Learning Multilingual Subjective Language via Cross-Lingual Projections

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

Janyce Wiebe
Department of Computer Science
University of Pittsburgh
wiebe@cs.pitt.edu

Abstract

This paper explores methods for generating subjectivity analysis resources in a new language by leveraging on the tools and resources available in English. Given a bridge between English and the selected target language (e.g., a bilingual dictionary or a parallel corpus), the methods can be used to rapidly create tools for subjectivity analysis in the new language.

1 Introduction

There is growing interest in the automatic extraction of opinions, emotions, and sentiments in text (*subjectivity*), to provide tools and support for various natural language processing applications. Most of the research to date has focused on English, which is mainly explained by the availability of resources for subjectivity analysis, such as lexicons and manually labeled corpora.

In this paper, we investigate methods to automatically generate resources for subjectivity analysis for a new target language by leveraging on the resources and tools available for English, which in many cases took years of work to complete. Specifically, through experiments with cross-lingual projection of subjectivity, we seek answers to the following questions.

First, can we derive a subjectivity lexicon for a new language using an existing English subjectivity lexicon and a bilingual dictionary? Second, can we derive subjectivity-annotated corpora in a new language using existing subjectivity analysis tools for English and a parallel corpus? Finally, third, can we build tools for subjectivity analysis for a new target language by relying on these automatically generated resources?

We focus our experiments on Romanian, selected as a representative of the large number of languages that have only limited text processing resources developed to date. Note that, although we work with Romanian, the methods described are applicable to any other language, as in these experiments we (purposely) do not use any language-specific knowledge of the target language. Given a bridge between English and the selected target language (e.g., a bilingual dictionary or a parallel corpus), the methods can be applied to other languages as well.

After providing motivations, we present two approaches to developing sentence-level subjectivity classifiers for a new target language. The first uses a subjectivity lexicon translated from an English one. The second uses an English subjectivity classifier and a parallel corpus to create target-language training data for developing a statistical classifier.

2 Motivation

Automatic subjectivity analysis methods have been used in a wide variety of text processing applications, such as tracking sentiment timelines in online forums and news (Lloyd et al., 2005; Balog et al., 2006), review classification (Turney, 2002; Pang et al., 2002), mining opinions from product reviews (Hu and Liu, 2004), automatic expressive text-to-speech synthesis (Alm et al., 2005), text semantic analysis (Wiebe and Mihalcea, 2006; Esuli and Sebastiani, 2006), and question answering (Yu and Hatzivassiloglou, 2003).

Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 976–983,
Prague, Czech Republic, June 2007.

While much recent work in subjectivity analysis focuses on *sentiment* (a type of subjectivity, namely positive and negative emotions, evaluations, and judgments), we opt to focus on recognizing subjectivity in general, for two reasons.

First, even when sentiment is the desired focus, researchers in sentiment analysis have shown that a two-stage approach is often beneficial, in which subjective instances are distinguished from objective ones, and then the subjective instances are further classified according to polarity (Yu and Hatzivassiloglou, 2003; Pang and Lee, 2004; Wilson et al., 2005; Kim and Hovy, 2006). In fact, the problem of distinguishing subjective versus objective instances has often proved to be more difficult than subsequent polarity classification, so improvements in subjectivity classification promise to positively impact sentiment classification. This is reported in studies of manual annotation of phrases (Takamura et al., 2006), recognizing contextual polarity of expressions (Wilson et al., 2005), and sentiment tagging of words and word senses (Andreevskaia and Bergler, 2006; Esuli and Sebastiani, 2006).

Second, an NLP application may seek a wide range of types of subjectivity attributed to a person, such as their motivations, thoughts, and speculations, in addition to their positive and negative sentiments. For instance, the opinion tracking system Lydia (Lloyd et al., 2005) gives separate ratings for subjectivity and sentiment. These can be detected with subjectivity analysis but not by a method focused only on sentiment.

