Annotating Sentiment and Irony in the Online Italian Political Debate on #labuonascuola

Marco Stranisci[1,3], Cristina Bosco[1], Delia Irazú Hernández Farías[1,2], Viviana Patti[1]

[1]Università degli Studi di Torino,
Dipartimento di Informatica
Corso Svizzera 185 – 10149 Torino, Italy,
[2]Universitat Politècnica de València
Departamento de Sistemas Informáticos y Computación
Camino de Vera s/n. 46022 Valencia, Spain
[3]weLaika,
Via Salgari 7, Torino, Italy
marco.stranisci@welaika.com, {bosco,patti}@di.unito.it, dhernandez1@dsic.upv.es

Abstract

In this paper we present the TWitterBuonaScuola corpus (TW-BS), a novel Italian linguistic resource for Sentiment Analysis, developed with the main aim of analyzing the online debate on the controversial Italian political reform "Buona Scuola" (Good school), aimed at reorganizing the national educational and training systems. We describe the methodologies applied in the collection and annotation of data. The collection has been driven by the detection of the hashtags mainly used by the participants to the debate, while the annotation has been focused on sentiment polarity and irony, but also extended to mark the aspects of the reform that were mainly discussed in the debate. An in-depth study of the disagreement among annotators is included. We describe the collection and annotation stages, and the in-depth analysis of disagreement made with Crowdflower, a crowdsourcing annotation platform.

Keywords: sentiment analysis, Italian, political debates, Twitter

1. Introduction and Motivation

The application of sentiment analysis and opinion mining techniques to the political domain has encountered rising interests in recent years, due to the exponential increase of user-generated contents through social media, such as Twitter, which have been extensively exploited for debating politics and public policies. We can outline two tendencies in this area. The first one focuses on studying electoral campaigns to gauge the political sentiment (Tumasjan et al., 2011; Sang and Bos, 2012), political trust analysis (Ceron et al., 2014), or users political alignment (Conover et al., 2011). The second one, instead, is less related to potential commercial applications, but mainly aimed at investigating socio-political issues (Maynard and Funk, 2011), user's stance on controversial topics (Rajadesingan and Liu, 2014), or debates on public policies (Gloor et al., 2009; Lai et al., 2015).

The present work is in line with the second tendency. We propose a novel data-driven approach to the creation of a manually annotated corpus for Italian sentiment analysis on a popular political debate. Our dataset is composed by Twitter messages about the controversial school reform called *La Buona Scuola* (Good School), discussed in Italy in 2014. The TWitterBuonaScuola corpus (TW-BS henceforth) is manually annotated at the message level for sentiment polarity and irony, in line of previous work on Italian (Bosco et al., 2013; Basile et al., 2014b)[1]. Moreover, our annotation scheme has been extended in order to mark the 'aspects' of the reform that were mainly discussed in the debate. This opens the way to the exploitation of the corpus

for a fine-grained sentiment analysis, where it is possible to go beyond the detection of a generic sentiment in the political debate, by identifying different aspects of the reform at issue and the sentiment expressed towards each aspect.

On the one hand, the project provides a new linguistic resource which meaningfully enriches the scenario of datasets available for Italian, enabling also a deeper and finer grained analysis of sentiment related phenomena. On the other hand, it allowed us to investigate communication dynamics between the Italian government and civil society. We took advantage of this analysis in the definition of a data-driven annotation scheme, to be tested also in the development of other resources in a cross-linguistic perspective (Bosco et al., 2015b; Lai et al., 2015). Indeed, together with a project on the debate about the homosexual wedding in France and an ongoing project on the debate on Catalan separatism held at the end of 2015 in Spain in both Catalan and Spanish, the project presented in this paper contributes to the definition of a wide scenario on political debate to be exploited in a multilingual perspective.

The paper is organized as follows. The next section surveys the area, while the following describes the collection of the dataset and some analysis about the contribution of users to the discussion and the communicative strategies implemented in the context of the debate. In section four, instead, the annotation schema is presented and its application to the data is discussed in particular for what concerns the detected inter-annotator disagreement.

