

EU-Border crisis on Twitter: sentiments and misinformation analysis

Elena Ambrosetti^a, Cecilia Fortunato^a, and Sara Miccoli^b

^a Sapienza University of Rome; elena.ambrosetti@uniroma1.it, cecilia.fortunato@uniroma1.it

^b Istat; sara.miccoli@istat.it

Abstract

The objective of this paper is to investigate the information and to detect the presence of misinformation on Twitter posts circulating in relation to migration events happened in 2020 at the Greek-Turkish border and in 2021 at the Polish-Belarusian border. Data were retrieved through API by using keywords referring to the two border events. The study was carried out by applying text mining and sentiment analysis techniques on tweets and retweets related to these two events, and by conducting a qualitative analysis on specific subsets of tweets. Our results show that in both borders' crises migration is perceived as an emergency issue, migration-related narratives mainly refer to "war", "attacks", "tension", "invasion" and the emotions expressed are mostly negative. In addition, in outbreaking crisis, the identification of misinformation in social media is extremely challenging, because of the rapid circulation of rumours related to facts that are rather difficult to ascertain.

Key words: misinformation, twitter, migration, Europe, border

1. Migration events at the Greece-Turkey and Poland-Belarus borders

This study focuses on the analysis of information spreading on social media during specific migration-related events in Europe and it is part of the work conducted within the H2020 Perceptions project¹. The study focuses on a relevant topic related to migration phenomenon, i.e., the spread of misinformation and rumours that can have a great impact on the lives of migrants and influence the ideas circulating among migrants themselves and the inhabitants of receiving countries.

We focus on two borders' crises. The first one is the border between Greece and Turkey. This border has been the scene of several humanitarian crises involving migrants in recent years. One of the last events originated in late February 2020, when the Turkish president declared the opening of the border for migrants directed to Europe. During spring 2020 videos, pictures, and news on migrants at the Greek-Turkish border started to circulate. The two sides involved, Greece and Turkey, began to make accusations against each other regarding the treatment of migrants and the circulation of misinformation. The Greek and Turkish governments themselves have actively participated in this exchange of accusations, also by using social media. In this context, not only hardly verifiable news began to be spread, but real news was possibly being pointed as fake news or made confusing for political purposes.

Beyond specific ad hoc created fake news then, there was an underlying misinformation and confusion about real events. Misinformation also arose regarding the size of the phenomenon, with the Turkish side tending to overestimate the number of migrants who left Turkish territory. In addition to distorted news, real fake news was artfully created.

The second border's crisis that we analyse is the one on Poland-Border border in 2021. It was triggered by the tensions' escalation in Belarus-European Union (EU) relations started in August 2020, following the Belarusian presidential election, the repression of mass-protests in the country, and other events such as the hijacking of the Ryanair Flight 4978 for the arrest of Belarusian journalist Roman Protasevich. As a response, United States Treasury Department, and the EU imposed sanctions to President Aleksander Lukashenko and to exponents of Belarusian administration and economy. In May 2021, President Aleksander Lukashenko warned the EU of the possibility for Belarus to stop patrolling

¹ The project PERCEPTIONS is funded by the European Commission H2020 Research & Innovation Action under Grant Agreement No 833870. It aims to identify and understand narratives, images, and perceptions of the EU outside Europe, and to study ways, channels and actors through which narratives are distributed.

the EU border, threatening EU to provoke a migrant's crises in response to sanctions. From that moment, Frontex reports thousands of border crossing attempts from Belarus into EU, with the growing apprehension of Polish and European authorities and public opinion as well as the rising tension in the whole region, involving Russia, Latvia, Lithuania and Ukraine. The president of Belarus has been publicly accused by European authorities of offering tourists' visas and flights to migrants from Iraq and Syria. While thousands of migrants were stuck on the Belarus side of the border, tensions between Belarusian government and EU increased also in contents posted on Twitter. Both sides were accusing the other to use migrants as a destabilizing weapon, neglecting international protection's procedures and keeping migrants in life danger.

Both the contexts under study involved two EU border countries (Poland and Greece) facing migrant's events perceived as crises, under the guidance on the EU in a controversy with third countries (Belarus and Turkey). Both crises have echoed widely in the mainstream media, prompted discussions and news circulations (accurate or distorted) in social media. The spread of misinformation on social media has undoubtedly exacerbated the tension in crisis phases and could have also misled migrants and potential migrants' idea about the migration issue in Europe.

