

A manually-annotated Italian corpus for fine-grained sentiment analysis

Marilena Di Bari, Serge Sharoff, Martin Thomas

University of Leeds

School of Languages, Cultures and Societies

LS2 9JT, Leeds (UK)

{mlmdb, s.sharoff, m.thomas}@leeds.ac.uk

Abstract

English. This paper presents the results of the annotation carried out on the Italian section of the *SentiML corpus*, consisting of both originally-produced and translated texts of different types. The two main advantages are that: (i) the work relies on the linguistically-motivated assumption that, by encapsulating opinions in pairs (called *appraisal groups*), it is possible to annotate (and automatically extract) their sentiment in context; (ii) it is possible to compare Italian to its English and Russian counterparts, as well as to extend the annotation to other languages.

Italiano. *Questo lavoro presenta i risultati dell'annotazione effettuata sulla sezione italiana del corpus "SentiML", che consiste di testi sia originali che tradotti appartenenti a diversi tipi. I due vantaggi principali sono che: (i) il lavoro si fonda sull'assunzione motivata linguisticamente che, codificando le opinioni in coppie (chiamate appraisal groups), è possibile annotare (ed estrarre automaticamente) il loro sentiment tenendo in considerazione il contesto; (ii) è possibile confrontare l'italiano con le sue controparti inglese e russa, ed estendere l'annotazione ad altre lingue.*

1 Introduction

Overall, the field of Sentiment Analysis (SA) aims at automatically classifying opinions as positive, negative or neutral (Liu, 2012). While at first the focus of SA was on the document level (coarse-grained) classification, with the years it has become more and more at the sentence level or below the sentence (fine-grained). This shift has

been due to both linguistic and application reasons. Linguistic reasons arise because sentiment is often expressed over specific entities rather than an overall document. As for practical reasons, SA tasks are often aimed at discriminating between more specific aspects of these entities. For example, if an opinion is supposed to be on the plot of a movie, it is not unusual that the user also evaluates actors' performance or director's choices (Shastri et al., 2010). For SA applications these opinions need to be assessed separately. Also opinions are not expressed as simple and direct assertions, but by using a number of stylistic devices such as pronominal references, abbreviations, idioms and metaphors. Finally, the automatic identification of sarcasm, irony and humour is even more challenging (Carvalho et al., 2009).

For all these reasons, fine-grained sentiment analysis is looking at entities that are usually chains of words such as "noun+verb+adjective" (e.g. the house is beautiful) or "adverb+adjective+noun" (e.g. very nice car) (Yi et al., 2003; Popescu and Etzioni, 2005; Choi et al., 2006; Wilson, 2008; Liu and Seneff, 2009; Qu et al., 2010; Johansson and Moschitti, 2013).

In addition to the multitude of approaches to fine-grained SA, there is also shortage of multilingual comparable studies and available resources. To close this gap, we designed the *SentiML* annotation scheme (Di Bari et al., 2013) and applied it to texts in three languages, English, Italian and Russian. The proposed annotation scheme extends previous works (Argamon et al., 2007; Bloom and Argamon, 2009) and allows multi-level annotations of three categories: *target* (T) (expression the sentiment refers to), *modifier* (M) (expression conveying the sentiment) and *appraisal group* (AG) (couple of modifier and target). For example in:

"Gli uomini hanno il potere di
[[sradicare]_M la [[povertà]_T]_{AG},
ma anche di [[sradicare]_M le

[tradizioni]_T]_{AG}”.

(Men have the power to eradicate poverty, but also to eradicate traditions)

the groups “*sradicare povertà*” (eradicate poverty) and “*sradicare tradizioni*” (eradicate traditions) have an opposite sentiment despite including the same word *sradicare* (to eradicate).

This scheme has been developed in order to facilitate the annotation of the sentiment and other advanced linguistic features that contribute to it, but also the appraisal type according to the *Appraisal Framework* (AF) (Martin and White, 2005) in a multilingual perspective (Italian, English and Russian). The AF is the development of the *Systemic Functional Linguistics* (Halliday, 1994) specifically concerned with the study of the language of evaluation, attitude and emotion. It consists of *attitude*, *engagement* and *graduation*. Of these, *attitude* is sub-divided into *affect*, which deals with personal emotions and opinions (e.g. *excited*, *lucky*); *judgement*, which concerns author’s attitude towards people’s behaviour (e.g. *nasty*, *blame*); *appreciation*, which considers the evaluation of things (e.g. *unsuitable*, *comfortable*). The *engagement* system considers the positioning of oneself with respect to the opinions of others, while *graduation* investigates how the use of language amplifies or diminishes attitude and engagement. In particular, *force* is related to intensity, quantity and temporality. To the best of our knowledge the AF has only been applied in the case of Italian for purposes not related to computation (Pounds, 2010; Manfredi, 2011).

