Sentiment Analysis: An Overview from Linguistics

Maite Taboada

Department of Linguistics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada; email: mtaboada@sfu.ca

Annu. Rev. Linguist. 2016. 2:325–47

First published online as a Review in Advance on September 21, 2015

The *Annual Review of Linguistics* is online at linguist.annualreviews.org

This article's doi:
10.1146/annurev-linguistics-011415-040518

Keywords

opinion mining, negation, speculation, appraisal, evaluation, social media

Abstract

Sentiment analysis is a growing field at the intersection of linguistics and computer science that attempts to automatically determine the sentiment contained in text. Sentiment can be characterized as positive or negative evaluation expressed through language. Common applications of sentiment analysis include the automatic determination of whether a review posted online (of a movie, a book, or a consumer product) is positive or negative toward the item being reviewed. Sentiment analysis is now a common tool in the repertoire of social media analysis carried out by companies, marketers, and political analysts. Research on sentiment analysis extracts information from positive and negative words in text, from the context of those words, and from the linguistic structure of the text. This brief review examines in particular the contributions that linguistic knowledge can make to the task of automatically determining sentiment.

1. SENTIMENT, SUBJECTIVITY, OPINION, APPRAISAL, AFFECT, EMOTION

"I feel, therefore I am" could have preceded Descartes's statement. Feelings seem more primitive than thoughts, yet they constitute a significant portion of our lives, and the linguistic expression of emotions and opinions is one of the most fundamental human traits. Martin & White (2005) suggest that the expression of emotional states, or affect, comprises two further categories. The first one is the expression of judgment toward other people, and the second is the expression of appreciation, or aesthetic opinion. Together, affect, judgment, and appreciation capture how we convey our feelings and opinions, the object of study of sentiment analysis.

This expression of emotions and evaluations is studied under different umbrella terms in linguistics and other social sciences. In linguistics, studies of affect (Batson et al. 1992), subjectivity and point of view (Banfield 1982; Langacker 1990; Traugott 1995, 2010), evidentiality (Aikhenvald 2004, Chafe & Nichols 1986), attitudinal stance (Biber & Finegan 1988, 1989), modality (Bybee & Fleischman 1995, Palmer 1986, Portner 2009), and appraisal (Martin & White 2005) all aim to explain how we use language to convey emotions, evaluation, and subjectivity. Defining each of those terms could easily take up an entire paper. For the purposes of this review, I refer to subjectivity as the linguistic expression of belief, emotion, evaluation, or attitude (Wiebe 1994). In contrast, objective statements present events or describe the state of the world.

Research in linguistics, communication, and psychology has explored how we express, understand, and are affected by the expression of subjectivity (Caffi & Janney 1994, Krippendorf 2004); how we associate emotions and opinion with certain linguistic aspects, such as specific words or syntactic patterns (Biber & Finegan 1989, Hunston 2011, Stein 1995); and how we can classify linguistic expressions according to the type of opinion that they convey (Martin & White 2005). In this review, I concentrate on what has come to be called sentiment, the expression of subjectivity as either a positive or negative opinion. Many of the techniques and approaches, however, are applicable to closely related areas like the study and classification of emotions (anger, surprise, fear, etc.) as they are expressed in language.

Purely theoretical interest in the study of subjectivity and evaluation has been accompanied, in the last few years, by increased attention to how we express opinion online. This has opened up the field of sentiment analysis in computer science and computational linguistics, wherein subjectivity, opinion, and evaluation are captured for various purposes. This area of research is also referred to as opinion mining, perhaps due to interest from researchers in data mining and big data.

In this review, I briefly summarize the state of the art in extracting sentiment and opinion automatically, focusing on aspects of sentiment analysis that are most relevant to linguistics and drawing attention to areas where interaction would be beneficial. General surveys from a computational point of view are presented by Pang & Lee (2008), Liu (2012, 2015) and Sonntag & Stede (2014); Feldman (2013) offers a short overview for a lay audience.

2. THE "ANALYSIS" PART: COMPUTATIONAL METHODS

The main goal of sentiment analysis is to determine whether a text, or a part of it, is subjective or not and, if subjective, whether it expresses a positive or negative view. The direction of the opinion (i.e., whether positive or negative) is sometimes referred to as semantic orientation. Esuli & Sebastiani (2006) define the problem as having three different aspects: (*a*) determining the text's subjectivity (i.e., whether the text is factual in nature or whether it expresses an opinion on its subject matter); (*b*) determining the text's polarity, or deciding if a given subjective text expresses a positive or negative opinion on its subject matter; and (*c*) determining the strength of the text's polarity (i.e., deciding whether the positive opinion expressed by a text on its subject matter is weakly positive, mildly positive, or strongly positive).

Kim & Hovy (2004) suggest the importance of a fourth aspect as well: the source or holder of the opinion. They therefore define opinion as a quadruple, [Topic, Holder, Claim, Sentiment], in which the Holder believes a Claim about the Topic and associates a (positive or negative) Sentiment with that belief.

Much of this research has focused on analyzing reviews of movies, books, and consumer products (Dave et al. 2003, Hu & Liu 2004, Kennedy & Inkpen 2006, Turney 2002). Such research can find applications in search engines: Users searching for reviews of a movie can have the reviews automatically classified into positive or negative. Sentiment is relevant to producers as well as consumers. Because companies are interested in their reputation, and that of their product, it is in their interest to track online discussions and evaluate whether they are positive or negative. There is also an emerging field of analysis of political discourse (Mullen & Malouf 2006, Bakliwal et al. 2013, Ceron et al. 2015), including opinion pieces in newspapers. The applications in political life and policy making are obvious: A new form of polling, in which pollsters track online discussions rather than ask questions, could emerge. Other projects track the evolution of financial markets by following discussions online (Ahmad et al. 2006) or investor sentiment in message boards (Das & Chen 2007, Bollen et al. 2013). We have used these methods to track literary reputation using historical reviews (Taboada et al. 2006). Most recent applications have involved forms of blogging and microblogging, such as Twitter or Facebook messages (Thelwall et al. 2010, Mohammad et al. 2013, Kiritchenko et al. 2014, Ortigosa et al. 2014, Vilares et al. 2015b), including the Hedonometer project, an attempt to measure happiness on Twitter (Dodds et al. 2015).

The basic task in sentiment analysis, then, is to have enough information so that, when a new item (tweet, sentence, headline, excerpt, or whole text) needs to be processed, its characteristics can be extracted in order to decide whether it contains positive or negative sentiment, on the basis of existing information (see the sidebar). The crucial aspect of this task is where that information comes from. Two main approaches to the problem exist: machine learning and lexicon based.

In the machine learning approach, a classifier is built that can determine the polarity of new texts. The classifier is created by training on sentiment-labeled instances of other items (sentences, documents, etc.). This process is referred to as supervised learning because the classifier is given direction in terms of what good or bad examples of the class are. The classifier learns that certain characteristics distinguish a positive text from a negative one. Those characteristics are parameters in the learning and tend to be unigrams, that is, individual words or tokens that are present in the training data set. The classification may be binary (positive versus negative) or may include a third, neutral category.

The advantages of the machine learning approach are that, given a labeled data set—that is, one where documents have been previously determined to be positive or negative—training is

THE NATURE OF ONLINE TEXT

Recent sentiment analysis research in the domain of tweets, blogs, and Facebook posts has shown that adaptations are always necessary. Most researchers perform a first-pass cleaning of the data, correcting spelling mistakes, removing hashtags and URLs, and in general making the text more like formal written text, which is what most taggers and parsers expect. If dictionaries are used, they are also adapted to include emoticons and common online abbreviations. In machine learning approaches, however, the very nature of online text is exploited as a feature. The presence of capitalization and extra punctuation often indicates strong opinion, and can be added as a feature in classification.

trivial, and a classifier can be built quite quickly with existing classifier tools, such as WEKA by Witten & Frank (2005). For instance, a number of classifiers have been built using a set of 2,000 movie reviews, labeled according to whether their evaluation of the movie discussed is positive or negative (Pang et al. 2002). The resulting classifiers in most cases are able to correctly determine the polarity of unseen reviews 80% of the time (see Bloom et al. 2007, Andreevskaia & Bergler 2008, Prabowo & Thelwall 2009, Yessenalina et al. 2010, Socher et al. 2011, and Dinu & Iuga 2012, among many others).

Though machine learning approaches are desirable because of their accuracy, they often suffer from a number of disadvantages. First, because they are trained on very specific data, they are typically not portable to new sources of text. Applying the model to new contexts and data sets typically requires new training data and thus extensive human coding. For instance, most of the classifiers built using movie review data suffer from bias toward that genre, and they would not be able to capture the particular characteristics of other types of text, such as formal reviews or blog posts.

One of the most successful models for sentiment analysis is the Stanford Deep Learning for Sentiment Analysis model (Socher et al. 2013). This is quite a different machine learning approach because the labels are not documents or sentences, but phrases and their analysis in a parse tree. Because parsing information is used, the classifier in effect learns grammatical information along with clues to identify the polarity of individual words.

The other main approach to sentiment analysis, the lexicon-based or dictionary-based method, is often referred to as rule based because the dictionaries are applied following certain rules. On this approach, sentiment values of text are derived from the sentiment orientation of the individual words in the text, and using an existing dictionary. The dictionary contains words and their polarity (*excellent* is positive; *horrible* is negative). When a new text is encountered, words in the text are matched to words in the dictionary, and then their values are aggregated using one of various algorithms. An aggregation of the positive/negative values of the words in the text produces the semantic orientation for the entire text. A simplified representation of the two methods is provided in **Figure 1**. Serrano-Guerrero et al. (2015) also provide a visual classification of the different methods for sentiment analysis.

