

EMOTIONAL TEXT MINING OF SOCIAL NETWORKS: THE FRENCH PRE-ELECTORAL SENTIMENT ON MIGRATION

Francesca Greco, Dario Maschietti, Alessandro Polli

1. Introduction

Migration has actually gained considerable relevance both in national and European political agendas and in general public debate. The migratory phenomenon, as well as its humanitarian and health relevance, is presented nowadays as a challenge for national and supranational governments, which needs coordinated responses to ensure citizens' security (Coppola & Macioti, 2017). The issue of security is also related to the terrorist attacks of religious matrix perpetrated on western countries since the beginning of the new millennium that have shaken public opinion, highlighting the issue of freedom of movement and residence for people within the European Union.

During the presidential elections in France, the populist rhetoric had largely exploited the perception of the freedom of movement as a risk factor, taking advantage of this issue in the political propaganda. The *Front National*, a far-right party led by Marine Le Pen, who was considered one of the favourite presidential candidates, quadrupled its consensus getting 25% of the vote in the European elections of 2014, focusing its political campaign on topics such as border closure and exiting from the EU.

In recent years, social media analysis has become a fast and cheap device, compared to the traditional survey, to explore the political and electoral opinions and sentiments of citizens. Moreover, social network analysis was used for several purposes, such as demonstrations and revolt organization, the engagement of individuals in mobilization, and the construction of social movements and political parties (e.g. the *Cinque Stelle* Italian political party). For this reason, social media and social network sites, like Facebook and Twitter, have started to play a growing role in real-world politics (Ceron, Curini, Iacus & Porro, 2014).

The wide diffusion of the internet increases the opportunity for millions of people to surf the web, create account profiles and search or share information daily. The constant rise in the number of users of social media platforms, such as Twitter, make a large amount of data available that represents one of the primary sources to explore people's opinions, sentiments, and emotions (Ceron, Curini &

Iacus, 2013; Pelagalli, Greco & De Santis, 2017). Therefore, texts can be analysed in order to explain and anticipate the dynamics of different events, such as stock market activity, elections, etc. (e.g., Schoen et al., 2013; Ceron, Curini, Iacus & Porro, 2014), potentially producing useful results applicable in different contexts. There are a variety of procedures used to extract such information from different types of textual data focusing on several procedures as shown by the literature (Reinert, 1983; Halfon et al., 2016; Hopkins and King, 2010; Ceron, Curini & Iacus, 2016; Pelagalli, Greco & De Santis, 2017).

In this paper, we analyse the sentiment about migration in social media during the French presidential campaign of 2017. We perform an emotional text mining (Cordella, Greco & Raso, 2014; Greco, 2016; Pelagalli, Greco & De Santis, 2017) in order to explore the emotional content of the Twitter messages concerning migration written in French in the last two weeks before the first round of the presidential election on April 23th, 2017. The aim is to analyse the opinions, feelings and shared comments, classifying the contents and measuring the sentiments. This procedure allows for the detection of the emotional representation of migration emerging from tweets in the pre-election period.

The paper is composed of 5 sections, as follows: in section 2, we describe the methodology of the emotional text mining; in section 3, we describe the collection and analysis of data; in section 4, we illustrate the main results; in section 5, we discuss the conclusions.

2. The emotional text mining methodology

We know that a person's behaviour depends not only on their rationale thinking but also, and sometimes most of all, on their emotional and social way of mental functioning (Carli, 1990; Moscovici, 2005). That is, people consciously categorize reality and, at the same time, unconsciously symbolize it emotionally (Fornari, 1976). These two thinking processes are the product of the double-logic way of the functioning of the mind (Matte Blanco, 1981) which allows people to adapt to their social environment. According to this socio-constructivist approach based on a psychodynamic model, the unconscious processes are social, as people generate interactively and share the same emotional meanings. The socially shared emotional symbolization sets the interactions, behaviours, attitude, expectation and communication processes, and for this reason, the analysis of the narrations allows for the acquisition of the latent emotional meaning of the text (Salvatore & Freda, 2011).