2. Related Works

Annotated corpora for sentiment analysis and opinion mining are useful to train machine learning statistical tools for classifying sentiment. Focusing on Italian, currently

---

[1]A preliminary version of project is presented in (Stranisci et al., 2015).

still an under-resourced language, let us mention, among the existing resources, the Senti-TUT corpus (Bosco et al., 2013; Bosco et al., 2014; Bosco et al., 2015a), which has been exploited together with the TWITA corpus (Basile and Nissim, 2013) in the SENTIment POLarity Classification (SENTIPOLC) shared task (Basile et al., 2014b), proposed during the last edition of the evaluation campaign for Italian NLP tools and resources (Attardi et al., 2015). Sentipolc's dataset includes tweets with politics as topic. The TW-BS new corpus extends the available Italian Twitter data in the domain of political communication, and it is created with an annotation schema which is compatible with the Sentipolc's one. For what concerns, instead, the development of resources about political debates, let us mention a corpus of tweets in French language about the reform of the homosexual wedding centered on the hashtag *Mariage pour tous* (Bosco et al., 2015b; Lai et al., 2015), which has been developed by following a similar methodology.

3. Collection of the Dataset

The collection of the corpus described in this paper has been driven by the need of representing the particular features of the sort of *conversational context* which is currently realizing in social media communication. Indeed, we mainly focused on an evident feature of Italian political debate: the Italian current government's systematic use of *frames* to impose a narration of reforms to its opponents (Conoscenti, 2011). The prime minister Matteo Renzi, leader of the Italian Democratic Party, has indeed shown a great ability in exploiting social media effectively, managing to shape the public discourse on the basis of the characteristics and constraints imposed by such platforms. Observing that such communication strategy applied both in Twitter and in the online consultation organized by the government on a specific website developed for this purpose, we collected data from both the sources.

Three major reform proposals have driven the Italian political debate in 2014: electoral reform, labor reform and the reform of the school, associated with the hashtags *#italicum*, *#jobsact* e *#labuonascuola*, respectively; but, according to a quantitative study, since February 22, 2014, when the current government took office, to 31 December, 2014, *#labuonascuola* was the most popular among the three hashtags and this influenced our decision to focus on the political debate about school reform.

Moreover, since the government organized an online consultation on this reform by developing a specific web platform – labuonascuola.gov.it – for this purpose, we decided to collect data from both Twitter and the platform, with the aim to compare the analysis of opinions expressed by users in different online environments. Summarizing we collected two datasets:

WEB-BS corpus We collected all texts from the web consultation (WEB-BS corpus henceforth) organized by the government on the website `http://labuonascuola.gov.it`. It includes 4,129 posts published from September 15th, 2014, to November 15th, 2014, on the consultation website. In particular, we collected 2,043 posts from focus groups ("dibattiti diffusi"

in Italian) composed of student, teachers, schools, citizens, who joined the consultation about the *laBuonaScuola* reform in order to express their opinion on the subject. Each comment posted was manually tagged by authors with one among 53 topic labels made available on the platform (e.g., *evaluation*, *merit*, etc.). Moreover, users manually tagged the post according to the polarity and nature of the comment, by using one of the following predefined categories: 'what I like', 'what I don't like', 'what is missing', and 'new integration'. Therefore, we can consider such texts as already annotated both for what concerns the polarity of the opinion on the reform, and for what concerns the finer-grained information about the specific topics (or *aspects* of the reform) that were addressed in their comments.

TW-BS corpus The TWitter-BuonaScuola corpus (TW-BS) include texts collected from Twitter by filtering the tweets posted 3rd September 2014 (when the reform was announced by a governmentÕs press conference) - 15th November 2014, filtered for the presence of the hashtag #labuonascuola and of keywords like 'la buona scuola', 'buona scuola', 'riforma scuola', 'riforma istruzione', and without retweets and replies. It includes 35,148 tweets (extracted from the 218,928,483 gathered in all 2014 with stream API with Italian filter proposed in (Basile and Nissim, 2013)), firstly reduced to by automatically removing retweets, and subsequently to 8,594, after a manual revision devoted to further deleting duplicates and partial duplicates which can be hardly automatically detected.

The availability of these two datasets on the same topic - the first one from the web consultation and the other one from the microblogging platform Twitter - makes the collection especially interesting because of the variety of comparisons which is possible to explore. Even if in the current phase of the project we were mainly focused on the TW-BS corpus, which has been indeed annotated and more deeply analyzed, the usefulness of the other dataset cannot be reduced to the extraction of the semantic areas applied in the annotation, but also in the definition of a wider notion of context for the study of the communicative strategies involved in the debate.

3.1. Observing users' behavior

An interesting perspective for the observation of the dataset is that offered by users' behaviors within the debate. For what concerns the TW-BS corpus we can observe, in particular, the presence of a small set of users that proposed opinions and contributed to the debate. The distribution in Figure 1 shows that in the TW-BS corpus 20% of users posted near 60% of tweets.

