Tweeting and being ironic in the debate about a political reform: the French annotated corpus TWitter-MariagePourTous

Cristina Bosco[1], Mirko Lai[1], Viviana Patti[1], Daniela Virone[2]
[1]Dipartimento di Informatica – Università degli Studi di Torino,
Corso Svizzera 185 – 10149 Torino (Italy),
[2]Dipartimento di Studi Umanistici – Università degli Studi di Torino
Via Verdi 10 – 10124 Torino (Italy)
{bosco,lai,patti}@di.unito.it, dvirone@unito.it

Abstract

The paper introduces a new annotated French data set for Sentiment Analysis, which is a currently missing resource. It focuses on the collection from Twitter of data related to the socio-political debate about the reform of the bill for wedding in France. The design of the annotation scheme is described, which extends a polarity label set by making available tags for marking target semantic areas and figurative language devices. The annotation process is presented and the disagreement discussed, in particular, in the perspective of figurative language use and in that of the semantic oriented annotation, which are open challenges for NLP systems.

Keywords: sentiment analysis, French, political debates, Twitter, figurative language

1. Introduction

Texts generated by users within the context of social media can be a great opportunity for moving onward the development of corpus-based techniques for Sentiment Analysis and Opinion Mining (SA&OM). But their annotation raises several challenging issues. Even more attention than in classical tasks must indeed be paid to the development of such kind of data sets.

First of all, it should be observed that we are still far from sharing standardized annotation schemes and tag sets, and the big effort devoted to the development of a resource may also not result in a proportional advancement for the area e.g. if is not enough reusable of comparable with others. A suitable solution is adopting schemes which includes comparable annotations, which can be matched in particular with those exploited in evaluation campaigns (see e.g. SemEval[1], or Evalita[2]), but also, considering the relative recentness of the area, enriching the annotation with new tags for describing aspects that potentially have impact on the results of sentiment engines, even if not previously annotated.

Second, the topics of texts of interest for SA&OM can be several, and we must take into account also the coverage of domains when we design a resource. The same problem of coverage can be related to languages, since several resources are available for a few languages when they are almost missing at all for several others.

Finally, a variety of phenomena features the language of social media, e.g. figurative devices, which can worst the performance of SA systems (Ghosh et al., 2015b); by contrast, the availability of data where they are annotated can positively influence results, but it is very difficult to evaluate the suitability of this annotation task.

In this paper, we present a new annotated corpus for French SA&OM called TWitter-MariagePourTous (henceforth TW-MPT), collecting texts from Twitter about the debate in France on the homosexual wedding and the related reform.

This resource has been developed within the context of a project aiming at investigating the communicative behaviors featured by political debates, and at shedding some light on the way communities of users with different roles in the society and different political sentiment interact. The contribute of the resource and of the analyses here presented must be therefore considered within the frame of this larger project for studying in a multilingual perspective political debates in social media. This wider project indeed currently includes a data set for Italian, collecting texts from the debate about the school reform (Stranisci et al., 2015; Stranisci et al., 2016). Furthermore, a corpus for Spanish and Catalan is under development, with a specific focus on the debate about the separatism of Catalonia, a controversial issue which is raising a growing interest in Spain the last months. The same annotation schema here proposed has been adopted for manually annotating all these resources, and the same set of analyses has been applied in order to develop a framework for the comparison of results on political debates in Twitter in different languages.

The novelty of the resource presented in this paper consists in being both a currently missing resource for French, which is a under resourced language under this respect, and an extension of political texts (Conover et al., 2011a; Li et al., 2012; Conover et al., 2011b; Skilters et al., 2011) towards the field of discussions about controversial topics. In this last sense, the corpus can be of some interest for training systems in stance detection, i.e. the task of automatically determining from text whether the author is in favor, against or neutral with respect to a given target when the topic is controversial, which is currently considered as a crucial issue for sentiment analysis systems

[1]http://aclweb.org/aclwiki/index.php?title=SemEval_Portal
[2]http://www.evalita.it/

