



**HAL**  
open science

## The flow of ideologies between a political figure and a militant community: a CMC corpora analysis

Julien Longhi, Claudia Marinica, Zakarya Després, Clément Plancq

### ► To cite this version:

Julien Longhi, Claudia Marinica, Zakarya Després, Clément Plancq. The flow of ideologies between a political figure and a militant community: a CMC corpora analysis. 6th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora, Sep 2018, Anvers, Belgium. hal-01872152

**HAL Id: hal-01872152**

**<https://hal.science/hal-01872152>**

Submitted on 24 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Flow of Ideologies Between a Political Figure and a Militant Community: A CMC Corpora Analysis

Julien Longhi<sup>1</sup>, Claudia Marinica<sup>2</sup>, Zakarya Després<sup>3</sup>, Clément Plancq<sup>4</sup>

<sup>1</sup> Paris Seine University, EA7392 AGORA, IDHN

<sup>2</sup> Paris Seine University, ETIS (ENSEA/UCP/CNRS UMR 8051), IDHN

<sup>3</sup> Paris Seine University, IDHN

<sup>4</sup> Paris 3 University, Lattice (ENS/ CNRS UMR 8094)

julien.longhi@u-cergy.fr, claudia.marinica@u-cergy.fr, zakarya.despres@gmail.com, clement.plancq@ens.fr

## Abstract

This paper presents an exploratory work on the valorisation of political ideologies in CMC corpora, and more precisely their flow between a political figure and a militant community. Our main goal is to put forward a methodology for analysis and to present our preliminary results. We consider that the CMC corpora we used are traces that allow us to answer to political and social interrogations: first we present the context of the study, the corpora, and the addressed themes. Then, we present our methodology, and the intermediary results obtained with two softwares (Iramuteq and TXM), that led us to build an original analysis method. Finally, we compare the results from both corpora to provide hypothetical answers to the political question of the construction and circulation of political ideas from a movement with a charismatic leader and an important militant community.

**Keywords:** CMC corpora, textometry, digital spaces, tweets, forums.

## 1. Introduction

This article is the continuation of Djemili et al. (2014), Longhi (2017), Longhi et al. (2017), Plancq et al. (2018) and Marinica et al. (2018), and it aims to complete the analyses of the digital political discourse on Twitter through other CMC communication mediums, like the forums. To this end, we chose to study the tweets and forum threads during the 2017 French presidential election. We collected all tweets from the account @JLMelenchon, which belongs to the leader of the movement and candidate to the election, as well as posts from the "Blabla 18-25" forum on the website "Jeuxvideo.com", and messages from the "Discord des Insoumis".

We consider that the CMC corpora we use are traces that allow us to answer to political and social interrogations; by doing so, we deal with complex questions (defining ideologies, identifying their possible flow, creating a timeline for the emergence of ideas and political assertions) by computer processing, on a basis of structured, comparable social data.

More precisely, this paper presents exploratory work on the valorisation of political ideologies in CMC corpora, especially their potential flow between a political figure and a militant community. Our main goal is to put forward methods and analyses, supported by a set of results that open up new questions.

This paper is structured as follows: first, we present the context of the study, the corpora, and the addressed issues. Then, we present our methodology, and the intermediary results obtained with two softwares, that led us to build an original analysis method. Finally, we compare the results from both corpora to answer the political question of the construction and flow of political ideas from a movement with a charismatic leader and an important militant community.

## 2. Context, data and goals

### 2.1 Context

In this work, we focus on the candidate Jean-Luc Mélenchon and on the *France Insoumise* militant community, who named their members the *Insoumis* (the insubordinates), as they expressed online and offline.

Mélenchon is different from the other candidates to the 2017 French presidential elections, first, because he placed his digital strategy at the centre of his campaign: he was active on social media and he had a YouTube channel. Second, he was supported by self-organised militant relays with the *Discord* server created by the *Insoumis*, an idea which started in the *Jeuxvideo.com* forums. These two distinct digital spaces, an official one and a participative one, fit our objectives of analysing the circulation of *ideologies*.

