ORIGINAL PAPER



# The good, the bad and the implicit: a comprehensive approach to annotating explicit and implicit sentiment

Marjan Van de Kauter · Bart Desmet · Véronique Hoste

Published online: 15 February 2015 © Springer Science+Business Media Dordrecht 2015

**Abstract** We present a fine-grained scheme for the annotation of polar sentiment in text, that accounts for explicit sentiment (so-called private states), as well as implicit expressions of sentiment (polar facts). Polar expressions are annotated below sentence level and classified according to their subjectivity status. Additionally, they are linked to one or more targets with a specific polar orientation and intensity. Other components of the annotation scheme include source attribution and the identification and classification of expressions that modify polarity. In previous research, little attention has been given to implicit sentiment, which represents a substantial amount of the polar expressions encountered in our data. An English and Dutch corpus of financial newswire text, consisting of over 45,000 words each, was annotated using our scheme. A subset of this corpus was used to conduct an interannotator agreement study, which demonstrated that the proposed scheme can be used to reliably annotate explicit and implicit sentiment in real-world textual data, making the created corpora a useful resource for sentiment analysis.

**Keywords** Corpus annotation · Polarity · Sentiment analysis · Natural language processing

V. Hoste e-mail: Veronique.Hoste@UGent.be

M. Van de Kauter  $(\boxtimes) \cdot B$ . Desmet  $\cdot V$ . Hoste

LT3, Language and Translation Technology Team, Ghent University, Groot-Brittanniëlaan 45, 9000 Ghent, Belgium

e-mail: Marjan.VandeKauter@UGent.be

B. Desmet e-mail: Bart.Desmet@UGent.be

# 1 Introduction

Due to the explosive growth of the World Wide Web in the past few decades, Internet users today are exposed to a vast amount of information, from which they can no longer manually distill the information that is relevant to them. This evolution has resulted in a rapidly growing interest in text mining applications, by means of which users can automatically extract and analyze information from the web and other collections of text. One of these text mining tasks is sentiment analysis, also referred to as opinion mining, which is aimed at the automatic identification and analysis of "people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes" (Liu 2012). The automatic extraction of opinions from text has a wide range of application possibilities: it can be used by companies to track how their brand is perceived by consumers (Zabin and Jefferies 2008), by individuals who need advice on purchasing the right product or service (Dabrowski et al. 2010), by nonprofit organizations [e.g. for the detection of suicidal messages (Desmet and Hoste 2014)], etc. An extensive overview of existing work on sentiment analysis is given by Pang and Lee (2008) and Liu (2012).

A large amount of the research on sentiment analysis and opinion mining focuses on user-generated content, e.g. product reviews (Turney 2002; Pang et al. 2002; Dave et al. 2003; Hu and Liu 2004; Popescu and Etzioni 2007), blogs (Ounis et al. 2006; Boldrini et al. 2009; O'Hare et al. 2009), tweets (Kouloumpis et al. 2011; Roberts et al. 2012; Nakov et al. 2013), etc. These types of texts are characterized by the use of subjective language, by means of which opinions and other types of sentiment are expressed. Correspondingly, most sentiment analysis research is dedicated to the detection and analysis of subjective words and phrases. However, objective utterances can also express sentiment, be it in an indirect way. It is possible for readers of a text to infer a positive or negative impression of a certain topic from factual information using world knowledge or common sense. Implicit expressions of sentiment like this are particularly common in more general, factual text types, such as newswire for example. As a consequence, researchers working on sentiment analysis in these kinds of texts have started to pay attention to implicit ways of expressing sentiment (Musat and Trausan-Matu 2010; Balahur et al. 2011a, b; Zhang and Liu 2011; Feng et al. 2013). The amount of work dedicated to this issue, however, is limited.

In this paper, we present a new annotation scheme for the detection of explicit as well as implicit expressions of sentiment on a sub-sentential level. In addition, the annotation scheme also allows to analyze these expressions in a fine-grained manner through the annotation of sentiment sources, targets, modifiers, etc. To our knowledge, none of the other existing annotation schemes for sentiment combine the identification of explicit and implicit sentiment expressions with the fine-grained analysis of these expressions below sentence level. We applied the annotation scheme to an English and Dutch corpus of financial news articles, in which we expected sentiment to be uttered both explicitly and implicitly. The resulting annotated corpus can serve as a training and evaluation dataset for the detection of explicit and implicit sentiment. It can also be useful for tackling sentiment analysis subtasks like polarity classification and source and target identification. We conducted an inter-annotator agreement study on a subset of the corpus to assess whether the annotation scheme can be reliably applied to real-world textual data. The results of this study show that overall, the agreement between annotators is high. Furthermore, we calculated some statistics on the entire annotated corpus, from which we can conclude that the annotation of explicit as well as implicit expressions of sentiment is instrumental to capturing the full spectrum of sentiment conveyed in financial newswire text.

The remainder of this paper is structured as follows. Section 2 gives an overview of existing work on the creation of annotated corpora for sentiment analysis. In Sect. 3, we present our annotation scheme. We discuss how it differs from other existing schemes, and present the different properties of our scheme, including examples for each annotation step. Section 4 elaborates on the results of our annotation efforts. First, we describe the data and annotation procedure used. Then, we discuss the results of the inter-annotator agreement study that was conducted to assess the annotation scheme. Next, we present some statistics calculated on the full annotated corpus. Finally, Sect. 5 gives some conclusions and prospects for future work.

