

# Multilingual Sentiment and Subjectivity Analysis

Carmen Banea and Rada Mihalcea  
Department of Computer Science  
University of North Texas  
rada@cs.unt.edu, carmen.banea@gmail.com

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

May 12, 2011

## 1 Introduction

Subjectivity and sentiment analysis focuses on the automatic identification of private states, such as opinions, emotions, sentiments, evaluations, beliefs, and speculations in natural language. While subjectivity classification labels text as either subjective or objective, sentiment classification adds an additional level of granularity, by further classifying subjective text as either positive, negative or neutral.

To date, a large number of text processing applications have already used techniques for automatic sentiment and subjectivity analysis, including automatic expressive text-to-speech synthesis [1], tracking sentiment timelines in on-line forums and news [22, 2], and mining opinions from product reviews [11]. In many natural language processing tasks, subjectivity and sentiment classification have been used as a first phase filtering to generate more viable data. Research that benefitted from this additional layering ranges from question answering [48], to conversation summarization [7] and text semantic analysis [41, 8].

Much of the research work to date on sentiment and subjectivity analysis has been applied to English, but work on other languages is growing, including Japanese [19, 34, 35, 15], Chinese [12, 49], German [18], and Romanian [23, 4]. In addition, several participants in the Chinese and Japanese Opinion Extraction tasks of NTCIR-6 [17] performed subjectivity and sentiment analysis in languages other than English.<sup>1</sup>

As only 29.4% of Internet users speak English,<sup>2</sup> the construction of resources and tools for subjectivity and sentiment analysis in languages other than English is a growing need. In this chapter, we review the main directions of research focusing on the development of resources and tools for multilingual subjectivity and sentiment analysis. Specifically, we identify and overview three main categories of methods: (1) those focusing on word and phrase level annotations, overviewed in Section 4; (2) methods targeting the labeling of sentences, described in Section 5; and finally (3) methods for document-level annotations, presented in Section 6.

We address both multilingual and cross-lingual methods. For multilingual methods, we review work concerned with languages other than English, where the resources and tools have been specifically developed for a given target language. In this category, in Section 3 we also briefly overview the main directions of work on English data, highlighting the methods that can be easily

---

<sup>1</sup>NTCIR is a series of evaluation workshops sponsored by the Japan Society for the Promotion of Science, targeting tasks such as information retrieval, text summarization, information extraction, and others. NTCIR-6, 7 and 8 included an evaluation of multilingual opinion analysis on Chinese, English and Japanese.

<sup>2</sup>[www.internetworldstats.com/stats.htm](http://www.internetworldstats.com/stats.htm), June 30, 2008

ported to other languages. For cross-lingual approaches, we describe several methods that have been proposed to leverage on the resources and tools available in English by using cross-lingual projections.

## 2 Definitions

An important kind of information that is conveyed in many types of written and spoken discourse is the mental or emotional state of the writer or speaker or some other entity referenced in the discourse. News articles, for example, often report emotional responses to a story in addition to the facts. Editorials, reviews, weblogs, and political speeches convey the opinions, beliefs, or intentions of the writer or speaker. A student engaged in a tutoring session may express his or her understanding or uncertainty. Quirk et al. give us a general term, *private state*, for referring to these mental and emotional states [29]. In their words, a private state is a state that is not open to objective observation or verification: “a person may be observed to assert that God exists, but not to believe that God exists. Belief is in this sense ‘private’.” A term for the linguistic expression of private states, adapted from literary theory [5], is *subjectivity*. *Subjectivity analysis* is the task of identifying when a private state is being expressed and identifying attributes of the private state. Attributes of private states include who is expressing the private state, the type(s) of attitude being expressed, about whom or what the private state is being expressed, the polarity of the private state, (i.e., whether it is positive or negative), and so on. For example, consider the following sentence:

The choice of Miers was praised by the Senate’s top Democrat, Harry Reid of Nevada.

In this sentence, the phrase “was praised by” indicates that a private state is being expressed. The private state, according to the writer of the sentence, is being expressed by Reid, and it is about the choice of Miers, who was nominated to the Supreme Court by President Bush in October 2005. The type of the attitude is a sentiment (an evaluation, emotion, or judgment) and the polarity is positive [44].

This chapter is primarily concerned with detecting the presence of subjectivity, and further, identifying its polarity. These judgments may be made along several dimensions. One dimension is context. On the one hand, we may judge the subjectivity and polarity of words, out of context: *love* is subjective and positive, while *hate* is subjective and *negative*. At the other extreme, we have “full” contextual interpretation of language as it is being used in a text or dialog. In fact, there is a continuum from one to the other, and we can define several natural language processing tasks along this continuum.

The first is developing a word-level subjectivity lexicon, a list of keywords which have been gathered together because they have subjective usages; polarity information is often added to such lexicons. In addition to *love* and *hate*, other examples are *brilliant* and *interest* (positive polarity), and *alarm* (negative polarity).

We can also classify word *senses* according to their subjectivity and polarity. Consider, for example, the following two senses of *interest* from WordNet [24]:

- “Interest, involvement – (a sense of concern with and curiosity about someone or something; ‘an interest in music’)”
- “Interest – a fixed charge for borrowing money; usually a percentage of the amount borrowed; ‘how much interest do you pay on your mortgage?’ ”

The first sense is subjective, with positive polarity. But the second sense is not (non-subjective senses are called *objective* senses) – it does not refer to a private state. For another example, consider the senses of the noun *difference*:

- difference – (the quality of being unlike or dissimilar) ‘there are many differences between jazz and rock’
- deviation, divergence, departure, difference (a variation that deviates from the standard or norm) ‘the deviation from the mean’
- dispute, difference, difference of opinion, conflict (a disagreement or argument about something important) ‘he had a dispute with his wife’
- difference (a significant change) ‘his support made a real difference’
- remainder, difference (the number that remains after subtraction)

The first, second, and fifth of these definitions are objective. The others are subjective. Interestingly, the third sense has negative polarity (referring to conflict between people), while the fourth sense has positive polarity.

Word- and sense-level subjectivity lexicons are important because they are useful resources for *contextual subjectivity analysis* [45] – recognizing and extracting private state expressions in an actual text or dialog. We can judge the subjectivity and polarity of texts at several different levels. At the document level, we can ask if a text is opinionated and, if so, whether it is mainly positive or negative. We can perform more fine-grained analysis, and ask if a sentence contains any subjectivity. For instance, consider the following examples from [45]. The first sentence below is subjective (and has positive polarity), but the second one is objective, because it does not contain any subjective expressions:

- He spins a riveting plot which grabs and holds the reader’s interest.
- The notes do not pay interest.

