Porting Multilingual Subjectivity Resources across Languages

Carmen Banea, Rada Mihalcea, and Janyce Wiebe

Abstract—Subjectivity analysis focuses on the automatic extraction of *private states* in natural language. In this paper, we explore 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.

Index Terms—Natural language processing, text analysis, sentiment and subjectivity analysis, multilingual subjectivity

1 Introduction

Subjectivity analysis, the automatic extraction of information related to attitudes, opinions, and sentiments in text, is a rapidly growing area of Natural Language Processing (NLP). To date, a large number of NLP applications have employed techniques for automatic subjectivity analysis, including automatic expressive text-to-speech synthesis [1], tracking sentiment timelines in online forums and news [2], [3], mining opinions from product reviews [4], question answering [5], and summarization [6].

Much of the work in subjectivity analysis has been applied to English data. In general, efforts to build subjectivity analysis tools for other languages have been hampered by the high cost involved in creating corpora and lexical resources for a new language. Despite this aspect, work on other languages is growing: e.g., Japanese data are used in [7], [8], [9], Chinese data are used in [10], and German data are used in [11]. In addition, several participants in the Chinese and Japanese opinion extraction tasks of NTCIR-6 [12] performed subjectivity and sentiment analysis in languages other than English.

As only 29.4 percent of Internet users speak English,[1] it is important to focus on subjectivity research in other languages as well. This goal can initially be met through leveraging on the laboriously developed resources and tools available in English, thus avoiding expending valuable resources at this preliminary stage. Later on, unique structures appearing in individual languages can be employed to detect more subtle subjectivity clues. For example, in languages such as Chinese, researchers have been looking at the ability of characters to carry sentiment information. In Romanian, due to markers of politeness, formal and informal registers, and inflected verbal modes, experiments have hinted that subjectivity detection may be easier to achieve. These additional sources of information may not be available across all languages, yet, ultimately investigating a synergistic approach to detecting subjectivity and sentiment in multiple languages at the same time, may trigger improvements not only in the other languages, but in English as well, as research conducted by [13], [14], [15], [16] suggests. This paper explores the initial link in the chain, by employing a multilingual path to extrapolate resources available in English in a target language. Note that although one could initially rely only on machine translation to produce basic sentiment and subjectivity analysis methods in another language, this does not circumvent the need to build resources that are specific to the task such as sentiment and subjectivity tools and lexicons. The translation-based method is static, it cannot be easily expanded and improved, and is entirely dependent on improvements in machine translation, while the task-specific resources can be used as an initial step to build increasingly accurate tools that also account for the specifics of the target language.

1. www.internetworldstats.com/stats.htm, June 30, 2008.

- *C. Banea and R. Mihalcea are with the Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203. E-mail: carmen.banea@gmail.com.*
- *J. Wiebe is with the Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15260.*

Manuscript received 19 Aug. 2011; revised 29 Nov. 2012; accepted 11 Dec. 2012; published online 23 Jan. 2013.
Recommended for acceptance by J. Gratch.
For information on obtaining reprints of this article, please send e-mail to: taffc@computer.org, and reference IEEECS Log Number TAFFC-2011-08-0058.
Digital Object Identifier no. 10.1109/T-AFFC.2013.1.

1.1 Background

Before we describe our research goals and experiments, we first provide some background about subjectivity analysis.

Subjective expressions [17] are words and phrases used to express *private states*, where private states are mental and emotional states such as speculations, evaluations, sentiments, and beliefs [18]. Following are some examples, with subjective expressions in bold [19]:

1. His *alarm* grew.
2. He *absorbed* the information quickly.

Polarity (also called *semantic orientation*) is also important for NLP applications in sentiment analysis and opinion extraction. In review mining, for example, we want to know whether an opinion about a product is positive or negative. Even so, there are strong motivations for a separate

1949-3045/13/$31.00 © 2013 IEEE Published by the IEEE Computer Society

subjective/objective (S/O) classification as well. First, expressions may be subjective but not have any particular polarity. An example given by Wilson et al. [20] is *Jerome says the hospital* **feels** *no different than a hospital in the states.* An NLP application system may want to find a wide range of private states attributed to a person, such as their motivations, thoughts, and speculations, in addition to their positive and negative sentiments.

Second, distinguishing subjective and objective instances has often proven to be more difficult than subsequent polarity classification. Researchers have found this at various levels of analysis, including the manual annotation of phrases [8], sentiment classification of phrases [20], sentiment tagging of words [21], and sentiment tagging of word senses [22]. Thus, effective methods for S/O classification promise to improve performance for sentiment analysis. Researchers in sentiment analysis have realized benefits by decomposing the problem into S/O and polarity classification [5], [23], [20], [11]. One reason is that different features may be relevant for the two subproblems. For example, negation features are more important for polarity classification than for subjectivity classification. A promising approach is twofold: classifying instances as objective or subjective, and further classifying the subjective instances by polarity. The paper addresses the first of these subproblems. However, because the subproblems are interrelated, we consider both when discussing previous work below.

This paper is primarily concerned with detecting the presence of subjectivity. Many approaches to this problem rely on subjectivity lexica, groups of words and phrases that have subjective usages. Such lexica may be formed manually, such as the General Inquirer [24], or they may be created using automatic processes.

Several methods have been proposed for learning subjective words (also known as words with *semantic orientation*, *opinion bearing* words, *sentiment bearing* words, and so on). In early work, Hatzivassiloglou and McKeown [25] exploit linguistic constraints on the semantic orientations of adjectives used in conjunctions: In the construct <adjective> and <adjective>, the adjectives must be the same polarity. We can view the pattern <adjective-known-to-be-positive> and <adjective> as an extraction pattern for finding positive adjectives (similarly for negative ones). This idea was later extended by several other researchers, including Turney and Littman [26], who use five patterns to extract positive and negative words; Gamon and Aue [27], who exploit the constraint that words of opposite polarity tend not to occur in the same sentence; Kanayama and Nasukawa [9], who exploit the tendency for the same polarities to appear successively in context; and Riloff et al. [28], who identify extraction patterns for subjective nouns automatically.

Lexicon entries need not be single words. First, the entries may be phrases. The authors of [29], [30] first experimented with n-grams, and the authors of [31], [32] proposed methods for automatically identifying subjective clues that are lexico-syntactic patterns rather than individual words or n-grams. Since then, pattern representations of subjective expressions have become more common (e.g., [9], [33], [34]). Second, lexicon entries may be word senses, to capture the ambiguity of words with both subjective and objective (or both positive and negative) senses [19], [22].

How are subjectivity lexicons used to recognize subjective sentences or documents? They may be exploited in rule-based approaches that calculate scores reflecting the presence of clues (e.g., [35]), or they may be used in machine-learning approaches that include features built on the lexicon entries (e.g., [32], [20]).

Subjectivity lexicons are not required; it is also common to build machine learning systems using features based only on the training data (e.g., [29]). Our experiments involve both lexicon-based and purely corpus-based approaches.

1.2 Research Questions

In this paper, we explore methods for generating subjectivity analysis resources in a new language by leveraging on the tools and resources available in English.

We seek to answer the following questions: First, assuming a subjectivity annotated English lexicon, can we use machine translation to generate a high-quality lexical resource in the target language? Second, if we translate a small number of seeds from this subjectivity lexicon, can we grow the set while maintaining the quality of the lexicon in the target language? Third, if an English corpus manually annotated for subjectivity is available, can we use machine translation to generate a subjectivity-annotated corpus in the target language and train a subjectivity classifier in the target language? Fourth, 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 or automatic annotations of automatic parallel text generated through machine translation; can these automatically generated corpora be used to train a subjectivity classifier in the target language?

1.3 Paper Contributions

Our work focuses on leveraging three types of resources developed for English: Subjectivity annotated lexicons, manually annotated corpora for subjectivity, and automatic subjectivity annotation tools. To demonstrate the portability of the methods we explore, we run several experiments on two different languages, Romanian and Spanish.

Subjectivity annotated lexicons. We focus on two potential paths to leverage on subjectivity annotated lexicons. One is based on attempting to automatically translate a source language lexicon into the target language by using a bridging multilingual dictionary. The resources required by this experiment are an annotated subjectivity lexicon and a multilingual dictionary. The second one is based on selecting a small set of subjective seeds from the source language lexicon and manually translating them into the target language. We can then expand the lexicon by solely using material in the target language through a bootstrapping mechanism. This scenario requires the availability of a small set of subjective seeds, an electronic dictionary and a raw corpus.

Manually annotated corpora for subjectivity. We propose leveraging on a corpus manually annotated for subjectivity at the sentence level. To create a version of the corpus in the

target language, we employ a statistical machine translation engine that allows transfer between the source and target language. Since the resulting corpus in the target language is not annotated, we project the human annotations from the source corpus onto the machine translated corpus in the target language. The resulting annotated corpus can then be used to train a machine learning algorithm and, therefore, create an automatic subjectivity analysis tool in the target language. This experimental setup requires the availability of a manually annotated corpus for subjectivity in the target language, as well as a machine translation engine between the source and the target language.