(see e.g. the novel Semeval 2016's Task about *Detecting Stance in Twitter* within the Sentiment analysis Track[3]).
In order to improve the reusability and portability of our effort, we applied in the data set an annotation which is compliant with that exploited in the SENTIment POLarity Classification (SENTIPOLC) shared task held within the evaluation campaign for Italian natural language processing (see at `www.evalita.it`), which is consistent but extends the annotation proposed for English in the context of the *Sentiment Analysis in Twitter* shared task, which has been periodically re-proposed in the last years (Rosenthal et al., 2015). Moreover we also extended the annotation in two main directions, describing data in the perspective of irony and metaphor, and giving tags for classifying them according to a few semantic lines emerging in the debate. On the one hand, we approached the task of the annotation of figurative devices in order to evaluate in a real data set the suitability of a task known as very hard (Reyes and Rosso, 2014; Filatova, 2012; Reyes et al., 2013; Maynard and Greenwood, 2014). On the other hand, we propose a semantic-oriented annotation assuming that it can give more precise hints about the conversational context, considering that often the meaning of a text varies according to the topic and, in the case of political debates, also according to the specific aspects discussed or to the author (e.g. a politician joined to a particular party). This latter feature makes our annotation scheme more adequate for dealing with the contents of the debate and for detecting the communicative strategies involved in it.

The paper is organized as follows. The next section briefly surveys the related work, while the following describes the collection of the data set and the methodology applied for defining the annotation scheme. The fourth section discusses instead the application of the scheme on the collected data set, focusing in particular on the analysis of disagreement related to the annotation of irony and that of semantic areas.

2. Related work

Several works focused in the last few years on political domain as debated in various ways within the variety of social media. For what concerns Twitter, the analysis of political debates has been often related to election campaigns and became quite popular (Mohammad et al., 2015; Sang and Bos, 2012; Tumasjan et al., 2011; Bermingham and Smeaton, 2011), since there is a lot of interest in developing tools to automatically gauge the political sentiment in order to predict the election outcome. Some works focused also on aspects concerning the political polarization in this social medium (Conover et al., 2011b; Skilters et al., 2011), while others addressed the issues related to the arguments accompanying the political messages. In (Eensoo and Valette, 2014), an analysis devoted to discover in tweets the argumentation related to evaluative discourse applied to the racism anti-Rom in the Web, it is shown that a discourse where a form of evaluation is expressed does not necessarily exploits semantic and linguistic markers traditionally linked to the evaluation, but it can be also based on dialogical and dialectical components.
Also online debates are a large source of informal and opinion-sharing dialogue on current socio-political issues, and several works rely on sentiment analysis techniques (Pang and Lee, 2008) to analyze politics (Tumasjan et al., 2011; Li et al., 2012; He et al., 2012). Among these works some is dedicated in particular to the classification of users' stance, i.e. the detection of positions pro or con topics that users assume within debates, applied to data from Web sites ranging from forums to other dedicated platforms like `convinceme.net`. In these studies, the interest is usually focused on dual-sided debates (Somasundaran and Wiebe, 2010; Sridhar et al., 2014) where two possible polarizing sides can be taken by participants, and on expressed forms of subjectivity that can be the signal of stance, like e.g. arguing (since supporting their side, people not only express their sentiment but also argue about what is true). On this line, also the social news website Reddit[4] has been recently taken as an object of study (Wallace et al., 2015) since it comprises many, often polarized, user communities – called *subreddits* – centered around specific topics of interest, which constitute a natural source of data for the analysis of debates on controversial issues.
As observed in (Ranade et al., 2013), online debates differ from public debates because participants assert their opinion towards either side something ignoring discourse coherence and generally using strong degree of sentiment words including insulting or sarcastic remarks for greater emphasis of their point. The frequent exploitation of figurative language devices in social media and website like Reddit (Wallace et al., 2015), especially in the political domain, is described in several papers, among which (Maynard and Greenwood, 2014; Bosco et al., 2013; Reyes et al., 2012; Reyes et al., 2013; Davidov et al., 2011), and has been addressed in the Semeval 2015 shared task 11 on *Sentiment Analysis of Figurative Language in Twitter* (Ghosh et al., 2015a). Dealing properly with the presence of ironic devices is indeed crucial when the goal is to analyze the polarity of the opinions on a topic of interest, and in political domains irony is very often used in conjunction with a seemingly positive statement, to reflect a negative one, due to a phenomenon known in literature as *polarity reversal* (Bosco et al., 2013). Therefore, the issue has to be tackled in order to correctly label the polarity of an opinion, going beyond the literal meaning of the posts.
Most of the works carried on so far in this area focus their analysis on English datasets and rely on the use of lexical and affective resources which are available only for English, while French is currently considered among the under-resourced languages. Nevertheless, in the last few years some effort has been devoted to the development of new annotated data to be exploited in this area, see e.g. (Fraisse and Paroubek, 2014a; Fraisse and Paroubek, 2014b; Bestgen, 2008) for French.
The work on the new French corpus presented here aims at investigating the presence of sentiment and ironic devices, but our interest goes beyond that, since we are interested in the encoding of a more fine-grained knowledge which is

