

Tools and Resources for Detecting Hate and Prejudice against Immigrants in Social Media

Cristina Bosco and Viviana Patti and Marcello Bogetti and Michelangelo Conoscenti
and Giancarlo Ruffo and Rossano Schifanella¹ and Marco Stranisci²

Abstract. This position paper describes the early stages of an on going project for the development of artificial intelligence tools and resources to detect and analyze hate speech and prejudice against immigrants: IHATEPREJUDICE³. It is the result of a multi-disciplinary team that includes skills from computational linguistics and sentiment analysis, to social network analysis and visualization.

The approach follows a holistic and multi-lingual perspective, which encompasses different knowledge sources to be meaningfully aggregated. Thus, although the project is currently mainly focussed on a local case-study (Piedmont, a region of Italy), it is scalable over larger territories and different languages which can be considered as representative, namely the European Union. Since this is quite an innovative approach to the problem, this position paper aims at reaching out colleagues with a potential interest in this area and/or to solicitate feedback by those who are in the early stages of their projects and wish to share their experience or join efforts.

1 INTRODUCTION

Nearly a quarter of a billion people, throughout the world, currently live in a country other than their place of birth. This is an increase of 41% from 2000 to 2015. This figure includes more than 21 million refugees - often vulnerable and dissatisfied. Since 2015 Europe is facing an unprecedented refugee crisis, the by-effect of the Syrian civil war and the terrible living conditions in equatorial Africa. A 1,300,000 people have generated this increased migration flow to Europe which can only but increase, putting European stable societies, so far, under pressure.

Therefore, the implications for the European society and the way we behave towards immigration, immigrant integration and social inclusion for newcomers and their children, are becoming more decisive and must be addressed either at a local or global level, considering a political and social perspective. While this phenomenon stimulates the generation and diffusion of hate speech and hate crimes, at the same time several initiatives are promoted, but they should be further improved to increase the awareness and empathy of receiving populations while avoiding polarization against immigrants.

In this paper we describe objectives, motivations and methodology of a new research project, called *IHATEPREJUDICE*, aimed at addressing hate speech and prejudice against immigrants by exploiting Artificial Intelligence (AI) methods and tools. Its novelty consists in the application, mainly to texts extracted from social media, of

a multi-disciplinary and holistic approach which encompasses tools and resources from different and complementary disciplines related to data analysis, in particular referring to computational linguistics, Sentiment Analysis (SA), Discourse Analysis (DA), computational social science and data visualization.

This is done in order to analyze and detect contents related to immigration to better understand the social dynamics of immigration and thus managing the public debate on migrants' integration by supporting social workers' competences.

AI can play, indeed, an important role on both understanding and managing the complexity of migration issues in order to define better policies towards migrants' integration and community building. On the one hand, computational linguistics and SA techniques can be twofold crucial for accessing and analyzing knowledge about migration and to detect in them hate contents and communication dynamics useful to investigate the birth and diffusion of prejudice against migrants. Since this knowledge is diluted in a huge variety of texts from different sources, and because these texts can be considered a case of big data, which humans cannot directly access in a reasonable amount of time and manner, our integrated approach will facilitate this task. On the other hand, data aggregation and visualization can be applied for sharing such knowledge with operators working in the local area or in a global perspective, but also to make citizens more aware of these issues. Even policy makers can be helped in order to make more informed decisions on matters related to migration management.

The way we address these challenges is innovative, focusing on identifying different patterns of hate and prejudice in on line expressions. The involved phenomena will be studied in a bidirectional perspective, e.g. movements and communication dynamics of immigrants, or the reactions of local community, media and single citizens.

Furthermore, we integrate NLP, data and network analytics, together with visualization approaches based on maps showing the diffusion of phenomena also as related to geographical locations. Data aggregation and visualization will be applied for sharing such knowledge with operators working in the local area, but also with citizens for increasing awareness. A control sample from a multicultural community (UK based) will be generated as well.

