CorEA:
Italian News Corpus with Emotions and Agreement

Fabio Celli
University of Trento
Via Sommarive 5,
Trento, Italy
fabio.celli@unitn.it

Giuseppe Riccardi
University of Trento
Via Sommarive 5,
Trento, Italy
riccardi@disi.unitn.it

Arindam Ghosh
University of Trento
Via Sommarive 5,
Trento, Italy
arindam.ghosh@unitn.it

Abstract

English. In this paper, we describe an Italian corpus of news blogs, including bloggers' emotion tags, and annotations of agreement relations amongst blogger-comment pairs. The main contributions of this work are: the formalization of the agreement relation, the design of guidelines for its annotation, the quantitative analysis of the annotators' agreement.

Italiano. *In questo articolo descriviamo la raccolta di un corpus di blog giornalistici in Italiano che include le emozioni etichettate dai blogger e l'annotazione manuale con la relazione di approvazione tra commenti. I contributi principali di questo articolo sono: la formalizzazione della relazione di approvazione, le linee guida per la sua annotazione e l'analisi quantitativa dell'accordo tra annotatori.*

1 Introduction

Online news media, such as journals and blogs, allow people to comment news articles, to express their own opinions and to debate about a wide variety of different topics, from politics to gossips. In this scenario, commenters express approval and dislike about topics, other users and articles, either in a linguistic form and/or using like precoded actions (e.g. *like* buttons). Corriere is one of the most visited Italian news websites, attracting over 1.6 million readers everyday[1]. The peculiarity of *corriere.it* with respect to most news websites, is that it contains metadata on emotions expressed by the readers about the articles. The emotions (amused, satisfied, sad, preoccupied and indignated) are annotated directly by the readers on a voluntary basis. They can express one emotion per article. In this paper, we describe the collection of a corpus from *corriere.it*, that combines emotions and agreement/disagreement.

The paper is structured as follows: in section 2 we will provide an overview of related work, in sections 3 and 4 we will define the agreement/disagreement relation, describe the corpus, comparing it to related work, and provide the annotation guidelines. In section 5 we will draw some conclusions.

2 Background and Related Work

The CorEA corpus combines emotions and agreement/disagreement in a social media domain. Emotions and sentiment in corpora are usually annotated manually or automatically at message level. Examples of manually annotated corpora are Affective Text (Strapparava and Mihalcea, 2007), that contains annotation of news titles with emotion labels (anger, disgust, fear, joy, sadness, surprise), and sentiTUT (Bosco et al., 2013), that combines sentiment (positive/negative message polarity) and irony. Automatically and semi-automatically annotated corpora, like TWITA (Basile and Nissim, 2013), usually exploit external resources such as senticNet (Cambria et al., 2012). The peculiarity of CorEA is that emotions are annotated directly by commenters on a voluntary basis. These ground truth emotion labels (amused, satisfied, sad, preoccupied and indignated) are not at message level, but at author level. In other words are part of the bloggers' personal profile and describe all the emotions they declared after reading articles.

There are not many corpora of agreement/disagreement. The ICSI corpus of multi-party conversation (Shriberg et al., 2004), is a collection of 75 meetings between 53 unique speakers, annotated with

[1] source 'http://en.wikipedia.org/wiki/Corriere della Sera' retrieved in Jan 2014.

dialogue acts (including 4 labels for strong and weak agreement/disagreement) by 2 raters. A specific inter-annotator agreement for the agreement/disagreement relation is not reported. More recent corpora with agreement/disagreement labels are the AAWD corpus of Wikipedia talk pages (Bender et al., 2011), the AACD chat corpus (Morgan et al., 2013) and the IAC/ARGUE corpus of political debates (Abbott et al., 2011) (Walker et al., 2012). AAWD is a collection of asynchronous conversations from Wikipedia in English, Russian and Mandarin Chinese (about 500 threads and 325k tokens in total). It is annotated with 2 classes (agreement/disagreement, called positive/negative alignment) and authority claims by 2 annotators. AACD is a small corpus (12 threads, 14k tokens in total) of elicited chat dialogues in the same languages, annotated in the same way. The average inter-annotator agreement for alignment over the three languages of AAWD is Cohen's *k*=0.5 (Cohen, 1977). IAC/ARGUE is a large corpus in English (about 2700 authors, 11k threads) sampled from 4forums.com and annotated with Amazon's Mechanical Turk[2]. It combines agreement/disagreement, emotionality (subjective/objective), sarcasm, attack (objective/offensive language) and attitude (nice/nasty). Agreement/disagreement in IAC/ARGUE has been annotated with a scale +5, -5 and the inter-annotator agreement is α=0.62 (Krippendorff, 2004).