# 2 Related work

Existing approaches to sentiment analysis can be roughly divided into two main categories: lexicon-based and machine learning approaches. Hybrid methods combine both approaches. Lexicon-based methods make use of sentiment or subjectivity lexicons (Turney 2002; Hu and Liu 2004; Esuli and Sebastiani 2006; Ding et al. 2008). These lexicons contain sentiment words, also called opinion words, listed with their polarity and strength. Taboada et al. (2011) provide an overview of work on lexicon-based sentiment analysis. Machine learning-based systems are trained on datasets in which the sentiment is labeled. Sentiment labels can be extracted from existing resources such as product reviews assigned a rating by consumers (Pang et al. 2002; Turney 2002; Dave et al. 2003), or they can be created through manual annotation. The latter approach allows researchers to tailor the annotations to their domain, level of granularity or application of interest. In the following paragraphs, we give an overview of existing efforts in the manual annotation of sentiment in text. In Sect. 3, we present our new annotation scheme and compare it to existing annotation studies. Furthermore, we refer to sentiment analysis tasks for which the different properties of our annotation scheme could be useful.

When annotating sentiment, the first step is usually determining whether a piece of text is opinionated (i.e. subjective) or not. This annotation task is often performed at the sentence level (Yu and Hatzivassiloglou 2003; Kim and Hovy 2005; Seki et al. 2007; Strapparava and Mihalcea 2007; Bermingham and Smeaton 2009; Abdul-Mageed and Diab 2011), or sometimes at the document or paragraph level (Ounis et al. 2006; Devitt and Ahmad 2007; Macdonald et al. 2007; Ferguson et al.

2009; Roberts et al. 2012). However, documents, paragraphs and even sentences can be made up of a mixture of subjective and objective content. Furthermore, they can contain multiple opinions.

Wiebe et al. (2005) suggested that for information extraction systems, it would be useful to identify the individual clauses containing opinions. They developed a detailed annotation scheme for the identification of key components and properties of opinions and emotions in text. The goal of the scheme is to distinguish subjective from factual information and is centered around the notion of so-called private states, which Quirk et al. (1985) define as states that are "not open to objective observation or verification" (e.g. opinions, emotions, etc.). Three types of expressions of private states are annotated in context, at the word and phrase level: explicit mentions of private states, speech events expressing private states and expressive subjective elements (Banfield 1982). For each of these expressions, a private state frame is defined, which includes the source of the private state, its intensity, the type of attitude (positive, negative, other or none) and other properties of the private state. The annotation scheme has been used to annotate the MPOA corpus<sup>1</sup>, a freely available English corpus consisting mainly of newswire text, which has been used for sentiment analysis experiments at the expression level (Wilson et al. 2005; Breck et al. 2007). In other work, the annotations have been extended with attitude frames tied to the private frames that represent a wider set of attitude types (e.g. positive/negative arguing, positive/negative intentions, speculation) and target frames (Wilson and Wiebe 2005). Furthermore, topic annotations have been added to part of the MPQA corpus by Stoyanov and Cardie (2008). The linguistic work most related to the annotation scheme of Wiebe et al. (2005) is the Appraisal Theory framework (Martin and White 2005), which stems from systemic functional linguistics (SFL) (Halliday 1994) and is concerned with construing interpersonal meaning in written text. The framework consists of three interacting domains: attitude (feelings), engagement (taking positions with regards to attitudes) and graduation (the grading of attitudes).

Aside from the fine-grained sentiment annotation scheme of Wiebe et al. (2005), which is the most well-known, a number of other studies exist on the annotation of sentiment expressions, their properties and components. The work of Read and Carroll (2012), for instance, can also be linked to the Appraisal Theory framework. The typology of Appraisal is applied to a corpus of book reviews, in which different units of appraisal are annotated. Different emotions are also identified by the EmotiBlog annotation scheme (Boldrini et al. 2009), which is inspired by Wiebe et al. (2005) and aims at a finer-grained annotation of emotions in non-traditional textual genres. Building on the work of Scherer (2005), Boldrini et al. (2009) selected groups of emotions to be identified such as criticism, happiness, guilt, surprise, etc. Polarity, sources and targets are annotated as well. The annotation scheme has been used to annotate the EmotiBlog corpus, a collection of English, Spanish and Italian blog posts (Boldrini et al. 2012). Somasundaran et al. (2008) identify two types of opinions in a corpus of group meetings: sentiment and arguing. For these opinions, the polarity and target are determined. Opinion frames are

<sup>&</sup>lt;sup>1</sup> http://mpqa.cs.pitt.edu/corpora/mpqa\_corpus/

annotated which are composed of two opinions with targets that are related. For the NTCIR-7 MOAT Task (Seki et al. 2008), opinion annotation was performed on the sub-sentence level. Annotations were made for different subtasks, including determining the polarity of each opinion expression and detecting opinion holders and targets. In the corpora of tweets and SMS collected for the SemEval-2013 task on Sentiment Analysis in Twitter (Nakov et al. 2013), subjective words and phrases were annotated with their respective polarity. For the SemEval-2014 task on Aspect Based Sentiment Analysis (Pontiki et al. 2014), aspects i.e. features of certain target entities (e.g. the screen of a laptop) and the polarity of the opinions expressed towards these aspects were annotated in a corpus of restaurant and laptop reviews. Kessler et al. (2010) annotate sentiment expressions with their polarity, targets, modifiers and opinion holder. Finally, Asher et al. (2008) annotate opinions using discourse relations. As opposed to the other studies mentioned, in which expressions of sentiment can take various forms, the annotation scheme of Asher et al. (2008) only allows the annotation of an opinion expression if it contains an opinion word from their lexicon, or if it is related to an opinionated expression via a rhetorical relation.