If we compare this result with the traditional 80/20 rule, known as the Pareto principle, applied to social media, i.e. 20 percent of a group will produce 80 percent of the activity, user activity in TW-BS corpus seems to be somehow less unbalanced. This can be explained by the fact that in the TW-BS corpus there is a greater number of regular
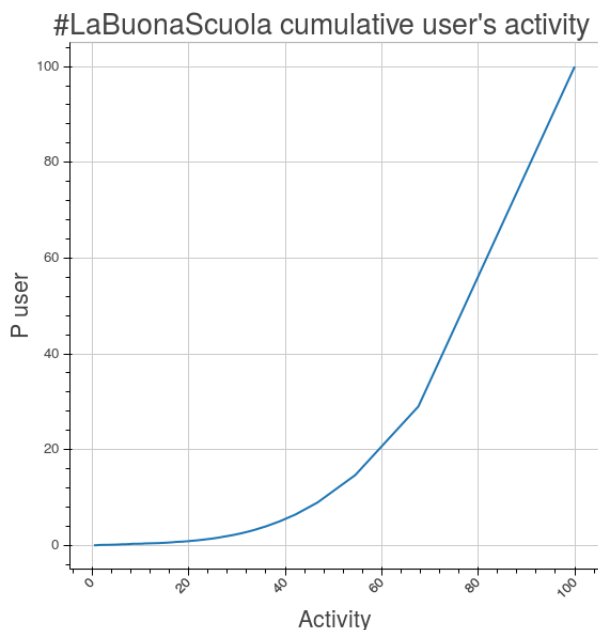
Figure 1: Cumulative distribution of user activity in *laBuonaScuola*

user	tweets
robertoboggiani	162
LipScuola	141
MiurSocial	139
patriziaprestip	101
dominicipi	101
v4leW	94
AnnaMonia_A	77
herrbenanton	70
FLuccisano	70
TecnicaScuola	65

Table 1: The users that used more frequently the hashtag *#labuonascuola* in TW-BS.

user	tweets	retweets	retweets/tweets
matteorenzi	2	806	403
Rai_Arte	1	63	63
anbarone	1	58	58
ItsCetty	1	55	55
VujaBoskov	1	32	32
MazzOnedaje	1	31	31
ilmanifesto	1	30	30
fnicodemo	3	89	29.66
FBastardInside	3	85	28
Giovan73	6	165	27.5
mariannamadia	1	27	27
SteGiannini	43	1153	26.814

Table 2: The users whose posts has been mostly retweeted in TW-BS.

users, whereas sporadic users are less present in this corpus.

In order to extract information related to the user influence, it is useful to observe not only the most active users, i.e. those posting the most of tweets, but also how much each tweet influenced the debate, i.e. how many times it has been retweeted by other users. The users that used more frequently the hashtag #labuonascuola in TW-BS are listed in Table 1. Among the detected users we can see, in particular, the first in the list that is a teacher who describes himself as opponent of the reform and communist. The second name corresponds to a public account of a group of teachers, students, and parents proposing a reform alternative to *laBuonaScuola*. Furthermore, there are two profiles who belong to the Italian Democratic Party's entourage (*patriziaprestip* and *FLucisano*). At the same time we can observe that the third most prolific account is the MIUR official profile. Then, we have the account of another teacher (*vle4W*), a University professor (*dominicipi*), a journalist (*herrbenanton*), a nun (*AnnaMonia_A*), and the profile of a thematic blog (*TecnicaScuola*).

Another measure of the users' influence is given in table 2 which shows the list of users whose posts has been mostly retweeted in the TW-BS corpus (considering only tweets including the #buonascuola hashtag). Looking at the users' profile, we observe that some accounts refer to associations against the school reform, institutional accounts of members of the government, teachers or bloggers which exploit the hashtag's virality to increase their popularity.

Instead, a sample of the most retweeted messages containing the hashtags #labuonascuola can be seen in Table 3[2].

Those tweets were the only posts overriding the 100 threshold in terms of sharing. It is interesting to notice that all users posting the most influential tweets, according to this measure, are related to the government which is proposing the bill. In particular two of the three messages have been posted from the Prime Minister Matteo Renzi, whereas the third one has been published from the official account of the Ministry of Education. However, this does not necessarily imply a broad agreement on the bill *laBuonaScuola*, as we observed that numerous replies to those tweets convey a negative opinion on the proposal.