[3] `http://alt.qcri.org/semeval2016/task6/`

[4] `http://www.reddit.com`

related to more specific targets within the debates showing the relationship linking these targets and the opinions about them.

3. Collection and definition of the annotation scheme

As usual for drawing attention to topics that vary from commercial products to political elections (see e.g. (Bollen et al., 2011; Buscaldi and Irazú, 2015; Hu and Liu, 2004; Sang and Bos, 2012)) some new hashtag has been used by French Twitter users for making widely known information and opinions about the bill on the homosexual wedding (Cunha et al., 2011).
Since Twitter does not offer specific functions for creating or registering hashtags, they only assume the status of language artifacts. Nevertheless, taking a diachronic perspective, we can observe in social media texts, on the one hand, the birth and life of hashtags, and, on the other hand, the behavior of users that exploit them. In this sense we can recognize a sort of creation of several hashtags which is followed by a negotiation time where some is accepted, and then extensively used by the community of users (as attested by a frequency analysis), while others perish.
Among the hashtags proposed with the debate about the wedding bill reform in France, *#mariagepourtous* has been accepted within the dialogical and social context growing around the topic, and largely exploited, in order to assume opinions about the content of the hashtag itself.
We exploited therefore the hashtag *#mariagepourtous* as a keyword for filtering data to be included in the TW-MPT corpus. It allowed the selection of 254,366 original messages. Thanks to the association of each message with the metadata related to the author and posting time, and in order to better understand the conversational context growing around the debate, we performed a set of analysis described in (Lai et al., 2015) and in (Bosco et al., 2015b). These metadata can also be usefully exploited in association with the annotation we are here describing.

The annotation scheme encompasses three different kinds of knowledge, that is the polarity, the target semantic area of the message, and the exploitation of figurative devices. All them may give a significant contribute to the detection of meaning, especially for what concerns the content of posts related to sentiments and opinions, but they must be considered as interrelated. For instance, regardless of the affective words exploited, the polarity of a post cannot be reliable determined without taking into account the possible presence of irony, which can reverse the polarity as in the following example:

'@MGrossiord le #mariagepourtous est validé, le monde est sauvé, plus de discrimination ni de chômage, enfin la paix dans le mon...'
(@MGrossiord the #mariagepourtous is validated, the world has been saved, no more discrimination neither strike, finally the peace in the world ...)

3.1. Annotation of polarity

As far as the design of the annotation scheme is involved, we applied the annotation exploited in (Gianti et al., 2012; Bosco et al., 2013; Bosco et al., 2014; Bosco et al., 2015a) for marking the polarity of opinions and sentiments, extended with the labels UN and RP for marking

label	polarity
POS	positive
NEG	negative
NONE	neutral
MIXED	both positive and negative
UN	unintelligible content
RP	repetition of a post

Table 1: Polarity tags annotated in the TW-MPT corpus.

unintelligible and repeated content respectively. The following examples show how the labels, presented in Table 1 have been used:

- POS: *#mariagepourtous est une manifestation historique pour la France dans l'histoire de l'égalité! Participez à cet événement historique!* (#mariagepourtous is a historical manifestation for France into the hystory of equality! Participate to this historical event!)
- NEG: *Le projet #mariagepourtous est l'arbre qui cache la forêt! Déconstruction de la personne humaine ... #manifpourtous* (The #mariagepourtous project is the tree which hide the wood! Deconstruction of the human being ... #manifpourtous)
- NONE: *Le débat pour le projet de loi du #Mariagepourtous est actuellement focalisé sur la #PMA et la #GPA... #DirectAN* (The debate on #mariagepourtous bill project is actually focusing #PMA and #GPA ... #DirectAN)
- MIXED: *Autant je ne comprends pas pourquoi le #MariagePourTous est critiqué, autant l'#adoption est une autre chose vraiment discutable.* (One the one hand I don't understand why #mariagepourtous is so criticized, on the other hand #adoption is another thing, hardly disputable).

