



HAL
open science

MigrTwit Corpora. (Im)migration Tweets of French Politics

Sangwan Jeon

► **To cite this version:**

Sangwan Jeon. MigrTwit Corpora. (Im)migration Tweets of French Politics. Louis Cotgrove; Laura Herzberg; Harald Lungen; Ines Pisetta. Proceedings of the 10th International Conference on CMC and Social Media Corpora for the Humanities, University of Mannheim; Leibniz-Institut für Deutsche Sprache (IDS), 2023, 978-3-937241-95-1. 10.14618/1z5k-pb25 . hal-04286931

HAL Id: hal-04286931

<https://hal.science/hal-04286931>

Submitted on 17 Nov 2023

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.

MigrTwit Corpora. (Im)migration Tweets of French Politics.

Sangwan Jeon

Université de Lille

E-mail: sangwn.jeon@univ-lille.fr

Abstract

Since the early 2010s, French politicians have steadily utilized Twitter as a communication tool. Political tweets about immigration are prolifically produced by Marine Le Pen (former leader of the French far-right populist Party Rassemblement National). Their biased social representation of migrants, immigrants, and asylum seekers seems to be instilled in manifold voters through online public debate and media. To characterize political immigration discourse on Twitter, we developed the diachronic bilingual corpus of political tweets posted throughout the last 12 years, from 2011 to 2022. The whole MigrTwit corpus consists of three subcorpora, for a total of 23869 tweets, with 703016 words. The constitution of the French MigrTwit corpora enabled us to study the evolution of immigration discourse in comparative and corpus-based approaches.

Keywords: Corpus linguistics, Twitter, Political discourse, Critical discourse analysis, Immigration discourse, Online hate speech

1. Introduction

Immigration issues continuously constitute the core of the political agenda of far-right populist parties in Europe and in the United States, increasingly appealing to manifold voters (Wodak, 2021; Aït Abdeslam, 2021). In France, anti-immigration discourses have been tremendously produced by the far-right Party Rassemblement National¹ (henceforth RN). International and national events at various levels seem to fuel anti-immigration arguments (Wodak, 2021; Pietrandrea & Battaglia, 2022). In this era of Web 2.0, electoral campaigns and results are affected by disinformation and hate speech spread throughout social media platforms (Badouard, 2017; Allcott & Gentzkow, 2017). Since the early 2010s, French politicians have steadily utilized Twitter as a communication tool. Manipulative discourse should be considered to detect abusive language in political discourse (Macagno, 2022). The shortness of messages and the decontextualized feature of political tweets may contribute to the “imperfect and biased way our mechanisms of information processing work” (Hart, 2013). Within the framework of the research project OLINDiNUM (Observatoire LINGuistique du DIscours NUMérique [Linguistic Observatory of Online Debate]), the MigrTwit corpus has been developed, annotated, and analyzed to characterize political immigration discourses on Twitter in collaboration with Elena Battaglia, Guido Blandino, Paola Pietrandrea, and with the participation of Adelina Stoian. The MigrTwit corpus consists of 23869 tweets posted between January 2011 and June 2022, from 51 French and British political figures and parties. It has three components, i.e., the corpus of French right-wing political migr-tweets, the corpus of British right-wing political migr-tweets (cf. Blandino,

2023), and the corpus of French left-wing political migr-tweets. The whole corpus is published and downloadable². In this paper³, focusing on the French MigrTwit corpus, I will argue the methodological framework in Section 2. In Section 3.1, I will discuss the production rate and the frequency of migr-tweets of French right-wing politics. In Sections 3.2 and 3.3, I will briefly discuss how the topic of immigration is framed through collocational and topic analyses of tweets.

2. Methods

Within the theoretical framework of Critical Discourse Analysis (Reisigl and Wodak, 2015; Van Dijk, 1991; 1980), we proceed to the investigation of the collocations of the migr-lexicon, and the identification of topics, i.e., hashtags of migr-tweets. Hart argued that anti-immigration discourse is a manipulative discourse in the sense that its “argument acts [namely topoi⁴] may automatically yield decisions in favor of discrimination” (2013: 204). We formulate the hypothesis according to which the semantic connotation of migr-lexicon results from the topoi established “in perceived truth-status as a consequence of the frequency with which it is repeated” (Hart, 2013: 204). In other words, the pejoration has been aggravated by the spread of biased social representation of migrants, asylum seekers, and immigrants through right-wing politicians’ tweets containing words derived from the Latin root *-migr-* of *migrare* (henceforth migr-tweets). Since political tweets accelerate based on the political conjuncture, statistical analysis is needed to measure their visibility and frequency depending on time. Analyzing predications reveals the “discursive qualification of social actors, objects, phenomena, events/processes and actions (more or less positively or negatively)” (Reisigl & Wodak, 2014:95).