This paper is organized as follows: Section 2 describes the annotation scheme and the annotated Italian corpus, Section 3 reports the results and finally Section 4 our conclusions.

2 Annotation scheme and corpus

The scheme, described in (Di Bari et al., 2013), specifies different attributes for the categories *target*, *modifier* and *appraisal group*.

A target is usually a noun. Targets have 2 attributes: *type* (‘person’, ‘thing’, ‘place’, ‘action’ and ‘other’), and prior (out-of-context) *orientation* (‘positive’, ‘negative’, ‘neutral’ and ‘ambiguous’).

A modifier is what *modifies* the target. It can be an adjective, a verb, an adverb or a noun in the case of two nouns linked by a preposition, e.g. “libertà di parola” (freedom of speech). Modifiers have 4

attributes: *attitude* (‘affect’, ‘judgement’ and ‘appreciation’); *force* referring to the intensity of the modifier, i.e. high like in the case of “molto bella” (very beautiful), ‘low’ like in the case of “poco elegante” (little elegant), ‘reverse’ like in the case of “contro la guerra” (against the war) or ‘normal’; *polarity* if there is a negation (‘marked’) or not (‘unmarked’), and prior (out-of-context) *orientation* (‘positive’, ‘negative’, ‘neutral’ and ‘ambiguous’).

Appraisal groups have 1 attribute: contextual *orientation* (‘positive’, ‘negative’, ‘neutral’ and ‘ambiguous’).

In the example sentence shown in Section 1, the modifier *sradicare* would thus have attitude ‘judgement’, force ‘normal’, polarity ‘unmarked’, orientation ‘ambiguous’; the target *povertà* would have type ‘thing’ and orientation ‘negative’, whereas the target *tradizioni* would have type ‘thing’ and orientation ‘positive’; the appraisal group “*sradicare povertà*” would have orientation ‘positive’, while the appraisal group “*sradicare tradizioni*” would have orientation ‘negative’.

SentiML has been applied to the text types different from those taken into account in previous works in Italian (Casoto et al., 2008; Basile and Nissim, 2013; Bosco et al., 2013; Sorgente et al., 2014):

- **Political speeches.** Translations of American presidents’ addresses.
- **Talks.** Translations of TED (Technology, Entertainment, Design) talks (Cettolo et al., 2012).
- **News.** Belonging to the newspaper *Sole24ore*.

The corpora have been annotated by using MAE (Stubbs, 2011), a freely available software annotation environment. The Italian corpus contains 328 sentences for a total of 9080 tokens. To deal with the limitation of having only one annotator, different confidence-rated machine learning classifiers were used to spot inconsistencies and thus revise the annotations accordingly ((Di Bari et al., 2014)).

3 Results of the annotation

In Table 1 details about the number of the appraisal groups, targets and modifiers are shown,

Language	Text type	Appraisal groups	Targets	Modifiers	% of words included in appraisal groups
ITA	Political	486	411	437	25%
	News	254	203	244	22%
	TED	341	292	323	24%
	tot	1081	906	1004	24%

Table 1: Statistics on the annotated data. A different amount of appraisal groups has been annotated according to the text type, but on average the 24% of words are sentiment-loaded.

along with the percentages of words embedded in appraisal groups for each text type.

Figure 1 shows that ‘positive’ orientation is the predominant one for appraisal groups with 67%, followed by ‘negative’ with 32%. These data are consistent with the assumption that appraisal groups should not be ‘neutral’ nor ‘ambiguous’ because they carry appraisal and their orientation should be clear in context. At the same time, targets and modifiers can be ‘ambiguous’ because their orientation depends on the context and ‘neutral’ in case they are not the element carrying appraisal in the group.

Figure 2 shows the statistics on the other attributes: ‘appreciation’ is the most common attitude, which is consistent with the fact that this value is associate to ‘thing’ in the AF (see Section 1), which is the most common target type; polarity, which indicates that a negation has been encountered, has been ‘marked’ 4% times; force, an important feature for a more accurate prediction of the sentiment, is ‘reverse’ 4% of times.

We have also compared the contextual orientation manually annotated by us with the prior orientation included in the translation of the ‘positive’ and ‘negative’ values in the *NRC Word-Emotion Association Lexicon* (Mohammad, 2011), whose English annotations were manually done through *Amazon’s Mechanical Turk*, and the *Roget Thesaurus* and it has entries for about 14200 word types. We calculated that, in the case of Italian, only 29.39% of the words belonging to the appraisal groups were present in the sentiment dictionary, with higher percentage for political speeches (33.54%), followed by news (27.66%) and TED talks (26.98%). As previously found in the case of English, most of these are nouns reasonably not carrying sentiment on their own, but still part of an appraisal group (e.g., *brevetti* (patents), *computer*, *confini* (borders), *nostro* (our)). There are also cases, ad-

jectives in particular, that should probably be included in a dictionary with prior orientation (e.g., *necessario* (necessary), *negativo* (negative), *overato* (overburdened), *ideale* (ideal)).