If the conscious process sets the manifest content of the narration, that is *what* is narrated, the unconscious process can be inferred through *how* it is narrated, that is,

the words chosen to narrate and their association within the text. We consider that people emotionally symbolize an event, or an object, and socially share this symbolisation. The words they choose to talk about this event, or object, is the product of the socially-shared unconscious symbolization (Greco, 2016).

According to this, it is possible to detect the associative links between the words to infer the symbolic matrix determining the coexistence of these terms in the text. To this aim, we perform a multivariate analysis based on a bisecting k -means algorithm (Savaresi & Boley, 2004) to classify the text, and a correspondence analysis (Lebart, Salem & Berry, 1997; Bolasco, 2013) to detect the latent dimensions setting the cluster per keywords matrix. The interpretation of the cluster analysis results allows for the identification of the elements characterizing the emotional representation of the migrants, while the results of correspondence analysis reflect its emotional symbolization (Cordella, Greco & Raso, 2014; Greco, 2016; Pelagalli, Greco & De Santis, 2017). The advantage connected with this approach is to interpret the factorial space according to words polarization, thus identifying the emotional categories that generate migration representations, and to facilitate the interpretation of clusters, exploring their relationship within the symbolic space.

3. Data collection and analysis

In order to explore the sentiment on migration in Twitter communications in the period immediately preceding the first round of the French presidential elections, we scraped all the messages in French produced from April 10th to April 22nd, 2017, containing the word "migration", "migrant" or "migrants", from the Twitter repository. The data extraction was carried out with the *twitteR* package of R Statistics (Gentry, 2016).

The sample of 111.767 tweets was made up of 77,7% of retweets, and resulted in a large corpus of 2.154.194 tokens. In order to check whether it was possible to statistically process data, two lexical indicators were calculated: the type-token ratio and the percentage of hapax (TTR = 0,01; Hapax percentage = 40,4). According to the large size of the corpus, both lexical indicators highlight its richness and indicate the possibility of proceeding with the analysis.

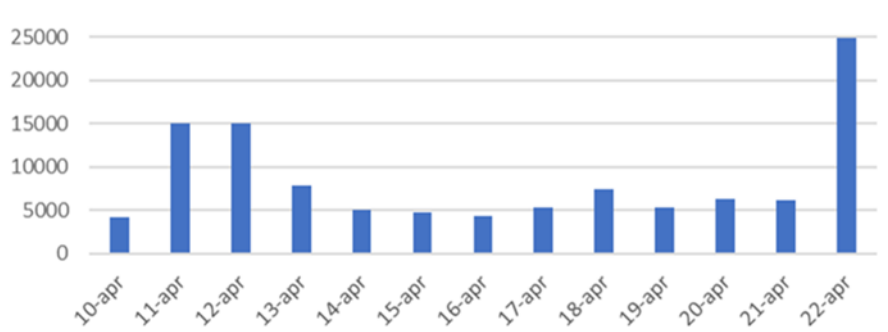
First, data were cleaned and pre-processed with the software T-Lab (Lancia, 2017) and keywords selected. In particular, we used lemmas as keywords instead of type, filtering out the lemma "migrant" and those of the low rank of frequency (Greco, 2016). Then, on the tweets per keywords matrix, we performed a cluster analysis with a bisecting k -means algorithm (Savaresi & Boley, 2004) limited to ten partitions, excluding all the tweets that did not have at least two keywords co-

occurrence. The eta squared value was used to evaluate and choose the optimal solution. To finalize the analysis, a correspondence analysis on the keywords per clusters matrix was made (Lebart, Salem & Berry, 1997) in order to explore the relationship between clusters, and to identify the emotional categories setting migrant's representations (see paragraph 2).

