

# Sentiment analysis on Italian tweets

**Valerio Basile**

University of Groningen  
v.basile@rug.nl

**Malvina Nissim**

University of Bologna  
malvina.nissim@unibo.it

## Abstract

We describe TWITA, the first corpus of Italian tweets, which is created via a completely automatic procedure, portable to any other language. We experiment with sentiment analysis on two datasets of tweets from TWITA: a generic collection and a topic-specific collection. The only resource we use is a polarity lexicon, which we obtain by automatically matching three existing resources thereby creating the first polarity database for Italian. We observe that albeit shallow, our simple system captures polarity distinctions matching reasonably well the classification done by human judges, with differences in performance across polarity values and on the two sets.

## 1 Introduction

Twitter is an online service which lets subscribers post short messages (“tweets”) of up to 140 characters about anything, from good-morning messages to political stands.

Such micro texts are a precious mine for grasping opinions of groups of people, possibly about a specific topic or product. This is even more so, since tweets are associated to several kinds of meta-data, such as geographical coordinates of where the tweet was sent from, the id of the sender, the time of the day — information that can be combined with text analysis to yield an even more accurate picture of who says what, and where, and when. The last years have seen an enormous increase in research on developing opinion mining systems of various sorts applying Natural Language Processing techniques.

Systems range from simple lookups in polarity or affection resources, i.e. databases where a polarity score (usually positive, negative, or neutral) is associated to terms, to more sophisticated models built through supervised, unsupervised, and distant learning involving various sets of features (Liu, 2012).

Tweets are produced in many languages, but most work on sentiment analysis is done for English (even independently of Twitter). This is also due to the availability of tools and resources. Developing systems able to perform sentiment analysis for tweets in a new language requires at least a corpus of tweets and a polarity lexicon, both of which, to the best of our knowledge, do not exist yet for Italian.

This paper offers three main contributions in this respect. First, we present the first of corpus of tweets for Italian, built in such a way that makes it possible to use the exact same strategy to build similar resources for other languages without any manual intervention (Section 2). Second, we derive a polarity lexicon for Italian, organised by senses, also using a fully automatic strategy which can be replicated to obtain such a resource for other languages (Section 3.1). Third, we use the lexicon to automatically assign polarity to two subsets of the tweets in our corpus, and evaluate results against manually annotated data (Sections 3.2–3.4).

## 2 Corpus creation

We collected one year worth of tweets, from February 2012 to February 2013, using the Twitter filter API<sup>1</sup> and a language recognition strategy which

<sup>1</sup><https://dev.twitter.com/docs/api/1/post/statuses/filter>

we describe below. The collection, named TWITA, consists of about 100 million tweets in Italian enriched with several kinds of meta-information, such as the time-stamp, geographic coordinates (when present), and the username of the twitter. Additionally, we used off-the-shelf language processing tools to tokenise all tweets and tag them with part-of-speech information.

## 2.1 Language detection

One rather straightforward way of creating a corpus of language-specific tweets is to retrieve tweets via the Twitter API which are matched with strongly language-representative words. Tjong Kim Sang and Bos (2012) compile their list of highly typical Dutch terms manually to retrieve Dutch-only tweets. While we also use a list of strongly representative Italian words, we obtain such list *automatically*. This has the advantage of making the procedure more objective and fully portable to any other language for which large reference corpora are available. Indeed, we relied on frequency information derived from ItWac, a large corpus of Italian (Baroni et al., 2009), and exploited Google n-grams to rule out cross-language homographs. For boosting precision, we also used the publicly available language recognition software *langid.py* (Lui and Baldwin, 2012). The details of the procedure are given below:

1. extract the 1.000 most frequent lemmas from ItWac;
2. extract tweets matched by the selected representative words and detect the language using a freely available software;<sup>2</sup>
3. filter out the terms in the original list which have high frequency in a conflicting language. Frequency is obtained from Google N-grams;
4. use high frequency terms in the resulting cleaner list to search the Twitter API.

The 20 top terms which were then used to match Italian-only tweets are: *vita Roma forza alla quanto amore Milano Italia fare grazie della anche periodo bene scuola dopo tutto ancora tutti fatto*. In the

<sup>2</sup>Doing so, we identify other languages that share character sequences with Italian. The large majority of tweets in the first search were identified as Portuguese, followed by English, Spanish and then Italian.

extraction, we preserved metadata about user, time, and geographical coordinates whenever available.