We must apply both resources and technologies to these challenges in creative and innovative ways. This means increasing the size of the analyzed big data and the power of AI tools to analyze texts, flows and patterns. By developing computational tools for accessing and analyzing small and big data from different sources/media we will accomplish two goals: on the one hand, **transforming analyzed data in actionable knowledge** for territorial entities working in raising awareness about discrimination and taking action against

¹ University of Turin, Italy, email: {bosco,patti,ruffo,schifane}@di.unito.it {michelangelo.conoscenti,marcello.bogetti}@unito.it

² Acmos, Italy, email: marco.stranisci@acmos.net

³ Fund CSTO160598, Compagnia San Paolo, IT.

it; on the other hand, **democratizing data**, by improving their accessibility and opening new opportunities for enhancing the quality of citizen life.

The paper is organized as follows. In the next section motivations are presented. Then, section three clarifies how the different disciplines involved in the project contribute to the project in an innovative way. In the fourth section we describe the project and we illustrate the case study for the first and prototypical application of our techniques. We end with a section on the expected results and future development of the project.

2 MOTIVATIONS

Population change occurs as a result of two factors: natural change, defined as the difference between the number of births and the number of deaths, and net migration, commonly defined as the difference between immigration and emigration.

Data released by European Community about population change [19] show that from the 1990s onwards natural population change had a diminishing role in EU demographic developments, while the role of net migration became increasingly important. In the period 2011 to 2013, net migration contributed more than 80% to total population growth, drawing an overall pattern of growth of EU's populations driven increasingly by changes in migratory flows, which hides a range of demographic situations among the EU Member States. Between 2004 and 2013, indeed the population of 11 EU Member States decreased, with the biggest reductions recorded in Germany and Romania, but a high overall increase in population numbers was recorded in the other countries like UK (a gain of 4.51 million inhabitants), Spain (3.96 million), France (3.54 million) and Italy (3.29 million). Among these countries, characterized by a negative natural population change, also compounded by negative net migration, Italy is affected by a negative natural change, that was completely offset by net migration which accounted for 108% of the total population change. This makes especially crucial to deal with immigration and related phenomena in this country and motivates the selection of Italy (and of an Italian region) as our case study for the beginning of our project.

Given this scenario, European policies encourage to enforce integration at country, city and community level. Integration is the successful outcome of acculturation, defined as the dynamic, two-way process of mutual accommodation by immigrants and receiving communities. In this perspective immigration can be a source of cultural enhancement. However, negative attitudes to immigration increased in recent years together with prejudice and more or less direct feeling of hostility to foreign people: recent episodes related to the Brexit electoral campaign in UK, to the refugee referendum in Hungary or to the hosting of a small group of refugees in Gorino in Italy are symptomatic of this very worrying trend. Therefore it is gradually becoming more decisive to address in local and global, political and social perspective the implications on Italian and European society of our behavior towards immigration, immigrant integration and social inclusion for newcomers and their children. Anti-immigration attitudes frequently fuels the spread of hate expressions in the variety of media exploited nowadays. Like a mirror of the political and legal choices, of attitudes of the population and feeling encoded in the collective memory of a nation, categories and words exploited for defining migrants are coined and used by humans when faced with the challenges represented by migration phenomenon. They are relevant for determining who is migrant and who is instead not, and for classifying different forms of migrants to be related to different

positions in society, often to be fossilized in prejudices. Language is crucial for determining boundaries among groups and minorities, to confer them specific social roles, rights and opportunities, in practice to describe migrants as others, as outsiders of our society, often as dangerous and unwanted people notwithstanding their real nature. But, on the other hand, language can be also exploited by immigrants for react to the challenges of integration in a different society. Starting from the relevance of language in the immigration phenomena and the huge variety of linguistic data available as digital traces, the main objective of our project is the development of computational tools for accessing and analyzing small and big data from different social media to acquire a deeper knowledge about anti-immigration attitudes, related to **prejudices and hate speech**. We will first focus on Italy and in particular on our region, Piedmont, which is the case study of the project. The local context is crucial when studying attitudes towards immigrants: the type of neighbourhood, city, region, country where an individual lives creates a filter which conditions perceptions and attitudes towards immigrants [13]. We integrate NLP and data analytics, together with visualization approaches based on **maps** showing the diffusion of phenomena in geographical locations. Computational linguistics and sentiment analysis techniques will be crucial for analyzing knowledge about migration diluted in the variety of texts from different sources, for detecting in them hate contents and communication dynamics useful to observe birth and diffusion of prejudice against migrants. Data aggregation and visualization will be applied for sharing such knowledge with operators working in the Piedmont local area, but also with citizens for increasing awareness. A control sample from a multicultural community (UK based) will be generated as well.