In all these corpora agreement/disagreement is at message level (post or utterance). There is also a corpus that combines LiveJournal and Wikipedia (118 threads) (Andreas et al., 2012), annotated with agreement/disagreement labels at sentence level (segments or chunks of messages). They reported inter-annotator agreement on 3 classes (agree/disagree/neutral) between 2 annotators as Cohen's *k*=0.73. CorEA corpus aggregates self-reported anntotations, such as emotions and likes, and metadata information, (ids, time stamps, etc.) about the conversation and human annotation of the agreement/disagreement relation.

3 Definition of Agreement/Disagreement

During debates in social media, participants attack or support the content of other participants' messages (Herring, 2007). This practice can be modeled in two different actions: 1) refer-to-message and 2) expression of agreement/disagreement. Refer-to-message, depicted as connection lines with round heads in figure 1, are directed links between pairs of messages. This information can be encoded as metadata in the message exchange strucure - as in Corriere - or as surface realizations in text in the form of coreference expressions (i.e. @ Lettore_10108563, see figure 1). Here we define agreement/disagreement as a relation, built on refer-to-message links, between a set of participants $P = \{p_i, ..., p_n\}$ to a conversation C that generate a set of messages $M = \{m_{ij}, ..., m_{nm}\}$, where m_{ij} is the j^{th} message of participant p_i. The conversation contains a set of

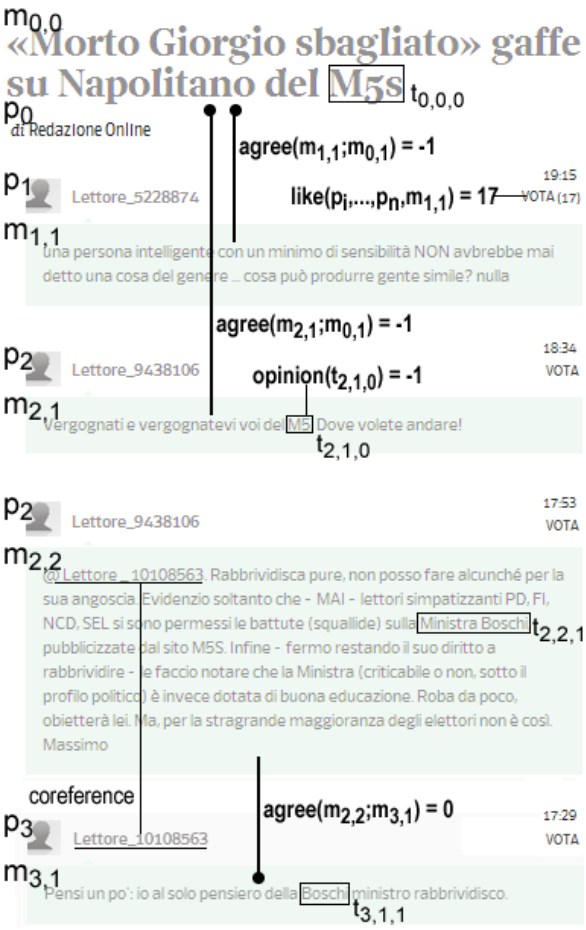

Figure 1: Example of asynchronous conversation in Corriere with participants $P = \{p_i, ..., p_n\}$, the messages they produce $M = \{m_{ij}, ..., m_{nm}\}$, sorted by time from bottom to top, and topics within messages $T = \{t_{ijk}, ..., t_{nmo}\}$. Connection lines with round heads are refer-to-message links, occasionally corresponding to coreferences. The agreement/disagreement relation is defined at message level as the *agree* functions $agree(m_{ij}; m_{i'j'})$ that maps pairs of participants/messages to values (+1,0,-1). *Opinion* is a function that maps a topic to a positive or negative polarity (+1,-1). *Like* is a function that measures the appreciation of participants to a message.

[2]https://www.mturk.com/mturk/welcome

topics $T = \{t_{ijk}, ..., t_{nmo}\}$, where t_{ijk} is the k^{th} topic of the conversation contained into the j^{th} message of participant p_i. We define topics as recurrent chunks or named entities appearing in different messages of C (see figure 1). We formalize agreement/disagreement as the *agree* function, that maps pairs of participants and messages to values between 1 (agree) and -1 (disagree), where 0 is neutral, as reported below:

$agree(m_{ij}; m_{i'j'}) = \{-1, 0, 1\}$

where m_{ij} is the parent participant/message pair, and $m_{i'j'}$ is the child participant/message pair. The parent m_{ij} precedes the child $m_{i'j'}$ in a time sequence. The child $m_{i'j'}$ is the j'^{th} message of $p_{i'}$ referred to the j^{th} message of p_i. The *agree* function is different from opinion expression and from like. The *opinion* function maps a topic to a positive or negative polarity (+1,-1):