Both precision and recall of this method are hard to assess. We cannot know how many tweets that are in fact Italian we're actually missing, but the amount of data we can in any case collect is so high that the issue is not so relevant.<sup>3</sup> Precision is more important, but manual checking would be too time-consuming. We inspected a subset of 1,000 tweets and registered a precision of 99.7% (three very short tweets were found to be in Spanish). Considering that roughly 2.5% of the tweets also include the geographical coordinates of the device used to send the message, we assessed an approximate precision indirectly. We plotted a one million tweets randomly chosen from our corpus and obtained the map shown in Figure 1 (the map is clipped to the Europe area for better identifiability). We can see that Italy is clearly outlined, indicating that precision, though not quantifiable, is likely to be satisfactory.

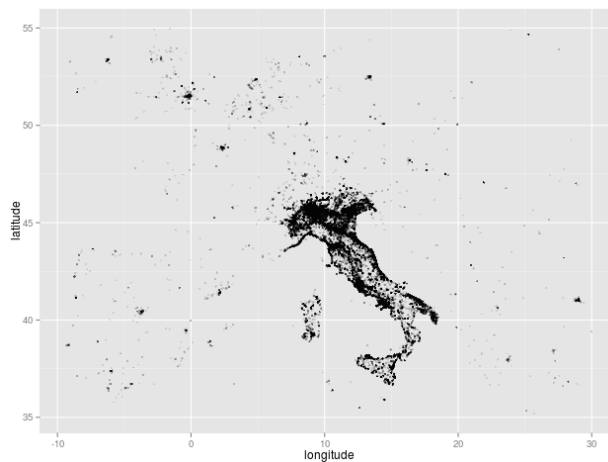


Figure 1: Map derived by plotting geo-coordinates of tweets obtained via our language-detection procedure.

## 2.2 Processing

The collected tweets have then been enriched with token-level, POS-tags, and lemma information. Meta-information was excluded from processing. So for POS-tagging and lemmatisation we substituted hashtags, mentions (strings of the form *@user-*

<sup>3</sup>This is because we extract generic tweets. Should one want to extract topic-specific tweets, a more targeted list of characterising terms should be used.

*name* referring to a specific user) and URLs with a generic label. All the original information was re-inserted after processing. The tweets were tokenised with the UCTO rule-based tokeniser<sup>4</sup> and then POS-tagged using TreeTagger (Schmid, 1994) with the provided Italian parameter file. Finally, we used the morphological analyser morph-it! (Zanchetta and Baroni, 2005) for lemmatisation.

### 3 Sentiment Analysis

The aim of sentiment analysis (or opinion mining) is detecting someone’s attitude, whether positive, neutral, or negative, on the basis of some utterance or text s/he has produced. While a first step would be determining whether a statement is objective or subjective, and then only in the latter case identify its polarity, it is often the case that only the second task is performed, thereby also collapsing objective statements and a neutral attitude.

In SemEval-2013’s shared task on “Sentiment Analysis in Twitter”<sup>5</sup> (in English tweets), which is currently underway, systems must detect (i) polarity of a given word in a tweet, and (ii) polarity of the whole tweet, in terms of positive, negative, or neutral. This is also what we set to do for Italian. We actually focus on (ii) in the sense that we do not evaluate (i), but we use and combine each word’s polarity to obtain the tweet’s overall polarity.

Several avenues have been explored for polarity detection. The simplest route is detecting the presence of specific words which are known to express a positive, negative or neutral feeling. For example, O’Connor et al. (2010) use a lexicon-projection strategy yielding predictions which significantly correlate with polls regarding ratings of Obama. While it is clear that deeper linguistic analysis should be performed for better results (Pang and Lee, 2008), accurate processing is rather hard on texts such as tweets, which are short, rich in abbreviations and intra-genre expressions, and often syntactically ill-formed. Additionally, existing tools for the syntactic analysis of Italian, such as the DeSR parser (Attardi et al., 2009), might not be robust enough for processing such texts.

Exploiting information coming from a polarity

lexicon, we developed a simple system which assigns to a given tweet one of three possible values: *positive*, *neutral* or *negative*. The only input to the system is the prior polarity coded in the lexicon per *word sense*. We experiment with several ways of combining all the polarities obtained for each word (sense) in a given tweet. Performance is evaluated against manually annotated tweets.

