EmotiBlog: un esquema de anotación detallado para la sujetividad en los nuevos géneros textuales de la Web 2.0

EmotiBlog: a fine-grained annotation schema for labelling subjectivity in the new-textual genres born with the Web 2.0

Ester Boldrini
University of Alicante, Department of Language and Computing Systems
eboldrini@dlsi.ua.es

Alexandra Balahur
University of Alicante, Department of Language and Computing Systems
abalahur@dlsi.ua.es

Patricio Martínez-Barco
University of Alicante, Department of Language and Computing Systems
patricio@dlsi.ua.es

Andrés Montoyo
University of Alicante, Department of Language and Computing Systems
montoyo@dlsi.ua.es

Resumen: El crecimiento exponencial de la información subjetiva en el marco de la Web 2.0 ha creado la necesidad de producir herramientas de Procesamiento del Lenguaje Natural que sean capaces de analizar y procesar estos datos para aplicaciones concretas. Estas herramientas requieren un entrenamiento con corpus anotados con este tipo de información a nivel muy detallado para poder capturar aquellos fenómenos lingüísticos que contienen una carga emotiva. El presente artículo describe *EmotiBlog*, un modelo detallado para la anotación de la subjetividad. Presentamos el proceso de creación y demostramos que aporta mejoras a los sistemas de aprendizaje automático. Para ello, empleamos distintos corpus que presentan textos de diversos géneros – una colección de noticias periodísticas en estilo indirecto, la colección de títulos de noticias anotados con la polaridad y emoción del SemEval 2007 (Tarea 14) e ISEAR, un corpus de expresiones reales de emociones. Además, demostramos que otros recursos pueden integrarse con *EmotiBlog*. Los resultados prueban que gracias a su estructura y parámetros de anotación, el modelo propuesto, *EmotiBlog,* proporciona ventajas considerables para el entrenamiento de sistemas que trabajan con minería de opiniones y detección de emoción.
Palabras clave: Esquema de anotación, subjetividad, corpus, Análisis de Sentimientos, Aprendizaje Automático.

Abstract: The exponential growth of the subjective information in the framework of the Web 2.0 has led to the need to create Natural Language Processing tools able to analyse and process such data for multiple practical applications. These applications require training on specifically annotated corpora, whose level of detail must be fine enough to capture the phenomena involved. This paper presents *EmotiBlog* – a fine-grained annotation scheme for subjectivity. We show the manner in which it is built and demonstrate the benefits it brings to the systems using it for training, through the experiments we carried out on opinion mining and emotion detection. We employ corpora of different textual genres –a set of annotated reported speech extracted from news articles, the set of news titles annotated with polarity and emotion from the *SemEval* 2007 (Task 14) and ISEAR, a corpus of real-life self-expressed emotion. We also show how the model built from the *EmotiBlog* annotations can be enhanced with external resources. The results demonstrate that EmotiBlog, through its structure and annotation paradigm, offers high quality training data for systems dealing both with opinion mining, as well as emotion detection.
Keywords: Annotation scheme, subjectivity, corpus, Sentiment Analysis, Machine Learning.

1 *Introduction*

The huge and rapid growth in the quantity of subjective data in the framework of the Web 2.0 created the need to develop new Natural Language Processing (NLP) tools for the treatment of such heterogeneous information in the new-textual genres. "*The State of the Blogosphere 2009*" survey[1] assesses that there is a growing influence of the blogosphere on a wide range of subjects. Moreover, due to the growing interest and usage of this text type, the subjective content of the Web is increasing on a