As we discussed in Sect. 1, most of the existing research on sentiment focuses on the detection and analysis of explicit sentiment. The amount of work dedicated to implicit expressions of sentiment is, at this moment, limited (Musat and Trausan-Matu 2010; Balahur et al. 2011a, b; Zhang and Liu 2011; Feng et al. 2013). Implicit sentiment has, however, been annotated in consumer reviews by Toprak et al. (2010) and in meeting content by Wilson (2008) [the latter uses an adaptation of the scheme of Wiebe et al. (2005)]. The annotation schemes proposed in these two studies allow for the detection of not only explicit expressions of opinions, but also of factual information that implies a positive or negative evaluation. Wilson (2008) and Toprak et al. (2010) refer to objective utterances like this as resp. *objective polar utterances* and *polar facts*.

# 3 Annotation scheme for polar expressions

The most important motivation for developing a new sentiment annotation scheme was the fact that, to our knowledge, no annotation scheme exists for the fine-grained annotation of explicit as well as implicit expressions of sentiment below sentence level. Toprak et al. (2010) and Wilson (2008) detect so-called polar facts/objective polar utterances, but these are only identified on the sentence and utterance<sup>2</sup> level, respectively. Similar to explicit expressions of opinions, we want to pinpoint the particular phrases that express sentiment in an implicit way. Drury and Almeida (2011) identified so-called event words in business news stories; these words indicate a positive or negative event, depending on the context (e.g. 'drop', 'increase', etc.). However, we want to consider all kinds of implicit sentiment

 $<sup>^2</sup>$  Wilson (2008) states that an utterance "may be a single phrase or expression, but whenever possible it is a sentence or proposition with references to the source and target of the subjectivity included in the span that is marked".

expressions (not just expressions of so-called events), including those containing more than one word. We therefore propose a new fine-grained annotation scheme, for which we build on the insights of Wiebe et al. (2005), Wilson (2008), Toprak et al. (2010) and other studies.

The main goal of our annotation scheme is to detect and analyze explicit as well as implicit expressions of positive and negative sentiment, which we call polar expressions. These expressions are annotated and assessed at expression level. Two types of polar expressions are identified:

- *private state expressions* explicit expressions of positive or negative sentiment towards a target entity or entities. Private states are internal states that cannot be directly observed by others, e.g. opinions, beliefs, thoughts, feelings, emotions, judgments, etc. (Quirk et al. 1985; Wiebe et al. 2005). They represent someone's personal attitude towards a target entity. An example of a private state expression can be found in Fig. 1. Here, Stephen Robertson explicitly utters a negative opinion about the target 'the data' by means of the expression 'were "disappointing"'.<sup>3</sup> Contrary to Wiebe et al. (2005), we do not make a distinction between different types of private state expressions (i.e. explicit mentions of private states, speech events expressing private states and expressive subjective elements), in order to facilitate the application of the annotation scheme.
- polar fact expressions implicit expressions of positive or negative sentiment, i.e. . expressions conveying a piece of factual information that, when relating to a certain target entity or entities, results in a positive or negative impression of that target/those targets. Sometimes, a person does not explicitly express his/her personal opinion about a certain topic, but an evaluation of it can nonetheless be inferred from what he/she says using common sense, world knowledge and context. Objectively verifiable sentences from which a positive or negative evaluation can be inferred are referred to by Toprak et al. (2010) as polar facts<sup>4</sup>. Whereas Toprak et al. (2010) identify full sentences as polar facts, we pinpoint the particular phrases that convey the relevant factual information. When regarding these so-called polar fact expressions in combination with their target entities, the polarity of the expressed sentiment can then be determined. For instance, in Fig. 2, 'is still facing action from US federal, state and local governments' is a polar fact expression with 'BP' as its target entity, because the fact that BP is facing legal action results in a negative impression of BP.

Because the distinction between private state expressions and polar fact expressions (i.e. between subjective expressions and factual information) is not always an easy one to make, the type of a polar expression is not specified by

<sup>&</sup>lt;sup>3</sup> When applying our fine-grained annotation scheme to text, we make use of the brat annotation tool (see Sect. 4.2). In brat, the polarity of the sentiment expressed by a polar expression about a certain target is denoted by the colour of the arrow pointing from the polar expression to that target (viz. green for positive, red for negative, purple for unknown and orange for other) and a symbol (viz. +, -, ? and  $\sim$ ).

<sup>&</sup>lt;sup>4</sup> Other sources use the terms *objective polar utterances* (Wilson 2008) or *evaluative factuals* (Nigam and Hurst 2004) for polar facts.





Fig. 2 Example of a polar fact expression

assigning it to one of these two categories, but by locating the expression on a continuum ranging from objective to subjective (see Sect. 3.1).