2. Research design, data and method

Focusing on misinformation this concept has numerous, and sometimes contradictory, definitions (Treen et al. 2020, p. 3). According to the Oxford English Dictionary 2018, misinformation is defined as "wrong or misleading information". In many cases, misinformation is used as a synonym of false facts, rumours (e.g., Donato et al, 2022). Slightly different is the concept of disinformation, defined in literature as a "deliberately false information". This differentiation has been embraced by many authors (e.g. Stahl 2006, Alonso et al. 2021, Treen et al. 2020), while others, such as Ruokolainen & Widén (2020), do not rigidly classify the terms misinformation and disinformation. In the framework of this analysis, we decided to adopt the umbrella concept of *misinformation*, as ambivalent, distorted or falsified information (Zhou and Zhang 2007), without distinction as to intention or source of origin of the information. Rumours and fake news are thus equally included in this concept.

The objective of the work is to investigate the information and misinformation circulating in Twitter during specific geopolitical crisis related to migration flows at the abovementioned border areas, by examining tweets and retweets retrieved using targeted keywords.

In particular, the first research question of the study is: (RQ1) what words are most frequently expressed in Twitter environment during migration-related events at border areas and what sentiments do they communicate about the discussion taking place in social media? Our aim is to examine the nature of discourse developed on Twitter around these issues. We hypothesize that the words used, and the sentiment expressed, may indicate the presence of certain topics related to specific actors of these migration events.

The second research question is: (RQ2) is it possible to identify misinformation in Twitter contents about migration-related events at the borders? We hypothesize that tweets posted during the uprising of these kind of crises are a huge, heterogeneous and fragmented corpus of text and they can be hardly analysed as whole. Therefore, a qualitative approach on subsets of tweets containing selected keywords could be more informative for the analysis of patterns in narratives and misinformation about specific topics.

The data used for the analysis were downloaded using the Academic Research Twitter API, by searching for the hashtags #GreeceTurkeyBorder and #PolandBorder and #BelarusBorder. Our analysis is restricted to the content in English language, since English messages are more likely to cross national borders and spread all over the world.

For these three complete datasets, we applied text mining and sentiment analysis techniques to analyse the words used and to understand what kind of sentiment they express.

Using the *TM* package in R, we have built a term-document matrix that describes the frequency of single terms that occur in the corpus of documents. To perform the analysis, we cleaned the text data by applying a series of filters and we only selected semantically significant terms. We analysed the association between words through a correlation analysis, a statistical technique that can demonstrate whether, and how strongly, pairs of words could be related, helping to map the discourses around these words.

We investigated the most recurrent words related to the migrant situation at the borders to evaluate the prevalence of specific topics in tweets' contents.

We then conducted a sentiment analysis to understand what sentiments and emotions the tweets expressed, by using the *syuzhet* R package. We used the NRC Emotion Lexicon, developed by Saif and Turney (2010), that assigns words to 8 different types of emotions (Naldi 2019; Widyaningrum et al. 2019). The 8 emotions are anger, fear, sadness, disgust (negative ones) and anticipation, trust, surprise, and joy (positive ones). For each tweet the method counts the number of words associated with each category (Naldi 2019, Widyaningrum et al. 2019), providing a cumulative score representing the sentiments for the whole corpus of text.

We then created various subsets of specific tweets (excluding retweets). First, we selected tweets containing the words “misinformation”, “disinformation”, and “fake”. On these we conducted a manual qualitative analysis, trying to understand which news items were categorized as misinformation by the users in social media. Then, we created targeted subsets of tweets, by searching for selected keywords related to specific topics that generated a major flow of information during the two events. We manually analyzed the contents to understand what kind of information was conveyed and in what ways, in order to identify biased or inaccurate information.

3. Section Heading

The analysis produced some interesting results, useful in understanding what kind of discourse emerges during critical situations concerning the migration phenomenon, and what role misinformation can play. Although these results derive from contents posted on Twitter only in English and related to specific geographic contexts, they can be representative of general discourses spread via social media related to migration phenomena.

We found common patterns among the contents concerning the situations on the Greek-Turkish and Polish-Belarusian borders. In both cases, the words circulating in social media about these migration-related events refer to wars, attacks, tension, crises, invasion. In addition to terminology related to the theme of conflict, the use of words connected with the most vulnerable groups, such as children and women, is also frequent. In both cases, the perception of migration as an emergency issue emerges. The feelings and emotions expressed by the words contained in these tweets are mostly negative. The messages and information existing in social media regarding these issues are therefore dominated by negative elements, mainly related to anger and fear (RQ1).