Hate speech analysis and **hate maps** allow both a greater understanding of social phenomena linked to the integration of migrants, that more targeted actions to improve it. The integration of migrants is strongly linked to the new cultural context where they try to rebuild their lives. The process of acculturation depends on personal and social variables of the migrant, in large part in turn dependent on the cultural context of his/her origin, on the characteristics of the context of resettlement and on events occurring during this life period. The different migrant's strategies firstly affect the different outcomes achieved. In particular, he can decide whether or not to maintain the cultural identity of origin and whether or not to establish and maintain new relationships within the new context. This gives rise to four possible different outcomes: integration, assimilation, separation/segregation, marginalization [3]. The analysis of migrants' ego-network allows a representation of these outcomes. Part of the network of the country of origin tends to disappear and these relationships need to be redefined in form and contents. In the country of destiny new relationships with co-ethnics, natives and/or other groups are formed. Social networks can thus take different configuration [20]: integration in the country of resettlement, by building relationships with native people; regression, by increasing in relationships with people living in the countries of origin; local concentration, by increasing in relationships in the country of residence, primarily with coethnics but to a lower degree with native people too; ethnic segregation, by increasing in relationships with coethnics, at the cost of relationships with native people; "transnationalization", by increasing in relationships with others, either living in different countries and / or with people from other countries; stability in composition (no change in composition). The integration in particular is the outcome of a process of maintaining the own identity of origin but at the same time of the expansion of relational system with natives, and then of the exposure to the cultural context of insertion. How-

ever, it requires the presence of a multicultural context. The outcome of the process of acculturation and therefore the possibility that integration will occur, depend on a number of factors: in addition to the strategy and personal characteristics (demographic, socio-economic, country of origin) are in fact very important those of society of settlement. Among them, a significant role is given by ethnic attitudes of the native people, as the opening level or prejudice or even overt hostility towards those who come from different cultures, especially if very different from their own. Their interaction with the strategies and the characteristics of newcomers, generates relational systems of the migrants, which in turn are the basis of how the acculturation process flows in one and the other of the possible outcomes. To know, through the analysis of the online language, if, where and how intensively the cultural context expresses opening / closing and even hate towards migrants, allows, on the analysis plan, to understand given the others characteristics- what kind of relationship they will probably create and, accordingly, what kind of social networks they will rebuild and how that can in turn influence the result of the acculturation process. From a practical point of view, there are at least two possible uses. The first concerns the possibility to a certain extent to intervene on migratory flows, in terms of addressing towards areas at lower risk of hostility, or in any case greater awareness of the possible risks of insertion in potentially hostile environments. The second concerns the possibility, as already mentioned, to program adequate action enforcement efforts, that - knowing as hostility is expressed and manifested in the language, and where the focus is more acute - allows to design targeted education and communication campaigns.

3 OBJECTIVES AND METHODOLOGY

The methodology will be based on the development of a computational framework for data semantic exploration. It will follow a holistic approach, which integrates several sources of knowledge to describe a multi-faceted reality. The cooperation of a multi-disciplinary team, but also the participation of associations working on immigration in our local area as stakeholders and domain experts, will ensure an inclusive approach that can produce combined multidisciplinary results greater than the sum of their separate effects. The success of the project strongly depends on the synergetic application of the methodologies from the different research areas, and in particular those related to the linguistic and sociological analysis of communication (see sec. 3.1), those related to the automatic detection of hate speech with the purpose of visualize their aggregations in hate maps (see sec. 3.2), and those coming from the application of Discourse Analysis techniques (see sec. 3.3).