A sentence can contain multiple polar expressions, and a mixture of private state expressions and polar fact expressions is possible. Furthermore, polar expressions can be identified on different sentence levels. This means that one polar expression can be embedded in another. Our annotation scheme does not only allow for the detection of each of these polar expressions, it also enables annotators to analyze the expressions in a comprehensive manner. For every polar expression, several related elements and attributes are identified, e.g. targets, sources, modifiers, etc. Because private state expressions and polar fact expressions are not regarded as two strictly distinct categories of expressions, but rather as two overlapping categories located on a continuum ranging from objective to subjective, the same annotation procedure is used for both types of expressions. This ensures that the same types of elements and attributes are identified for each polar expression, and makes the annotation scheme as a whole more consistent and easier to apply. Besides polar expressions, the scheme covers the identification of a few types of relations that are not directly related to sentiment analysis (viz. coreference and feature relations), but could be useful for certain subtasks in the sentiment analysis field. To apply the full annotation scheme to the test corpus described in Sect. 4, we made use of the brat rapid annotation tool (Stenetorp et al. 2012).

The following subsections describe the different properties of our annotation scheme. Figure 3 provides a schematic overview of the scheme. The annotation procedure is illustrated by means of example sentences and phrases from the test corpus, annotated in the brat environment. Note that not every polar expression in the example sentences and phrases is (fully) discussed.

### 3.1 Polar expressions

Polar expressions are the core element of our annotation scheme. They are linguistic expressions that (explicitly or implicitly) convey positive or negative sentiment towards a certain entity (or entities). As can be seen in Figs. 4, 5 and 6, polar expressions can take different forms: they can be adjectives, verbal constructions, noun phrases, etc. Therefore, no strict rules were defined about boundary detection



Fig. 3 Schematic overview of the annotation scheme

The impact of higher energy prices also squeezed corporate margins across the eurozone.

Fig. 4 Example of a polar expression in the form of an adjective



Fig. 5 Example of a polar expression in the form of a verb



Fig. 6 Example of a polar expression in the form of a noun

or the types of words or phrases to be annotated as polar expressions. Furthermore, a polar expression can consist of non-consecutive tokens<sup>5</sup>. Note that we do not identify polar expressions below word level or above sentence level.

For each polar expression, three attributes are determined: type (with corresponding confidence score), insubstantiality and causativity.

*Type:* Two types of polar expressions are defined: private state expressions and polar fact expressions. However, categorizing a polar expression as either a private state or a polar fact expression can be a difficult task, since the distinction between subjectivity and objectivity is not always an easy one to make. The type of a polar expression is therefore specified by locating it on a continuum ranging from objective to subjective (cf. Riloff and Wiebe (2003), who divide subjectivity clues into strongly and weakly subjective clues). The further an expression is situated on the scale, the more subjective it is. Polar expressions can take 1 of 4 positions on the

<sup>&</sup>lt;sup>5</sup> The same applies to the other elements covered by our annotation scheme, viz. modifiers, sources, source expressions, targets and causes.



In spite of these problems , there , Vodafone 's pre-tax profits are expected to rise modestly from Pounds 10bn to Pounds 10.2bn



subjectivity continuum, denoted by the values 0, 1, 2 and 3. Polar fact expressions are objective and consequently receive the value 0. Private state expressions are situated on the other side of the subjectivity scale and are labeled 3. If a polar expression cannot unambiguously be classified as a polar fact or a private state expression, the values 1 and 2 can be used to locate the polar expression somewhere between the two extremes of the subjectivity continuum. 1 is chosen when annotators have a slight preference for the type *polar fact expression*, whereas the use of 2 indicates a preference for *private state expression*. In Fig. 7, the polar expression 'are expected to rise modestly from Pounds 10 bn to Pounds 10.2 bn' has received the value 1. A rise from Pounds 10 to 10.2 bn can be objectively measured, but the word 'modestly' makes the information conveyed by this expression slightly subjective.

These type annotations can be useful for research on subjectivity classification (Wiebe et al. 1999; Riloff and Wiebe 2003). The four-point scale used by the annotators allows users of the data to perform subjectivity classification on a fine-grained or coarse-grained (by reducing the annotations to a two-point objective/subjective scale) level. When specifying the type of a polar expression, annotators also determine a **confidence score** (viz. *low, medium* or *high*), which indicates how certain they are about the type value assigned to the polar expression. This gives users of the annotations the possibility to only take into account polar expressions which are assigned a type label with high confidence.

*Insubstantiality* For the annotation of private states, Wiebe et al. (2005) introduced the notion of insubstantiality. Insubstantial private states are private states that are not real, because the presupposition that the state exists is removed via the context (or the state is explicitly asserted not to exist). Because most sentiment analysis applications want to ignore insubstantial expressions of sentiment, it is useful to be able to detect insubstantiality and thus to incorporate this notion in our annotation scheme. Polar expressions conveying a sentiment that is not real, are marked using the polar expression attribute *Is\_insubstantial*. An example of an insubstantial polar expression is the phrase 'will be found guilty of gross negligence over the spill' in Fig. 8.

*Causativity* Polar expressions sometimes express causality in that they refer to a positive or negative effect being exerted on a target entity or entities by another entity. Polar expressions like this are marked using the attribute *Is\_causative*, which is discussed in detail in Sect. 3.5.

# 3.2 Modifiers

Modifiers are lexical items that cause a shift in the prior polarity of other nearby lexical items. To determine whether a polar expression conveys positive or negative sentiment, we need to take into account the presence of modifiers (Polanyi and



Shrink to grow is fine , but the problem is they re still shrinking and not yet growing .

ative targe

Fig. 10 Example of a negation modifier

Zaenen 2004; Ikeda et al. 2008; Li et al. 2010). As an example, Fig. 9 shows the polar expression 'is n't too rosy'. Here, the word 'rosy' has a positive prior polarity. Due to the presence of the negation modifier 'n't', however, the polarity of the sentiment expressed towards the target 'the outlook for the consumer electronics industry' shifts to negative.