Even further, individual expressions may be judged, for example that “spins”, “riveting” and “interest” in the first sentence above are subjective expressions. A more interesting example appears in this sentence “Cheers to Timothy Whitfield for the wonderfully horrid visuals”. While “horrid” would be listed as having negative polarity in a word-level subjectivity lexicon, in this context, it is being used positively: “wonderfully horrid” expresses a positive sentiment toward the visuals (similarly, “Cheers” expresses a positive sentiment toward “Timothy Whitfield”).

### 3 Sentiment and Subjectivity Analysis on English

Before we describe the work that has been carried out for multilingual sentiment and subjectivity analysis, we first briefly overview the main lines of research carried out on English, along with the most frequently used resources that have been developed for this language. Several of these English resources and tools have been used as a starting point to build resources in other languages, via cross-lingual projections or monolingual and multi-lingual bootstrapping. As described in more detail below, in cross-lingual projection, annotated data in a second language is created by projecting the annotations from a source (usually major) language across a parallel text. In multi-lingual bootstrapping, in addition to the annotations obtained via cross-lingual projections, mono-lingual corpora in the source and target languages are also used in conjunction with bootstrapping techniques such as co-training, which often lead to additional improvements.

### 3.1 Lexicons

One of the most frequently used lexicons is perhaps the subjectivity and sentiment lexicon provided with the OpinionFinder distribution [42]. The lexicon was compiled from manually developed resources augmented with entries learned from corpora. It contains 6,856 unique entries, out of which 990 are multi-word expressions. The entries in the lexicon have been labeled for part of speech as well as for reliability – those that appear most often in subjective contexts are *strong* clues of subjectivity, while those that appear less often, but still more often than expected by chance, are labeled *weak*. Each entry is also associated with a polarity label, indicating whether the corresponding word or phrase is positive, negative, or neutral. To illustrate, consider the following entry from the OpinionFinder lexicon *type=strongsubj word1=agree pos1=verb mpqapolarity=weakpos*, which indicates that the word *agree* when used as a *verb* is a strong clue of subjectivity and has a polarity that is weakly positive.

Another lexicon that has been often used in polarity analysis is the General Inquirer [32]. It is a dictionary of about 10,000 words grouped into about 180 categories, which have been widely used for content analysis. It includes semantic classes (e.g., animate, human), verb classes (e.g., negatives, becoming verbs), cognitive orientation classes (e.g., causal, knowing, perception), and other. Two of the largest categories in the General Inquirer are the valence classes, which form a lexicon of 1,915 positive words and 2,291 negative words.

SentiWordNet [9] is a resource for opinion mining built on top of WordNet, which assigns each synset in WordNet with a score triplet (positive, negative, and objective), indicating the strength of each of these three properties for the words in the synset. The SentiWordNet annotations were automatically generated, starting with a set of manually labeled synsets. Currently, SentiWordNet includes an automatic annotation for all the synsets in WordNet, totaling more than 100,000 words.

### 3.2 Corpora

Subjectivity and sentiment annotated corpora are useful not only as a means to train automatic classifiers, but also as resources to extract opinion mining lexicons. For instance, a large number of the entries in the OpinionFinder lexicon mentioned in the previous section were derived based on a large opinion-annotated corpus.

The MPQA corpus [43] was collected and annotated as part of a 2002 workshop on Multi-Perspective Question Answering (thus the MPQA acronym). It is a collection of 535 English-language news articles from a variety of news sources manually annotated for opinions and other private states (i.e., beliefs, emotions, sentiments, speculations, etc.). The corpus was originally annotated at clause and phrase level, but sentence-level annotations associated with the dataset can also be derived via simple heuristics [42].

Another manually annotated corpus is the collection of newspaper headlines created and used during the recent SEMEVAL task on “Affective Text” [33]. The data set consists of 1000 test headlines and 200 development headlines, each of them annotated with the six Eckman emotions (anger, disgust, fear, joy, sadness, surprise) and their polarity orientation (positive, negative).

Two other data sets, both of them covering the domain of movie reviews, are a polarity data set consisting of 1,000 positive and 1,000 negative reviews, and a subjectivity data set consisting of 5,000 subjective and 5,000 objective sentences. Both data sets have been introduced in [27], and have been used to train opinion mining classifiers. Given the domain-specificity of these collections, they were found to lead to accurate classifiers for data belonging to the same or similar domains.

### 3.3 Tools

There are a large number of approaches that have been developed to date for sentiment and subjectivity analysis in English. The methods can be roughly classified into two categories: (1) rule-based systems, relying on manually or semi-automatically constructed lexicons; and (2) machine learning classifiers, trained on opinion-annotated corpora.

Among the rule-based systems, one of the most frequently used is OpinionFinder [42], which automatically annotates the subjectivity of new text based on the presence (or absence) of words or phrases in a large lexicon. Briefly, the OpinionFinder high-precision classifier relies on three main heuristics to label subjective and objective sentences: (1) if two or more strong subjective expressions occur in the same sentence, the sentence is labeled *Subjective*; (2) if no strong subjective expressions occur in a sentence, and at most two weak subjective expressions occur in the previous, current, and next sentence combined, then the sentence is labeled *Objective*; (3) otherwise, if none of the previous rules apply, the sentence is labeled *Unknown*. The classifier uses the clues from a subjectivity lexicon and the rules mentioned above to harvest subjective and objective sentences from a large amount of unannotated text; this data is then used to automatically identify a set of extraction patterns, which are then used iteratively to identify a larger set of subjective and objective sentences.

In addition to the high-precision classifier, OpinionFinder also includes a high-coverage classifier. This high-precision classifier is used to automatically produce an English labeled data set, which can then be used to train a high-coverage subjectivity classifier.

When evaluated on the MPQA corpus, the high-precision classifier was found to lead to a precision of 86.7% and a recall of 32.6%, whereas the high-coverage classifier has a precision of 79.4% and a recall of 70.6%.

Another unsupervised system worth mentioning, this time based on automatically labeled words or phrases, is the one proposed in [36], which builds upon earlier work by [10]. Starting with two reference words, “excellent” and “poor,” Turney classifies the polarity of a word or phrase by measuring the fraction between its pointwise mutual information (PMI) with the positive reference (“excellent”) and the PMI with the negative reference (“poor”).<sup>3</sup> The polarity scores assigned in this way are used to automatically annotate the polarity of product, company, or movie reviews. Note that this system is completely unsupervised, and thus particularly appealing for application to other languages.

Finally, when annotated corpora is available, machine-learning methods are a natural choice for building subjectivity and sentiment classifiers. For example, Wiebe et al. [40] used a data set manually annotated for subjectivity to train a machine learning classifier, which led to significant improvements over the baseline. Similarly, starting with semi-automatically constructed data sets, Pang and Lee [27] built classifiers for subjectivity annotation at sentence level, as well as a classifier for sentiment annotation at document level. To the extent that annotated data is available, such machine-learning classifiers can be used equally well in other languages.