3.2. Annotation of figurative language

We included in our schema also tags for marking figurative language devices, i.e. irony and metaphor (see table 2). HUMPOS is the label we used for marking the presence of irony featured by positive polarity, HUMNEG for negative irony, and a yes/no feature is used for METAPHOR.
Also other figurative devices (e.g. hyperbole) may be of interest for sentiment analysis, but the extension of the schema in this direction will be object of future work, and also the annotation of metaphor in this corpus can be considered as a preliminary issue because of the detected complexity of this phenomenon. See the following examples for the exploitation of the figurative devices' tags presented in Table 2:

Figure 1: A cloud-style representation of words distribution in the TW-MPT dataset. It includes four sections respectively showing (from left to right and from top to bottom) the more used words for the following semantic areas: family, socio-political debate, legal aspects, public manifestations.

label	figurative device
HUMPOS	positive irony
HUMNEG	negative irony
METAPH	metaphorical expression

Table 2: Tags annotated in the TW-MPT corpus for figurative language uses.

- HUMPOS: *#lesavieztu: Le #mariagepourtous est responsable de la crise financière et du réchauffement globale.* (The #mariagepourtous is responsible for the financial crisis and for the global warming.)
- HUMNEG: *Comme la loi sur #mariagepourtous est officielle un vendredi, on peut donc dire que "le vendredi c'est sodomie"?* (Since the bill about #mariagepourtous is official on Friday, we can say that "Friday it's sodomy"?)
- METAPH: *Le #mariagepourtous est passé car la France est toujour sous anesthésie TiboRimo Touchepasalafam. Le réveil va être brutal en septembre.* (The #mariagepourtous is passed since France is always under anesthesia TiboRimo Touchepasalafam. The awakening will be rude in September.)

3.3. Annotation of aspects discussed in the debate

Observing the corpus and the other data collected within the project, we hypothesized that the debate developed around a few particular topic. We classified words occurring in the TW-MPT corpus according to their frequency, and we extracted four distinguishable semantic areas by applying a tag cloud extraction and we determined for each of them a label to be annotated: *family* (we labeled as FAMILLE), *legal aspects* (we labeled as LOI), *public manifestations* (we labeled as MANIF), *socio-political debate* (we labeled as DEBAT). The tag cloud can be seen in Figure 1. Notice that the labeling of each area has been done by using the more frequent and representative word of the cloud area itself. In particular, in the upper left side of the cloud, that labeled as family, we can see *mariage* (marriage), *enfants* (children), *parents* (parents) and *pére* (father); *loi* (law), *égalité* (equality) and *droits* (rights) feature instead the lower left areas, which has been labeled as legal aspects. Among the more frequent words for the area labeled as public manifestation, i.e. the lower right one, there are *manif* and *manifestation* (manifestation); and finally the area labeled as socio-political debate includes words as *debat* (debate), *deputés* (deputies) and *opposants* (opponents).

4. Annotation and discussion of disagreement

In an analysis of the full corpus, we observed that, by using the verb *être* (to be) in the left-side context of the hashtag, users tend to express a kind of rough polarization about the reform, while when they use it in the right-side context of the hashtag they usually express a more articulated evaluation or definition of the hashtag content. We selected therefore all the messages featured by the co-occurrence in the right or left side of the verb *être* (to be) with the hashtag driving data collection, *#mariagepourtous*, and we obtained the corpus to be annotated, composed of 2,872 posts.

The annotation process involved two human annotators,

both skilled in French language and graduated in linguistics, that annotated both all the tweets of our collection, thus producing a pair of annotation for each tweet.
Following our previous experiences and the examples of other projects devoted to the development of annotated linguistic data sets, first of all, we developed a set of guidelines for the annotation, then we discussed them with the annotators, after a first stage of annotation when they applied the tags to a small portion of data (around 100 tweets). An upgraded version of the guidelines, which is the result of the discussion among the annotators and the designers of the schema, have been used for the annotation of the rest of the corpus.
The detected inter-annotator agreement at this stage was $\kappa = 0.612$[5], but in order to solve the disagreement and to further validate the annotation done, we applied a third annotation via Crowdflower – a crowdsourcing platform for manual annotation which has been recently widely used for accomplishing annotation tasks, see e.g. (Ghosh et al., 2015b). This latter stage of the annotation process allowed the reduction of the detected disagreement and the improvement of the reliability of the data set for the release of the corpus[6].