In line with our previous experiments in English (Di Bari et al., 2013), we used the following categories for the comparison:

Agreeing words: words whose dictionary orientation agrees with that of the appraisal group they belong to. They cover 69.63% of the total times words were found in the dictionary. This means that we can rely to a certain extent to the dictionary orientation, but not if we aim at more accuracy. The list includes reasonable out-of-context positive words (e.g., *alleati* (allies), *comprensione* (comprehension), *dotato* (gifted), *felicità* (happiness)), as well as out-of-context negative words (e.g., *debolezza* (weakness), *malattia* (sickness), *stagnante* (stagnant), *violenza* (violence)).

Disagreeing words: words whose dictionary orientation does not agree with that of the appraisal group they belong to. They cover 28.18% of the total times words were found in the dictionary, a percentage that demonstrates how crucial the context is. For example reversals such as *abolire* (abolish) and *diminuire* (diminish), and *sfida* (challenge), *sopportare* (to bear), *tendenza* (trend). However, it was interesting to notice that also words normally considered positive (e.g. *prosperare* (to prosper) and *risorse* (resources)) or negative (e.g. and *tensione* (tension) and *rischio* (risk)) became included in groups with opposite orientation.

Ambiguous words: words which already have both positive and negative values in the dictionary. They are *resta* (stays), *rivoluzione* (revolution), *sciogliere* (to unleash), *umile* (humble), and they cover 1.07%.

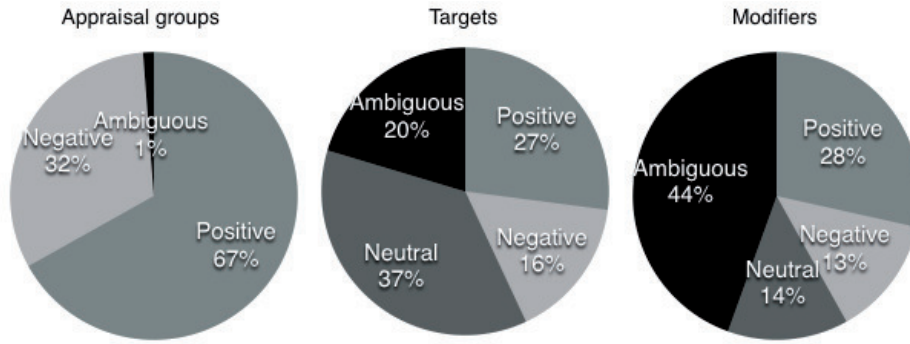


Figure 1: Values for the attribute orientation for appraisal groups, targets and modifiers. In the case of appraisal groups, positive is the most common value, followed by negative.

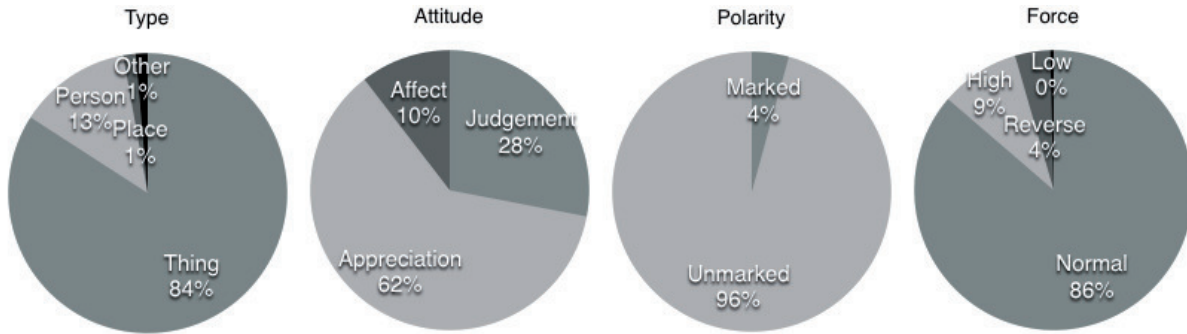


Figure 2: Values for the attributes attitude, polarity, force and type.

4 Conclusions

In this paper we have described a manually-annotated corpus of Italian for fine-grained sentiment analysis. The manual annotation has been done in order to include important linguistic features. Apart from extracting statistics related to the annotations, we have also compared the manual annotations to a sentiment dictionary and demonstrated that (i) the dictionary includes only 29.29% of the annotated words, and (ii) the prior orientation given in the dictionary is different from the correct one given by the context in 28.18% of the cases.

