Multilingual Subjectivity Analysis Using Machine Translation

Carmen Banea and **Rada Mihalcea**
University of North Texas
carmenb@unt.edu, rada@cs.unt.edu

Janyce Wiebe
University of Pittsburgh
wiebe@cs.pitt.edu

Samer Hassan
University of North Texas
samer@unt.edu

Abstract

Although research in other languages is increasing, much of the work in subjectivity analysis has been applied to English data, mainly due to the large body of electronic resources and tools that are available for this language. In this paper, we propose and evaluate methods that can be employed to transfer a repository of subjectivity resources across languages. Specifically, we attempt to leverage on the resources available for English and, by employing machine translation, generate resources for subjectivity analysis in other languages. Through comparative evaluations on two different languages (Romanian and Spanish), we show that automatic translation is a viable alternative for the construction of resources and tools for subjectivity analysis in a new target language.

1 Introduction

We have seen a surge in interest towards the application of automatic tools and techniques for the extraction of opinions, emotions, and sentiments in text (*subjectivity*). A large number of text processing applications have already employed techniques for automatic subjectivity analysis, including automatic expressive text-to-speech synthesis (Alm et al., 2005), text semantic analysis (Wiebe and Mihalcea, 2006; Esuli and Sebastiani, 2006), tracking sentiment timelines in on-line forums and news (Lloyd et al., 2005; Balog et al., 2006), mining opinions from product reviews (Hu and Liu, 2004), and question answering (Yu and Hatzivassiloglou, 2003).

A significant fraction of the research work to date in subjectivity analysis has been applied to English, which led to several resources and tools available for this language. In this paper, we explore multiple paths that employ machine translation while leveraging on the resources and tools available for English, to automatically generate resources for subjectivity analysis for a new target language. Through experiments carried out with automatic translation and cross-lingual projections of subjectivity annotations, we try to answer the following questions.

First, assuming an English corpus manually annotated for subjectivity, can we use machine translation to generate a subjectivity-annotated corpus in the target language? Second, assuming the availability of a tool for automatic subjectivity analysis in English, can we generate a corpus annotated for subjectivity in the target language by using automatic subjectivity annotations of English text and machine translation? Finally, third, can these automatically generated resources be used to effectively train tools for subjectivity analysis in the target language?

Since our methods are particularly useful for languages with only a few electronic tools and resources, we chose to conduct our initial experiments on Romanian, a language with limited text processing resources developed to date. Furthermore, to validate our results, we carried a second set of experiments on Spanish. Note however that our methods do not make use of any target language specific knowledge, and thus they are applicable to any other language as long as a machine translation engine exists between the selected language and English.

2 Related Work

Research in sentiment and subjectivity analysis has received increasingly growing interest from the natural language processing community, particularly motivated by the widespread need for opinion-based applications, including product and movie reviews, entity tracking and analysis, opinion summarization, and others.

Much of the work in subjectivity analysis has been applied to English data, though work on other languages is growing: e.g., Japanese data are used in (Kobayashi et al., 2004; Suzuki et al., 2006; Takamura et al., 2006; Kanayama and Nasukawa, 2006), Chinese data are used in (Hu et al., 2005), and German data are used in (Kim and Hovy, 2006). In addition, several participants in the Chinese and Japanese Opinion Extraction tasks of NTCIR-6 (Kando and Evans, 2007) performed subjectivity and sentiment analysis in languages other than English.

In general, efforts on building subjectivity analysis tools for other languages have been hampered by the high cost involved in creating corpora and lexical resources for a new language. To address this gap, we focus on leveraging resources already developed for one language to derive subjectivity analysis tools for a new language. This motivates the direction of our research, in which we use machine translation coupled with cross-lingual annotation projections to generate the resources and tools required to perform subjectivity classification in the target language.

The work closest to ours is the one reported in (Mihalcea et al., 2007), where a bilingual lexicon and a manually translated parallel text are used to generate the resources required to build a subjectivity classifier in a new language. In that work, we found that the projection of annotations across parallel texts can be successfully used to build a corpus annotated for subjectivity in the target language. However, parallel texts are not always available for a given language pair. Therefore, in this paper we explore a different approach where, instead of relying on manually translated parallel corpora, we use machine translation to produce a corpus in the new language.