4. Annotation: the Schema, its Application and the Analysis of Disagreement

In this section we will focus on the annotation of TW-BS observing in particular the annotation process, the schema applied on the data set and finally the knowledge that can be extracted analyzing the annotated data.

According to our annotation schema, each tweet needed to be tagged for polarity, presence of irony and topic (i.e. aspect of the reform discussed), exploiting the tags described below. All them may give a significant contribute to the de-

---

[2]The English translation of the posts follows: matteorenzi - These are our 12 points for #labuonascuola. From September 15th to November 15th we will be in each institute http://t.co/CPF3DF9gd4; MiurSocial - #labuonascuola in #12points #stepafterstep @SteGiannini @matteorenzi http://t.co/EYja76pIWX; matteorenzi - appointment at 10 on http://t.co/p5c1lKamaZ #Italyleavesagain #labuonascuola

user	tweet	retweets
matteorenzi	Ecco i nostri 12 punti per #labuonascuola. Dal 15 settembre al 15 novembre saremo in ogni scuola http://t.co/CPF3DF9gd4	600
MiurSocial	#labuonascuola in #12punti #passodopopasso @SteGiannini @matteorenzi http://t.co/EYja76pIWX	381
matteorenzi	Alle 10 appuntamento su http://t.co/p5c1lKamaZ #italiariparte #labuonascuola	206

Table 3: The most retweeted posts during the consultation on *laBuonaScuola*.

tection of meaning, especially for what concerns the content of posts related to sentiments and opinions, but they must be considered as interrelated.

4.1. Annotation of polarity

For the annotation of polarity, we exploited the same labels of the Senti-TUT project[3], POS, NEG, NONE and MIXED reported in Table 4. Moreover, the annotation schema was completed by a tag for unintelligible tweets (UN), one for duplicates (RT), and one for posts irrelevant to the debate at issue (NP). The latter label has been in practice used only in the annotation of the TW-BS corpus, since some of the hashtags and keywords used for filtering this data set have been also improperly exploited by users, e.g., *'Vi sto rompendo con troppi tweets, buona scuola a tutte e ci vediamo dopo, vi lovvo muchissimo *come capire che sono del linguistico* [emoji]'* (I'm annoying you with to many tweets, good school to everybody and see you later, I lovve you very-much *to show that I from the linguistic high school*).

label	polarity
POS	positive
NEG	negative
NONE	neutral
MIXED	both positive and negative
UN	unintelligible content
RP	repetition of a post
NP	post about a non pertinent topic

Table 4: Polarity tags annotated in TW-BS.

The MIXED label has been introduced to deal with cases where the tweet includes some sentiment or opinion which is positive for a target entity, but negative for another, or where a mixed sentiment on the reform is expressed:

TW-BS-507942755805708288
'parere su #labuonascuola: belle idee e tentativo di politica di espansione del valore della scuola, ma un grande problema e alcune critiche'
(opinion on #labuonascuola: good ideas and an attempt of strategy of expansion of the value of the school, but a big problem and some criticisms)

4.2. Annotation of irony

We also annotated the presence of ironic devices. Indeed figurative devices are well known in sentiment analysis and

opinion mining literature because their presence in texts may undermine the accuracy of sentiment classifiers not aware of them, but the more suitable strategy to deal with them currently consists in referring to annotated data sets. In fact, as shown in several works in literature (Bosco et al., 2013), regardless of the affective words exploited, the polarity of a post cannot be reliably determined without taking into account the possible presence of irony, which can reverse the polarity as in the following example:

TW-BS-507091575764418560
'@matteorenzi dare la scelta dei docenti ai presidi? Dopo le lobby universitarie arrivano quelle scolastiche? Fantastico! #labuonascuola'
(@matteorenzi to give the choice of teachers to the headmasters? After the academic lobbies are coming those from schools? Fantastic! #labuonascuola)

Annotating the presence of ironic devices is a challenging task because the inferring process of this figure of speech does not always lie on semantic and syntactic elements of texts (Reyes et al., 2013; Reyes and Rosso, 2014; Maynard and Greenwood, 2014; Ghosh et al., 2015), but often requires contextual knowledge (Sperber and Wilson, 1986; Wilson, 2006). In order to mark irony, we introduced two polarized ironic labels: HUM NEG, for negative ironic tweets, and HUM POS for positive ironic tweets.