¹ Formerly the *Front National* till 2018 was founded by Jean-Marie Le Pen in 1972.

² Detailed information, such as the list of the selected Twitter accounts, is provided with the published versions of the corpus. You can use the links in the header of Table 1 to reach the site.

³ As part of my doctoral research in preparation *Le discours de haine sur le web. Analyse linguistique et sociolinguistique [Hate*

speech on the web. Linguistic and sociolinguistic analyses].

⁴ Following Hart’s work (2013:201), the notion of topoi is defined as “formal and content-related warrants which connect premises with conclusions ([Wodak,] 2001:75)” whose a key feature “is that they are ‘common-sense’ reasoning schemes typical for specific issues (Van Dijk, 2000b: 98)”.

Linguistic constructions such as collocations, which attribute, among others, to this qualification, were automatically annotated *ad hoc* by means of the corpus analysis platform Sketch Engine with the function of Word Sketch Difference. The contrastive approach may shed light on the fallacious side of topoi of the immigration discourse, in particular, “persuasive definition (implicit modification of the meaning of words)” and “quasi-definition (unshared or not commonly accepted inferences from the use of a word taken for granted)” (Macagno, 2022: 72).

2.1 Data collection

Concerning the FrRMigr-Twit corpus (henceforth FrR corpus), Pietrandrea and Battaglia (2022) created its first version to analyze migr-tweets⁵ ($n=5689$) of selected 10 French right-wing and far-right politicians. Furthermore, I have opted for the contrastive approach to point out formal and linguistic characteristics of political (far) right-wing politicians’ migr-tweets. 39 French political figures and parties were selected, based on four non-mutually exclusive criteria: high number of migr-tweets, political affiliation, political careers, that is, members of the European Parliament, and Presidential candidate between 2011 and 2022. To extend the FrR corpus using the Twitter API⁶, Battaglia and I carried out data collection at the beginning of the fourth quarter of 2021. The collection of tweets ($n=11761$) and metadata from 16 (far) right-wing political figures and parties ended in July 2022. I extracted migr-tweets ($n=5636$) of 23 (far) left-wing political figures and parties during March and April 2023 to create the FrLMigr-Twit corpus (henceforth FrL corpus). Table 1 illustrates essential information about these two corpora constituting the French MigrTwit corpus.

	FrLMigr-Twit	FrRMigr-Twit
Migr-tweets	5636	11761
Average number of migr-tweets per account	296.50	853.27
Twitter accounts	23	16
words	169818	358491

Table 1: FrMigr-Twit corpora

3. Results

3.1 Frequency analysis

Concerning the raw frequency of results, right-wing

⁵ 5689 tweets were retrieved through Europresse.com before the Twitter API v2 Academic Research was launched in 2021.

⁶ The Academic Research service has been deprecated since May 2023.

⁷ The average number of migr-tweets is a sum of the average numbers of every single year. Since not every account did produce migr-tweets for all of the dozen years, the average number of

politicians produced more migr-tweets than left-wing politicians (Table 1). Overall, approximately 558 more migr-tweets per account have been posted by right-wing politics over the last dozen years, i.e., 853.27 versus 296.5 migr-tweets per account⁷. Moreover, we observed the considerable increases in Marine Le Pen’s and the other right-wing politicians’ migr-tweet productions in 2015 and 2018 (Figure 1). The peak year of 2015 for Marine Le Pen’s migr-tweet production is also a year of its highest growth (1123% increase). Incidentally, 2015 corresponds to the year of the highest growth for both FrR and FrL corpora, but their peak year is 2018.