3 Machine Translation for Subjectivity Analysis

We explore the possibility of using machine translation to generate the resources required to build subjectivity annotation tools in a given target language. We focus on two main scenarios. First, assuming a corpus manually annotated for subjectivity exists in the source language, we can use machine translation to create a corpus annotated for subjectivity in the target language. Second, assuming a tool for automatic subjectivity analysis exists in the source language, we can use this tool together with machine translation to create a corpus annotated for subjectivity in the target language.

In order to perform a comprehensive investigation, we propose three experiments as described below. The first scenario, based on a corpus manually annotated for subjectivity, is exemplified by the first experiment. The second scenario, based on a corpus automatically annotated with a tool for subjectivity analysis, is subsequently divided into two experiments depending on the direction of the translation and on the dataset that is translated.

In all three experiments, we use English as a source language, given that it has both a corpus manually annotated for subjectivity (MPQA (Wiebe et al., 2005)) and a tool for subjectivity analysis (OpinionFinder (Wiebe and Riloff, 2005)).

3.1 Experiment One: Machine Translation of Manually Annotated Corpora

In this experiment, we use a corpus in the source language manually annotated for subjectivity. The corpus is automatically translated into the target language, followed by a projection of the subjectivity labels from the source to the target language. The experiment is illustrated in Figure 1.

We use the MPQA corpus (Wiebe et al., 2005), which is a collection of 535 English-language news articles from a variety of news sources manually annotated for subjectivity. Although the corpus was originally annotated at clause and phrase level, we use the sentence-level annotations associated with the dataset (Wiebe and Riloff, 2005). From the total of 9,700 sentences in this corpus, 55% of the sentences are labeled as subjective while the rest are objective. After the automatic translation of the cor-
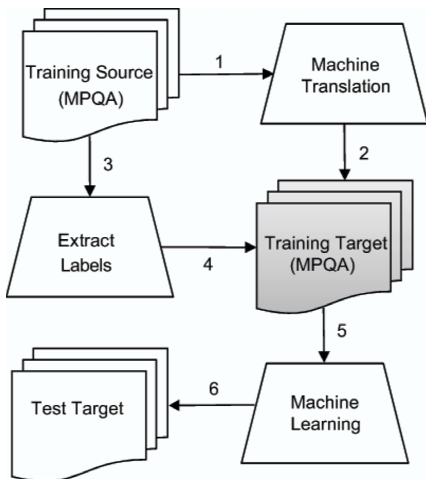
Figure 1: Experiment one: machine translation of manually annotated training data from source language into target language

Figure 2: Experiment two: machine translation of raw training data from source language into target language

pus and the projection of the annotations, we obtain a large corpus of 9,700 subjectivity-annotated sentences in the target language, which can be used to train a subjectivity classifier.

3.2 Experiment Two: Machine Translation of Source Language Training Data

In the second experiment, we assume that the only resources available are a tool for subjectivity annotation in the source language and a collection of raw texts, also in the source language. The source language text is automatically annotated for subjectivity and then translated into the target language. In this way, we produce a subjectivity annotated corpus that we can use to train a subjectivity annotation tool for the target language. Figure 2 illustrates this experiment.

In order to generate automatic subjectivity annotations, we use the OpinionFinder tool developed by (Wiebe and Riloff, 2005). OpinionFinder includes two classifiers. The first one is a rule-based high-precision classifier that labels sentences based on the presence of subjective clues obtained from a large lexicon. The second one is a high-coverage classifier that starts with an initial corpus annotated using the high-precision classifier, followed by several bootstrapping steps that increase the size of the lexicon and the coverage of the classifier. For most of our experiments we use the high-coverage classifier.
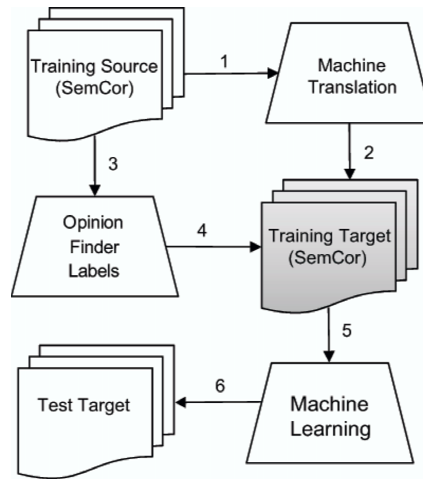
Table 1 shows the performance of the two Opinion-Finder classifiers as measured on the MPQA corpus (Wiebe and Riloff, 2005).