4.3. Annotation of aspects of the reform addressed in the debate

For the annotation of topic we considered the 13 tags more exploited by users during the government consultation on the online platform (see Table 5).
We also created a general topic-tag for tweets that ain't fell in none of the above, and for tweets just indirectly targeted to school reform. The task for each annotator was the selection of one or two tags for each post to be annotated.
The final dataset, cleaned out from unintelligible, duplicates and irrelevant tweets, includes 7,049 items.

4.4. Annotation Process and Analysis

As far as the annotation process is concerned, eight human skilled annotators (different ages, varying from 25 to 50, three males and five females) have been involved in the annotation process producing for each tweet at least two independent annotations.
Considering the complexity of the task, we developed a set of guidelines for annotators, examples of ambiguous cases, and a glossary of the reform terminology, that is available

---

[3]www.di.unito.it/~tutreeb\corpora.html

semantic tag	translation
docenti	teachers
valutazione	evaluation
formazione	training
alternanza scuola/lavoro	school-work
investimenti	investments
reclutamento	recruitment
curricolo	curriculum
innovazione	innovation
lingue	languages
merito	merit
presidi	headmasters
studenti	students
retribuzione	remuneration

Table 5: Aspects of the reform annotated in TW-BS.

Figure 2: The percentage distribution of labels after the third annotation

together with the corpus. These guidelines are the result of a discussion with the annotators, devoted to the reduction of their biases, that took place after the annotation by all them of a small portion of the data set.

Overall, for each tweet two independent annotations were provided, and 4,813 tweets were labeled with the same tag. The detected inter-annotator agreement at this stage was $\kappa = 0.492$. Polarity labels were distributed as follows: NONE (2,649), POS (497), NEG (1,381), HUM POS (18), HUM NEG (404), MIX (44), as shown in figure 2.

The analysis of the disagreement shows two important trends.

The first trend concerns the detection of irony. In 535 messages annotators agreed about polarity but one of them did detected irony, while the other did not, confirming the difficulty of this task and how the perception of irony can vary among humans. Conversely, there were some cases with a disagreement both on polarity and irony detection (72 occurrences). More generally, the co-occurrence of irony and positive polarity was very rare, hardly recognizable (18 occurrences), and very subjective, in line with psychological findings. Moreover, as observed in (Karoui et al., 2015), also in Twitter there are cases where information about external context can be essential in order to identify irony, since the incongruity between the literal and the intended meaning of an utterance cannot be detected by relying exclusively on the information internal to the tweet.

Another important issue concerns the NONE label; 1,448 out of the 2,236 tweets in the disagreement set were cases in which one of the two annotators selected the NONE label, while the other perceived a defined polarity. This can be motivated by the political nature of the corpus which includes several reform proposals and headlines, whose classification in polarity classes can be very hard for annotators. The main motivation of the relatively high disagreement seems therefore to be related more to an effective difficulty of the task rather than to some limit in guidelines or in the annotators' training.

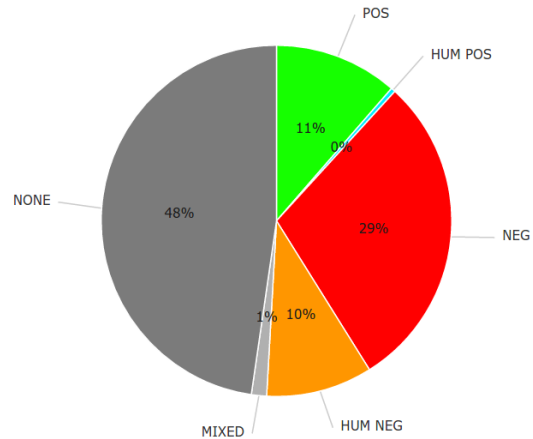In order to solve the disagreement and to further validate the annotation done, we applied a third annotation, using