Lexicon-based methods are robust across different domains without changing the dictionaries (Taboada et al. 2011). Furthermore, Brooke et al. (2009) showed that porting dictionaries to a new language or a new domain is not an onerous task, probably less onerous than labeling data in a new domain for a classifier. Lexicon-based models make use of the linguistic information contained in the text. Because they show the most promise in terms of a good synergy between computational and linguistic approaches, I focus on these methods for the rest of this section, describing which linguistic aspects contribute to accurate extraction of sentiment.

The lexicon-based approach entails determining which words or phrases are relevant (i.e., Which words capture the evaluative meaning of a sentence or text?); which sentences are relevant (i.e., Are some sentences or parts of a text more representative of its orientation?); and how to aggregate the individual words or phrases extracted. I discuss each in turn below.

2.1. Which Words and Phrases

Adjectives convey much of the subjective content in a text, and a great deal of effort has therefore been devoted to extracting semantic orientation (i.e., positive and negative values) for adjectives. Hatzivassiloglou & McKeown (1997) pioneered the extraction of semantic orientation by association, using coordination: The phrase *excellent and X* predicts that *X* will be a positive adjective, in a situation where we do not know the polarity of *X*. Researchers have increasingly

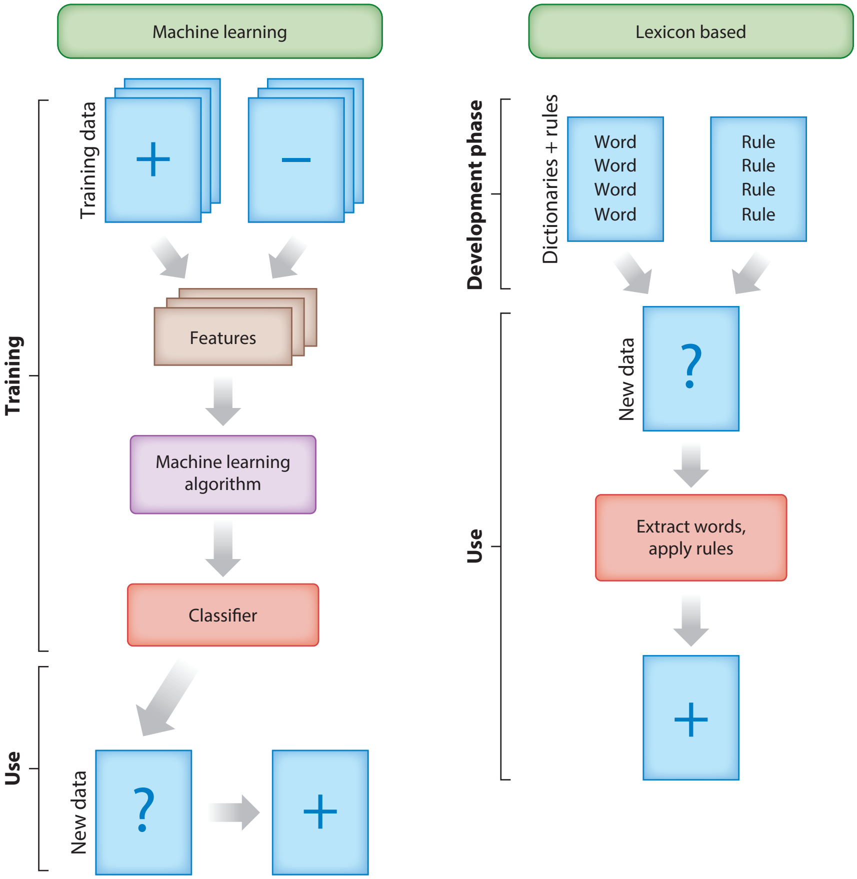

Figure 1

Machine learning and lexicon-based approaches to sentiment analysis.

noticed, however, that a great deal of sentiment is conveyed through other parts of speech, such as nouns (*masterpiece*, *disaster*), verbs (*love*, *hate*), adverbs (*skillfully*, *poorly*), and phrases that contain those words (Benamara et al. 2007, Subrahmanian & Reforgiato 2008).

Regardless of whether they include different parts of speech or only adjectives, dictionaries tend to contain lists of positive and negative words (i.e., specifying the polarity for each word). Many lexicon-based approaches also include information about strength, that is, how positive or negative the word is. For instance, in the subjectivity dictionary created by Wiebe and colleagues (Wiebe et al. 2004, Wilson et al. 2009), words can fall into the following categories:

- strong positive (*absolve*, *accolade*, *altruistic*),
- weak positive (*accept*, *abundance*, *affluent*),
- neutral (*accentuate*, *alliance*, *alert*),
- weak negative (*abolish*, *addiction*, *alienated*), and
- strong negative (*abuse*, *abomination*, *afraid*).

Table 1 Sample semantic orientation values for different dictionaries

	Subjectivity dictionary	SO-CAL	SentiWordNet	Macquarie dictionary
good	Positive (weak)	3	Positive	Positive
excellent	Positive (strong)	5	Positive	Positive
masterpiece	Positive (strong)	5	Positive	Positive
bad	Negative (strong)	−3	Negative	Negative
terrible	Negative (strong)	−5	Negative	Negative
disaster	Negative (strong)	−4	Negative	Negative

Other dictionaries have a more fine-grained scale. The dictionary in our system, the Semantic Orientation Calculator (SO-CAL), has a 10-point scale, from −5 to +5, which has shown good agreement with the judgments of human subjects (Taboada et al. 2011). A sample list of words and their values in different dictionaries is provided in **Table 1**, where the subjectivity dictionary refers to the dictionary created by Wiebe and colleagues from which the words above are taken. SentiWordNet (Baccianella et al. 2010) is a set of words extracted from WordNet, with positive, negative, and objective values added. In SentiWordNet, strength is associated with these words, but it needs to be computed across different senses and parts of speech for the same word. For simplicity, only polarity is noted in **Table 1**. Finally, the Macquarie Semantic Orientation Lexicon is a large collection of words annotated with semantic orientation by traversing *Roget's Thesaurus* (Mohammad et al. 2009).

Sentiment dictionaries vary widely in coverage. It is difficult to estimate the size of the evaluative lexicon of a language. Dictionaries for English (the object of the vast majority of the research; see Section 4, below) range from the roughly 5,000 words of SO-CAL (Taboada et al. 2011) or the 8,000 subjectivity clues proposed by Wilson et al. (2009) to the 38,000 words of SentiWordNet (Baccianella et al. 2010) or the almost 76,000 words of the Macquarie Semantic Orientation Lexicon (Mohammad et al. 2009). It is not clear what the optimal size is, or whether a language can possibly contain tens of thousands of evaluative terms. Our research group found that a large dictionary tends to capture more noise, leading to inaccurate results in automatic extraction of sentiment (Taboada et al. 2011).

Close examination of sentiment dictionaries, and of the opinions expressed (particularly) online, has revealed a relatively higher frequency of positive than negative terms. Such a phenomenon has been described as a form of the Pollyanna Principle (Boucher & Osgood 1969), whereby positive words have a higher frequency, in terms of both tokens and types, because we tend to remember past events positively. Some indicators suggest that, indeed, a great deal of online reviews are positive. TripAdvisor's own analysis indicates that hotel and destination reviews are largely positive (an average of 4.08 out of 5 points; see **http://resources.reviewpro.com/webinars/tripadvisor-how-to-improve-your-hotel-ranking-thanks**). Positive reviews account for 73% of the 1.36 million reviews in a corpus from the Internet Movie Database (Potts 2011).

The counterpart is the Negativity Bias, which postulates that negative events have a stronger effect on our psychological state and behavior (Rozin & Royzman 2001). If a negativity bias exists, then the lower frequency of negative terms can be accounted for because of their stronger effect. Jing-Schmidt (2007) argues that we use fewer negative than positive terms because of euphemism and political correctness. It is also possible that negative terms are simply positive terms that are negated. If one counts only evaluative words, without taking negation into account (see the next section), then naturally *good* and *not good* would both be tallied as positive. More generally, and with respect to the task of sentiment analysis, an important part of the process of deciding which words to include in the dictionary has to do with how to weigh them relative to one another. If the presence of a negative word is more indicative of a negative review, then negative words should perhaps carry more weight in the final aggregation.

2.2. Intensification and Downtoning; Irrealis and Nonveridicality

Whatever parts of speech are chosen as conveying sentiment, they can be intensified and downtoned by being modified. The general term intensifier is used for devices that change the intensity of an individual word, whether by bringing it up or down. These have also been described as valence shifters (Zaenen & Polanyi 2004), and as amplifiers versus downtoners (Quirk et al. 1985). Taking modifiers into account (whether intensifiers or downtoners) has consistently been shown to improve the performance of sentiment analysis systems (Kennedy & Inkpen 2006, Taboada et al. 2011, Morsy & Rafea 2012, Carrillo de Albornoz & Plaza 2013).