## 4 Word and Phrase-level Annotations

The development of resources and tools for sentiment and subjectivity analysis often starts with the construction of a lexicon, consisting of words and phrases annotated for sentiment or subjectivity. Such lexicons are successfully used to build rule-based classifiers for automatic opinion annotation,

---

<sup>3</sup>The PMI of two words  $w_1$  and  $w_2$  is defined as the probability of seeing the two words together divided by the probability of seeing each individual word:  $PMI(w_1, w_2) = \frac{p(w_1, w_2)}{p(w_1)p(w_2)}$

by primarily considering the presence (or absence) of the lexicon entries in a text.

There are three main directions that have been considered so far for word and phrase level annotations: (1) manual annotations, which involve human judgment of selected words and phrases; (2) automatic annotations based on knowledge sources such as dictionaries; and (3) automatic annotations based on information derived from corpora.

## 4.1 Dictionary-based

One of the simplest approaches that have been attempted for building opinion lexicons in a new language is the **translation of an existing source language lexicon by using a bilingual dictionary**. Mihalcea et. al [23] generate a subjectivity lexicon for Romanian by starting with the English subjectivity lexicon from OpinionFinder (described in Section 3.1) and translating it using an English-Romanian bilingual dictionary.

Several challenges were encountered in the translation process. First, although the English subjectivity lexicon contains inflected words, the lemmatized form is required in order to be able to translate the entries using the bilingual dictionary. However, words may lose their subjective meaning once lemmatized. For instance, the inflected form of *memories* becomes *memory*. Once translated into Romanian (as *memorie*), its main meaning is objective, referring to the ability of retaining information.

Second, neither the lexicon nor the bilingual dictionary provides information concerning the sense of the individual entries, and therefore the translation has to rely on the most probable sense in the target language. Fortunately, some bilingual dictionaries list the translations in reverse order of their usage frequencies, which is a heuristic that can be used to partly address this problem. 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.

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

Table 1: Examples of entries in the Romanian subjectivity lexicon

Using this direct translation process, Mihalcea et al. were able to obtain a subjectivity lexicon in Romanian containing 4,983 entries. Table 1 shows examples of entries in the Romanian lexicon, together with their corresponding original English form. The table also shows the reliability of the expression (*weak* or *strong*) and the part of speech – attributes that are provided in the English subjectivity lexicon.

To evaluate the quality of the lexicon, two native speakers of Romanian annotated the subjectivity of 150 randomly selected entries. Each annotator independently read approximately 100 examples of each drawn from the Web, including a large number from news sources. The subjectivity of a word is consequently judged in the contexts where it most frequently appears, accounting for its most frequent meanings on the Web. After the disagreements were reconciled through discussions, the final set of 123 correctly translated entries included 49.6% (61) subjective entries, but as many as 23.6% (29) entries were found to have primarily objective uses (the other 26.8% were mixed).

The study from [23] suggests that the Romanian subjectivity clues derived through translation are less reliable than the original set of English clues. In several cases, the subjectivity is lost in

the translation, mainly due to word ambiguity in either the source or target language, or both. For instance, the word *fragile* correctly translates into Romanian as *fragil*, yet this word is frequently used to refer to breakable objects, and it loses its subjective meaning of *delicate*. Other words, such as *one-sided*, completely loses subjectivity once translated, as it becomes in Romanian *cu o singura latură*, meaning *with only one side* (as of objects).

Using a similar translation technique, Kim and Hovy [18] build a **lexicon for German starting with a lexicon in English, this time focusing on polarity** rather than subjectivity. They use an English polarity lexicon semi-automatically generated starting with a few seeds and using the WordNet structure [24]. Briefly, for a given seed word, its synsets and synonyms are extracted from WordNet, and then the probability of the word belonging to one of the three classes is calculated based on the number and frequency of seeds from a particular class appearing within the word’s expansion. This metric thus represents the closeness of a word to the seeds. Using this method, Kim and Hovy are able to generate an English lexicon of about 1,600 verbs and 3,600 adjectives, classified as positive or negative based on their polarity.

The lexicon is then translated into German, by using an automatically generated translation dictionary obtained from the European Parliament corpus via word alignment [26]. To evaluate the quality of the German polarity lexicon, the entries in the lexicon were used in a rule-based system that was applied to the annotation of polarity for 70 German emails. Overall, the system obtained an F-measure of 60% for the annotation of positive polarity, and 50% for the annotation of negative polarity.

Another method for **building subjectivity lexicons** is proposed by Banea et al. [3], by **bootstrapping from a few manually selected seeds**. At each iteration, the seed set is expanded with related words found in an online dictionary, which are filtered by using a measure of word similarity. The bootstrapping process is illustrated in Figure 1.

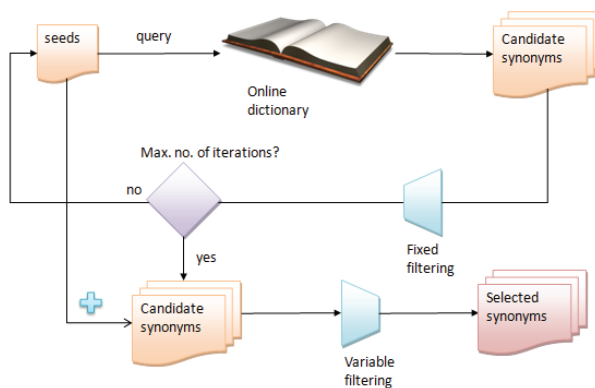


Figure 1: Bootstrapping process

Starting with a seed set of subjective words, evenhandedly sampled from verbs, nouns, adjectives and adverbs, new related words are added based on the entries found in the dictionary. For each seed word, all the open-class words appearing in its definition are collected, as well as synonyms and antonyms if available. Note that word ambiguity is not an issue, as the expansion is done with all the possible meanings for each candidate word. The candidates are subsequently filtered for incorrect meanings by using a measure of similarity with the seed words, calculated using a latent semantic analysis system trained on a corpus in the target language.

In experiments carried out on Romanian, starting with 60 seed words, Banea et al. are able to build a subjective lexicon of 3,900 entries. The quality of the lexicon was evaluated by embedding it into a rule-based classifier used for the classification of subjectivity for 504 manually annotated

sentences. The classifier led to an F-measure of 61.7%, which is significantly higher than a simple baseline of 54% that can be obtained assigning a majority class by default.