According to Knight (2006), "specific *ideologies* crystallize and communicate the many beliefs, opinions and values of an identifiable group". The use of a statistical method will allow us to compare corpora, and to infer their ideology, without having to subjectively decide on the polysemous status of terms, or their possible interpretation.

### 2.2 Data

We used two corpora for our study:

- 3,036 tweets from the Twitter account @JLMelenchon, representing 51,552 words;
- messages from *Insoumis* militants on different platforms.

This second corpus comes from two sources: a first part comes from a series of threads dedicated to Jean-Luc Mélenchon and France Insoumise on the *Blabla 18-25* forum of *Jeuxvideo.com*, a generalist forum on a gaming website. Then, we extracted the messages that militants

sent on the Discord server, an instant messaging platform which is also originally gaming-related. In total, the militant corpus represents 383,403 messages with 6,850,823 words.

### 2.3 Goals of the paper

The objective of the study is to create a methodology that, according to the temporality of the messages, would allow us to see how the themes, terms and ideologies from the candidate Mélenchon are spreading in the militant community, and/or how the discussions of the militants can bring substance to the candidate’s digital discourses.

## 3. Methodology and results

In order to assess the flow of terms and to quantify it, we used a method bringing together two tools and two different perspectives.

In the first place, a textometric study (statistical text analysis) was performed with the software Iramuteq (<http://iramuteq.org>), allowing us to detect a set of lexical classes. These classes, along with the terms composing them, are used in the second software, TXM (<http://textometrie.ens-lyon.fr/>), in order to understand the temporality and the relations between the terms.

The Iramuteq software offers a set of analysis procedures for the description of a textual corpus. One of its principal methods is Alceste. This allows a user to segment a corpus into *context units*, to make comparisons and groupings of the segmented corpus according to the lexemes contained within it, and then to seek *stable distributions* (Reinert, 1998).

In addition to the Alceste method, Iramuteq provides other analysis tools including prototypical analysis, similarities analysis, and word clouds analysis. All of these methods allow the users of this tool to map out the dynamics of the discourses of the different subjects engaged in interaction (Reinert, 1999).

One method used by Alceste is the hierarchical descending classification (HDC). This method offers a global approach to a corpus. The HDC, after partitioning the corpus, identifies statistically independent word classes (forms). These classes are interpreted through their profiles, which are characterized by specific correlated forms. The HDC provides as a result a dendrogram.

More precisely, as explained by Camargo & Justo (2016), text segments (TS) are clustered according to their vocabularies and distributed according to the reduced forms frequencies. The descending hierarchical analysis uses matrices that cross reduced forms with TS (in repeated texts of X2 type). This method allows users to obtain a definitive classification. We obtain TS clusters with similar vocabulary within, but different from other segments. The software computes descriptive results of each cluster conforming to its main vocabulary and words with asterisk (variables). This analysis gives another way of presenting data, derived from a correspondence factor analysis. Based on the chosen clusters, the software calculates and provides the most typical TS of each cluster, giving context to them. These word clusters and

TS integrate several segments according to the vocabulary distribution.

The authors explain that on the interpretative level, it depends on the theoretical scope of the research. For example, Reinert (1990), when studying French literature, considered each cluster as a “world”, a cognitive-perceptive framework with a certain temporal stability related to a complex environment. From another point of view, research in linguistics considers these clusters as lexical fields or semantic contexts.



Figure 1: The result of the Hierarchical Descending Classification

This analysis highlights four disjoint classes presented in Figure 1, which can be interpreted as four major themes / areas of Mélenchon's tweets: Class 1 (nearly 29.9% of the vocabulary) concerns the economy; Class 2 (29.2%) concerns Europe and related defense issues; Class 3 (15.6%) concerns institutional issues; and Class 4 (25.2%) mainly includes SEO or communication terms. Each class is described by a set of words that will be used further in the analysis; one word belongs to only one class. In the following analyses we focus on the first three classes.