One model of intensification uses simple addition and subtraction (Kennedy & Inkpen 2006, Polanyi & Zaenen 2006). For example, if a positive adjective has a value of 2, an amplified (or positively intensified) adjective would become 3, and the downtoned version a 1. Intensifiers, however, do not all intensify at the same level—consider the difference between *extraordinarily* and *rather*. The value of the word being intensified also plays a role. A word at the higher end of the scale is probably intensified more intensely, as can be seen in the difference between *truly fantastic* and *truly okay*. In fact, the latter is probably often used ironically. A method to model these differences is to use multiplication rather than addition, that is, to place intensifiers on a percentage scale. Taboada et al. (2011) propose the following values:

- *most*, +100%,
- *really*, +25%,
- *very*, +15%
- *arguably*, −20%
- *somewhat*, −30%.

Polanyi & Zaenen (2006) also include other elements as valence shifters, such as presuppositional items (*even*, *barely*). Consider *It is barely sufficient*, which sets up the presupposition that, although *sufficient* is moderately positive, it is not in this case, because something better was expected.

Within the context of downtoning are typically discussed a host of phenomena that indicate that individual words and phrases may not be reliable for the purposes of sentiment analysis. Irrealis in general refers to expressions which indicate that the events mentioned in an utterance are not factual. Nonveridicality is wider, including all contexts that are not veridical, that is, not based on truth or existence (Giannakidou 1995, Zwarts 1995). In previous research, we defined the class of nonveridical operators as including negation (see the next section), modal verbs, intensional verbs (*believe*, *think*, *want*, *suggest*), imperatives, questions, protasis of conditionals, habituals, and the subjunctive (in languages that have an expression of subjunctive) (Trnavac & Taboada 2012). Consider the effect of the intensional verb *thought* and the modal *would* in example 1 and the modal plus question in example 2, which completely discounts any positive evaluation that may be present in *suitable* or *more suitable*.

(1) I thought this movie would be as good as the Grinch.

(2) Couldn't you find a more suitable ending?

A general consensus in sentiment analysis is that nonveridicality and irrealis result in the unreliability of any expression of sentiment in the sentences containing it (Wilson et al. 2009, Taboada et al. 2011, Benamara et al. 2012, Morante & Sporleder 2012, Denis et al. 2014), but not enough research has explored exactly how evaluation is affected. De Marneffe et al. (2012) characterize nonveridicality as a distribution over veridicality categories, rather than a binary classification, which would make accurate identification of nonveridical statements even more challenging.

A related area of study in the field of biomedical text processing is the role of speculation and negation. In biomedical text processing, the goal is to extract factual information from research literature; differentiating factual information from opinion or speculation is extremely important. A great deal of research in this area has focused on detecting speculation and negation, some of it with the help of the BioScope corpus. The BioScope corpus (Vincze et al. 2008) is a collection of abstracts, papers, and clinical reports annotated with cues that signal negation and speculation, as well as the scope of those cues. For instance, the verbs *suggest* and *indicate* introduce a signal that the finding is not completely reliable. Cues of speculation identified in this field partially overlap with the nonveridicality operators discussed above. Examples of cues are adjectives and adverbs (*probable*, *likely*, *possible*), modal verbs (*may*, *might*, *could*), verbs of speculation (*suggest*, *suspect*, *suppose*, *seem*), and a range of multiword cues (*no evidence/proof that*, *raise the possibility/question*, *whether or not*) (Farkas et al. 2010). Work in this field is increasingly making use of full sentence parsing or dependency parsing to identify the scope of cues (Velldall et al. 2012).

2.3. Negation

As with the asymmetry in frequency between positive and negative terms, negation in general shows interesting asymmetries, with important consequences for sentiment analysis. Negation detection usually involves finding a negator or an indication of negation, such as a negative-polarity item (words such as *any* or *at all* that appear in the presence of negation). The most crucial task, however, is to accurately capture the scope of negation, as it is important to negate only the evaluative item affected by negation.

In addition to the usual negator *not*, other negative words such as *no*, *none*, *nobody*, *nothing*, and *never* should be considered. Other words that may have a negative effect are *without*, *almost*, and *lack* (both as a noun and as a verb). Descriptions of negation and their scope, and how it can be identified computationally, can be found in Saurí (2008) and Blanco & Moldovan (2013).

Aspects of negation that are well known to linguists are syntactic versus morphological negation (examples 3*a* versus 3*b*), negation raising (example 3*c*), and negation scope and partial negation of only an argument (examples 3*d* and 3*e*). The following examples, unless otherwise indicated, are taken from the Simon Fraser University Review Corpus (Taboada 2008).

(3a) Mike Myers recycled his entire CV of SNL characters to create a Cat in the Hat that is unworthy of his name.

(3b) Mike Myers recycled his entire CV of SNL characters to create a Cat in the Hat that is not worthy of his name.

(3c) Our Sony phones died after 7 years. . . which I don't think it's too bad for a cordless phone.

(3d) I had stayed at Westin hotels before, and was never disappointed until now.

(3e) Propaganda doesn't succeed because it is manipulative, it works because people WANT it, NEED it, it gives their life a direction and meaning and guards against change. (*The Last Psychiatrist* 2013)

Assuming that negation and its scope have been adequately identified, the next problem is to decide how negation affects dictionary values for sentiment words. A straightforward strategy is to reverse the polarity of the lexical item in the scope of a negative item. For instance, in a system where dictionary words have both polarity and strength, *good* may have a value of +3, and under negation, *not good* may become −3. This approach is usually referred to as switch negation (Saurí 2008). Switch negation, however, does not capture well the subtleties of negation (Benamara et al. 2012, Liu & Seneff 2009). In highly positive words, a negation seems to imply a downtoning, rather than a reversal. For example, assuming that *excellent* may be a +5 adjective, *not excellent* hardly seems worthy of a −5, the polar opposite. In fact, it seems more positive than our −3 *not good* example. It just seems difficult to negate a strongly positive word without implying that a less positive one is to some extent possible (*not excellent*, *but not horrible either*). A possible solution is to use shift negation, in which the effect of a negator is to shift the negated term in the scale by a certain amount, but without making it the polar opposite of the original term. In my group's implementation of SO-CAL, shift negation moves the polarity by four points, resulting in the changes shown in example 4.

(4a)	excellent (+5)	→	not excellent (+1)
(4b)	terrific (+5)	→	not terrific (+1)
(4c)	sleazy (−3)	→	not sleazy (+1)
(4d)	horrid (−5)	→	not horrid (−1)

Litotes poses a particularly interesting challenge. This phenomenon involves conveying a mild positive by negating a negative item (*not bad*), or the opposite, using a negated positive to express a negative evaluation (*not my best day*). The effect seems to be one of downtoning the overall effect of the evaluation, whether positive or negative.

Another notable aspect of negation is its markedness. Negative statements tend to be perceived as more marked than their affirmative counterparts, both pragmatically and psychologically (Horn 1989, Osgood & Richards 1973). Negative forms are marked in terms of their linguistic form across languages (Greenberg 1966), and as mentioned above, they are less frequent. Potts (2011) posits an "emergent expressivity" for negation and negative polarity, observing that negative statements are less frequent and pragmatically more negative, with emphatic and attenuating polarity items modulating such negativity in a systematic way. Research in sentiment analysis has found that accurately identifying negative sentiment is more difficult, perhaps because we use fewer negative terms and because negative evaluation is couched in positive terms (Pang & Lee 2008, chapter 3). One way to solve this problem is to, in a sense, follow the Negativity Bias: If a negative word appears, then it has more impact. This has been achieved by weighing negative words more heavily than positives in aggregation (Taboada et al. 2011).

Yet another form of negation poses a particularly difficult challenge for sentiment analysis: irony. Thus far, no successful proposals exist for how to deal with (verbal) irony, which typically involves stating the opposite of what is meant, and can be understood as a narrower form of sarcasm (more generally, a sharp and aggressive remark). The intention to convey irony is not often expressed overtly. Some attempts have been made at using emoticons, where the emoticon carries what could be interpreted as the opposite polarity of the preceding statements (Carvalho et al. 2009, Tsur et al. 2010). Other surface indicators include acronyms or onomatopoeic expressions that indicate laughter (*LOL*, *he he*), heavy use of exclamation marks, and quotation marks. Riloff et al. (2013) achieve good success identifying ironic statements where a positive evaluator is associated with a negative situation (e.g., *love being ignored*, *enjoy shoveling the driveway*). Irony, however, draws upon a much more varied pool of resources than just a few surface indicators and is sometimes difficult even for humans to detect (Utsumi 2000). Indeed, the following examples, all titles of reviews from Tsur et al. (2010), would be difficult to interpret without some knowledge of context and, most importantly, without world knowledge (e.g., the fact that mentioning a book's cover as the main positive feature of the book implies a negative evaluation of its contents).

(5a) Love the cover (book)
(5b) Where am I? (GPS device)
(5c) Trees died for this book? (book)
(5d) Be sure to save your purchase receipt (smartphone)
(5e) Great for insomniacs (book)
(5f) Defective by design (music player)

Classic work in corpus linguistics has shown that certain patterns can be used to detect irony. Louw (1993) demonstrates that a clash in what he termed semantic prosody is indicative of irony. By semantic prosody, Louw refers to the positive or negative connotations that a word carries, and that go beyond the mere polarity described here. For instance, the verb *set in* is, at first sight, a neutral word. Upon corpus inspection, however, one can determine that it only collocates with negative events; in other words, only bad things set in. Similarly, one is always *bent on* pursuing negative actions. Louw uses this concept to show that *utterly* also carries negative prosody, in that it intensifies only negative words. When it accompanies a positive word, it is used ironically. Louw shows the same principle at play when discussing how David Lodge, in the novel *Small World*, characterizes academics attending conferences as "bent on self-improvement."