Subjectivity annotation tools. This scenario explores the potential of using parallel text paired with an automatic tool for subjectivity analysis to generate subjectivity annotated text in the target language. We look at the differences between employing manual parallel text as well as parallel text generated automatically through machine translation. The annotated corpus developed in the target language will be able to act as a training data set for a machine learning algorithm. This setup requires either a manually translated parallel text, or data either in the source or target language paired with a machine translation engine.

1.4 Source Language Resources

We introduce below the individual English resources that this work employs to conduct subjectivity analysis in a target language.

OpinionFinder subjectivity lexicon. The subjectivity lexicon we use is employed in OpinionFinder [36], 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 multiword 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 considered *strong* clues of subjectivity, while those that appear less often, but still more often than expected by chance, are labeled *weak*, and the remaining are considered *neutral*. Each entry in the lexicon is extended by a group of fields that provide additional information, such as length, polarity, part-of-speech, stem, source, and so on.

MPQA. The MPQA corpus [17] 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 data set [37]. From the total of 9,700 sentences in this corpus, 55 percent of the sentences are labeled as subjective while the rest are objective.

OpinionFinder. OpinionFinder [36] is a system that automatically annotates the subjectivity of new text based on the presence (or absence) of words or phrases in a large lexicon. OpinionFinder includes both a high-precision, rule-based classifier and a high-coverage, machine learning classifier. The high-precision classifier uses the clues of a subjectivity lexicon to harvest subjective and objective sentences from a large amount of unannotated text; this data are 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. The high-precision classifier is also used to automatically produce an English labeled data set, which is used in turn to train the high-coverage subjectivity classifier. Table 1 shows the performance of the two OpinionFinder classifiers as measured on the MPQA corpus [17].

TABLE 1
Precision (P), Recall (R), and F-Measure (F) for the Two OpinionFinder Classifiers, as Measured on the English MPQA Corpus

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

SemCor. As a raw corpus, we use a subset of the English SemCor corpus [38], 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 4.3).

The paper is organized as follows: We first review related work in Section 2. Section 3 describes several methods to automatically generate a subjectivity lexicon in the target language, and evaluates the lexicons using both in vitro (manual evaluation of the lexicon entries) and in vivo (embedding of the lexicons in a subjectivity annotation tool) approaches. Section 4 describes methods to perform subjectivity analysis by relying on sentence-level annotations. All the methods proposed are thoroughly evaluated on gold standard data sets in Romanian (Ro) and Spanish (Es), and the results are compared and discussed. Finally, Section 5 gives concluding remarks and provides a "big picture" comparison of the approaches discussed in the paper.

2 Related Work

Recently, resources and tools for subjectivity and sentiment analysis developed for English have been used as a starting point to build resources in other languages, via cross-lingual projections or monolingual and multilingual bootstrapping. Several directions were followed, focused on leveraging annotation schemes, lexica, corpora, and automated annotation systems. Since most researchers have focused on sentiment analysis due to the popular task of analyzing product or movie reviews (further motivated by the large amount of online data in multiple languages that can be used in this subarea of subjectivity analysis), where we cannot reference papers related to subjectivity, we cite some articles focusing on polarity, because the methods are often times comparable, even though the results are not.

The English annotation scheme developed by Wiebe et al. [17] for opinionated text lays the groundwork for the research carried out by Esuli et al. [39] when annotating expressions of private state in the Italian Content Annotation

Bank. Sentiment and subjectivity lexica such as the one included with the OpinionFinder distribution [36], the General Inquirer [24], or the SentiWordNet [22] were transferred into Chinese [40], [41] and into Romanian [42]. English corpora manually annotated for subjectivity or sentiment such as MPQA [17] or the multidomain sentiment classification corpus [43] were subjected to experiments in Spanish, Romanian, or Chinese upon automatic translation [44], [16]. Furthermore, tools developed for English were used to determine sentiment or subjectivity labeling for a given target language by transferring the text to English and applying an English classifier on the resulting data. The labels were then transferred back into the target language [45], [44], [13]. These experiments were carried out in Arabic, Chinese, English, French, German, Italian, Japanese, Korean, Spanish, and Romanian.

2.1 Subjectivity Lexicon Construction

Many subjectivity and sentiment analysis tools rely on manually or semiautomatically constructed lexicons [5], [32], [11]. The availability of such lexicons enables the construction of efficient rule-based subjectivity and sentiment classifiers that rely on the presence of lexicon entries in the text.

Most of the work to date on subjectivity lexicon construction has assumed advanced natural language processing tools such as syntactic parsers [46] or tools for information extraction [32], or the availability of broad-coverage rich lexical resources such as WordNet [47]. However, such tools and resources are available only for a handful of languages, which limits the applicability of these approaches. Instead, in the methods introduced in Section 3, we try to minimize the resources required to build a subjectivity lexicon. Thus, these methods are potentially applicable to a large number of the languages spoken worldwide.

Kim and Hovy [11] present a method for lexicon induction that is embedded in an automatic system that identifies opinions in e-mails. They use a set of seeds in English that are propagated through the WordNet structure [47]. The seeds are grouped based on polarity into positive, negative, and neutral classes, and they are expanded by navigating the synonym relationship encoded in WordNet while enforcing their closeness to the most probable polarity class. The enriched set is then transferred into German through a statistical word alignment obtained by applying the Giza++ toolkit[2] on the European Parliament corpus.[3] Building on this approach, Rao and Ravichandran [48] introduce a sentiment lexicon construction method that uses the WordNet graph and the relationships it entails to extend polarity classification to other words using graph based semisupervised learning algorithms, such as mincuts, randomized mincuts, and label propagation. The latter algorithm is the best performing one and was applied to Hindi (employing the Hindi WordNet[4]) and to French (using the OpenOffice thesaurus[5]).

Our approach describing bootstrapping a subjectivity lexicon in a target language from a small set of seed words (see Section 3.2) relates most closely to the method proposed by Turney and Littman [26] for the construction of lexicons annotated for polarity in English. His algorithm starts with a few positive and negative seeds, and then uses data from the web together with a similarity method (pointwise mutual information) to automatically grow this seed list. Our approach differs from [26] in three important ways: First, we address the task of *subjectivity* lexicon construction, and not polarity lexicons. Second, our method learns a subjectivity lexicon in a target language relying on fewer, smaller-scale resources compared to Turney's method (which uses a very large terabyte corpus consisting of all the English web documents indexed by the Altavista search engine).[6] Third, we also explore using different groupings of seeds to understand what variation allows for the most comprehensive similarity (from a subjectivity perspective) to transpire.

2.2 Automatic Subjectivity Tagging

Automatic subjectivity classification has followed two paths, namely lexicon based and corpus based. The former uses a set of predefined heuristics that give a particular weight to lexicon entries encountered in a span of text and decide their cumulative effect in labeling the text as subjective or objective [32] or polar [11]. We have followed a similar technique when performing in vivo evaluations of the lexicons, we extracted in the target language in Section 3.

The corpus-based methods build a model directly from the annotated corpora and infer heuristics automatically based on the learning algorithm of choice. Following this latter approach, Wan [16] constructs a polarity cotraining system by using the multilingual views obtained through the automatic translation of product reviews into Chinese and English. His algorithm uses two SVM classifiers, one in Chinese and one in English based on unigram and bigram language models that make high-precision polarity predictions by targeting primarily those unlabeled examples that are most similar to the examples already in the training set. Initially, the training set consists of a data set of labeled examples in English and their counterpart in Chinese. Then, the first iteration of cotraining is performed, and a set of unlabeled instances is classified and added to the training set so that it can be used to retrain the two classifiers at the next iteration. In a similar vein, Prettenhofer and Stein [49] investigate cross-lingual sentiment classification using a method inspired by the domain adaptation approach with structural correspondence learning proposed by Blitzer et al. [43].

Also focusing on sentiment analysis are the methods proposed in [50], [51], where multiple languages are used to build more robust sentiment classifiers. In [50], multilingual topics are learned using Latent Dirichlet Allocation. The model identifies vocabulary partitions that are consistent across different languages in terms of meaning and relevance to sentiment. The method does not require neither parallel text nor machine translation, and it only needs a few resources to bridge the languages (e.g., WordNets in

2. http://www.fjoch.com/GIZA++.html.
3. http://www.statmt.org/europarl/.
4. http://www.cfilt.iitb.ac.in/wordnet/webhwn/.
5. http://www.openoffice.org/.
6. http://www.altavista.com/.

multiple languages or bilingual dictionaries). Instead, [51] relies on unlabeled parallel data to augment existing labeled data sets using a joint bilingual bootstrapping model. Both methods are found to improve significantly over the monolingual baselines, demonstrating the usefulness of the multilingual representations.

3 Leveraging on Word-Level Subjectivity Annotations

Many subjectivity and sentiment analysis tools rely on manually or semiautomatically constructed lexicons [5], [32], [11]. Given the success of such techniques, the first approach described in Section 3.1 is based on creating a subjectivity lexicon in the target language by translating a source language lexicon. The second approach explored in Section 3.2 focuses on extracting a subjectivity lexicon within a given target language starting from a small seed set of manually translated subjective entries. Both methods are evaluated by building a rule-based classifier that relies on the generated lexicons to perform subjectivity annotations in the target language.

