Open access • Journal Article • DOI:10.1146/ANNUREV-LINGUISTICS-011415-040518

Sentiment Analysis: An Overview from Linguistics — Source link

Maite Taboada

Institutions: Simon Fraser University

Published on: 14 Jan 2016 - Social Science Research Network (Annual Reviews)

Topics: Sentiment analysis

Sentiment Analysis: An overview from Linguistics*

Annual Review of Linguistics

Maite Taboada

Department of Linguistics

Simon Fraser University

8888 University Dr.

Burnaby, BC, V5A 1S6

Canada

mtaboada@sfu.ca

Table of Contents

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

* Taboada, M. (2016) Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics.* 2: 325-347. Pre-publication version.

Abstract: Sentiment analysis is a growing field at the intersection of linguistics and computer science, which attempts to automatically determine the sentiment, or positive/negative opinion, 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 towards 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 the linguistic structure of the text. This brief survey examines in particular the contributions that linguistic knowledge can make to the problem of automatically determining sentiment.

1 Sentiment, subjectivity, opinion, appraisal, affect, emotion

"I feel, therefore I am" could have preceded Descartes' statement. Feelings seem more primitive than thought, yet they constitute a significant portion of our lives. Emotions, opinion, and their expression in language are probably one of the most fundamental human traits. Martin and White (2005) suggest that the expression of emotional states, or affect, is institutionalized into two further categories. The first one is the expression of judgement towards other people, and the second the categorization 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, Traugott 2010), evidentiality (Aikhenvald 2004, Chafe & Nichols 1986), attitudinal stance (Biber & Finegan 1988, Biber & Finegan 1989), modality (Bybee & Fleischman 1995, Palmer 1986, Portner 2009) and appraisal (Martin & White 2005), to mention just a few in each area, all aim at explaining how we use language to convey emotions, evaluation and subjectivity. Defining each of those terms could easily take up an entire paper. For the current purposes, I will refer to subjectivity as the linguistic expression of belief, emotion, evaluation, or attitude (Wiebe 1994). This in contrast to objective statements, which present events or describe the state of the world.

Research in linguistics, communication and psychology has studied how we express, understand and are affected by the expression of subjectivity (Caffi & Janney 1994, Krippendorf 2004); how we associate emotions and opinion to 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 survey I will concentrate on what has come to be called *sentiment*, the expression of subjectivity as either a positive or negative opinion. A closely related area is the study of emotion and emotive terms, in particular their classification (anger, surprise, fear, etc.). Some projects attempt at capturing both, but the research I will describe here mostly deals with sentiment. Many of the techniques and approaches, however, are applicable to the study and classification of emotions as they are expressed in language.

The purely theoretical interest in the study of subjectivity and evaluation has been accompanied, in the last few years, by an increased attention to how we express opinion online. This has opened up the field of sentiment analysis in computer science and computational linguistics, whereby 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 paper, I briefly summarize the different approaches to extracting sentiment and opinion automatically, and present the state of the art. In particular, I will discuss the aspects of sentiment analysis most relevant to linguistics, and where interaction would be beneficial. General surveys from a computational point of view are presented by Pang and Lee (2008), Liu (2012) and Sonntag and Stede (2014), whereas Feldman (2013) is a short overview for a lay audience.

2 The 'analysis' part: Computational methods

The approaches and terminology vary, but the main goal is to determine whether a text, or a part of it, is subjective or not and, if subjective, whether it expresses a positive or a negative view. The direction of the opinion (i.e., whether positive or negative) is sometimes referred to as semantic orientation. Esuli and Sebastiani (2006) define the problem as having three different aspects: (i) determining the text's subjectivity (i.e., whether the text is factual in nature or whether it expresses an opinion on its subject matter); (ii) determining the text's polarity, or deciding if a given subjective text expresses a positive or negative opinion on its subject matter; and (iii) 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 and Hovy (2004) further incorporate the source, and define opinion as a quadruple, [Topic, Holder, Claim, Statement], in which the Holder believes a Claim about the Topic, and associates a Sentiment with that belief. The sentiment may be positive or negative.