A similar bootstrapping technique was used by Pitel and Grefenstette [28], for the **construction of affective lexicons for French**. They classify words into 44 affect classes (e.g., *morality, love, crime, insecurity*), each class being in turn associated with a positive or negative orientation. Starting with a few seed words (two to four seed words for each affective dimension), they use synonym expansion to automatically add new candidate words to each affective class. The new candidates are then filtered based on a measure of similarity calculated with latent semantic analysis, and machine learning trained on seed data. Using this method, Pitel and Grefenstette are able to generate a French affective lexicon of 3,500 words, which is evaluated against a gold standard data set consisting of manually annotated entries. As more training samples are available in the training lexicon, the F-measure classification increases from 12% to 17%, up to a maximum of 27% F-measure for a given class.

## 4.2 Corpus-based

In addition to dictionaries, textual corpora were also found useful to derive subjectivity and polarity information associated with words and phrases. Much of the corpus-based research carried out to date follows the work of Turney [36] (see Section 3.3), who presented a method to measure the polarity of a word based on its PMI association with a positive or a negative seed (e.g., *excellent* and *poor*).

In [14], Kaji and Kitsuregawa propose a method to **build sentiment lexicons for Japanese**, by measuring the strength of association with positive and negative data automatically collected from Web pages. First, using structural information from the layout of HTML pages (e.g., list markers or tables that explicitly indicate the presence of the evaluation sections of a review, such as “pros”/“cons”, “minus”/“plus”, etc.), as well as Japanese-specific language structure (e.g., particles used as topic markers), a corpus of positive and negative statements is automatically mined from the Web. Starting with one billion HTML documents, about 500,000 polar sentences are collected, with 220,000 being positive and the rest negative. Manual verification of 500 sentences, carried out by two human judges, indicated an average precision of 92%, which shows that reasonable quality can be achieved using this corpus construction method.

Next, Kaji and Kitsuregawa use this corpus to automatically acquire a set of polar phrases. Starting with all the adjectives and adjectival phrases as candidates, they measure the chi-squared and the PMI between these candidates and the positive and negative data, followed by a selection of those words and phrases that exceed a certain threshold. Through experiments, the PMI measure was found to work better as compared to chi-squared. The polarity value of a word or phrase based on PMI is defined as:

$$PV_{PMI}(W) = PMI(W, pos) - PMI(W, neg)$$

where

$$PMI(W, pos) = \log_2 \frac{P(W, pos)}{P(W)P(pos)} \quad PMI(W, neg) = \log_2 \frac{P(W, neg)}{P(W)P(neg)}$$

*pos* and *neg* representing the positive and negative sentences automatically collected from the Web.

Using a data set of 405 adjective phrases, consisting of 158 positive phrase, 150 negative, and 97 neutral, Kaji and Kitsuregawa are able to build a lexicon ranging from 8,166 to 9,670 entries, depending on the value of the threshold used for the candidate selection. The precision for the positive phrases was 76.4% (recall 92.4%) when a threshold of 0 is used, and went up to 92.0%



(recall 65.8%) when the threshold is raised to 3.0. For the same threshold values, the negative phrases had a precision ranging from 68.5% (recall 84.0%) to 87.9% (recall 62.7%).

Another corpus-based method for the **construction of polarity lexicons in Japanese, this time focusing on domain-specific propositions**, is proposed in [15]. Kanayama and Nasukawa introduce a novel method for performing domain-dependent unsupervised sentiment analysis through the automatic acquisition of polar atoms in a given domain by building upon a domain-independent lexicon. In their work, a polar atom is defined as “the minimum human-understandable syntactic structures that specify the polarity of clauses,” and it typically represents a tuple of polarity and a verb or an adjective along with its optional arguments. The system uses both intra- and inter-sentential coherence as a way to identify polarity shifts, and automatically bootstraps a domain-specific polarity lexicon.

First, candidate propositions are identified by using the output of a full parser. Next, sentiment assignment is performed in two stages. Starting from a lexicon of pre-existing polar atoms based on an English sentiment lexicon, the method finds occurrences of the entries in the propositions extracted earlier. These propositions are classified as either positive or negative based on the label of the atom they contain, or its opposite in case a negation is encountered. The next step involves the extension of the initial sentiment labeling to those propositions that are not labeled. To this end, context coherency is used, which assumes that in a given context the polarity will not shift unless an adversative conjunction is encountered, either between sentences and/or within sentences. Finally, the confidence of each new polar atom is calculated, based on its total number of occurrences in positive and negative contexts.

The method was evaluated on Japanese product reviews extracted from four domains: digital cameras, movies, mobile phones and cars. The number of reviews in each corpus ranged from 155,130 (mobile phones) to 263,934 (digital cameras). Starting with these data sets, the method is able to extract 200-700 polar atoms per domain, with a precision evaluated by human judges ranging from 54% for the mobile phones corpus to 75% for the movies corpus.

Kanayama and Nasukawa’s method is similar to some extent to an approach proposed earlier by Kobayashi et al., which **extracts opinion triplets from Japanese product reviews mined from the Web** [19]. An opinion triplet consists of the following fields: product, attribute and value. The process involves a bootstrapping process consisting of two steps. The first step consists of the generation of candidates based on a set of co-occurrence patterns, which are applied to a collection of Web reviews. Three dictionaries that are updated at the end of each bootstrapping iteration are also provided (dictionaries of subjects, attributes, and values). Once a ranked list of candidates is generated, a human judge is presented with the top ranked candidates for annotation. The manual step involves identifying the attributes and their values and updating their corresponding dictionaries with the newly extracted entities.

For the experiments, Kobayashi et al. use two data sets, consisting of 15,000 car reviews and 10,000 game reviews respectively. The bootstrapping process starts with a subject dictionary of 389 car names and 660 computer games names, an initial attribute list with seven generic descriptors (e.g., cost, price, performance), and a value list with 247 entries (e.g., good, beautiful, high). Each extraction pattern is scored based on the frequency of the extracted expressions and their reliability. For the evaluation, a human annotator tagged 105 car reviews and 280 computer game reviews, and identified the attributes and their corresponding values. Overall, using the semi-automatic system, Kobayashi et al. found that lexicons of opinion triplets can be built eight times faster as compared to a fully manual set-up. Moreover, the semi-automatic system is able to achieve a coverage of 35-45% with respect to the manually extracted expressions, which represents a significant coverage.

The **semantic orientation of phrases in Japanese** is also the goal of the work of [35] and [34], both using an expectation maximization model trained on annotated data. Takamura et al.

consider the task of finding the polarity of phrases such as “light laptop,” which cannot be directly obtained from the polarity of individual words (since, in this case, both “light” and “laptop” are neutral). On a data set of 12,000 adjective-noun phrases drawn from a Japanese newspaper, they found that a model based on triangle and “U-shaped” graphical dependencies leads to an accuracy of approximately 81%.

