Exploring the Predominant Targets of Xenophobia-motivated Behavior: A Longitudinal Study for Greece

Maria Pontiki, Konstantina Papanikolaou, Haris Papageorgiou
Institute for Language and Speech Processing, Athena Research and Innovation Center
Artemidos 6 & Epidavrou, 15125, Athens, Greece
{mpontiki, konspap, xaris}@ilsp.gr

Abstract

We present a data-driven linguistic approach for exploring the predominant targets of xenophobia-motivated behavior in Greece over time focusing on specific Target Groups of interest. We employ two principal data analytics workflows to produce the corresponding data insights; Event Analysis using news data from 7 different sources for the last two decades (1995-2016) capturing physical attacks along with the involved social actors, and Verbal Aggressiveness using Twitter data, locating xenophobic stances as expressed by Greeks in social media for the time period 2013-2016. The results indicate that examining physical and verbal aggression as indicators of xenophobic attitudes and combining News and Twitter data can provide important insights allowing to measure, monitor and comprehend xenophobia as a violent practice in Greece over time. Hence, our work constitutes a source of valuable information for journalists, political and social scientists, policy makers and all stakeholders interested in this complex social phenomenon, and can also serve as a storytelling and narrative framework.

Keywords: Xenophobia, Event Analysis, Verbal Aggressiveness

1. Introduction

The recent refugee/immigrant crisis in Europe gave burst to xenophobic sentiments, attitudes and practices ranging from individual (re)actions to official state policies. Xenophobia is associated with feelings of dislike implying superiority, or feelings of fear/vulnerability implying the perception of threat (Van der Veer et al., 2013), and is often examined as a violent practice (Harris, 2002). Focusing on the violence aspect, we present a data-driven linguistic approach for exploring the predominant targets of xenophobia-motivated behavior in Greece over time.

In collaboration with political scientists, we defined 10 Target Groups (TGs) of interest (TG1: Pakistani, TG2: Albanians, TG3: Romanians, TG4: Syrians, TG5: Muslims, TG6: Jews, TG7: Germans, TG8: Roma, TG9: Immigrants in general, TG0: Refugees in general). The selection was based on specific criteria such as the population of specific ethnic groups in Greece (e.g. Albanians and Pakistani are the two most populated national groups living in Greece) or the existence of established prejudices and stereotypes in Greece about specific groups (e.g. Jews were selected in order to examine anti-Semitism within the Greek society given that, according to the "ADL Global 100" survey, which elaborated an index of anti-Semitism based on the strength of anti-Semitic stereotypes, Greece was the most anti-Semitic country in Europe scoring 69%).

Then, we employed two principal data analytics workflows; [**A**] **Event Analysis** (Section 4) using news data aiming to capture physical attacks against the predefined TGS of interest. In order to explore whether violent incidents are mostly directed to foreigners or not, Greeks were used as a control group. [**B**] **Sentiment Analysis** (Section 5) in Twitter data aiming to detect Verbal Aggressiveness (VA) targeting the predefined groups of foreigners. An overview of the overall methodology and the data sources is provided in Sections 2 and 3, respectively. The results, presented in Section 6, indicate that a large-scale study combining news data –generated by journalists to report physical attacks against foreigners– and Twitter data –generated by Social Media users in order to verbally attack to foreigners– can provide valuable insights allowing to analyze the complex phenomenon of xenophobia, and to confirm or debunk and disprove the existence of certain stereotypes and prejudices. Moreover, conclusions can be drawn regarding the implications and the various dimensions that are attributed to xenophobia. Hence, our work could serve as a storytelling and narrative framework for journalists interested in this complex social phenomenon. In addition, given the high correlation between verbal and physical aggression (Berkowitz, 1993), the proposed method can provide valuable insights also to political scientists and policy makers.

2. Methodology Overview

The overall methodology consists of 5 steps: [**A**] **Knowledge Representation.** Design of a coding framework covering a wide spectrum of physical and verbal attacks along with their complementary elements. [**B**] **Data Collection**. An important dataset of News and Twitter data was collected, prepared and curated (Section 3). [**C**] **Data Exploration**. Valuable insights were extracted helping to finalize the coding framework, and to create focused data collections. [**D**] **Content Analysis**. The data was modelled according to the information types that our research focuses on (Event Analysis and VA detection). [**E**] **Data Visualization**. The content analysis results, having been revised, were visualized in different ways making them explorable, comprehensible and interpretable.