Much of the work has focused on analyzing reviews of movies, books and consumer products (Dave et al 2003, Hu & Liu 2004, Kennedy & Inkpen 2006, Turney 2002). There is also an emerging field of analysis of political discourse (Efron 2004, Mullen & Malouf 2006), including opinion pieces in newspapers. Such work can find applications in search engines: when searching for reviews of a movie, one could ask for a further classification of the reviews into positive or negative. Companies are also interested in their reputation, and that of their products. It is in their interest to track on-line discussions and evaluate whether they are positive or negative. The applications in political life and policy-making are obvious: A new form of polling, in which pollsters track on-line discussions, rather than ask questions, could emerge. Some projects also track the evolution of financial markets by following discussion on-line (Ahmad et al 2006), or investor sentiment from message boards (Das & Chen 2001). In another project, we used these methods to track literary reputation using historical reviews (Taboada et al 2006). The methods can also be applied to e-mail messages (Spertus 1997) and customer service enquiries (Gamon 2004). Most recent applications have involved forms of blogging and microblogging, such as Twitter or Facebook messages (Kiritchenko et al 2014, Mohammad et al 2013, Ortigosa et al 2014, Thelwall et al to appear, Vilares et al 2015), including the Hedonometer project, an attempt to measure happiness in 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 to decide whether it contains positive or negative sentiment, based on existing information. The crucial aspect of the task is where that information comes from. Two main approaches exist to the problem: machine learning or lexicon-based. In the machine learning approach, a classifier is built that can determine the polarity of new texts. The classifier is built thanks to labelled instances of other items (sentences, documents, etc.). This is referred to as supervised learning, because the classifier is given direction in terms of which are good or bad examples of the class. The classifier learns that certain characteristics distinguish a positive from a negative text. Those characteristics are parameters in the learning, and tend to be unigrams, that is, individual words or tokens that are present in the training dataset. The classification may be binary (positive and negative), or may include a neutral category. The advantages of the machine learning approach are that, given a labelled dataset, that is, one where documents have been previously determined to be positive or negative, training is trivial, and a classifier can be built quite quickly with existing tools, e.g., WEKA, by Witten and Frank (2005). For instance, a number of classifiers have been built using a set of 2,000 movie reviews, labelled according to whether their evaluation of the movie discussed is positive or negative (Pang et al 2002). Their performance in most cases is around 80% or above. That is, the resulting classifiers are able to correctly determine the polarity of unseen data 80% of the time (Andreevskaia & Bergler 2008, Bloom et al 2007, Dinu & Iuga 2012, Prabowo & Thelwall 2009, Socher et al 2011, Yessenalina et al 2010, among many others)

While machine learning approaches are desirable because of their accuracy, they often suffer from a number of disadvantages. First of all, because they are trained on very specific data, they are typically not portable to new types. Applying the model to new contexts and datasets will typically require new training data and thus extensive human coding. For instance, most of the classifiers built using movie review data suffer from bias towards that data, and would not be able to capture some of the nuances and 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 (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. Socher et al (2013: 1633) argue that "[f]rom a linguistic or cognitive standpoint, ignoring word order in the treatment of a semantic task is not plausible".

The other main approach to sentiment analysis is the lexicon-based, or dictionary-based method. These are often also 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 their values aggregated, using various algorithms for aggregation. 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 (in press) also provide a visual classification of the different methods for sentiment analysis.