	P	R	F
high-precision	86.7	32.6	47.4
high-coverage	79.4	70.6	74.7

Table 1: Precision (P), Recall (R) and F-measure (F) for the two OpinionFinder classifiers, as measured on the MPQA corpus

As a raw corpus, we use a subset of the SemCor corpus (Miller et al., 1993), consisting of 107 documents with roughly 11,000 sentences. This is a balanced corpus covering a number of topics in sports, politics, fashion, education, and others. The reason for working with this collection is the fact that we also have a manual translation of the SemCor documents from English into one of the target languages used in the experiments (Romanian), which enables comparative evaluations of different scenarios (see Section 4).

Note that in this experiment the annotation of subjectivity is carried out on the original source language text, and thus expected to be more accurate than if it were applied on automatically translated text. However, the training data in the target language is produced by automatic translation, and thus likely to contain errors.

3.3 Experiment Three: Machine Translation of Target Language Training Data

The third experiment is similar to the second one, except that we reverse the direction of the translation. We translate raw text that is available in the target language into the source language, and then use a subjectivity annotation tool to label the automatically translated source language text. After the annotation, the labels are projected back into the target language, and the resulting annotated corpus is used to train a subjectivity classifier. Figure 3 illustrates this experiment.
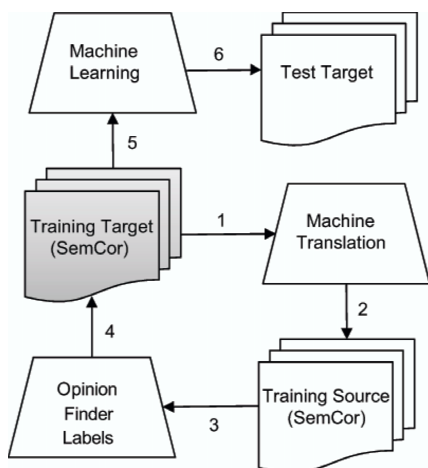

Figure 3: Experiment three: machine translation of raw training data from target language into source language

As before, we use the high-coverage classifier available in OpinionFinder, and the SemCor corpus. We use a manual translation of this corpus available in the target language.

In this experiment, the subjectivity annotations are carried out on automatically generated source text, and thus expected to be less accurate. However, since the training data was originally written in the target language, it is free of translation errors, and thus training carried out on this data should be more robust.

3.4 Upper bound: Machine Translation of Target Language Test Data

For comparison purposes, we also propose an experiment which plays the role of an upper bound on the methods described so far. This experiment involves the automatic translation of the test data from the target language into the source language. The source language text is then annotated for subjectivity using OpinionFinder, followed by a projection of the resulting labels back into the target language.

Unlike the previous three experiments, in this experiment we only generate subjectivity-annotated *resources*, and we do not build and evaluate a standalone subjectivity analysis *tool* for the target language. Further training of a machine learning algorithm, as in experiments two and three, is required in order to build a subjectivity analysis tool. Thus, this fourth experiment is an evaluation of the *resources* generated in the target language, which represents an upper bound on the performance of any machine learning algorithm that would be trained on these resources. Figure 4 illustrates this experiment.
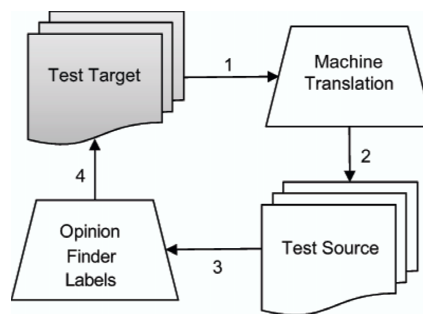

Figure 4: Upper bound: machine translation of test data from target language into source language

4 Evaluation and Results

Our initial evaluations are carried out on Romanian. The performance of each of the three methods is evaluated using a dataset manually annotated for subjectivity. To evaluate our methods, we generate a Romanian training corpus annotated for subjectivity on which we train a subjectivity classifier, which is then used to label the test data.