3. Data Collection

3.1 News Data

A total of **3.638.137** news articles for a time span of more than 20 years, specifically 1995-2016, was collected from 7 news national-wide agencies in Greece (Table 1). All articles are in Greek and metadata (section labels, headlines, names of authors) were gathered for each along with the text itself. Data preparation included tackling normalization problems and transforming the data to a human readable corpus. During the Data Exploration phase, event-oriented data clusters were created (one collection for each event type) to filter the collected bulk of data. To this end, complex queries were constructed comprising words and phrases in which each event type was expressed and lexicalized, also leveraging Boolean operators.

Source	Articles	Time span
Avgi	792.715	1996-2015
Kathimerini	282.621	2002-2006, 2009-2012
Eleftherotypia	429.364	2002-2006, 2008-2014
Rizospastis	725.108	1995-07/2016
TaNea	330.190	1997-2007
In.gr	428.880	1999-21/09/2016
Naftemporiki	649.259	2000-21/09/2016

Table 1: News data

3.2 Twitter Data

For each TG of interest we retrieved from Twitter relevant Tweets using related queries/keywords (e.g. "ισλάμ" (="islam"), "Πακιστανός" (="Pakistani"), etc.). Given that the search function in the database configuration is stemmed, the queries returned also tweets containing compound words and morphological variations of the selected keywords (e.g. "*ισλαμοποίηση*" for "*ισλάμ*"). The search resulted in 10 collections (1 per TG) containing in total **4.490.572** Tweets covering the time period 2013-2016 (Fig. 1). The peak in the mentions of refugees during the last two years coincides with the refugee crisis, whilst Germans are continuously in the limelight since, along with the IMF and the EU, they have a central role in the Greek crisis. The next most discussed TGs are immigrants and Syrians, who are also related to the refugee crisis. Muslims and Islam follow in the 5th place, with a peak from 2014 onward which coincides with the rise of ISIS.

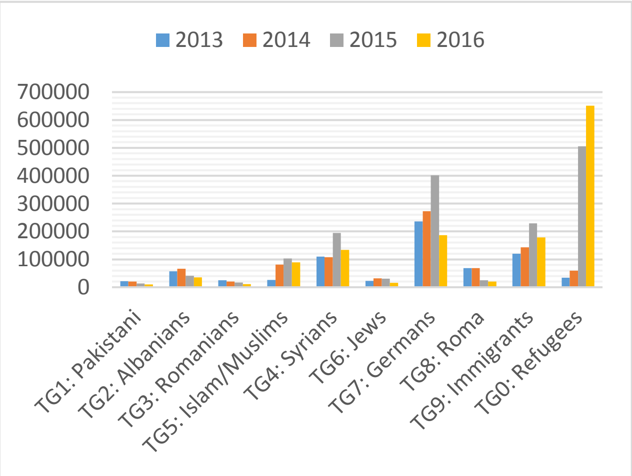

Figure 1: Per-year number of Tweets collected for each TG

4. Event Analysis

4.1 Codebook

A coding schema (q.v. Papanikolaou et al., 2016) covering a wide spectrum of event types related to xenophobia along with their structural components was set up. In this context, the coding unit of the analysis is the event. The proposed event taxonomy includes a major event category, namely Physical Attacks, encompassing various event types like *Violent Attack* and *Sexual Assault*. In the schema, an event comprises a tuple containing five types of information, each of which is also attributed several features as illustrated below:

1. **EVENT**. The word or phrase representing an event type under examination, which is located within the text. Features: *Event type*.
2. **ACTOR**. The entity that performs each event instance. Features: *Summary*, *TG*, *Nationality*, *Age*, *Sex*, *Status*.
3. **TARGET**. The entity to whom the action is addressed. Features: *Summary*, *TG*, *Nationality*, *Age*, *Sex*, *Status*.
4. The **LOCATION** where the event took place. Features: *Category*.
5. The **TIME** at which the event happened. Features: *Day*, *Month*, *Year*.
6. The **CONFIDENCE** element which captures whether in the article there is any indication that an Actor of an assault may not be the actual perpetrator. Features: *Degree*.