3.1 Translating a Subjectivity Lexicon

We employ the OpinionFinder subjectivity lexicon (see Section 1.4). Since this subjectivity lexicon is compiled based on occurrences in actual text, our first task is to extract words in the form in which they may appear in a dictionary. Due to the fact that some of the entries present discrepancies, we implement a voting mechanism that takes into consideration the additional annotations such as part-of-speech, polarity, and so on, accompanying every entry to decide on the correct part-of-speech information and lemma. In case of a tie, we query WordNet to decide which form to consider. To exemplify, let us consider the word "atrocities."

$$\begin{aligned}
&\text{atrocities: } \{len = 1, word1 = \text{atrocities},\\
&\quad othertypes = \text{metaboot} - \text{2000nouns-strongsubj},\\
&\quad pos1 = \text{adj}, highprec = \text{yes},\\
&\quad polannsrc = \text{ph}, stem1 = \text{atrocity:noun},\\
&\quad origtypes = \text{metaboot} - \text{2000nouns-strongsubj},\\
&\quad RFE = \text{tff}, origpats = \%\text{atrocities}\|\text{adj}\|\text{n}\%,\\
&\quad mpqapolarity = \text{strongneg}, stemmed1 = \text{n},\\
&\quad MISS = \text{t}, intensity = \text{high},\\
&\quad type = \text{metaboot} - \text{2000nouns-strongsubj}\}.
\end{aligned}$$

For this entry, the *origpats* field suggests an erroneous adjective classification, further supported by the *pos1* field. Our voting mechanism is, however, able to accurately decide on the correct part-of-speech, by taking into consideration the *othertypes*, *stem1*, *origtypes*, *origpats*, *stemmed1*, and *type* fields. Therefore, this subjective entry is corrected, and its lemma becomes "atrocity," while its part-of-speech is updated to *noun*.

We select from the lexicon only those entries composed of a single word, and labeled as either strong or weak (as provided by the *mpqapolarity* field). Upon optionally adjudicating the POS discrepancies (for approximately 2 percent of the entries), we obtain two finalized versions of the subjectivity lexicon in the source language, each containing 5,339 entries. Next, to perform the translation in the target language, we use the Ectaco online dictionary.[7] This service provides translations into 25 languages, and each dictionary features more than 400,000 entries. We chose this resource so we can conduct a quality assessment of our approach, by carrying out translation experiments from the source language into two languages that are both supported by the Ectaco online dictionary, namely Romanian and Spanish.

There were several challenges encountered in the translation process. First, although the English subjectivity lexicon contains inflected words, we must use the lemmatized form 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."[8] Therefore, it is very difficult if not impossible to recreate the inflection, as Romanian does not have a synonym for "memories" from the same lexical family.

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 seem 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 may not make this distinction, and therefore, we may have again to rely on the *most frequent* heuristic captured by the translation order in the bilingual dictionary.

Third, Romanian and Spanish do not offer direct translations for the multitude of adverbs suffixed by *-ly* in English (e.g., the adverb "freely" obtained from the noun "free" can be translated into Romanian or Spanish only as a phrase (Ro: "în libertate," Es: "con libertad"; En: "in/with freedom")). Others, such as "staunchly" or "solicitously" do not return any translations.

The Ectaco dictionary provided similar coverage for both our target languages, as 1,580 entries were translated into Romanian (29.6 percent) and 2,009 (37.6 percent) into Spanish, respectively. Table 2 shows examples of entries in the Romanian and Spanish lexicons, together with their corresponding original English form. The table also shows the reliability of the expression (*weak* or *strong*) and the part-of-speech, both attributes being provided in the English subjectivity lexicon.

Our method involving translating a subjectivity lexicon was initially published in [42]. Here, we expand on that work in several ways: 1) we provide a method for correcting erroneous part-of-speech annotations provided by OpinionFinder; 2) we conduct additional experiments to evaluate

7. http://www.ectaco.co.uk/free-online-dictionaries/.

8. The correct translation in this case for the plural "memories" would have been "amintiri."

TABLE 2
Sample Entries from the Subjectivity Lexica

English	attributes	Romanian	Spanish
admonish	strong, verb	preveni	amonestar
beautify	strong, verb	împodobi	embellecer
credence	weak, noun	crezare	creencia
diligent	strong, adj	sârguitor	aprovechado
excuse	weak, verb	scuză	disculpa

TABLE 3
Evaluation of 150 Entries in the Romanian (Ro) and Spanish (Es) Lexicons, and Comparison with the OpinionFinder English Lexicon (OF_{lex})

Lang	Subj	Both	Obj	Wrong
Ro	0.7 (105)	0.19 (29)	0.07 (10)	0.04 (6)
Es	0.72 (108)	0.18 (27)	0.05 (8)	0.05 (7)
OF_{lex}	1 (150)			

the performance achieved by translating using the most frequent sense versus a part of speech-based sense; and 3) we duplicate our experiments in a new language (Spanish) and compare the results.

3.1.1 Manual Evaluation

We want to assess the quality of the translated lexicons, and compare them to the quality of the original English lexicon. The English subjectivity lexicon was evaluated in [37] against a corpus of English-language news articles manually annotated for subjectivity (the *MPQA* corpus [17]). According to that evaluation, 85 percent of the instances of the clues marked as *strong* and 71.5 percent of the clues marked as *weak* appear in subjective sentences in the MPQA corpus, where 55 percent of the sentences in this corpus are subjective.

Since there are no comparable Romanian or Spanish corpora, an alternate way to judge the subjectivity of the translated lexicon entries is needed. Two native speakers, one of Romanian and one of Spanish, annotated the subjectivity of 150 randomly selected entries in the generated lexica. They were presented with the original English subjective entry and the automatically translated one. Due to word ambiguity, and the inability of a human to immediately recall all the possible uses of a given word, we assisted the judges by providing the first approximately 100 snippets containing the translated word, based on a query to the Google search engine (restricted to either Spanish or Romanian). Since many of the sites in languages with fewer electronic resources publish news, whose content changes more frequently thus influencing the Google search results ranking, a large number of snippets originates from the news domain. The subjectivity of a word was consequently judged from the contexts in which it most frequently appeared, thus, accounting for its most frequent meaning on the web. The tag set used for the annotations consists of *Subj(ective)*, *Obj(ective)*, and *Both*.[9] A *Wrong* label is also used to indicate a wrong translation. Additionally, for the *Subj(ective)* and *Both* labels, the judges added strength granularity, resulting in *weak* and *strong* annotations. An entry is considered *strong* when its appearance in a given context would render the entire text subjective. In contrast, a *weak* entry contains a certain level of subjectivity, yet a sentence in which a weak entry appears may not be labeled as subjective based on this clue alone. Table 3 summarizes the two annotators' judgments on this data.

Thus, the study presented in Table 3 suggests that the Romanian and Spanish subjectivity clues derived through translation are less reliable than the original set of English clues. Only 70 percent of the translated entries into Romanian and 72 percent of those translated into Spanish are considered unequivocally subjective by the judges. Also, about 19 percent of the entries automatically ported to both Romanian and Spanish have ambiguous subjective meanings. It is also interesting to note that the behavior of the two languages is very similar, as they differ by at most three annotations for each tag category.

In several cases, the subjectivity is lost in 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 completely lose subjectivity once translated (such as "one sided," which becomes "cu o singură latură" in Romanian, meaning "with only one side" (as of objects)). In the case of verb "appreciate" translated into Romanian as "aprecia," which is a polysemous verb denoting a frequent objective meaning of "gaining value" (as of currencies); in Spanish, the word was translated as "estimar" ("estimate"), which involves a far more clear subjective judgment.

In other cases, the translation adds frequent objective meanings through part-of-speech transfer. One example is the adverb "icy," which the dictionary translates into the noun "gheață" ("ice") in Romanian; due to the transfer in part-of-speech, the translation candidate has only an objective meaning.

In a similar way, the word *strut* (see definition below) appears in the subjectivity lexicon as a verb. Upon translation, its correspondent in Spanish becomes the noun *puntal*, with the meaning of *strut* (noun, first dictionary sense), with a clear objective meaning and use.

strut (*Merriam Webster online dictionary*)

intransitive verb: 1) to become turgid, swell; 2a) To walk with a proud gait, 2b) to walk with a pompous and affected air

transitive verb: to parade (as clothes) with a show of pride

noun: 1) a structural piece designed to resist pressure in the direction of its length; 2) a pompous step or walk; and 3) arrogant behavior, swagger

3.2 Growing a Subjectivity Lexicon

A different path to generate a subjectivity lexicon in a target language is to acquire a large subjectivity lexicon by bootstrapping from a few manually selected seeds. At each iteration, the seed set is expanded with related words found in an on line dictionary, which are filtered by using a measure of word similarity. The bootstrapping process is

9. *Both* is used when the word does not have a clear subjective or objective predominant use, but can rather appear in both types of contexts equally.