Despite these linguistic insights, however, most current research on irony and sarcasm detection is restricted to detecting irony using features from places where it is already present. For instance, a common approach involves collecting tweets with the tag #sarcasm, and then using those as labeled instances to learn the features that distinguish them from nonsarcastic comments (Bamman & Smith 2015). The advantage of applying machine learning and classification techniques to this problem is that it often helps reveal features of the text that are not easily accessible to the analyst. Commonly used features include the presence of certain words and expressions (*dare*, *clearly*, *lol*, *how dare*, *I'm shocked*), lexical density, capitalization and emoticons, and intensifiers. Bamman & Smith (2015) found that, although tweet features are useful, it is a combination of features of the author, the audience, and the characteristics of the tweet that works best at detecting irony. Bamman and Smith suggest that the #sarcasm tag is used when the author and the recipient do not actually know each other and have not interacted before. This means that authors feel compelled to add a tag when they think they will be misunderstood because of lack of context. There are probably instances of sarcasm among friends or peers that exhibit different features, and thus would not be detected with the classifier resulting from such a method.

A similar argument has been made in the detection of discourse relations when they are implicitly or explicitly marked by a conjunction or connective (Sporleder & Lascarides 2008). Using relations that are typically explicitly marked through a conjunction as training examples (with the conjunction removed) to detect typically implicit examples results in poor performance, probably because explicit relations do not share many features with implicit ones.

Underlying all of the research in detecting words and phrases, and the effects of valence shifters, is the principle of compositionality. Researchers take for granted that the sentiment of a document, a sentence, or a tweet is the sum of its parts. Some of the parts contribute more than others and some reduce or cancel out the sentiment, but the assumption is often that components can be added up, subtracted, or multiplied to yield a reliable result. As can easily be seen in the case of irony, such an assumption is not always correct. Words take on new meanings in context that are not predictable from what Wilson et al. (2005, 2009) have described as prior polarity. Haas & Versley (2015) point out that seemingly neutral adjectives can become polar when combined with aspects of a movie (*elaborate continuation*, *expanded vision*), as can words that are intensified (*simply intrusive* was considered negative, but *intrusive* was neutral).

2.4. Sentence and Clause Patterns

Evaluation and subjectivity are not only expressed by individual words and phrases but also often conveyed through entire sentences, and particular patterns in sentences. Pattern-based descriptions of language are especially relevant here because they avoid a distinction between lexis and grammar, instead treating them as part of the same object of description (Hunston & Francis 2000). Subjectivity spans the two, sometimes being conveyed by a single word, sometimes by a phrase, and sometimes by an entire grammatical structure. Hunston & Francis (2000, p. 37) define patterns of a word as "all the words and structures which are regularly associated with the word and which contribute to its meaning." In this section, I also include more general descriptions of grammatical structures, such as inversion.

The most in-depth description of patterns and evaluation is offered by Hunston (2011), who makes the case that certain patterns contribute to evaluative meanings, with a distinction between patterns that perform the function of evaluation, namely "performative" patterns (Hunston 2011, p. 139) and patterns that report evaluation. Examples of performative patterns are *it* and *there* patterns, as in *It is amazing that* . . . or *There is something admirable about* . . . . Hunston also discusses phrases that accompany evaluation, such as *as* (*is*) *humanly possible*, *to the point of*, or *bordering on*.

Many other researchers have noticed the potential of certain patterns to express subjectivity. Andersen & Fretheim (2000) discuss the *I think* (*that*) . . . pattern, characterized by a verb in the matrix clause such as *think*, *hope*, *understand*, or *wonder* and a complement clause. They discuss how the structure communicates the subject's attitude to the complement clause. Although the subject is typically a first person, some of the verbs allow third-person subjects (*She thought that the lock had been changed* but not *She took it that the lock had been changed*). Thompson (2002) proposes to consider introductory verbs such as *think* as markers of epistemic stance or evidentiality, and to reconsider the status of the clause as an "object complement." Verhagen (2005) argues that complement clauses are not "objects" but rather the main point of the complex sentence, and that the so-called matrix clause (*I think* . . .) instructs the addressee how to construe the complement.

Scheibman (2002), in a study of American English conversation, discusses the subjective content of certain syntactic structures, such as relational clauses (the most frequently occurring utterance type in her corpus). The predicates in those structures are typically adjectives (expressing an evaluation of the subject) and predicate nominals (expressing a relation between subject and predicate that is identifiable on the basis of subjective criteria). Scheibman (2002, p. 157) argues that both adjectives and predicate nominals in relational constructions function subjectively "in the sense that the relations conveyed by these utterances are contingent upon speaker point of view." Other features used to convey point of view are the first-person singular pronoun (*I*), present tense, modals, verbs of cognition, intensifiers, and modal adverbs.

Word order often plays a role in conveying stance. Socher et al. (2013, p. 1633) argue that "[f]rom a linguistic or cognitive standpoint, ignoring word order in the treatment of a semantic task is not plausible." Stein (1995) discusses the role of word order in expressing subjective meanings in

English. According to Stein (1995, p. 132), examples 6–8 represent a cline of emotional expression, the first one being the most subjective.

(6) Bitterly did they repent their decision.

(7) Bitterly they repented their decision.

(8) They repented their decision bitterly.

Wiebe and colleagues have devoted considerable effort to finding indicators of subjectivity in sentences (e.g., Wiebe & Riloff 2005, Wiebe et al. 2004, Wilson et al. 2006). They propose a set of clues to subjectivity, some lexical and some syntactic. Among the lexical clues are psychological verbs and verbs of judgment (*dread*, *love*, *commend*, *reprove*), verbs and adjectives that usually involve an experiencer (*fuss*, *worry*, *pleased*, *upset*, *embarrass*, *dislike*), and adjectives that have previously been annotated for polarity (Hatzivassiloglou & McKeown 1997). The syntactic clues are learned from manually annotated data (Riloff et al. 2003, Wiebe et al. 2003).

2.5. Relevant Sentences

It is obvious that not all parts of a text contribute equally to the possible overall opinion expressed therein. A movie review may contain sections relating to other movies by the same director, or with the same actors. Those sections have little or no bearing on the author's opinion toward the movie under discussion. A more complicated case involves texts where the author discusses a completely irrelevant topic (such as the restaurant she visited before the movie). In general, this is a topic-detection problem, to which solutions have been proposed (see, e.g., Allan 2002).

A slightly different problem is that of a text that contains mostly relevant information, some of which is more relevant than other. Less relevant aspects include background on the plot of the movie or book, or additional factual information on any aspect of the product. This problem has to do with distinguishing opinion from fact, or subjective from objective information. Wiebe and colleagues have annotated corpora with expressions of opinion (Wiebe et al. 2005) and have developed classifiers to distinguish objective from subjective sentences (Wiebe & Riloff 2005). Another way of weighing the text consists of identifying which parts consist of evaluation and which are mostly description. In reviews in particular, there may be a description of the product or of the context that is irrelevant to the evaluation. A movie review may describe the plot or the actors' previous roles, for instance. Taboada et al. (2009) proposed a method to automatically classify paragraphs in the text as description or evaluation, and showed that it improved the accuracy of the sentiment analysis.

Finally, another aspect of relevance is related to parts of the text that summarize or capture an overall opinion. Thus, among parts that contain an opinion related to the movie, some may be more useful than others. Adjectives (if those are the primary words used) in different parts of the text may have different weights (Pang et al. 2002, Taboada & Grieve 2004). Taboada & Grieve (2004) improved the performance of a semantic orientation calculator by weighing more heavily the words appearing toward the end of the text. This result is in line with an observation by Hunston & Thompson (2000, p. 11), attributed to John Sinclair, that "evaluation, in writing as in speech, tends to occur at the boundary points in a discourse."

2.6. Discourse Patterns

Once we have extracted words and phrases from a text, with or without having used a pruning method for sentences, the next step is to aggregate the semantic orientation, or evaluative value, of those individual words. The most commonly used method for this purpose is to average the semantic orientation of the words found in the text (Turney 2002). A text with 10 positive and 2 negative words would then be labeled positive. This technique obviously fails in many cases where discourse structure plays an important role in the construction of an argument. Consider the following example, a portion of a review of the movie *The Last Samurai*. Positive words are in bold, and words expressing negative evaluation are underlined (here I am considering mostly words, not their wider context, such as the modal verbs and perfective aspect in *could have been*).

(9) It could have been a **great** movie. It could have been **excellent**, and to all the people who have forgotten about the older, **greater** movies before it, will think that as well. It does have **beautiful** scenery, some of the **best** since Lord of the Rings. The acting is **well done**, and I **really liked** the son of the leader of the Samurai. He was a **likeable** chap, and I hated to see him die. But, other than all that, this movie is nothing more than hidden rip-offs.

This evaluation is clearly negative, but it is written in a style we have characterized as vernacular argumentation (Taboada & Gómez-González 2012), whereby a series of positive aspects are presented before a final fatal flaw or flaws, which summarize the opinion. Such examples make a compelling case for taking discourse structure into account, in particular discourse, coherence, or rhetorical relations (Mann & Thompson 1988). Such relations within and across sentences may change the polarity of sentiment words.

Relations of concession and condition are some of the relations proposed under various theories of discourse to account for the structure of discourse. For instance, a condition relation will limit the extent of a positive evaluation. In example 10, the positive evaluation in *interesting* is tempered by the condition that readers have to be able to change their expectations about the author's typical style and previous books.