3.1 Linguistic and Sociological Analysis of Hate Speech

The complexity of communication dynamics related to immigration makes challenging the development of novel forms of computational analysis which can sustain the traditional ones and be also adequate for approaching big data [22]. The study of this phenomenon can be indeed fruitfully addressed by analyzing large-scale data, like those represented by social media texts and their users' digital traces, where causes and consequences of hate and prejudice are diluted. The project will apply for achieving this purpose a methodology well known in corpus-based NLP, where data collections are carefully annotated by teams of experts to provide systems with ground-truth information about the observed phenomena. In our project, this methodology consists therefore in: i) collecting several text samples

from several media where hate speech expressions against immigrants are exploited; ii) analyze them for detecting a set of relevant expressions; iii) select a set of texts to be annotated according to a scheme that makes explicit those expressions, their targets and polarization.

These corpora will drive the development of SA tools and data aggregation, since these tools will be trained and tested on them.

3.2 Automatic Hate Speech Analysis and Hate Maps

The framework to be developed during the project integrates innovative techniques for collecting social data and analysing them by data mining and aspect-based sentiment analysis, by applying highly developed natural language processing to deal with the variety of fine-grained aspects involved in hate speech and hostility towards migrants, like e.g. irony and figurative uses of language. It will aggregate and analyze data under several semantic perspectives, including geographical and temporal components. Furthermore, geo-social data will be exploited both to derive a geography of hate and prejudices about immigrants and to study immigrants mobility across our territory.

This web-based platform will be moreover designed for supporting interactive access to analyzed data and visualization of hate maps: a powerful tool to be exploited in the variety of decisions humans must take everyday with respect to their behavior towards community or other people, i.e. for democratizing the knowledge about immigration. The platform will provide indeed a visual easy to read representation of reports coming from the modules devoted to analysis, showing the sentiment, the social and political reactions and the dynamics of the information flow on the topic. The dashboard will allow the different end users to gather the intelligence required, for example, to make informed decisions on local policies or support investigations on on line hate speech and prevent crimes that would potentially affect the stability of local communities.

With pervasive diffusion of mobile devices, social media data provides a spatial dimension that characterizes where a content is generated, providing an additional layer to study the geography of on line information processes. We plan to implement an alternative cartography of a city that visually maps the spreading of hate speech in time and space. This would allow, for example, to observe which neighborhoods of a city are subject to a negative/positive sentiment towards the immigration phenomenon, with the possibility to link the online discourse to offline characteristics of the built environment like demographics, ethnicity, socio-economic conditions of the population. The use of different spatial, e.g., at street, block, neighborhood, city, region, and time aggregations, e.g., day, week, months, years, will provide an effective tool to visually explore the complexity of the phenomenon at different scales.

3.3 Discourse Analysis

In parallel, the Discourse Analyst, will apply and extend the protocol used in [9] to verify how the principles of Audience Architecture [16] and Reverse Language Engineering [8] can integrate the methodological framework discussed above with a double-goal perspective. Exploiting the concept of the "majority illusion, i.e. "a state that is globally rare in a network [is] dramatically over-represented in the local neighbourhood of many individuals" [12] the DA approach will make sure that data gathered with AI techniques are complemented with a specific human filter that aligns the decoded and en-

coded messages with the desired Audience Architecture through a specific language engineering activity. In this way, management of the public debate on migrant integration can be maximised in such a way to increase awareness and empathy while avoiding polarization. The methodological blend of working frameworks will thus ensure an improvement of the responsive platform, even at natural language level, so that message production on Social Media is aligned with the desired master message and preventing a “weaponization” of Social Media themselves. Thus, this approach will develop conceptual maps that will favour a better memetic writing that resonates with users expectations of a responsive and customized compelling narrative experience [21]. This approach offers several advantages: it is flexible as it adjusts to the users needs being the innovative approach that is able to accommodate the technological innovation. It is multipurpose as it tracks a) shifts in language use with customised, detailed focus (time and space); b) antagonists strategies; c) it is able to detect and interpret rumours and background noise that are not considered by standard software/reports based on big data, but that can generate, because of the majority illusion undesired effect in the language of the neighbourhood and thus, hate speech. In this way the project should contribute to a better understanding of how belief systems work as networks [10], [11] and, on the other hand, how Social Media networks foster certain belief systems to the point that they develop into hate systems.