#### 3.1 Polarity lexicon for Italian

Most polarity detection systems make use, in some way, of an affection lexicon, i.e. a language-specific resource which assigns a negative or positive prior polarity to terms. Such resources have been built by hand or derived automatically (Wilson et al., 2005; Wiebe and Mihalcea, 2006; Esuli and Sebastiani, 2006; Taboada et al., 2011, e.g.). To our knowledge, there isn’t such a resource already available for Italian. Besides hand-crafting, there have been proposals for creating resources for new languages in a semi-automatic fashion, using manually annotated sets of seeds (Pitel and Grefenstette, 2008), or exploiting twitter emoticons directly (Pak and Paroubek, 2011). Rather than creating a new polarity lexicon from scratch, we exploit three existing resources, namely MultiWordNet (Pianta et al., 2002), SentiWordNet (Esuli and Sebastiani, 2006; Baccianella et al., 2010), and WordNet itself (Fellbaum, 1998) to obtain an annotated lexicon of senses for Italian. Basically, we port the SentiWordNet annotation to the Italian portion of MultiWordNet, and we do so in a completely automatic fashion.

Our starting point is SentiWordNet, a version of WordNet where the independent values *positive*, *negative*, and *objective* are associated to 117,660 synsets, each value in the zero-one interval. MultiWordNet is a resource which aligns Italian and English synsets and can thus be used to transfer polarity information associated to English synsets in SentiWordNet to Italian synsets. One obstacle is that while SentiWordNet refers to WordNet 3.0, MultiWordNet’s alignment holds for WordNet 1.6, and synset reference indexes are not plainly carried over from one version to the next. We filled this gap using an automatically produced mapping between synsets of Wordnet versions 1.6 and 3.0 (Daud et al., 2000), making it possible to obtain SentiWordNet annotation for the Italian synsets of MultiWordNet. The

<sup>4</sup><http://ilk.uvt.nl/ucto/>

<sup>5</sup>[www.cs.york.ac.uk/semeval-2013/task2/](http://www.cs.york.ac.uk/semeval-2013/task2/).

coverage of our resource is however rather low compared to the English version, and this is due to the alignment procedure which must exploit an earlier version of the resource. The number of synsets is less than one third of that of SentiWordNet.

### 3.2 Polarity assignment

Given a tweet, our system assigns a polarity score to each of its tokens by matching them to the entries in SentiWordNet. Only matches of the correct POS are allowed. The polarity score of the complete tweet is given by the sum of the polarity scores of its tokens.

Polarity is associated to synsets, and the same term can occur in more than one synset. One option would be to perform word sense disambiguation and only pick the polarity score associated with the intended sense. However, the structure of tweets and the tools available for Italian do not make this option actually feasible, although we might investigate it in the future. As a working solution, we compute the positive and negative scores for a term occurring in a tweet as the means of the positive and negative scores of all synsets to which the lemma belongs to in our lexical resource. The resulting polarity score of a lemma is the difference between its positive and negative scores. Whenever a lemma is not found in the database, it is given a polarity score of 0.

One underlying assumption to this approach is that the different senses of a given word have similar sentiment scores. However, because this assumption might not be true in all cases, we introduce the concept of “polypathy”, which is the characterising feature of a term exhibiting high variance of polarity scores across its synsets. The polypathy of a lemma is calculated as the standard deviation of the polarity scores of the possible senses. This information can be used to remove highly polypathic words from the computation of the polarity of a complete tweet, for instance by discarding the tokens with a polypathy higher than a certain threshold. In particular, for the experiments described in this paper, a threshold of 0.5 has been empirically determined. To give an idea, among the most polypathic words in SentiWordNet we found *weird* (.62), *stunning* (.61), *conflicting* (.56), *terrific* (.56).

Taboada et al. (2011) also use SentiWordNet for polarity detection, either taking the first sense of a term (the most frequent in WordNet) or taking the

average across senses, as we also do — although we also add the polypathy-aware strategy. We cannot use the first-sense strategy because through the alignment procedure senses are not ranked according to frequency anymore.

