the year lifetime of many Countries when the President (or the King or Queen of a Country) gives a wishes speech to the whole citizens (Cortelazzo and Tuzzi, 2007).

The corpus collected and arranged for this research might be a good testing field in order to check whether the fracture opened in 1992 within the Italian society can be detected also in the President's End of Year speeches. And more specifically whether it is possible to do it by means of machine learning methods aimed at text classification. All the analyses included in this work have been processed within the version 3.6.1 of R (R Core Team, 2019) environment for statistical computing.

## 2. Corpus

The corpus of the End of Year addresses of the Presidents of the Italian Republic (Table 1) is 120,531 word tokens in size and includes all the 71 speeches delivered by all the 11 Presidents: Luigi Einaudi (1948-1955), Giovanni Gronchi (1955-1962), Antonio Segni (1962-1964), Giuseppe Saragat (1964-1971), Giovanni Leone (1971-1978), Sandro Pertini (1978-1985), Francesco Cossiga (1985-1992), Oscar Luigi Scalfaro (1992-1999), Carlo Azeglio Ciampi (1999-2006), Giorgio Napolitano (2006-2015), and Sergio Mattarella (2015-current). We have not included Enrico De Nicola (Provisional Head of State and first President of the Italian Republic in the time span 1946-48) because his successor, Luigi Einaudi, inaugurated the tradition of the Italian End of Year speech in Italy and his first discourse took place during the second year of his office in 1949.

The Italian Constitution envisages a presidential office of 7 years and our corpus includes seven speeches for most Presidents. Luigi Einaudi (6 speeches), Antonio Segni (2 speeches, since he resigned from his position after two years), Giorgio Napolitano (9 speeches, since in 2013 he obtained a second office and, then, resigned after two years), and Sergio Mattarella (5 speeches, as he is currently in office) represent exceptions.

Table 1. Basic measures of the End of Year speeches of the Italian Presidents

	Office	No. disc.	size in word tokens			
			total	min	max	mean
Luigi Einaudi	1948-1955	6	1,203	150	260	201
Giovanni Gronchi	1955-1962	7	5,829	388	1,252	833
Antonio Segni	1962-1964	2	1,795	738	1,057	898
Giuseppe Saragat	1964-1971	7	8,476	465	1,929	1,211
Giovanni Leone	1971-1978	7	7,379	262	1,604	1,054
Sandro Pertini	1978-1985	7	15,592	1,340	3,746	2,227
Francesco Cossiga	1985-1992	7	13,890	418	3,345	1,984
Oscar Luigi Scalfaro	1992-1999	7	24,675	2,085	5,012	3,525
Carlo Azeglio Ciampi	1999-2006	7	12,597	1,193	2,129	1,800
Giorgio Napolitano	2006-2015	9	20,389	1,717	2,643	2,265
Sergio Mattarella	2015-current	5	8,706	1,102	2,127	1,741
Corpus		71	120,531	150	5,012	1,698

Figure 1 highlights the changing length in word tokens of speeches over time. As already illustrated in previous studies (cfr. Cortelazzo and Tuzzi 2007) the size of speeches is highly mutable: at the very beginning short speeches were delivered at the radio (in 1959 the fifth address by Gronchi was broadcast on television), some Presidents decided to be very concise in the first addresses (e.g. Gronchi, Saragat, Leone), in 1991 Cossiga chose to deliver a very short last message that was only an anticipation of his resignation from office a few months later (and 1992 emerges again as a troubled time in the history of democratic institutions in Italy); on the contrary, Scalfaro's discourses are wide and eloquent.

Previous studies have also already explained that the text genre is only apparently flat and homogeneous. The individual traits of the President are more relevant than other variables, for example, more relevant than the temporal proximity among Presidents, as well as ideological proximity, generation proximity, etc. (Pauli and Tuzzi 2009, Bernardi and Tuzzi 2007, Cortelazzo and Tuzzi 2016). An analysis of chronological data highlighted that the time effect is not as important as Presidents' individual choices and offered a portrait of some Presidents as representatives that were able to affirm their personality beyond contingent issues (Trevisani and Tuzzi 2012).