The original and annotated texts in Italian (along with English and Russian) and the Document Type Definition (DTD) of *SentiML* to be used with MAE are publicly available¹.

In the meanwhile, the authors are already working on an automatic system to identify and classify appraisal groups multilingually.

¹<http://corpus.leeds.ac.uk/marilena/SentiML>

Acknowledgments

The first author would like to thank Michele Filannino (The University of Manchester) for his insights throughout the research.

References

- S. Argamon, K. Bloom, A. Esuli, and F. Sebastiani. 2007. Automatically Determining Attitude Type and Force for Sentiment Analysis. In *Proceedings of the 3rd Language and Technology Conference (LTC'07)*, pages 369–373, Poznan, Poland.
- Valerio Basile and Malvina Nissim. 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.
- Kenneth Bloom and Shlomo Argamon. 2009. Automated learning of appraisal extraction patterns. *Language and Computers*, 71(1):249–260.
- Cristina Bosco, Viviana Patti, and Andrea Bolioli. 2013. Developing corpora for sentiment analysis: The case of irony and senti-tut. *IEEE Intelligent Systems*, 28(2):55–63.
- Paula Carvalho, Luís Sarmento, Mário J. Silva, and Eugénio de Oliveira. 2009. Clues for detecting

- irony in user-generated contents: oh...!! it's "so easy" ;-). In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, TSA '09, pages 53–56, New York, NY, USA. ACM.
- Paolo Casoto, Antonina Dattolo, and Carlo Tasso. 2008. Sentiment classification for the italian language: A case study on movie reviews. *Journal of Internet Technology*, 9(4):365–373.
- Mauro Cettolo, Christian Girardi, and Marcello Federico. 2012. Wit³: Web inventory of transcribed and translated talks. In *Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT)*, pages 261–268, Trento, Italy, May.
- Yejin Choi, Eric Breck, and Claire Cardie. 2006. Joint extraction of entities and relations for opinion recognition. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, pages 431–439, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Marilena Di Bari, Serge Sharoff, and Martin Thomas. 2013. SentiML: Functional annotation for multilingual sentiment analysis. In *DH-case 2013: Collaborative Annotations in Shared Environments: metadata, vocabularies and techniques in the Digital Humanities*, ACM International Conference Proceedings.
- Marilena Di Bari, Serge Sharoff, and Martin Thomas. 2014. Multiple views as aid to linguistic annotation error analysis. In *Proceedings of the 8th Linguistic Annotation Workshop (LAW VIII)*. ACL SIGANN Workshop held in conjunction with Coling 2014.
- M.A.K. Halliday. 1994. *An Introduction to Systemic Functional Linguistics*. London:Arnold, 2 edition.
- Richard Johansson and Alessandro Moschitti. 2013. Relational features in fine-grained opinion analysis. *Computational Linguistics*, 39(3).
- Jingjing Liu and Stephanie Seneff. 2009. Review sentiment scoring via a parse-and-paraphrase paradigm. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, EMNLP '09, pages 161–169, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- Marina Manfredi. 2011. Systemic functional linguistics as a tool for translation teaching: towards a meaningful practice. *Rivista Internazionale di Tecnica della Traduzione*, 13:49 – 62.
- James R Martin and Peter RR White. 2005. *The language of evaluation*. Palgrave Macmillan, Basingstoke and New York.
- Saif Mohammad. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114, Portland, OR, USA, June.
- Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 339–346, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Gabrina Pounds. 2010. Attitude and subjectivity in italian and british hard-news reporting: The construction of a culture-specific 'reporter'voice. *Discourse Studies*, 12(1):106–137.
- Lizhen Qu, Georgiana Ifrim, and Gerhard Weikum. 2010. The bag-of-opinions method for review rating prediction from sparse text patterns. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, pages 913–921, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Lokendra Shastri, Anju G. Parvathy, Abhishek Kumar, John Wesley, and Rajesh Balakrishnan. 2010. Sentiment extraction: Integrating statistical parsing, semantic analysis, and common sense reasoning. In *IAAI*.
- Antonio Sorgente, Via Campi Flegrei, Giuseppe Vetigli, and Francesco Mele. 2014. An italian corpus for aspect based sentiment analysis of movie reviews. In *First Italian Conference on Computational Linguistics CLiC-it*.
- Amber Stubbs. 2011. Mae and mai: Lightweight annotation and adjudication tools. In *Linguistic Annotation Workshop*, pages 129–133.
- Theresa Ann Wilson. 2008. *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*. Ph.D. thesis, University of Pittsburgh.
- Jeonghee Yi, Tetsuya Nasukawa, Razvan Bunescu, and Wayne Niblack. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pages 427–434. IEEE.