### 3.3 Gold standard

For evaluating the system performance we created two gold standard sets, both annotated by three independent native-speakers, who were given very simple and basic instructions and performed the annotation via a web-based interface. The value to be assigned to each tweet is one out of *positive*, *neutral*, or *negative*. As mentioned, the neutral value includes both objective statements as well as subjective statements where the twitter’s position is neutral or equally positive and negative at the same time (see also (Esuli and Sebastiani, 2007)).

All data selected for annotation comes from TWITA. The first dataset consists of 1,000 randomly selected tweets. The second dataset is topic-oriented, i.e. we randomly extracted 1,000 tweets from all those containing a given topic. Topic-oriented, or target-dependent (Jiang et al., 2011), classification involves detecting opinions about a specific target rather than detecting the more general opinion expressed in a given tweet. We identify a topic through a given hashtag, and in this experiment we chose the tag “Grillo”, the leader of an Italian political movement. While in the first set the annotators were asked to assign a polarity value to the message of the tweet as a whole, in the second set the value was to be assigned to the author’s opinion concerning the hashtag, in this case Beppe Grillo. This is a relevant distinction, since it can happen that the tweet is, say, very negative about someone else while being positive or neutral about Grillo at the same time. For example, the tweet in (1), expresses a negative opinion about Vendola, another Italian politician, but is remaining quite neutral towards Grillo, the target of the annotation exercise.

- (1) #Vendola dà del #populista a #Grillo è una barzelletta o ancora non si è accorto che il #comunismo è basato sul populismo?

Thus, in the topic-specific set we operate a more subtle distinction when assigning polarity, some-

thing which should make the task simpler for a human annotator while harder for a shallow system.

As shown in Table 1, for both sets the annotators detected more than half of the tweets as neutral, or they were disagreeing – without absolute majority, a tweet is considered neutral; however these cases account for only 7.7% in the generic set and 6.9% in the topic-specific set.

Table 1: Distribution of the tags assigned by the absolute majority of the raters

set	positive	negative	neutral
generic	94	301	605
topic-specific	293	145	562

Inter-annotator agreement was measured via Fleiss’ Kappa across three annotators. On the generic set, we found an agreement of  $Kappa = 0.321$ , while on the topic-specific set we found  $Kappa = 0.397$ . This confirms our expectation that annotating topic-specific tweets is actually an easier task. We might also consider using more sophisticated and fine-grained sentiment annotation schemes which have proved to be highly reliable in the annotation of English data (Su and Markert, 2008a).

### 3.4 Evaluation

We ran our system on both datasets described in Section 3.3, using all possible variations of two parameters, namely all combinations of part-of-speech tags and the application of the threshold scheme, as discussed in Section 3.2. We measure overall accuracy as well as precision, recall, and f-score per polarity value. In Tables 2 and 3, we report best scores, and indicate in brackets the associated POS combination. For instance, in Table 2, we can read that the recall of 0.701 for positive polarity is obtained when the system is run without polypathy threshold and using *nouns*, *verbs*, and *adjectives* (nva).

We can draw several observations from these results. First, a fully automatic approach that leverages existing lexical resources performs better than a wild guess. Performance is boosted when highly polypathic words are filtered out.

Second, while the system performs well at recognising especially neutral but also positive polarity, it is really bad at detecting negative polarity. Especially in the topic-specific set, the system assigns

Table 2: Best results on the generic set. In brackets POS combination: (n)oun, (v)erb, (a)djective, adve(r)b.

without polypathy threshold, best accuracy: 0.505 (a)			
	positive	negative	neutral
best prec	0.440 (r)	0.195 (v)	0.664 (nar)
best rec	0.701 (nva)	0.532 (var)	0.669 (a)
best F1	0.485 (nvar)	0.262 (vr)	0.647 (a)
with polypathy threshold, best accuracy: 0.554 (r)			
	positive	negative	neutral
best prec	0.420 (r)	0.233 (v)	0.685 (nar)
best rec	0.714 (nvar)	0.457 (var)	0.785 (r)
best F1	0.492 (nar)	0.296 (vr)	0.698 (r)

Table 3: Best results on the topic-specific set. In brackets POS combination: (n)oun, (v)erb, (a)djective, adve(r)b.