4. Main results and discussion

The number of messages produced in the period from April 10th to April 22nd, 2017 changed over the time. As shown in figure 1, there are two peaks in the distribution that correspond to two relevant events connected to the presidential campaign. The first was on April 11th in an interview when Mélenchon, president of the far-left wing political party *La France Insoumise* and one of the four favourite candidates to the presidential election, suggested the possibility of hosting immigrants in French historical buildings. This proposal shocked public opinion, and probably was used in the electoral debates in the days following. The second peak was the day before the election, on April 22nd, when all the candidates could perform their last electoral discourses before the vote, as during the vote it is forbidden to publicly comment on the presidential election.

Figure 1 – Number of tweets collected per day from April 10th to April 22nd, 2017



In the evening, three days before the election, there were two small peaks of messages: the first after a terrorist attack where a policeman was killed at Champs Elysée in Paris, and two hours later, when the President commented on the event (figure 2). It is interesting to note that Twitter's users produced a very small number of messages with the term *migrant* after the attack.

The results of the cluster analysis show that the keywords selection criteria allow the classification of 89,8% of the tweets. The eta squared value was calculated on partitions from 3 to 9, and it shows that the optimal solution is seven

clusters ($\eta^2 = 0,57$). The correspondence analysis detected six latent dimensions, and the explained inertia for each factor is reported in table 1.

Figure 2 – Number of tweets produced from 8pm to 10pm on April 20th, 2017, time of the terrorist attack in Paris.

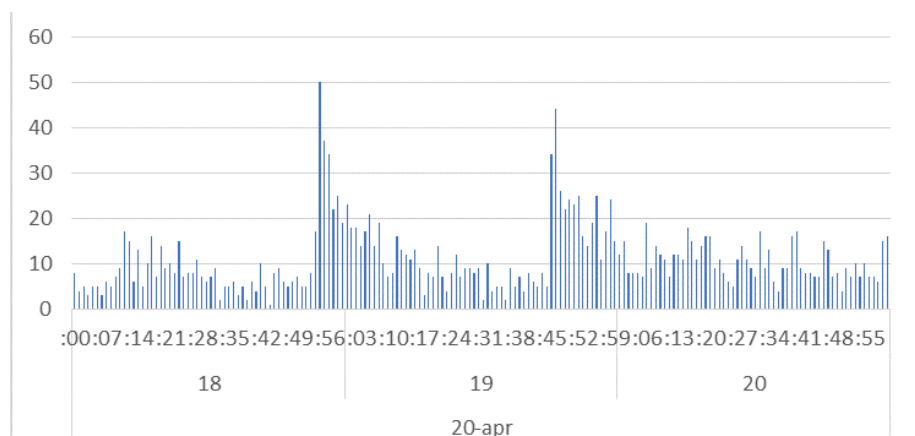


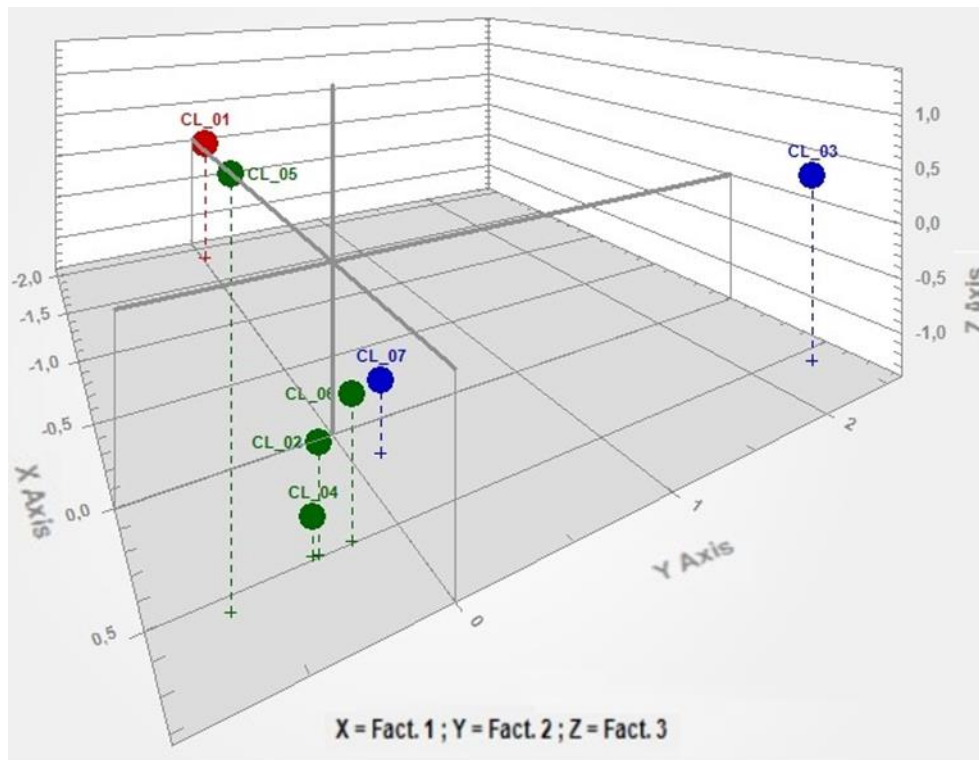
Table 1 – Explained inertia for each factor.

