Sentiment Analysis: What's your Opinion?

Jonathan Sonntag and Manfred Stede

Abstract For more than ten years now, sentiment analysis has enjoyed enormous popularity in Computational Linguistics, one main reason being its great potential for practical applications, predominantly (but not only) for industrial purposes. We observe a tendency that early work referred to certain theoretical notions of subjectivity, whereas a lot of the later approaches follow the 'engineering' perspective that can include using terminology somewhat indiscriminately and are not aiming at making progress with the underlying theoretical issues. In this paper, we first survey some important notions surrounding "subjectivity" in linguistics and psychology, trying to broaden the perspective of standard opinion analysis. Thereafter, we take a snapshot of the state of the art in computational sentiment analysis, as it has developed since roughly 2000. Combining these two viewpoints leads us to assessing the gap between the broader notion of subjectivity analysis and the subfields that language technology research tends to focus on. We suggest a few potential research directions that could help narrowing this gap.

1 Introduction

Sentiment Analysis has become popular over the last 15 years, due to various reasons: *a*) the rise of social media; *b*) the technological developments; especially the possibilities and problems of "big data" and, lastly, *c*) the progress of natural language processing tools, which lead to a shift of attention towards more complicated

Jonathan Sonntag
Universität Potsdam, Department Linguistik, Haus 14, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, e-mail: jonathan.sonntag@yahoo.de

Manfred Stede
Universität Potsdam, Department Linguistik, Haus 14, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, e-mail: stede@uni-potsdam.de

and thus more semantic/pragmatic problems, such as Sentiment Analysis, Question Answering, or Textual Entailment.

Mining product reviews is one of the most promising NLP problems for industrial uses today. The grand goal of being able to automatically detect customer feedback in large quantities would help merchandisers and manufacturers in developing specialised marketing campaigns tailored to the standing a product has among its customers, and also contribute to improve the products.

Besides the industrial application, Sentiment Analysis is now also attracting attention in the Social Sciences. For example, some researchers investigate the ways of conveying opinion in parliamentary debates; others are interested in automatically gathering standpoints on political issues from newspapers or from social media.

Sentiment Analysis is only one, albeit an important, element of a battery of Text Mining tools necessary to extract the relevant information from large amounts of arbitrary text. Focusing again on the commercial application, the issues are: who buys your products? What other products do your customers buy? How do they get attracted to your products? How do they learn about your products? How often do they use them? Which attributes or *aspects* of the products are evaluated? The role of Sentiment Analysis within the larger Text Mining task in industrial uses is twofold: $a)$ to detect polar statements, which can be interpreted as relations between entities,[1] and $b)$ to provide a clear distinction between sentences conveying or revealing a sentiment on the one hand, and objective statements on the other. Traditional Question Answering systems, for example, are designed to extract not opinions and evaluations, but *facts* from a text. Consider the query in Example 1 and the two possible answers, which a QA System might find in a large corpus.

(1) Who was the 15th president of the United States of America?

(2) a. James Buchanan was the 15th, and a horrible, president of the U.S.
    b. James Buchanan was the 15th president of the U.S.

Here, the QA system should prefer the pure 'factual' statement in 2b to the subjective one in 2a.

The term 'Sentiment Analysis' is used in different ways in the literature. Even attempts at clarifying the term seem to ignore the underlying problem: [3], for instance, suggests that the terms 'Sentiment Analysis' and 'Opinion Mining' can be used interchangeably. Yet, the author bases this assessment on [24, p. 10], who say:

> A sizable number of papers mentioning "Sentiment Analysis" focus on the specific application of classifying reviews as to their polarity, a fact that appears to have caused some authors to suggest that the phrase refers specifically to this narrowly defined task. However, nowadays many construe the term more broadly to mean the computational treatment of opinion, sentiment, and subjectivity in text.

---

[1] We provide just a minimal property of sentiment at this point, which goes beyond coarse-grained Sentiment Analysis, but it is deliberately a rather abstract description. We come back to this issue in Section 4.

Yet, consumer reviews are not restricted to opinions; they also contain factual, yet polar, statements – see the discussion below in Section 4. Therefore, the two terms cannot mean the same.

In this chapter, we use 'Sentiment Analysis' in a broad sense that subsumes opinions, evaluations, emotions, judgements, polar facts, and other kinds of subjective utterances.

The rest of this chapter is organised as follows: in Linguistics (and related disciplines), the term 'Sentiment Analysis' is highly uncommon. Instead, 'Subjectivity' is a general notion being studied, inter alia, in Linguistics. In Section 2, we describe various facets of this concept and how they relate to each other. Then, Section 3 provides a brief survey of work in computational Sentiment Analysis, before in Section 4 we give our personal opinion on how certain notions should be defined, and suggest some directions for future work in computational analysis, partly inspired by the insights provided by the theoretical disciplines. Finally, Section 5 summarizes the chapter.

2 The Counterpart of *Sentiment* in Linguistics and Psychology

Rather than aiming at a "grand overview", in this section we will take a look at a number of prominent aspects of subjectivity, as they are being studied in Linguistics and Psychology. The selection is largely motivated by our judgement of relevance to Computational Linguistics. In the second part, we briefly touch upon the relation between subjectivity and its dual notion: objectivity.

2.1 Subjectivity

> The term Subjectivity refers to the way in which natural languages, in their structure and their normal manner of operation, provide for the locutionary agents expression of himself and his own attitudes and beliefs. (Lyons 1982: 102)

This well-known definition reminds us of the fact that human language fulfills a variety of purposes, and it also goes some way in suggesting a particular classification of those purposes. An earlier proposal in this vein had been made by Bühler (1934), who assessed that language has three different functions:

- Representation *(Darstellung)*: the speaker describes a state of affairs in the world.
- Expression *(Ausdruck)*: the speaker conveys his or her own feelings or state of mind.
- Appeal *(Appell)*: the speaker wants the addressee to change their mind, or to act in a certain way.

Expression corresponds quite closely to the main point of Lyons's definition, which can be called the "internal" view of Subjectivity; it will be discussed below in Sections 2.1.1 and 2.1.2. Bühler's *appeal* function, on the other hand, points to an additional role: communication involves multiple partners, and aligning with them is a part of an interlocutor's linguistic behavior. This aspect is nowadays sometimes called "Inter-Subjectivity", and we will address it in Section 2.1.3.

2.1.1 The 'Private State'

If the "objective" is in principle observable to everybody, then a reasonable reading of "subjective" is that of a particular agent's "inner world", which is not observable to anyone except for that agent him- or herself. [27] used the term *private state* for this and illustrate the idea with this example (p. 1181; emphasis by those authors): "A person may be observed to *assert that God exists*, but not to *believe that God exists*. Belief is in this sense 'private'." Then, agents may choose to communicate (certain aspects of) their private states, and at this point, the linguistic interest sets in: What are the linguistic means for verbalizing different aspects of a private state? In order to study this, the fairly general concept of 'private state' needs to be broken up into a number of distinct, simpler realms.