When a modifier is detected, it is linked to the polar expression at hand and assigned a **modification type**. We define 11 types of modifiers, for each of which a description and example is provided below. The selection of modifier categories was based on the types of modifier categories considered in existing sentiment analysis research. We further altered and extended the list of modifier categories based on a preliminary annotation study conducted on a test set of newswire text. Our annotation study is the first to cover such a broad range of modifiers. Toprak et al. (2010) mark modifiers of sentiment expressions as well, but focus on the detection of negators, intensifiers and diminishers. Kessler et al. (2010) identify negators, intensifiers, diminishers, committers (expressing the author's certainty) and neutralizers (found in, for instance, hypothetical and conditional sentences).

- *negation modifiers* (see Polanyi and Zaenen 2004; Kennedy and Inkpen 2006; Wiegand et al. 2010) These modifiers invert the polarity of the expressed sentiment (e.g. 'not yet' in Fig. 10). The identification of negation modifiers is indispensable to determine whether the opinion expressed by a certain polar expression is positive or negative.
- *intensifiers*<sup>6</sup> (see Polanyi and Zaenen 2004; Kennedy and Inkpen 2006) These modifiers increase the intensity of the expressed sentiment (e.g. 'very' in Fig. 11). The detection of intensifiers helps us determine the strength of a certain polar expression.
- *diminishers*<sup>7</sup> (see Polanyi and Zaenen 2004; Kennedy and Inkpen 2006) These modifiers decrease the intensity of the expressed sentiment (e.g. 'slightly' in

<sup>&</sup>lt;sup>6</sup> These modifiers are also sometimes referred to as amplifiers (Quirk et al. 1985).

<sup>&</sup>lt;sup>7</sup> These modifiers are also sometimes referred to as downplayers or downtoners (Quirk et al. 1985).





Fig. 12 Example of a diminisher



Fig. 13 Example of a modality modifier



Fig. 14 Example of a question modifier

Fig. 12). Similar to intensifiers, they need to be taken into account when determining the strength of a certain polar expression.

- *modality modifiers* (see Polanyi and Zaenen 2004; Taboada et al. 2011) The identification of these modifiers (e.g. 'could' in Fig. 13) helps us distinguish between opinions that are asserted to be real (*realis*) and not real (*irrealis*) (Wilson et al. 2005).
- *question modifiers* (see Taboada et al. 2011) Question modifiers (e.g. 'whether' in Fig. 14) indicate that the polar expression at hand is not real i.e. insubstantial (see Sect. 3.1).
- *conditional modifiers* (see Taboada et al. 2011) Conditional modifiers (e.g. 'if wage growth accelerates' in Fig. 15) are an indication that the polar expression at hand is only real i.e. substantial under certain conditions.
- *past modifiers* Past modifiers (e.g. 'last year' in Fig. 16) do not have a direct influence on the polarity and intensity of a polar expression. However, we believe that positive and negative opinions (or facts) expressed in the past are typically perceived as less salient than those asserted to exist at this moment.
- *future modifiers* Our motivation for the detection of future modifiers (e.g. 'is forecast to' and 'in the next three years' in Fig. 17) is similar to the reason why we identify past modifiers. Opinions and facts referred to as coming about in the





Fig. 18 Example of a perspective modifier

future are believed to influence people in a different way compared to current opinions and facts.

- *perspective modifiers* When using perspective modifiers (e.g. 'for continuing investors' in Fig. 18), the writer of a text indicates that a polar expression conveys positive or negative sentiment when assessed from a certain perspective.
- *specifying modifiers* Specifying modifiers (e.g. 'by 50 %' in Fig. 19) provide more specific information about the positive or negative opinion/situation described by the polar expression at hand.
- *other modifiers* This modifier type is provided for cases where annotators are not sure which category to assign a certain modifier to.

# 3.3 Sources and source expressions

The source of a polar expression is the person or entity expressing or implying positive or negative sentiment about the target entity (or entities). Like Bethard et al. (2004), Wiebe et al. (2005), Seki et al. (2008), Somasundaran et al. (2008), Boldrini et al. (2009), Kessler et al. (2010) and Toprak et al. (2010), we identify the source, sometimes also referred to as the opinion holder, of each sentiment expression. Corpora in which the sources of sentiment expressions are marked are a



Fig. 22 Example of an opinionated source expression

valuable resource for the task of source attribution (Choi et al. 2005; Kim and Hovy 2006).

The source of a polar expression is sometimes explicited in the text. If this is the case, it is marked. The source expression is the linguistic expression linking the polar expression to its source. Identifying this source expression can be useful when linking sentiment expressions to their sources (Choi et al. 2006). Figure 20 shows the polar expression 'were "disappointing"', of which the source is 'Stephen Robertson, the BRC 's director-general'. The source expression is 'said'; this linguistic expression indicates that Stephen Robertson is the person expressing negative sentiment about the target 'the data' by means of the polar expression 'were "disappointing"'. For some polar expressions, no separate source expression is found, because the polar expression can directly be linked to its source. These polar expressions are also marked as source expressions through the use of the polar expression attribute Is\_also\_source\_expression. An example of this is shown in Fig. 21, where 'welcomed' is a private state expression that expresses positive sentiment about 'Beijing 's decision'. At the same time, it is the source expression linked to the source of this positive sentiment, 'The American Chamber of Commerce in China'.