4.2 Content Analysis

The overall event extraction framework is data-driven. The adopted approach is to first detect each structural element of an event instance and afterwards to bind the right elements and create the event tuples. The methodology employed is semi-supervised, in the sense that a small fraction of data was labeled and used for the development of the system. Moreover, it is linguistically driven, thus morphosyntactic information from basic NLP tools is used to approach the information types comprising the event tuple as it is defined in the Codebook. The general approach for extracting events is incremental, as every module builds over the annotations produced by previous modules (Stathopoulou et al., 2017). At the first stage, the ILSP NLP tools suite (Papageorgiou et al., 2002; Prokopidis et al., 2011) is used for performing pre-processing over raw text and producing annotations for tokens, lemmas, chunks, Syntactic relations and named entities. In the next phase, the pre-processing output is given as input to the Event Analysis Unit, which performs two subsequent tasks. First, the elements comprising an Event are detected and then linguistic rules based on shallow syntactic relations bind the right elements in an event tuple. The system is implemented as Finite State Transducers (FSTs) using Gate JAPE patterns (Cunningham et al., 2000). These FSTs process annotation streams utilizing regular expressions to create generalized rules. Moreover, they are ordered in a cascade, so that the output of an FST is given as input to the next transducer. Subsequently, the output of the above described workflow, is a set of tuples, each depicting an Event with its structural elements. An illustrative example of the system output is the following:

<**Actor**: *A 24-year-old American*, **Confidence**: *is accused of involvement*, **Event**: *in the arson*, **Target**: *of the Synagogue*, **Location**: *at Chania*, **Time**: *on Thursday*>.

5. Verbal Aggressiveness (VA)

5.1 VA Framework

Based on literature review and explorative analysis findings we developed a linguistically-driven framework where VA messages (VAMs) are classified based on: **A.**

Their focus (i.e. distinguishing between utterances focusing on the target's attributes, and utterances focusing on the attacker's thoughts). **B.** The type of linguistic weapon used for the attack (e.g. formal evaluations, dirty language, humor). **C.** The content of the attack (e.g. threats/calls for physical violence or for deportation). The detailed typology is illustrated in Fig. 2.