The following analyses were processed with the textometry software TXM, with which we partitioned the corpora according to the dates of each tweet or message. This allowed us to get the daily frequency of each word extracted from the classes found with Iramuteq. In order to visualise the evolution in time of the utilisation of these terms, we used RAWGraphs (<https://rawgraphs.io>) to produce the following horizon graphs, readable as a heatmap-like timeline.

#### 4. Results and interpretations of the results

Following this method, we obtained several types of results: general results about the presence of classes during our time period (without differentiating specific words), and specific class by class analyses with a visualisation of the repartition of each word in the class. For this paper, we will focus on the two last weeks of February 2017, which provided the most relevant results. In this period, we collected 659 tweets from Mélenchon, and 50954 messages from the Insoumis.

##### 4.1 Global analysis of the repartition of classes

The global analysis that we carried out is presented in Figure 2. Over the month, we found 147 occurrences of words from Class 1 in Mélenchon’s tweets (JLM) and 1700 in Insoumis’ messages, 24 words of Class 2 from JLM and 1033 from Insoumis, and 85 words of Class 3 from JLM and 705 from Insoumis. In order to better represent the frequency peaks in each corpus despite their size discrepancy, the following figures use different scales for JLM tweets and Insoumis messages.

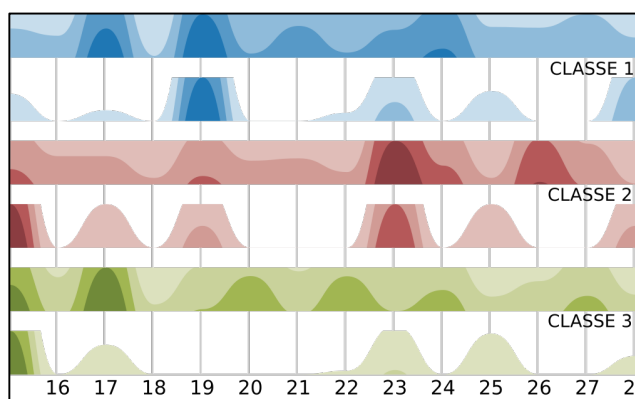


Figure 2: Frequency of the classes in February 2017. Top: Insoumis, bottom: Mélenchon.

Over the period, we can observe the temporal repartition of the apparition of classes, that allows us to evaluate if a dependence between the candidate Mélenchon and the Insoumis community exists with the use of lexicons that can assess the presence or absence of some themes.

If we take a closer look at Class 1 in Figure 2, dealing with economical issues with words like *impôt* (taxes), *euro*, *payer* (to pay), etc., we can see that terms which are frequent in the candidate’s tweets (on the 19th of February) are used later in the community (on the 19th, 23rd and 25th of February). But, this class was also used on the 17th of February in the community, thus, we cannot conclude on the flow of Class 1 without having a more specific analysis of the class word by word. We can just admit that the dynamic is not unilateral, and the discursive relationships between Mélenchon and the Insoumis are more complex than they appear. To verify this, we propose to focus on the terms of Class 1.

##### 4.2 Specific analysis of the economy-themed words

By zooming in on Class 1, we find the following terms: *coûter* (to cost), *euro*, *impôt* (tax), *milliard* (billion), *payer* (to pay), *retraite* (retirement), *salaire* (salary), *santé* (health), *SMIC* (minimum salary), *social*, *sécurité* ‘*sociale*’ (social security system), *travail* (work/job).

By projecting the frequency of several of these terms (the more relevant/interesting ones) on a temporal axis, we obtain the visualisation in Figure 3.