(10) It is an interesting book if you can look at it with out expecting the Grisham "law and order" style.

Example 11 presents a concessive relation, marked by *while*. The polarity of the subordinate clause could be negative (a book being different, especially for prolific authors, tends to cause anxiety in loyal readers). The polarity of the main clause is clearly positive (*disappoint* + *not*). The change that the relation brings about in the combination of the subordinate and main clauses is one of reversal of the potential negative in the first clause.

(11) While this book is totally different than any other book he has written to date, it did not disappoint me at all.

Coherence relations interact with negation in interesting ways. Verhagen (2005) points out the negative–positive relation between concessive and causal relations, as in examples 12*a* and 12*b*, where the negation of the causal relation in the first sentence leads to a concessive reading in the second.

(12a) John is the best candidate because he happens to have a PhD.

(12b) John is not the best candidate because he happens to have a PhD.

The negation in example 12*b* does not necessarily imply a negation of the positive evaluation conveyed by *best*. Rather, it is a negation of the causal relation; that is, John is still the best candidate, but the reason is not that he has a PhD. Blanco & Moldovan (2013) refer to this phenomenon as partial negation.

Thus far, making use of coherence relations in sentiment analysis is mostly a proposal; methods to automatically parse the discourse structure of text are still in development, although significant advances have been made in the last few years (Hernault et al. 2010, Feng & Hirst 2014, Feng 2015, Joty et al. 2015). A related line of research involves investigating exactly how polarity words change in the context of a discourse relation (Trnavac & Taboada 2012, Benamara et al. 2013, Chardon et al. 2013).

3. A SMALL SAMPLE OF INTERESTING PROJECTS IN SENTIMENT ANALYSIS

This review is not intended to be comprehensive and include all examples of sentiment analysis to date. There are simply too many, both within academia and research settings and in commercial applications. Here, I discuss a few that are particularly interesting because of their approach or because of the subject matter or type of text being studied. Some of these projects deal with emotion rather than sentiment proper. While the computational study of emotions merits its own survey, I point out areas where there is an overlap with sentiment analysis.

First, many types of texts beyond the well-studied online reviews are being analyzed in terms of their sentiment. Politics is, of course, another ripe area for consideration, and early research focused on debates, blogs, and online discussions (Durant & Smith 2006, Mullen & Malouf 2006, Thomas et al. 2006). Tumasjan et al. (2010) exploit the potential of tweets about political parties to determine how well tweets align with the parties' stated values. Most interesting is their finding that the volume of messages may be a good indicator of election results, although this has been criticized as an artifact of data collection (Jungherr et al. 2012). Current approaches make use not only of the text but also of characteristics of the author and the author's online interactions (Qiu et al. 2015). Much of the research on political discourse uses Twitter and online media as a source (Yano et al. 2013) but also exploits the availability of large-scale corpora, such as the newswire articles collected in the English Gigaword corpus (O'Connor et al. 2013). As with other forms of social media, researchers have found that sarcasm poses a particularly difficult problem (Bakliwal et al. 2013).

Many other texts contain evaluation, sometimes of a personal and sensitive nature but still worthy of analysis. An interesting recent study by Stewart (2015) analyzed students' written comments in course evaluations, from a quantitative point of view, using the Appraisal framework (Martin & White 2005). I am not aware of any large-scale automatic analysis of student evaluations. Provided that issues of confidentiality can be solved, this area can lead to interesting applications. Of an even more sensitive nature are the suicide notes that Pestian et al. (2012) made available as part of a shared task. An important outcome of such analyses is determining who among those who attempt suicide are likely to try again. This is properly an emotion identification task, rather than simple polarity. The authors annotated certain emotions because they are considered to be good predictors, including, among others, abuse, anger, sorrow, forgiveness, love, and pride.

Bobicev et al. (2015) studied feelings expressed in online medical forums. They annotated a corpus of discussions about personal health (experiences with in vitro fertilization) with five types of feelings, which they described as sentiments: encouragement, gratitude, confusion, facts, and endorsement. Using the corpus as training data, they built a classifier to automatically identify those sentiments, showing that reliable identification of the sentiments is possible. This is a particularly interesting problem because the usual classification of the messages into positive and negative polarity would not provide enough fine-grained information for the authors' purposes, which include extracting sentiment from discussions on health care policy. It is also possible to annotate text below the level of the sentence. Socher et al. (2013) labeled not documents or sentences but phrases, and their analysis in a parse tree.

The vast majority of research on sentiment is conducted on text (unlike research on emotions, where speech is often analyzed in terms of prosody, pitch, and intonation). Some work, however, involves the detection of sentiment from images. Both Borth et al. (2013) and Wang et al. (2015) use a combination of characteristics of images posted online and text (comments and tags) about the images in order to identify the sentiment conveyed by images.

A related area of interest is the detection of opinion spam, or fake reviews. The popularity of reviews and the weight they carry in purchasing decisions have resulted in attempts to change ratings. Companies sometimes pay writers to produce either a large number of positive reviews or negative reviews about a rival's business. The practice has led to court cases and settlements; companies have been found guilty of paying for positive reviews or of writing them themselves (Streitfeld 2013). TripAdvisor was recently fined €500,000 for failing to prevent the posting of fake reviews on their site (Scott 2014). Detection of fake reviews employs many features that prove very useful for the task but are linguistically not so interesting, such as user IDs, user activity, URLs, and temporal patterns (Li et al. 2014). Some of the research, however, relies on the same principles that are deployed in authorship attribution: genre identification through part-of-speech distribution, similarity of linguistic patterns, and style characteristics of the text (Feng et al. 2012, Ott et al. 2011). This research allows systems to determine whether the same review is being posted on different sites, and whether certain stock phrases are being used repeatedly. Bing Liu has been a leader in this field, and his survey includes a chapter on how to detect fake reviews (Liu 2012; see also Liu 2015).

4. SENTIMENT ANALYSIS IN LANGUAGES OTHER THAN ENGLISH

English has been the main object of study in sentiment analysis. English is not, however, the only language in which opinions are expressed online. Accordingly, researchers are attempting to identify sentiment for other languages. Approaches vary. One obvious path is the native development of either lexicon-based or machine learning methods for the language in question. In dictionary-based approaches, this task involves creating a dictionary of polarity words in the language, together with appropriate rules to identify phenomena such as negation and intensification. In supervised learning methods, the main component needed is a labeled set of examples (texts, sentences, etc.).

The other main avenue, if "from scratch" development is not desirable or feasible, involves translation. One could translate texts in other languages, and then use an English-based sentiment analysis system. Alternatively, one could take English dictionaries and translate them into the target language, but that involves also adapting any rules being used.

Languages being studied with respect to sentiment analysis include Arabic (El-Beltagy & Ali 2013, Salameh et al. 2015), Chinese (Huang et al. 2012, Wan 2008, Wang et al. 2012, Ziyan et al. 2015), French (Ghorbel 2012, Marchand 2012, Benamara et al. 2013), German (Clematide & Klenner 2010, Waltinger 2010, Haas & Versley 2015), and Spanish (López et al. 2012; Moreno-Ortiz & Pérez Hernández 2012; Molina-González et al. 2013; Vilares et al. 2015a,b). In some cases, the focus is a combination of different languages (Mihalcea et al. 2007; Banea et al. 2008, 2014; Popat et al. 2013).

5. THE FUTURE AHEAD

The literature on sentiment analysis seems to be proliferating at an alarming rate, and it is often difficult to keep up with new developments in the field. There are many exciting and interesting projects in active development right now. There are also many small contributions, sometimes cumulative, sometimes derivative. For the field to prosper, I believe linguistic insight needs to be seriously considered, and a principled way of measuring progress has to be established. Ultimately, the real test is how useful the automatic classifications are. It is the sort of test that Google Translate provides. On one hand, if one can use the translations obtained through Google, then they are good enough. On the other hand, if either translations or sentiment classification are better than some baseline but otherwise useless for some practical purpose, then we need to rethink the direction the field is taking. One of the applications of sentiment analysis is in matching sentiment of markets and stocks to stock price (Feldman 2013). The real test here is whether one is willing to bet money that the sentiment–stock price correlation is accurate.

As with many other computational applications, many sentiment systems are proprietary and not available to the public. Some make their code publicly available, like the sentiment toolkit available in the Stanford CoreNLP package (Socher et al. 2013, **http://nlp.stanford.edu/sentiment/**). Readers who want to test sentiment analysis for themselves should consult Serrano-Guerrero et al. (2015), whose paper lists 15 different web services that allow textual input and output various types of sentiment information.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Ahmad K, Gillam L, Cheng D. 2006. Sentiments on a grid: analysis of streaming news and views. In *Proceedings of the 5th International Conference on Language Resources and Evaluation* (*LREC 5*), ed. N Calzolari, K Choukri, A Gangemi, B Maegaard, J Mariani, et al., pp. 2517–20. Genoa, Italy: Eur. Lang. Resour. Assoc.

Aikhenvald A. 2004. *Evidentiality*. Oxford, UK: Oxford Univ. Press

Allan J, ed. 2002. *Topic Detection and Tracking: Event-Based Information Organization*. New York: Springer

Andersen G, Fretheim T. 2000. Introduction. In *Pragmatic Markers and Propositional Attitude*, ed. G Andersen, T Fretheim, pp. 1–16. Amsterdam/Philadelphia: Benjamins

Andreevskaia A, Bergler S. 2008. When specialists and generalists work together: domain dependence in sentiment tagging. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pp. 290–98. Stroudsburg, PA: Assoc. Comput. Linguist.