Crowdflower[4], a crowd-sourcing platform for manual annotation already used in literature for similar annotation tasks (Ghosh et al., 2015). After this further stage we obtained the final version of the corpus which includes 6,659 tweets, to be released and made available to the community[5]. The use of Crowdflower also confirmed the distribution of polarity tags found in the previous part of the annotation process. Figure 2 shows the distribution of sentiment polarity tags after the third annotation and in the released resource: NONE: 3,177, NEG: 1,951, POS: 756, HUM NEG: 655, HUM POS: 28, MIX: 92.
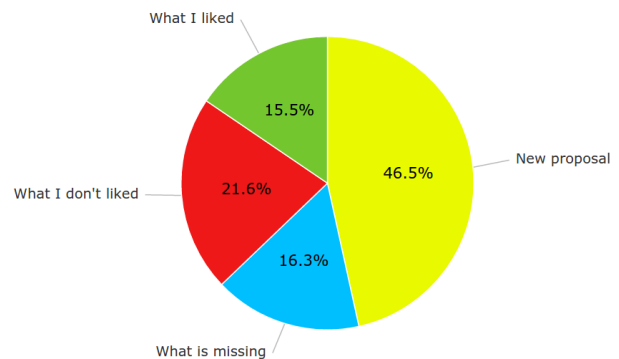
Figure 3: Distribution of categories in the WEB-BS corpus.

Notice that, when we consider polarized opinions on the reform, most of the them are negative, with a significative presence of messages exploiting irony as figurative devices to express criticisms with a sarcastic flavor. Furthermore, also the previous findings reported above about disagreement are confirmed. We observed indeed in Crowdflower's annotators, during the gold standard quality test, the same

---

[4]http://crowdflower.com

[5]The corpus will available at the following webpage http://www.di.unito.it/~tutreeb/corpora.html. We plan to release it by adhering to the Sentipolc 2014 format (Basile et al., 2014b), see guidelines in (Basile et al., 2014a).
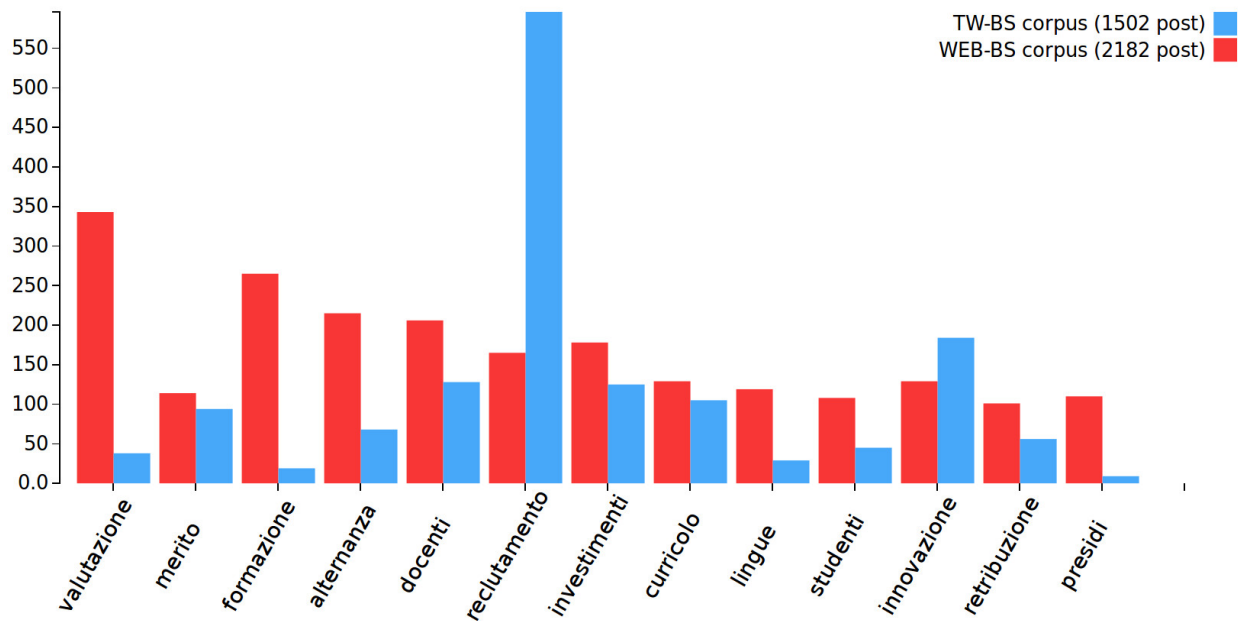
Figure 4: Comparison between the topic distribution in the two LBS corpora: TW-BS and WEB-BS.

difficulty in detection of irony and in agreeing about the exploitation of the label NONE.

Such results about the TW-BS corpus can be compared with the ones which emerged from the analysis of the texts collected in the WEB-BS corpus during the online consultation, which have been labeled by users of the focus groups by using the labels 'what I like' (642), 'what I don't like' (892), 'what is missing' (675), and 'new integration' (1920). Observing the graph in Figure 3, we can see that the category 'what I like', which refers to a positive stance about the reform, has been chosen for labeling the 15.5% of the posts, whereas the 'what I don't like' is the chosen label for the 21.6% of cases.