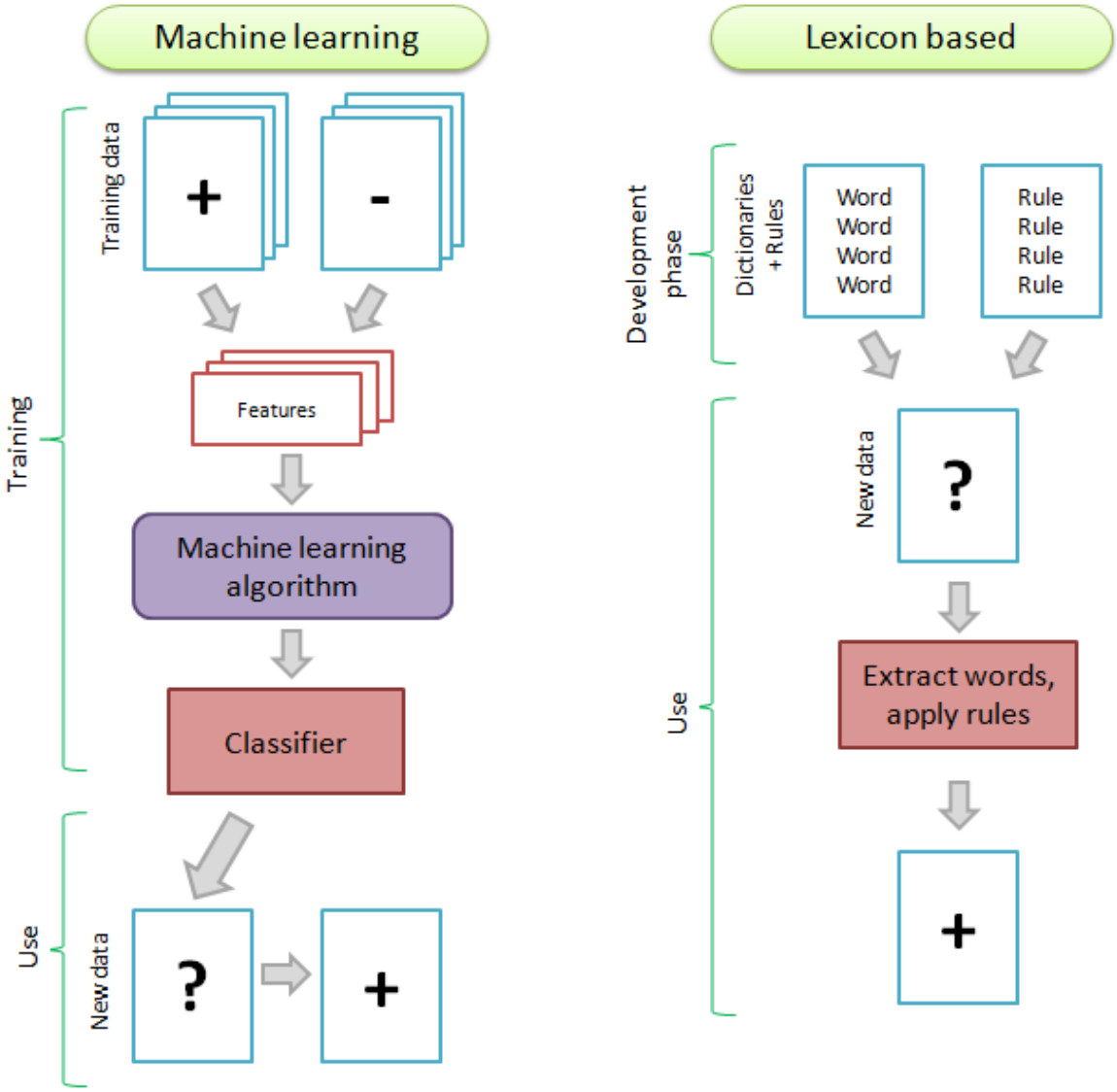

Figure 1. Machine learning and lexicon-based approaches to sentiment analysis

Lexicon-based methods have been shown to be 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, and probably one less onerous than labelling 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 will focus on these methods for the rest of the section, describing which linguistic aspects contribute to accurate extraction of sentiment.

The lexicon 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.

[Somewhere here, add the following side bar]

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, by including 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 is often indicative of strong opinion, and can be added as a feature in classification.

2.1 Which words and phrases

Most research in sentiment analysis has focused on the evaluative nature of adjectives. Adjectives convey much of the subjective content in a text, and a great deal of effort has 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*. Turney (2002), and Turney & Littman (2002, 2003) used a similar method, but this time using the Web as corpus. In their method, the adjective *X* is positive if it appears mostly in the vicinity of other positive adjectives, not only in a coordinated phrase.

Researchers have increasingly noticed, however, that a great deal of sentiment is conveyed through other parts of speech, such as nouns (*masterpiece, disaster*), verbs (*love, hate*) or adverbs (*skilfully, poorly*), and phrases that contain those words (Benamara et al 2007, Subrahmanian & Reforgiato 2008). Interesting is the exclusive use of verbs by Sokolova and Lapalme (2008), perhaps because, in some contexts, direct evaluation as expressed by positive and negative adjectives is avoided. (The authors studied consumer reviews and US Congressional debates.)

Dictionaries tend to contain lists of positive and negative words (i.e., polarity). Many lexicon-based approaches also include information about strength, i.e., how positive or negative the word is. For instance, in the subjectivity dictionary of Wiebe and colleagues (Wiebe et al 2004, Wilson et al 2009), words can fall into the following categories:

- Positive – strong absolve, accolade, altruistic
- Positive – weak accept, abundance, affluent
- Neutral accentuate, alliance, alert
- Negative – weak abolish, addiction, alienated
- Negative – strong abuse, abomination, afraid

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 been shown to be consistent with the judgements of human subjects (Taboada et al 2011). A sample list of words and their values in different dictionaries is provided in Table 1. The Subjectivity dictionary refers to the dictionary of Wiebe and colleagues from which the words above are taken. SO-CAL is the dictionary of Taboada et al (2011), whereas SentiWordNet (Baccianella et al 2010) is a set of words extracted from WordNet, with positive, negative and objective values added. SentiWordNet does have strength associated to these words, but it needs to be computed across different senses and parts of speech for the same word. For simplicity, we have simply noted polarity here. Finally, the Macquarie dictionary is a large collection of words annotated with semantic orientation by traversing the Roget's Thesaurus (Mohammad et al 2009).