The use of a certain source expression can be a way for the author of the text to express an opinion about the sentiment expressed by the polar expression at hand. In Fig. 22, 'BP' expresses positive sentiment about itself by means of the private state





Fig. 24 Example of a polar expression with the author of the text as its source

expression 'is recovering from the near-fatal blow it suffered in 2010'. By using the source expression 'insists', the author of this sentence implies that he might not agree with this positive sentiment, i.e. he voices an opinion about it. Source expressions like this can be marked as **opinionated** using the source expression attribute *Is\_also\_opininionated*.

The source of a polar expression is not always explicited in the text. It is possible that sentiment is expressed or implied by an impersonal source or by the author of the text. Consider the example in Fig. 23. Here, the private state expression 'fears of' expresses negative sentiment towards the target 'a downturn in mining capital expenditure'. The source of the polar expression is the impersonal 'the people', but it is not explicited in the text. In cases like this, the source of the polar expression is denoted by the polar expression attribute *Impersonal\_source*. In Fig. 24, the author of the text is the source of the polar expression 'is n't too rosy'. This can be indicated by means of the polar expression attribute *Author\_is\_source*. If a lexicalization of the author (e.g. 'I') can be found in the text, however, it is marked as the source of the polar expression. In cases where the source of a polar expression is not lexicalized in the proximity of the polar expression (i.e. only in one of the previous or following sentences), it is linked to the polar expression but considered implicit. This is indicated by means of the attribute *Implicit\_source*.

3.4 Targets and sentiment polarity

For each polar expression, at least one target is identified. This is the entity sentiment is being expressed or implied about. The annotation of these targets can be useful for topic-dependent sentiment analysis (Hu and Liu 2004; Zhuang et al. 2006; Popescu and Etzioni 2007). In Fig. 25, the polar expression 'were "disappointing"' expresses negative sentiment about the entity 'the data', which is marked as the target.

Stephen Robertson , the BRC 's director-general , said the data were " disappointing " .

Fig. 25 Example of a target





Fig. 30 Example of sentiment with ambiguous polarity expressed about a target

It is possible for a polar expression to have multiple targets. This is, for example, the case if the polar expression at hand describes a situation which can be assessed from the perspective of different entities. Figure 26 shows the polar expression 'has outperformed', which has two targets.

For each identified target, the sentiment expressed or implied about it by the polar expression is analyzed by determining two attributes: the polarity and the intensity of the sentiment. Sentiment polarity and intensity are marked on the target level, because a polar expression can express a different sentiment towards each of its targets (see Fig. 26). Like Wilson et al. (2005), our annotation scheme allows annotators to attribute several targets and corresponding sentiment polarity values to a single polar expression.

The **polarity** of the expressed sentiment can take the values positive (Fig. 27), negative (Fig. 28), other (Fig. 29) or unknown (Fig. 30). The value other is used for private states conveying a sentiment that is neither positive nor negative, such as the emotion surprise in example 29. Unlike Read and Carroll (2012); Boldrini et al.

Entity Cause PolarExpr 3. tgt ---- Entity

Meanwhile , investor worries that higher oil prices could damage the global economy gathered pace .

Fig. 31 Example of a polar resultative causative

(2009), our annotation scheme is not aimed at the detection of various types of emotions, but focuses on the distinction between positive and negative sentiment, since most sentiment analysis studies aim at polarity classification (Wilson et al. 2005; Choi and Cardie 2008). Since polar facts do not explicitly express sentiment towards a target entity, a positive or negative evaluation of this entity needs to be inferred from the factual information conveyed by these polar expressions. This requires interpretation, which can pose a problem because some polar facts can be interpreted from different perspectives. If a polar expression conveys ambiguous sentiment towards a target entity, the polarity of that sentiment is labeled unknown. Figure 30 shows the polar expression 'lowers' with its target 'interest rates'. Whether the lowering of the interest rates results in a positive or negative impression of this target, is hard to determine. A lower interest rate is positive for people who want to take out a loan, but can also be considered negative in that it may cause inflation. For this reason, the polarity attribute unknown is assigned.

The **intensity** of the sentiment can be low, medium or high.<sup>8</sup> As stated by Wiebe et al. (2005), intensity ratings are useful for "distinguishing inflammatory messages from reasoned arguments, and for recognizing when rhetoric is ratcheting up or cooling down in a particular forum. In addition, intensity ratings help in distinguishing borderline cases from clear cases of subjectivity and objectivity."

When analyzing the sentiment expressed by a polar expression, modifiers affecting the polarity and intensity of the sentiment are taken into account (as can be seen in Fig. 28).

#### 3.5 Causativity

Private state or polar fact expressions sometimes express causativity in that they refer to a positive or negative effect being exerted on a target entity or entities by another entity. In Fig. 31, for instance, the private state expression 'could damage' indicates that a negative effect is exerted on the target 'the global economy' by the entity 'higher oil prices'. We refer to private state and polar fact expressions like this as polar resultative causatives. In contrast to simple causatives, which only refer to the causal link between two entities, resultative causatives are causative constructions that refer to a causal link plus a part of the resulting situation (Girju 2003)<sup>9</sup>. If a causative construction is resultative, and the resulting situation referred to by that resultative causative results in a positive or negative impression of the

<sup>&</sup>lt;sup>8</sup> In brat, the intensity of the positive or negative sentiment is denoted by the number of plus or minus signs accompanying the arrow pointing from the polar expression to the target at hand.