Suzuki et al. target instead evaluative expressions, similar to those addressed by [19]. They use an expectation maximization algorithm and a Naïve Bayes classifier to bootstrap a system to annotate the polarity of evaluative expressions consisting of subjects, attributes and values. Using a data set of 1,061 labeled examples and 34,704 unlabeled examples, they obtain an accuracy of 77%, which represents a significant improvement over the baseline of 47% obtained by assigning the majority class from the set of 1,061 labeled examples.

Finally, another line of work concerned with the polarity analysis of words and phrases is presented in [6]. Instead of targeting the derivation of subjectivity or sentiment lexicon in a new language, the goal of Bautin et al.’s work is to measure the polarity of given entities (e.g., George Bush, Vladimir Putin) in a text written in a target language. Their approach relies on the translation of documents (e.g., newswire, European parliament documents) from the given language into English, followed by a calculation of the polarity of the target entity by using association measures between the occurrence of the entity and positive/negative words from a sentiment lexicon in English.

The experiments presented in [6] focus on nine different languages (Arabic, Chinese, English, French, German, Italian, Japanese, Korean, Spanish), and fourteen entities covering country and city names. They show that large variations can be achieved in the measures of polarity or subjectivity of an entity across languages, ranging from very weak correlations (close to 0), to strong correlations (0.60 and higher). For instance, an aggregation of all the polarity scores measured for all fourteen entities in different languages leads to a low correlation of 0.08 between mentions of such entities in Japanese and Chinese text, but as high as 0.63 when the mentions are collected from French and Korean texts.

## 5 Sentence-level Annotations

Corpus annotations are often required either as an end goal for various text processing applications (e.g., mining opinions from the Web; classification of reviews into positive and negative; etc.), or as an intermediate step toward building automatic subjectivity and sentiment classifiers. Work in this area has considered annotations at either sentence or document level, depending mainly on the requirements of the end application (or classifier). The annotation process is typically done following one of two methods: (1) dictionary-based, consisting of rule-based classifiers relying on lexicons built with one of the methods described in the previous section; or (2) corpus-based, consisting of machine learning classifiers trained on pre-existing annotated data.

### 5.1 Dictionary-based

Rule-based classifiers, such as the one introduced Riloff and Wiebe in [30], can be used in conjunction with any opinion lexicon to develop a sentence-based classifier. These classifiers mainly look for the presence (or absence) of lexicon clues in the text, and correspondingly decide on the classification of a sentence as subjective/objective or positive/negative.

One of the lexicons described in the previous section that has been evaluated in a rule-based classifier is the Romanian subjectivity lexicon built by translating an English lexicon [23] (see Section 4.1). The classifier relied on three main heuristics to label subjective and objective sentences:

(1) if two or more strong subjective expressions occur in the same sentence, the sentence is labeled *subjective*; (2) if no strong subjective expressions occur in a sentence, and at most three weak subjective expressions occur in the previous, current, and next sentence combined, then the sentence is labeled *objective*; (3) otherwise, if none of the previous rules applied, the sentence is labeled *unknown*.

The quality of the classifier was evaluated on a Romanian gold-standard corpus annotated for subjectivity, consisting of 504 sentences from the Romanian side of an English-Romanian parallel corpus, annotated according to the annotation scheme in [43]. The classifier had an overall precision of 62% and a recall of 39%; the precision for the subjective annotations only was evaluated at 80%, for a recall of 21%.

Another subjectivity lexicon that was evaluated in a rule-based approach is the one from [3] (Section 4.1). Using a lexicon of 3,900 entries in Romanian, obtained after several bootstrapping iterations, Banea et al. build a rule-based classifier with an overall precision and recall of 62%, when evaluated on the same data set of 504 manually annotated Romanian sentences. This is significantly higher than the results obtained based on the translated lexicons, indicating the importance of language-specific information for subjectivity analysis.

Besides Romanian, a lexicon approach is also used for the classification of polarity for sentences in Japanese [16]. Kanayama et al. use a machine translation system based on deep parsing to extract “sentiment units” with high precision from Japanese product reviews, where a sentiment unit is defined as a tuple between a sentiment label (positive or negative) and a predicate (verb or adjective) with its argument (noun). The sentiment analysis system uses the structure of a transfer-based machine translation engine, where the production rules and the bilingual dictionary are replaced by sentiment patterns and a sentiment lexicon, respectively.

The system is ultimately able to not only mine product reviews for positive/negative product attributes, but also to provide a user friendly interface to browse product reviews. The sentiment units derived for Japanese are used to classify the polarity of a sentence, using the information drawn from a full syntactic parser in the target language. Using about 4,000 sentiment units, when evaluated on 200 sentences, the sentiment annotation system was found to have high precision (89%) at the cost of low recall (44%).

## 5.2 Corpus-based

Once a corpus annotated at sentence level is available, with either subjectivity or polarity labels, a classifier can be easily trained to automatically annotate additional sentences.

This is the approach taken by Kaji and Kitsuregawa [13, 14], who **collect a large corpus of sentiment-annotated sentences from the Web, and subsequently use this data set to train sentence-level classifiers**. Using the method described in Section 4.2, which relies on structural information from the layout of HTML pages, as well as Japanese-specific language structure, Kaji and Kitsuregawa collect a corpus of approximately 500,000 positive and negative sentences from the Web. The quality of the annotations was estimated by two human judges, who found an average precision of 92% as measured on a randomly selected sample of 500 sentences.

A subset of this corpus, consisting of 126,000 sentences, is used to build a Naïve Bayes classifier. Using three domain specific data sets (computers, restaurants and cars), automatically collected by selecting manually annotated reviews consisting of only one sentence, the precision of the classifier was found to have an accuracy ranging between 83% (computers) and 85% (restaurants), which is comparable to the accuracy obtained by training on in-domain data. These results demonstrate the quality of the automatically built corpus, and the fact that it can be used to train reliable sentence-level classifiers with good portability to new domains.

Another corpus-based approach is explored by Mihalcea et al. [23], where a **Romanian corpus annotated for subjectivity at sentence level is built via cross-lingual projections across parallel texts**. Mihalcea et al. use a parallel corpus consisting of 107 documents from the English SemCor corpus [25] and their manual translation into Romanian. The corpus consists of roughly 11,000 sentences, with approximately 250,000 tokens on each side. It is a balanced corpus covering a number of topics in sports, politics, fashion, education, and others.

To annotate the English side of the parallel corpus, the two OpinionFinder classifiers (described in Section 3.3) are used to label the sentences in the corpus. Next, the OpinionFinder annotations are projected onto the Romanian training sentences, which are then used to develop a Naïve Bayes classifier for the automatic labeling of subjectivity in Romanian sentences. The quality of the classifiers was evaluated on a corpus of 504 sentences manually annotated for subjectivity (the same gold-standard corpus used in the experiments described in the previous sections). When the high-precision classifier is used to produce the annotations for the English corpus, the overall accuracy was measured at 64%. When the high-coverage classifier is used, the accuracy raised to 68%. In both cases, the accuracy was found to be significantly higher than the majority-class baseline of 54%, indicating that cross-lingual projections represent a reliable technique for building subjectivity annotated corpora in a new language.