[1] http://technorati.com/

daily basis, becoming a reflection of people's points of view (Cui, Mittal and Datar, 2006). Blogs became a source of real-time and spontaneous information that can be exploited for a wide range of useful applications. At present, NLP tools and methods for processing objective information have better performance than the new ones the research community is creating for dealing with subjective content. The NLP task in charge of the treatment of subjective data is called Sentiment Analysis (SA). Subjectivity can be expressed in text by means of emotions, beliefs, views (a way of considering something)[2] or opinions, generally denominated "private states" (Uspensky, 1973), which are not open to verification (Wiebe, 1994). Starting from (Wiebe, Wilson and Cardie, 2005) we conceived an annotation model able to capture a wide range of key linguistic subjective elements, moving a step ahead to the mere polarity recognition. In particular, the experiments concern expressions of emotion, as a finer-grained analysis of affect in text and a subsequent task to Opinion Mining (OM) and classification. We carried out a series of evaluations focused on demonstrating that *EmotiBlog* represents a step forward to previous research in this field. In fact, its use allows for a finer-grained and more detailed learning of subjectivity expression models. We employ corpora of different textual genres (a set of annotated reported speech extracted from news articles, denominated JRC quotes) (Balahur et al., 2010), the set of news titles annotated with polarity and emotion from the SemEval 2007 Task No. 14 (Strapparava and Mihalcea, 2007), as well as a corpus of real-life self-expressed emotion entitled ISEAR (Scherer and Walbott, 1999). We subsequently show, through the quality of the results obtained, that *EmotiBlog*, through its structure and annotation paradigm, offers high quality training data for systems dealing both with OM, as well as emotion detection.

2 Motivation and contribution

Given the proven relevance of the Web 2.0 and its content, new methods and tools must be developed to effectively process subjective data. The opinionated content is in most cases complex to extract and classify employing only grammatical static rules. Expression of subjectivity is spontaneous and even if a large part of the content of blogs is quite formal, new means of expressivity can be encountered, such as the use of colloquialisms, sayings, collocations or anomalies in the use of punctuation. Nowadays, it is common that when taking a decision, people search for information and opinions expressed on the Web on their matter of interest and the final decision is influenced by the information found. At the same time, when using a product, people often write reviews on it, so that others can have a better idea of the performance of that product before purchasing it. Retrieving opinion information requires the discrimination of different discussion topics and subsequently their classification, according to the corresponding polarity. Determining points of view expressed in dialogues with the mixture of quotes and pastes from newspapers on a topic can, additionally, involve determining the persons involved and whether or not the opinion expressed is on the required topic or on a point previously made by another speaker. This difficult NLP problem requires the use of specialized data for system training and tuning to be gathered, annotated and tested within the different text spheres. At the present moment, these specialized resources are scarce or they are rather simplistically annotated or highly domain-dependent and most of them are created for English. The main contributions of this paper are to present and test the *EmotiBlog* annotation scheme and to assess the validity of the hypothesis that a finer-grained model can help to improve OM and emotion detection. To this aim, we evaluate the distinct features extracted from the *EmotiBlog* annotations on very different corpora. Subsequently, we analyse the weak points of the model, as well as the characteristics of different textual genres. Through the obtained results, we demonstrate the applicability of the created resources to different NLP tasks.

3 Related work

In recent years, different researchers have addressed the needs and possible methodologies involved in dealing with subjective information. The first approaches aimed at building lexical resources of affect, such as WordNet Affect (Strapparava and Valitutti, 2004), SentiWordNet (Esuli and Sebastiani, 2006), Micro-WNOP (Cerini et. Al, 2007) or "emotion triggers" (Balahur and Montoyo, 2009). These lexicons contain single words, whose polarity and emotions are not necessarily the ones annotated within the resource in a larger

[2] http://dictionary.cambridge.org/

context. The starting point of research in emotion is represented by (Wiebe, 1994), who centered the idea of subjectivity around that of private states, and set the benchmark for subjectivity analysis, in order to distinguish it from objective language. Subsequenly, the author proposed a method to annotate a corpus depending on these two aspects – MPQA (Wiebe, Wilson and Cardie, 2005). Further on, subsequent research has shown that this initial discrimination is crucial for the sentiment task, as part of Opinion Information Retrieval (last three editions of the TREC Blog tracks[3] competitions, the TAC 2008 competition[4]), Information Extraction (Riloff and Wiebe, 2003) and Question Answering (Stoyanov et al., 2004) systems. Once this discrimination is done, or in the case of texts containing only or mostly subjective language (such as e-reviews), Opinion Mining (OM) becomes a polarity classification task. Related work also includes customer review classification at a document level, sentiment classification using unsupervised methods (Turney, 2002), Machine Learning techniques (Pang and Lee, 2002), scoring of features (Dave, Lawrence and Pennock, 2003), using PMI, syntactic relations and other attributes with SVM (Mullen and Collier, 2004), sentiment classification considering rating scales (Pang and Lee, 2002), supervised and unsupervised methods (Chaovalit and Zhou, 2005) and semi-supervised learning (Goldberg and Zhou, 2006). Other research has been conducted in analysing sentiment at a sentence level using bootstrapping techniques (Riloff and Wiebe, 2003), considering gradable adjectives (Hatzivassiloglou and Wiebe, 2000), semi-supervised learning with the initial training some strong patterns and then applying NB or self-training (Wiebe and Riloff, 2005) finding strength of opinions (Wilson, Wiebe and Hwa, 2004) or summing up orientations of opinion words (Kim and Hovy, 2004), (Wilson and Wiebe, 2004). All these approaches concentrate on finding and classifying the polarity of opinion words, mostly adjectives, without considering modifiers or their context. Our research is focused on the creation of a linguistic resource that is a fine-grained annotation schema for emotion detection in non-traditional textual genres. Our schema allows a deeper analysis of subjective content, an adequate discrimination with the objective discourse and a study of the importance of the elements it contemplates.