We evaluate the results against a gold-standard corpus consisting of 504 Romanian sentences manually annotated for subjectivity. These sentences represent the manual translation into Romanian of a small subset of the SemCor corpus, which was removed from the training corpora used in experiments two and three. This is the same evaluation dataset as used in (Mihalcea et al., 2007). Two Romanian native speakers annotated the sentences individually, and the differences were adjudicated

through discussions. The agreement of the two annotators is 0.83% ($\kappa = 0.67$); when the uncertain annotations are removed, the agreement rises to 0.89 ($\kappa = 0.77$). The two annotators reached consensus on all sentences for which they disagreed, resulting in a gold standard dataset with 272 (54%) subjective sentences and 232 (46%) objective sentences. More details about this dataset are available in (Mihalcea et al., 2007).

In order to learn from our annotated data, we experiment with two different classifiers, Naïve Bayes and support vector machines (SVM), selected for their performance and diversity of learning methodology. For Naïve Bayes, we use the multinomial model (McCallum and Nigam, 1998) with a threshold of 0.3. For SVM (Joachims, 1998), we use the LibSVM implementation (Fan et al., 2005) with a linear kernel.

The automatic translation of the MPQA and of the SemCor corpus was performed using Language Weaver,[1] a commercial statistical machine translation software. The resulting text was post-processed by removing diacritics, stopwords and numbers. For training, we experimented with a series of weighting schemes, yet we only report the results obtained for binary weighting, as it had the most consistent behavior.

The results obtained by running the three experiments on Romanian are shown in Table 2. The baseline on this data set is 54.16%, represented by the percentage of sentences in the corpus that are subjective, and the upper bound (UB) is 71.83%, which is the accuracy obtained under the scenario where the test data is translated into the source language and then annotated using the high-coverage OpinionFinder tool.

Perhaps not surprisingly, the SVM classifier outperforms Naïve Bayes by 2% to 6%, implying that SVM may be better fitted to lessen the amount of noise embedded in the dataset and provide more accurate classifications.

The first experiment, involving the automatic translation of the MPQA corpus enhanced with manual annotations for subjectivity at sentence level, does not seem to perform well when compared to the experiments in which automatic subjectivity classi-

Romanian				
Exp	Classifier	P	R	F
E1	Naïve Bayes	60.91	60.91	60.91
	SVM	66.07	66.07	66.07
E2	Naïve Bayes	63.69	63.69	63.69
	SVM	69.44	69.44	69.44
E3	Naïve Bayes	65.87	65.87	65.87
	SVM	67.86	67.86	67.86
UB	OpinionFinder	71.83	71.83	71.83

Table 2: Precision (P), Recall (R) and F-measure (F) for Romanian experiments

fication is used. This could imply that a classifier cannot be so easily trained on the cues that humans use to express subjectivity, especially when they are not overtly expressed in the sentence and thus can be lost in the translation. Instead, the automatic annotations produced with a rule-based tool (OpinionFinder), relying on overt mentions of words in a subjectivity lexicon, seems to be more robust to translation, further resulting in better classification results. To exemplify, consider the following subjective sentence from the MPQA corpus, which does not include overt clues of subjectivity, but was annotated as subjective by the human judges because of the structure of the sentence: *It is the Palestinians that are calling for the implementation of the agreements, understandings, and recommendations pertaining to the Palestinian-Israeli conflict.*

We compare our results with those obtained by a previously proposed method that was based on the manual translation of the SemCor subjectivity-annotated corpus. In (Mihalcea et al., 2007), we used the manual translation of the SemCor corpus into Romanian to form an English-Romanian parallel data set. The English side was annotated using the Opinion Finder tool, and the subjectivity labels were projected on the Romanian text. A Naïve Bayes classifier was then trained on the subjectivity annotated Romanian corpus and tested on the same gold standard as used in our experiments. Table 3 shows the results obtained in those experiments by using the high-coverage OpinionFinder classifier.

Among our experiments, experiments two and three are closest to those proposed in (Mihalcea et al., 2007). By using machine translation, from

OpinionFinder classifier	P	R	F
high-coverage	67.85	67.85	67.85

Table 3: Precision (P), Recall (R) and F-measure (F) for subjectivity analysis in Romanian obtained by using an English-Romanian parallel corpus

English into Romanian (experiment two) or Romanian into English (experiment three), and annotating this dataset with the high-coverage OpinionFinder classifier, we obtain an F-measure of 63.69%, and 65.87% respectively, using Naïve Bayes (the same machine learning classifier as used in (Mihalcea et al., 2007)). This implies that at most 4% in F-measure can be gained by using a parallel corpus as compared to an automatically translated corpus, further suggesting that machine translation is a viable alternative to devising subjectivity classification in a target language leveraged on the tools existent in a source language.