Following the same idea of using cross-lingual projections across parallel texts, Banea et al. [4] propose a method based on machine translation to generate the required parallel texts. The **English sentence-level subjectivity annotations are projected across automatically translated texts, in order to build subjectivity classifiers for Romanian and Spanish**. Using first Romanian as a target language, several translation scenarios are considered, with various results as measured on the same gold-standard data set of 504 sentences described before. First, a classifier is trained on annotations projected across the automatic translation of an English manually annotated corpus (MPQA; see Section 3.2); this resulted in an accuracy of 66% using an SVM classifier [37]. Second, an English corpus is automatically annotated with the high-coverage OpinionFinder classifier, and the annotations were projected across machine translated text. Again, an SVM classifier is trained on the resulting annotations in the new language, this time resulting in an accuracy of 69%. Finally, a Romanian corpus is automatically translated into English, followed by an annotation of the English version using the OpinionFinder classifier, and a projection of the subjectivity labels back into Romanian. The SVM classifier trained on this data had an accuracy of 67%.

The same experiments were replicated on Spanish, which led to 68% accuracy when the source language text had manual subjectivity annotations, and 63% when the annotations were automatically generated with OpinionFinder. Overall, the results obtained with machine translated text were found to be just a few percentages below the results obtained with manually translated text, which shows that machine translation can be effectively used to generate the required parallel texts for cross-lingual projections.

## 6 Document-level Annotations

Natural language applications, such as review classification or Web opinion mining, often require corpus-level annotations of subjectivity and polarity. In addition to sentence-level annotations, described in the previous section, there are several methods that have been proposed for the annotation of entire documents. As before, the two main directions of work have considered: (1) dictionary-based annotations, which assume the availability of a lexicon, and (2) corpus-based annotations, which mainly rely on classifiers trained on labeled data.

## 6.1 Dictionary-based

Perhaps the simplest approach for document annotations is to use a **rule-based system based on the clues available in a language-specific lexicon**. One of the methods proposed by Wan [38] consists of **annotating Chinese reviews by using a polarity lexicon**, along with a set of negation words and intensifiers. The lexicon contains 3,700 positive terms, 3,100 negative words, and 148 intensifier terms, all of them collected from a Chinese vocabulary for sentiment analysis released by HowNet, as well as 13 negation terms collected from related research. Given this lexicon, the polarity of a document is annotated by combining the polarity of its constituent sentences, where in turn the polarity of a sentence is determined as a summation of the polarity of the words found in the sentence. When evaluated on a data set of 886 Chinese reviews, this method was found to give an overall accuracy of 74.3%.

The other method proposed by Wan [38] is to **use machine translation to translate the Chinese reviews into English, followed by the automatic annotation of the English reviews** using a rule-based system relying on English lexicons. Several experiments are run with two commercial machine translation systems, using the OpinionFinder polarity lexicon (see Section 3.1). For the same test set mentioned before, the translation method achieves an accuracy of up to 81%, significantly higher than the one achieved by directly analyzing the reviews using a Chinese lexicon. Moreover, an ensemble combining different translations and methods leads to an even higher accuracy of 85%, demonstrating that a combination of different knowledge sources can exceed the performance obtained with individual resources.

Another approach, proposed by Zagibalov and Carroll [49], consists of a **bootstrapping method to label the polarity of Chinese text** by iteratively building a lexicon and labeling new text. The method starts by identifying “lexical items” in text, which are sequences of Chinese characters that occur between non-character symbols and which include a negation and an adverbial; a small hand-picked list of six negations and five adverbials is used, which increase the portability of the method to other languages. In order to be considered for candidacy in the seed list, the lexical item should appear at least twice in the data that is being considered.

Next, “zones” are identified in the text, where a zone is defined as the sequence of characters occurring between punctuation marks. The sentiment associated with an entire document is calculated as the difference between the number of positive and negative zones that the review entails. In turn, the sentiment of a zone is computed by summing the polarity scores of their component lexical items. Finally, the polarity of a lexical item is proportional with the square of its length (number of characters), and with its previous polarity score, while being inversely proportional to the length of the containing zone. This score is multiplied by -1 in case a negation precedes the lexical item.

The bootstrapping process consists of iterative steps that result in an incrementally larger set of seeds, and an incrementally larger number of annotated documents. Starting with a seed set consisting initially of only one adjective (“good”), new documents are annotated as positive and negative, followed by the identification of new lexical items occurring in these documents that can be added to the seed set. The addition to the seed set is determined based on the frequency of the lexical item, which has to be at least three times larger in the positive (negative) documents for it to be considered. The bootstrapping stops when over two runs no new seeds are found.

The method was evaluated over a balanced corpus of Chinese reviews compiled from ten different domains. The average accuracy at document level was measured at 83%. Moreover, the system was also able to extract a set of 50-60 seeds per domain, which may be helpful for other sentiment annotation algorithms.

Another method, used by Kim and Hovy [18], consists of the **annotation of German doc-**

**uments using a lexicon translated from English.** A lexicon construction method, described in detail in Section 4.1, is used to generate an English lexicon of about 5,000 entries. The lexicon is then translated into German, by using an automatically generated translation dictionary obtained from the European Parliament corpus using word alignment. The German lexicon is used in a rule-based system that is applied to the annotation of polarity for 70 German emails. Briefly, the polarity of a document is decided based on heuristics: a number of negative words above a particular threshold renders the document negative, whereas a majority of positive words triggers a positive classification. Overall, the system obtained an F-measure of 60% for the annotation of positive polarity, and 50% for the annotation of negative polarity.

## 6.2 Corpus-based

The most straight-forward approach for corpus-based document annotation is to **train a machine learning classifier, assuming that a set of annotated data already exists.** Li and Sun [21] use a data set of Chinese hotel reviews, on which they apply several classifiers, including SVM, Naïve Bayes and maximum entropy. Using a training set consisting of 6,000 positive reviews and 6,000 negative reviews and a test set of 2,000 positive reviews and 2,000 negative reviews, they obtain an accuracy of up to 92%, depending on the classifier and on the features used. These experiments demonstrate that if enough training data are available, it is relatively easy to build accurate sentiment classifiers.

