Emotion recognition in Italian political language to predict positionings and crises government

Il riconoscimento delle emozioni nel linguaggio politico italiano per predire posizionamenti e crisi di governo

Alessia Forciniti and Emma Zavarrone

Abstract The paper aims to analyze the political language adopted on Twitter by the main Italian parties' leaders during the first two waves of Covid-19 pandemic.
A two-step model based on sentiment emotion recognition (ER) and Correspondence analysis detected which emotions characterized the political language and which changes happened between the two waves. The results showed the use of a language with a strong emotional weight for some political actors as opposed to others who used a neutral register of political language in both waves. The comparison between two waves denoted a shift from anger to sadness and fear for Meloni and a moving away Salvini by predicting through ER the rift of the right-wing.

Abstract *L'articolo mira ad analizzare il linguaggio politico adottato su Twitter dai principali leader di partito italiani durante le prime due ondate della pandemia di Covid-19. Un modello di analisi a 2-step basato su sentiment emotion recognition (ER) e Analisi delle Corrispondenze, ha rilevato quali emozioni hanno caratterizzato il linguaggio politico e quali cambiamenti sono avvenuti tra le due ondate. I risultati hanno mostrato l'uso di un linguaggio con un forte peso emozionale per alcuni attori politici in opposizione ad altri che hanno utilizzato un registro del linguaggio politico neutro. Il confronto tra le due ondate ha denotato il passaggio dalla rabbia alla tristezza e paura per la Meloni e un distacco da Salvini preannunciando con l'ER la spaccatura del centro-destra.*

---

[1]   Alessia Forciniti, IULM University; alessia.forciniti@iulm.it

Emma Zavarrone, IULM University; emma.zavarrone@iulm.it

1 Introduction

The health emergency of Covid-19 represented a breaker point in terms of stability both in private life of individuals and in public domain, creating vulnerability in social, economic and political field. In this phase marked by a general perception of disorientation, the communication has assumed a decisive role in order to inform the citizens about the evolution of the pandemic and to explain the governments' containment measures aimed at addressing the behaviors of national communities. In Italy, communication has been broadcast both through traditional media and by means of social media, which have become the main channels for disseminating information and sharing opinions. In fact, during the lockdown period, in Italy the use of social media grown by 30% (Comscore, 2020). Institutional communication and specifically political parties' leaders have also followed this trend, concentrating their activities of communication on social media which, above all on Twitter, have found a large space of narration and representation (Paolillo & Forciniti, 2021). The language of politics represents the language of power (Lasswell, 1979) as exercise of persuasion and with the progressive affirmation of *media logic* has generated the definition of two registers of political language: one didactic - argumentative and the other polemical (Cepernich & Novelli, 2018) based on a very strong *affective* dimension. The register of political language used by institutional actors contributed to affect the so-called *connected public culture* (Boccia Artieri, 2012) through the fragmentation of the public sphere.

In this paper, we propose the study of the political language adopted on Twitter by the main Italian parties' leaders during the first two waves of the first phase of Covid-19 by means of a 2-step analysis model which combines methods for semantic orientation and emotion classification with the multivariate approach of Correspondence Analysis. Our contribution considers the difference existing between sentiment and emotion, the need to go beyond the classification of positive - negative polarity to better understand the semantic aspect enclosed in a text. More specifically, emotions have a central role in political discourse (Huguet Cabot *et al.*, 2020). Therefore, the research aims to classify the political language adopted by Italian leaders through the emotion recognition and the factorial projection to assess if the emotional approach is useful to predict changes in the political positioning and to anticipate crises of government.

The paper explores four main research questions:

RQ1) Which emotions characterized the Italian politicians' language during the first and second waves of pandemic on Twitter?

RQ2) The emotions detected are useful to characterize the use of register didactic - argumentative or polemical?

RQ3) Were there changes in the emotions adopted by the political leaders between the first and second wave?

RQ4) Can the recognition of emotions about political language predicts the political positioning and crises of government?

To follow, section 2 presents data; section 3 describes the methodology used to reach the research questions; section 4 shows results and discussion.

2 Data

We extracted the text content released by each political actor on own official Twitter account to investigate the four research questions. Data extraction was performed using the open-source programming interface Twitter *API* and the *library rtweet* (Kearney, 2019). The official accounts of seven political figures whose hierarchy and popularity dominated the national scene during the first two pandemic waves: Silvio Berlusconi, Giuseppe Conte, Luigi Di Maio, Giorgia Meloni, Matteo Renzi, Matteo Salvini and Nicola Zingaretti have been followed and downloaded. Two temporal slots have been considered: the first wave, from first February to 30 May 2020, and the second one, from first September to 30 November 2020. The monitoring period represents the main media hype about the pandemic, where both citizens and politicians are living uncertainty and bewilderment. Data extraction in two separate periods determined that two different corpora were obtained, thus, each corpus represents a different case of analysis. During the first wave 32,551 tweets have been obtained and the greater number, about it 69.16% (*Table 1*), comes from Matteo Salvini. With the second wave, we extracted 47,094 tweets, recording also in this case the Salvini's leadership in terms of posting on Twitter (74.10%).