4 CASE STUDIES AND PROJECT SCENARIO

In this section, an application of our methodology to a case study, where the availability of a computational framework for automatically detecting hate speech against immigrants may contribute in transforming information in actionable knowledge, is described. According to the Council of Europe, the term “hate speech” shall be understood as covering all expressions which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance⁴. So hate speech is a twofold concept: it has both a semantic (the insult), and a pragmatic level (the public incitement).

If we focus on the pragmatic perspective, considering the ability to spread online hate contents to a wide audience, it is important to analyse if, how, and how successfully newspapers incite their readers to hatred, by spreading online hate contents. Italian authorities are very sensitive to this topic: in 2008 the *Carta di Roma*, a code of conducts about how to treat the immigration topic, was adopted by the Order of Journalists⁵. Nevertheless, the presence of hate speech carried out by Italian newspapers is still pervasive. Several cases can be found in newspapers everyday. For instance, September 23, 2016 the Italian newspaper *Libero* published the hatred content on immigrants shown in Figure 1, both on its website and on its Twitter account.

Starting from the publication of this information, we can study how hate speech spreads in Twitter. After being published on Twitter, the news provoked negative reactions against refugees among *Libero*’s readers, who are in synch with the incitement expressed in the news. They expressed online their intolerance towards refugees, by posting messages like the following:

‘Clandestini non profughi’
‘Illegals not refugees’

⁴ Recommendation no.R (97) 20 of the Committee of Ministers to Member States on Hate Speech, October 30th 1997.

⁵ <http://www.cartadiroma.org/>

Vittorio Veneto, il Paese dove fanno votare anche i profughi



Figure 1. The title of an article that incites to hate: “Vittorio Veneto, the city where even refugees can vote”

Another reader, acting similarly, interpreted the possibility of voting for refugees as a simple stratagem applied by the political party that introduced this novelty, and expressed therefore his/her hate against it by the following post:

‘Il PD (Partito Democratico) vuol vincere barando’
‘PD (Democratic Party) wants to win [the elections] by cheating’

This example shows what kind of dynamics feature the spread of hate speech, and what kind of paths we have to follow in order to detect hate and prejudice against immigrants. The main lesson learned is that we need a direct access to data, and to data from different sources and media. Following this type of indication, we are building a methodology aimed at creating a dataset of hate speeches that spread from newspapers to social, thus focusing either on the semantic and the pragmatic aspects of hate speeches. The methodology is structured in four steps: first, the identification of *Carta di Roma* violations on the most important online newspapers; second, the analysis of their disseminations among readers. e.g., how many Twitter users reply or retweet these news; third, the linguistic analysis of data, aimed at informing SA tools; four, the visualization of these discourses.

At the present stage we have built a small dataset of 1,312 crime news of the Turin Metropolitan Area occurred in 2016, and published by 6 national newspapers: *La Stampa*, *La Repubblica*, *Il Corriere della Sera*, *Il Fatto Quotidiano*, *Il Giornale*, *Libero*. We collected articles using the Google Custom Search API⁶, then we searched for the ones which contain references to the most prevalent nationalities among immigrants: Romanian, Albanian, Moroccan, Tunisian, Chinese; and two generic terms referring to immigration: immigrants and illegal immigrants. Table 1 reports some statistics about the articles which contain at least one term referred to nationality. Different columns refer to different news’ portions where references to nationality occur (i.e. in the title, in the tagline, and in the body of the news).