without polypathy threshold, best accuracy: 0.487 (r)			
	positive	negative	neutral
best prec	0.164 (a)	0.412 (a)	0.617 (nar)
best rec	0.593 (nva)	0.150 (nr)	0.724 (a)
best f-score	0.251 (nv)	0.213 (nr)	0.637 (a)
with polypathy threshold, best accuracy: 0.514 (r)			
	positive	negative	neutral
best prec	0.163 (nvar)	0.414 (a)	0.623 (nar)
best rec	0.593 (nvar)	0.106 (nar)	0.829 (r)
best f-score	0.256 (nvar)	0.166 (nar)	0.676 (r)

too many positive labels in place of negative ones, causing positive’s precision and negative’s recall to drop. We believe there are two explanations for this. The first one is the “positive-bias” of SentiWordNet, as observed by Taboada et al. (2011), which causes limited performance in the identification of negative polarity. The second one is that we do not use any syntactic clues, such as negated statements. Including some strategy for dealing with this should improve recognition of negative opinions, too.

Third, the lower performance on the topic-specific dataset confirms the intuition that this task is harder, mainly because we operate a more subtle distinction when assigning a polarity label as we refer to one specific subject. Deeper linguistic analysis, such as dependency parsing, might help, as only certain words would result as related to the intended target.

As far as parts of speech are concerned, there is a tendency for adverbs to be good indicators towards overall accuracy, and best scores are usually obtained exploiting adjectives and/or adverbs.

## 4 Related work

We have already discussed some related work concerning corpus creation, the development of an affection lexicon, and the use of such polarity-annotated resources for sentiment analysis (Section 3). As for results, because this is the first experiment on detecting polarity in Italian tweets, comparing performance is not straightforward. Most work on sentiment analysis in tweets is on English, and although there exist relatively complex systems based on statistical models, just using information from a polarity resource is rather common. Su and Markert (2008b) test SentiWordNet for assigning a subjectivity judgement to word senses on a gold standard corpus, observing an accuracy of 75.3%. Given that SentiWordNet is the automatic expansion over a set of manually annotated seeds, at word-level, this can be considered as an upper bound in sense subjectivity detection. Taboada et al. (2011) offer a survey of lexicon-based methods which are evaluated on adjectives only, by measuring overall accuracy against a manually annotated set of words. Using SentiWordNet in a lexicon-projection fashion yields an accuracy of 61.47% under best settings. These are however scores on single words rather than whole sentences or microtexts.

Considering that we assign polarity to tweets rather than single words, and that in the creation of our resource via automatic alignment we lose more than two thirds of the original synsets (see Section 3.1), our results are promising. They are also not that distant from results reported by Agarwal et al. (2011), whose best system, a combination of unigrams and the best set of features, achieves an accuracy of 60.50% on a three-way classification like ours, evaluated against a manually annotated set of English tweets. Best f-scores reported for positive, negative, and neutral are comprised between 59% and 62%. Similar results are obtained by Pak and Paroubek (2010), who train a classifier on automatically tagged data, and evaluate their model on about 200 English tweets. Best reported f-score on a three-way polarity assignment is just over 60%.

## 5 Conclusions and future work

We have presented the first corpus of Italian tweets obtained in a completely automatic fashion, the first

polarity lexicon for Italian, and the first experiment on sentiment analysis on Italian tweets using these two resources. Both the corpus and the lexicon are as of now unique resources for Italian, and were produced in a way which is completely portable to other languages. In compliance with licensing terms of the sources we have used, our resources are made available for research purposes after reviewing.

Simply projecting the affection lexicon, using two different polarity scoring methods, we experimented with detecting a generic sentiment expressed in a microtext, and detecting the twitter's opinion on a specific topic. As expected, we found that topic-specific classification is harder for an automatic system as it must discern what is said about the topic itself and what is said more generally or about another entity mentioned in the text.

Indeed, this contribution can be seen as a first step towards polarity detection in Italian tweets. The information we obtain from SentiWordNet and the ways we combine it could obviously be used as feature in a learning setting. Other sources of information, to be used in combination with our polarity scores or integrated in a statistical model, are the so-called *noisy labels*, namely strings (such as emoticons or specific hashtags (Go et al., 2009; Davidov et al., 2010)) that can be taken as positive or negative polarity indicators as such. Speriosu et al. (2011) have shown that training a maximum entropy classifier using noisy labels as class predictors in the training set yields an improvement of about three percentage points over a lexicon-based prediction.