Baccianella S, Esuli A, Sebastiani F. 2010. SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th Conference on International Language Resources and Evaluation* (*LREC 10*), ed. N Calzolari, K Choukri, B Maegaard, J Mariani, J Odijk, et al., pp. 2200–4. Valletta, Malta: Eur. Lang. Resour. Assoc.

Bakliwal A, Foster J, van der Puil J, O'Brien R, Tounsi L, Hughes M. 2013. Sentiment analysis of political tweets: towards an accurate classifier. In *Proceedings of the Workshop on Language Analysis in Social Media* (*LASM 2013*), pp. 49–58. Stroudsburg, PA: Assoc. Comput. Linguist.

Bamman D, Smith NA. 2015. Contextualized sarcasm detection on Twitter. In *Proceedings of the 9th International Conference on Web and Social Media*, pp. 574–77. Menlo Park, CA: AAAI

Banea C, Mihalcea R, Wiebe J. 2014. Sense-level subjectivity in a multilingual setting. *Comput. Speech Lang.* 28:7–19

Banea C, Mihalcea R, Wiebe J, Hassan S. 2008. Multilingual subjectivity analysis using machine translation. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2008*), pp. 127–35. Stroudsburg, PA: Assoc. Comput. Linguist.
Banfield A. 1982. *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Boston: Routledge/Kegan Paul
Batson CD, Shaw LL, Oleson KC. 1992. Differentiating affect, mood, and emotion: toward functionally based conceptual distinctions. In *Emotion: Review of Personality and Social Psychology*, ed. MS Clark, pp. 294–326. Newbury Park, CA: Sage
Benamara F, Cesarano C, Picariello A, Reforgiato D, Subrahmanian V. 2007. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *Proceedings of the 1st International Conference on Weblogs and Social Media* (*ICWSM*). **http://www.icwsm.org/papers/paper31.html**
Benamara F, Chardon B, Mathieu YY, Popescu V, Asher N. 2012. How do negation and modality impact opinions? In *Proceedings of the ACL-2012 Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics* (*ExProM-2012*), pp. 10–18. Stroudsburg, PA: Assoc. Comput. Linguist.
Benamara F, Popescu V, Chardon B, Asher N, Mathieu Y. 2013. Assessing opinions in texts: Does discourse really matter? In *Nonveridicality and Evaluation: Theoretical, Computational and Corpus Approaches*, ed. M Taboada, R Trnavac, pp. 127–50. Leiden, Neth.: Brill
Biber D, Finegan E. 1988. Adverbial stance types in English. *Discourse Process.* 11:1–34
Biber D, Finegan E. 1989. Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text* 9:93–124
Blanco E, Moldovan D. 2013. Retrieving implicit positive meaning from negated statements. *Nat. Lang. Eng.* 20:501–35
Bloom K, Garg N, Argamon S. 2007. Extracting appraisal expressions. In *Proceedings of the 2007 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* (*HLT/NAACL*), pp. 308–15. Stroudsburg, PA: Assoc. Comput. Linguist.
Bobicev V, Sokolova M, Oakes M. 2015. What goes around comes around: learning sentiments in online medical forums. *Cogn. Comput.* 7:609–21
Bollen J, Mao H, Zeng X. 2013. Twitter mood predicts the stock market. *J. Comput. Sci.* 2:1–8
Borth D, Ji R, Chen T, Breuel T, Chang S-F. 2013. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM International Conference on Multimedia*, pp. 223–32. New York: ACM
Boucher JD, Osgood CE. 1969. The Pollyanna hypothesis. *J. Verbal Learn. Verbal Behav.* 8:1–8
Brooke J, Tofiloski M, Taboada M. 2009. Cross-linguistic sentiment analysis: from English to Spanish. In *Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing*, ed. G Angelova, K Bontcheva, R Mitkov, N Nicolov, pp. 50–54. Shoumen, Bulg.: Incoma
Bybee J, Fleischman S, ed. 1995. *Modality in Grammar and Discourse*. Amsterdam/Philadelphia: Benjamins
Caffi C, Janney RW. 1994. Towards a pragmatics of emotive communication. *J. Pragmat.* 22:325–73
Carrillo de Albornoz J, Plaza L. 2013. An emotion-based model of negation, intensifiers, and modality for polarity and intensity classification. *J. Am. Soc. Inf. Sci. Technol.* 64:1618–33
Carvalho P, Sarmento L, Silva MJ, Oliveira Ed. 2009. Clues for detecting irony in user-generated contents: Oh..!! It's "so easy";-). In *Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement*, p. 53. New York: ACM
Ceron A, Curini L, Iacus SM. 2015. Using sentiment analysis to monitor electoral campaigns: Method matters—evidence from the United States and Italy. *Soc. Sci. Comput. Rev.* 33:3–20
Chafe W, Nichols J. 1986. *Evidentiality: The Linguistic Coding of Epistemology*. Norwood, NJ: Ablex
Chardon B, Benamara F, Mathieu YY, Popescu V, Asher N. 2013. Measuring the effect of discourse structure on sentiment analysis. In *Proceedings of the 14th International Conference on Computational Linguistics and Intelligent Text Processing* (*CICLing 2013*), pp. 25–37. Berlin: Springer
Clematide S, Klenner M. 2010. Evaluation and extension of a polarity lexicon for German. In *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis* (*WASSA*), pp. 7–14. Lisbon: Univ. Alicante
Das SR, Chen MY. 2007. Yahoo! for Amazon: opinion extraction from small talk on the web. *J. Manag. Sci.* 53:1375–88

Dave K, Lawrence S, Pennock DM. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International World Wide Web Conference* (*WWW 2003*), pp. 519–28. New York: ACM

Denis A, Cruz-Lara S, Bellalem N, Bellalem L. 2014. Synalp-Emphatic: a valence shifting hybrid system for sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation* (*SemEval 2014*), pp. 605–9. **http://alt.qcri.org/semeval2014/cdrom/pdf/SemEval106.pdf**

Dinu LP, Iuga I. 2012. The Naive Bayes Classifier in opinion mining: in search of the best feature set. In *Computational Linguistics and Intelligent Text Processing*, ed. A Gelbukh, pp. 556–67. Berlin: Springer

Dodds PS, Clark EM, Desu S, Frank MR, Reagan AJ, et al. 2015. Human language reveals a positivity bias. *PNAS* 112:2389–94

Durant KT, Smith MD. 2006. Mining sentiment classification from political web logs. In *Proceedings of the Workshop on Web Mining and Web Usage Analysis of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM. **http://webmining.spd.louisville.edu/webkdd06/papers/paper-17-Mining%20Sentiment%20Classification%20from%20Political%20Web%20Logs-WM_1069%5b1%5d.pdf**

El-Beltagy SR, Ali A. 2013. Open issues in the sentiment analysis of Arabic social media: a case study. In *Proceedings of the 9th International Conference on Innovations in Information Technology*, pp. 215–20. Piscataway, NJ: IEEE

Esuli A, Sebastiani F. 2006. SentiWordNet: a publicly available lexical resource for opinion mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation* (*LREC 5*), ed. N Calzolari, K Choukri, A Gangemi, B Maegaard, J Mariani, et al., pp. 417–22. Genoa, Italy: Eur. Lang. Resour. Assoc.

Farkas R, Vincze V, Móra G, Csirik J, Szarvas G. 2010. The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In *Proceedings of the 14th Conference on Computational Natural Language Learning: Shared Task*, pp. 1–12. Stroudsburg, PA: Assoc. Comput. Linguist.

Feldman R. 2013. Techniques and applications for sentiment analysis: the main applications and challenges of one of the hottest research areas in computer science. *Commun. ACM* 56:82–89

Feng S, Banerjee R, Choi Y. 2012. Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pp. 171–75. Stroudsburg, PA: Assoc. Comput. Linguist.

Feng VW. 2015. *RST-style discourse parsing and its applications in discourse analysis*. PhD thesis, Grad. Dep. Comput. Sci., Univ. Toronto, Can.

Feng VW, Hirst G. 2014. A linear-time bottom-up discourse parser with constraints and post-editing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 1:511–21. Stroudsburg, PA: Assoc. Comput. Linguist.

Ghorbel H. 2012. Experiments in cross-lingual sentiment analysis in discussion forums. In *Proceedings of the 4th International Conference on Social Informatics*, ed. K Aberer, A Flache, W Jager, L Liu, J Tang, C Guéret, pp. 138–51. Berlin: Springer

Giannakidou A. 1995. On the semantic licensing of polarity items. In *Studies in Greek Linguistics 15: Proceedings of the 15th Annual Meeting of the Department of Linguistics*, ed. A-P Christidis, M Margariti-Roga, A Arhakis, pp. 406–18. Thessaloniki, Greece: Univ. Thessaloniki

Greenberg JH. 1966. *Language Universals, with Special Reference to Feature Hierarchies*. The Hague: Mouton

Haas M, Versley Y. 2015. Subsentential sentiment on a shoestring: a crosslingual analysis of compositional classification. In *Proceedings of the 2015 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* (*HLT/NAACL*), pp. 694–704. Red Hook, NY: Curran

Hatzivassiloglou V, McKeown K. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Meeting of the Association for Computational Linguistics*, pp. 174–81. Stroudsburg, PA: Assoc. Comput. Linguist.