There is world-wide interest in text analysis applications. While work on subjectivity analysis in other languages is growing (e.g., Japanese data are used in (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)), much of the work in subjectivity analysis has been applied to English data. Creating corpora and lexical resources for a new language is very time consuming. In general, we would like to leverage resources already developed for one language to more rapidly create subjectivity analysis tools for a new one. This motivates our exploration and use of cross-lingual lexicon translations and annotation projections.

Most if not all work on subjectivity analysis has been carried out in a monolingual framework. We are not aware of multi-lingual work in subjectivity analysis such as that proposed here, in which subjectivity analysis resources developed for one language are used to support developing resources in another.

3 A Lexicon-Based Approach

Many subjectivity and sentiment analysis tools rely on manually or semi-automatically constructed lexicons (Yu and Hatzivassiloglou, 2003; Riloff and Wiebe, 2003; Kim and Hovy, 2006). Given the success of such techniques, the first approach we take to generating a target-language subjectivity classifier is to create a subjectivity lexicon by translating an existing source language lexicon, and then build a classifier that relies on the resulting lexicon.

Below, we describe the translation process and discuss the results of an annotation study to assess the quality of the translated lexicon. We then describe and evaluate a lexicon-based target-language classifier.

3.1 Translating a Subjectivity Lexicon

The subjectivity lexicon we use is from OpinionFinder (Wiebe and Riloff, 2005), an English subjectivity analysis system which, among other things, classifies sentences as subjective or objective. The lexicon was compiled from manually developed resources augmented with entries learned from corpora. It contains 6,856 unique entries, out of which 990 are multi-word expressions. The entries in the lexicon have been labeled for part of speech, and for reliability – those that appear most often in subjective contexts are *strong* clues of subjectivity, while those that appear less often, but still more often than expected by chance, are labeled *weak*.

To perform the translation, we use two bilingual dictionaries. The first is an authoritative English-Romanian dictionary, consisting of 41,500 entries,[1] which we use as the main translation resource for the lexicon translation. The second dictionary, drawn from the Universal Dictionary download site (UDP, 2007) consists of 4,500 entries written largely by Web volunteer contributors, and thus is not error free. We use this dictionary only for those entries that do not appear in the main dictionary.

[1] Unique English entries, each with multiple Romanian translations.

There were several challenges encountered in the translation process. First, although the English subjectivity lexicon contains inflected words, we must use the lemmatized form in order to be able to translate the entries using the bilingual dictionary. However, words may lose their subjective meaning once lemmatized. For instance, the inflected form of *memories* becomes *memory*. Once translated into Romanian (as *memorie*), its main meaning is objective, referring to the power of retaining information as in *Iron supplements may improve a woman's memory*.

Second, neither the lexicon nor the bilingual dictionary provides information on the sense of the individual entries, and therefore the translation has to rely on the most probable sense in the target language. Fortunately, the bilingual dictionary lists the translations in reverse order of their usage frequencies. Nonetheless, the ambiguity of the words and the translations still seems to represent an important source of error. Moreover, the lexicon sometimes includes identical entries expressed through different parts of speech, e.g., *grudge* has two separate entries, for its noun and verb roles, respectively. On the other hand, the bilingual dictionary does not make this distinction, and therefore we have again to rely on the "most frequent" heuristic captured by the translation order in the bilingual dictionary.

Finally, the lexicon includes a significant number (990) of multi-word expressions that pose translation difficulties, sometimes because their meaning is idiomatic, and sometimes because the multi-word expression is not listed in the bilingual dictionary and the translation of the entire phrase is difficult to reconstruct from the translations of the individual words. To address this problem, when a translation is not found in the dictionary, we create one using a word-by-word approach. These translations are then validated by enforcing that they occur at least three times on the Web, using counts collected from the AltaVista search engine. The multi-word expressions that are not validated in this process are discarded, reducing the number of expressions from an initial set of 990 to a final set of 264.

The final subjectivity lexicon in Romanian contains 4,983 entries. Table 1 shows examples of entries in the Romanian lexicon, together with their corresponding original English form. The table also shows the reliability of the expression (*weak* or *strong*) and the part of speech – attributes that are provided in the English subjectivity lexicon.