Factor	Eigenvalues	%	Cumul. %
1	0,956	24,9	24,9
2	0,737	19,2	44,1
3	0,659	17,2	61,3
4	0,548	14,3	75,5
5	0,504	13,1	88,6
6	0,436	11,4	100,0

In figure 3, we can appreciate the emotional map of the migration emerging from the French tweets. It shows how the clusters are placed in the factorial space produced by the first three components, explaining about 60% of the inertia. The first component represents the flow of people that arrive and stay; the second factor reproduces the double-aspect of policy, abstract and concrete distinguishing the political rhetoric and the social reality; the third component represents the empowerment of the citizen and the effectiveness of the action, which can be effective or ineffective. The fourth component is death, the tragic aspect of the mobility; the fifth factor is the medialization of the migration phenomenon, the spectacle, distinguishing between serious (politics) and recreational (sport); and the sixth component is people's reaction, that moves from rejection to the hospitality.

These six components set the symbolic space in which clusters are located, facilitating their interpretation. For example, both cluster 1 and cluster 3 are isolated from the rest of the clusters in the factorial space set by the first and the second component. These clusters are probably connected to spot events largely covered by the media that made a buzz: cluster 1, as it refers to Mélenchon interview, where he suggests the hosting solution for refugees; and cluster 3 is probably connected to the spread of information about human trafficking report promoted by the OSCE (Organization for Security and Co-operation in Europe).

Figure 3 – Factorial space set by the first three components



The seven clusters are of different sizes (figure 4) and reflect different representations of migration that correspond to three different sentiments: positive, negative for the community, and negative for migrants (table 2). The first cluster reflects the reaction of public opinion on the proposal to host migrants in historical buildings, suggested by Mélenchon. Migrants seem to be perceived as squatters, or

undesirable guests. The second label highlights the association that is made between migration and security.

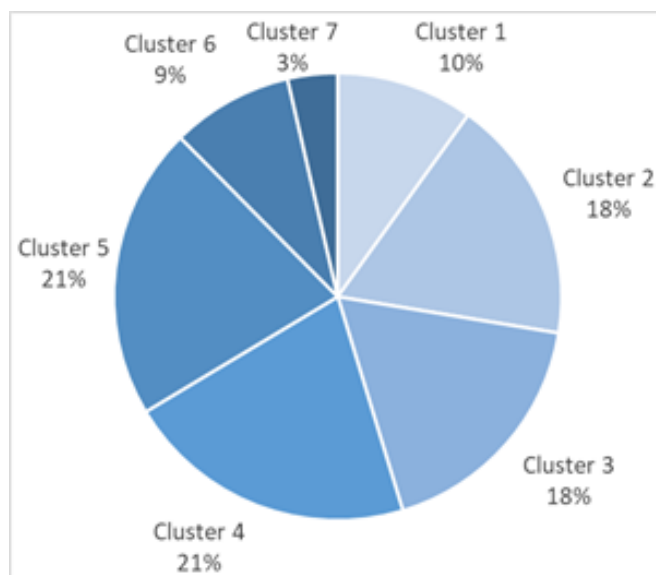
Table 2 – Migration representations and sentiments.