One proposal to this end can be found in 'Appraisal Theory' stemming from systemic-functional linguistics (SFL), and proposed by Martin and White [18]. The overall goal of the SFL approach to language is to delineate and taxonomise the semantic and pragmatic dimensions that are assumed to be responsible for the spectrum of syntactic variety within a language. The portion of this endeavor that is relevant for our purposes is the following sub-taxonomy:

- ATTITUDE encompasses different options for expressing positive or negative evaluation
  - AFFECT: emotional evaluation of things, processes or states of affairs; main subclasses: un/happiness, in/security, dis/satisfaction
  - JUDGEMENT: ethical evaluation of human behavior (e.g., good/bad)
  - APPRECIATION: aesthetic or functional evaluation of things, processes and states of affairs (e.g., beautiful/ugly, useful/useless)
- ENGAGEMENT addresses options for expanding and contracting space for other voices (i.e. how much does the writer endorse the statements of others)
- GRADUATION: adjustments of attitude and engagement in terms of strength

For our purposes, the central part of the taxonomy is ATTITUDE with its three daughter nodes, which all revolve around a speaker *evaluating* something. The first way of doing this is by expressing an emotion or affect; since this has been studied extensively in Psychology, we discuss it in somewhat more detail in Section 2.1.2. For the non-emotional evaluations, Martin and White distinguish two classes of targets of the evaluation: ethical judgment of human behavior versus aesthetic/functional evaluation of "things" in a wide sense. – It is this particular notion that is at the heart of the vast majority of computational work on opinion mining.

The third term in the Martin/White sub-taxonomy is GRADUATION, which refers to the linguistic means of marking a strengthening or down-toning of a subjective utterance. Which of these means are appropriate depends on the particular type of utterance. Some distinctions to be made among so-called *epistemic stances* are:

- marking the degree of precision or truth or appropriateness of a category label
  She is almost a PhD now.
- marking the probability of truth
  Most likely she is a PhD by now.
- marking an expectation on the probability of a statement
  She ought to be a PhD soon.

2.1.2 Emotions and their Reflection in Language

Emotions are quite clearly distinct from non-emotions,[2] and the automatic identification and displaying of emotions has become a research discipline in its own right, also with close ties to Computational Linguistics (CL) [28].

A major foundation of emotion research in Psychology is the work of [52]. Their theory has become influential also in CL, under the term "OCC Model of Emotions" [33]. Here, the realm of emotion is split into two different concepts: *arousal* and *appraisal*[3]. Arousal is described as the "hot" part of emotion and has a similar meaning to the term "stress" in everyday language. It involves all bio-chemical reactions to stimuli. For computational linguistics, the second term is more interesting. Appraisal is the cold, calculating side of emotions; it can be characterised as reactions to three types of entities: objects, agents, and events. (i) Objects are being evaluated in terms of their *appealingness*: How well does the object fit to a person's attitudes? If one's favourite colour is 'green', then a green bicycle is perceived more favourably than a pink bicycle. (ii) Reactions to the actions of another agent are evaluated in terms of *praiseworthiness*, which refers to normative expectations on how an agent should act in a certain situation. (iii) Finally, events with respect to their *desirability*, i.e. lead to positive emotions if they support the agents goals.

All reactions that are relevant to appraisal share one basic feature: they are *valenced reactions* – we respond either positively or negatively to the stimulus. We discuss this commonality and possible implications for Sentiment Analysis in Section 4.

In addition to the different types of appraisal, Ortony et al. describe *intensity* as a major factor: the strength with which a person experiences the emotion. Intensity variables can be either global, i.e., relevant for every appraisal type, or local, i.e., dependent on the appraisal type at hand. An interesting variable affecting the intensity is the *sense of reality*. The authors hypothesise that emotion-inducing situations

[2] Ortony et. al. [52, pp. 29 – 32] give examples for non-emotional events giving rise to emotions. Yet, the distinction is clear.

[3] Notice the identity to the linguistic term discussed in the previous section (Martin/White). There seems to be no direct (established) connection between the two approaches.

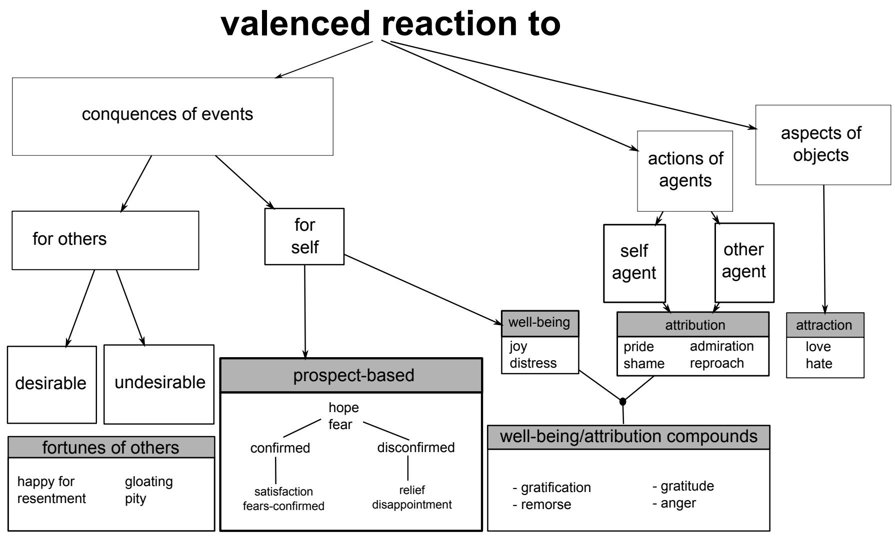

Fig. 1 Emotion schema described in [52, p. 69]

have a certain temporal, locational and psychological proximity: distant situations induce the same emotions as near situations, but the intensity changes. Another factor worth mentioning is *unexpectedness*: situations that have a high probability of happening are deemed less intense than surprising ones.

Another interesting proposal by Ortony et al. is the *balance principle* for valenced reactions. It originates from [11] and describes how relations between people evolve when they form a triangle, i.e., when three people interact with each other (*triadic relationships*). Balance is achieved if and only if the product of the edge weights is positive. Applied to text analysis, this means: if all valenced reactions are combined into a weighted graph in which the nodes represent discourse entities and the edges the intensity of the reactions between them (say, -1 to 0 for negative reactions, and 0 to 1 for positive reactions), then for each triangle in the graph, the product of their edge weights tends to be positive.