As shown in the correspondence analysis in Figure 2, two Presidents clearly emerges from the others: Sandro Pertini and Oscar Luigi Scalfaro. Pertini can be considered unique as he often delivered speeches without following a written canvas word by word. Scalfaro's differentiation shows that he can't be easily assimilated to any other presidential style. The current President, Sergio Mattarella, is clearly a mix of the characteristics of his two predecessors: Giorgio Napolitano and Carlo Azeglio Ciampi (Figure 3).

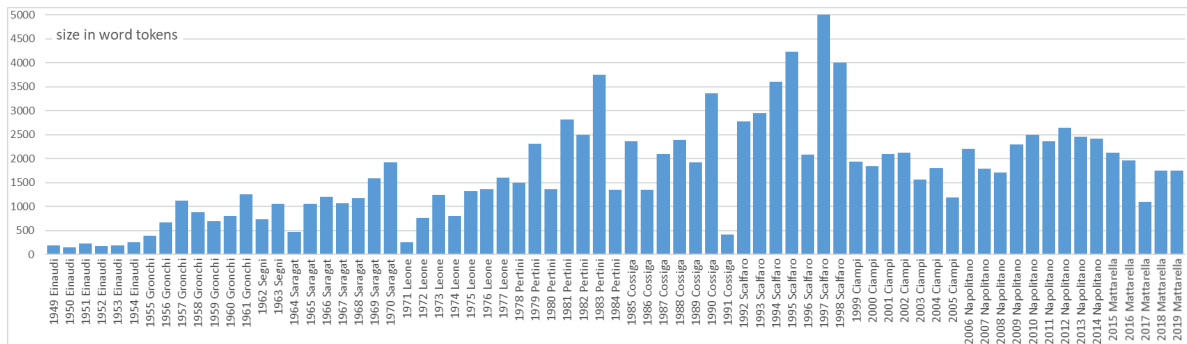


Fig.1. Size of the End of Year speeches in word tokens.

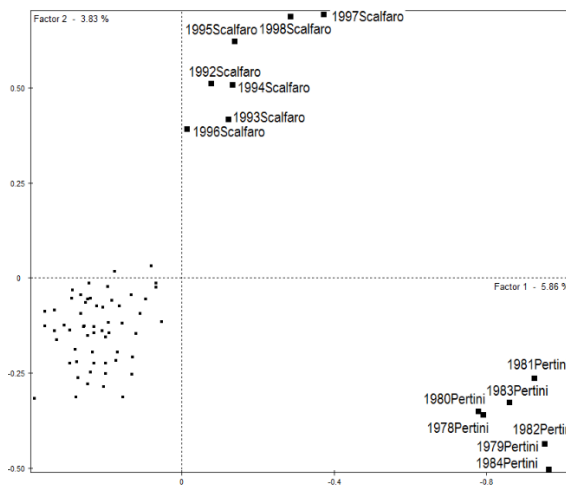


Fig.2 First factorial plan of Correspondence Analysis - word types with a frequency  $\geq 5$ , projection of speeches

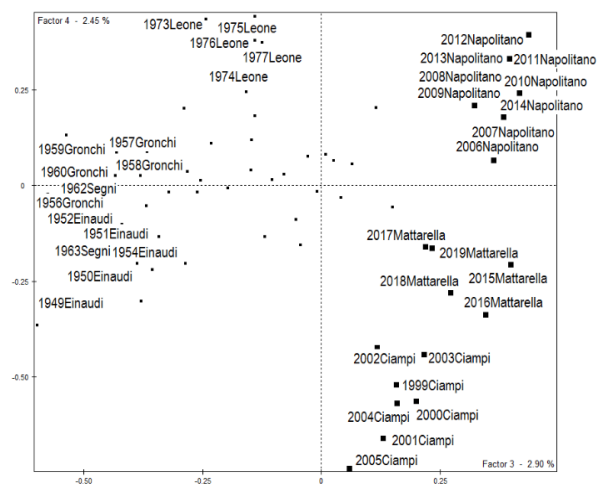


Fig.3 Second factorial plan of Correspondence Analysis - word types with a frequency  $\geq 5$ , projection of speeches