	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

Table 1. Sample semantic orientation values for different dictionaries

An important question, regardless of the parts of speech being considered, is the coverage of sentiment dictionaries. It is difficult to estimate how large the evaluative lexicon of a language is. 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 of SentiWordNet (Baccianella et al 2010) or the almost 76,000 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 dozens 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 Pollyanna Principle (Boucher & Osgood 1969), whereby positive words have a higher frequency, both in terms of tokens and types, because we tend to remember past events positively. Some indicators suggest that, indeed, a great deal of online review is positive. TripAdvisor's own analysis indicates that hotel and destination reviews are largely positive (an average of 4.08 out of 5 points)[1]

The counterpart is the Negativity Bias, which postulates that negative events have a stronger effect on our psychological state and behaviour (Rozin & Royzman 2001). If a negativity bias exists, then the lower frequency of negative terms is 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 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 each other. If the presence of a negative word is more

[1] Webinar: "Climbing TripAdvisor's Popularity Index", http://resources.reviewpro.com/webinars/tripadvisor-how-to-improve-your-hotel-ranking-thanks. Retrieved April 22, 2015

indicative of a negative review, then negative words should maybe 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* vs. *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 (Carrillo de Albornoz & Plaza 2013, Kennedy & Inkpen 2006, Morsy & Rafea 2012, Taboada et al 2011).

The effect of intensification has been accounted for using 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*. Another consideration is that 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/subtraction, i.e., placing intensifiers in a percentage scale. Taboada et al (2011) propose values such as:

- *most* +100%

- *really* +25%
- *very* +15%
- *somewhat* -30%
- *arguably* -20%

Polanyi and Zaenen (2004) 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 which are not *veridical*, i.e., which are not based on truth or existence (Giannakidou 1995, Zwarts 1995). In previous work, we have defined the class of nonveridical operators as including negation (see next section), modal verbs, intensional verbs (*believe, think, want, suggest*), imperatives, questions, protasis of conditionals, habituals and the subjunctive, in languages which have an expression of subjunctive (Trnavac & Taboada 2012). Consider the effect of the intentional verb *thought* and the modal *would* in (1), and the modal plus question in (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 (Benamara et al 2012, Denis et al 2014, Morante & Sporleder 2012, Taboada et al 2011, Wilson et al 2009), but not enough research has explored exactly *how* evaluation is affected.

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. In this case, connections to sentiment research are established because 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* introducing a finding 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, it turns out that 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*, which appear in the presence of negation). The most important 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 and Moldovan (2013).

Aspects of negation that are well known to linguists are syntactic vs. morphological negation (3a vs. 3b), negation raising (3c), negation scope and partial negation of only an argument (3d and 3e). Examples, unless otherwise indicated, are taken from the SFU Review Corpus (Taboada 2008).

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

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

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

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

e. 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. (Anonymous 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*, a method where 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.

(4) a. excellent (+5) $\rightarrow$ not excellent (+1)

b. terrific (+5) $\rightarrow$ not terrific (+1)

c. sleazy (-3) → not sleazy (+1)

d. horrid (-5) → not horrid (-1)

Litotes poses a particularly interesting challenge. The 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.

An aspect of negation that is worth discussing 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 we mentioned earlier, they are less frequent. 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 approach 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).

A further 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, in most cases, 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 are acronyms or onomatopoeic expressions that indicate laughter (*LOL, he he*), heavy use of exclamation marks, or quotation marks. Irony, however, draws upon a much more varied pool of resources than just a few surface indicators and is sometimes even difficult to detect by humans (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 the contents of the book).

(5) a. Love the cover (book)

b. Where am I? (GPS device)

c. Trees died for this book? (book)

d. Be sure to save your purchase receipt (smart phone)

e. Great for insomniacs (book)

f. Defective by design (music player)

Classic work in corpus linguistics has shown that certain patterns can be used to detect irony. Louw (1993) showed that a clash in what he termed semantic prosody is indicative of irony. By semantic prosody, he 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, i.e., only bad things set in. Similarly, one is always *bent on*

pursuing negative actions. He 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 some of these linguistic insights, however, most current work on irony and sarcasm detection is restricted to detecting it 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 labelled instances to learn features that distinguish them from non-sarcastic 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. Features often used include the presence of certain words and expressions (*dare, clearly, lol, how dare, I'm shocked*), lexical density, capitalization and emoticons, and intensifiers. Bamman and 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'll be misunderstood because of lack of context. There are probably instances of sarcasm among friends or peers that exhibit different features, and would thus not be detected with the classifier resulting from such 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 detected typically implicit examples results in poor performance, probably because explicit relations do not share many features with implicit ones.

Underlying all of the work in detecting words an 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, 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 assumption is not always correct. Words take on new meanings in context that are not predictable from what Wilson, Wiebe and Hoffmann have described as *prior polarity* (Wilson et al 2005, Wilson et al 2009). Haas and 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 often conveyed through entire sentences, and particular patterns in sentences. Pattern-based descriptions of language are of special relevance here, because they avoid a distinction between lexis and grammar, but rather treat them as part of the same object of description (Hunston & Francis 2000). Subjectivity spans over the two, sometimes being conveyed by a

single word, sometimes by a phrase, and sometimes by an entire grammatical structure. Hunston and Francis define patterns of a word as "all the words and structures which are regularly associated with the word and which contribute to its meaning." (2000: 37). Here, I also include more general descriptions of grammatical structures, such as inversion.