Romanian	English	attributes
înfrumuseţa	beautifying	strong, verb
notabil	notable	weak, adj
plin de regret	full of regrets	strong, adj
sclav	slaves	weak, noun

Table 1: Examples of entries in the Romanian subjectivity lexicon

Manual Evaluation.

We want to assess the quality of the translated lexicon, and compare it to the quality of the original English lexicon. The English subjectivity lexicon was evaluated in (Wiebe and Riloff, 2005) against a corpus of English-language news articles manually annotated for subjectivity (the *MPQA* corpus (Wiebe et al., 2005)). According to this evaluation, 85% of the instances of the clues marked as *strong* and 71.5% of the clues marked as *weak* are in subjective sentences in the MPQA corpus.

Since there is no comparable Romanian corpus, an alternate way to judge the subjectivity of a Romanian lexicon entry is needed.

Two native speakers of Romanian annotated the subjectivity of 150 randomly selected entries. Each annotator independently read approximately 100 examples of each drawn from the Web, including a large number from news sources. The subjectivity of a word was consequently judged in the contexts where it most frequently appears, accounting for its most frequent meanings on the Web.

The tagset used for the annotations consists of *S(ubjective)*, *O(bjective)*, and *B(oth)*. A *W(rong)* label is also used to indicate a wrong translation. Table 2 shows the contingency table for the two annotators' judgments on this data.

	S	O	B	W	Total
S	53	6	9	0	68
O	1	27	1	0	29
B	5	3	18	0	26
W	0	0	0	27	27
Total	59	36	28	27	150

Table 2: Agreement on 150 entries in the Romanian lexicon

Without counting the wrong translations, the agreement is measured at 0.80, with a Kappa κ =

0.70, which indicates consistent agreement. After the disagreements were reconciled through discussions, the final set of 123 correctly translated entries does include 49.6% (61) subjective entries, but fully 23.6% (29) were found in the study to have primarily objective uses (the other 26.8% are mixed).

Thus, this study suggests that the Romanian subjectivity clues derived through translation are less reliable than the original set of English clues. In several cases, the subjectivity is lost in the translation, mainly due to word ambiguity in either the source or target language, or both. For instance, the word *fragile* correctly translates into Romanian as *fragil*, yet this word is frequently used to refer to breakable objects, and it loses its subjective meaning of *delicate*. Other words, such as *one-sided*, completely lose subjectivity once translated, as it becomes in Romanian *cu o singura latură*, meaning *with only one side* (as of objects).

Interestingly, the reliability of clues in the English lexicon seems to help preserve subjectivity. Out of the 77 entries marked as *strong*, 11 were judged to be objective in Romanian (14.3%), compared to 14 objective Romanian entries obtained from the 36 *weak* English clues (39.0%).

3.2 Rule-based Subjectivity Classifier Using a Subjectivity Lexicon

Starting with the Romanian lexicon, we developed a lexical classifier similar to the one introduced by (Riloff and Wiebe, 2003). At the core of this method is a high-precision subjectivity and objectivity classifier that can label large amounts of raw text using only a subjectivity lexicon. Their method is further improved with a bootstrapping process that learns extraction patterns. In our experiments, however, we apply only the rule-based classification step, since the extraction step cannot be implemented without tools for syntactic parsing and information extraction not available in Romanian.

The classifier relies on three main heuristics to label subjective and objective sentences: (1) if two or more strong subjective expressions occur in the same sentence, the sentence is labeled *Subjective*; (2) if no strong subjective expressions occur in a sentence, and at most two weak subjective expressions occur in the previous, current, and next sentence combined, then the sentence is labeled *Objective*; (3) otherwise, if none of the previous rules apply, the sentence is labeled *Unknown*.

The quality of the classifier was evaluated on a Romanian gold-standard corpus annotated for subjectivity. Two native Romanian speakers (Ro_1 and Ro_2) manually annotated the subjectivity of the sentences of five randomly selected documents (504 sentences) from the Romanian side of an English-Romanian parallel corpus, according to the annotation scheme in (Wiebe et al., 2005). Agreement between annotators was measured, and then their differences were adjudicated. The baseline on this data set is 54.16%, which can be obtained by assigning a default *Subjective* label to all sentences. (More information about the corpus and annotations are given in Section 4 below, where agreement between English and Romanian aligned sentences is also assessed.)