As English is a language with fewer inflections when compared to Romanian, which accommodates for gender and case as a suffix to the base form of a word, the automatic translation into English is closer to a human translation (experiment three). Therefore labeling this data using the OpinionFinder tool and projecting the labels onto a fully inflected human-generated Romanian text provides more accurate classification results, as compared to a setup where the training is carried out on machine-translated Romanian text (experiment two).
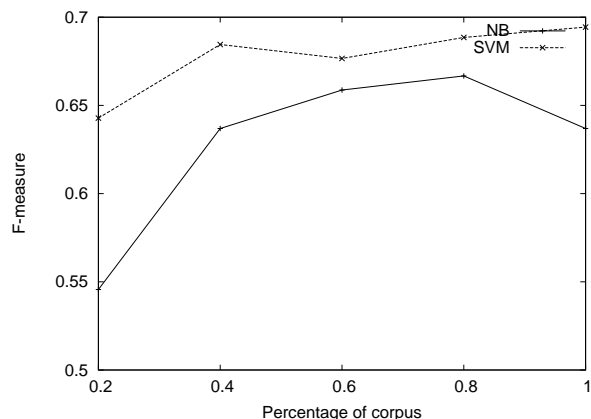

Figure 5: Experiment two: Machine learning F-measure over an incrementally larger training set

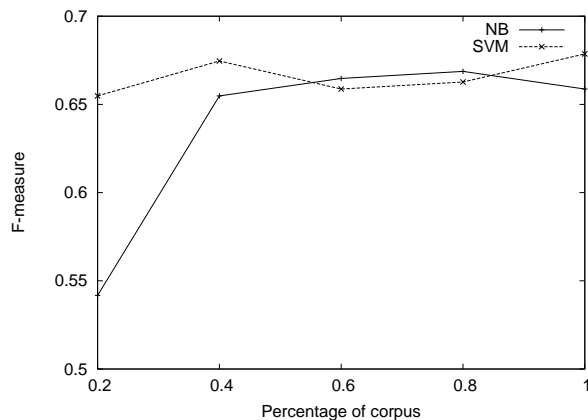We also wanted to explore the impact that the cor-

Figure 6: Experiment three: Machine learning F-measure over an incrementally larger training set

pus size may have on the accuracy of the classifiers. We re-ran experiments two and three with 20% corpus size increments at a time (Figures 5 and 6). It is interesting to note that a corpus of approximately 6000 sentences is able to achieve a high enough F-measure (around 66% for both experiments) to be considered viable for training a subjectivity classifier. Also, at a corpus size over 10,000 sentences, the Naïve Bayes classifier performs worse than SVM, which displays a directly proportional trend between the number of sentences in the data set and the observed F-measure. This trend could be explained by the fact that the SVM classifier is more robust with regard to noisy data, when compared to Naïve Bayes.

5 Portability to Other Languages

To test the validity of the results on other languages, we ran a portability experiment on Spanish.

To build a test dataset, a native speaker of Spanish translated the gold standard of 504 sentences into Spanish. We maintain the same subjectivity annotations as for the Romanian dataset. To create the training data required by the first two experiments, we translate both the MPQA corpus and the Sem-Cor corpus into Spanish using the Google Translation service,[2] a publicly available machine translation engine also based on statistical machine translation. We were therefore able to implement all the experiments but the third, which would have required

---

[2]http://www.google.com/translate_t

a manually translated version of the SemCor corpus. Although we could have used a Spanish text to carry out a similar experiment, due to the fact that the dataset would have been different, the results would not have been directly comparable.

The results of the two experiments exploring the portability to Spanish are shown in Table 4. Interestingly, all the figures are higher than those obtained for Romanian. We assume this occurs because Spanish is one of the six official United Nations languages, and the Google translation engine is using the United Nations parallel corpus to train their translation engine, therefore implying that a better quality translation is achieved as compared to the one available for Romanian. We can therefore conclude that the more accurate the translation engine, the more accurately the subjective content is translated, and therefore the better the results. As it was the case for Romanian, the SVM classifier produces the best results, with absolute improvements over the Naïve Bayes classifier ranging from 0.2% to 3.5%.