The most in-depth description of patterns and evaluation is Hunston (2011), where a case is clearly made that certain patterns contribute to evaluative meanings, with a distinction between patterns that perform the function of evaluation, i.e., 'performative' patterns, according to Hunston (2011: 139), and patterns that report evaluation. Examples of performative patterns are *'it'* and *'there'* patterns, as in *It is amazing that…; 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 and Fretheim (2000) discuss the *I think (that)…* pattern, characterized by a verb in the matrix clause such as *think, hope, understand, 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) has discussed the need 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 are typically adjectives (expressing an evaluation of the subject) and predicate nominals (expressing a relation between subject and predicate identifiable based on subjective criteria). Scheibman (2002: 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: first person singular pronoun (*I*), present tense, modals, verbs of cognition, intensifiers and modal adverbs.

Word order often plays a role in conveying stance. Stein (1995) discusses its role in expressing subjective meanings in English. According to Stein, the examples in (6) to (8) represent a cline of emotional expression, the first one being the most subjective (Stein 1995: 132).

(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 of them lexical and some syntactic. Among the lexical clues are psychological verbs and verbs of judgement (*dread, love, commend, reprove*); verbs and adjectives that usually involve an experiencer (*fuss, worry, pleased, upset, embarrass, dislike*);

and adjectives that have been previously 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 no or little bearing on the author's opinion towards the movie under discussion. A worse case involves texts where the author discusses a completely irrelevant topic (such as the restaurant they visited before the movie). In general, this is a topic-detection problem, to which solutions have been proposed (e.g., Yang 1999 for statistical approaches).v

A slightly different problem is that of a text that contains mostly relevant information, but where some information 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. Janyce 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 description of the product or of the context that is irrelevant to the evaluation. A movie may describe the plot, and 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, within parts that contain opinion related to the movie, some may be more useful than others. It has been pointed out that 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 and Grieve (2004) improved the performance of a semantic orientation calculator by weighing more heavily the words appearing towards the end of the text. This is line with an observation by Hunston and Thompson (2000: 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 SO of the words found in the text (Turney 2002). A text with ten positive and two negative words would then be labelled as positive. This 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 have been rendered in bold, and negative evaluation is underlined (for the moment, we are mostly considering 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 is clearly a negative evaluation, but the it is presented in a style of writing that 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 the reader has 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.

In (11) below, 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 (12), where the negation of the causal relation in (12a) leads to a concessive reading in (12b).

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

b. John is not the best candidate because he happens to have a Ph.D.

The interesting aspect of this example is that the negation in (12b) does not necessarily imply a negation of the positive evaluation conveyed by *best*. It is rather a negation of the causal relation, that is, John is still the best candidate, but the reason is not that he has a Ph.D. Blanco and Moldovan (2013) refer to this phenomenon as partial negation.

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

3 A small sample of interesting projects in sentiment analysis

It is not the goal of this survey 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 will just select a few that are particularly interesting because of their approach or because of the subject matter or type of text being studied.

First of all, and with regards to text types, beyond the well-studied online reviews, many other types of texts are being analyzed in terms of their sentiment content. Politics is of course another ripe area for consideration, and early work 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 the data collection (Jungherr et al 2012). New work is being produced in this area, and current approaches make use not only of the text, but of characteristics of the author, and their online interactions (Qiu et al 2015). Much of the work on political discourse uses Twitter and online media as a source. 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. One interesting recent study by Stewart (2015) analyzes students' written comments in course evaluations, from a quantitative point of view, and using the Appraisal framework (Martin & White 2005). I am not aware of any large-scale automatic analysis of

student evaluations. Provided issues of confidentiality can be solved, this is an area that 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. One important outcome of such analysis is determining who among those who attempt suicide are likely to try again. This is properly an emotion identification task, rather than simple polarity. Emotions that were annotated, because they are considered to be good predictors, included, among others: abuse, anger, sorrow, forgiveness, love, pride, and also instructions for others.

Bobicev et al (in press) study 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, what they describe as sentiments: encouragement, gratitude, confusion, facts and endorsement. Using the corpus as training data they build 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 into positive and negative polarity of the messages would not pride enough fine-grained information for their purposes, which include extracting sentiment from discussions on healthcare policy.

As an aside, some of these projects study emotion rather than sentiment (polarity). The study of emotions in general, and their automatic identification, probably merits another survey. In this paper, I merely point out areas where there is an overlap with sentiment analysis.

The vast majority of the 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 is being carried out, however, in 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 image, 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 a large number of positive reviews, or negative reviews about a rival's business. The practice has led to court cases and settlements, with companies being found guilty of paying for positive reviews, or of writing them themselves (Streitfeld 2013). TripAdvisor was recently fined €500,000 for failing to prevent fake reviews on their site (Scott 2013). Fake review detection 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 allows systems to determine if the same review is being posted in different sites, and if certain stock phrases are being used repeatedly. Bing Liu has been a leader in this field, and included a chapter on how to detect fake reviews in his survey (Liu 2012).