As mentioned earlier, due to the lexicon projection process that is performed via a bilingual dictionary, the entries in our Romanian subjectivity lexicon are in a lemmatized form. Consequently, we also lemmatize the gold-standard corpus, to allow for the identification of matches with the lexicon. For this purpose, we use the Romanian lemmatizer developed by Ion and Tufiş (Ion, 2007), which has an estimated accuracy of 98%.[2]

Table 3 shows the results of the rule-based classifier. We show the precision, recall, and F-measure independently measured for the subjective, objective, and all sentences. We also evaluated a variation of the rule-based classifier that labels a sentence as objective if there are at most three weak expressions in the previous, current, and next sentence combined, which raises the recall of the objective classifier. Our attempts to increase the recall of the subjective classifier all resulted in significant loss in precision, and thus we kept the original heuristic.

In its original English implementation, this system was proposed as being high-precision but low coverage. Evaluated on the MPQA corpus, it has subjective precision of 90.4, subjective recall of 34.2, objective precision of 82.4, and objective recall of 30.7; overall, precision is 86.7 and recall is 32.6 (Wiebe and Riloff, 2005). We see a similar behavior on Romanian for subjective sentences. The subjective precision is good, albeit at the cost of low

[2]Dan Tufiş, personal communication.

Measure	Subjective	Objective	All
subj = at least two strong; obj = at most two weak			
Precision	80.00	56.50	62.59
Recall	20.51	48.91	33.53
F-measure	32.64	52.52	43.66
subj = at least two strong; obj = at most three weak			
Precision	80.00	56.85	61.94
Recall	20.51	61.03	39.08
F-measure	32.64	58.86	47.93

Table 3: Evaluation of the rule-based classifier

recall, and thus the classifier could be used to harvest subjective sentences from unlabeled Romanian data (e.g., for a subsequent bootstrapping process). The system is not very effective for objective classification, however. Recall that the objective classifier relies on the weak subjectivity clues, for which the transfer of subjectivity in the translation process was particularly low.

4 A Corpus-Based Approach

Given the low number of subjective entries found in the automatically generated lexicon and the subsequent low recall of the lexical classifier, we decided to also explore a second, corpus-based approach. This approach builds a subjectivity-annotated corpus for the target language through projection, and then trains a statistical classifier on the resulting corpus (numerous statistical classifiers have been trained for subjectivity or sentiment classification, e.g., (Pang et al., 2002; Yu and Hatzivassiloglou, 2003)). The hypothesis is that we can eliminate some of the ambiguities (and consequent loss of subjectivity) observed during the lexicon translation by accounting for the context of the ambiguous words, which is possible in a corpus-based approach. Additionally, we also hope to improve the recall of the classifier, by addressing those cases not covered by the lexicon-based approach.

In the experiments reported in this section, we use a parallel corpus consisting of 107 documents from the SemCor corpus (Miller et al., 1993) and their manual translations into Romanian.[3] The corpus consists of roughly 11,000 sentences, with approximately 250,000 tokens on each side. It is a balanced corpus covering a number of topics in sports, politics, fashion, education, and others.

[3]The translation was carried out by a Romanian native speaker, student in a department of "Foreign Languages and Translations" in Romania.

Below, we begin with a manual annotation study to assess the quality of annotation and preservation of subjectivity in translation. We then describe the automatic construction of a target-language training set, and evaluate a classifier trained on that data.

Annotation Study.
We start by performing an agreement study meant to determine the extent to which subjectivity is preserved by the cross-lingual projections. In the study, three annotators – one native English speaker (*En*) and two native Romanian speakers (Ro_1 and Ro_2) – first trained on 3 randomly selected documents (331 sentences). They then independently annotated the subjectivity of the sentences of two randomly selected documents from the parallel corpus, accounting for 173 aligned sentence pairs. The annotators had access exclusively to the version of the sentences in their language, to avoid any bias that could be introduced by seeing the translation in the other language.