A related, yet more sophisticated technique is proposed in [39], where a co-training approach is used to leverage resources from both a source and a target language. The technique is tested on the automatic sentiment classification of product reviews in Chinese. For a given product review in the target language (Chinese), an alternative view is obtained another language (English) via machine translation. The algorithm then uses two SVM classifiers, one in Chinese and one in English, to start a co-training process that iteratively builds a sentiment classifier. Initially, the training data set consists of a set of labeled examples in Chinese and their English translations. Next, the first iteration of co-training is performed, and a set of unlabeled instances is classified and added to the training set if the labels assigned in the models built on the languages agree. The newly labeled instances are used to re-train the two classifiers at the next iteration. Reviews with conflicting labels are not considered. As expected, the performance initially grows with the number iterations, followed by a degradation when the number of erroneously labeled instances exceeds a certain threshold. The best results are reported at the 40th iteration, for an overall F-measure of 81%, after adding five negative and five positive reviews at each iteration. The method is successful because it makes use of both cross-language and within-language knowledge.

## 7 What Works, What Doesn't

When faced with a new language, what is the best method that one can use to create a sentiment or 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<sup>4</sup> such as English, e.g., bilingual dictionaries or parallel texts.

---

<sup>4</sup>I.e., a language for which many resources and tools are already available

## 7.1 Best Scenario: Manually Annotated Corpora

The best scenario is when a corpus manually annotated for sentiment or 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 [43].

The alternative is to derive such corpora from online data, such as collections of product or movie reviews. This is a viable solution for languages where a significant number of online reviews exist. For instance, there are a number of methods that rely on online review repositories, including movie or product reviews in English [27, 11], product reviews in Japanese [15, 19], or hotel reviews in Chinese [21].

Once a large annotated data set is available, either manually annotated or mined from publicly available Web sites, 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, decision trees, or SVM,<sup>5</sup> can be used to annotate the subjectivity or the sentiment of new text.

## 7.2 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. This method was first introduced in [23] for the projection of subjectivity labels across parallel Romanian-English texts, and later used in conjunction with machine translation for the projection of subjectivity labels onto Romanian [4] or sentiment annotations onto Chinese [38].

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 or sentiment (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 or sentiment analysis tools such as OpinionFinder [42]. 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 by using a machine translation system.

Regardless of the direction of the translation, and regardless of the use of manually created parallel corpora or machine translated text, the result is a data set in the target language annotated for subjectivity or sentiment, which can be used to train an automatic classifier as described in the previous section.

## 7.3 Third Best: Bootstrapping a Lexicon

There are a number of methods that rely on the availability of subjectivity and sentiment 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 [42], which is based on a large subjectivity lexicon [46]. Similarly, the method suggested in [36] relies on a dictionary of sentiment words to automatically annotate the polarity of text.

---

<sup>5</sup>Usually available in off-the-shelf packages such as Weka [47].

One of the most successful approaches for the construction of subjectivity or sentiment lexicons is to bootstrap from a few manually selected seeds. If language processing tools such as part-of-speech taggers and parsers are available for the language at hand, the bootstrapping process can rely on information extraction patterns [31]. The alternative is to perform the bootstrapping using the synonyms and definitions found in an electronic dictionary [3]. In this case, no advanced language processing tools are required, 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 closely related to the seeds are kept for the next bootstrapping iteration, with the relatedness being measured with a similarity metric such as latent semantic analysis [20]. The bootstrapping process is illustrated in Figure 1. Running the process for several iterations can result in large lexicons with several thousands entries.

The typical use of a subjectivity or sentiment lexicon is within rule-based classifiers for sentence-level or document-level annotations. For instance, the classifier used in [3] labels a sentence as subjective if it contains three or more entries that appear in the subjective lexicon, and as objective if it has two or less entries. Document-level annotations can also be performed, by determining the frequency of subjective or sentiment words in the document.

#### 7.4 Fourth Best: Translating a Lexicon

If none of the previous methods is applicable to the target language, the last resort is to construct a lexicon by automatically translating an already existing lexicon from a major language. The only requirements for this approach are a sentiment or subjectivity lexicon in a source language, and a bilingual dictionary used to automatically translate the lexicon into the target language. The method was first used for building a sentiment lexicon in German [18], and later applied to the construction of a subjectivity lexicon in Romanian [23].

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 that are typical to a context-free translation process: difficulty in selecting the most appropriate translation for words that are ambiguous; small coverage for phrase translations; mismatch between the inflected forms appearing in the lexicon and the lemmatized forms from the bilingual dictionary. Even so, a lexicon constructed this way can be easily corrected by hand, and may therefore provide a building block for the generation of subjectivity or sentiment resources in a given target language.

#### 7.5 Comparing the Alternatives

It is difficult to make a full comparison among the four methods presented above, since there is no single language on which all four alternatives have been applied. There are however several experiments that were run for the construction of a subjectivity annotation tool for Romanian, which were all evaluated on a common data set thus allowing for a partial comparison.

	Precision	Recall	F-measure	Accuracy
parallel texts	<b>69.35</b>	78.75	73.76	<b>69.64</b>
machine translation, source-target	67.76	<b>83.15</b>	<b>74.57</b>	69.44
machine translation, target-source	76.06	59.34	66.67	67.86
lexicon bootstrapping	68.98	61.90	65.25	64.29
lexicon translation	65.84	38.83	48.85	55.95

Table 2: Comparison of several subjectivity annotation methods for Romanian.



Table 2 shows the results of these experiments, which were run on a data set of 504 sentences manually annotated for subjectivity (see Section 4.1). The only method that is not included in this comparison is the one relying on manually annotated corpora, since no manually annotated data set is available for Romanian. Not surprisingly, the highest accuracy is obtained when the corpus annotations are projected across manually constructed parallel texts, followed closely by the methods that rely on source-target or target-source machine translation. When no annotated data is available, a rule-based classifier can be built based on subjectivity lexicons. In this case, the best results are obtained using a bootstrapping method, followed by the projection of the lexicon based on a bilingual dictionary.

## 8 Conclusions

Sentiment and subjectivity analysis is a rapidly growing area. Although most of the work to date has focused on English, work that is targeting other languages is growing, with resources or tools for subjectivity or sentiment analysis currently available in Chinese, German, Japanese, Romanian, Spanish, and others.

This chapter described some of the most recent methods for multilingual sentiment and subjectivity analysis. The methods differ in the approach they take: supervised corpus-based algorithms or unsupervised rule-based approaches, as well as in the span of text they target: single words or phrases, entire sentences, or entire documents. While it is difficult to draw a comparison encompassing all the methods that have been proposed to date, the chapter also attempted to provide an overview of the most commonly used techniques and showed how they compare to each other when applied on the specific task of subjectivity analysis in Romanian.

Although still far from the performance that can be obtained on rich-resource languages such as English, the growing body of work on multilingual methods promises to lead to resources and tools in a continuously increasing number of languages.