Table 1: Percentage of tweets produced by leaders during the first and second wave

Leader	*Percentage of tweets*	
	First wave	*Second wave*
Silvio Berlusconi	7.47%	5.31%
Giuseppe Conte	3.39%	2.19%
Luigi Di Maio	5.64%	2.99%
Giorgia Meloni	9.02%	9.97%
Matteo Renzi	1.57%	2.28%
Matteo Salvini	69.16%	74.10%
Nicola Zingaretti	3.75%	3.16%

3 Methodology

To study the political language used by party leaders during the first and second pandemic wave and to determine if there were changes between the two waves, we adopted a 2-step analysis model.

At first step, we used emotion classification algorithms for Italian language. The literature on sentiment analysis is greater developed in English language than others which instead suffer the lack of resources, but Italian represents an interest evolving field (e.g.; Nozza *et al.,* 2020; Bianchi *et al*., 2021). In fact, after we cleaned the texts from link, URLs, emoji and other special characters, we used UmBERTo, a recent Italian Language Model trained with Whole Word Masking trained on

Commoncrawl ITA, a large number of Italian corpora implemented by means of Facebook Research Artificial Intelligence codes (Ott *et al*., 2019) and annotated with four basic emotions (Ekman, 1992): anger, fear, joy, sadness. The datasets consist in corpora of Italian tweets in a broad topic and domain coverage - from football matches to politics - for emotion and sentiment classification able to predict sentiments and emotions in text by training prediction models. UmBERTo uses two innovative approaches: the first, called Sentence Piece Model (SPM) is a language-independent sub-word tokenizer which creates sub-word units specifically to the size of the vocabulary through neural-based text processing; the second, denominated Whole Word Masking (WWM) applies a mask to an entire word, if at least one of all tokens was originally chosen as mask, so to maintain the sub-words. The task of emotion classification used is fine-tune called UmBERTo-FT, which obtains the best results in terms of the overall performance metrics (e.g.; accuracy, precision, recall) (Bianchi *et al.,* 2021).

At the second step, we coded information coming from the emotion recognition (ER) in a contingency table $\mathbf{T}_{lxe}$ where l indicates the leaders and e denotes the categories of emotions. The generic element t_{le} presents the total value of emotion detected into tweets of each leader for each emotion category.

Finally, the table $\mathbf{T}_{lxe}$ has been analyzed by the Correspondence Analysis (CA) (Benzécri, 1973; Greenacre, 1984). The performing of CA on $\mathbf{T}_{lxe}$ allowed to observe the association between leaders and categories of emotions in order to describe which emotions better characterized their political language at the time of Covid-19. At the same time, this allowed to detect similarities among leader according to emotions of political language adopted on Twitter.

This 2-step analysis has been performed separately both for the first wave corpus and for that the second wave in order to intercept changes in political language between the two periods examined.

4   Results and discussion

The CA map (*Figure 1*) shows the association between the leaders and the four categories of emotions (refers to Ekman, 1992) for the first pandemic wave. The factorial projection explains 77.3% of variance through the first dimension and 21.7% by means of the second dimension. We can observe the plane how split in four sections, each of which is characterized by a category of emotion ascribable to each leader. The emotional dimension can be interpreted by looking at the map from left to right. This interpretation makes it possible to determine a political language characterized by the shifting from anger to fear. More specifically, the exploratory analysis suggests that Silvio Berlusconi's political language was better represented by sadness; that adopted by Luigi di Maio and Nicola Zingaretti was instead imbued with fear; finally, Giorgia Meloni and Matteo Salvini's language were both associated to anger (RQ1).

These three association areas between political actors and emotions allowed to identify how the use of linguistic expressions with a strong emotional weight have a

central role in their political discourses, hence we can attribute to these leaders the use of register of political language very affective. However, we cannot ascribe any category of emotions to Matteo Renzi and Giuseppe Conte, thus we can find the use of a political language register that was not based on affective dimension during the first pandemic wave (RQ2).

Figure 1: Correspondence Analysis Map "leader by emotion" for the first wave
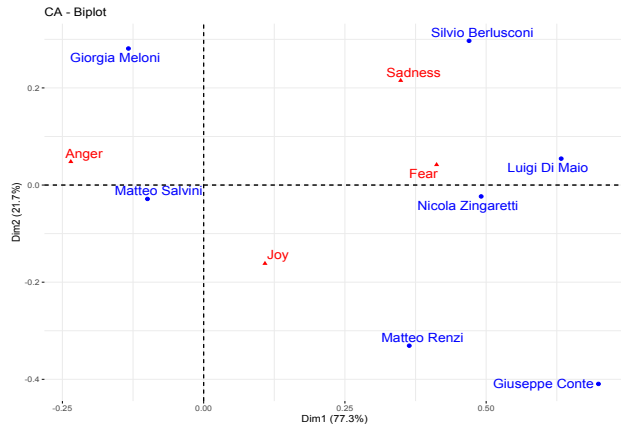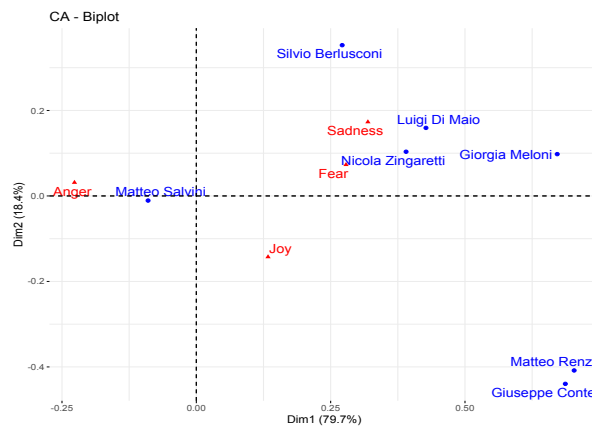