Cluster	N tweets classified	Size	label	keyword	CU	Sentiment
1	951	9,9%	Squatters	Mélenchon	20856	Negative for the community
				installer	20305	
				réponse	20228	
				énorme	20218	
				établissement	20165	
				Iglesias	152	
2	17770	17,7%	Terrorists	terroriste	1664	Negative for the community
				enfant	1562	
				Bataclan	1178	
				juif	1168	
				moi	1201	
				Le Pen	1082	
3	3571	3,6%	Slaves of the Trafficking	vendre	6111	Negative for the migrant
				Libye	5029	
				marché	3943	
				africain	3720	
				esclave	3653	
				esclavage	3046	
4	21178	21,2%	EU Solidarity Target	accueil	2619	Positive
				UE	2448	
				France	2169	
				pays	1990	
				accueillir	1889	
				migration	1446	
5	17770	17,7%	Invaders	camp	3581	Negative for the community
				cons	3238	
				main	3048	
				attraper	2940	
				deconne	2938	
				flux	1808	
6	21315	21,1%	Sport Heroes	Europe	12657	Positive
				mettre	2251	
				France	2242	
				français	1756	
				champion	1660	
				seul	1653	
7	8795	8,8%	Migration victims	Méditerranée	1261	Negative for the migrant
				mourir	1209	
				mort	879	
				silence	769	
				secourir	733	
				minute	699	

The first six keywords of the cluster are ordered by the number of context units (CU) classified in the cluster.

Migrants are represented as terrorists, as aggressors, and the variety of citizens' religious choices are perceived as an obstruction to cohabitation. Migration is represented as a risk factor for security because it favours terrorism, and this association seems to be driven by the speeches of Marine Le Pen. The third cluster refers to human trafficking, that changes the sense of mobility from a voluntary choice to an unfortunate destiny, transforming the journey of hope into a nightmare and the migrants into slaves. The fourth cluster reflects the EU reception policy, in which receiving, hosting, and integrating migrants are some of the actions needed that call for solidarity from French citizens. In the fifth cluster, migrants are perceived as invaders, which is a thematic that often appears in Marine Le Pen's electoral discourses.

Figure 4 – Cluster size

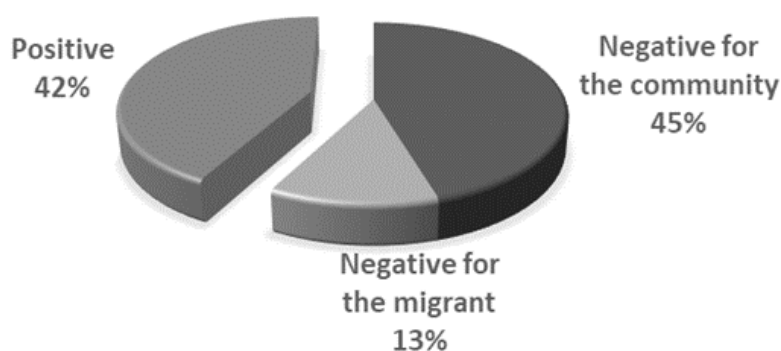


Among the words of this cluster, there are words such as *border*, *stop*, *wave*, *to control* (*frontière*, *arrêter*, *vague*, *maîtriser*) which are associated insults and swearing, highlighting the elevated level of anger. The sixth cluster represents migrants such as sport heroes, bringing prestige to the country. There are the first names, or surnames of famous football players and artistic gymnastic champions among the word of this cluster that are associated with *goal*, *cup* but also *diversity*. In this cluster, diversity seems to be a positive value that distinguishes sport players. Lastly, the seventh cluster reflects the dangers entailed in the migration

journey that puts people's life in jeopardy. Only a part of the sea travellers through the Mediterranean Sea will land, some of them will never arrive nor return.