Note that the Romanian annotations (after all differences between the Romanian annotators were adjudicated) of all 331 + 173 sentences make up the gold standard corpus used in the experiments reported in Sections 3.2 and 4.1.

Before presenting the results of the annotation study, we give some examples. The following are English subjective sentences and their Romanian translations (the subjective elements are shown in bold).

> [en] **The desire to** give Broglio as many starts as possible.
> [ro] **Dorinţa de** a-i da lui Broglio cât mai multe starturi posibile.

> [en] **Suppose** he did lie beside Lenin, **would it be permanent ?**
> [ro] **Să presupunem** că ar fi aşezat alături de Lenin, **oare va fi pentru totdeauna?**

The following are examples of objective parallel sentences.

> [en]The Pirates have a 9-6 record this year and the Redbirds are 7-9.
> [ro] Piraţii au un palmares de 9 la 6 anul acesta si Păsările Roşii au 7 la 9.

> [en] One of the obstacles to the easy control of a 2-year old child is a lack of verbal communication.
> [ro] Unul dintre obstacolele în controlarea unui copil de 2 ani este lipsa comunicării verbale.

The annotators were trained using the MPQA annotation guidelines (Wiebe et al., 2005). The tagset consists of *S(ubjective)*, *O(bjective)* and *U(ncertain)*. For the *U* tags, a class was also given; *OU* means, for instance, that the annotator is uncertain but she is leaning toward *O*. Table 4 shows the pairwise agreement figures and the Kappa (κ) calculated for the three annotators. The table also shows the agreement when the borderline uncertain cases are removed.

	all sentences		Uncertain removed		
pair	agree	κ	agree	κ	(%) removed
Ro_1 & Ro_2	0.83	0.67	0.89	0.77	23
En & Ro_1	0.77	0.54	0.86	0.73	26
En & Ro_2	0.78	0.55	0.91	0.82	20

Table 4: Agreement on the data set of 173 sentences. Annotations performed by three annotators: one native English speaker (*En*) and two native Romanian speakers (Ro_1 and Ro_2)

When all the sentences are included, the agreement between the two Romanian annotators is measured at 0.83 (κ = 0.67). If we remove the borderline cases where at least one annotator's tag is *Uncertain*, the agreement rises to 0.89 with κ = 0.77. These figures are somewhat lower than the agreement observed during previous subjectivity annotation studies conducted on English (Wiebe et al., 2005) (the annotators were more extensively trained in those studies), but they nonetheless indicate consistent agreement.

Interestingly, when the agreement is conducted cross-lingually between an English and a Romanian annotator, the agreement figures, although somewhat lower, are comparable. In fact, once the *Uncertain* tags are removed, the monolingual and cross-lingual agreement and κ values become almost equal, which suggests that in most cases the sentence-level subjectivity is preserved.

The disagreements were reconciled first between the labels assigned by the two Romanian annotators, followed by a reconciliation between the resulting Romanian "gold-standard" labels and the labels assigned by the English annotator. In most cases, the disagreement across the two languages was found to be due to a difference of opinion about the sentence subjectivity, similar to the differences encountered in monolingual annotations. However, there are cases where the differences are due to the subjectivity being lost in the translation. Sometimes, this is due to several possible interpretations for the translated sentence. For instance, the following sentence:

> [en] They **honored** the battling Billikens last night.
> [ro] Ei i-au celebrat pe Billikens seara trecută.

is marked as *Subjective* in English (in context, the English annotator interpreted *honored* as referring to praises of the Billikens). However, the Romanian translation of *honored* is *celebrat* which, while correct as a translation, has the more frequent interpretation of *having a party*. The two Romanian annotators chose this interpretation, which correspondingly lead them to mark the sentence as *Objective*.

In other cases, in particular when the subjectivity is due to figures of speech such as irony, the translation sometimes misses the ironic aspects. For instance, the translation of *egghead* was not perceived as ironic by the Romanian annotators, and consequently the following sentence labeled *Subjective* in English is annotated as *Objective* in Romanian.