Figure 2: Correspondence Analysis Map "leader by emotion" for the second wave

The CA map referred to second pandemic wave (*Figure 2*) explains 79.7% of variance by means of the first dimension and 18.4% through the second dimension. On the horizontal axis, we affirm the transition from a political language based on anger to a language based on fear. In this case, we can intercept on the plane three sections which describe the emotions and political actors. More precisely, we observe the association between Luigi Di Maio, Nicola Zingaretti and Giorgia Meloni with emotions as fear and sadness; while Matteo Salvini's language remained characterized by anger (RQ1). No emotional category can be associated

with Matteo Renzi and Giuseppe Conte (RQ1), confirming a non-affective political language different from the other leaders mentioned above (RQ2).

Through the comparison between the CA maps of the first and second wave (*Figure 1* and *2*), it emerges a change in the political language, since in the *Figure 1* the language of Di Maio and Zingaretti was best characterized by fear, while in *Figure 2* we can see also sadness; as well as Meloni's language shifted from anger to sadness and fear. In fact, Meloni approached Di Maio and Zingaretti's language and moved away from Salvini. No change for Salvini, who remained characterized by anger, as well as Conte and Renzi that they are not associated with any category of emotions between the first and second wave (RQ3). ER in the political language adopted on Twitter between Meloni and Salvini in the second wave allowed to predict the rift and disaggregation of the right-wing coalition before the Draghi's government established in February 2021 (RQ4). An interesting aspect is the proximity between Conte and Renzi detected in the second wave, given that this last determined the Conte's government crisis. The first impression can indicate closeness in political strategic positioning as in the first wave. However, the map may show a different interpretation: the tactical outsider represented by Renzi, who based his language on emotionless register in the opposite to Meloni, Salvini, Zingaretti, to obtain consensus without polemics. Thus, in the second period, Renzi and Conte continued to use a similar neutral language but for different strategies.

References

1. Benzécri, J. P.: L'Analyse des données. Tome 2: l'analyse des correspondances. Dunod, Paris (1973).
2. Bianchi,F., Nozza, D., Hovy, D.: FEEL-IT: Emotion and Sentiment Classification for the Italian Language. Proceedings of the 11th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 76--83 (2021).
3. Boccia Artieri, G.: Stati di connessione. Pubblici, cittadini e consumatori nella (Social) Network Society, Franco Angeli, Milano (2012).
4. Cepernich, C., Novelli, E.: Sfumature del razionale. La comunicazione politica emozionale nell'ecosistema ibrido dei media. Comunicazione politica (**19**)(1), 13--30 (2018).
5. Comscore: Coronavirus e nuovi comportamenti online: aggiornamento del 30 marzo 2020 (2020). Available via https://www.comscore.com/ita/Public-Relations/Blog/Coronavirus-e-nuovi-comportamenti-online-aggiornamento-del-30-marzo-2020. Last access: August 2021.
6. Ekman, P.: An argument for basic emotions. Cognition and Emotion, **6**(3-4):169--200 (1992).
7. Greenacre, M.: Theory and Applications of Correspondence Analysis. Academic Press (1984).
8. Kearney, M. W.: Rtweet: Collecting and analyzing Twitter data", Journal of Open Source Software, **4**(42), 18--29 (2019).
9. Lasswell, H.D.: Il linguaggio della politica: studi di semantica quantitativa, Nuova Eri, Torino (1979).
10. Nozza, D., Bianchi, F., & Hovy, D.: What the [MASK]? Making Sense of Language-Specific BERT Models. arXiv preprint Available via: https://arxiv.org/pdf/2003.02912.pdf (2020).
11. Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., Auli, A.: fairseq: A Fast, Extensible Toolkit for Sequence Modeling. Proceedings of NAACL-HLT 2019: Demonstrations (2019).
12. Paolillo, M., Forciniti, A.: L'impatto del Covid-19 sull'opinione pubblica: una strategia di analisi per lo studio della comunicazione su Twitter. In: in Favretto A.R., Maturo A., Tomelleri S. (eds.), L'impatto sociale del Covid-19, Franco Angeli, Milano (2021).