$opinion(t_{ijk}) = \{-1, 1\}$

The *like* function measures the appreciation of a subset of participants to a message:

$like(p_i, ..., p_n; m_{ij}) = \{0, inf\}$

It is possible to define a more fine-grained function at topic level $agree(t_{ijk}; t_{i'j'k}) = \{-1, 0, 1\}$, where two (portions of) different messages m_{ij} and $m_{i'j'}$, connected by a refer-to-message link, are generated by two different participants (p_i and $p_{i'}$), and contain the same topic. The annotation of agreement/disagreement at topic level requires much more effort than at message level, we plan to annotate CorEA at topic level in the future. The agreement/disagreement relation concerns participants, messages and topics. Since participants are an important part of the relation, the *agree* function should exploit also information about them. This is why we combined emotions and agreement/disagreement relations in a single corpus. In Corriere, and social media in general, users/commenters/bloggers/authors are participants, comments/posts are messages, threads are conversations and articles are the first message of a conversation. In the next section we describe the procedure for the annotation of agreement/disagreement in CorEA.

4 Data, Annotation Schema and Guidelines

The CorEA corpus is a collection of news articles and comments from Corriere. It contains 27 news articles, about 1660 unique authors and more than 2900 posts (comments and articles) for a total of 135.6k tokens. Details are reported in table 1. We selected articles from all the main categories of news, in order to have a balance between categories that generate many comments, such as politics, and categories that generate few comments, such as culture and science. The corpus contains the data reported in table 2.

topics	articles	tokens	comments
technology	4	11.6k	266
culture	3	9.3k	215
politics	3	39.2k	876
science	2	2.6k	70
economics	3	30.1k	578
news	6	31.6k	560
gossip	3	4.4k	168
sport	3	6.8k	154
total	**27**	**135.6k**	**2887**

Table 1: Details of the CorEA corpus.

We performed a manual annotation of the agreement/disagreement relations at message level on each child participant/message pair, using the following guidelines:

field	description
Mid	message Id
Pid	participant Id
Pname	participant's nickname
Mtype	article/comment
text	text
timestamp	date/time
category	macro-topic
refer-to-P	Id of parent participant
refer-to-M	Id of parent message
avatar	link to participant's picture
replies-count	replies to the message
likes	like count of the message
agree	agree/disagre labels
Pday-activity	participant's activity score
Pinterests	count of interests of participant
Pviews	participant page views
Pcomments	count of messages of participant
Pshares	count of shares
Pcomments-votes	count of participant's votes
emo-indig	indignation score
emo-disapp	disappointment score
emo-worried	preoccupation score
emo-amused	amusement score
emo-satisfied	satisfaction score

Table 2: Corpus data schema.

1) Read and understand the content of the article and its title.

2) Read the messages of each child pair one by one, sorted by time from the oldest to the newest.

3) For each child pair, check the refer-to-message link finding the corresponding parent pair.

4) read the parent pair, understand the semantics of the relation between child and parent.

5) Annotate with a "NA" label (not applicable) if the child falls under one or both the following conditions: a) **broken refer-to-mesage**: cannot find the parent (e.g. the message is not referred to any other); b) **mixed agreement** (e.g. "I partly agree with you but ..").

6) Judge the agreement/disagreement expressed in the child with respect to the parent. Annotate the child pair with the corresponding label: agree (1), disagree (-1) neutral (0). We did not use any annotation tool. An example of annotation follows:

```
  1:  5 Stars Movement party
returns 2.5 milions Euros to
Italian citizens.
  2:  great!!!.  [agree(2,1)=1]
  3:  http://xyz.com see this :)
ha ha [NA]
  4:  what has to do this link
with the topic?  [agree(4,3)=-1]
  5:  if only every party did
it!..  [agree(5,1)=1]
  6:  would not change anything.
[agree(6,5)=-1]
  7:  what do you mean?
[agree(7,6)=0]
```

We computed the inter-annotator agreement between two Italian native speaker raters, for 50 and 100 instances with 2 (+1, -1) and 3 classes (+1, -1, 0). The "NA" labels were reannotated into the other classes to include all cases into the evaluation. We used Fliess' k (Fleiss et al., 1981), comparable to Cohen's k (used in most of previous work) but generalized over individual raters, like Krippendorf's α (Artstein and Poesio, 2008). Results are reported in table 3. In particular, we noticed that the neutral class is the main source of disagreement between annotators. Figure 2 reports the distribution of the agreement/disagreement labels between annotators and in the corpus. We annotated again the examples using only 2 classes: inter-annotator agreement rose from moderate, in line with (Morgan et al., 2013), to substantial.

type	classes	instances	score
inter	3	50	k=0.6
inter	3	100	k=0.58
inter	2	50	k=0.87
inter	2	100	k=0.93
intra	3	100	k=0.87
intra	2	100	k=0.91

Table 3: Inter- and intra- annotator agreement for the agreement/disagreement relation annotation in CorEA.