By the clusters interpretation, we detected seven different representations of migrants that correspond to three different sentiments: positive, negative for the community, and negative for migrants (table 2). We have considered as negative the representation of migrants as squatters, invaders, terrorists, trafficking slaves and migration victims, and positive the sport heroes and the EU solidarity target. Among the negative clusters, we distinguished negativity according to the direction of the action: squatters, terrorists and invaders are negative for the community and trafficking slaves and migration victims are negatives for migrants themselves. In figure 5, the size of the three sentiments are displayed.

Figure 5 – Sentiment on migration size



5. Conclusion

In this study, we show that the representation of migration in social media presents seven distinctive profiles and three sentiments that seem to be connected to the French presidential campaign. While the terrorist attack three days before the first round of voting in the centre of Paris has slightly modified the production of messages, the candidates' interviews have a higher impact. Most probably, it is the medialization that impacts on the number of tweets produced.

The results highlight that a negative sentiment prevails towards migration, as only 42% of the messages classified are positive. Among negative sentiments, 45% of discourses classified reflect citizen perception of being invaded or attacked, while only 13% consider migrants as victims. Moreover, in line with the literature, the social media analysis seems to be a fast and cheap device to explore political and electoral opinions. In fact, although the analysis is made only on the messages

produced by Twitter users on a specific topic and cannot be extended to all the voters, it is interesting to note that Marine Le Pen, who has largely used the negative sentiment on migration in the presidential election campaign, lost the election with 34% of votes on the second round. This percentage almost corresponds to the sum of clusters 2 and 5 (35,4%) in which migrants are represented as terrorists and invaders.¹

In conclusion, the emotional text mining presented in this research defines a clear framework for the sentiment that French citizens have on migrations. It allows for the identification of the themes connected with positive and negative sentiments, offering a different way of looking to this phenomenon.

References

- BOLASCO S. 2013. *L'analisi automatica dei testi: fare ricerca con il text mining*. Roma: Carocci Editore.
- CARLI R. 1990. Il processo di collusione nelle rappresentazioni sociali. *Rivista di Psicologia Clinica*, Vol. 4, pp. 282-296.
- CARLI R., PANICCIA R. M. 2002. *Analisi Emozionale del Testo*. Milano: Franco Angeli.
- CERON A., CURINI L., IACUS S. M. 2016. iSA: a fast, scalable and accurate algorithm for sentiment analysis of social media content. *Information Sciences*, Vol. 367, pp. 105-124.
- CERON A., CURINI L., IACUS S. M., PORRO G. 2014. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France, *New Media & Society*, Vol 16, No. 2, pp. 340-358.
- CERON A., CURINI L., IACUS S.. 2013. *Social Media e Sentiment Analysis. L'evoluzione dei fenomeni sociali attraverso la Rete*. Milano: Springer.
- COPPOLA N. & MACIOTI M. I. 2017. Migrations and Gender. In DEL RE, E.C., LARÉMONT R. R. (Eds) *Pursuing Stability and a Shared Development in Euro-Mediterranean Migrations*, Rome: Aracne, pp. 183-201.
- CORDELLA B., GRECO F., RASO A. 2014. Lavorare con Corpus di Piccole Dimensioni in Psicologia Clinica: Una Proposta per la Preparazione e l'Analisi dei Dati. In NEE E., DAUBE M., VALETTE M., FLEURY S. (Eds) *Actes JADT 2014, 12es Journées internationales d'Analyse Statistique des Données Textuelles, Paris, France, Juin 3-6, 2014, Lexicometrica*, pp. 173-184.
- FORNARI F. 1976. *Simbolo e codice: Dal processo psicoanalitico all'analisi istituzionale*. Milano: Feltrinelli.
- GENTRY J. 2016. R Based Twitter Client. R package version 1.1.9.