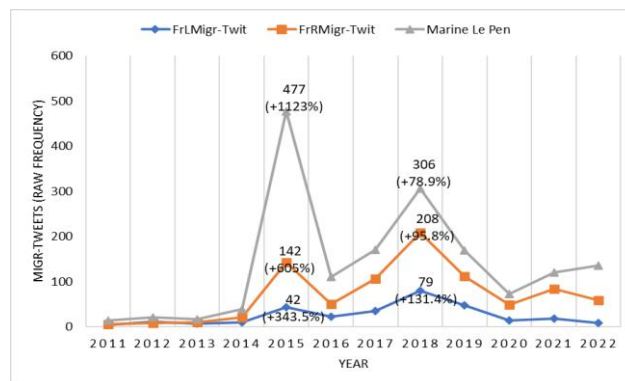


Figure 1: Average number of migr-tweets per account

We also monitored the monthly production of all tweets ($n=212977$) of selected accounts ($n=32$) to calculate the annual proportion of migr-tweets⁸ (Table 2). The visibility of migr-tweets is calculated by dividing the total amount of retweets, likes, replies, and quotes by the number of migr-tweets (Table 3). Despite the relative low frequency of left-wing politics, their migr-tweets were more visible than those of right-wing politics in 2011.

	11	12	13	14	15	16	17	18	19	20	21	22
FrL	0.3	0.4	0.2	0.6	2.5	1.5	2.2	6.5	3.6	1.3	1.9	0.9
FrR	0.7	0.6	0.5	1.5	5.1	2.8	3.1	7.4	5.3	2.8	3.9	3.8

Table 2: Proportion of migr-tweets

The topic of immigration seems to have become viral on Twitter for both groups since 2012, albeit 2013 is the only year where the virality drops for both groups. However, since 2013, right-wing politicians’ migr-tweets have become more viral, and their virality has kept increasing until 2020.

migr-tweets is greater than the quotient of the total number of migr-tweets and a total number of Twitter accounts.

⁸ The number of accounts ($n=16$) is paired to calculate the proportion of migr-tweets of each group. Based on their productivity of migr-tweets, 16 left-wing political accounts were selected.

	FrL	FrR
11	7.0 (+3.0)	4.0
12	60.4	69.8 (+9.4)
13	8.7	39.1 (+30.4)
14	28.0	87.8 (+59.8)
15	25.5	134.8 (+109.3)
16	54.2	297.7 (+243.5)
17	253.8	325.5 (+71.7)
18	235.0	341.2 (+106.2)
19	195.3	527.2 (+331.1)
20	276.7	958.4 (+681.7)
21	295.2	908.3 (+613.1)
22	424.8	1001.3 (+576.5)

Table 3: Virality of migr-tweets

The first year of the highest increase (i.e., 2015) will be further investigated in Section 3.3 with a closer investigation of hashtags of June 2015 because the monthly frequency of migr-tweets shows two different peak months for each group, i.e., June for the FrL corpus and September for the FrR corpus (Figure 2).

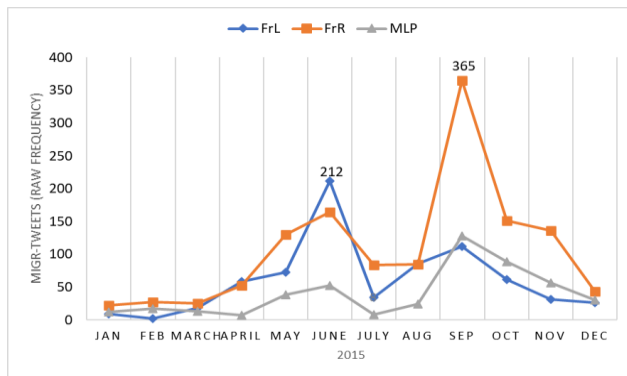


Figure 2: Number of migr-tweets of 2015 by month

3.2 Collocational analysis

Until 2014, a low frequency of MIGRANT⁹ was observed through the FrR corpus, even with no occurrence in 2012. Following the 2015 refugee movement, migr-tweets containing MIGRANT increased through both subcorpora. Within the FrR corpus, the topos of abuse¹⁰ is depicted from 2015 by modifiers, such as *economic*, *clandestine*, and *illegal*, as in (1). The explicitly threat-connoting modifier, i.e., *terrorist* occurred once in 2016 in the tweet (2) referring to the Daesh terrorist attack posted by Marion

Maréchal, retweeted 921 times.