## Acknowledgments

This material is based in part upon work supported by National Science Foundation awards #0917170 and #0916046. Any opinions, findings, and 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. In *Proceedings of the Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, pages 347–354, Vancouver, Canada, 2005.
- [2] K. Balog, G. Mishne, and M. de Rijke. Why are they excited? identifying and explaining spikes in blog mood levels. In *Proceedings of the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, 2006.
- [3] C. Banea, R. Mihalcea, and J. Wiebe. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In *Proceedings of the Learning Resources Evaluation Conference (LREC 2008)*, Marrakech, Morocco, 2008.
- [4] C. Banea, R. Mihalcea, J. Wiebe, and S. Hassan. Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*, Honolulu, Hawaii, 2008.

- [5] A. Banfield. *Unspeakable Sentences*. Routledge and Kegan Paul, Boston, 1982.
- [6] M. Bautin, L. Vijayarenu, and S. Skiena. International sentiment analysis for news and blogs. In *Proceedings of the International Conference on Weblogs and Social Media*, Seattle, WA, 2008.
- [7] G. Carenini, R. Ng, and X. Zhou. Summarizing emails with conversational cohesion and subjectivity. In *Proceedings of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2008)*, Columbus, Ohio, 2008.
- [8] A. Esuli and F. Sebastiani. Determining term subjectivity and term orientation for opinion mining. In *Proceedings the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pages 193–200, Trento, IT, 2006.
- [9] A. Esuli and F. Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, Genova, IT, 2006.
- [10] V. Hatzivassiloglou and K. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, pages 174–181, 1997.
- [11] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD 2004)*, pages 168–177, Seattle, Washington, 2004.
- [12] Y. Hu, J. Duan, X. Chen, B. Pei, and R. Lu. A new method for sentiment classification in text retrieval. In *IJCNLP*, pages 1–9, 2005.
- [13] N. Kaji and M. Kitsuregawa. Automatic construction of polarity-tagged corpus from html documents. In *Proceedings of the International Conference on Computational Linguistics / Association for Computational Linguistics*, Sydney, Australia, 2006.
- [14] N. Kaji and M. Kitsuregawa. Building lexicon for sentiment analysis from massive collection of html documents. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Prague, Czech Republic, 2007.
- [15] H. Kanayama and T. Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Sydney, Australia, 2006.
- [16] H. Kanayama, T. Nasukawa, and H. Watanabe. Deeper sentiment analysis using machine translation technology. In *International Conference on Computational Linguistics*, 2004.
- [17] N. Kando, T. Mitamura, and T. Sakai. Introduction to the ntcir-6 special issue. *ACM Transactions on Asian Language Information Processing (TALIP)*, 7(2), 2008.
- [18] S.-M. Kim and E. Hovy. Identifying and analyzing judgment opinions. In *Proceedings of the Human Language Technology Conference - North American chapter of the Association for Computational Linguistics*, New York City, NY, 2006.
- [19] N. Kobayashi, K. Inui, K. Tateishi, and T. Fukushima. Collecting evaluative expressions for opinion extraction. In *Proceedings of IJCNLP 2004*, pages 596–605, 2004.
- [20] T. K. Landauer, P. Foltz, and D. Laham. Introduction to latent semantic analysis. *Discourse Processes*, 25, 1998.
- [21] J. Li, , and M. Sun. Experimental study on sentiment classification of chinese review using machine learning techniques. In *International Conference on Natural Language Processing and Knowledge Engineering*, 2007.
- [22] L. Lloyd, D. Kechagias, and S. Skiena. Lydia: A system for large-scale news analysis. In *String Processing and Information Retrieval (SPIRE 2005)*, 2005.
- [23] R. Mihalcea, C. Banea, and J. Wiebe. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of the Association for Computational Linguistics*, Prague, Czech Republic, 2007.
- [24] G. Miller. Wordnet: A lexical database. *Communication of the ACM*, 38(11), 1995.
- [25] G. Miller, C. Leacock, T. Randee, and R. Bunker. A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, Plainsboro, New Jersey, 1993.
- [26] F. Och and H. Ney. Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hongkong, October 2000.

- [27] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics*, Barcelona, Spain, July 2004.
- [28] G. Pitel and G. Grefenstette. Semi-automatic building method for a multidimensional affect dictionary for a new language. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, 2008.
- [29] R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. *A Comprehensive Grammar of the English Language*. Longman, New York, 1985.
- [30] E. Riloff and J. Wiebe. Learning extraction patterns for subjective expressions. In *Conference on Empirical Methods in Natural Language Processing (EMNLP-03)*, pages 105–112, 2003.
- [31] E. Riloff, J. Wiebe, and T. Wilson. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the Seventh Conference on Natural Language Learning (CoNLL-2003)*, 2003.
- [32] P. Stone. *General Inquirer: Computer Approach to Content Analysis*. MIT Press, 1968.
- [33] C. Strapparava and R. Mihalcea. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on the Semantic Evaluations (SemEval 2007)*, Prague, Czech Republic, 2007.
- [34] Y. Suzuki, H. Takamura, and M. Okumura. Application of semi-supervised learning to evaluative expression classification. In *Proceedings of the 7th International Conference on Intelligent Text Processing and Computational Linguistics*, 2006.
- [35] H. Takamura, T. Inui, and M. Okumura. Latent variable models for semantic orientations of phrases. In *Proceedings of the 11th Meeting of the European Chapter of the Association for Computational Linguistics*, 2006.
- [36] P. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, pages 417–424, Philadelphia, 2002.
- [37] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
- [38] X. Wan. Using bilingual knowledge and ensemble techniques for unsupervised chinese sentiment analysis. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 2008.
- [39] X. Wan. Co-training for cross-lingual sentiment classification. In *Proceedings of the Joint Conference of the Association of Computational Linguistics and the International Joint Conference on Natural Language Processing*, Singapore, August 2009.
- [40] J. Wiebe, R. Bruce, and T. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 246–253, 1999.
- [41] J. Wiebe and R. Mihalcea. Word sense and subjectivity. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia, 2006.
- [42] J. Wiebe and E. Riloff. Creating subjective and objective sentence classifiers from unannotated texts. In *Proceedings of the 6th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2005) (invited paper)*, Mexico City, Mexico, 2005.
- [43] J. Wiebe, T. Wilson, and C. Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210, 2005.
- [44] T. Wilson. *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of private states*. PhD thesis, Intelligent Systems Program, University of Pittsburgh, 2007.
- [45] T. Wilson. *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*. PhD thesis, University of Pittsburgh, 2008.
- [46] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, Canada, 2005.
- [47] I. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.
- [48] H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2003)*, pages 129–136, Sapporo, Japan, 2003.
- [49] T. Zagibalov and J. Carroll. Automatic seed word selection for unsupervised sentiment classification of chinese text. In *Proceedings of the Conference on Computational Linguistics*, 2008.