4 Sentiment analysis in languages other than English

English is, without a doubt, the main object of study in sentiment analysis. English is not, however, the only language in which opinions are expressed online. There are, accordingly,

efforts 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 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 labelled 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. Or 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 in press), French (Benamara et al 2013, Ghorbel 2012, Marchand 2012), German (Clematide & Klenner 2010, Haas & Versley 2015, Waltinger 2010), Spanish (López et al 2012, Molina-González et al 2013, Moreno-Ortiz & Pérez Hernández 2012, Vilares et al 2013, Vilares et al 2015). In some cases, the focus is a combination of different languages (Banea et al 2014, Banea et al 2008, Mihalcea et al 2007, Popat et al 2013).

5 The future ahead

Literature on sentiment analysis seems to be multiplying at alarming rates, 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. If one can use the translations obtained through the Google service, then they are good enough. If, on the other hand, 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, the systems being developed are in-house and not available to the public. But what if you want to test sentiment analysis for yourself? Serrano-Guerrero et al (in press) list 15 different web services that allow textual input and output various types of sentiment information. The Stanford Deep Learning Model (Socher et al 2013)[2] allows free text input, and also user input on values already provided by the system.

References

Ahmad K, Gillam L, Cheng D. 2006. Sentiments on a grid: Analysis of streaming news and views. In *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC)*, pp. 2517-20. Genoa, Italy

Aikhenvald A. 2004. *Evidentiality*. Oxford: Oxford University Press

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

Andreevskaia A, Bergler S. 2008. When specialists and generalists work together: Domain dependence in sentiment tagging. In *Proceedings of 46th Annual Meeting of the Association for Computational Linguistics*, pp. 290-98. Columbus, OH

[2] http://nlp.stanford.edu/sentiment/

Anonymous. 2013. How does the shutdown relate to me? In *The Last Psychiatrist*, pp. Blog post

Baccianella S, Esuli A, Sebastiani F. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pp. 2200-04. Valletta, Malta

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 in Social Media (LASM 2013)*, pp. 49-58. Atlanta, GA

Bamman D, Smith NA. 2015. Contextualized sarcasm detection on Twitter. In *Proceedings of the 9th International Conference on Web and Social Media*. Oxford, UK

Banea C, Mihalcea R, Wiebe J. 2014. Sense-level subjectivity in a multilingual setting. *Computer Speech and Language* 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*, pp. 127-35. Honolulu

Banfield A. 1982. *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Boston: Routledge and 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 International Conference on Weblogs and Social Media, ICWSM*. Boulder, CO

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. Jeju, Korea

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: Brill

Biber D, Finegan E. 1988. Adverbial stance types in English. *Discourse Processes* 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. *Natural Language Engineering* 20: 501-35

Bloom K, Garg N, Argamon S. 2007. Extracting appraisal expressions. In *Proceedings of HLT/NAACL*, pp. 308-15. Rochester, NY

Bobicev V, Sokolova M, Oakes M. in press. What goes around comes around: Learning sentiments in online medical forums. *Cognitive Computation*

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. Barcelona, Spain

Boucher JD, Osgood CE. 1969. The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behaviour* 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*, pp. 50-54. Borovets, Bulgaria

Bybee J, Fleischman S, eds. 1995. *Modality in Grammar and Discourse*. Amsterdam and Philadelphia: John Benjamins

Caffi C, Janney RW. 1994. Towards a pragmatics of emotive communication. *Journal of Pragmatics* 22: 325-73

Carrillo de Albornoz J, Plaza L. 2013. An emotion-based model of negation, intensifiers, and modality for polarity and intensity classification. *Journal of the American Society for Information Science and Technology* 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 First International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement*. Hong Kong

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 *Proceeding of the 14th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, pp. 25-37. Samos, Greece

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)*. Lisbon, Portugal

Das SR, Chen MY. 2001. Yahoo! for Amazon: Opinion extraction from small talk on the web. In *Proceedings of 8th Asia Pacific Finance Association Annual Conference*. Bangkok, Thailand

Dave K, Lawrence S, Pennock DM. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the Twelfth International World Wide Web Conference (WWW 2003)*. Budapest, Hungary

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. 605-09. Dublin, Ireland

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. *Proceedings of the National Academy of Sciences* 112: 2389-94

Durant KT, Smith MD. 2006. Mining sentiment classification from political web logs. In *Proceedings of Workshop on Web Mining and Web Usage Analysis of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Philadelphia, PA

Efron M. 2004. Cultural orientation: Classifying subjective documents by cociation analysis. In *Proceedings of AAAI Fall Symposium on Style and Meaning in Language, Art, and Music*. Washington, D.C.