Contrary to the desire of Computational Linguists to have a clearly defined inventory of emotions, Ortony et al. refrain from defining such an inventory, due to the disagreement on the term *basic emotions* in Psychology. The only possibly 'minimal' emotions those authors would agree with are positive and negative emotions, as already proposed by [34]. Another, rather pessimistic claim for the engineering task is the authors' diagnosis that the words of English are underspecified with respect to their emotion type. This is an issue that has been studied under the heading of lexical *connotation*: the idea is that words have a 'kernel meaning', essentially the real-world entity that they stand for (denotation), plus additional traits of meaning, the so-called connotations. The challenge is to produce a list of connotative dimensions that can be productively used to differentiate between words that have the same denotation. The best-known such dimension is *formality* (e.g., motion picture vs. movie); others include *pejorative* (e.g., man, jerk) and *euphemism* (e.g., genocide, ethnic cleansing).

2.1.3 Intersubjectivity

In some situations, a speaker may convey a private state just for herself, as by uttering "ouch!" or "phew!". Typically, however, communication is directed toward some addressee, which in our context leads to the notion of "Intersubjectivity". One aspect of this notion, which is particularly relevant for Sentiment Analysis, is the question to what extent a speaker assumes responsibility for her statement: am I stating my own conviction, or am I attributing the responsibility to somebody else? Quoted speech is the clearest case here: its boundaries are unambiguously marked. For indirect speech, this need not be the case. If it stretches over more than one sentence, it can be ambiguous whether some material is still attributed to a source cited earlier, or whether the speaker has resumed to stating his own position. The linguistic notion at stake here is *evidentiality*, which refers to the variety of means that languages offer for marking this relationship between statement and alleged responsibility. For an overview on the linguistic discussion, see [8]. For some languages, this marking is obligatory by means of grammatical categories; but for English or German, speakers have the choice of choosing lexical expressions to mark Evidentiality, in particular via modal verbs:

(3) Es wird morgen regnen.
It will tomorrow rain.
'It will be raining tomorrow.'

a. Es soll morgen regnen.
It should tomorrow rain.
'It is said to be raining tomorrow.'

There is thus a continuum between explicitly stating the source of a statement and clearly marking the boundaries of that statement on the one hand, and vaguely hinting at "some" external source. In fact, one special case of an "external" source or viewpoint can be the speaker's own perception, as in:

(4) It seems to be raining.

This can be paraphrased as "My sensory organs indicate that it is raining, but I don't fully commit to the truth of the statement." The speaker thus puts some distance between himself and the statement, and we can see that there is a fuzzy boundary between the realms of evidentiality and what we have discussed above as 'epistemic stance', in particular the marking of reliability of information.

The term 'intersubjectivity' has many more facets, and here we want to just briefly mention the work of cognitively-oriented linguists such as Langacker [17] or

Verhagen [41]. In contrast to the more standard linguistic analysis "pipeline" (syntax followed by semantics followed by pragmatics/context), they emphasize that basically any linguistic utterance should be seen *foremost* as being directed to an addressee and as managing the relation between the interlocutors. In Langacker's theory, an utterance has to be analyzed in tandem in terms of the interlocutor relationship and the real-word states of affairs that is being talked about. In a similar vein, Verhagen elaborates on the idea of Anscombe and Ducrot [1], who posit that language use is essentially always 'argumentative' in the sense that a speaker by making an utterance intends to influence the mental state of the addressee. These (and other) authors demonstrate with many examples that linguistic constructions are sensitive to the 'argumentative orientation' of individual statements and, hence, that subjectivity is deeply built into the linguistic system.

2.2 Factuality

We use the term *factuality* for the linguistic marking whether a certain event happened or an object exists. Factuality is relevant to Sentiment Analysis because it can contribute to the decision whether a sentence is understood as conveying a sentiment.

2.2.1 The Semantic Viewpoint: Evidentiality and Veridicity

Linguistic semantics is interested in modality in general, and the marking of *evidentiality* is an important subgroup here (cf. Section 2.1.3). Coming from the CL perspective, [13] use the term *Vericidity*, which is to deal with these questions on a certain event:

1. has the event really occured?
2. who said that the event occured?
3. does the author believe the event occured?
4. how does the author of the text refer to it?

The first question is the central, practical question. Questions 2 to 4 are part of the first question and are observable in text, while the answer to the first one is not.

Sentences 5a and 5b demonstrate how different reporting verbs can convery different stances of the author. While the author does not take any stance in Sentence 5a, he supports Bush's claim in Sentence 5b.

(5) a. Bush *said* that Iraq had aided al Qaida.
b. Bush *acknowledged* that Iraq had aided al Qaida.

Karttunen et. al. [13] embed their research within the Advanced Question Answering for Intelligence (AQUAINT) project. In Section 1 we pointed out why knowledge about subjective versus objective statements is important for QA systems.

2.2.2 Interpretation

Before a Sentiment evolves in a human being as a reaction to some real world event, the event has to be interpreted. This interpretation can be straightforward or involve some further inference. Example 6 typically creates some negativity inside a reader: a reasonable interpretation is that 'Carlo' wants[4] to do bad things to the people inside the cafe. Then, a complex interaction evolves: the reader probably develops some sympathy for the people inside the cafe, since, for all he knows, they were ordinary people just like the reader himself. Additionally, Carlo obviously wants to harm the people inside the cafe. Then, the reader reacts to the negativity of Carlo towards people like himself and the last sentiment relation emerges: the reader dislikes Carlo. This interpretation also coincides with the balance principle described above in Section 2.1.2.

(6) Carlo threw a hand-granade inside the cafe.

This example shows that facts are an integral part of emotion detection and thus also of sentiment analysis. Not every interpretation of the real world is as straightforward as the example above, though. As readers of news in political discourse we rely on the activity of the media providing interpretations for us.

In a newspaper article, Zastrow [50] criticises the role of the media in political discourse. Analyzing the media coverage of the 2013 elections in the German state of Lower Saxony, he describes how the analysis and reflection in the media changes significantly due to small changes in the numerical results of the elections. Commenting on the interpretation of the results of the election, he says:

> And, thus, just about everything that has been said was turned into the opposite a little later.

Further debating the interpretations of the media, Zastrow wonders:

> What is going on there? Nothing special, it is simply the good old manipulation. The analyses only pretend to be analyses. In fact, they are political demands masked as objective analysis.

Finally, the author makes a very strong point about the usage of facts in argumentation:

> There is no bigger success in a political debate than to convince a majority that your opinion, evaluations or demands are facts.

These quotations not only deal with the interpretation of events but also with their veridicity. A possible reason is that assessing the veridicity of facts is part of constructing the mental representation of real-world affairs.

---

[4] Note that the use of the term 'want' suggests Subjectivity.

3 Sentiment Analysis in Computational Linguistics

We now turn to describing the major developments of sentiment analysis within the "engineering" part of Computational Linguistics. For the reader interested in a more extensive introduction, Pang et. al. [24] provide an overview over early work in the field, with some tools having been developed as early as 1979 (the POLITICS software for sentiment analysis on political text [6]).