A team of linguists is now analyzing the dataset with two main aims: find recurring linguistic expressions that convey hate contents; understand which of these news are potential hate speeches and

⁶ Google strongly limits the number of searches a user can do, so we are developing a number of scripts that collect news directly from newspapers pages.

newspaper	title	tagline	article	total
<i>La Stampa</i>	10	18	147	175
<i>La Repubblica</i>	1	2	24	27
<i>Il Giornale</i>	12	12	46	70
<i>Libero</i>	3	6	19	28
<i>Il Fatto Quotidiano</i>	3	4	18	25
<i>Il Corriere della Sera</i>	3	3	23	29

Table 1. Number of references to the most prevalent nationalities among immigrants in six Italian newspapers

which not. The latter aim is crucial. Indeed it allows to focus on counter-narratives: news that foster tolerance among citizens. For instance, the newspaper *Il Corriere della Sera* published a news about a young Moroccan student and peddler, that is almost achieving his master degree at the Faculty of Engineering:

‘Rachid Khadiri, studente d’ingegneria marocchino
Vendeva accendini, ora dopo la laurea al Politecnico di Torino, sta frequentando il biennio della specialistica’

‘*Rachid Khadiri, Moroccan Engineering student*
He sold lighters and now, after his graduation at Turin’s Polytechnic, he is attending the two-year master course’

Thus, the ongoing research provides both a system of hate speech reporting, and a set of counter narratives that can be also used to train journalists to avoid hatred contents and to spread tolerance.

5 RELATED WORK

The immigration phenomenon stimulates the production and diffusion of hate speech and hate crimes. Against it several initiatives are promoted, let us mention the *No hate speech movement*, a youth campaign of the Council of Europe to reduce the levels of acceptance of hate speech, the *Research - Report - Remove* project⁷, and the *Map of Intolerance* promoted by *Vox Diritti*⁸. Such initiatives are the most often devoted to the detection of the target/diffusion of hate speech or to their containment.

Few studies addressed the detection of sentiment about immigrants, like e.g. [7]. Our approach is novel and unique, to the best of our knowledge, since 1) we propose to combine fine-grained computational linguistics analysis, including sentiment analysis, and data visualization based on interactive maps; 2) we address hate and prejudice about the specific immigration phenomenon in Italy (in particular in our region) and define strategies for increasing tolerance and integration and enabling stakeholders to understand hate spreading within communities. This has never been studied before.

As far as finer-grained tasks related to sentiment analysis are concerned, let us mention two important tendencies that we will consider in our work: on the one hand the recent and growing interest of the NLP community in the development of aspect-based techniques where the detection of sentiment targets (and orientation towards/against them) is addressed beyond that of generic polarity of expressions (see e.g. stance detection, [15, 14]); on the other hand the reflections on the impact of using figurative language on sentiment analysis. They rise issues on the importance of taking into

account even non literal meanings expressed in texts whose presence may undermine the results of SA systems, with a particular emphasis on ironic contents [6, 18].

Among the aims of the project, the management of the public debate on migrant integration devoted to increasing integration, awareness and empathy with migrants, is crucial. Since debates are contexts where opinions are usually strongly polarized, in literature we can find works about debates with a strong social impact where social media have been analyzed and annotated to be exploited in training and testing sentiment analysis systems. For Italian, which is the language of the case study addressed first of all in our project, see e.g. [6, 17, 1]. For French we developed a corpus on the debate on the homosexual wedding [4]. Also for Catalan and Spanish will be soon available a corpus about the debate on the separatism of Catalonia [5]. The most of these corpora have been usefully exploited also within the context of shared task on SA in evaluation campaigns for NLP tools and resources [1, 2]. This is a guarantee of their quality and usability as benchmarks on which the tools developed in the project can be trained and tested. In continuity with this line of research about the analysis of sentiment and irony in political debates, we will contribute to the development of lexical resources with a specific and novel focus devoted to detect sentiment, hate and prejudice about immigrants in Italian social media in a multi- and cross-linguistic perspective.

Visualization techniques will be applied for an interactive access to maps and other data aggregation displays, which is advanced and novel w.r.t. the state-of-the art related to the Hate Maps. Few initiatives for visualizing hate speech indeed exist, see e.g. the *Geography of Hate* project in US⁹ and the *Map of Intolerance* promoted by *Vox Diritti* in Italy, but they do not have a special focus on immigrants as in the present project and do not provide any possibility to dynamically interact with the map and with the source data.