The qualitative analysis carried out on more specific subsets of data made it possible to integrate and deepen the results emerged from the text mining analysis and to understand whether misinformation could be identified in tweets concerning specific topics. Regarding the different topics related to critical situations at the borders, confused discourses emerged, with polarized messages and viewpoints, in which it is very difficult to distinguish between misinformation and accurate information (RQ2). This is true both in the context of events on the Greek-Turkish and on the Polish-Belarusian borders.

Opposing views emerged regarding migrants. On the one hand, they are depicted as victims of a cynical policy in a game for power assets, on the other hand, they are often identified as an invasion force and a source of violent crime. As emerged by the analysis of words' associations in the entire datasets, and confirmed by the qualitative analysis of specific subsets, refugees and asylum seekers, as being subject to a more careful regulation and therefore more entitled to enter Europe, presented a slightly different representation, more empathetic and supportive in comparison to generic migrants. Around migrants' arrival to Europe, their crossing the borders, even on social media (as happens in the public debate), a war on numbers has been fought. Also on this issue, social media acts as an echo chamber of the confusion existing in society regarding the real numbers of migratory flows.

In the analysed tweets, Europe as a political entity is often defined as a weak, silent, role-playing spectator in crisis involving migrants. In this context, both the idea of Europe as a political entity completely helpless and disinterested in migratory events for political ends, and the idea of Europe as an entity that is victim of the events and under attacks, emerges. Discourses that can be based on real facts are therefore mixed with preconceived ideas that do not sink into real information.

In some tweets concerning situations on both borders, Europe is also considered weak in countering the spread of misinformation.

Names of politicians frequently appear in circulating social media discourse. Social media seem to function as echo chambers for politicians' declarations, and this implies a fundamental role that they should play in providing correct and non-distorted information. The reconstruction of what happened on the Greek-Turkish and Polish-Belarusian borders showed that some public actors have been found to push misinformation, by accident or not, spreading confusion, fear, anger, or prejudice.

In social media, we found discourses and accusations very similar to those that emerged and were disseminated in the mainstream media during the days of the two crises. On some specific episodes mentioned in tweets, we found different interpretations regarding the protagonists and the parties involved, demonstrating the various narratives that can circulate and the confusion that can arise about what is true and what is fake. How difficult is identifying misinformation from real information was evidenced not only by the presence of various points of view, but also by the lack of reliable reference points. Indeed, the information circulating in social media in relation to the migration-related events at the borders are often "black box information", without mentioning any verifiable source, often found to be supported by misleading or manipulative images and videos, providing inaccurate messages or marked by simplistic or sensationalist slogans.

Our findings underlined the difficulty in identifying misinformation from real news in social media. The coexistence of so many points of view, often unsupported by reliable data and sources, is an indication of a tendency to provide inaccurate information in social media. Indirectly, this leads to the idea that misinformation, now understood in a very broad sense, can influence the lives of migrants and potential migrants and the perceptions of migrants travelling to Europe, but also of Europeans themselves.

Certainly, the results that emerged from these analyses cannot be considered as definitive. The issue of misinformation can be investigated with further methods and by drawing on further data. Still, these findings are an important indication of how information is developing and circulating on social media, shedding further light on the theme of narratives and discourses, their construction and dissemination, and their potential role on people lives.

References

- [1] Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., & Vilares, J.: Sentiment analysis for fake news detection. In: *Electronics*, 10(11), 1348 (2021)
- [2] Donato, K. M., Singh, L., Arab, A., Jacobs, E., & Post, D.: Misinformation about COVID-19 and Venezuelan migration: Trends in Twitter conversation during a pandemic. In: *Harvard Data Science Review*, 4(1) (2022)
- [3] Naldi, M.: A review of sentiment computation methods with R packages. In: *arXiv preprint arXiv:1901.08319* (2019)
- [4] Ruokolainen, H., Widén, G.: Conceptualising misinformation in the context of asylum seekers. In: *Information Processing & Management*, 57(3), 102127 (2020)
- [5] Saif M., Turney P.: Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In: *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, LA, California. (2010)
- [6] Stahl, B. C.: On the difference or equality of information, misinformation, and disinformation: A critical research perspective. In: *Informing Science*, 9, 83 (2006)
- [7] Treen, K. M. D. I., Williams, H. T., & O'Neill, S. J.: Online misinformation about climate change. In: *Wiley Interdisciplinary Reviews: Climate Change*, 11(5), e665 (2020)
- [8] Zhou, L., & Zhang, D.: An ontology-supported misinformation model: Toward a digital misinformation library. In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(5), 804-813 (2007)