[3] http://trec.nist.gov/data/blog.html

[4] http://www.nist.gov/tac/

4 Corpora

At present, blogs are extremely useful and a poll for discussion about any topic with the world. For this reason, the first corpus object of our study is a collection of blog posts extracted from the Web. The texts we selected have distinctive features, different from traditional textual ones. Bloggers can employ an informal language style - colloquialisms, emoticons, etc. to express their feelings. It is not rare to find a mix of sources in the same post since they usually mention some facts to express their opinion. In this case, source detection represents one of the most complex, as well as relevant tasks. We carried out a multilingual research, collecting texts in three languages: Spanish, Italian, and English, about three subjects of interest. The first one contains blog posts commenting upon the signing of the Kyoto Protocol against global warming, the second collection consists of blog entries about the Mugabe government in Zimbabwe, and finally we selected a series of blog posts about the 2008 USA presidential elections. For each of these topics, we have gathered 100 texts, summing up to a total of 30.000 words approximately for each language. The second corpus we employed for this research is a collection of 1592 quotes extracted from the news in April 2008. As a consequence they are about many different topics and in English (Balahur and Steinberger, 2009). Both of these corpora have been annotated with *EmotiBlog*.

5 EmotiBlog Annotation Model

EmotiBlog (Boldrini et al., 2009) is a fine-grained model for the subjectivity annotation in the context of new textual genres born with the Web 2.0. As mentioned above, it represents a step forward to previous research and it is focused on detecting the linguistic elements, which give text a subjective nature. This annotation model contemplates different levels of annotation: sentence and word level (Boldrini et al., 2009). The first distinction made in the model is between objective and subjective speech. The list below shows which attributes have to be contemplated for the elements we are labelling.

- Objective speech: annotator's confidence (high, medium law), comment (if necessary), source (writer) and target (discourse topic);

In some cases writers use rhetoric strategies to state something that is apparently objective but it is expressing a personal point of view. In order to be able to contemplate these cases we inserted in the model the following elements:

- Reader Interpretation: annotator's confidence, comment, level, emotion, phenomenon, polarity, source and target. It is employed for capturing the impression/feeling/reaction the reader has going through the intervention and what s/he can deduce from the piece of
- Author Interpretation: annotator's confidence, comment, level, emotion, phenomenon, polarity, source and target. This element is used to understand what we can deduce from the author (politic orientation, preferences) thanks the words and language s/he chooses.

For both objective and subjective speech, the annotator has to specify the nature of the sentence s/he is labelling:

- Phenomenon: annotator's confidence, comment, type. This element explains the nature of the sentence we are labelling. They can be collocation, saying, slang, title, and rhetoric. A saying is a well-known and wise statement, which often has a meaning, different from the simple meanings of the words it contains[5]; while a collocation is a word or phrase, which is frequently used with another word or phrase, in a way that sounds correct to native speakers, but might not be expected from the individual words' meanings[6].

In case the annotator is labelling a subjective sentence, the first thing is to label the entire sentence underlining its nature, using the following tag:

- Subjective speech: annotator's confidence, comment, level, emotion, phenomenon, polarity, source and target.