### 3. Methods

By means of two supervised machine learning methods, we aim to classify the End of Year messages of the Italian Presidents and observe whether the so called “transition president” Oscar Luigi Scalfaro fits better in the First or in the Second Republic. In order to investigate our research aim, we arranged three different configurations of the corpus (Figure 4):

- 1- the first one is a bipartition between Presidents of the First Republic and Presidents of the Second one, with Oscar Luigi Scalfaro included in the second half;
- 2- the second one is a bipartition between Presidents of the First Republic and Presidents of the Second one, with Oscar Luigi Scalfaro included in the first half;
- 3- the third is a bipartition between Presidents of the First Republic and Presidents of the Second one, without Oscar Luigi Scalfaro.

	1949 - 1991	1992 - 1998	1999 - 2019
First Arrangement	First Republic	Second Republic	
Second Arrangement	First Republic		Second Republic
Third Arrangement	First Republic		Second Republic

Figure 4. The three different arrangements of the corpus (71 speeches)

For each arrangement, we ran an analysis through a model named *Author's Multilevel N-Gram Profiles* (Mikros and Perifanos, 2013; Cortelazzo, Mikros and Tuzzi; 2018), culminating in the application of two machine learning algorithms: *Support Vector Machine*, and *Random Forest*.

Before running AMNP procedure, we decided to exclude all Luigi Einaudi's speeches. The reason is twofold: we had to take into account an historical viewpoint as Einaudi inaugurated the tradition of the End of Year speech when messages were broadcasted at the radio. Then, when television started being widespread among Italian population, Presidents were able to extend and deepen their speeches (Zotti Minici, 2007). This condition is connected to our second reason, a methodological one: Einaudi's speeches are very short (most of them are shorter than 200 words) and we decided to process chunks of 200 words in length (Figure 5).

	1949 - 1954	1955 - 1991	1992 - 1998	1999 - 2019
First Arrangement	Einaudi	First Republic	Second Republic	
Second Arrangement	Einaudi	First Republic		Second Republic
Third Arrangement	Einaudi	First Republic		Second Republic

Figure 5. The three different arrangements of the corpus without Einaudi (65 speeches)

### 3.1. Author's Multilevel N-Gram Profiles model

AMNP can be seen as a language independent model (Mikros and Perifanos 2011; 2013) useful to prepare a text corpus for upcoming analyses with classification algorithms. Aside from loading, parsing and then splitting the corpus in a specific number of text chunks chosen by the researchers, it allows to set many parameters in order to be able to reconstruct a corpus readable for the selected algorithms.

In this work, we decided to split the text corpus into chunks of 200 words in length and to analyse their most frequent linguistic features at three different levels of the language, identified by bigrams, trigrams, words and word bigrams; with word bigrams we refer to couples of contiguous words (e.g. *Repubblica Italiana* [Italian Republic], *per ogni* [for each], etc.). Since many speeches included in the corpus are short in length, we had to choose a size for the chunks that wasn't as large to force us to exclude many speeches from the analysis. Furthermore, previous works that used AMNP script set a higher number of features, so we decided as well to not cut the size of the chunks more than necessary (e.g. using chunks of 100 tokens). In trying to find a compromise between these two needs, we saw as reasonable chunks of 200 words in length, which have already been used in Cortelazzo, Mikros and Tuzzi (2018). During the training phase, each text chunk has been assigned to two classes: *Pre* and *Post*. The border between the two classes represents the fracture we aim to investigate (First vs. Second Republic).

After loading and parsing the text corpus, we counted the stylometric features and split the files into 200 words text chunks. The total amount of text chunks resulting in this process is 562 (for the analyses that include Scalfaro's speeches) and 442 (for the ones without them). Then, when

computed a list of 200 most frequent features for each bigrams, trigrams, words and word bigrams, we produced a table of relative frequencies for each type of these feature. Starting from these 4 tables, we interlocked them in one single matrix featuring the 200 most frequent bigrams, the 200 most frequent trigrams, the 200 most frequent words and the 200 most frequent word bigrams.