Hernault H, Prendinger H, duVerle DA, Ishizuka M. 2010. HILDA: a discourse parser using Support Vector Machine classification. *Dialogue Discourse* 1:1–33

Horn LR. 1989. *A Natural History of Negation*. Chicago: Univ. Chicago Press

Hu M, Liu B. 2004. Mining and summarizing customer reviews. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD-2004*), pp. 168–77. New York: ACM
Huang T-H, Yu H-C, Chen H-H. 2012. Modeling Polyanna phenomena in Chinese sentiment analysis. In *Proceedings of the 12th Annual Conference on Computational Linguistics* (*COLING 2012*)*: Demonstration Papers*, pp. 231–38. Stroudsburg, PA: Assoc. Comput. Linguist.
Hunston S. 2011. *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. New York: Routledge
Hunston S, Francis G. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam/Philadelphia: Benjamins
Hunston S, Thompson G. 2000. Evaluation: an introduction. In *Evaluation in Text: Authorial Distance and the Construction of Discourse*, ed. S Hunston, G Thompson, pp. 1–27. Oxford, UK: Oxford Univ. Press
Jing-Schmidt Z. 2007. Negativity bias in language: a cognitive-affective model of emotive intensifiers. *Cogn. Linguist.* 18:417–43
Joty S, Carenini G, Ng R. 2015. CODRA: a novel discriminative framework for rhetorical analysis. *Comput. Linguist.* 41:385–435
Jungherr A, Jürgens P, Schoen H. 2012. Why the Pirate Party won the German election of 2009, or the trouble with predictions: a response to Tumasjan, A., Sprenger, T.O., Sander, P.G., & Welpe, I.M. "Predicting elections with Twitter: What 140 characters reveal about political sentiment". *Soc. Sci. Comput. Rev.* 30:229–34
Kennedy A, Inkpen D. 2006. Sentiment classification of movie and product reviews using contextual valence shifters. *Comput. Intell.* 22:110–25
Kim S-M, Hovy E. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th Annual Conference on Computational Linguistics* (*COLING 2004*), artic. number 1367. Stroudsburg, PA: Assoc. Comput. Linguist.
Kiritchenko S, Zhu X, Mohammad S. 2014. Sentiment analysis of short informal texts. *J. Artif. Intell. Res.* 50:723–62
Krippendorf K. 2004. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage
Langacker RW. 1990. Subjectification. *Cogn. Linguist.* 1:5–38
Li H, Mukherjee D, Liu B, Kornfield R, Emery S. 2014. Detecting campaign promoters on Twitter using Markov Random Fields. In *Proceedings of the IEEE International Conference on Data Mining* (*ICDM-2014*), pp. 290–99. Piscataway, NJ: IEEE
Liu B. 2012. *Sentiment Analysis and Opinion Mining*. San Rafael, CA: Morgan & Claypool
Liu B. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions.* Cambridge, UK: Cambridge Univ. Press.
Liu J, Seneff S. 2009. Review sentiment scoring via a parse-and-paraphrase paradigm. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2009*), pp. 161–69. Stroudsburg, PA: Assoc. Comput. Linguist.
López R, Tejada J, Thelwall M. 2012. Spanish SentiStrength as a tool for opinion mining Peruvian Facebook and Twitter. In *Artificial Intelligence Driven Solutions to Business and Engineering Problems*, ed. G Setlak, M Alexandrov, K Markov, pp. 82–85. Sofia, Bulg.: Ithea
Louw B. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In *Text and Technology: In Honour of John Sinclair*, ed. M Baker, G Francis, E Tognini-Bonelli, pp. 157–76. Amsterdam: Benjamins
Mann WC, Thompson SA. 1988. Rhetorical Structure Theory: toward a functional theory of text organization. *Text* 8:243–81
Marchand M. 2012. État de l'art: l'influence du domaine sur la classification de l'opinion. In *Proceedings of the Joint Conference JEP-TALN-RECITAL*, ed. L Basacier, B Lecouteux, G Sérasset, 1:177–90. Grenoble, France: AFCP/ATALA
de Marneffe M-C, Manning C, Potts C. 2012. Did it happen? The pragmatic complexity of veridicality assessment. *Comput. Linguist.* 38:301–33
Martin JR, White PRR. 2005. *The Language of Evaluation*. New York: Palgrave
Mihalcea R, Banea C, Wiebe J. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 976–83. Stroudsburg, PA: Assoc. Comput. Linguist.

Mohammad S, Dorr B, Dunne C. 2009. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2009*), pp. 599–608. Stroudsburg, PA: Assoc. Comput. Linguist.

Mohammad S, Kiritchenko S, Zhu X. 2013. NRC-Canada: building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of the 7th International Workshop on Semantic Evaluation Exercises* (*SemEval 2013*), pp. 443–47. **http://www.aclweb.org/anthology/S14-2077**

Molina-González MD, Martínez-Cámara E, Martín-Valdivia M-T, Perea-Ortega JM. 2013. Semantic orientation for polarity classification in Spanish reviews. *Expert Syst. Appl.* 40:7250–57

Morante R, Sporleder C. 2012. Modality and negation: an introduction to the special issue. *Comput. Linguist.* 38:223–60

Moreno-Ortiz A, Pérez Hernández L. 2012. Lexicon-based sentiment analysis of Twitter messages in Spanish. In *Taller de Análisis de Sentimientos en la SEPLN* (*Sociedad Española para el Procesamiento del Lenguaje Natural*), pp. 93–100. **http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4664**

Morsy SA, Rafea A. 2012. Improving document-level sentiment classification using contextual valence shifters. In *Natural Language Processing and Information Systems: Proceedings of the 17th International Conference on Applications of Natural Language to Information Systems*, ed. G Bouma, A Ittoo, E Métais, H Wortmann, pp. 253–58. Groningen, Neth.: Springer

Mullen T, Malouf R. 2006. A preliminary investigation into sentiment analysis of informal political discourse. In *Proceedings of the AAAI 2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, pp. 125–26. Menlo Park, CA: AAAI

O'Connor B, Stewart BM, Smith NA. 2013. Learning to extract international relations from political context. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 1094–104. Stroudsburg, PA: Assoc. Comput. Linguist.

Ortigosa A, Martín JM, Carro RM. 2014. Sentiment analysis in Facebook and its applications to e-learning. *Comput. Hum. Behav.* 31:527–41

Osgood CE, Richards MM. 1973. From Yang and Yin to *and* or *but*. *Language* 49:380–412

Ott M, Choi Y, Cardie C, Hancock JT. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pp. 309–19. Stroudsburg, PA: Assoc. Comput. Linguist.

Palmer F. 1986. *Mood and Modality*. Cambridge, UK: Cambridge Univ. Press

Pang B, Lee L. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* 2:1–135

Pang B, Lee L, Vaithyanathan S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2002*), pp. 79–86. Stroudsburg, PA: Assoc. Comput. Linguist.

Pestian JP, Matykiewicz P, Linn-Gust M, South B, Uzuner O, et al. 2012. Sentiment analysis of suicide notes: a shared task. *Biomed. Inf. Insights* 5:3–16

Polanyi L, Zaenen A. 2006. Contextual valence shifters. In *Computing Attitude and Affect in Text: Theory and Applications*, ed. JG Shanahan, Y Qu, J Wiebe, pp. 1–10. Dordrecht, Neth.: Springer

Popat K, Balamurali AR, Battacharyya P, Haffari G. 2013. The haves and the have-nots: leveraging unlabelled corpora for sentiment analysis. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 412–22. Stroudsburg, PA: Assoc. Comput. Linguist.

Portner P. 2009. *Modality*. Oxford, UK: Oxford Univ. Press

Potts C. 2011. On the negativity of negation. In *Proceedings of the 20th Semantics and Linguistic Theory* (*SALT 20*), pp. 636–59. Ithaca, NY: CLC

Prabowo R, Thelwall M. 2009. Sentiment analysis: a combined approach. *J. Informetr.* 3:143–57

Qiu M, Sim Y, Smith NA, Jiang J. 2015. *Modeling user arguments, interactions and attributes for stance prediction in online debate forums*. Presented at SIAM Int. Conf. Data Mining, Vancouver, Can.

Quirk R, Greenbaum S, Leech G, Svartvik J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman

Riloff E, Qadir A, Surve P, De Silva L, Gilbert N, Huang R. 2013. Sarcasm as contrast between a positive sentiment and negative situation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2013*), pp. 704–14. Stroudsburg, PA: Assoc. Comput. Linguist.

Riloff E, Wiebe J, Wilson T. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the 7th Conference on Natural Language Learning* (*CoNLL 7*), pp. 25–32. **http://aclweb.org/anthology/W/W13/W13-1602.pdf**

Rozin P, Royzman EB. 2001. Negativity bias, negativity dominance, and contagion. *Personal. Soc. Psychol. Rev.* 5:296–320

Salameh M, Mohammad S, Kiritchenko S. 2015. Sentiment analysis after translation: a case-study on Arabic social media posts. In *Proceedings of the 2015 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* (*HLT/NAACL*), pp. 767–77. Red Hook, NY: Curran

Saurí R. 2008. *A factuality profiler for eventualities in text*. PhD thesis, Comput. Sci. Dep., Brandeis Univ., Waltham, MA

Scheibman J. 2002. *Point of View and Grammar: Structural Patterns of Subjectivity in American English*. Amsterdam/Philadelphia: Benjamins

Scott M. 2014. TripAdvisor fined $610,000 in Italy for failing to prevent fake reviews. *New York Times*, Dec. 23, p. B8

Serrano-Guerrero J, Olivas JA, Romero FP, Herrera-Viedma E. 2015. Sentiment analysis: a review and comparative analysis of web services. *Inf. Sci.* 311:18–38

Socher R, Pennington J, Huang EH, Ng AY, Manning C. 2011. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2011*), pp. 151–61. Stroudsburg, PA: Assoc. Comput. Linguist.