In case of a subjective sentence, the annotator has to detect the elements, which give the subjectivity shadow to the discourse. *EmotiBlog* contemplates the ones below:

- Adjective/Adverbs: annotator's confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target.
- Verbs: annotator's confidence, comment, level, emotion, phenomenon, polarity, mode, source and target.
- Nouns: annotator's confidence, comment, level, emotion, phenomenon, modifier/not, polarity, and source.
- Anaphora: annotator's confidence, comment, type, source and target. This element underlines the correference phenomena at a cross-post level. Usually, blog posts are like a multi-party conversation and thus this element can be useful to follow the discourse in case of multiple posts or when it is interrupted with other posts about a subtopic or related topic.
- Capital Letter: annotator's confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target. Bloggers generally produce a genuine and spontaneous language and it is frequent to find complete words that are meant as a sign of a special user attitude.
- Punctuation: annotator's confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target. This phenomenon is similar to the previous one. An exceptional use of punctuation could mean a special feeling of the writer.
- Emotions: annotator's confidence, comment, accept, anger, anticipation, anxiety, etc.

Regarding the list of emotions employed, we grouped all sentiments into subgroups to facilitate the evaluation process. Emotions of the same subgroup will have less impact when calculating the inter-annotation agreement. In order to make this subdivision proper and effective, we were inspired by (Scherer, 2005). We started form this classification, grouping sentiments into positive and negative, and we also divided them as high/low power control, obstructive/conductive and active/passive. Further on, we distributed the sentiments within our list into the Scherer slots, creating other smaller categories included in the abovementioned general ones.

6 Experiments and Evaluation

In order to evaluate the appropriateness of the *EmotiBlog* annotation scheme and to prove that the fine-grained level of annotation has a positive impact on the performance of the systems employing it as training, we carried out several experiments. Given that *EmotiBlog* contains annotations at a word level, as well as

[5] Definition according to the Cambridge Advanced Learner's Dictionary

[6] Definition according to the Cambridge Advanced Learner's Dictionary

for multi-word expressions and at a sentence level, and they are labelled with polarity, but also emotion, our experiments show how the annotated elements can be used as training for the opinion mining and polarity classification task, as well as for emotion detection. Moreover, since *EmotiBlog* labels the intensity level of the annotated elements, we performed a brief experiment on determining the sentiment intensity, measured on a three-level scale: low, medium and high. In order to perform these three different evaluations, we chose three different corpora. The first one is a collection of quotes (reported speech) from newspaper articles presented in (Balahur et al., 2010), enriched with the manual fine-grained annotation of *EmotiBlog*; the second one is the collection of newspaper titles in the test set of the SemEval 2007 task number 14 – Affective Text, the third one, is a corpus of self-reported emotional response – ISEAR (Scherer and Walbott, 1999), while the last one is the NTCIR 8 MOAT corpus. The intensity classification task is evaluated only on the second corpus, given that it is the only one in which scores between -100 and 0 and 0 and 100, respectively, are given for the polarity of the titles.

6.1 Creation of training models

For the OM and polarity classification task, we first extracted the Named Entities contained in the annotations using Lingpipe and united through a "_" all the tokens pertaining to the NE. All the annotations of punctuation signs that had a specific meaning together were also united under a single punctuation sign. After that, we processed the annotated data, using Minipar. We compute, for each word in a sentence, a series of features (some of these features are used in (Choi et al., 2005):

- the part of speech (POS)
- capitalization (if all letters are in capitals, if only the first letter is in capitals, and if it is a NE or not)
- opinionatedness/intensity/emotion - if the word is annotated as opinion word, its polarity, i.e. 1 and -1 if the word is positive or negative, respectively and 0 if it is not an opinion word, its intensity (1.2 or 3) and 0 if it is not a subjective word, its emotion (if it has, none otherwise)
- syntactic relatedness with other opinion word – if it is directly dependent of an opinion word or modifier (0 or 1), plus the polarity/intensity and emotion of this word (0 for all the components otherwise)
- role in 2-word, 3-word and 4-word annotations: opinionatedness, intensity and emotion of the other words contained in the annotation, direct dependency relations with them if they exist and 0 otherwise.