> [en] I have lived for many years in a Connecticut commuting town with a high percentage of [...] business executives of **egghead** tastes.
> [ro] Am trăit mulţi ani într-un oraş din apropiere de Connecticut ce avea o mare proporţie de [...] oameni de afaceri cu gusturi intelectuale.

4.1 Translating a Subjectivity-Annotated Corpus and Creating a Machine Learning Subjectivity Classifier

To further validate the corpus-based projection of subjectivity, we developed a subjectivity classifier trained on Romanian subjectivity-annotated corpora obtained via cross-lingual projections.

Ideally, one would generate an annotated Romanian corpus by translating English documents manually annotated for subjectivity such as the MPQA corpus. Unfortunately, the manual translation of this corpus would be prohibitively expensive, both timewise and financially. The other alternative – automatic machine translation – has not yet reached a level that would enable the generation of a high-quality translated corpus. We therefore decided to use a different approach where we automatically annotate the English side of an existing English-Romanian corpus, and subsequently project the annotations onto the Romanian side of the parallel cor-

	Precision	Recall	F-measure
high-precision	86.7	32.6	47.4
high-coverage	79.4	70.6	74.7

Table 5: Precision, recall, and F-measure for the two OpinionFinder classifiers, as measured on the MPQA corpus.

pus across the sentence-level alignments available in the corpus.

For the automatic subjectivity annotations, we generated two sets of the English-side annotations, one using the high-precision classifier and one using the high-coverage classifier available in the OpinionFinder tool. The high-precision classifier in OpinionFinder uses the clues of the subjectivity lexicon to harvest subjective and objective sentences from a large amount of unannotated text; this data is then used to automatically identify a set of extraction patterns, which are then used iteratively to identify a larger set of subjective and objective sentences.

In addition, in OpinionFinder, the high-precision classifier is used to produce an English labeled data set for training, which is used to generate its Naive Bayes high-coverage subjectivity classifier. Table 5 shows the performance of the two classifiers on the MPQA corpus as reported in (Wiebe and Riloff, 2005). Note that 55% of the sentences in the MPQA corpus are subjective – which represents the baseline for this data set.

The two OpinionFinder classifiers are used to label the training corpus. After removing the 504 test sentences, we are left with 10,628 sentences that are automatically annotated for subjectivity. Table 6 shows the number of subjective and objective sentences obtained with each classifier.

Classifier	Subjective	Objective	All
high-precision	1,629	2,334	3,963
high-coverage	5,050	5,578	10,628

Table 6: Subjective and objective training sentences automatically annotated with OpinionFinder.

Next, the OpinionFinder annotations are projected onto the Romanian training sentences, which are then used to develop a probabilistic classifier for the automatic labeling of subjectivity in Romanian sentences.

Similar to, e.g., (Pang et al., 2002), we use a Naive Bayes algorithm trained on word features co-occurring with the subjective and the objective classifications. We assume word independence, and we use a 0.3 cut-off for feature selection. While recent work has also considered more complex syntactic features, we are not able to generate such features for Romanian as they require tools currently not available for this language.

We create two classifiers, one trained on each data set. The quality of the classifiers is evaluated on the 504-sentence Romanian gold-standard corpus described above. Recall that the baseline on this data set is 54.16%, the percentage of sentences in the corpus that are subjective. Table 7 shows the results.

	Subjective	Objective	All
projection source: OF high-precision classifier			
Precision	65.02	69.62	64.48
Recall	82.41	47.61	64.48
F-measure	72.68	56.54	64.68
projection source: OF high-coverage classifier			
Precision	66.66	70.17	67.85
Recall	81.31	52.17	67.85
F-measure	72.68	56.54	67.85

Table 7: Evaluation of the machine learning classifier using training data obtained via projections from data automatically labeled by OpinionFinder (OF).

Our best classifier has an F-measure of 67.85, and is obtained by training on projections from the high-coverage OpinionFinder annotations. Although smaller than the 74.70 F-measure obtained by the English high-coverage classifier (see Table 5), the result appears remarkable given that no language-specific Romanian information was used.