Socher R, Perelygin A, Wu JY, Chuang J, Manning C, et al. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2013*), pp. 1631–42. Stroudsburg, PA: Assoc. Comput. Linguist.

Sonntag J, Stede M. 2014. Sentiment analysis: What's your opinion? In *Text Mining: From Ontology Learning to Automated Text Processing Applications*, ed. C Biemann, A Mehler, pp. 177–99. Berlin: Springer

Sporleder C, Lascarides A. 2008. Using automatically labelled examples to classify rhetorical relations: an assesment. *Nat. Lang. Eng.* 14:369–416

Stein D. 1995. Subjective meanings and the history of inversions in English. In *Subjectivity and Subjectivisation: Linguistic Perspectives*, ed. D Stein, S Wright, pp. 129–50. Cambridge, UK: Cambridge Univ. Press

Stewart M. 2015. The language of praise and criticism in a student evaluation survey. *Stud. Educ. Eval.* 45:1–9

Streitfeld D. 2013. Give yourself 5 stars? Online, it might cost you. *New York Times*, Sept. 22, p. A1

Subrahmanian V, Reforgiato D. 2008. AVA: adjective–verb–adverb combinations for sentiment analysis. *Intell. Syst.* 23:43–50

Taboada M. 2008. *SFU Review Corpus*. Vancouver, Can.: Simon Fraser Univ. **http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html**

Taboada M, Brooke J, Stede M. 2009. Genre-based paragraph classification for sentiment analysis. In *Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue*, ed. P. Healey, R Pieraccini, D Byron, S Young, M Purver, pp. 62–70. Stroudsburg, PA: Assoc. Comput. Linguist.

Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. 2011. Lexicon-based methods for sentiment analysis. *Comput. Linguist.* 37:267–307

Taboada M, Gillies MA, McFetridge P. 2006. Sentiment classification techniques for tracking literary reputation. In *Proceedings of the 5th International Conference on Language Resources and Evaluation* (*LREC 5*), ed. N Calzolari, K Choukri, A Gangemi, B Maegaard, J Mariani, et al., pp. 36–43. Genoa, Italy: Eur. Lang. Resour. Assoc.

Taboada M, Gómez-González M. 2012. Discourse markers and coherence relations: comparison across markers, languages and modalities. *Linguist. Hum. Sci.* 6:17–41

Taboada M, Grieve J. 2004. Analyzing appraisal automatically. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, ed. Y Qu, JG Shanahan, J Wiebe, pp. 158–61. Menlo Park, CA: AAAI

The Last Psychiatrist. 2013. How does the shutdown relate to me? *The Last Psychiatrist* blog, Sept. 25. **http://thelastpsychiatrist.com/2013/09/how_does_the_shutdown_relate_t.html**

Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. 2010. Sentiment strength detection in short informal text. *J. Am. Soc. Inf. Sci. Technol.* 61:2544–58

Thomas M, Pang B, Lee L. 2006. Get out the vote: determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2006*), pp. 327–35. Stroudsburg, PA: Assoc. Comput. Linguist.

Thompson SA. 2002. "Object complements" and conversation: towards a realistic account. *Stud. Lang.* 26:125–64

Traugott EC. 1995. Subjectification in grammaticalisation. In *Subjectivity and Subjectivisation: Linguistic Perspectives*, ed. D Stein, S Wright, pp. 31–54. Cambridge, UK: Cambridge Univ. Press

Traugott EC. 2010. (Inter)subjectivity and (inter)subjectification: a reassessment. In *Subjectification, Intersubjectification and Grammaticalization*, ed. K Davidse, L Vandelanotte, H Cuyckens, pp. 29–74. Berlin: de Gruyter

Trnavac R, Taboada M. 2012. The contribution of nonveridical rhetorical relations to evaluation in discourse. *Lang. Sci.* 34:301–18

Tsur O, Davidov D, Rappoport A. 2010. ICWSM—a great catchy name: semi-supervised recognition of sarcastic sentences in online product reviews. In *Proceedings of the 4th International Conference on Weblogs and Social Media* (*ICWSM*), pp. 162–69. Menlo Park, CA: AAAI

Tumasjan A, Sprenger TO, Sandner PG, Welpe IM. 2010. Predicting elections with Twitter: what 140 characters reveal about political sentiment. In *Proceedings of the 4th International Conference on Weblogs and Social Media* (*ICWSM*), pp. 178–85. Menlo Park, CA: AAAI

Turney P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics*, pp. 417–24. Stroudsburg, PA: Assoc. Comput. Linguist.

Utsumi A. 2000. Verbal irony as implicit display of ironic environment: distinguishing ironic utterances from nonirony. *J. Pragmat.* 32:1777–806

Velldall E, Ovrelid L, Read J, Oepen S. 2012. Speculation and negation: rules, rankers, and the role of syntax. *Comput. Linguist.* 38:369–410

Verhagen A. 2005. *Constructions of Intersubjectivity: Discourse, Syntax, and Cognition*. Oxford, UK: Oxford Univ. Press

Vilares D, Alonso MA, Gómez-Rodríguez C. 2015a. A syntactic approach for opinion mining on Spanish reviews. *Nat. Lang. Eng.* 21:139–63

Vilares D, Alonso MA, Gómez-Rodríguez C. 2015b. On the usefulness of lexical and syntactic processing in polarity classification of Twitter messages. *J. Assoc. Inf. Sci. Technol.* 66:1799–816

Vincze V, Szarvas G, Farkas R, Móra G, Csirik J. 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinform.* 9(Suppl. 11):S9

Waltinger U. 2010. GermanPolarityClues: a lexical resource for German sentiment analysis. In *Proceedings of the 7th Conference on International Language Resources and Evaluation* (*LREC 10*), ed. N Calzolari, K Choukri, B Maegaard, J Mariani, J Odijk, et al., pp. 1638–42. Valletta, Malta: Eur. Lang. Resour. Assoc.

Wan X. 2008. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2008*), pp. 553–61. Stroudsburg, PA: Assoc. Comput. Linguist.

Wang F, Wu Y, Qiu L. 2012. Exploiting discourse relations for sentiment analysis. In *Proceedings of the 24th International Conference on Computational Linguistics* (*COLING 2012*), pp. 1311–20. Stroudsburg, PA: Assoc. Comput. Linguist.

Wang Y, Hu Y, Kambhampati S, Li B. 2015. Inferring sentiment from web images with joint inference on visual and social cues: a regulated matrix factorization approach. In *Proceedings of the 9th International Conference on Web and Social Media*, pp. 473–82. Menlo Park, CA: AAAI

Wiebe J. 1994. Tracking point of view in narrative. *Comput. Linguist.* 20:233–87

Wiebe J, Breck E, Buckley C, Cardie C, Davis P, et al. 2003. Recognizing and organizing opinions expressed in the world press. In *Working Notes of the AAAI Spring Symposium in New Directions in Question Answering*, pp. 12–19. Menlo Park, CA: AAAI

Wiebe J, Riloff E. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *Proceedings of the 6th International Conference on Intelligent Text Processing and Computational Linguistics* (*CICLing 2005*), pp. 486–97. Berlin: Springer

Wiebe J, Wilson T, Bruce R, Bell M, Martin M. 2004. Learning subjective language. *Comput. Linguist.* 30:277–308

Wiebe J, Wilson T, Cardie C. 2005. Annotating expressions of opinions and emotions in language. *Lang. Resour. Eval.* 39:165–210

Wilson T, Wiebe J, Hoffmann P. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the 2005 Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing* (*HLT/EMNLP-05*), pp. 347–54. Stroudsburg, PA: Assoc. Comput. Linguist.

Wilson T, Wiebe J, Hoffmann P. 2009. Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. *Comput. Linguist.* 35:399–433

Wilson T, Wiebe J, Hwa R. 2006. Recognizing strong and weak opinion clauses. *Comput. Intell.* 22:73–99

Witten IH, Frank E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann

Yano T, Yogatama D, Smith NA. 2013. A penny for your tweets: campaign contributions and Capitol Hill microblogging. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, pp. 737–40. Menlo Park, CA: AAAI

Yessenalina A, Choi Y, Cardie C. 2010. Automatically generating annotator rationales to improve sentiment classification. In *Proceedings of the Association for Computational Linguistics 2010: Conference Short Papers*, pp. 336–41. Stroudsburg, PA: Assoc. Comput. Linguist.

Zaenen A, Polanyi L. 2004. Contextual valence shifters. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, ed. Y Qu, JG Shanahan, J Wiebe, pp. 106–11. Menlo Park, CA: AAAI

Ziyan C, Huang Y, Tian J, Liu X, Fu K, Huang T. 2015. Joint model for subsentence-level sentiment analysis with Markov logic. *J. Assoc. Inf. Sci. Technol.* 66:1913–22

Zwarts F. 1995. Nonveridical contexts. *Linguist. Anal.* 25:286–312