### **3.2. Support Vector Machines and Random Forest**

The first algorithm we decided to apply is *Support Vector Machine* (SVM). In our case, it uses a polynomial kernel that makes 27 combinations of three parameters (*scale*, *degree*, *c*) with 3 possible values each, in order to find the hyperplane that marks the largest difference between the two classes. Once the execution of the first algorithm was complete, we applied a 5-fold cross-validation procedure in order to project the results of the classification performed within the corpus.

The second algorithm we decided to apply is *Random Forest* (RF). Unlike SVM, *Random Forest* works with decision trees instead of hyperplanes- The output is similar to the first one in terms of data readability and we submitted its results to a 5-fold cross-validation procedure as well.

After repeating the procedure above for three times (one for each corpus partition, cfr. Figure 4), we selected the third configuration since it obtained the highest accuracy. Moreover, since in this configuration the 7 speeches delivered by Scalfaro are not part of the training set, we had the opportunity to split the corpus into a *training set* and a *test set* in order to predict the classification Scalfaro's speeches.

## **4. Results**

As previously explained, the machine learning analysis in this work is divided in two phases: the output of the first one is a couple of cross-validation matrices (one for SVM and one for RF), for each configuration of the corpus (Table 2: first arrangement, Table 3: second arrangement, Table 4: third arrangement, cfr. Figure 4). In each task we took into account the corpus split into text chunks of 200 words in length and the 200 features with the highest relative frequencies for bigrams, trigrams, words and word bigrams. This means that the total number of features considered in each task amounted to 800 total features (200 most frequent bigrams, 200 most frequent trigrams, 200 most frequent words and 200 most frequent word bigrams).

Table 2. First arrangement of the corpus (1992 fracture): 65 speeches, 562 text chunks.

<b>Accuracy: 0.904</b> <b>SVM</b>			<b>Accuracy: 0.812</b> <b>RF</b>		
Reference			Reference		
Prediction	<b>Pre</b>	<b>Post</b>	Prediction	<b>Pre</b>	<b>Post</b>
<b>Pre</b>	43.6	5.4	<b>Pre</b>	31.8	2.0
<b>Post</b>	5.4	58.0	<b>Post</b>	17.2	61.4
<b>Precision (Pre / Post): 0.90 / 0.92</b> <b>Recall (Pre / Post): 0.90 / 0.92</b> <b>F1 Score (Pre / Post): 0.90 / 0.92</b>			<b>Precision (Pre / Post): 0.94 / 0.78</b> <b>Recall (Pre / Post): 0.65 / 0.97</b> <b>F1 Score (Pre / Post): 0.77 / 0.87</b>		

Table 3. Second arrangement of the corpus (1999 fracture): 65 speeches, 562 text chunks.

<b>Accuracy: 0.924</b> <b>SVM</b>			<b>Accuracy: 0.813</b> <b>RF</b>		
Reference			Reference		
Prediction	<b>Pre</b>	<b>Post</b>	Prediction	<b>Pre</b>	<b>Post</b>
<b>Pre</b>	68.8	4.2	<b>Pre</b>	68.2	16.2
<b>Post</b>	4.2	35.2	<b>Post</b>	4.8	23.2
<b>Precision (Pre / Post): 0.94 / 0.90</b> <b>Recall (Pre / Post): 0.94 / 0.90</b> <b>F1 Score (Pre / Post): 0.94 / 0.90</b>			<b>Precision (Pre / Post): 0.81 / 0.85</b> <b>Recall (Pre / Post): 0.94 / 0.59</b> <b>F1 Score (Pre / Post): 0.87 / 0.70</b>		

Table 4. Third arrangement of the corpus (without Scalfaro): 58 speeches, 442 text chunks.