TABLE 4
Sample Entries from the Initial Seed Set in Romanian (Ro)/ Spanish (Es) Accompanied by Their English Translations

POS	Sample seeds
Noun	blestem / maldiciòn (curse), despot / tirano (tyrant), furie / furia (fury), idiot / idiota (idiot), fericire / felicidad (happiness)
Verb	iubi / amar (love), aprecia / apreciar (appreciate), spera / esperar (hope), dori / desear (wish), urî / odiar (hate)
Adj	frumos / bello (beautiful), dulce / dulce (sweet), urât / feo (ugly), fericit / feliz (happy), fascinant / facinante (fascinating)
Adv	posibil / posiblemente (possibly), probabil / probablemente (probably), desigur / seguramente (of course), enervant / irritante (unnerving)

illustrated in Fig. 1 in the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/T-AFFC.2013.1).

3.2.1 Seed Set

We use a preliminary set of 60 seeds, evenhandedly sampled from verbs, nouns, adjectives, and adverbs. This allows for a good coverage in each part-of-speech grouping (approximately 15 words), thus, enabling experiments in a variety of settings when growing the subjectivity lexicon (*seed versus candidate*, *POS group versus candidate*, *all versus candidate*). While seeds can easily be obtained directly in the target language, without the need of manual translation from a source language, to maintain similar experimental settings across languages, we opted to manually translate the seeds into Romanian and Spanish starting from hand-picked strong subjective entries appearing in the English subjectivity lexicon. Table 4 shows a sample of the entries in the initial seed set translated into the target languages, accompanied by the initial seed word in English.

3.2.2 Dictionary

For each seed word we collect *candidates*, namely all open-class words appearing in the definition, as well as synonyms and antonyms, if available.[10]

In our experiments, for the Romanian dictionary we use *Dex online*,[11] while for the Spanish dictionary we query the *Diccionario de la lengua española*[12] maintained by the Real Academia Española institution. Similar dictionaries are available for many other languages; when online dictionaries are not available, they can be obtained at relatively low cost through optical character recognition (OCR) performed on a hard copy dictionary.

3.2.3 Bootstrapping Iterations

For each seed word, a query is formulated against the explicative dictionary available in the target language (see Fig. 1, available in the online supplementary material). From the definitions obtained, a list of related words is extracted and added to the list of candidates if they were not already encountered, if they are longer than three characters, and if they do not appear in a list of stopwords.

We used three different variants to filter the candidate words. The first focuses on capturing the similarity between the original seed word that extracted the candidate and the candidate (*seed versus candidate*). The second variation groups together all seeds with the same part-of-speech, and proposes calculating the similarity between the candidate and the group with the part-of-speech that extracted it (*POS group versus candidate*). The third variation filters candidates based on their similarity with the entire original seed set (*all versus candidate*).[13] The bootstrapping process continues to the next iteration until a maximum number of iterations is reached.

Note that the part-of-speech information is not maintained during the bootstrapping process, as candidate words occurring in the definitions belong to different parts-of-speech and we do not use a POS tagger (because we want the method to be easily portable to other languages). Although the initial seed set is balanced with respect to syntactic categories, as candidate words are extracted, the balance may tilt toward one of the categories by the end of the bootstrapping process.

3.2.4 Filtering

To remove noise from the lexicon, we implemented a filtering step that takes into account the similarity between the original seeds (in the three variations mentioned earlier) and each of the candidates. We experimented with two corpus-based measures of similarity, namely the Pointwise Mutual Information [52] and Latent Semantic Analysis (LSA) [53], [54]. We ultimately decided to use only LSA, as both methods provided similar results, but the LSA-based method was significantly faster and required less training data. In our experiments, we used the Infomap NLP[14] implementation of LSA. After each iteration, only candidates with a LSA score higher than 0.4 (determined empirically) are considered to be expanded in the next iteration.

Upon bootstrapping termination, the subjectivity lexicons constructed incrementally after each iteration consist of a ranked list of candidates in decreasing order of similarity to the three seed set variations. A variable filtering threshold can be used to enforce the selection of only the most closely related candidates, resulting in more restrictive and pure subjectivity lexicons. In our experiments, we used the following thresholds: 0.40 (i.e., the lexicon resulting after the bootstrapping process without additional filtering), 0.45, 0.50, and 0.55.

LSA modules were trained on two approximately 55 million word corpora constructed for Romanian and Spanish, respectively. Smaller corpora are also feasible [55], yet to obtain a more accurate LSA similarity measure, larger data sets are desirable. Corpora can be obtained for many low-resource languages by using semiautomatic methods for corpus construction [56]. The Romanian corpus was created from a set of editorials collected from Romanian

10. Since we expand and process all the possible meanings for each candidate word, word ambiguity is not an issue for our method.
11. http://www.dexonline.ro.
12. http://buscon.rae.es/draeI/.
13. The intuition behind using P0S groupings is that we will be able to capture functionally similar words which, since they are compared to a subjective set of seeds, will also be subjective. In the case of the "all" grouping, we expect a stronger measure of subjectivity to emerge, as subjective dimensions end up being added together.
14. http://infomap-nlp.sourceforge.net.

newspapers, a publicly available Romanian literature accessible from WikiSource, the Romanian Wikipedia, and the manual translation of a subset of the SemCor data set [38] into Romanian. The Spanish corpus is similar in composition to the Romanian data set, yet we only used one-seventh of the Spanish Wikipedia, no editorials, and no SemCor data. We limited the size of this corpus to create settings that are as similar as possible to the experiments conducted in Romanian. Also, our choice to add literature works or editorials to the mix is motivated by the need to increase the occurrence of potentially subjective entries in the corpora. Since Wikipedia is considered to be an encyclopedic resource, its usage of subjective entries may be limited, therefore impairing LSA's ability in calculating similarity among subjective words.

We initially proposed this method in [55]. The current work adds the following improvements: 1) we expand the measure of word relatedness to include either a subset of subjective seeds (based on part of speech information), or the entire set; 2) we extend our experiments to Spanish; and 3) we use a larger corpus to train the LSA model to obtain more accurate relatedness measures.

3.3 Gold Standard

We evaluate our results against a gold standard consisting of 504 sentences extracted from the English SemCor corpus [38]. These sentences were manually translated into Romanian panish, resulting in two parallel test sets for the two languages. Two Romanian native speakers annotated the Romanian sentences individually, and the differences were adjudicated through discussions. The agreement of the two annotators is 0.83 percent ($\kappa = 0.67$); 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 data set with 272 (54 percent) subjective sentences and 232 (46 percent) objective sentences. The same subjectivity annotations developed for Romanian are also ported to the Spanish test set. The test set is further processed by removing diacritics and any nonliteral characters. The corpus-based methods described in Section 4 use this version of the test sets. For the lexicon-based methods proposed earlier in this section, the test data are further lemmatized to allow for a match with the automatically extracted candidates from the dictionaries. Lemmatization for Romanian is performed automatically using the module provided by the LanguageWeaver translation engine.[15] For Spanish, we use the TreeTagger program.[16] More details about this data set are available in [42].

3.4 Evaluation and Discussion

Our experiments suggest that five bootstrapping iterations are sufficient to extract a subjectivity lexicon, as the number of features saturates during the last iteration. Figs. 2, 3, and 4, available in the online supplementary material, exemplify the lexicon acquisition in both Romanian and Spanish through all five iterations, using different filtering variations as proposed in Section 3.2.3. As suggested by the graphs, as a stricter similarity between the seeds and the candidates is enforced, a lower number of entries are extracted. Also, the number of entries extracted in Romanian are consistently fewer when compared to Spanish by a factor of at least 1 to 2, resulting in a lower lexicon coverage in this language. For the various evaluations presented further, we use an LSA threshold of 0.5.

The variations we proposed for the filtering step during the bootstrapping process resulted in three lexicons for Romanian and Spanish, respectively, which are evaluated by using them with a rule-based sentence-level subjectivity classifier. Briefly, the rule-based algorithm labels as subjective a sentence that contains two or more entries that appear in the subjectivity lexicon, and as objective a sentence that has one or fewer entries, respectively. The algorithm is derived based on the rules described in [37], which were modified to account for the fact that no strong/weak confidence labels are available.

The sentence-level subjectivity classification results are shown in Tables 7 and 8, available in the online supplementary material.[17] By using the extracted lexicons alone, we were able to obtain a rule-based subjectivity classifier with an overall F-measure of 64.29 (*seed*), 56.94 (*POS group*), and 52.38 percent (*all*), for Romanian, and 57.54 (*seed*), 64.09 (*POS group*), 69.64 percent (*all*), for Spanish. The *seed* variation of the bootstrapping process entails a more lax similarity, because it focuses on extracting candidates based on their closeness to an individual seed word, and therefore, is able to extract the largest lexicons in our experiments both in Romanian and Spanish (see Fig. 2, available in the online supplementary material). The other two variations gradually enforce a stricter similarity, as the comparison is first made against 15 seeds for the *POS group*, and then against 60 seeds for the *all* set. Each candidate in this case not only has to display similarity with one of the seeds, but with each and every element composing the set. For this reason, the *POS group* variation is able to extract a medium size lexicon (Fig. 3, available in the online supplementary material), while the *all* variation derives the most compact and highly correlated lexicon (Fig. 4, available in the online supplementary material), in both Romanian and Spanish.