6 IMPACT

The expected result of the project is the acquisition and spreading of a wider knowledge about immigration, to be exploited in the education to citizenship. The automatic analysis of hate speech and the detection of immigration networks can be indeed exploited for better designing activities related to the awareness of people, for improving the quality of the information diffused by newspapers and other media, to more adequately deal with vocational training for young people (born both into Italian or immigrant families), as a means to support the teaching methodology for preventing disadvantages and discrimination. This knowledge can be exploited in preventing conflicts between groups, in promoting the tolerance and integration among citizens and immigrants, a dynamic, two-way process of mutual accommodation by immigrants and receiving communities. Since communities featured by successful integration of immigrants are proved to be stronger economically and more inclusive socially and culturally, and that adequate knowledge allows to prevent immigrants and citizens to be exposed to inhuman conditions and crimes, this may produce an enhancement of citizen life quality in our territory.

We will develop tools and resources scalable across a European perspective (e.g. ready to be applied to different languages), but the application on a local case study is twofold crucial: for testing and trying the approach and for developing immediately impacting actions.

⁷ <http://www.inach.net/RRR-description.pdf>

⁸ <http://www.voxdiritti.it>

⁹ http://users.humboldt.edu/mstephens/hate/hate_map.html

The cooperation with three local associations will enhance the impact of the project: Acmos¹⁰, Babelica¹¹ and Fondazione Piazza dei Mestieri¹². They will be stakeholders expressing needs to be satisfied by the project, but also domain experts, testers and will have a crucial role for in dissemination of project's results. We have a long-time collaboration with Acmos for promoting an educational path about hate speech in schools. The interest of Babelica on *IHATEPREJU-DICE* is in line with their ongoing project for involving a new audience in the creation of a new culture of inclusion. Our cooperation with Piazza dei Mestieri will be instead related to their work in training young people and supporting teaching methodologies aiming at social inclusion policies. Beyond the activities devoted to the education of citizens, our actions on territory will also impact on long-term monitoring of the evolution of immigration phenomena, producing a repository of knowledge to be exploited by policy makers for improving integration.

7 CONCLUSION

Summarizing, the aim of the project we present is to build an interactive web tool for the collection, analysis and visualization of the discourse on immigrants and the reactions of the community talking in social media, with the goal of capturing in a quantitative fashion through visuals how people perceive and share the hate and prejudice against foreigners living in our area.

Our approach will be based on a holistic and multi-disciplinary perspective, since the addressed topic inherently requires the exploitation of techniques from several different disciplines and the collection of information also from sources other than social media useful for defining the semantic interpretation context for social media posts. The results of the project will be evaluated according to several perspectives, observing the quality of the sentiment engine and of the interface for displaying results, but also considering the usability and adequateness of these tools as detected by the users (first of all the associations involved in the project). As far as usability is concerned, we will organize trials with groups of users where they will be asked to fill in a questionnaire to track user's experience and collect feedbacks. Following well-established evaluation methodologies, we will compare the results generated by the automatic tools in SA, data analysis and aggregation with those produced by a team of humans on the same task and having access to the same gold standard data set. This evaluation will constantly run during all the phases of the project. The role of local associations in the activities foreseen by the project will be crucial, as they will act as domain experts along three main lines: needs assessment; testing and evaluations; dissemination of the results of the project in their area of action and integration of such results with their local activities related to the immigration phenomenon.

ACKNOWLEDGEMENTS

The work of Cristina Bosco, Viviana Patti and Giancarlo Ruffo have been partially funded by Fondazione CRT (*Hate Speech and Social Media*, project n. 2016.0688).