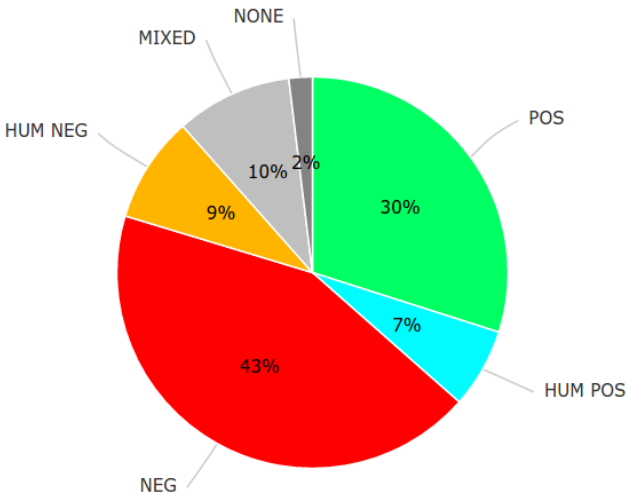

Figure 2: The distribution of polarity tags in the 1,958 agreed annotated posts: 43% NEG, 30% POS, 10% MIXED, 9% HUM NEG, 7 % HUM POS and 2 % NONE.

We think that some main trend can be observed and some interesting lesson learned from the data. Because of the multi-faceted annotation applied in the corpus, the significance of the detected disagreement can best be grasped by looking at the data according to three main directions of the annotation separately, namely polarity, irony and semantic areas.
For what concerns polarity, we found differing annotations in about 30% of cases after the application of the the third annotation, but a real disagreement (when the annotators select opposite polarity POS and NEG) is indeed very low since it consists in 121 cases (4.2%) only. Similar degrees of disagreement we detected also for labels like MIXED or NONE, whose classification can be less sharp and harder for annotators: one annotator selected a defined polarity and the other used the label NONE (neutral polarity) in 137 posts or the label MIXED (positive and negative polarity both) in 86.
The rest of the disagreement observed is related to irony, i.e. the exploitation of the labels HUMNEG and HUMPOS. Nevertheless, there is a limited amount of hard disagreement, i.e. 56 cases where both the annotators detected irony but marking different polarity, while in 184 posts one annotator marked irony while the other doesn't, but them both detected the same polarity POS or NEG. A confirmation of the fact that the annotators are not biased with respect to irony, and of the inherent difficulty of the task, derived from the application of the third annotation cited above. A particular feature of this corpus that we observed is the higher frequency of positive irony with respect to what we can observe in other corpora on political debate, e.g. (Stranisci et al., 2016), where the use of negative irony is usually highly predominant. This specific characteristics of the TW-MPT corpus can be motivated by the topic involved or by the features of the particular users' community involved in the debate, and it is surely an aspect that deserves some further investigation as it can shed some light on a type of irony which is less common and less investigated than bitter sarcasm.
All the 2,872 tweets are annotated with semantic areas, but their annotations are different in 914 cases, while coincide in the remaining 1,958, whose distribution is represented in figure 3. The disagreement is mainly referred to the label *loi* (536), followed by *debat* (172), *famille* (158) and *manif* (47). Putting together all these observations, we can see that the debate has been mainly oriented to the topics summarized by the label *debat*, and on it also the higher agreement has been achieved. More difficult has been the detection of topics related to the label *loi*, which has been annotated in around the same times in agreement and in disagreement. But the proportionally higher disagreement must be referred to the *famille*'s topics.
We conclude with some preliminary observation about the annotation of metaphor. Metaphor has been found in only 52 posts by the first annotator, while the other does the same in 109, showing the limited use of this device in the corpus. The difficulty and subjectivity involved in this phenomenon is instead confirmed by the low agreement achieved by annotators: only 31 are the cases of agreement and they are simple very common instances of this figurative device like the following:

Le #MariagePourTous est en train de devenir la "patate chaude" dont le gouvernement voudrait vite se débarrasser avant qu'elle explose!

The #MariagePourTous is fast becoming the "hot potato" that the

[5]The value is calculated by considering the labels related to polarity and irony: POS, NEG, NONE, MIXED, HUMPOS, HUMNEG.