As noticed from Tables 7 and 8, available in the online supplementary material, Romanian seems to exhibit an opposite F-measure pattern over the method variations when compared to Spanish. Yet, this is only a superficial assessment. As shown in Figs. 5 and 6, available in the online supplementary material, the subjectivity precision and recall curves perform very similar in both Romanian and Spanish. The only factor that creates a discrepancy is the low recall level obtained for Romanian when using the *all* variation. This anomaly results from the extremely low coverage of the Romanian subjectivity lexicon extracted by this method (at most 300 entries), and should we have had access to a more elaborate Romanian dictionary, and through it to a reacher set of candidates, we expect the behavior of the recall curves to correct and be consistent in the three variations of the method across the two languages.

15. http://www.languageweaver.com.

16. Developed by Helmut Schmid as part of the TC project at the Institute for Computational Linguistics, University of Stuttgart.

17. We used the *Statistics::Contingency* Perl module for all the evaluations; the reported *overall* precision, recall and F-measure are microaveraged.

TABLE 5
Precision (P), Recall (R), and F-Measure (F) for the Automatic Translation of the Subjectivity Lexicon of Strength Weak and Strong

	Romanian			Spanish		
	All	Subj.	Obj.	All	Subj.	Obj.
Most Frequent Sense						
P	55.95%	65.84%	51.31%	58.53%	66.00%	53.62%
R	55.95%	38.83%	76.19%	58.53%	48.35%	70.56%
F	55.95%	**48.85**%	61.32%	58.53%	55.81%	60.93%
Part-of-speech Based Sense						
P	56.75%	67.97%	51.85%	59.72%	66.99%	54.70%
R	56.75%	38.10%	78.79%	59.72%	50.55%	70.56%
F	**56.75**%	48.83%	**62.54**%	**59.72**%	**57.62**%	**61.63**%

We can, therefore, conclude that if all the bootstrapping variations are able to extract a lexicon of around 1,000 entries, then the best results should be obtained by the setup enforcing the strictest similarity (*all*).

Furthermore, Tables 7 and 8, available in the online supplementary material, show that the best overall F-measure and the best subjective F-measure seem to be obtained using the lexicon generated after the fifth iteration, which provides a consistent classification pattern. For Spanish, for example, the highest subjective F-measure obtained in the last iteration is 74.20 percent, whereas for Romanian is 65.25 percent.

We compare our results with those obtained by automatically translating a subjectivity lexicon, as described in Section 3.1. In that method, a subjectivity lexicon is automatically obtained through the translation of the *strong* and *weak* entries composing the English subjectivity lexicon available in OpinionFinder. The translation focuses on automatically acquiring *the most frequent sense*, regardless of the part-of-speech information adjoining the original English entry. Two lexicons are obtained using this method: A Romanian lexicon consisting of 1,580 entries, and a Spanish lexicon of 2,009 entries, respectively. We apply the same rule-based classifier and use the same gold-standard data set as used in testing the bootstrapping method. The results are shown in the top part of Table 5.

We also conduct an additional experiment (*part-of-speech-based sense*) to test whether the automatic translation of the lexicon may perform better if we utilize the part-of-speech annotation from the English subjectivity lexicon to disambiguate among candidate translations in the target language. We start with the subjectivity lexicon annotated with the corrected part-of-speech tag (explained in Section 3.1). We then use the same Ectaco dictionary for performing lexicon translations into Romanian and Spanish. If the online dictionary offers a translation candidate for the part-of-speech projected from English, then we accept the first entry among the translations. Otherwise, the English entry is disregarded. The resulting lexicons in both Romanian and Spanish are about 50 words shorter (Romanian: 1,509 entries, Spanish: 1,953 entries) when compared to the more generic first sense translation counterparts. Despite the additional effort entailed in correcting the part-of-speech and enforcing it in a bilingual dictionary, the results show only a marginal 1 percent improvement in F-measure for both languages (see the bottom part of Table 5).

To asses whether the best option to extract a subjectivity lexicon in a target language is by direct translation of a source language subjectivity lexicon, or by growing the target language lexicon from a small number of seeds, we compare Tables 7 and 8, available in the online supplementary material, showing the bootstrapping results, with Table 5, which shows the lexicon translation results. The overall F-measure obtained for Romanian after the fifth iteration, a LSA threshold of 0.5 and the *seed* variation is 64.29 percent, being higher by 8.34 or 7.54 percent when compared to the *most frequent sense* or the *part-of-speech-based sense* overall F-measure, respectively. The results for Spanish are even more compelling, as the best bootstrapping results achieved under the *POS* variation is 69.64 percent F-measure, while the *most frequent sense* and the *part-of-speech-based sense* reach 58.53 and 59.72 percent. It is also important to note that the bilingual dictionary extracts more candidates than the bootstrapping process is able to, yet they appear less frequently in the target language, therefore hampering the recall of the rule-based classifier, and undermining a high-subjectivity F-measure. Based on these aspects, we can, therefore, conclude that a more reliable subjectivity lexicon can be extracted directly in a target language, instead of requiring a sizeable subjectivity lexicon in a source language and a bridging bilingual dictionary.

4 Leveraging on Sentence-Level Subjectivity Annotations

In this section, we explore the potential of leveraging on sentence-level subjectivity annotations ported from the source language to generate the resources required to build subjectivity annotation tools in a given target language. We, therefore, explore two scenarios we presented in [44]. 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 a manually or automatically generated parallel text to create a corpus annotated for subjectivity in the target language.

To perform a comprehensive investigation, we propose four 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 three experiments depending on the type of parallel text (manual or automatic), and the direction of the translation. Given a source and a target language, the choice of the best scenario to be used depends on the resources available for that source and target language.

In all four experiments, we use English as a source language, given that it has both a corpus manually annotated for subjectivity (MPQA [17]) and a tool for subjectivity analysis (OpinionFinder [36]).

4.1 Manually Annotated Corpora

Experiment 1: Machine translation of manually annotated corpora. In this experiment, we start from a corpus in the source language, which is manually annotated for subjectivity. The data are transferred into the target language through the intercession of a machine translation engine, and then it is augmented with subjectivity labels projected from the source language annotations.

We use the MPQA corpus (see Section 1.4). After the automatic translation of the corpus 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.

4.2 Automatically Annotated Corpora

As a raw corpus, we use a subset of the English SemCor corpus (introduced in Section 1.4) that excludes the gold-standard data set described in Section 3.3. This text or a machine translated version of it into English is processed with the aid of the high-coverage classifier embedded in the OpinionFinder tool (see Section 1.4), resulting in automatic sentence-level subjectivity annotations.

Experiment 2: Manually translated parallel text. In the second experiment, we assume we have access to a tool for subjectivity annotation developed for the source language, and to a manually constructed parallel text in both the source and target language. We are, therefore, able to generate automatic subjectivity annotations for the source data set, which we can project via the parallel text into the target language. This scenario results in an automatically annotated subjectivity data set in the target language, which can be used to train an automatic classifier in that language.

Since this experiment uses manually translated parallel text, both the subjectivity annotation tool in the source language and the training of the classifier in the target language are performed on correct human generated text. Therefore, this setup should provide better results when compared to the machine translation experiments.

Experiment 3: Machine translation of source language training data. In the third experiment, we hypothesize 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.

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 are produced by automatic translation, and thus likely to contain errors.

Experiment 4: Machine translation of target language training data. The fourth experiment is similar to the third 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.

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

Upper bound (UB): Machine translation of target language test data. For comparison purposes, we also propose an experiment that plays the role of an UB on the machine translation methods proposed in Section 4. It 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 the projection of the resulting labels back into the target language.

Unlike the previous experiments, in this setup we only generate subjectivity-annotated *resources*, and we do not build and evaluate a stand-alone subjectivity analysis *tool* for the target language. Further training of a machine learning algorithm, as in experiments three and four, is required to build a subjectivity analysis tool. Thus, this study is an evaluation of the *resources* generated in the target language, which represents an UB on the performance of any machine learning algorithm that would be trained on these resources.

4.3 Evaluation and Results

Our evaluations are carried out on Romanian and Spanish (where data availability allows). The performance of each of the four experiments proposed in this section is evaluated using the same gold-standard described in Section 3.3. To evaluate our methods, we generate a training corpus annotated for subjectivity based on projections from the source language, for both Romanian and Spanish. We feed the document-label pair to a machine learner under the hypothesis that these labels are the accurate annotations for the target sentences. We assume that any possible translation or annotation errors in the training data will be eventually voted out during the training of the classifiers. For learning, we use a state-of-the-art learning algorithm, namely SVM.

Support vector machines (SVM) [57], [58] is a machine learning approach based on decision planes. The algorithm tries to render the optimal hyper-plane that separates the set of points associated with different class labels resulting in a maximum-margin. The unlabeled examples are then classified by deciding on which side of the hypersurface they reside. This type of learning is advantageous with a noisy training data such as the one provided in our scenario. We use the implementation available in the AI::Categorizer PERL module with a linear kernel, because it was proved to be as powerful as other kernels in text classification experiments [59].