We compute the length of the longest sentence in *EmotiBlog*. The feature vector for each of the sentences contains the feature vectors of each of its words and 0s for the corresponding feature vectors of the words, which the current sentence has less than the longest annotated sentence. Finally, we add for each sentence as feature binary features for subjectivity and polarity, the value corresponding to the intensity of opinion and the general emotion. These feature vectors are fed into the Weka[7] SVM SMO Machine Learning (ML) algorithm and a model is created (*EmotiBlog I*). A second model (*EmotiBlog II*) is created by adding to the collection of single opinion and emotion words annotated in *EmotiBlog*, the Opinion Finder lexicon and the opinion words found in MicroWordNet, the General Inquirer resource and WordNet Affect. The idea behind this approach is to measure the impact that a larger set of opinion words would have on the results – but under the EmotiBlog annotation paradigm. That is – if we were to annotate more, would we gain or lose on accuracy?

6.2 Evaluation of models on test sets

With the purpose of evaluating the performance of the models extracted from the features of the annotations in *EmotiBlog*, we perform different tests. The first one regarded the evaluation of the polarity and intensity classification task using the *EmotiBlog* I and II constructed models on two test sets – the JRC quotes collection and the SemEval 2007 Task Number 14 test set. Since the quotes often contain more than a sentence, we consider the polarity and intensity of the entire quote as the most frequent result in each class, corresponding to its constituent sentences. Also, given the fact that the SemEval Affective Text headlines were given intensity values between -100 and 100, we mapped the values contained in the Gold Standard of the task into three categories: [-100, -67] is high (value 3 in intensity) and negative (value -1 in polarity), [-66, 34] medium negative and [33, 1] is low negative. The values between [1 and 100] are mapped in the same manner to the positive category. 0 was considered objective, so containing the value 0

[7] http://www.cs.waikato.ac.nz/ml/weka/

for intensity. The results are presented in Table 2 (the values I and II correspond to the models EmotiBlog I and EmotiBlog II):

Test Corpus	Evaluation type	Precision	Recall
JRC quotes I	Polarity	32.13	54.09
	Intensity	36.00	53.2
JRC quotes II	Polarity	36.4	51.00
	Intensity	38.7	57.81
SemEval I	Polarity	38.57	51.3
	Intensity	37.39	50.9
SemEval II	Polarity	35.8	58.68
	Intensity	32.3	50.4

Table 2. Results for polarity and intensity classification using the models built from the EmotiBlog annotations

The results shown in Table 2 show a significantly high improvement over the results obtained in the SemEval task in 2007. This is explainable, on the one hand, by the fact that systems performing the opinion task did not have at their disposal the lexical resources for opinion employed in the *EmotiBlog II* model, but also because of the fact that they did not use ML on a corpus comparable to *EmotiBlog* (as seen from the results obtained when using solely the *EmotiBlog I* corpus). Compared to the NTCIR 8 Multilingual Analysis Task this year, we obtained significant improvements in precision, with a recall that is comparable to most of the participating systems. In the second experiment, we tested the performance of emotion classification using the two models built using *EmotiBlog* on the three corpora – JRC quotes, SemEval 2007 Task No.14 test set and the ISEAR corpus. The JRC quotes are labelled using *EmotiBlog,* while the other two with a reduced set of emotions – 6 in the case of the SemEval data (joy, surprise, anger, fear, sadness, disgust) and 7 in ISEAR (joy, sadness, anger, fear, guilt, shame, disgust). Moreover, the SemEval data contains more than one emotion per title in the Gold Standard, therefore we consider as correct any of the classifications containing one of them. In order to unify the results and obtain comparable evaluations, we assessed the performance of the system using the alternative dimensional structures defined in Table 1. The ones not overlapping with the category of any of the 8 different emotions in SemEval and ISEAR are considered as "Other" and are not included either in the training, nor test set. The results of the evaluation are presented in Table 3. Again, the values I and II correspond to the models EmotiBlog I and II. The "Emotions" category contains the following emotions: joy, sadness, anger, fear, guilt, shame, disgust, surprise.

Test corpus	Evaluation type	Precision	Recall
JRC quotes I	Emotions	24.7	15.08
JRC quotes II	Emotions	33.65	18.98
SemEval I	Emotions	29.03	18.89
SemEval II	Emotions	32.98	18.45
ISEAR I	Emotions	22.31	15.01
ISEAR II	Emotions	25.62	17.83

Table 3. Results for emotion classification using the models built from the EmotiBlog annotations.