Since the Spanish automatic translation seems to be closer to a human-quality translation, we are not surprised that this time the first experiment is able to generate a more accurate training corpus as compared to the second experiment. The MPQA corpus, since it is manually annotated and of better quality, has a higher chance of generating a more reliable data set in the target language. As in the experiments on Romanian, when performing automatic translation of the test data, we obtain the best results with an F-measure of 73.41%, which is also the upper bound on our proposed experiments.

Spanish				
Exp	Classifier	P	R	F
E1	Naïve Bayes	65.28	65.28	65.28
	SVM	68.85	68.85	68.85
E2	Naïve Bayes	62.50	62.50	62.50
	SVM	62.70	62.70	62.70
UB	OpinionFinder	73.41	73.41	73.41

Table 4: Precision (P), Recall (R) and F-measure (F) for Spanish experiments

6 Discussion

Based on our experiments, we can conclude that machine translation offers a viable approach to generating resources for subjectivity annotation in a given target language. The results suggest that either a manually annotated dataset or an automatically annotated one can provide sufficient leverage towards building a tool for subjectivity analysis.

Since the use of parallel corpora (Mihalcea et al., 2007) requires a large amount of manual labor, one of the reasons behind our experiments was to asses the ability of machine translation to transfer subjective content into a target language with minimal effort. As demonstrated by our experiments, machine translation offers a viable alternative in the construction of resources and tools for subjectivity classification in a new target language, with only a small decrease in performance as compared to the case when a parallel corpus is available and used.

To gain further insights, two additional experiments were performed. First, we tried to isolate the role played by the quality of the subjectivity annotations in the source-language for the cross-lingual projections of subjectivity. To this end, we used the high-precision OpinionFinder classifier to annotate the English datasets. As shown in Table 1, this classifier has higher precision but lower recall as compared to the high-coverage classifier we used in our previous experiments. We re-ran the second experiment, this time trained on the 3,700 sentences that were classified by the OpinionFinder high-precision classifier as either subjective or objective. For Romanian, we obtained an F-measure of 69.05%, while for Spanish we obtained an F-measure of 66.47%.

Second, we tried to isolate the role played by language-specific clues of subjectivity. To this end, we decided to set up an experiment which, by comparison, can suggest the degree to which the languages are able to accommodate specific markers for subjectivity. First, we trained an English classifier using the SemCor training data automatically annotated for subjectivity with the OpinionFinder high-coverage tool. The classifier was then applied to the English version of the manually labeled test data set (the gold standard described in Section 4). Next, we ran a similar experiment on Romanian, using a classifier trained on the Romanian version of the same

SemCor training data set, annotated with subjectivity labels projected from English. The classifier was tested on the same gold standard data set. Thus, the two classifiers used the same training data, the same test data, and the same subjectivity annotations, the only difference being the language used (English or Romanian).

The results for these experiments are compiled in Table 5. Interestingly, the experiment conducted on Romanian shows an improvement of 3.5% to 9.5% over the results obtained on English, which indicates that subjective content may be easier to learn in Romanian versus English. The fact that Romanian verbs are inflected for mood (such as indicative, conditional, subjunctive, presumptive), enables an automatic classifier to identify additional subjective markers in text. Some moods such as conditional and presumptive entail human judgment, and therefore allow for clear subjectivity annotation. Moreover, Romanian is a highly inflected language, accommodating for forms of various words based on number, gender, case, and offering an explicit lexicalization of formality and politeness. All these features may have a cumulative effect in allowing for better classification. At the same time, English entails minimal inflection when compared to other Indo-European languages, as it lacks both gender and adjective agreement (with very few notable exceptions such as *beautiful girl* and *handsome boy*). Verb moods are composed with the aid of modals, while tenses and expressions are built with the aid of auxiliary verbs. For this reason, a machine learning algorithm may not be able to identify the same amount of information on subjective content in an English versus a Romanian text. It is also interesting to note that the labeling of the training set was performed using a subjectivity classifier developed for English, which takes into account a large, human-annotated, subjectivity lexicon also developed for English. One would have presumed that any classifier trained on this annotated text would therefore provide the best results in English. Yet, as explained earlier, this was not the case.