Much work in early sentiment analysis focuses on the assignment of polarity values on the level of words. We touch upon this in Section 3.1, in which we briefly discuss different lexicons and corpora, which form the basis for any sentiment system. Afterwards, Section 3.2 discusses rule-based analysis systems, and then Sections 3.3 and 3.4 talk about *aspect analysis* and machine learning approaches, respectively. The two are closely related because most fine-grained sentiment approaches relying on machine learning are aspect analysis systems.

3.1 Resources: Lexicons and Corpora

For all approaches to Sentiment Analysis, annotated corpora are required. Their minimal usage is the evaluation of automatic systems, but of course, for machine-learning approaches, corpora are also essential as training data.

Of the various corpora that have been built and annotated with sentiment, we mention only a few. The MPQA corpus [42] is one of the most prominent for English, while the MLSA corpus [7] is the first publicly available resource for German.

The MPQA corpus consists of Chinese newspaper articles translated into English and articles from U.S. newspapers. They have been annotated for 'subjective frames', which are based on the notion of 'private state' as introduced in Section 2.1.1. Very briefly, a frame consists of the opinion holder, opinion target, and the expression of the sentiment. The original corpus consists of 535 documents, corresponding to 11.114 sentences.

The MLSA corpus is a fine-grained corpus based on the DeWaC Corpus [5]. It is annotated with three different layers: a) at the sentence-level, objective/subjective[5] and positive/neutral/negative are specified, b) polarity and modifiers are annotated at the phrase-level and c) private states are annotated at the expression-level similar to the annotation of the MPQA corpus.

A recent corpus of amazon.com reviews, the USAGE corpus for aspect analysis [16], consists of 800 German and 800 English reviews and is annotated for aspects and evaluative expressions.

Apart from the three corpora mentioned above, various others were assembled from reviews. Since many review sites provide textual comments as well as a nu-

---

[5] In this case, the authors intention is to specify factuality, which relates to 'Evidentiality' and 'Veridicity' (cf. Section 2.2.1)

merical or star-based rating, information on sentiment in the text can be inferred straightforwardly assuming minimal simplifications [20, 21, 22].

One type of information that is required for virtually every solution to sentiment analysis is the so-called *prior polarity* of words. Simply put, *bad* is inherently negative, while *joy* is inherently positive. Due to its importance, it makes sense to factor polarity information out of the individual approach and represent it externally in a lexicon – so that the result of the expensive production of such a lexicon can be re-used and shared with other researchers. Even machine-learning approaches to Sentiment Analysis typically use a lexicon containing information about the polarity of words. Often the polarity is supplied not as just a binary distinction but in terms of intensity values (either numerical or nominal).

Table 1 Example entries from a subjectivity lexicon [47]. Notice that some entries seem to be rather objective but polar, and that some entries seem to be very underspecified with respect to their polarity and strength. The authors explicitly mention that the entries in their lexicon *can* have subjective meanings.

word	prior-polarity	reliability
abhorrent	negative	strong
absence	negative	weak
obsolete	negative	weak
bankrupt	negative	weak
lack	negative	strong
obstacle	negative	strong
odd	negative	weak
opportunity	positive	strong
originality	weak	strong

While sentiment lexicons can be written by hand, they are usually extracted from large corpora and possibly hand-corrected afterwards. In general, the smaller the piece of text a human annotator has to judge is, the more difficulties they have in their judgement. Marking single words in isolation as being positive or negative, and maybe even providing a score, is therefore very difficult and hardly leads to high precision. (Our illustrative examples of *bad* and *joy* are comparatively easy cases.)

One of the first approaches to generate such a lexicon automatically is [10]. Adjectives are differentiated into positive and negative ones in order to detect antonyms and to distinguish near synonyms in text. The authors' algorithm is based upon the intuition that words with the same semantic orientations occur in coordinated constructions while words with different semantic orientations do not. For example (taken from [10]):

(7) a. fair and legitimate
    b. corrupt and brutal

(8) a. # fair and brutal

b. # corrupt and legitimate

After the detection of conjunctions, a regression model is used to establish relationships between conjunctions with respect to their semantic orientation and the result is a graph with the words being the nodes and the edges representing whether the two connected nodes are of the same or different semantic orientation. Finally, the graph is clustered into two sets: one being the class of positive adjectives and the other one being the class of negative adjectives.

More recent methods of lexicon generation typically use seed words (just to name some pioneer work: [39, 9]). The idea behind seeding methods is the following: at first, a rather small, hand-crafted set of reliable instances of negative and positive semantically oriented words are built. In a second step, a similarity measure is established, and words that are sufficiently similar to the seed words are added to the respective sets of positive and negative words.

An import sentiment lexicon of English is [47]. An excerpt from that dictionary is presented in Table 1. The first major lexicon for German is [29] which contains entries consisting of the lemma, the PoS-Tag using the Stuttgart-Tbingen Tagset[38], a weight and inflected forms. The weights are machine generated from various sources. A first step was to machine translate entries from the General Inquirer [32] into German and review them manually to remove bad entries. To extend these entries, a co-occurrence analysis is performed on a corpus of product reviews. The machine-translated entries are added to the product reviews and high co-occurrence words are extracted and, again, manually inspected and selected to be added to the lexicon. Finally, the German Collocation Dictionary [26] is used to extract polar noun clusters. The German Collocation Dictionary groups words by their semantic similarity and the groups with a strong relation to sentiment are calculated and added to the lexicon. Semantic orientation and the strength is then calculated using Pointwise Mutual Information.

3.2 Rule-based Approaches

Rule-based, or symbolic, approaches to sentiment analysis have the advantage to work fairly reliable; besides, it is easier to repair unintended behaviour of symbolic systems than to fix models for statistical classifications. Furthermore, especially companies are interested in tracking the continuous improvement of their systems over time, and it is easier to achieve consistent improvement of a system if it is rule-based, since additional rules for false negatives can be added to the system. Of course, machine-learning systems can also be improved, by way of providing more training data. However, increasing the size of the training set does not automatically lead to a better performance. Instead, the performance can reach a plateau or even drop, and it is unknown how much more training data is required to leave the plateau. Also, when the desire is to fix a particular problem or class of problems, it can be very difficult to obtain precisely the “right” training data for it.

This is important because companies need to respond to customer feedback. Clients may complain about missing or wrong sentiment relations, and then it is important to work on those cases in particular. Those corrections might not make the system much better, nor might there be any change in f-score, but the customer satisfaction may be more important than that.

On the other hand, the central disadvantage of rule-based systems is that the number of hits per rule is usually pretty low and follows a Zipfian distribution. A large percentage of the rules may deal with hapax legomena. And the rules which hit very frequently can easily be too general and produce many mistakes.

A prominent example for a symbolic system is the Semantic Orientation Calculator (SO-CAL) [37]. It is based on prior polarities of words and on rules for combining them to an aggregate sentence polarity, which account for the effects of specific contexts involving irrealis blocking, negations, diminishers and intensifiers.

(9) I do $[\text{not}]_{\sim 4}$ $[\text{like}]_{+1}$ this dishwasher, although the dishes are $[\text{really}]_{\text{intensifier:1.15}}$ $[\text{clean}]_{+2}$ afterwards.