<sup>&</sup>lt;sup>9</sup> While Girju (2003) uses the term resultative causative for verbal constructions, we also identify other causative constructions.



Fig. 32 Example of a coreferential relation

Entity has medium positive target Polar Expression 0. BP said that by 2014 its operating cash flow would have increased by 50 per cent compared with 2011.

#### Fig. 33 Example of a feature relation

target entity or entities (i.e. expresses a private state or polar fact), it is a polar resultative causative.

If a private state or polar fact expression is a polar resultative causative, it is marked using the attribute *Is\_causative*. Additionally, the entity exerting a positive or negative effect on the target entity or entities (for instance 'higher oil prices' in Fig. 31) is annotated as the **cause** and linked to the private state or polar fact expression at hand.

In related work, Deng et al. (2013) annotated so-called benefactive and malefactive events, also referred to as goodFor/badFor (gfbf) events, which positively/negatively affect entities.

# 3.6 Coreferential relations

The source, target or cause entities of a polar expression can be referred to in a text with multiple, coreferring linguistic expressions. If this is the case, we mark the linguistic expression that is in a syntactic relation with the polar expression or is closest to it. Consequently, expressions marked as sources, targets or causes can be anaphora, items with little or no intrinsic meaning. These sources, targets and causes are linked to the closest meaningful antecedent by means of a coreferential relation. Coreferential relations are also incorporated in the annotation schemes of Kessler et al. (2010) and Toprak et al. (2010). The use of coreference resolution can help improve the performance of topic-dependent sentiment analysis approaches. An example of a coreferential relation can be found in the sentence in Fig. 32. Here, the pronoun 'it' is marked as the target of the polar expression 'is recovering from the near-fatal blow it suffered in 2010'. It is linked to its antecedent 'BP' through a coreferential relation.

# 3.7 Feature relations

Sometimes, the sentiment being expressed or implied about a certain target is also indirectly aimed at another entity, because the target is a feature of this entity, i.e. a part or property of it. An example of a feature relation can be found in Fig. 33. The polar fact expression 'would have increased by 50 % compared with 2011' results in a positive impression of the target entity 'its operating cash flow' and, furthermore, of 'BP', since the target is a feature of this company. For this reason, these two entities are linked by means of a feature relation. Feature-based sentiment

analysis has been an important topic of interest in the domain of product reviews (Hu and Liu 2004; Zhuang et al. 2006; Popescu and Etzioni 2007; Pontiki et al. 2014), where researchers detect and analyze the sentiment expressed towards the features of a certain topic i.e. product (e.g. the picture quality and size of a digital camera). Kessler et al. (2010), for instance, have annotated feature relations in a corpus of blog posts about automobiles. However, as can be seen in Fig. 33, topics in other text genres (e.g. companies in financial news) can also have features (for instance the profits and liquidity of a company). Feature relations can be annotated for targets, causes and sources of polar expressions.

### 4 Annotation study

# 4.1 Data

The sentiment annotation scheme described in Sect. 3 was applied to a corpus of English and Dutch financial newswire text<sup>10</sup>. We hypothesized that this type of data contains explicit expressions of sentiment (e.g. opinions of financial analysts about certain companies, markets or events) as well as implicit ones (viz. factual news content resulting in positive or negative opinions), making it an appropriate text genre for the evaluation of our annotation scheme. English news articles were taken from The Financial Times, Dutch articles from the Belgian economic newspaper De Tijd. They were sampled between November 2004 and May 2005, and between November 2011 and May 2012, and include macroeconomic, sector and company news coverage.

Each document was split into sentences and tokenized using the LeTs Preprocess toolkit (Van de Kauter et al. 2013). Table 1 gives an overview of the amount of material that has currently been annotated for both languages. The inter-annotator agreement study discussed in Sect. 4.3 was conducted on a subset of this corpus. The statistics presented in Sect. 4.4 were calculated on the full corpus.

For both languages together, 185 documents and 4790 sentences have been annotated. The corpus size approaches that of the fully annotated section of the MPQA v2.0 corpus, which contains 344 documents and 5957 sentences, but which, contrary to our annotated dataset, is a monolingual corpus.

#### 4.2 Annotation process

The sentiment annotation scheme we developed was applied to the English and Dutch financial newswire corpus by means of the brat rapid annotation tool (Stenetorp et al. 2012), a web-based tool for text annotation. It is freely available and open source and allows for the annotation of text spans as well as relations. Examples of excerpts from the corpus annotated with polar expressions and related elements using the brat tool can be found in Sect. 3.

Because the annotation of explicit and implicit sentiment in particular requires a substantial amount of interpretation and thus a substantial knowledge of the

<sup>&</sup>lt;sup>10</sup> The resulting annotations will be made available for research purposes.