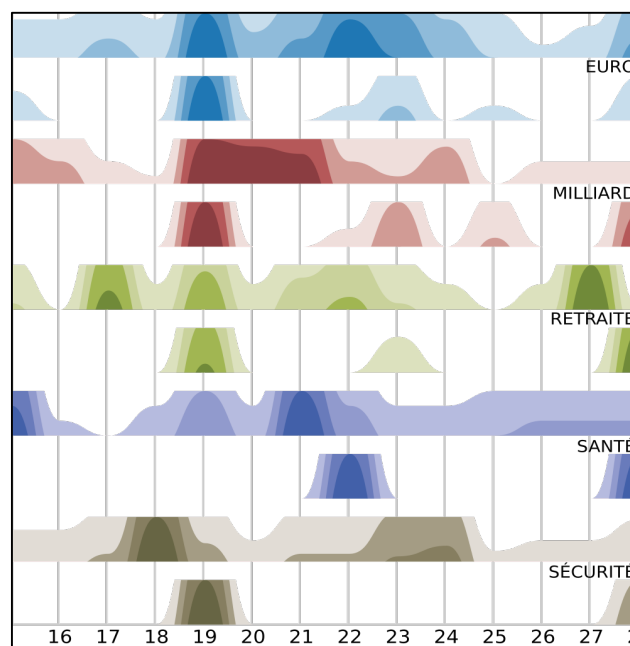


Figure 3: Frequency of the words from Class 1 in February 2017. Top: Insoumis, bottom: Mélenchon.

Economy-themed words are used a lot on the 19th of February, and continue to make echo in the community, like *milliards* or *euros*. Other terms, like *retraite*, *santé* and *sécurité* ‘*sociale*’, seem to come from the community first, and are then reused by the candidate. In Figure 4 we can see an example of a tweet by Mélenchon, posted on February 19th, which contains the words “milliard” and “euros”, saying “We are going to invest 7 billion euros in public service”.



Figure 4: Tweet of @JLMelenchon on 19/02/17.

Three days after, we find discussions on the same themes using the same terms on Discord, such as the one presented in Figure 5 saying “If you do the math, for less than 10 billion, you can raise the RSA to 1100 euros per month”.

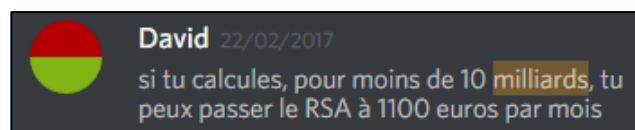


Figure 5: Message on Discord on 22/02/17.

For comparison purposes, the terms from Class 3, that refers to institutional issues, are presented in a different way, as shown in Figure 6. The temporality of these terms, but also their intensity, show that they concern broader discussions for the campaign, and that the candidate uses them occasionally.

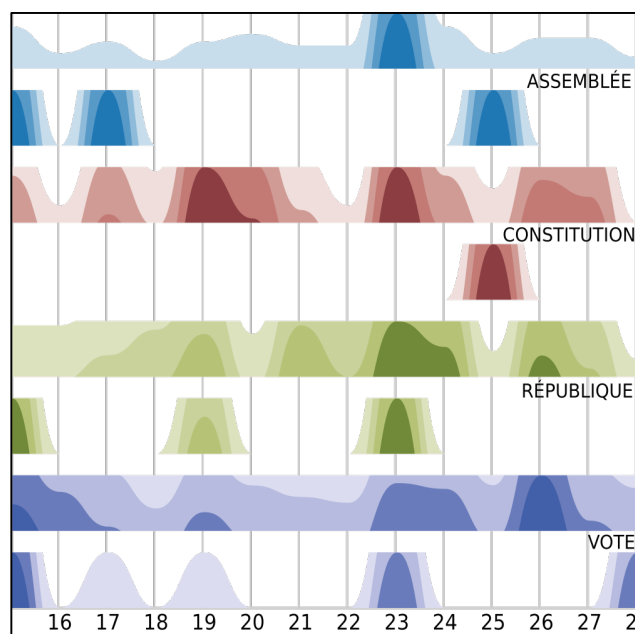


Figure 6: Frequency of words from Class 3 in February 2017. Top: Insoumis; bottom: Mélenchon.