El-Beltagy SR, Ali A. 2013. Open issues in the sentiment analysis of Arabic social media: A case study. In *Proceedings of 9th International Conference on Innovations in Information Technology*. Al Ain, UAE

Esuli A, Sebastiani F. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC)*, pp. 417-22. Genoa, Italy

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 Fourteenth Conference on Computational Natural Language Learning: Shared Task*, pp. 1-12. Uppsala, Sweden

Feldman R. 2013. Techniques and applications for sentiment analysis: The main applications and challenges of one of the hottest research areas in computer science. *Communications of the 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. Jeju, Korea

Feng VW. 2015. *RST-Style Discourse Parsing and its Applications in Discourse Analysis*. Ph.D. dissertation thesis. University of Toronto, Toronto

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*. Baltimore, MA

Gamon M. 2004. Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis. In *Proceedings of COLING 2004*, pp. 841-47. Geneva, Switzerland

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: University of 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 Conference of the North American Chapter of the Association for Computational Linguistics*. Denver, CO

Hatzivassiloglou V, McKeown K. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of 35th Meeting of the Association for Computational Linguistics*, pp. 174-81. Madrid, Spain

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

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

Hu M, Liu B. 2004. Mining and summarizing customer reviews. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004)*. Seattle, WA

Huang T-H, Yu H-C, Chen H-H. 2012. Modeling Polyanna phenomena in Chinese sentiment analysis. In *Proceedings of COLING 2012: Demonstration papersi*, pp. 231-38. Mumbai, India

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 and Philadelphia: John 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: Oxford University Press

Jing-Schmidt Z. 2007. Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics* 18: 417-43

Joty S, Carenini G, Ng R. 2015. CODRA: A novel discriminative framework for rhetorical analysis. *Computational Linguistics*: in press

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". *Social Science Computer Review* 30: 229-34

Kennedy A, Inkpen D. 2006. Sentiment classification of movie and product reviews using contextual valence shifters. *Computational Intelligence* 22: 110-25

Kim S-M, Hovy E. 2004. Determining the sentiment of opinions. In *Proceedings of COLING 2004*, pp. 1367-73. Geneva, Switzerland

Kiritchenko S, Zhu X, Mohammad S. 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research* 50: 723-62

Krippendorf K. 2004. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage

Langacker RW. 1990. Subjectification. *Cognitive Linguistics* 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 IEEE International Conference on Data Mining (ICDM-2014)*. Shenzhen, China

Liu B. 2012. *Sentiment Analysis and Opinion Mining*: Morgan & Claypool

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*, pp. 161-69. Singapore

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: 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 2012*, pp. 177-90. Grenoble, France

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. Prague, Czech Republic

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

Mohammad S, Kiritchenko S, Zhu X. 2013. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*. Atlanta, GA

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

Morante R, Sporleder C. 2012. Special Issue on Modality and Negation. *Computational Linguistics* 38

Moreno-Ortiz A, Pérez Hernández L. 2012. Lexicon-based sentiment analysis of twitter messages in Spanish. In *TASS, Taller de Análisis de Sentimientos en la SEPLN (Sociedad Española para el Procesamiento del Lenguaje Natural)*. Castellón de la Plana, Spain

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, The Netherlands: 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. Stanford, CA

Ortigosa A, Martín JM, Carro RM. 2014. Sentiment analysis in Facebook and its applications to e-learning. *Computers in Human Behavior* 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. Portland, OR

Palmer F. 1986. *Mood and Modality*. Cambridge: Cambridge University Press

Pang B, Lee L. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* 2: 1-135

Pang B, Lee L, Vaithyanathan S. 2002. Thumbs up? Sentiment classification using Machine Learning techniques. In *Proceedings of Conference on Empirical Methods in NLP*, pp. 79-86

Pestian JP, Matykiewicz P, Linn-Gust M, South B, Uzuner O, et al. 2012. Sentiment analysis of suicide notes: A shared task. *Biomedical Informatics 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: 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. Sofia, Bulgaria

Portner P. 2009. *Modality*. Oxford: Oxford University Press

Prabowo R, Thelwall M. 2009. Sentiment analysis: A combined approach. *Journal of Informetrics* 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. In *Proceedings of the SIAM International Conference on Data Mining*. Vancouver, Canada

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

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)*, pp. 25-32. Edmonton, Canada

Rozin P, Royzman EB. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* 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 North American Chapter of the Association for Computational Linguistics (NAACL-2015)*. Denver, CO

Saurí R. 2008. *A Factuality Profiler for Eventualities in Text*. Ph.D. dissertation thesis. Brandeis University, Waltham

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

Scott M. 2013. TripAdvisor fined $610,000 in Italy for failing to prevent fake reviews. In *The New York Times*, pp. B8. New York

Serrano-Guerrero J, Olivas JA, Romero FP, Herrera-Viedma E. in press. Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*

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

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 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*. Melbourne, Australia

Sokolova M, Lapalme G. 2008. Verbs speak loud: Verb categories in learning polarity and strength of opinions. In *Canadian AI 2008*, ed. S Bergler, pp. 320-31. Berlin: Springer