The best results for emotion detection were obtained for the "anger" category, where the precision was around 35 percent, for a recall of 19 percent. The worst results obtained were for the ISEAR category of "shame", where precision was around 12 percent, with a recall of 15 percent. We believe this is due to the fact that the latter emotion is a combination of more complex affective states and it can be easily misclassified to other categories of emotion. Moreover, from the error analysis we realized that many of the affective phenomena presented were more explicit in the case of texts expressing strong emotions such as "joy" and "anger", and were mostly related to common-sense interpretation of the facts presented in the weaker ones. As it can be seen in Table 3, results for the texts pertaining to the news category obtain better results, most of all news titles. This is due to the fact that they contain a few words and more direct and stronger emotional charge than direct speech. Finally, the error analysis showed that emotion that is directly reported by the persons experiencing is more "hidden", in the use of words carrying special signification or related to general human experience. This fact makes emotion detection in such texts a harder task. Nevertheless, the results in all corpora are comparable, showing that the approach is robust enough to handle different text types. The results obtained using the fine and coarse-grained annotations in *EmotiBlog* increased the performance of emotion detection as compared to the systems in the SemEval competition. Due to space limitations, they are not repeated here; for further reference, please see (Strapparava and Mihalcea, 2007).

6.3 Discussion

From the results obtained, we can see that this approach combining the features extracted from the *EmotiBlog* fine and coarse-grained annotation helps to balance between the results obtained for precision and recall. The impact of using additional resources with opinion words is that of increasing the recall of the system, at the cost of a slight drop in precision, which proves that the approach is robust enough so that additional knowledge sources can be added. Although the corpus is small, the results obtained show that the phenomena it captures is relevant in the OM task, not only for the blogosphere, but also for other text-types (newspaper articles, self-reported affect).

7 Conclusions

The exponential increase of the subjective information on the Web 2.0 originates the need of NLP able to process this data. In this paper we presented the procedure by which we compiled a multilingual corpus of blog posts on different topics of interest in three languages: Spanish, Italian and English. We explained the need to create a finer-grained annotation schema that can be used to improve the performance of subjectivity mining systems. Thus, we presented the new annotation model, *EmotiBlog* and justified the benefits of this fine-grained annotation schema, presenting the sources and the reasons taken into consideration when building up the corpus and its labelling. Furthermore, we addressed the presence of “copy and pastes” from news articles or other blogs, the frequent quotes. In order to solve this possible ambiguity the annotation model contemplates both the directly indicated source, as well as the anaphoric references at cross-document level. We performed several experiments on three corpora, aimed at finding and classifying both opinions and expressions of emotion. We demonstrated that the fine and coarse-grained levels of annotation that *EmotiBlog* contains offer important information on the structure of affective texts, leading to an improvement of the performance of systems trained on it. Although the *EmotiBlog* corpus is small, the results obtained are promising and show that the phenomena it captures are relevant in the OM task, not only for the blogosphere, but also for other textual-genres. It is well known that OM is an extremely challenging task and a young discipline, thus there is room for improvement above all to solve linguistic phenomena such as the correference resolution at a cross document level and temporal expression recognition. In addition to this, more experiments would need to be carried out to verify the complete robustness of *EmotiBlog*. Last but not least, our idea is to include the existing tools for a more effective semi-supervised annotation. After the training of the ML system we obtain automatically some markables to be validated by the annotator and the ideal option would be to connect these terms the system detects automatically with tools, mapping with an opinion lexicon based on WordNet (SentiWordNet, WordNet Affect, MicroWordNet), in order to automatically annotate all the synonyms and antonyms with the same or the opposite polarity respectively and assigning them some other elements contemplated into the *EmotiBlog* annotation schema. This would represent an important step forward for saving time during the annotation process and it will also assure a high quality labelling assured by human supervision.

Acknowledgements

This paper has been supported partially by Ministerio de Ciencia e Innovación - Spanish Government (grant no. TIN2009-13391-C04-01), and Conselleria d'Educación - Generalitat Valenciana (grant no. PROMETEO/2009/119 and ACOMP/2010/288.

References

Balahur A., Steinberger R., Kabadjov M., Zavarella V., van der Goot E., Halkia M., Pouliquen B., and Belyaeva J. 2010. Sentiment Analysis in the News. In Proceedings of LREC 2010.

Balahur A., Boldrini E., Montoyo A., Martínez-Barco P. 2009. A Comparative Study of Open Domain and Opinion Question Answering Systems for Factual and Opinionated Queries. In Proceedings of the Recent Advances in Natural Language Processing.