REFERENCES

- [1] F. Barbieri, V. Basile, D. Croce, M. Nissim, N. Novielli, and V. Patti, 'Overview of the Evalita 2016 SENTiment POLarity Classification Task', in *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016), Napoli, Italy, December 5-7, 2016.*, volume 1749. CEUR-WS.org, (2016).
- [2] V. Basile, A. Bolioli, M. Nissim, V. Patti, and P. Rosso, 'Overview of the evalita 2014 sentiment polarity classification task', in *Proceedings of the 4th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA'14)*, pp. 50–57. Pisa University Press, (2014).
- [3] J. W. Berry, 'Immigration, acculturation, and adaptation', *Applied Psychology*, **46**(1), 5–34, (1997).
- [4] C. Bosco, M. Lai, V. Patti, and D. Virone, 'Tweeting and being ironic in the debate about a political reform: the french annotated corpus TWitter-MariagePourTous', in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pp. 1619–1626. ELRA, (2016).
- [5] C. Bosco, M. Lai, F. M. Rangel Pardo, and P. Rosso, 'Tweeting in the debate about catalan elections', in *Proceedings of Workshop on Emotion and Sentiment Analysis, at LREC 2016*. ELRA, (2016).
- [6] C. Bosco, V. Patti, and A. Bolioli, 'Developing Corpora for Sentiment Analysis: The Case of Irony and Senti-TUT', *IEEE Intelligent Systems*, **28**(2), 55–63, (2013).
- [7] J.D. Carr, 'Measuring Twitter sentiment and implications for social psychological research', *American Journal of Applied Psychology*, **2**, (2014).
- [8] M. Conoscenti, *Language Engineering and Media Management Strategies in Recent Wars*, Roma: Bulzoni, 2004.
- [9] M. Conoscenti, *La Stampa Locale in Piemonte nell'Anno Europeo del Dialogo Interculturale (2008). Un'analisi discorsiva secondo i principi della corpus linguistics*, volume 22, Europa e Mediterraneo. Storia e immagini di una comunit internazionale, Cagliari: CNR, Istituto di Storia dell'Europa Mediterranea, 2011.
- [10] N.E. Friedkin, A.V. Proskurnikov, R. Tempo, and S.E. Parsegov, 'Network science on belief system dynamics under logic constraints', *Science*, **354**, 321–326, (2016).
- [11] Z. Karpinski and J. Skvoretz, 'Repulsed by the 'other': Integrating theory with method in the study of intergroup association', *Sociological Theory*, **33**, 20–43, (2015).
- [12] K. Lerman, X. Yan, and X-Z Wu, 'The 'majority illusion' in social networks', *PLoS ONE*, **11**, (2016).
- [13] Y. Markaki and S. Longhi, 'What determines attitudes to immigration in european countries? an analysis at the regional level', Technical report, (2012). discussion paper no. 2012-32.
- [14] S. M. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, 'A dataset for detecting stance in tweets', in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, (may 2016). ELRA.
- [15] S. M. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, 'Semeval-2016 task 6: Detecting stance in tweets', in *Proceedings of the International Workshop on Semantic Evaluation, SemEval '16*, San Diego, California, (June 2016).
- [16] B. Pearson, *Storytizing: What's Next After Advertising*, Austin (TX): 1845 Publishing, 2016.
- [17] M. Stranisci, C. Bosco, D. I. Hernández Farias, and V. Patti, 'Annotating sentiment and irony in the online italian political debate on #labuonascuola', in *Proceedings of LREC 2016*, pp. 2892–2899. ELRA, (2016).
- [18] E. Sulis, D. I. Hernández Farias, P. Rosso, V. Patti, and G. Ruffo, 'Figurative messages and affect in Twitter: Differences between #irony, #sarcasm and #not', *Knowledge-Based Systems*, **108**, 132 – 143, (2016). New Avenues in Knowledge Bases for Natural Language Processing.
- [19] European Union, *People in the EU: who are we and how do we live? - 2015 edition*, Eurostat Statistical Books, Luxembourg: Publications Office of the European Union, 2015.
- [20] Raffaele Vacca, Giacomo Solano, Miranda Jessica Lubbers, Jos Luis Molina, and Christopher McCarty, 'A personal network approach to the study of immigrant structural assimilation and transnationalism', *Social Networks*, (2016). In Press. Available online 23 September 2016.
- [21] D. Westen, *The Political Brain*, New York: Public Affairs, 2007.
- [22] E. Zagheni, V.R.K. Garimella, I. Weber, and B. State, 'Inferring international and internal migration patterns from Twitter data', in *Proceedings of the 23rd International World Wide Web Conference, WWW '14 (Companion Volume)*. ACM, (2014).

¹⁰ <http://acmos.net>

¹¹ <http://www.babelica.it/>

¹² <http://www.piazzadeimestieri.it/en/>