<b>Accuracy: 0.925</b> <b>SVM</b>			<b>Accuracy: 0.864</b> <b>RF</b>		
Reference			Reference		
Prediction	<b>Pre</b>	<b>Post</b>	Prediction	<b>Pre</b>	<b>Post</b>
<b>Pre</b>	45.0	2.6	<b>Pre</b>	46.0	9.0
<b>Post</b>	4.0	36.8	<b>Post</b>	3.0	30.4
<b>Precision (Pre / Post): 0.96 / 0.90</b> <b>Recall (Pre / Post): 0.92 / 0.95</b> <b>F1 Score (Pre / Post): 0.94 / 0.92</b>			<b>Precision (Pre / Post): 0.84 / 0.91</b> <b>Recall (Pre / Post): 0.94 / 0.77</b> <b>F1 Score (Pre / Post): 0.88 / 0.83</b>		

As the three couples of matrices show clearly, the precision in the classification tasks is pretty high for all the arrangements of the corpus, with SVM performing better than RF in each and every run. However, when shifting the fracture from 1992 to 1999 (i.e. considering Oscar Luigi Scalfaro a better fit in the First Republic or in the First one), the precision grows from 0.904 to 0.924 with Support Vector Machine and from 0.812 to 0.813 with Random Forest (cfr. Table 2 and Table 3).

A further significant step forward in performance is then registered when we totally exclude President Scalfaro's speeches from the corpus, as the accuracy grows from 0.924 to 0.925 with SVM, and from 0.813 up to 0.864 with RF. In other words, with Random Forest applied to the third arrangement of the corpus, the number of correctly classified chunks of Scalfaro's successors rises from 23 to 30 (cfr. Table 3 and Table 4).

While accuracy is the overall percentage of chunks correctly classified, it is not the only measure available to track algorithms performance. In both SVM and RF we calculated other three commonly used validation measures: precision, recall, and F1 score, each one focusing on different values of the cross-validation matrices. Since accuracy is an overall percentage, its value is the same whether calculated on the first or on the second diagonal of the table, while precision, recall and F1 score have to be calculated on each class.

As for the precision, it is the value related to the percentage of chunks predicted to be in a specific class, that actually belong to that class. So, precision is calculated dividing the number of chunks belonging to a specific class, for the number of chunks that the prediction assigned to that class. On the other side, what recall aims to find is something different from precision, and it refers to the percentage of chunks belonging to a specific class, that have been correctly classified, so it is calculated dividing the number of correctly classified chunks in a single class, for the total number of chunks belonging to that class. Finally, since there is usually a trade-off between precision and recall, F1 score aims at calculating the relative performance between precision and recall:  $(precision \times recall) / (precision + recall)$ . Whether to choose accuracy, precision, recall or F1 score as a value to define how good a classification algorithm performs, it depends on the objective of the research. In this work, we believe that a raw but effective overall percentage of correctly classified chunks (accuracy) may be a good value to take into account in order to evaluate performance. In fact, considering the values of precision, recall and F1 score in each one of the three arrangements of the corpus, and comparing and contrasting these values, we notice that they generally increase along with accuracy, with the only exception for the performances of RF in the first and second arrangement.

The next phase in the research is the mere application of the algorithms to the corpus divided in training and testing set, in order to assess the performance of the classification tasks (for both SVM and RF) on a group of speeches kept apart from the those included in the training process of the algorithms.

In our case, the training set is composed of speeches pronounced by the Presidents before and after Scalfaro (i.e. the third arrangement), and the test set is made of the 7 end of year speeches of Scalfaro himself. This operation is a further leap forward in the experiments, because the algorithms now work not with samples taken from the corpus itself, but aim to classify our "transition president" taking samples from his speeches that are outside the corpus. This means that the text chunks derived from Scalfaro's speeches did not contribute to training the algorithms that we then called to evaluate the best fit spot for his speeches (into the First or Second Republic), thus making their classification the exclusive result of the elaboration on features derived from the speeches of his successors and predecessors.



Table 5. Test set. Classification of the Scalfaro's speeches. 7 speeches, 120 text chunks.

	SVM	RF
<b>Pre</b>	76	105
<b>Post</b>	44	15

As shown above, both algorithms (and especially Random Forest) tend to indicate Scalfaro's speeches as a better fit among Presidents of the First Republic, but still keeping a cloud of uncertainty around his end of year messages. In particular, SVM puts 44 out of 120 chunks among the "post" period, defining him as a transition point among his predecessors and successors. (Table 5)

Furthermore, it is interesting to observe the distribution of Pre and Post chunks across the 7 years office of President Scalfaro (Figure 6). While at the beginning his speeches are more linked to the First Republic and around 1994, 1995 and 1996 we notice uncertainty in classifying them, at the end of his office President Scalfaro seemed to get back to his roots.