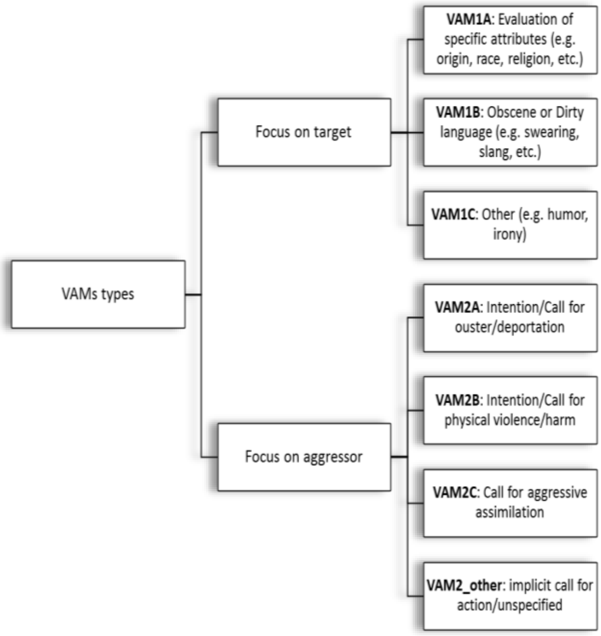

Figure 2 : Typology of VAMs

5.2 VA Analysis

We employed a rule-based method that comprises of a variety of lexical resources and grammars (sets of linguistic patterns). The VA analyzer is an FST cascade implemented as a JAPE grammar in the GATE framework. The input for the analyzer was preprocessed data. In particular, the Twitter collections described in Section 3.2 were tokenized, sentence splitted, part-of-speech tagged and lemmatized using the ILSP suite of NLP tools for the Greek language (Papageorgiou et al., 2002; Prokopidis et al., 2011). Given a preprocessed tweet, the VA analyzer detects candidate VAMs and candidate targets based on the respective lexical resources; if a token is recognized as a lexicon entry then it was annotated with the respective metadata (lexicon labels). In a subsequent step, the grammars determine which candidate VAMs and targets are correct. The grammars implement multi-phase algorithms, where the output of each phase is input for the next one. Each phase comprises several modules that contain a variety of contextual lexico-syntactic patterns. The patterns are templates that generate rules in the context around the candidate VAMs and targets. For each identified VAM, the method returns the type and the linguistic evidence of the attack as well as the id and the linguistic evidence of the object of the attack (TG). For example, for the tweet: "*Muslims should be baptized if they want to find a job in Greece*", the analyzer returns the following tuple:

<**TG_id**: "*TG5*", **TG_evid**: "*Muslims*", **VAM_type**: "*VAM2C*", **VA_evid**: "*baptized*">.

6. Results - Predominant Targets of Xenophobia in Greece

The targets of the xenophobic attitudes are examined in the context of physical attacks reported in News data and in terms of verbal attacks expressed in Twitter. News data allow to measure and monitor physical attacks as they are reported by journalists in various newspapers for a time span of more than 20 years, and to explore possible correlations with events like the financial crisis in Greece. Due to space restrictions, we only present results for Avgi and Naftemporiki. The particular sources were chosen due to their different political orientation and because they cover the longest time span. Twitter analysis captures users' instant and freely expressed sentiments, thoughts, intentions etc. providing a snapshot of the pulse of the Greek society for the time period 2013-2016.

The quantitative analysis of the physical and verbal attacks indicates that xenophobic behaviors do not seem to be dominant in Greece, since the proportion of physical attacks against foreigners (TGs) and those against Greeks (Control Group) showed that the increase of violent incidents targeting foreigners should be examined in the light of a rise of aggressiveness in general, irrespectively of national identity of the victim. Similarly, the VA rates (VAMs/Tweets) detected in Twitter regarding the specific TGs are low (i.e. the VA rate for the mostly attacked TG is approx. 4%). However, focusing on the research goal of this paper, the identity of the victims can provide valuable insights about the xenophobic behavior of Greeks; it can help to comprehend if this type of behavior is superiority or vulnerability-based as well as if it is driven by deeply rooted stereotypes and prejudices in Greek society or by specific events (e.g. financial crisis, refugee crisis).

6.1 Main Targets of Physical Attacks

The TGs against whom most attacks occur as they were recorded in Avgi and Naftemporiki are presented in Fig. 3 and 4, respectively.

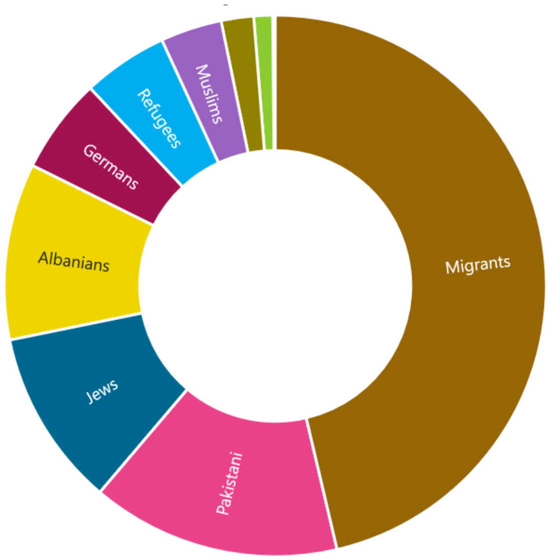

Figure 3 : Mostly Attacked TGs 2000-2015 (Avgi)

Despite some ranking differentiations the five mostly attacked TGs are the same in both sources. The general category Migrants seems to be most frequently referred to by the newspapers, which is consistent to the regulations about media reports that have been implemented over time. Albanians and Pakistani, two of the mostly attacked TGs, are the two most populated national groups living in Greece. Finally, Jews and Germans complete the top five TGs. What needs to be noted is that there is not a significant number of people from these two ethnicities living in the Greece.