This shows that also in the online consultation we have a higher amounts of posts expressing a negative stance on the reform, but with different proportions w.r.t. Twitter. In TW-BS, indeed negative tweets are 3 times more than positive ones, while in WEB-BS they are just 1.5 more than the positive ones.

This can be explained by considering that Twitter and the online platform should be considered as two different social contexts, where the context is defined by the imagined audience of the online user, which influences his/her online behavior (Boyd, 2014). The impression is that, on the one hand, the participation on Twitter has been more critical, free from the constraints of the debate and, especially, individual. On the other hand, the web platform provided for the online consultation contributions collected from *groups* of individuals, which *collectively* took the responsibility to synthesize and publish the contents. Furthermore it seems that participants of the focus groups approached the consultation animated by a more constructive mood. Elements in favor of this hypothesis are also the high percentage of 'new integration' posts (about 63%) and the almost total absence of ironic contents in WEB-BS, as could be observed by a

qualitative analysis of the messages.

For what concerns instead the analysis of topic annotation, it shows two important findings. A comparison for all 13 topics proposed by the government for categorizing the online contributions to the debate on *laBuonaScuola* can be seen in Figure 4. On the one hand, the most used topic labels during the debate on the Web captured by WEB-BS were rarely used on Twitter. Indeed, among 5,573 tweets in agreement by topic, the 80% were labeled with the generic topic tag. On the other hand, the distribution of topics in the two corpora, TW-BS and WEB-BS is different, as can be observed in Figure 4. For instance, the *evaluation* topic is present in many posts of the WEB-BS corpus, whereas it is barely used in the TW-BS corpus. Conversely, the topic *recruitment* occurs frequently on Twitter, but it is little used in the WEB-BS corpus.

5. Conclusions

In this paper we presented an analysis of the Italian online debate on the school reform "Buona Scuola", through the creation of an annotated corpus of tweets, called TW-BS. We described the collection and annotation stages, and we provided an in-depth analysis of the disagreement. Future works will take into account the issues raised by the disagreement analysis, providing to annotators with more specific guidelines about irony detection, with examples taken from the TW-BS corpus. Furthermore, in order to reduce the vagueness of the NONE label, we will better specify to annotators its boundaries.

Finally, we are planning a further in-depth analysis of the TW-BS corpus and of the three-level annotation (sentiment/irony/topics) aimed at identifying the presence of sentiment and irony w.r.t. to specific aspects discussed in the debate. This is a very interesting analysis which is enabled by the annotation scheme we applied, and is in principle

applicable to any kind of political debate also in other languages. Indeed it allows to extract information that can be exploited in the context of finer-grained sentiment analysis tasks, which are raising a growing interest in the very last years, such as for instance *stance detection*, a new task proposed at Semeval 2016 (Task 6[6]), where participants are asked to detect the position of the tweeter (against/in favor/neutral) with respect to a given target that is usually a controversial issue.

Acknowledgments

References

Attardi, G., Basile, V., Bosco, C., Caselli, T., Dell'Orletta, F., Montemagni, S., Patti, V., Simi, M., and Sprugnoli, R. (2015). State of the art language technologies for italian: The EVALITA 2014 perspective. *Journal of Intelligenza Artificiale*, 9(1):43–61.

Basile, V. and Nissim, M. (2013). Sentiment analysis on italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 100–107, Atlanta, Georgia. Association for Computational Linguistics.

Basile, V., Bolioli, A., Nissim, M., Patti, V., and Rosso, P. (2014a). Evalita 2014 Sentipolc task: Task guidelines. Technical report.

Basile, V., Bolioli, A., Nissim, M., Patti, V., and Rosso, P. (2014b). 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)*, pages 50–57, Pisa, Italy. Pisa University Press.

Bosco, C., Patti, V., and Bolioli, A. (2013). Developing corpora for sentiment analysis: The case of irony and Senti–TUT. *IEEE Intelligent Systems*, 28(2):55–63.

Bosco, C., Allisio, L., Mussa, V., Patti, V., Ruffo, G., Sanguinetti, M., and Sulis, E. (2014). Detecting happiness in italian tweets: Towards an evaluation dataset for sentiment analysis in Felicittà. In *Proc. of the 5th International Workshop on Emotion, Social Signals, Sentiment and Linked Opena Data, ESSSLOD 2014*, pages 56–63, Reykjavik, Iceland. ELRA.