[6]The corpus will be made available on the following webpage: `www.di.unito.it/~tutreeb/corpora.html`, together with the annotation guidelines mentioned above.

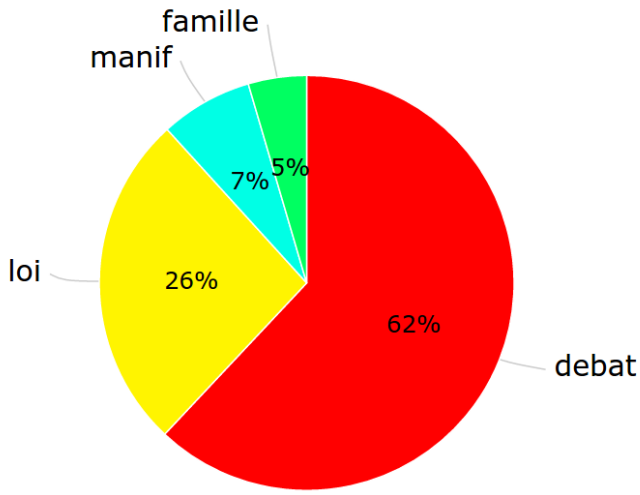

Figure 3: The distribution of semantic areas in the 1,958 agreed annotated posts: 62% *debat*, 26% *loi*, 7% *manif* and 5% *famille*.

government would quickly get rid before it explodes!).

A planned application of a further annotation of this aspect on this same data set can give us more information about the phenomenon. Nevertheless only an extension of our corpus with the collection of a larger amount of examples of metaphorical expressions can result in a more reliable analysis of this kind of figurative device and in a more useful annotated data set.

5. Conclusions

The paper presents a French annotated corpus for Sentiment Analysis, a currently missing resource.

The novelty of the resource consists in the application of a multi-faceted annotation scheme, which includes not only classical tags for polarity but also those for marking figurative uses of language and semantic target areas. The development of this corpus is part of a wider project for studying communicative strategies acting within political debates which strongly impact on citizens, like the debate about the school reform in Italy and the one about the separatism of Catalonia in Spain.

The resource here presented can be therefore of some interest for training systems e.g. in stance detection and other tasks which apply on political debates, but also for shedding some light on the way community of users with different roles in the society, different political sentiments and speaking different languages interact.

Finally, another feature of political debates that could be interesting to explore in future work is related to arguments exploited by users in order to support their positions about a controversial issue. On this line, we plan to add a further layer of annotation related to this aspect on top of the sentiment and topic layers, in order to explore possible fruitful relationships among sentiment and argument related information.

Acknowledgments

The authors thank all the persons who supported the work, and in particular Federica Ramires that meaningfully contributed to the annotation and analysis of the corpus as part of her Bachelor's degree thesis. The National Council for Science and Technology (CONACyT Mexico) has funded the research work of Delia Irazú Hernández Farias (Grant No. 218109/313683 CVU-369616). The work of Viviana Patti was partially carried out at the Universitat Politècnica de València within the framework of a fellowship of the University of Turin co-funded by Fondazione CRT (World Wide Style Program 2).

References

Bermingham, A. and Smeaton, A. F. (2011). On using Twitter to monitor political sentiment and predict election results. In *Sentiment Analysis where AI meets Psychology (SAAIP), Workshop at the International Joint Conference for Natural Language Processing (IJCNLP)*.

Bestgen, Y. (2008). Building affective lexicons from specific corpora for automatic sentiment analysis. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 496–500, Marrakech, Morocco. European Language Resources Association (ELRA).

Bollen, J., Huina, M., and Xiaojun, Z. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

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 B. Schuller, et al., editors, *Proceedings 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 IJCAI 2015*, pages 4188–4193.

Bosco, C., Patti, V., Lai, M., and Virone, D. (2015b). Building a corpus on a debate on political reform in twitter. In *Proceedings of CLIC-2015*, pages 171–176. Accademia University Press.

Buscaldi, D. and Irazú, H.-F. (2015). Sentiment analysis on microblogs for natural disasters management: A study on the 2014 genoa floodings. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 1185–1188, Geneva, Swiss. International World Wide Web Conferences Committee.

Conover, M., Gonçalves, B., and Ratkiewicz, J. (2011a). 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.