Sentence 9 contains an intensifier and a negation. SO-CAL models intensification by multiplication, and 'really' has a value of 1.15. All polar words in proximity of the intensifier are modified accordingly and thus, clean is increased from 2 to 2.3. For negations, SO-CAL estimates the negation scope by looking forward as well as backwards up to a potential clause boundary. The effect of a negation is a polarity shift: The value of 'like' is shifted by 4 from 1 to -3.

SO-CAL was evaluated on reviews from epinions.com, movie reviews [22], and camera reviews. The average accuracy is reported as 0.7874. The accuracy is quite stable across the tested corpora. It also relatively robust against change of domains, where accuracies between 0.7938 and 0.8898 are given.

3.3 Aspect Analysis

Aspect analysis is a compromise between text-level analysis, which is more suitable for machine learning algorithms, and phrase-level analysis, which is a requirement for accurate sentiment analysis, since leaving out the detection of opinion targets generates many mistakes. Typically, aspect analysis is employed in the analysis of product reviews.

(10) a. This is the quietest dishwasher I have ever owned.
b. And yes, it's so quiet that you can't tell it's running [. . .]
c. Another great surprise was to see how clean our glassware and dishes come out.

Sentences 10a and 10b both refer to the same aspect of a dishwasher: its loudness. Sentence 10c, on the other hand, evaluates a different aspect.[6] Thus, aspect analysis consists of *a*) knowing what possible aspects of a product are, *b*) detecting aspects in text and mapping them to their *aspect category* and *c*) deciding about the polarity, intensity and possible other attributes of the sentiment.

Hu et. al [12] describe the task of aspect analysis as a special instance of text summarisation. A set of product reviews has to be summarised in order for potential buyers to get a quick overview over all the reviews. Such summarisation is advantageous, because popular items on large online stores can have thousands of reviews. (As of June, 2014, the most reviewed book on a large online store has $17,500$ reviews). The authors define two stages in their approach to aspect analysis. The first is to extract the product features (or *aspects*) that the reviews comment upon. The second step detects the polarity of the statements within the sentences that talk about an aspect. This latter step is the same as in sentence- or phrase-level sentiment analysis, so we ignore it here. In step 1, Hu et al. make a distinction between frequent and infrequent aspects. This distinction is only based on the differences in finding the aspects and not in a different role within sentiment analysis. Frequent aspects have to occur in at least 1% of all sentences from the reviews of a product. All other aspects are treated as infrequent ones. A rough outline of the algorithm:

1. detect frequent aspects
2. prune frequent aspects to reduce the noise generated within the detection of frequent aspects, and to remove redundancy stemming from more or less coarse-grained features (e.g., 'battery' and 'battery life')
3. create a list of opinionated words from the contexts of the previous step: a modifying adjective close to a frequent aspect is an opinionated word
4. detect infrequent aspects based on the opinion words gathered in the last step: the nearest noun phrase is an aspect

In this approach, aspects and opinion words co-depend; if the aspects are known, it is easier to compute the opinion, and vice versa.

Klinger et. al. [15] directly investigate this dependency between the evaluative expressions and aspects using factor graphs. The major finding is that the knowledge about aspects significantly improves the detection of evaluative expressions: 0.54 f_1-score for the detection of evaluative expressions in isolation increases to 0.65 f_1-score for its detection with gold-knowledge of aspects. The independent detection of targets is reported with an f_1-score of 0.32 and rises to 0.58 with gold-knowledge of evaluative expressions. The f_1-score for partial overlap is higher, but the tendency remains.

[6] All Sentences are taken from http://www.amazon.com/Bosch-SHP65T55UC-Stainless-Integrated-Dishwasher/dp/B00CWX0KDA/ref=sr_1_2?ie=UTF8&qid=1401714426&sr=8-2&keywords=dishwasher

3.4 Machine Learning Approaches

The majority of systems for detecting and classifying sentiment uses machine learning (henceforth ML) approaches. These are ideal for very complex problems that are hard to describe or even to understand for the human analyst. And since no comprehensive and detailed theory for sentiment in natural language exists, it is an obvious candidate to be tackled with ML approaches.

Wiegand [49] reports an accuracy of 0.775 for sentence-level polarity classification with an optimal feature set consisting of prior-polarity information, bag of words and a range of linguistic features. Unfortunately, the author only reports results for experiments on the MPQA corpus and it is thus unclear how the results carry over to different corpora. The author relates his work to Pang et al. [20] and compares the bag-of-word feature classifications to his own work. Pang et. al. achieve an accuracy of 0.829 while Wiegand reports an accuracy of 0.672. The difference stems from the granularity of the analysis: Pang et al. work at document-level. Another (not mentioned) difference are the used domains and genres: newspaper articles on the one hand, and movie reviews on the other hand. Still, the comparison provides evidence for an intuitively obvious observation: it is easier to classify sentiment at document-level than at sentence-level.

Recently, sentiment analysis was set as a task in the SemEval-2013 and 2014 challenges. The task provides micro-blogging data annotated at message level with a four-way classification: 'objective', 'neutral', 'positive', and 'negative'. The results that were achieved range from 0.1628 to 0.6902 in 2013 and from 0.396 to 0.7484 for this task in 2014.

As indicated above, fine-grained sentiment analysis is harder than coarse-grained analysis. But, fine-grained analysis is also the more interesting and challenging problem and has become increasingly popular. Unfortunately, fine-grained sentiment analysis cannot easily be formulated as a set of classification problems. Therefore, fine-grained ML approaches can either try to model compositional sentiment relying on syntactic or semantic representations [31], or do aspect analysis (cf. Section 3.3).

Socher et. al. [31] introduce a sentiment treebank ($11,855$ sentences from movie reviews) that contains syntactic analyses where each constituent is assigned a polarity: very positive , positive, neutral, negative or very negative. The authors also describe a classification system trained on this corpus using neural networks and semantic vector spaces. The sentiment of a phrase is computed by applying a compositionality function to each pair of sister nodes in a binary tree. Semantic vector representations of the words are used to learn and compute prior-polarities for the word or phrase.

The current interest in aspect analysis led to another SemEval task in 2014 on sentiment detection in customer reviews of restaurants and laptops. The winners, Kiritchenko et. al. [14] cast the problem of aspect term extraction as a tagging task: every token in a sentence is tagged as either belonging to an aspect term or not. The second sub-task provides gold-standard aspect terms within sentences, and the polarity of the sentence towards the aspect is to be determined. They describe a *support vector machine* (SVM) using surface, lexicon, and parse features. For all three classes, they define features that are essentially uni- and bigrams anchored at the aspect term. The performance varies significantly between the restaurant and the laptop data-sets. 0.7049 accuracy is reported for the laptop reviews, and 0.8016 for the restaurant reviews.