## 5. Conclusions

This paper, still being exploratory, on the ideological circulation between political discourse and militants, allowed to build a method that highlights political themes, through the Iramuteq software, then showcases their quantitative and temporal values thanks to the TXM and RAW Graphs softwares. These preliminary results provide important perspectives on the work and ongoing analyses will bring us precisions on the link between the visualisation lines, and locate these phenomena accurately in the corpora. The current results provide already interesting insights on how to grasp this complex phenomenon, and to measure the porosity between two types of CMC corpora.

## 6. References

- Camargo, B., Justo, A. (2016). Iramuteq Tutorial [http://iramuteq.org/documentation/fichiers/IRaMuTeQ%20Tutorial%20translated%20to%20English\\_17.03.2016.pdf](http://iramuteq.org/documentation/fichiers/IRaMuTeQ%20Tutorial%20translated%20to%20English_17.03.2016.pdf)
- Djemili, S., Longhi, J., Marinica, C., Kotzinos, D., Sarfati G.-E. (2014). What does Twitter have to say about ideology? In Gertrud Faaß & Josef Ruppenhofer (Eds.). *NLP 4 CMC: Natural Language Processing for Computer-Mediated Communication / Social Media - Pre-conference workshop at Konvens Oct 2014, Germany*. Universitätsverlag Hildesheim, 1, pp. 16--25. <halshs-01058867v2>
- Kathaleen, K. (2006). Transformations of the concept of ideology in the twentieth century. *American Political Science Review*, pp. 619--626.
- Longhi, J., Marinica, C., Haddioui, N. (2016). Extraction automatique de phénomènes linguistiques dans un corpus de tweets politiques : quelques éléments méthodologiques et applicatifs à propos de la négation. In *Res per nomen 5, Négation et référence*, Reims: EPURE.
- Longhi, J., Saigh, D. (2016). A textometrical analysis of French arts workers “fr.Intermittents” on Twitter. In *Proceedings of the 4th Conference CMC and Social Media Corpora for the Humanities*, Ljubljana, Slovenia, pp. 44--47.
- Longhi, J., Marinica, C., Hassine, N., Alkhouli, A. Borzic, B. (2017). The #Idéo2017 Platform. *Conference on Computer-Mediated Communication and Social Media Corpora for the Humanities*, Oct 2017, Bolzano, Italy. <hal-01619236>
- Marinica, C., Longhi, J. Hassine, N., Alkhouli, A. Borzic, B. (2018). #Idéo2017: une plateforme citoyenne dédiée à l'analyse des tweets lors des événements politiques. In *Extraction et Gestion des Connaissances (EGC)*, 2018.
- Planq, C., Després, Z., Longhi, J. (2018). “L'avenir en commun” des Insoumis. Analyse des forums de discussion des militants de la France Insoumise. In *Atelier Fouille de Données Complexes*, EGC 2018, Jan 2018, Paris, France. <halshs-01719374>
- Reinert, M. (1990). ALCESTE, une méthodologie d'analyse des données textuelles et une application: *Aurélia* de G. de Nerval. *Bulletin de méthodologie sociologique*, 28, pp. 24--54.
- Reinert, M. (1998). Quel objet pour une analyse statistique du discours? Quelques réflexions à propos de la réponse Alceste. *Actes des 4èmes JADT*. URL : <http://lexicométrica.univ-paris3.fr/jadt/jadt1998/reinert.htm>.
- Reinert, M. (1999). Quelques interrogations à propos de l'objet d'une analyse de discours de type statistique et de la réponse « Alceste ». *Langage et société*, 90(1), pp. 57--70.
- Iramuteq: [www.iramuteq.org](http://iramuteq.org)  
TXM: <http://textometrie.ens-lyon.fr>