7 Conclusion

In this paper, we explored the use of machine translation for creating resources and tools for subjec-

Exp	Classifier	P	R	F
En	Naïve Bayes	60.32	60.32	60.32
	SVM	60.32	60.32	60.32
Ro	Naïve Bayes	67.85	67.85	67.85
	SVM	69.84	69.84	69.84

Table 5: Precision (P), Recall (R) and F-measure (F) for identifying language specific information

tivity analysis in other languages, by leveraging on the resources available in English. We introduced and evaluated three different approaches to generate subjectivity annotated corpora in a given target language, and exemplified the technique on Romanian and Spanish.

The experiments show promising results, as they are comparable to those obtained using manually translated corpora. While the quality of the translation is a factor, machine translation offers an efficient and effective alternative in capturing the subjective semantics of a text, coming within 4% F-measure as compared to the results obtained using human translated corpora.

In the future, we plan to explore additional language-specific clues, and integrate them into the subjectivity classifiers. As shown by some of our experiments, Romanian seems to entail more subjectivity markers compared to English, and this factor motivates us to further pursue the use of language-specific clues of subjectivity.

Our experiments have generated corpora of about 20,000 sentences annotated for subjectivity in Romanian and Spanish, which are available for download at http://lit.csci.unt.edu/index.php/Downloads, along with the manually annotated data sets.

References

C. Ovesdotter Alm, D. Roth, and R. Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of the Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, pages 347–354, Vancouver, Canada.

K. Balog, G. Mishne, and M. de Rijke. 2006. Why are they excited? identifying and explaining spikes in blog mood levels. In *Proceedings of the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2006)*.

A. Esuli and F. Sebastiani. 2006. Determining term subjectivity and term orientation for opinion mining. In *Proceedings the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pages 193–200, Trento, IT.

R. Fan, P. Chen, and C. Lin. 2005. Working set selection using the second order information for training svm. *Journal of Machine Learning Research*, 6:1889–1918.

M. Hu and B. Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD 2004)*, pages 168–177, Seattle, Washington.

Y. Hu, J. Duan, X. Chen, B. Pei, and R. Lu. 2005. A new method for sentiment classification in text retrieval. In *IJCNLP*, pages 1–9.

T. Joachims. 1998. Text categorization with Support Vector Machines: learning with mny relevant features. In *Proceedings of the European Conference on Machine Learning*, pages 137–142.

H. Kanayama and T. Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*, pages 355–363, Sydney, Australia.

N. Kando and D. Kirk Evans, editors. 2007. *Proceedings of the Sixth NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access*, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan, May. National Institute of Informatics.

S.-M. Kim and E. Hovy. 2006. Identifying and analyzing judgment opinions. In *Proceedings of the Human Language Technology Conference of the NAACL*, pages 200–207, New York, New York.

N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima. 2004. Collecting evaluative expressions for opinion extraction. In *Proceedings of the 1st International Joint Conference on Natural Language Processing (IJCNLP-04)*.

L. Lloyd, D. Kechagias, and S. Skiena. 2005. Lydia: A system for large-scale news analysis. In *String Processing and Information Retrieval (SPIRE 2005)*.

A. McCallum and K. Nigam. 1998. A comparison of event models for Naive Bayes text classification. In *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*.

R. Mihalcea, C. Banea, and J. Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of the Association for Computational Linguistics*, Prague, Czech Republic.

G. Miller, C. Leacock, T. Randee, and R. Bunker. 1993. A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, Plainsboro, New Jersey.

Y. Suzuki, H. Takamura, and M. Okumura. 2006. Application of semi-supervised learning to evaluative expression classification. In *Proceedings of the 7th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2006)*, pages 502–513, Mexico City, Mexico.

H. Takamura, T. Inui, and M. Okumura. 2006. Latent variable models for semantic orientations of phrases. In *Proceedings of the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, Trento, Italy.

J. Wiebe and R. Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of COLING-ACL 2006*.

J. Wiebe and E. Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *Proceedings of the 6th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2005) (invited paper)*, Mexico City, Mexico.

J. Wiebe, T. Wilson, and C. Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210.

H. Yu and V. Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2003)*, pages 129–136, Sapporo, Japan.