4 What is your Opinion, What is ours?

After relatively objective surveys of the linguistic notion of subjectivity and the field of sentiment analysis, we now turn to a relatively subjective synthesis. At first, we offer a set of definitions to clear up the terminology; then, we collect a number of questions that arise from the previous two sections and suggest some personal answers.

4.1 Terminology

By making a *factual* statement, the speaker asserts something about the real world that she regards as (in principle) verifiable by others. This is in contrast to *subjective* utterances. We distinguish the conveying of private states (subjective$_1$) from phenomena of intersubjectivity (perspective-taking etc.; subjective$_2$). In the remainder of this section, we will be concerned only with subjectivity$_1$. Language offers means of signalling the difference between factual and subjective$_1$, but speakers are not obliged to make it explicit; hence there is often room for interpretation by the hearer.

We regard *evaluations* as utterances that are often difficult to classify as either subjective or objective. They obligatorily mention a target, i.e., the entity being evaluated, and they seem to neutrally state (for example) an attribute of the target, usually situating it on some scale. The evaluation can have an underlying polarity (ex. 11a - 11d) but it need not (11e).

(11) a. The weather is nice.
b. The food is salty.
c. The dishwasher is quiet.
d. The dishes come out spotless.
e. This lecture hall is huge.

The speaker may introduce such a statement with *I think*, and accordingly, an addressee may dispute such a statement, e.g., by responding *Well, not quite.* This indicates that these cases are not as objective as *snow is white*, but at the same time they are no prototypical cases of private states: The two genuinely-subjective$_1$ types (emotions and opinions) cannot be disputed by an addressee. Emotional statements may have a target (12a) or not (12b), whereas opinions always have one (13a, 13b).

(12) a. I'm afraid of spiders.
b. I'm feeling great today.

(13) a. I like this kind of wine.
b. This has always been my favourite restaurant.

Notice that there is no point in the addressee replying *Not quite* or *That's not true* to any of the above utterances.

We thus see the evaluations as being situated on a middle ground between subjective and objective, but as leaning toward the objective: They are open to verification by others, and they will often be agreeable to a majority of the audience. The MacMillan Dictionary [30] defines evaluating as:

> to think carefully about something before making a judgment about its value, importance, or quality

Finally, we view *polarity* as a basic emotional category by which humans respond to experiences: positive or negative. Conveying an emotion usually includes polarity, but there is an ambiguous or non-polar range (e.g., *I am excited*). As stated above, opinions are always polar, whereas evaluations need not be. Utilizing lexical connotations is one way for a speaker to convey a polar evaluation: *The gentleman/man/jerk asked me a question*. Polarity, however, is not generally tied to subjectivity. So-called *polar facts* [44, 40] convey positive or negative consequences for some agent (as generally assumed, i.e. being the common knowledge), without evaluation or opinion being part of it:

(14) a. Joe Smith was murdered.
b. George Myers received the nobel prize.

Notice that adding *I think* does not have the same effect as with evaluations (with polar facts, it merely conveys degree of belief, not personal judgement).

Figure 4.1 summarizes the distinctions we have proposed.

4.2 Issues (1): Polarity and lexicons

How should subjectivity and polarity be handled in an ideal sentiment lexicon?

An inspection of various prior-polarity lexicons reveals a large number of entries which are not subjective according to our terminology landscape as introduced above, but very relevant for the analysis of product reviews or political discussions. Examples are shown in Table 2.

The examples indicate that subjectivity as a lexical feature is very difficult to agree on, and therefore we would suggest to eliminate it from (or at least, signficantly reduce its role in) lexical description. Instead, we would clearly focus on

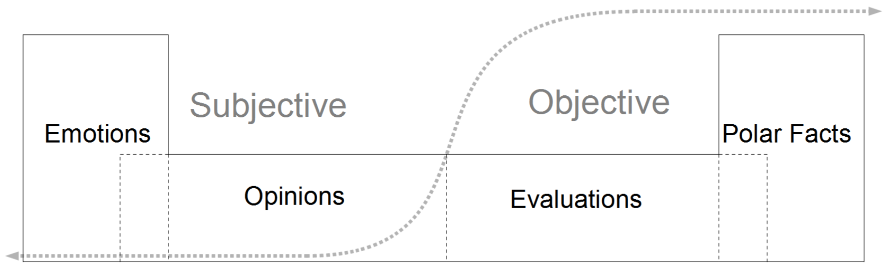

Fig. 2 Mapping the Sentiment terminology. Dotted lines indicate the room for interpretation by recipients of language. Solid lines represent the underlying model which we assume. The sloped line thus indicates how likely a recipient is to interpret an utterance supposed to convey an opinion, evaluation, etc. as subjective or objective, and even as an opinion, evaluation, etc.

word	Subjectivity Lexicon	SO-CAL
eliminate	w	-4
assassinate	w	-2
veto	s	n.a.
erosion	w	-2

Table 2 Comparing selected entries from the Subjectivity Lexicon [46] and the lexicon developed for SO-CAL [37]. ‘s’ = ‘strongly subjective’; ‘w’ = ‘weakly subjective’.

polarity, which is a central ingredient of all the three of our categories subjective opinion, semi-objective evaluation, and polar fact.

Evaluations, as in product reviews, behave much like facts – recall examples 11a - 11e. They do not include linguistic markers of subjectivity/opinion, so the question is how human readers understand that, e.g., examples 11c and 11d are positive evaluations of a dishwasher? Arguing in the terms of [52], we can say that the speaker’s expectations towards the dishwasher are exceeded, which in turn creates positive emotions. Then, something positive is stated about an object, and it is not an opinion.

If we reduce sentiment lexicon construction to assigning positive/negative values, the process becomes much easier, and the lexicon can then be used for a range of different tasks where “standard” opinion or domain-specific evaluations, or polar facts can play a role. These facts would not have to be encoded in separate lists anymore but simply were part of a prior-polarity lexicon. The rather artifical distinction between entries for a polar fact lexicon and an emotion/subjectivity lexicon vanishes.

However, using such a lexicon for a broad range of tasks makes it necessary to pay more attention to context and to calculate posterior polarity in appropriate ways; that is the issue we address next.

4.3 Issues (2): Context

Can sentiment analysis benefit from considering coarse-grained text structure?
For the task of text-level sentiment analysis, it can help to take the genre-specific text structure into account. In our work with movie reviews, we tested this by implementing a prior step of 'zone identification': Movie reviews usually provide information about what happens in the plot (description), and they present the author's opinion on various aspects (comment). We found that description and comment are most often clearly separate in the paragraphs of a review. Using a classifier making this distinction and then restricting the text given to the sentiment analyzer to the comment paragraphs yields to improvements ranging from 2 to 12% depending on the quality of the prior describe/comment classification [36].

What about argumentation and its structure?
By extension to the previous point, when sentences are composed of multiple clauses, the argumentative orientation can be modulated by connectives like *but* or *although*, which also renders the "single-number" sentiment analysis as a great simplification.