Bosco, C., Patti, V., and Bolioli, A. (2015a). Developing corpora for sentiment analysis: The case of irony and senti-tut (extended abstract). In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 4158–4162. AAAI Press / International Joint Conferences on Artificial Intelligence.

Bosco, C., Patti, V., Lai, M., and Virone, D. (2015b). Building a corpus on a debate on political reform in Twitter. In *Proceedings of Second Italian Conference on Computational Linguistics, CLIC-it 2015*, pages 171–176, Trento, Italy. Accademia University Press.

Boyd, D. (2014). *It's Complicated: the social lives of networked teens*. Yale University Press.

Ceron, A., Curini, L., and Stefano, I. M. (2014). *Social Media e Sentiment Analysis: l'evoluzione dei fenomeni sociali attraverso la rete*. Springer.

Conoscenti, M. (2011). *The Reframer: An Analysis of Barack Obama Political Discourse (2004-2010)*. Bulzoni Editore.

Conover, M., Gonçalves, B., and Ratkiewicz, J. (2011). Predicting the political alignment of twitter users. In *Proceeding of the IEEE Third Inernational Conference on Social Computing (SocialCom)*, pages 192–199, Los Angeles, CA, USA. Academy of Science and Engineering.

Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., and Reyes, A. (2015). Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 470–478, Denver, Colorado, June. Association for Computational Linguistics.

Gloor, P., Krauss, J., Nann, S., Fischbach, K., and Schoder, D. (2009). Web science 2.0: Identifying trends through semantic social network analysis. In *Proceedings of IEEE International Conference on Computational Science and Engineering (CSE-09)*, volume 4, pages 215–222, Vancouver, Canada. IEEE Computer Society.

Karoui, J., Farah, B., Moriceau, V., Aussenac-Gilles, N., and Hadrich-Belguith, L. (2015). Towards a contextual pragmatic model to detect irony in tweets. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 644–650, Beijing, China, July. Association for Computational Linguistics.

Lai, M., Virone, D., Bosco, C., and Patti, V. (2015). Debate on political reforms in Twitter: A hashtag-driven analysis of political polarization. In *Proc. of 2015 IEEE International Conference on Data Science and Advanced Analytics (IEEE DSAA'2015), Special Track on Emotion and Sentiment in Intelligent Systems and Big Social Data Analysis.*, pages 1–9, Paris, France. IEEE.

Maynard, D. and Funk, A. (2011). Automatic detection of political opinions in tweets. In *Extended Semantic Web Conference Workshop*, pages 88–99.

Maynard, D. and Greenwood, M. (2014). Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In *Proceedings of the Ninth International Conference on Language Resources and Evalua-*

---

[6]http://alt.qcri.org/semeval2016/task6/

tion (LREC'14), Reykjavik, Iceland, may. ELRA.

Rajadesingan, A. and Liu, H. (2014). Identifying users with opposing opinions in twitter debates. In William G. Kennedy, et al., editors, *Social Computing, Behavioral-Cultural Modeling and Prediction*, volume 8393 of *Lecture Notes in Computer Science*, pages 153–160. Springer International Publishing.

Reyes, A. and Rosso, P. (2014). On the difficulty of automatically detecting irony: beyond a simple case of negation. *Knowledge and Information Systems*, 40(3):595–614.

Reyes, A., Rosso, P., and Veale, T. (2013). A multidimensional approach for detecting irony in twitter. *Language Resources and Evaluation*, 47(1):239–268.

Sang, E. T. K. and Bos, J. (2012). Predicting the 2011 dutch senate election results with twitter. In *Proceedings of the Workshop on Semantic Analysis in Social Media*, pages 53–60, Stroudsburg, PA, USA. Association for Computational Linguistics.

Sperber, D. and Wilson, D. (1986). *Relevance: communication and cognition*. Basil Blackwell.

Stranisci, M., Bosco, C., Viviana, P., and Hernández Farias, D. I. (2015). Analyzing and annotating for sentiment analysis the socio-political debate on "La Buona Scuola". In *Proceedings of the 2th Italian Conference on Computational Linguistics (CLiC-IT 2015)*, pages 274–279, Trento, Italy. Accademia University Press. In Press.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2011). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the ICWSM-11*, pages 178–185, Barcelona, Spain.

Wilson, D. (2006). The pragmatics of verbal irony: Echo or pretence? *Lingua*, 116(10):1722–1743.