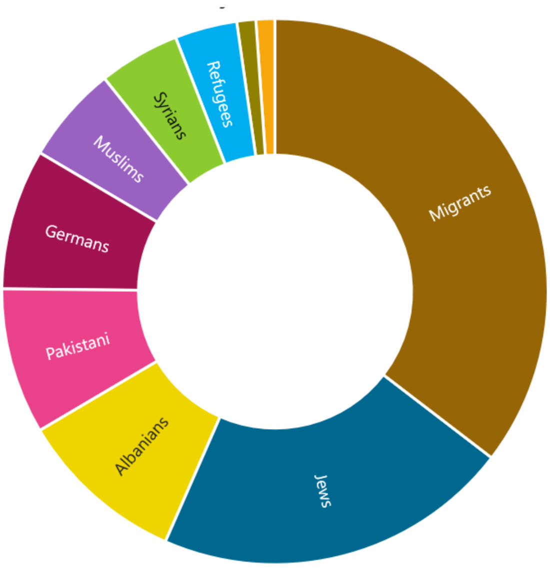

Figure 4 : Mostly Attacked TGs 2000-2016 (Naftemporiki)

We also examined how these TGs evolved over time, having 2009 as a reference point of the financial crisis' beginning as a major event that could affect the physical attacks against them. An increase of violence against foreigners during the financial crisis can be signalled, but should be related to an escalation of violent incidents in general, along with the emergence of far-right extremism in Greece. The results allow for two interesting conclusions. Firstly, there are three TGs that appear to be consistent over time, regardless of the financial crisis, viz Pakistani, Albanians and Jews. Consequently, a continuity is observed against these ethnic groups. On the other hand, new targets seem to emerge depending on the contemporary conditions affecting the country. Thus, Germans appear as TG, with the attacks against them rising as the economic crisis deepens. It is also important that there appears to be no differentiation between the two news agencies. The number of event mentions may differ, though the tendency is quite similar irrespectively of the newspaper's political orientation.

6.2 Main Targets of VA

The results of the VA analysis indicate that the most discussed/mentioned TGs in Twitter are not also the most attacked ones. In fact, refugees are the most discussed but the least attacked TG. The few verbal attacks that were captured are mostly attempts to challenge their identity implying that they are illegal immigrants. This notion of "illegality" or "lawlessness" is also dominant in the case of the generic TG Immigrants, where the most frequent terms used to attack it are the words "λαθρομεταναστες" and "λαθρο" (meaning illegal). According to the results of the analysis (Fig.5) the most attacked TGs are Jews (23%), Albanians (22%), Pakistani (15%), Muslims/Islam (14%), and Immigrants in general (10%). Antisemitism seems to be at the core of xenophobic discourse. Albanians are perhaps the most established group of foreigners in Greek public discourse, given that the image of foreigner as it was constructed in Greece during and after the first wave of migration flow (early 1990s-mid 1990s) was mainly associated with Balkan, and mainly Albanian, nationality.

Figure 5: Mostly Attacked TGs 2013-2016 (Twitter)

Focusing on the type of the verbal aggression, attacks that involve calls for physical extinction are far greater for Jews than for any other group. Moreover, aggressive messages related to this specific TG are revealing the emergence of threat perception based on biological and cultural terms, as well as the perception of a particular enmity towards the Greek nation. Threat perception seems to prevail also for Pakistani, Albanians and Immigrants, according to the share of VAM2 attacks and in particular the calls for ouster/deportation for the specific groups.

6.3 Discussion

An important observation, concluding the analysis of the targets of xenophobic attitudes, is that, in general terms, the VA results coincide with the Event Analysis results. In other words, physical and verbal aggression as indicators of xenophobic attitudes in Greece seem to be addressed to the same targets. Four out of five TGs that are mostly attacked both verbally and physically, are the same (Jews, Albanians, Pakistani and Immigrants in general). Germans don't seem to be one of the most prominent TGs of verbal attacks, while the physical attacks against them are more often. However, the qualitative analysis of the verbal attacks against them revealed the correlation to the economic crisis, in line with the physical attacks that are mainly addressed to politicians and diplomats.