Balahur A., Montoyo A. 2008. Applying a Culture Dependent Emotion Triggers Database for Text Valence and Emotion Classification. In Proceedings of the AISB 2008 Symposium on Affective Language in Human and Machine, Aberdeen, Scotland.

Balahur A., Steinberger R., Rethinking Sentiment Analysis in the News: from Theory to Practice and back. In Proceeding of WOMSA 2009. Seville.

Balahur A., Boldrini E., Montoyo A., Martínez-Barco P. 2009. Summarizing Threads in Blogs Using Opinion Polarity. In Proceedings of ETTS workshop. RANLP. 2009.

Boldrini E., Balahur A., Martínez-Barco P., Montoyo A. 2009. EmotiBlog: a fine-grained model for emotion detection in non-traditional textual genres. In Proceedings of WOMSA. Seville, Spain.

Boldrini E., Fernández J., Gómez J.M., Martínez-Barco P. 2009. Machine Learning Techniques for Automatic Opinion Detection in Non-Traditional Textual Genres. In Proceedings of WOMSA 2009. Seville, Spain.

Chaovalit P, Zhou L. 2005. Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches. In Proceedings of HICSS-05.

Carletta J. 1996. Assessing agreement on classification task: the kappa statistic. Computational Linguistics, 22(2): 249–254.

Cui H., Mittal V., Datar M. 2006. Comparative Experiments on Sentiment Classification for Online Product Reviews. In Proceedings of the 21st National Conference on Artificial Intelligence AAAI.

Cerini S., Compagnoni V., Demontis A., Formentelli M., and Gandini G. 2007. Language resources and linguistic theory: Typology, second language acquisition. English linguistics (Forthcoming), chapter Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. Franco Angeli Editore, Milano, IT.

Dave K., Lawrence S., Pennock, D. "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews". In Proceedings of WWW-03. 2003.

Esuli A., Sebastiani F. 2006. SentiWordNet: A Publicly Available Resource for Opinion Mining. In Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2006, Genoa, Italy.

Goldberg A.B., Zhu J. 2006. Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. In HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing.

Hu M., Liu B. 2004. Mining Opinion Features in Customer Reviews. In Proceedings of Nineteenth National Conference on Artificial Intelligence AAAI.

Hatzivassiloglou V., Wiebe J. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In Proceedings of COLING.

Kim S.M., Hovy E. 2004. Determining the Sentiment of Opinions. In Proceedings of COLING.

Mullen T., Collier N. 2006. Sentiment Analysis Using Support Vector Machines with Diverse Information Sources. In Proceedings of EMNLP. 2004. Lin, W.H., Wilson, T., Wiebe, J., Hauptman, A. "Which Side are You On? Identifying Perspectives at the Document and Sentence Levels". In Proceedings of the Tenth Conference on Natural Language Learning CoNLL.2006.

Pang B., Lee L., Vaithyanathan S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP-02, the Conference on Empirical Methods in Natural Language Processing.

Riloff E., Wiebe J. 2003. Learning Extraction Patterns for Subjective Expressions. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing.

Scherer K. R. 2005. What are emotions? And how can they be measured? Social Science Information, 44(4), 693–727.

Stoyanov V. and Cardie C. 2006. Toward Opinion Summarization: Linking the Sources. COLING-ACL. Workshop on Sentiment and Subjectivity in Text.

Stoyanov V., Cardie C., Litman D., and Wiebe J. 2004. Evaluating an Opinion Annotation Scheme Using a New Multi-Perspective Question and Answer Corpus. AAAI Spring Symposium on Exploring Attitude and Affect in Text.

Strapparava and Mihalcea, 2007 - SemEval 2007 Task 14: Affective Text. In Proceedings of the ACL.

Turney P. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. ACL 2002: 417-424.

Uspensky B. 1973. A Poetics of Composition. University of California Press, Berkeley, California.

Wiebe J. M. 1994. Tracking point of view in narrative. Computational Linguistics, vol. 20, pp. 233–287.

Wiebe J., Wilson T. and Cardie C. 2005. Annotating expressions of opinions and emotions in language. Language Resources and Evaluation.

Wilson T., Wiebe J., Hwa R. 2004. Just how mad are you? Finding strong and weak opinion clauses. In: Proceedings of AAAI.