(15) a. Big Brother is back these days, but in the meantime, the country has invested itself so deeply in its fantasy of cyber-liberation that no outrage will be sufficient to move it.[7]

Many systems compute sentiment score at sentence-level. Is that adequate?
The basic idea that is regularly implemented is to see sentence-level sentiment as the average of the lexical polarities in the sentence. This is a simple rule that often works, but it cannot do justice to sentences like 16a to 16c, where complex opinions are being stated that cannot just be "averaged". Likewise, the interesting recent approach of Socher et al. [31], which propagates sentiment values from node to node in the syntactic tree (see Section 3.4), does not capture the sentiment, because still, "overall sentiment" of the sentence is being reduced to a single number.

(16) a. The Germans are partners and adversaries at the same time.[8]
b. To snub and even to wound your most zealous supporters, as Obama has done, is regarded as a mark of maturity in Washington.[9]
c. Die Schweiz will die Zuwanderung von EU-Bürgern beschränken –
the Switzerland wants the immigration of EU-citizens restrict –
und Europa ist empört.
and Europe is outraged.
Switzerland wants to restrict the immigration of EU-citizens – and Europe is outraged.

[7] From http://www.faz.net/aktuell/feuilleton/debatten/the-u-s-and-the-n-s-a-scandal-freedom-the-big-american-lie-12263704.html?printPagedArticle=true.

[8] From http://www.spiegel.de/international/germany/why-spiegel-is-posting-leaked-nsa-documents-about-germany-a-975431.html

[9] From http://www.faz.net/aktuell/feuilleton/debatten/the-u-s-and-the-n-s-a-scandal-freedom-the-big-american-lie-12263704.html

Example 16c[10] is only interpretable in terms of sentiment if the relations between the entities involved in the sentence are examined. Entities are "Switzerland", "the immigration of EU-citizens", "EU-citizens", "EU", "Europe" and lastly, the author. Those relations can then be investigated and classified into being negative, positive, neutral in one dimension and being an opinion, evaluation, fact or emotion in another dimension.

Just as aspect analysis re-interprets the task of text-level sentiment analysis, compositional sentiment analysis needs to be re-interpreted as well: sources and targets should be a central part of sentiment analysis. It is neither feasible to assign the sentence-level sentiment to all entities within the sentence, nor to the topic of the text or sentence. Instead, what is missing is a systematic detection of sources and targets.

How can we model sentiment interaction when multiple entities are affected? When multiple entities are involved, interesting opportunities for sentiment classification arrise. Let us assume that each sentiment is a relation between a source S and a target T. In longer texts, we will probably encounter different relations with a common source and target. Our intuition is that all relations between the same source and target tend to have the same polarity. This principle could be used as an optimisation principle and therefore help out in dubious relations. If two entities share relation $R_1 \ldots R_n$ and R_3 to R_n are clearly positive and, additionally, the classification system is insecure about assigning a negative polarity to R_1, then the principle can influence the decision and R_1 gets a "neutral" label or none at all. If the author of the sentence wants to convey a different sentiment from the source towards the target, he can do it but he has to do it very explicitly. For example by using discourse markers as discussed above on argumentation structure.

(17) a. When Italian Prime Minister Matteo Renzi now offers the prospect of support for Juncker, what we are seeing is really part of a larger offensive against Berlin's so-called austerity diktat.

Extending this principle to multiple entities is possible via the Balance Principle, which we introduced in Section 2.1.2: All relations connecting three entities into a fully connected (sub-)graph follow the tendency that the product of their polarities is positive. To illustrate this, we consider the entities "Renzi", "Juncker" and "Berlin's so-call austerity diktat". An explicitly positive relation can be established between "Renzi" and "Juncker" because of "offers the prospect of support". And the relations from "Renzi" and "Juncker", respectively, to "Berlin's so-called austerity diktat" are both negative. Thus, this triadic relationship fullfills the balance principle. If a classifier for the polarity of the relations is unsure about one of its three decisions, it can use the principle to gain additional evidence.

[10] From http://www.sueddeutsche.de/politik/steuergeheimnis-und-zuzug-stopp-warum-die-schweiz-europas-liebster-pruegelknabe-ist-1.1659263

5 Summary

In this chapter, we provided an (arguably selective) overview of the central aspects of the computational sentiment analysis problem, and in particular pointed to some interesting recent work. We mentioned several performance results in order to give an impression on the extent to which the various problems can at present be solved with automatic methods.

The other survey was certainly very selective and discussed a number of notions from the Linguistics literature on subjectivity, as we see them to be relevant for sentiment analysis; prominent topics here were the 'private state' and the facets of 'evidentiality' and 'factuality', which deserve close attention when extracting sentiment-related information from text.

In the final section of the chapter, our goal was to identify several critical issues with sentiment analysis and at some points to suggest possible steps toward finding solutions. As a general sentiment (pardon the pun) we believe that more detailed linguistic analysis would be instumental for making progress with high-quality and fine-grained sentiment analysis, which requires careful analysis of contextual effects for identifying sources of opinions, for computing polarity in compositional ways, and for a more sophisticated identification of the entities that can be assigned sentiment values in complex sentences.

As a final remark, we want to point out that we touched only very superficially on the difference between opinions or emotions as mental states on the one hand, and the linguistic utterances speakers produce to express them on the other hand. This distinction is connected to various parts of the picture we presented. At the beginning of a text production process, there are mental states and stances of the author; they influence both the selection of information that gets verbalised (*what* to say) and the actual shape of the verbalisation (*how* to say it: the linguistic choices). The decision on what to say is already a matter of subjectivity, as we mentioned briefly at the end of Section 2.2.2. Then, when the author makes choices among lexical and grammatical options, she can opt to clearly mark the intended factuality or the various dimensions of subjectivity, or she can leave that undespecified (deliberately or inadvertently), which in turn leaves room for interpretation by the addressee. That is one major reason why Sentiment Analysis is *inherently* very challenging – not only for the machine.

Acknowledgements

We thank our anonymous reviewers for detailed and constructive criticism, which enabled us to improve this work at several points.