- (1) *We must block the flow of #migrants essentially composed, contrary to what is said, of economic migrants. #RadioClassique ;2015-06-17:@ECiotti;57;17;24;0*¹¹
- (2) *Remind yourself, we were told that it was false. Another terrorist migrant arrested today. How many others?; 2016-08-05:@MarionMarechal; 921;132;789;38*

Since 2019, ethnic modifiers used to depict crimes reported as committed by “a migrant” coming from Muslim ethnic groups have become typical¹² collocates, e.g., *Eritrean*, *Sudanese*, *Afghanistan*, *Algerian*, etc. Conversely, ethnic modifiers do not constitute the typical collocates of MIGRANT at the side of the FrL corpus. The inherent features of MIGRANT are rather observed with modifiers like *vulnerable*, *precarious*, and *climatic*. The humanitarian topos is triggered through the left-wing politicians’ migr-tweets as in (3), which are targeted by the far-right populist politicians. To give an example, when it comes to *minor migrants*, the humanitarian act of protecting children has been steadily pulled away as in (4) and (5).

- (3) Posted in February 2018 by Manon Aubry, member of the far-left populist party La France Insoumise:
RT@afpfr: Justice rules against the prefect and suspends the return of minor migrants to Italy #AFP; 2018-02-24:@ManonAubryFr;123;0;0;0
- (4) Posted in May 2019 by Jordan Bardella, member of the RN:
We have been witnessing, for the past few months, in the Bourgogne Franche Comté region, the TRIPLING of "isolated minor migrants", and these "minor #migrant" are as minor as I am I'm an archbishop! #Yonne #Le26MaiVotezRn ; 2019-05-21:@J_Bardella; 114;28;186;5
- (5) Posted in June 2022 by Marine Le Pen, former leader of RN:
At the #StadeDeFrance, it was the surge of hordes of ultra-violent migrant minors from the Porte de la Chapelle, these were raids perpetrated by city gangs acting with impunity. The government concealed the seriousness of the facts! #Legislative2022; 2022-06-05:@MLP_officiel;183;20;425;4

3.3 Topic analysis of hashtags

2015 was the transition year in terms of virality, productivity, and frequency of migr-tweets. The distribution of hashtags in migr-tweets posted during June and September showed that the highest increase in migr-tweet production in both subcorpora is not interrelated in 2015. Figure 3 illustrates the distribution of hashtags in June 2015. Regarding the FrL corpus, 66 out of 207 hashtags refer to the live issue of the violent evacuation of

⁹ By writing the term in capital letters, I refer to its lemmatic status.

¹⁰ “[T]he topos of abuse can be expressed as: “if a right or an offer for help is abused, the right should be changed, or the help should be withdrawn, or measures against the abuse should be taken” (Wodak, 2001: 77)” (Hart, 2013).

¹¹ The tweet example format is structured as follows: translation in English; posting date (year-month-day); user name; retweet; reply; like; quote.

¹² The typicality is measured utilizing the logDice score calculated by the analyzing tool Sketch Engine.

migrants from the Halle Pajol in the 18th district of Paris on 8 June 2015. In addition, 14 hashtags refer to the topic of ASYLUM, i.e., #réfugié (*refugees*), #asile (*asylum*), #Waterloomoral. However, none of the right-wing politicians' migr-tweets contain hashtags related to the topic of ASYLUM, even though 51 hashtags refer to the keyword of MIGRANT. Furthermore, the generic noun IMMIGRATION is more frequent in right-wing politicians' migr-tweets than in left-wing politicians' migr-tweets.

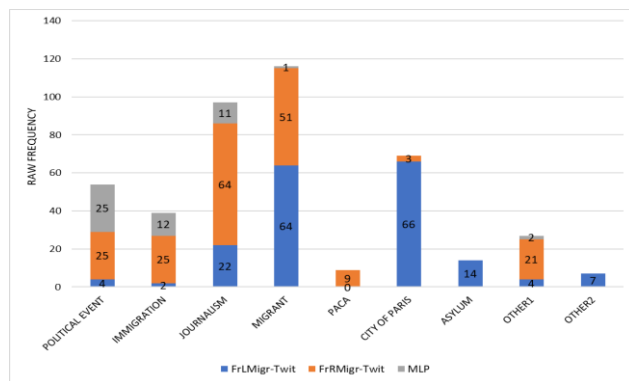


Figure 3: Topics of hashtags in migr-tweets of June 2015

The live issue of the evacuation of migrants is absent in right-wing politics' migr-tweets. Their hashtags refer to communication contexts such as the POLITICAL EVENT (e.g., #ConfMLP) and JOURNALISM (e.g., #BFMPolitique). 25 migr-tweets containing #ConfMLP were posted by Marine Le Pen on June 10, 2015. Hashtagging enables several “related” tweets to be linked together as in (6) and (7).