We also apply feature selection, and keep the top 50 percent discriminating features, according to a $tf.idf$ weighting scheme using raw term frequencies normalized by the length of the document vector [60].

For Romanian, the automatic translation of the MPQA and of the SemCor corpus was performed using Language Weaver,[18] a commercial statistical machine translation

18. http://www.languageweaver.com/.

TABLE 6
Precision (P), Recall (R), and F-Measure (F) for Section 4 Experiments; *Manual Subjectivity Annotations*: E1—Source to Target Language Machine Translation; *Automatic Subjectivity Annotations*: E2—Parallel Text, E3—Source to Target Language Machine Translation, E4—Target to Source Language Machine Translation; *Upper Bound*: UB

		Romanian			Spanish		
Exp.		All	Subj.	Obj.	All	Subj.	Obj.
E1	P	66.07%	59.00%	70.73%	68.85%	67.90%	70.56%
	R	66.07%	91.21%	25.11%	68.85%	80.59%	54.98%
	F	66.07%	71.65%	37.06%	**68.85**%	**73.70**%	61.80%
E2	P	69.64%	69.35%	70.10%			
	R	69.64%	78.75%	58.87%			
	F	69.64%	73.76%	64.00%			
E3	P	69.44%	67.76%	72.78%	63.89%	73.82%	57.83%
	R	69.44%	83.15%	53.25%	63.89%	51.65%	78.35%
	F	**69.44**%	**74.67**%	61.50%	63.89%	60.78%	**66.54**%
E4	P	67.86%	76.06%	61.86%			
	R	67.86%	59.34%	77.92%			
	F	67.86%	66.67%	**68.97**%			
UB	P	71.83%	71.91%	71.71%	73.41%	73.88%	72.77%
	R	71.83%	78.75%	63.64%	73.41%	78.75%	67.10%
	F	71.83%	75.17%	67.43%	73.41%	76.24%	69.82%

software. For Spanish, both corpora were translated using the Google Translation Service,[19] a publicly available machine translation engine also based on statistical machine translation. Due to the fact that we did not have access to a manual translation into Spanish of the SemCor training data, we were unable to replicate Experiments 2 and 4 for this language. We could have used another Spanish text instead, but, because the data set would have been different, the results would not have been directly comparable. The resulting text in the target languages was postprocessed by removing diacritics, stopwords, and numbers.

The results obtained by running the four experiments on Romanian and Spanish are shown in Table 6. The baseline on this data set is 54.16 percent, represented by the percentage of sentences in the corpus that are subjective, and the UB is 71.83 percent, which is the accuracy obtained under the scenario where the test data are translated into the source language and then annotated using the high-coverage OpinionFinder tool. All the results are statistically significant at $p < 0.05$ when compared to a random output following the class distribution present in the gold standard.

For Romanian, 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 classification is used on either manual or automatically generated text. 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, seem 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 its structure:

$$\textit{It is the Palestinians that are calling for the implementation of the agreements, understandings, and recommendations pertaining to the Palestinian-Israeli conflict.} \quad (1)$$

A unigram classifier will not be able to determine that the author of this sentence is using word topology to express his opinion about the conflict. The writer is able to state and emphasize his perspective even though it is not overtly expressed in the sentence by any given word. A learning algorithm will not find the above sentence very different from *Palestinians are calling for the implementation of the agreements, understandings, and recommendations pertaining to the Palestinian-Israeli conflict*, especially once stopwords are removed.

The results of the two experiments we were able to carry out on Spanish are shown in Table 6. Interestingly, the F-measure is higher for the first experiment involving the machine translated version of a corpora manually labeled for subjectivity, than its counterpart in Romanian. We assume this happens because Spanish is one of the six official United Nations languages; thus, larger amounts of parallel texts are available to train the machine translation system, which implies that a better quality translation can be achieved as compared to the one available for Romanian. 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 third experiment, surpassing the overall F-measure calculated for the third experiment by 4.96 percent. The MPQA corpus, because it is manually annotated and of better quality, has a higher chance of generating a more reliable data set in the target language. Unlike the results obtained for Romanian in Experiments 3 and 4, in Spanish, the classifier is not able to distinguish as well the subjective cases, reaching a subjectivity F-measure of only 60.78 percent, and thus penalizing the overall F-measure. 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 percent, which represents the UB on our proposed experiments.

Among the approaches proposed in this section, Experiments 3 and 4 are closest to the experiment based on parallel text (second experiment). By using machine translation, from English into Romanian or Spanish (Experiment 3) or Romanian into English (Experiment 4), and annotating this data set with the high-coverage OpinionFinder classifier using an SVM learning algorithm, we obtain an F-measure of 63.89/ 69.44 percent (Experiment 3), and 67.86 percent, respectively (Experiment 4). This implies that using a parallel corpus does not produce significantly better results when compared to automatically generated translations, especially when the training set was automatically annotated. This finding further suggests that machine translation is a viable alternative to devising subjectivity classification in a target language leveraged on the tools existent in a source language. Despite the loss in readability quality of the resulting automatic text, its subjectivity content seems to be mostly rendered correctly. The statistical machine translation engines are better equipped to transfer the subjective sense of a word because they disambiguate it based on n-gram language models.

19. http://www.google.com/translate_t.

To illustrate this argument, consider the following example:

I boil with anger.

Google translation in Romanian: *Am fiert cu furie.* (2)

Google translation in Spanish: *Me hierve con rabia.*

If the direct translation experiments described in Section 3.1 translate the verb *boil* into noun *Ro: fierbere* (as in *He brought a kettle of water to a boil.*[20]) and noun *Es: furúnculo* (with the meaning of *En: furuncle*), the machine translation correctly identifies that *boil* is used in its verb sense. Even though the preposition (*cu*) in Romanian is not the one a native would employ, the translation of the sentence ensures that its subjective meaning, and components, are accurately rendered into the target language.

Finally, we also wanted to explore the impact that the corpus size may have on the accuracy of the classifiers. We reran Experiments 3 and 4 with 20 percent corpus size increments at a time (Fig. 7, available in the online supplementary material). It is interesting to note that a corpus of approximately 6,000 sentences is able to achieve a high enough F-measure (around 66 percent for both experiments in Romanian and 63 percent for the Spanish experiment) to be considered viable for training a subjectivity classifier.

5 Conclusions and the Big Picture

In this paper, we explored various methods to transfer subjectivity information from a source language to a target language. We conclude the paper by summarizing our findings, and discussing what methods worked best and why.

We were surprised to find that automatically growing a subjectivity lexicon in a target language starting from a small number of manually translated seeds outperforms the automatic translation of a fully developed subjectivity lexicon in a source language. The amount of manual work entailed in constructing from scratch such resources is enormous, as it requires not only training annotators on identifying subjective entities in a language, but also manually tagging a large data set, and based on the occurrence of the subjective context decide whether it is reliable enough to be included in a specialized lexicon. Compared to this method, translating a small number of seeds is a trivial task. Furthermore, obtaining an explicative dictionary in the target language and a raw corpus that can be used for extracting similarity information should not pose significant problems. Upon implementing such a bootstrapping system, we can expect up to 7 percent improvement in the overall F-measure over the method based on direct translation. It is also interesting to observe that despite the fact that the bootstrapping method is almost completely unsupervised (except for the initial seed set), its results are competitive with those obtained by the machine learning methods.

The machine learning experiments we proposed suggest that the use of a manual or a machine translated parallel text annotated with an automatic subjectivity analysis tool in the major language provide similar results (within 2 percent overall F-measure variation for both Romanian and Spanish). This suggests that all the features that could have been leveraged from the automatic annotations are already taken into consideration by the machine learning algorithm.

It is also interesting to note that the best results obtained with our projection methods are only a few percentages below what has been previously reported for experiments on English with large manually annotated data [61]. This suggests that the crafting of manual resources to obtain performance figures similar to ours can be much more time consuming and expensive than using such projection methods.

When faced with a new language, what is the best method that one can use to create a subjectivity analysis tool for that language? The answer largely depends on the monolingual resources and tools that are available for that language, e.g., dictionaries, large corpora, natural language processing tools, and/or the cross-lingual connections that can be made to a major language[21] such as English, e.g., bilingual dictionaries or parallel texts. Of course, the quality/coverage of the resources also has a bearing on how the results will compare to those we have obtained here. For languages with very scarce electronic resources, options other than translating a subjectivity lexicon may not be feasible. We are encouraged, however, by the work conducted by Kim et al. [62] applying our methods to new languages such as Korean, Chinese, and Japanese, and obtaining comparable results. This supports the Big Picture crayoned here, at least for the more common languages.

Best scenario: Manually annotated corpora. The best scenario is when a corpus manually annotated for subjectivity exists in the target language. Unfortunately, this is rarely the case, as large manually annotated corpora exist only for a handful of languages, e.g., the MPQA corpus that is available for English [17]. If a large annotated data set is available, a tool for automatic annotation can be easily constructed by training a machine learning system. The task can be, thus, regarded as a text classification problem, and learning algorithms such as Naïve Bayes, or SVM can be used to annotate the subjectivity of new text.