The qualitative analysis of the content of the verbal attacks expressed in Twitter confirms the existence of stereotypes and prejudices that are deeply rooted in Greek society. For example, the dominant stereotypes in the construction of the image of Albanians are associated with "crime" and "cultural inferiority" indicating a continuity of the so-called stereotype of the Balkanian criminal. Crime and inferiority stereotypes are dominant also in the verbal attacks against Muslims and Islam, but with rather different aspects. In particular, the attacks are often lexicalized through evaluative and dysphemistic terms of insult or abuse to debase core Islamic values, principles, practices, etc. indicating irrationalism/inferiority, sexist behavior and fanaticism. The inferiority stereotype is also dominant for Pakistani; most of the verbal attacks against them are lexicalized through derogatory morphological variations of the nationality adjective. In the case of Jews, the verbal attacks entail a perception of a particular enmity towards the Greek nation and blame attribution patterns of the Greek crisis. Common themes in this group are the identification with the negative aspects of the banking system and global capitalism, as well as the frequent appeal to conspiracy theory elements. These observations coincide with surveys that establish a correlation between conspiratorial thinking and ethnocentricism, and elaborate an interpretation of Greek anti-Semitism building on aspects of national identity and by employing the concept of victimhood.

The results illuminate two different dimensions usually correlated to the conceptualization of the phenomenon of xenophobia. On the one hand, attacks against TGs like Germans and Jews, who are considered more powerful, are related to the concept of vulnerability, which implies the perception of threat. On the other hand, dominance is directed against Albanians and Pakistani who are thought of as inferior in socio-economic or cultural perspectives.

7. Conclusions

Focusing on the violence aspect of xenophobia, in this paper we presented a data-driven linguistic approach for exploring the predominant targets of xenophobia-motivated behavior towards specific TGs of interest in Greece over time. The results indicate that examining physical and verbal aggression as indicators of xenophobic attitudes and combining News and Twitter data can provide important insights regarding the nature of this type of behavior and also illuminate the possible reasons behind this complex social phenomenon.

9. Acknowledgements

We acknowledge support of this work by the project “Computational Science and Technologies: Data, Content and Interaction” (MIS 5002437) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund). Part of the work reported here was made possible by using the CLARIN infrastructure**.** The authors are grateful to the political scientists Vasiliki Georgiadou, Jenny Laliouti, Anastasia Kafe, and Ioannis Galariotis for their insights in the interpretation of the results and to Dimitris Galanis for his valuable help in the data processing phase.

8. Bibliographical References

Berkowitz, L. (1993). “Aggression: Its causes, consequences, and control*”.* New York, NY: McGraw-Hill.

Cunningham H., Maynard D. and Tablan V. (2000)/. “JAPE: a Java annotation patterns engine”. Research Memorandum CS–00–10, Department of Computer Science, University of Sheffield.

Harris, B. (2002). Xenophobia: A New Pathology for a New South Africa? Psychopathology and Social Prejudice. In D. Hook and G. Eagle, *Psychopathology and Social Prejudice*, 169-184. Cape Town: University of Cape Town Press.

Papageorgiou, H., Prokopidis, P., Demiros, I., Giouli, V., Konstantinidis, A., and Piperidis S. (2002). Multi–level XML–based Corpus Annotation. In: Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002), Las Palmas, Spain, pp. 1723–1728.

Stathopoulou, T., Papageorgiou, H., Papanikolaou, K., Kolovou, A. (2017). Exploring the dynamics of protest with automated computational tools. A Greek case study. In Computational Social Science in the Age of Big Data. Concepts, Methodologies, Tools, and Applications. German Society for Online Research.

Papanikolaou, K., Papageorgiou, H., Papasarantopoulos, N., Stathopoulou, T., Papastefanatos, G. (2016). “Just the Facts” with PALOMAR: Detecting Protest Events in Media Outlets and Twitter. In International AAAI Conference on Web and Social Media. North America.

Prokopidis, P., Georgantopoulos, B., and Papageorgiou, H. (2011). A suite of NLP tools for Greek. In: Proceedings of the 10th International Conference of Greek Linguistics, Komotini, Greece, pp. 373–383.

Van der Veer, K., Ommundsen, R., Yakushko, O., Higlers, L., Woelders, S., and Hagen K.A. (2013). “Psychometrically and qualitatively validating a cross-national cumulative measure of fear-based xenophobia”, Quality & Quantity 47.3 (2013): 1429-1444.