- (6) “Last February, a small town in #Burgundy was forced to take 60 migrants in order to unclog #Calais.” #ConfMLP; [2015-06-10:@MLP_officiel;59;7;26;0](https://twitter.com/MLP_officiel/status/597260)
- (7) “With his peopling policy, @manuelvalls wants to disseminate #immigration in our countryside in order to make it less visible” #ConfMLP; [2015-06-10:@MLP_officiel;103;9;34;0](https://twitter.com/MLP_officiel/status/1039340)

4. Discussion

The analysis of the monthly distribution of hashtags indicates what topics are selected and how they are organized according to specific events or current topics because “[t]opics not only suggest what information is most important in the text, but also what is most important ‘in the world’” (Van Dijk, 1991: 74). A closer investigation of hashtags of June 2015 showed that specific events put forward differ between right-wing and left-wing politicians. I suggest that even though MIGRANT is the most likely to be “what is the center or focus of the information” (Van Dijk, 1980: 98) as being highlighted by means of hashtagging and reiteration, the topic of MIGRANT is exploited by the far-right populist politicians to undermine

their political opponents.

5. Conclusion

Comparatively exploiting the FrMigr-Twit corpora, the statistical analysis showed that right-wing political migr-tweets have become more productive and viral since 2015. By analyzing the collocations of MIGRANT, we demonstrated that topoi of anti-immigration discourse (i.e., topoi of abuse and danger) are gradually conveyed through right-wing political migr-tweets defeating the humanitarian topoi. Twitter structure (i.e., hashtag, brevity of messages, decontextualization) and the frequency contribute to the semantic connotation of the migr-lexicon.

6. References

- Ait Abdeslam, A. (2021). Muslims and Immigrants in the Populist Discourse of the French Party Rassemblement National and Its Leader on Twitter, *Journal of Muslim Minority Affairs*, 41:1, 46-61.
- Allcott, H., Gentzkow, M. (2017). Social media and fake news in the 2016 election, *Journal of Economic Perspectives*, 31:2, 211-236.
- Badouard, R. (2017). *Le désenchantement de l'internet : Désinformation, rumeur et propagande*, FYP éditions.
- Battaglia, E., Blandino, G., Jeon, S., Pietrandrea, P. (2022). MIGR-TWIT CORPUS. Migration Tweets of right and far-right politics in Europe [Data set]. Zenodo, <https://doi.org/10.5281/zenodo.7347479>
- Blandino, G. (2023). 10 years of public debate on immigration: combining topic modeling and corpus linguistics to examine the British (far-)right discourse on Twitter, MA thesis. University of Wolverhampton.
- Hart, C. (2013). Argumentation meets adapted cognition: Manipulation in media discourse on immigration, *Journal of Pragmatics*, 59, 200-209.
- Macagno, F. (2022). Argumentation profiles and the manipulation of common ground. The arguments of populist leaders on Twitter, *Journal of Pragmatics*, 191, 67-82.
- Pietrandrea, P., Battaglia, E. (2022). “Migrants and the EU”. The diachronic construction of ad hoc categories in French far-right discourse, *Journal of Pragmatics*, 192, 139-157.
- Pietrandrea, P., Jeon, S. (2023). MIGR-TWIT CORPORA. Migration Tweets of French Left-wing Politics. [Data set]. Zenodo, <https://doi.org/10.5281/zenodo.7871602>
- Reisigl, M., Wodak, R. (2015). The Discourse-Historical Approach (DHA), In: Wodak, R., Meyer, M. (Eds.), *Methods of Critical Discourse Studies*, 87-121, SAGE.
- Scott, K. (2015). The pragmatics of hashtags: Inference and conversational style on Twitter, *Journal of Pragmatics*, 81, 8-20.
- Van Dijk, T-A., (1991). *Racism and the Press*, Routledge.
- Van Dijk, T-A., (1980). *Macrostructures: An Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition*. Lawrence Erlbaum Associates, Inc., Publishers.
- Wodak, R. (2021). *The Politics of Fear. The Shameless Normalization of Far-Right Discourse*, 2nd ed., SAGE.