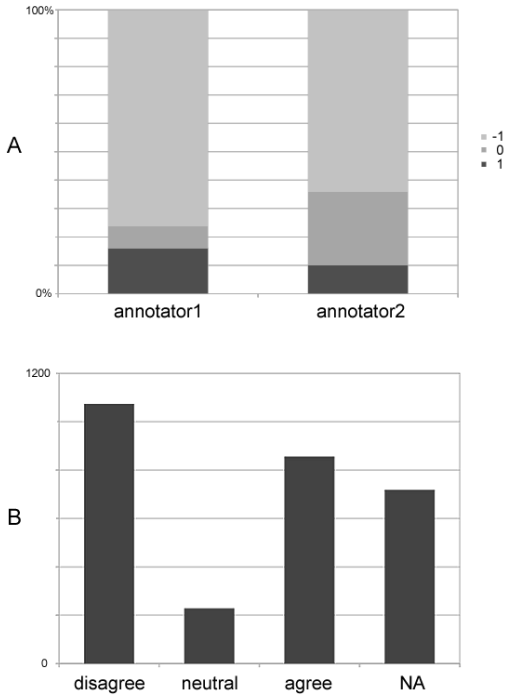

Figure 2: A) Distribution of agreement/disagreement labels between 2 annotators (50 comments, 3 classes) and B) distribution of labels in the corpus.

We labeled twice a set of 100 comments to compute intra-annotator agreement, reported in table 3 as well.

5 Conclusion

We presented the CorEA corpus, a resource that combines agreement/disagreement at message level and emotions at participant level. We are not aware of any other resource of this type for Italian. We found that the best way to annotate agreement/disagreement is with binary classes, filtering out "NA" and neutral cases.

In the future, we would like to annotate CorEA at topic level and develop classifiers for agreement/disagreement. We plan to make available the corpus at the end of the project.

Aknowledgements

The research leading to these results has received funding from the European Union - Seventh Framework Programme (FP7/2007-2013) under grant agreement n 610916 SENSEI.

References

Rob Abbott, Marilyn Walker, Pranav Anand, Jean E. Fox Tree, Robeson Bowmani, and Joseph King. 2011. How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proceedings of the Workshop on Languages in Social Media*, pages 2–11. Association for Computational Linguistics.

Jacob Andreas, Sara Rosenthal, and Kathleen McKeown. 2012. Annotating agreement and disagreement in threaded discussion. In *LREC*, pages 818–822.

Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Comput. Linguist.*, 34(4):555–596, dec.

Valerio Basile and Malvina Nissim. 2013. Sentiment analysis on italian tweets. *WASSA 2013*, page 100.

Emily M. Bender, Jonathan T. Morgan, Meghan Oxley, Mark Zachry, Brian Hutchinson, Alex Marin, Bin Zhang, and Mari Ostendorf. 2011. Annotating social acts: Authority claims and alignment moves in wikipedia talk pages. In *Proceedings of the Workshop on Languages in Social Media*, pages 48–57. Association for Computational Linguistics.

Cristina Bosco, Viviana Patti, and Andrea Bolioli. 2013. Developing corpora for sentiment analysis and opinion mining: the case of irony and senti-tut. *IEEE Intelligent Systems*, page 1.

Erik Cambria, Catherine Havasi, and Amir Hussain. 2012. Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis. In *FLAIRS Conference*, pages 202–207.

Jacob Cohen. 1977. *Statistical power analysis for the behavioral sciences*. Academic Press, New York.

Joseph L. Fleiss, Bruce Levin, and Myunghee Cho Paik. 1981. The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2:212–236.

Susan C. Herring. 2007. A faceted classification scheme for computer-mediated discourse. *Language@ internet*, 4(1):1–37.

Klaus Krippendorff. 2004. Measuring the reliability of qualitative text analysis data. *Quality & quantity*, 38:787–800.

Jonathan T. Morgan, Meghan Oxley, Emily Bender, Liyi Zhu, Varya Gracheva, and Mark Zachry. 2013. Are we there yet?: The development of a corpus annotated for social acts in multilingual online discourse. *Dialogue & Discourse*, 4(2):1–33.

Elizabeth Shriberg, Raj Dhillon, Sonali Bhagat, Jeremy Ang, and Hannah Carvey. 2004. The icsi meeting recorder dialog act (mrda) corpus. Technical report, DTIC Document.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics.

Marilyn A. Walker, Jean E. Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A corpus for research on deliberation and debate. In *LREC*, pages 812–817.