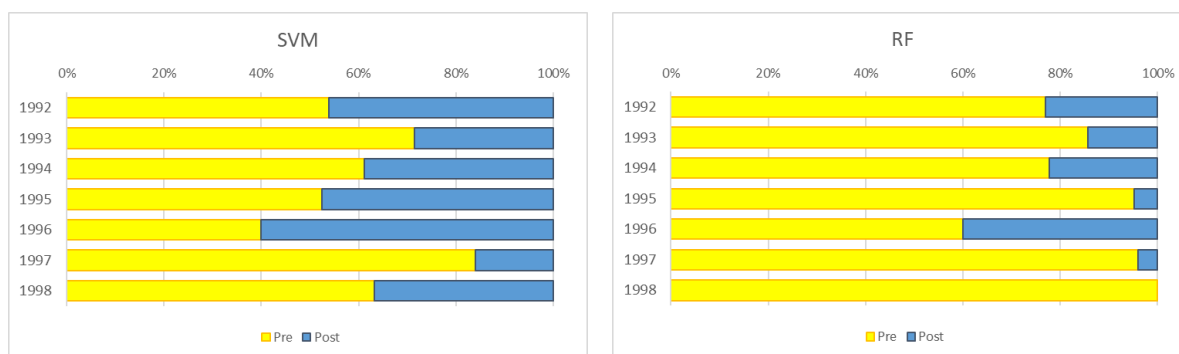


Figure 6. Results of the attribution task performed by SVM (left) and RF (right). No. of chunks of Scalfaro's speeches in percentage.

## 5. Discussion and conclusions

As President Oscar Luigi Scalfaro had a key role in leading the Country from First to Second Republic, it is reasonable to expect that this transitional characterization might be reflected in his speeches. After choosing the arrangement with the best accuracy in the classification tasks that we decided to apply, we used that specific configuration to test Scalfaro's speeches on a corpus trained exclusively with linguistic features taken from the speeches of his predecessors and his successors.

The results lead us to identify the linguistic features extracted from his texts as a mix of the traits of the two subcorpora. This may mean that he used to deliver unique speeches, that cannot be put just in the First and Second Republic, or rather confirm, once more, that he was certainly tied to the First Republic way to deliver speeches, but was also an innovator, as he was able to differentiate himself from that style.

However, even considering the transitional characterization of his End of Year messages, the best fit for his speeches seems located in the First Republic. In fact, as noticeable from increasing results of the classification performance when shifting the fracture from 1992 to 1999, and as confirmed by the classification of the test set composed of his 7 End of Year speeches, we were able to assign 76 chunks out of 120 with SVM, and 105 out of 120 with RF, to the First Republic.

Moreover, as shown in the distribution of the chunks across his 7 years Presidency, at the beginning and at the end of President Scalfaro's office, his speeches were mainly classified in the First Republic. The uncertainty in the classification takes place mostly in years 1994, 1995 and 1996, when the Country was rebuilding itself from the political and judicial earthquakes started in 1992. Those were also the years when Silvio Berlusconi stepped into the political panorama as the leader of the center-right coalition. Berlusconi's propaganda involved many communication techniques that were new to both his political opponents and the institutions (Bolasco et alii, 2009; Giuliano and Villani, 2015; Rodriguez, 2013), which had to confront with a candidate - and then, a Prime Minister - used to talk to Italian citizens from a horizontal perspective, opposed to the vertical way of communication typical of the First Republic (Antonelli, 2017). But the comeback to his roots that President Scalfaro makes it's not bizarre at all, if we think that he has been one of the main political figures throughout the whole First Republic, since when he was elected in the Assembly that wrote the Constitution of the Italian Republic between 1946 and 1948. So this is why we can say that, even if he led Italy in years when the Country had to cut from the past and to face new challenges and issues, President Scalfaro gave speeches that were more tied to the First Republic than to the Second one. Still, having a totally different and immediately recognisable style that opened the door to his successors.