References

1. Jean-Claude Anscombre, and Oswald Ducrot. 1983. *L'argumentation dans la langue.* Sprimont: Editions Margana.
2. A. Balahur, R. Steinberger. 2009. Rethinking Sentiment Analysis in the News: from Theory to Practice and back. *Proceedings of WOMSA*.
3. Alexandra Balahur. 2011. *Methods and Resources for Sentiment Analysis in Multilingual Documents of Different Text Types*. Dissertation. Department of Software and Computing Systems. Alacant, University of Alacant.
4. Luciano Barbosa, and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. 36 – 44. Association for Computational Linguistics.
5. Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*. 43(3):209–226. Springer Netherlands.
6. Jaime Carbonell. 1979. Subjective Understanding: Computer Models of Belief Systems. *Yale Computer Science Technical Report* 150
7. Simon Clematide, Stefan Gindl, Manfred Klenner, Stefanos Petrakis, Robert Remus, Josef Ruppenhofer, Ulli Waltinger, and Michael Wiegand. 2012. MLSA – A Multi-layered Reference Corpus for German Sentiment Analysis. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* 3551 – 3556
8. Patrick Dendale, and Liliane Tasmowski. 2001. Introduction: Evidentiality and related notions. *Journal of Pragmatics* 33:339-348
9. Andrea Esuli, and Fabrizio Sebastiani. 2005. Determining the Semantic Orientation of Terms through Gloss Classification. *CIKM'05. Proceedings of the 14th ACM International Conference on Information and Knowledge Management* 616 – 624
10. V. Hatzivassiloglou, K. R. McKeown. 1997. Predicting the semantic orientation of adjectives. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*. 174 – 181. Association for Computational Linguistics.
11. Fritz Heider. 1958. *The Psychology of Interpersonal Relations*. John Wiley & Sons Inc.
12. Minqing Hu, and Bing Liu. 2004. Mining opinion features in customer reviews. *AAAI* 4(4):755 – 760
13. Lauri Karttunen, and Annie Zaenen. 2005. Veridicity. In *Annotating, Extracting and Reasoning about Time and Events*, ed. Graham Katz, James Pustejovsky, and Frank Schilder. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.
14. Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. Detecting Aspects and Sentiment in Customer Reviews. *Proceedings of the eigth international workshop on Semantic Evaluation Exercises (SemEval-2014)*
15. Roman Klinger, and Philipp Cimiano. 2013. Bidirectional Inter-dependencies of Subjective Expressions and Targets and their Value for a Joint Model. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* 848 – 854. Association for Computational Linguistics.
16. Roman Klinger, and Philipp Cimiano. 2014. The USAGE Review Corpus for Fine Grained Multi Lingual Opinion Analysis. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* 2211 – 2218
17. Ronald W. Langacker. 1987. *Foundations of Cognitive Grammar, Volume 1, Theoretical Prerequisites.* Stanford University Press.
18. James R. Martin, and Peter R.R. White. 2007. *The Language of Evaluation: Appraisal in English.* Basingstoke: Palgrave Macmillan.
19. Saif Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. Building the State-of-the-Art in Sentiment Analysis of Tweets. *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*

20. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs Up? Sentiment Classification Using Machine Learning Techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*. 79 – 86
21. Bo Pang, and Lillian Lee. 2003. A Sentimental Education: Sentiment Analysis Using Subjectivity Based On Minimum Cuts. *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*. 271 – 278
22. Bo Pang, and Lillian Lee. 2004. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*. 115 – 124
23. Bo Pang, and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd annual meeting on Association for Computational Linguistics* 271. Association for Computational Linguistics.
24. Bo Pang, and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* 1 – 135
25. Rosalind Picard. 1995. Affective Computing. *MIT Media Laboratory Perceptual Computing Section Technical Report* 321
26. Uwe Quasthoff. 2010. *Deutsches Kollokationswörterbuch*. deGruyter, Berlin, New York.
27. Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1972. *A Grammar of Contemporary English.* London: Longman.
28. Siddarth Ramaswamy. 2011. *Visualization of the Sentiment of the Tweets.*. Department of Computer Science. Raleigh, North Carolina State University.
29. Robert Remus, Uwe Quasthoff, and Gerhard Heyer. 2010. SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. *Proceedings of the 7th International Language Resources and Evaluation (LREC)*.
30. Michael Rundell, and Gwyneth Fox. 2002. *Macmillan English dictionary for advanced leaners*. Macmillan Oxford.
31. Richard Socher, Alex Perelyglin, Jean Wu, Jason Chuang, Chrisopher Manning, Andrew Ng, and Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. *EMNLP 2013*
32. Philip Stone, Dunphy Dexter, and Marshall Smith. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press.
33. Mostafa Al Masum Shaikh, Helmut Prendinger, and Mitsuru Ishizuka. 2010. Emotion Sensitive News Agent (ESNA): A system for user centric emotion sensing from the news. *Web Intelligence and Agent Systems* 8(4):377 – 396
34. Benedictus de Spinoza. 1677. *Ethics, and on the Correction of Understanding*. Translated by Andrew Boyle, 1986. London: Dent.
35. Maite Taboada, and Jack Grieve. 2004. Analyzing appraisal automatically. *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report)*. SS 04 07:158 – 161. AAAI Press.
36. Maite Taboada, Julian Brooke, and Manfred Stede. 2009. Genre-based paragraph classification for sentiment analysis. Proceedings of the 10th SIGdial Workshop on Discourse and Dialogue, London.
37. Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. *Lexicon-based methods for sentiment analysis*. *Computational Linguistics* 37(2):267 – 307
38. Christine Thielen, and Anne Schiller. 1996. *Ein kleines und erweitertes Tagset frs Deutsche*. Lexikon & Text 215-226.
39. Peter Turney. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th annual meeting on Association for Computational Linguistics* 417 – 424. Association for Computational Linguistics.
40. Cigdem Toprak, Niklas Jakob, and Iryna Gurevych. 2010. Sentence and expression level annotation of opinions in user-generated discourse. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* 575 – 584. Association for Computational Linguistics.
41. Arie Verhagen. 2005. *Constructions of Intersubjectivity: Discourse, Syntax, and Cognition.* Oxford/New York: Oxford University Press.

42. Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation* 39:165 – 210
43. Janyce Wiebe. 1994. Tracking point of view in narrative. *Computational Linguistics* 20(2):233-287
44. Theresa Wilson. 2008. Annotating subjective content in meetings. Proceedings of the Sixth International Language Resources and Evaluation Conference (LREC), Marrakech, Morocco.
45. Theresa Wilson, Janyce Wiebe, and Paul Hoffman. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics* 3(35):399 – 433. MIT Press.
46. Theresa Wilson, Janyce Wiebe, and Paul Hoffman. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *Proceedings of HLT-EMNLP*.
47. Theresa Wilson, Janyce Wiebe, and Paul Hoffman. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *Proceedings of HLT-EMNLP-2005*
48. Ulli Waltinger. 2010. GERMANPOLARITYCLUES: A Lexical Resource for German Sentiment Analysis. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*
49. Michael Wiegand. 2011. *Hybrid approaches for sentiment analysis*. Universität des Saarlandes.
50. Volker Zastrow. 2013. Das Amalgam. *Frankfurter Allgemeine Sonntagszeitung*. http://www.faz.net/-gpg-762nc
51. Xiaodan Zhu, Svetlana Kiritchenko, and Saif Mohammad. 2014. Recent Improvements in Sentiment Analysis of Tweets. *Proceedings of the eigth international workshop on Semantic Evaluation Exercises (SemEval-2014)*
52. Andrew Ortony, Gerald L. Clore, and Allan Collins. 1990. *The cognitive structure of emotions*. Cambridge University Press.