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., and Menczer, F. (2011b). Political polarization on twitter. In *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Cunha, E., Magno, G., Comarela, G., Almeida, V., Goncalves, M. A., and Benevenuto, F. (2011). Analyzing the dynamic evolution of hashtags on twitter: a language-based approach. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, pages 58–65, Portland, Oregon. Association for Computational Linguistics.

Davidov, D., Tsur, O., and Rappoport, A. (2011). Semi-supervised recognition of sarcastic sentences in Twitter and Amazon. In *Proceedings of the CONLL'11*, pages 107–116, Portland, Oregon (USA).

Eensoo, E. and Valette, M. (2014). Approche textuelle pour le traitement automatique du discours evaluatif. *Langue francaise*, (4):109–124.

Filatova, E. (2012). Irony and sarcasm: Corpus generation and analysis using crowdsourcing. In *Proceedings of the LREC'12*, pages 392–398, Istanbul, Turkey.

Fraisse, A. and Paroubek, P. (2014a). Toward a unifying model for opinion, sentiment and emotion information extraction. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3881–3886, Reykjavik, Iceland. European Language Resources Association (ELRA).

Fraisse, A. and Paroubek, P. (2014b). Twitter as a comparable corpus to build multilingual affective lexicons. In *Proceedings of the LREC'14 Workshop on Building and Using Comparable Corpora*, pages 17–21, Reykjavik, Iceland. European Language Resources Association (ELRA).

Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Reyes, A., and Barnden, J. (2015a). Semeval-2015 task 11: Sentiment analysis of figurative language in Twitter. In *Proc. Int. Workshop on Semantic Evaluation (SemEval-2015), Co-located with NAACL and *SEM*.

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

Gianti, A., Bosco, C., Patti, V., Bolioli, A., and Caro, L. D. (2012). Annotating irony in a novel italian corpus for sentiment analysis. In *Proceedings of the 4th Workshop ES3*, pages 1–7, Istanbul, Turkey.

He, Y., Saif, H., Wei, Z., and Wong, K.-F. (2012). Quantising opinions for political tweets analysis. In *Proceedings of the LREC'12*, pages 3901–3906, Istanbul, Turkey.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD.

Lai, M., Virone, D., Bosco, C., and Patti, V. (2015). Debate on political reforms in Twitter: A hashtag-driven analysis of political polarization. In *Proceedings of 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.

Li, H., Cheng, X., Adson, K., Kirshboim, T., and Xu, F. (2012). Annotating opinions in German political news. In *Proceedings of the LREC'12*, pages 1183–1188, Istanbul, Turkey.

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 Evaluation (LREC'14)*, Reykjavik, Iceland, may. ELRA.

Mohammad, S. M., Zhu, X., Kiritchenko, S., and Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*, 51(4):480 – 499.

Pang, B. and Lee, L. (2008). *Opinion Mining and Sentiment Analysis (Foundations and Trends(R) in Information Retrieval)*. Now Publishers Inc.

Ranade, S., Sangal, R., and Mamidi, R. (2013). Stance classification in online debates by recognizing users' intentions. In *Proceedings of the SIGDIAL 2013 Conference*, pages 61–69.

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 Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data Knowledge Engineering*, 74:1–12.

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

Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S., Ritter, A., and Stoyanov, V. (2015). Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 451–463, Denver, Colorado, June. Association for Computational Linguistics.

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.

Skilters, J., Kreile, M., Bojars, U., Brikse, I., Pencis, J., and Uzule, L. (2011). The pragmatics of political messages in twitter communication. In Raul Garcia-Castro, et al., editors, *ESWC Workshops*, volume 7117 of *Lecture Notes in Computer Science*, pages 100–111. Springer.

Somasundaran, S. and Wiebe, J. (2010). Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124.

Sridhar, D., Getoor, L., and Walker, M. (2014). Collective stance classification of posts in online debate forums. In *Proceedings of the ACL Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, pages 109–117.

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.

Stranisci, M., Bosco, C., Viviana, P., and Hernández Farias, D. I. (2016). Annotating Sentiment and Irony in the Online Italian Political Debate on #labuonascuola. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC 2016)*, Portoroz (Slovenia). ELRA. 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.

Wallace, B. C., Choe, D. K., and Charniak, E. (2015). Sparse, contextually informed models for irony detection: Exploiting user communities, entities and sentiment. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 1035–1044.