Another important issue to deal with is figurative language. During manual annotation we have encountered many cases of irony or sarcasm, which is a phenomenon that must be obviously tackled. There have been attempts at identifying it automatically in the context of tweets (González-Ibáñez et al., 2011), and we plan to explore this issue in future work.

Finally, the co-presence of meta and linguistic information allows for a wide range of linguistic queries and statistical analyses on the whole of the corpus, also independently of sentiment information, of course. For example, correlations between parts-of-speech and polarity have been found (Pak and Paroubek, 2010), and one could expect also correlations with sentiment and time of the day, or month of the year, and so on.

## References

- Apoorv Agarwal, Boyi Xie, Iliia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, LSM '11, pages 30–38, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Giuseppe Attardi, Felice Dell’Orletta, Maria Simi, and Joseph Turian. 2009. Accurate dependency parsing with a stacked multilayer perceptron. In *Proceeding of Evalita 2009*, LNCS. Springer.
- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Nicoletta Calzolari et al., editor, *Proceedings of LREC 2010*.
- Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226.
- Jordi Daud, Llus Padr, and German Rigau. 2000. Mapping wordnets using structural information. In *38th Annual Meeting of the Association for Computational Linguistics (ACL’2000)*, Hong Kong.
- Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING ’10, pages 241–249, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06)*, pages 417–422.
- Andrea Esuli and Fabrizio Sebastiani. 2007. Pageranking wordnet synsets: An application to opinion mining. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 424–431, Prague, Czech Republic, June. Association for Computational Linguistics.
- Christiane Fellbaum, editor. 1998. *WordNet. An Electronic Lexical Database*. The MIT Press.
- Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment analysis using distant supervision. <http://cs.wmich.edu/~tllake/fileshare/TwitterDistantSupervision09.pdf>.
- Roberto González-Ibáñez, Smaranda Muresan, and Nina Wacholder. 2011. Identifying sarcasm in twitter: A closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 581–586, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 151–160, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- Marco Lui and Timothy Baldwin. 2012. langid.py: An off-the-shelf language identification tool. In *ACL (System Demonstrations)*, pages 25–30. The Association for Computer Linguistics.
- Brendan O’Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In William W. Cohen and Samuel Gosling, editors, *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26*. The AAAI Press.
- Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In Nicoletta Calzolari et al., editor, *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*. European Language Resources Association.
- Alexander Pak and Patrick Paroubek. 2011. Twitter for sentiment analysis: When language resources are not available. *23rd International Workshop on Database and Expert Systems Applications*, 0:111–115.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.
- Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, pages 21–25.
- Guillaume Pitel and Gregory Grefenstette. 2008. Semi-automatic building method for a multidimensional affect dictionary for a new language. In *Proceedings of LREC 2008*.
- Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees.
- Michael Speriosu, Nikita Sudan, Sid Upadhyay, and Jason Baldridge. 2011. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the First workshop on*

- Unsupervised Learning in NLP*, pages 53–63, Edinburgh, Scotland, July. Association for Computational Linguistics.
- Fangzhong Su and Katja Markert. 2008a. Eliciting subjectivity and polarity judgements on word senses. In *Proceedings of COLING 2008 Workshop on Human Judgements in Computational Linguistics, Manchester, UK*.
- Fangzhong Su and Katja Markert. 2008b. From words to senses: A case study of subjectivity recognition. In Donia Scott and Hans Uszkoreit, editors, *Proceedings of COLING 2008, Manchester, UK*, pages 825–832.
- Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Comput. Linguist.*, 37(2):267–307, June.
- Erik Tjong Kim Sang and Johan Bos. 2012. Predicting the 2011 dutch senate election results with twitter. In *Proceedings of the Workshop on Semantic Analysis in Social Media*, pages 53–60, Avignon, France, April. Association for Computational Linguistics.
- Janyce Wiebe and Rada Mihalcea. 2006. Word sense and subjectivity. In Nicoletta Calzolari, Claire Cardie, and Pierre Isabelle, editors, *ACL*. The Association for Computer Linguistics.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Human Language Technology and Empirical Methods in Natural Language Processing Conference, 6-8 October, Vancouver, British Columbia, Canada*.
- Eros Zanchetta and Marco Baroni. 2005. Morph-it! a free corpus-based morphological resource for the italian language. In *Proceedings of Corpus Linguistics 2005*.