¹ We did not consider the cluster 1 as it refers to the Mélenchon's proposition to host migrants in historical buildings.

- GRECO F. 2016. *Integrare la disabilità: Una metodologia interdisciplinare per leggere il cambiamento culturale*. Milano: Franco Angeli.
- HALFON S., ÇAVDAR A., ORSUCCI F., SCHIEPEK G. K., ANDREASSI S., GIULIANI A., DE FELICE, G. 2016. The non-linear trajectory of change in play profiles of three children in psychodynamic play therapy. *Frontiers in Psychology*, 7(OCT) doi:10.3389/fpsyg.2016.
- HOPKINS D., KING G. 2010. A method of automated nonparametric content analysis for social science, *American J. Pol. Sci.*, Vol. 54, No. 1, pp. 229-247.
- LANCIA F. 2017. User's Manual : Tools for text analysis. T-Lab version Plus 2017.
- LEBART L., SALEM A., BERRY L. 1997. *Exploring textual data. Vol. 4*. New York: Springer Science & Business Media.
- MATTE BLANCO I. 1981. *L'inconscio come insiemi infiniti: Saggio sulla bi-logica*, Torino: Einaudi.
- MOSCOVICI S. 2005. *Le rappresentazioni sociali*. Bologna: Il Mulino.
- PELAGALLI f., GRECO F., DE SANTIS E. 2017. Social emotional data analysis. The map of Europe. In PETRUCCI A. & VERDE R. (Eds) *SIS 2017. Statistics and Data Science: new challenges, new generations. Proceedings of the Conference of the Italian Statistical Society, Florence 28-30 June 2017*, Firenze: Firenze University Press, pp. 779-784.
- REINERT A. 1983. Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte, *Les cahiers de l'analyse des données*, Vol. 8, No. 2, pp. 187-198.
- SALVATORE S., FREDA M. F. 2011. Affect, unconscious and sensemaking: A psychodynamic, semiotic and dialogic model. *New Ideas, Psychology*, Vol. 29, pp. 119-135.
- SAVARESI, S. M., BOLEY, D. L. 2004. A comparative analysis on the bisecting *k*-means and the PDDP clustering algorithms. *Intelligent Data Analysis*, Vol. 8. No. 4, pp. 345-362.
- SCHOEN H., GAYO-AVELLO D., METAXAS P., MUSTAFARAJ E., STROHMAIER M., GLOOR P.. 2013. The power of prediction with social media, *Internet Res.*, Vol. 23, No. 5, pp. 528-543.

SUMMARY

Emotional text mining of social networks: The French pre-electoral sentiment on migration

Migration has actually gained considerable relevance both in the national and European political agendas and in general public debate. It is a challenge for governments, which needs coordinated responses to ensure citizens' security. Furthermore, the terrorist attacks against western countries have called into question freedom of movement and residence for people within the European Union. In the electoral campaign, the populist rhetoric on migration has largely exploited citizens' perception of insecurity, as in the French presidential electoral campaign of 2017. In order to analyse public sentiment on migration, we collected with *twittweR* of R Statistics a sample of 111767 messages containing the word "migrant" produced in the last two weeks before the first-round votes from the Twitter repository. The messages were collected in a large size corpus of over two million tokens to which we applied multivariate techniques, i.e. cluster analysis with a bisecting *k*-means algorithm and a correspondence analysis on the keyword per cluster matrix, in order to identify the contents and the sentiments behind the shared comments. The results show how the clusters and the factorial space are representative of the different ways of emotionally representing migration, highlighting the relevant aspects perceived by those who choose to express themselves through Twitter.

Francesca GRECO, Prisma S.r.l.; Dipartimento di Psicologia Dinamica e Clinica,
Sapienza Università di Roma, francesca.greco@uniroma1.it

Dario MASCHIETTI, Prisma S.r.l., d.maschietti@prisapaprogetti.it

Alessandro POLLI, Dipartimento di Scienze Sociali ed Economiche, Sapienza
Università di Roma, alessandro.polli@uniroma1.it