## References

- Antonelli, G. (2017) *Volgare eloquenza. Come le parole hanno paralizzato la politica*. Bari-Roma, Laterza.
- Bernardi, L., Tuzzi, A. (2007). Parole lette con misura. In Cortelazzo, M. A. and Tuzzi, A. editors, *Messaggi dal Colle*. Marsilio. pp. 109-134.
- Bolasco, S., Giuliano, L., Galli de' Paratesi, N. (2009). *Parole in libertà. Un'analisi statistica e linguistica dei discorsi di Berlusconi*. Manifestolibri, Roma.
- Cortelazzo, M. A. (2007). Continuità e discontinuità degli stili oratori. In Cortelazzo, M. A. and Tuzzi, A. editors, *Messaggi dal Colle*. Marsilio. pp. 207-230.
- Cortelazzo, M. A. and Tuzzi A. (2007). *Messaggi dal Colle*. Marsilio, Venezia.
- Cortelazzo, M. A., Mikros, G. K., Tuzzi, A. (2018). Profiling Elena Ferrante: a Look Beyond Novels, in Iezzi, D. F., Celardo L., Misuraca M. (eds.), JADT 2018. Proceedings of the 14th International Conference on Statistical Analysis of Textual Data. Roma: UniversItalia, pp. 165-173.
- Giuliano L., Villani P. (2015). *Il linguaggio della leadership politica tra la Prima e la Seconda Repubblica, problemi di metodo e linee di ricerca*. Camera dei Deputati, Roma.
- Grimaldi, S. (2012). *I Presidenti nelle forme di governo*. Carocci, Roma.
- Labbé, D., Monière, D. (2008), *Les mots qui nous gouvernent. Le discours des premiers ministres Québécois: 1960-2005*, Monière-Wollank Éditeurs, Montréal.
- Marchand, P. (2007), *Le grand oral. Les discours de politique générale de la Ve République*, De Boeck Université, Bruxelles.

- Mikros, G. K., Perifanos, K. (2013). Authorship attribution in Greek tweets using multilevel author's n-gram profiles. In Hovy E., Markman V., Martell C. H. and Uthus D. (eds.), Papers from the 2013 AAAI Spring Symposium "Analyzing Microtext", 25-27 March 2013, Stanford, California. Palo Alto, California, AAAI Press, pp. 17-23.
- Mikros, G.K., Perifanos, K. (2015). Gender Identification in Modern Greek Tweets. in Tuzzi A., Benešová M. & Macutek J. (Eds.), *Recent Contributions to Quantitative Linguistics* (Vol. 70, pp. 75-88). Berlin: De Gruyter.
- Pauli, F., Tuzzi, A. (2009). The end of year addresses of the Presidents of the Italian Republic (1948-2006): discursal similarities and differences, *Glottometrics*, 18, pp. 40-51.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.
- Rodriguez, M. (2013) *Consenso. La comunicazione politica tra strumenti e significati*. Guerini e associati, Milano.
- Savoy, J. (2010). Lexical analysis of US political speeches. *Journal of Quantitative Linguistics*, 17(2), pp. 123-141.
- Savoy, J. (2017). Analysis of the Style and the Rhetoric of the American Presidents Over Two Centuries. *Glottometrics*, 38(1), pp. 55-76.
- Trevisani, M., Tuzzi, A. (2013), Shaping the history of words, in: Obradović I., Kelih E., Köhler R. (eds), *Methods and Applications of Quantitative Linguistics: Selected papers of the VIIIth International Conference on Quantitative Linguistics (QUALICO)*, Belgrade, Serbia, April 16-19, 2012, Akademska Misao, Beograd, pp. 84-95.
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer-Verlag, New York.
- Zotti Minici, A. (2007). Dove guardano i presidenti a fine anno? In Cortelazzo, M. A. and Tuzzi, A. editors, *Messaggi dal Colle*. Marsilio. pp. 47-54.