The overall results obtained with the machine learning approach are considerably higher than those obtained from the rule-based classifier (except for the precision of the subjective sentences). This is most likely due to the lexicon translation process, which as mentioned in the agreement study in Section 3.1, leads to ambiguity and loss of subjectivity. Instead, the corpus-based translations seem to better account for the ambiguity of the words, and the subjectivity is generally preserved in the sentence translations.

5 Conclusions

In this paper, we described two approaches to generating resources for subjectivity annotations for a new language, by leveraging on resources and tools available for English. The first approach builds a target language subjectivity lexicon by translating an existing English lexicon using a bilingual dictionary. The second generates a subjectivity-annotated corpus in a target language by projecting annotations from an automatically annotated English corpus.

These resources were validated in two ways. First, we carried out annotation studies measuring the extent to which subjectivity is preserved across languages in each of the two resources. These studies show that only a relatively small fraction of the entries in the lexicon preserve their subjectivity in the translation, mainly due to the ambiguity in both the source and the target languages. This is consistent with observations made in previous work that subjectivity is a property associated not with words, but with word *meanings* (Wiebe and Mihalcea, 2006). In contrast, the sentence-level subjectivity was found to be more reliably preserved across languages, with cross-lingual inter-annotator agreements comparable to the monolingual ones.

Second, we validated the two automatically generated subjectivity resources by using them to build a tool for subjectivity analysis in the target language. Specifically, we developed two classifiers: a rule-based classifier that relies on the subjectivity lexicon described in Section 3.1, and a machine learning classifier trained on the subjectivity-annotated corpus described in Section 4.1. While the highest precision for the subjective classification is obtained with the rule-based classifier, the overall best result of 67.85 F-measure is due to the machine learning approach. This result is consistent with the annotation studies, showing that the corpus projections preserve subjectivity more reliably than the lexicon translations.

Finally, neither one of the classifiers relies on language-specific information, but rather on knowledge obtained through projections from English. A similar method can therefore be used to derive tools for subjectivity analysis in other languages.

References

Alina Andreevskaia and Sabine Bergler. Mining wordnet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In *Proceedings of EACL 2006*.

Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of HLT/EMNLP 2005*.

Krisztian Balog, Gilad Mishne, and Maarten de Rijke. 2006. Why are they excited? identifying and explaining spikes in blog mood levels. In *EACL-2006*.

Andrea Esuli and Fabrizio Sebastiani. 2006. Determining term subjectivity and term orientation for opinion mining. In *Proceedings the EACL 2006*.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD*.

Yi Hu, Jianyong Duan, Xiaoming Chen, Bingzhen Pei, and Ruzhan Lu. 2005. A new method for sentiment classification in text retrieval. In *Proceedings of IJCNLP*.

Radu Ion. 2007. *Methods for automatic semantic disambiguation. Applications to English and Romanian.* Ph.D. thesis, The Romanian Academy, RACAI.

Hiroshi Kanayama and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of EMNLP 2006*.

Soo-Min Kim and Eduard Hovy. 2006. Identifying and analyzing judgment opinions. In *Proceedings of HLT/NAACL 2006*.

Levon Lloyd, Dimitrios Kechagias, and Steven Skiena. 2005. Lydia: A system for large-scale news analysis. In *Proceedings of SPIRE 2005*.

George Miller, Claudia Leacock, Tangee Randee, and Ross Bunker. 1993. A semantic concordance. In *Proceedings of the DARPA Workshop on Human Language Technology*.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of ACL 2004*.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP 2002*.

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of EMNLP 2003*.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2006. Latent variable models for semantic orientations of phrases. In *Proceedings of EACL 2006*.

Peter Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of ACL 2002*.

Universal Dictionary. 2007. Available at www.dicts.info/uddl.php.

Janyce Wiebe and Rada Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of COLING-ACL 2006*.

Janyce Wiebe and Ellen Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *Proceedings of CICLing 2005 (invited paper)*. Available at www.cs.pitt.edu/mpqarequest.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2/3):164–210. Available at www.cs.pitt.edu/mpqa.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT/EMNLP 2005*.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of EMNLP 2003*.