Second best: Corpus-based cross-lingual projections. The second best option is to construct an annotated data set by doing cross-lingual projections from a major language that has such annotations readily available. This assumes that a "bridge" can be created between the target language and a major language such as English, in the form of parallel texts constructed via manual or automatic translations. By using this bridge, the corpus annotations available in the major language can be automatically transferred into the target language.

The translation can be performed in two directions. First, one can take a collection of texts in the major language and manually or automatically translate it into the target language. In this case, if the source text is already manually annotated for subjectivity (e.g., MPQA), then the manual labels can be projected into the target language. Alternatively, the text in the major language can be automatically annotated by using subjectivity analysis tools such as

20. Example obtained from http://dictionary.reference.com/.

21. That is, a language for which many resources and tools are already available.

OpinionFinder [36]. The other option is to start with texts in the target language and translate them into the major language. Again, the translation can be done either by hand, or automatically. Regardless of the direction of the translation or the way it was obtained, the result is a data set in the target language annotated for subjectivity, which can be used to train an automatic classifier, as shown in Section 4.2.

Third best: Bootstrapping a lexicon. There are several methods that rely on the availability of subjectivity lexicons to build rule-based classifiers for the annotation of new text. For instance, one of the most frequently used subjectivity annotation tools for English is OpinionFinder [36], which is based on a large subjectivity lexicon [20].

One of the most successful approaches for the construction of subjectivity lexicons is to bootstrap from a few manually selected seeds. The bootstrapping can be performed using the synonyms and definitions found in an electronic dictionary, as illustrated in Section 3.2. No advanced language processing tools are required for this method, only a dictionary in the target language. Starting with a set of seeds covering all open-class words, all the related words found in the dictionary are collected, including the synonyms, antonyms, and words found in the definitions. From this set of candidates, only those that are similar to the seeds are kept for the next bootstrapping iteration. Running the process for several iterations can result in large lexicons with several thousands entries.

Fourth best: Translating a lexicon. If none of the previous methods is applicable, a lexicon can be constructed by automatically translating an already existing lexicon from a major language. The only requirements for this approach are a subjectivity lexicon in a source language, and a bilingual dictionary used to automatically translate the lexicon into the target language. The method is described in Section 3.1. Although very simple and efficient (a lexicon of over 5,000 entries can be created in seconds), the accuracy of the method is rather low, mainly due to the challenges typical to a context-free translation process: Difficulty in selecting the most appropriate translation for ambiguous words; small coverage for phrase translations; mismatch between the inflected forms appearing in the lexicon and the bilingual dictionary entry. Even so, such a lexicon can be corrected by hand, and provide a building block for the generation of subjectivity resources in a given target language.

Acknowledgments

This material was based in part upon the work supported by the National Science Foundation Awards #0917170 and #0916046. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] C.O. Alm, D. Roth, and R. Sproat, "Emotions from Text: Machine Learning for Text-Based Emotion Prediction," *Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing (HLT-EMNLP '05)*, pp. 579-586, 2005.

[2] L. Lloyd, D. Kechagias, and S. Skiena, "Lydia: A System for Large-Scale News Analysis," *Proc. 12th Int'l Conf. String Processing and Information Retrieval*, pp. 161-166, 2005.

[3] K. Balog, G. Mishne, and M.D. Rijke, "Why Are They Excited? Identifying and Explaining Spikes in Blog Mood Levels," *Proc. 11th Conf. European Chapter of the Assoc. Computational Linguistics: Posters and Demonstrations (EACL '06)*, 2006.

[4] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," *Proc. ACM 10th SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '04)*, pp. 168-177, 2004.

[5] H. Yu and V. Hatzivassiloglou, "Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentence," *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP '03)*, pp. 129-136, 2003.

[6] G. Carenini, R.T. Ng, and X. Zhou, "Summarizing Emails with Conversational Cohesion and Subjectivity," *Proc. 46th Ann. Meeting Assoc. Computational Linguistics: Human Language Technologies (ACL-HLT '08)*, pp. 353-361, 2008.

[7] N. Kobayashi, K. Inui, K. Tateishi, and T. Fukushima, "Collecting Evaluative Expressions for Opinion Extraction," *Proc. First Int'l Joint Conf. Natural Language Processing (IJCNLP '04)*, pp. 596-605, 2004.

[8] H. Takamura, T. Inui, and M. Okumura, "Latent Variable Models for Semantic Orientations of Phrases," *Proc. Conf. European Chapter of the Assoc. Computational Linguistics (EACL '06)*, pp. 201-208, 2006.

[9] H. Kanayama and T. Nasukawa, "Fully Automatic Lexicon Expansion for Domain-Oriented Sentiment Analysis," *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP '06)*, pp. 355-363, 2006.

[10] Y. Hu, J. Duan, X. Chen, B. Pei, and R. Lu, "A New Method for Sentiment Classification in Text Retrieval," *Proc. Second Int'l Joint Conf. Natural Language Processing (IJCNLP '05)*, pp. 1-9, 2005.

[11] S.-M. Kim and E. Hovy, "Identifying and Analyzing Judgment Opinions," *Proc. Main Conf. Human Language Technology Conf. North Am. Chapter of the Assoc. Computational Linguistics (HLT-NAACL '06)*, pp. 200-207, 2006.

[12] Y. Seki, D. Evans, L. Ku, H. Chen, N. Kando, and C. Lin, "Overview of Opinion Analysis Pilot Task at NTCIR-6," *Proc. Sixth NTCIR Workshop Meeting Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access*, pp. 265-278, 2007.

[13] C. Banea, R. Mihalcea, and J. Wiebe, "Multilingual Subjectivity: Are More Languages Better?" *Proc. 23rd Int'l Conf. Computational Linguistics (COLING '10)*, pp. 28-36, 2010.

[14] S. Hassan, C. Banea, and R. Mihalcea, "Measuring Semantic Relatedness Using Multilingual Representations," *Proc. Workshop Semantic Evaluations (SemEval '12)*, 2012.

[15] W. Wei, J. Gulla, and Z. Fu, "Enhancing Negation-Aware Sentiment Classification on Product Reviews via Multi-Unigram Feature Generation," *Proc. Sixth Int'l Conf. Advanced Intelligent Computing Theories and Applications*, vol. 6215/2010, pp. 380-391, 2010.

[16] X. Wan, "Co-Training for Cross-Lingual Sentiment Classification," *Proc. Joint Conf. 47th Ann. Meeting ACL and Fourth Int'l Joint Conf. Natural Language Processing of the AFNLP (ACL-IJCNLP '09)*, vol. 1, pp. 235-243, 2009.

[17] J. Wiebe, T. Wilson, and C. Cardie, "Annotating Expressions of Opinions and Emotions in Language," *Language Resources and Evaluation*, vol. 39, no. 2/3, pp. 165-210, 2005.

[18] R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik, *A Comprehensive Grammar of the English Language.* Longman, 1985.

[19] J. Wiebe and R. Mihalcea, "Word Sense and Subjectivity," *Proc. 21st Int'l Conf. Computational Linguistics and 44th Ann. Meeting Assoc. Computational Linguistics (COLING-ACL '06)*, pp. 1065-1072, 2006.

[20] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing (HLT-EMNLP '05)*, pp. 347-354, 2005.

[21] A. Andreevskaia and S. Bergler, "Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses," *Proc. Conf. European Chapter of the Assoc. Computational Linguistics (EACL '06)*, pp. 209-216, 2006.

[22] A. Esuli and F. Sebastiani, "Determining Term Subjectivity and Term Orientation for Opinion Mining," *Proc. Conf. European Chapter of the Assoc. Computational Linguistics (EACL '06)*, vol. 2, pp. 193-200, 2006.

[23] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," *Proc. 42nd Ann. Meeting Assoc. Computational Linguistics (ACL '04)*, pp. 271-278, 2004.

[24] P.J. Stone, M.S. Smith, D.M. Ogilivie, and D.C. Dumphy, *The General Inquirer: A Computer Approach to Content Analysis*, first ed. MIT Press, 1966.

[25] V. Hatzivassiloglou and K.R. McKeown, "Predicting the Semantic Orientation of Adjectives," *Proc. Conf. European Chapter of the Assoc. Computational Linguistics (EACL '97)*, pp. 174-181, 1997.

[26] P.D. Turney and M.L. Littman, "Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus," Technical Report ERB-1094, Nat'l Research Council Canada, Inst. of Information Technology, 2002.

[27] M.A. Gamon and A. Aue, "Automatic Identification of Sentiment Vocabulary: Exploiting Low Association with Known Sentiment Terms," *Proc. ACL Workshop Feature Eng. for Machine Learning in Natural Language Processing*, pp. 57-64, 2005.

[28] E. Riloff, J. Wiebe, and T. Wilson, "Learning Subjective Nouns Using Extraction Pattern Bootstrapping," *Proc. Seventh Conf. Natural Language Learning (CoNLL '03)*, pp. 25-32, 2003.

[29] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP '02)*, vol. 10, pp. 79-86, 2002.