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

Spertus E. 1997. Smokey: Automatic recognition of hostile messages. In *Proceedings of AAAI/IAAI '97*, pp. 1058-65. Providence, RI

Sporleder C, Lascarides A. 2008. Using automatically labelled examples to classify rhetorical relations: An assesment. *Natural Language Engineering* 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: Cambridge University Press

Stewart M. 2015. The language of praise and criticism in a student evaluation survey. *Studies in Educational Evaluation* 45: 1-9

Streitfeld D. 2013. Give yourself 5 stars? Online, it might cost you. In *The New York Times*, pp. A1. New York

Subrahmanian V, Reforgiato D. 2008. AVA: Adjective-verb-adverb combinations for sentiment analysis. *Intelligent Systems* 23: 43-50

Taboada M. 2008. SFU Review Corpus. Vancouver: Simon Fraser University, 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*, pp. 62-70. London, UK

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

Taboada M, Gillies MA, McFetridge P. 2006. Sentiment classification techniques for tracking literary reputation. In *Proceedings of LREC Workshop, "Towards Computational Models of Literary Analysis"*, pp. 36-43. Genoa, Italy

Taboada M, Gómez-González MdlÁ. 2012. Discourse markers and coherence relations: Comparison across markers, languages and modalities. *Linguistics and the Human Sciences* 6: 17-41

Taboada M, Grieve J. 2004. Analyzing appraisal automatically. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07)*, ed. Y Qu, JG Shanahan, J Wiebe, pp. 158-61. Stanford University, CA: AAAI Press

Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. to appear. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*

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*, pp. 327-35. Sydney, Australia

Thompson SA. 2002. “Object complements” and conversation: Towards a realistic account. *Studies in Language* 26: 125-64

Traugott EC. 1995. Subjectification in grammaticalisation. In *Subjectivity and Subjectivisation: Linguistic Perspectives*, ed. D Stein, S Wright, pp. 31-54. Cambridge: Cambridge University 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 Mouton

Trnavac R, Taboada M. 2012. The contribution of nonveridical rhetorical relations to evaluation in discourse. *Language Sciences* 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*, pp. 162-69. Washington, DC

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*, pp. 178-85

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

Turney P, Littman M. 2002. *Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Rep. ERB-1094, NRC 44929*, National Research Council of Canada

Turney P, Littman M. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems* 21: 315-46

Utsumi A. 2000. Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony. *Journal of Pragmatics* 32: 1777-806

Velldall E, Ovrelid L, Read J, Oepen S. 2012. Speculation and negation: Rules, rankers, and the role of syntax. *Computational Linguistics* 38: 369-410

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

Vilares D, Alonso MA, Gómez-Rodríguez C. 2013. A syntactic approach for opinion mining on Spanish reviews. *Natural Language Engineering*: in press

Vilares D, Alonso MA, Gómez-Rodríguez C. 2015. On the usefulness of lexical and syntactic processing in polarity classification of Twitter messages. *Journal of the Association for Information Science and Technology*: in press

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

Waltinger U. 2010. GermanPolarityClues: A lexical resource for German sentiment analysis. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta, Malta

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*, pp. 553-61. Honolulu

Wang F, Wu Y, Qiu L. 2012. Exploiting discourse relations for sentiment analysis. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING), Posters*, pp. 1311-20. Mumbai, India

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*. Oxford, UK

Wiebe J. 1994. Tracking point of view in narrative. *Computational Linguistics* 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. Stanford, CA

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

Wiebe J, Wilson T, Bruce R, Bell M, Martin M. 2004. Learning subjective language. *Computational Linguistics* 30: 277-308

Wiebe J, Wilson T, Cardie C. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 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. Vancouver, Canada

Wilson T, Wiebe J, Hoffmann P. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics* 35: 399-433

Wilson T, Wiebe J, Hwa R. 2006. Recognizing strong and weak opinion clauses. *Computational Intelligence* 22: 73-99

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

Yang Y. 1999. An evaluation of statistical approaches to text classification. *Journal of Information Retrieval* 1: 67-88

Yessenalina A, Choi Y, Cardie C. 2010. Automatically generating annotator rationales to improve sentiment classification. In *Proceedings of the ACL 2010 Conference Short Papers*, pp. 336-41. Uppsala, Sweden

Zaenen A, Polanyi L. 2004. Contextual valence shifters. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07)*, ed. Y Qu, JG Shanahan, J Wiebe, pp. 106-11. Stanford University, CA: AAAI Press

Ziyan C, Huang Y, Tian J, Liu X, Fu K, Huang T. in press. Joint model for subsentence-level sentiment analysis with Markov logic. *Journal of the Association for Information Science and Technology*

Zwarts F. 1995. Nonveridical contexts. *Linguistic Analysis* 25: 286-312