[30] J. Wiebe and T. Wilson, "Learning to Disambiguate Potentially Subjective Expressions," *Proc. Conf. Natural Language Learning (CoNLL '02)*, pp. 1-7, 2002.

[31] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment Analyzer: Extracting Sentiments About a Given Topic Using Natural Language Processing Techniques," *Proc. IEEE Int'l Conf. Data Mining (ICDM '03)*, pp. 427-434, 2003.

[32] E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions," *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP '03)*, pp. 105-112, 2003.

[33] K. Bloom, N. Garg, and S. Argamon, "Extracting Appraisal Expressions," *Proc. Human Language Technologies/North Am. Assoc. Computational Linguists (NAACL-HLT '07)*, pp. 308-315, 2007.

[34] H. Takamura, T. Inui, and M. Okumura, "Extracting Semantic Orientations of Phrases from Dictionary," *Proc. Human Language Technologies/North Am. Assoc. Computational Linguists (NAACL-HLT '07)*, pp. 292-299, 2007.

[35] T. Zagibalov and J.A. Carroll, "Automatic Seed Word Selection for Unsupervised Sentiment Classification of Chinese Text," *Proc. 22nd Int'l Conf. Computational Linguistics (COLING '08)*, vol. 1, pp. 1073-1080, 2008.

[36] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, "OpinionFinder: A System for Subjectivity Analysis," *Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing (HLT-EMNLP) on Interactive Demonstrations*, pp. 34-35, 2005.

[37] J. Wiebe and E. Riloff, "Creating Subjective and Objective Sentence Classifiers from Unannotated Texts," *Proc. Sixth Int'l Conf. Computational Linguistics and Intelligent Text Processing (CICLing '05)*, pp. 486-497, 2005.

[38] G.A. Miller, C. Leacock, T. Randee, and R.T. Bunker, "A Semantic Concordance," *Proc. Workshop Human Language Technology (HLT '93)*, pp. 303-308, 1993.

[39] A. Esuli, F. Sebastiani, and I.C. Urciuoli, "Annotating Expressions of Opinion and Emotion in the Italian Content Annotation Bank," *Proc. Int'l Conf. Language Resources and Evaluation (LREC '08)*, pp. 2247-2251, 2008.

[40] L.-W. Ku, Y.-T. Liang, and H.-H. Chen, "Opinion Extraction, Summarization and Tracking in News and Blog Corpora," *Proc. AAAI Spring Symp. Computational Approaches to Analyzing Weblogs*, pp. 100-107, 2006.

[41] Y. Wu, "Classifying Attitude by Topic Aspect for English and Chinese Document Collections," PhD dissertation, Univ. of Maryland, College Park, MD, 2008.

[42] R. Mihalcea, C. Banea, and J. Wiebe, "Learning Multilingual Subjective Language via Cross-Lingual Projections," *Proc. 45th Ann. Meeting. Assoc. Computational Linguistics (ACL '07)*, pp. 976-983, 2007.

[43] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, Boom-Boxes and Blenders: Domain Adaptation for Sentiment Classification," *Proc. 45th Ann. Meeting. Assoc. Computational Linguistics (ACL '07)*, pp. 440-447, 2007.

[44] C. Banea, R. Mihalcea, J. Wiebe, and S. Hassan, "Multilingual Subjectivity Analysis Using Machine Translation," *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP '08)*, pp. 127-135, 2008.

[45] M. Bautin, L. Vijayarenu, and S. Skiena, "International Sentiment Analysis for News and Blogs," *Proc. Int'l Conf. Weblogs and Social Media (ICWSM '08)*, pp. 19-26, 2008.

[46] J. Wiebe, "Learning Subjective Adjectives from Corpora," *Proc. 17th Nat'l Conf. Artificial Intelligence and 12th Conf. Innovative Applications of Artificial Intelligence (AAAI '00)*, pp. 735-740, 2000.

[47] G.A. Miller, "WordNet: A Lexical Database for English," *Comm. Assoc. for Computing Machinery*, vol. 38, no. 11, pp. 39-41, 1995.

[48] D. Rao and D. Ravichandran, "Semi-Supervised Polarity Lexicon Induction," *Proc. 12th Conf. Empirical Methods in Natural Language Processing (EACL '09)*, pp. 675-682, 2009.

[49] P. Prettenhofer and B. Stein, "Cross-Language Text Classification Using Structural Correspondence Learning," *Proc. 48th Ann. Meeting Assoc. Computational Linguistics (ACL '10)*, pp. 1118-1127, 2010,

[50] J. Boyd-Graber and P. Resnik, "Holistic Sentiment Analysis across Languages: Multilingual Supervised Latent Dirichlet Allocation," *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP '10)*, pp. 45-55, 2010.

[51] B. Lu, C. Tan, C. Cardie, and B.K. Tsou, "Joint Bilingual Sentiment Classification with Unlabeled Parallel Corpora," *Proc. 49th Ann. Meeting Assoc. Computational Linguistics: Human Language Technologies (ACL-HLT '11)*, pp. 320-330, 2011,

[52] P.D. Turney, "Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL," *Proc. 12th European Conf. Machine Learning (ECML '01)*, pp. 491-502, 2001.

[53] S.T. Dumais, G.W. Furnas, T.K. Landauer, S. Deerwester, and R. Harshman, "Using Latent Semantic Analysis to Improve Access to Textual Information," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '88)*, pp. 281-285, 1988.

[54] T.K. Landauer, P.W. Foltz, and D. Laham, "An Introduction to Latent Semantic Analysis," *Discourse Processes*, vol. 25, pp. 259-284, 1998.

[55] C. Banea, R. Mihalcea, and J. Wiebe, "A Bootstrapping Method for Building Subjectivity Lexicons for Languages with Scarce Resources," *Proc. Int'l Conf. Language Resources and Evaluation (LREC '08)*, 2008.

[56] R. Ghani, R. Jones, and D. Mladenic, "Mining the Web to Create Minority Language Corpora," *Proc. 10th Int'l Conf. Information and Knowledge Management (CIKM '01)*, 2001.

[57] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.

[58] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Proc. European Conf. Machine Learning (ECML '98)*, pp. 137-142, 1998.

[59] Y. Yang and X. Liu, "A Reexamination of Text Categorization Methods," *Proc. 22nd Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (ACM-SIGIR '99)*, 1999.

[60] G. Salton and C. Buckley, "Term Weighting Approaches in Automatic Text Retrieval," *Readings in Information Retrieval*, Morgan Kaufmann, 1997.

[61] Y. Li, K. Bontcheva, and H. Cunningha, "Experiments of Opinion Analysis on two Corpora MPQA and NTCIR-6," *Proc. NTCIR Workshop Meeting. Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access (NTCIR '07)*, pp. 323-329, 2007.

[62] J. Kim, J.-J. Li, and J.-H. Lee, "Evaluating Multilanguage-Comparability of Subjectivity Analysis Systems," *Proc. Ann. Meeting Assoc. Computational Linguistics (ACL '10)*, pp. 595-603, 2010.

[63] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.

Carmen Banea received the doctoral degree in computer science from the University of North Texas in 2013. She is a postdoctoral fellow in the Department of Computer Science at the University of North Texas, where she works on research in the field of natural language processing. Her research work focuses primarily on multilingual approaches to subjectivity analysis, having developed both dictionary and corpus-based methods that leverage on languages with rich resources to create tools and data in other languages. She published numerous research papers in major NLP conferences (i.e., ACL, EMNLP, LREC), including a paper that received the IEEE Best Student Paper Award, and she coauthored a chapter on multilingual sentiment and subjectivity in the book entitled *Multilingual Natural Language Applications: From Theory to Practice* (Prentice Hall, 2010).

Rada Mihalcea is an associate professor in the Department of Computer Science and Engineering at the University of North Texas. Her research interests are in computational linguistics, with a focus on lexical semantics, graph-based algorithms for natural language processing, and multilingual natural language processing. She serves or has served on the editorial boards of the *Journals of Computational Linguistics, Language Resources and Evaluation*s, *Natural Language Engineering*, *Research in Language in Computation*, *IEEE Transactions on Affective Computing,* and *Transactions of the Association for Computational Linguistics*. She was a program cochair for the Conference of the Association for Computational Linguistics (2011), and the Conference on Empirical Methods in Natural Language Processing (2009). She is the recipient of a National Science Foundation CAREER Award (2008) and a Presidential Early Career Award for Scientists and Engineers (2009).

Janyce Wiebe is a professor of computer science and the director of the Intelligent Systems Program at the University of Pittsburgh. Her research with students and colleagues has been in discourse processing, pragmatics, and word-sense disambiguation. A major concentration of her research is "subjectivity analysis," recognizing and interpreting expressions of opinions and sentiments in text, to support NLP applications such as question answering, information extraction, text categorization, and summarization. Her professional roles have included ACL program cochair, NAACL program chair, NAACL Executive Board member, Computational Linguistics and Language Resources and Evaluation Editorial Board member, AAAI Workshop cochair, ACM Special Interest Group on Artificial Intelligence (SIGART) vice-chair, and ACM-SIGART/AAAI Doctoral Consortium chair.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.