-			
Language	No. of documents	No. of sentences	No. of words
English	87	1685	44,645
Dutch	98	3105	47,650

 Table 1
 Properties of the full annotated corpus

language at hand, the annotations for both English and Dutch were performed by annotators mastering these languages at a high level. All annotators of the newswire corpus had a background in linguistics and were native speakers of English or Dutch, or (higher education) students in the language at hand. Before starting on the annotation of the corpus, the annotators went through a training phase of about 3 working days in which they learned to apply the annotation scheme based on the guidelines detailed in Van de Kauter et al. (2014). This training phase also familiarized annotators with the brat annotation tool. They annotated a number of example texts (also taken from financial newspapers), were able to ask questions about the annotation scheme and tool, and received feedback about their annotations.

After completing their training, the annotators began to annotate articles from the newswire corpus describe above. Each text was annotated by two annotators. First, these annotators analyzed each article independently. In a second step, they compared their individual annotations and only retained those annotations on which they both agreed. In future work, these consensus annotations will be used as training data for automatic sentiment analysis in financial news articles.

# 4.3 Inter-annotator agreement study

To assess whether the sentiment annotation scheme described in Sect. 3 can be reliably applied to real-world textual data, we conducted an inter-annotator agreement study. A small subset of the news articles collected for English and Dutch was independently annotated by three annotators<sup>11</sup>, in order to avoid the results being dependent on the differences between two individual annotators. The individual annotations of these annotators were then compared to each other.

Table 2 shows the size of the sub-corpus used for the agreement study.

To measure agreement between the three annotators, we average the interannotator agreement scores of all three annotator pairs. These scores are calculated for various annotation tasks. First, we measure inter-annotator agreement for polar expression identification, which is the main task defined in our annotation scheme. For each of the polar expressions identified by both annotators of a pair, agreement scores are then calculated for subsequent annotation tasks such as polar expression type classification, target identification, source identification, etc. This allows us to measure the extent to which polar expressions are further analyzed in the same way by different annotators.

<sup>&</sup>lt;sup>11</sup> Note that not all annotators for Dutch (three Dutch native speakers) are the same persons who annotated the English corpus (two Master's students in English and one English native speaker). Only one annotator participated in the inter-annotator agreement study for both English and Dutch.

Language	No. of documents	No. of sentences	No. of words
English	4	119	2808
Dutch	6	179	2771

Table 2 Properties of the sub-corpus used for the inter-annotator agreement study

Inter-annotator agreement is measured using four different metrics, depending on the type of annotation task at hand. For classification tasks (e.g. assigning polar expressions to a category of the subjectivity continuum), Cohen's kappa (Cohen 1960) and Krippendorff's alpha (Krippendorff 1970, 2004) are used, two measures which correct for chance agreement. Kappa is used for nominal classifications (e.g. modifier classification), whereas Krippendorff's alpha is applied in cases where not all disagreements should be treated equally, with squared distance as a distance metric (e.g. for polar expression classification by subjectivity type). For identification tasks, accuracy or F-score ( $\beta = 1$ ) (van Rijsbergen 1979) is calculated. Accuracy is used for classification tasks where only one item can be identified (this is for example the case for source identification). For identification tasks permitting the identification of multiple items (e.g. modifier identification, target identification), F-score is the metric of choice, since annotators do not always identify the same set of items and agreement measures such as kappa or alpha can therefore not be used. F-scores are calculated by considering the annotations of one annotator as the gold standard, and measuring precision and recall of the second annotator on that gold standard set of annotations. This is equivalent to averaging precision or recall in both directions. The same approach is adopted by Wiebe et al. (2005).

When discussing the inter-annotator agreement results, kappa and alpha scores are interpreted using the conventions proposed by Landis and Koch (1977) and Krippendorff (1980). The former consider 0.6 and 0.8 the minimum values for resp. substantial and almost perfect agreement, whereas the latter regards 0.67 and 0.8 to be sufficient for resp. tentative conclusions and good reliability. However, we should note that, as discussed by Artstein and Poesio (2008), in more recent work, 0.8 or even 0.9 is considered the threshold for acceptable reliability.

# 4.3.1 Polar expressions

Polar expressions are the central component of our annotation scheme and, therefore, also the starting point for this inter-annotator agreement study. Tables 3 and 4 show the number of polar expressions identified by each of the annotators in resp. the English and Dutch sub-corpora presented in Table 2.

Annotator	No. of polar expressions
A	201
В	217
C	183

Table 3 Number of polar expressions identified by the annotators in English

Table	4	Number	of	polar	expressions	identified	by	the	annotators	in	Dutc	h
-------	---	--------	----	-------	-------------	------------	----	-----	------------	----	------	---

Six months ago , Bob Dudley , BP chief executive , stated that the company had reached a "

Polar Expression

 definite turning point " in its fortunes after the ravages of the Deepwater Horizon-disaster in 2010 .

 Six months ago , Bob Dudley , BP chief executive , stated that the company had reached a "
 PolarExpF

definite turning point " in its fortunes after the ravages of the Deepwater Horizon-disaster in 2010 .

Fig. 34 Example of a fuzzy polar expression match

To measure the extent to which the annotators have identified the same polar expressions, we compare the set of annotations made by each of the annotators to the annotations of the two other annotators and look for matching polar expressions. Similar to Wiebe et al. (2005), we do not focus on boundary detection when comparing polar expressions marked by two annotators, since this issue is less crucial for the task of sentiment analysis. It is most important that both annotators have annotated the same general expression. Hence, we do not only take into consideration exact polar expression matches, but also look for fuzzy matches. An example of a fuzzy match between polar expressions marked by two annotators can be found in Fig. 34, where the polar expressions 'the ravages of' and 'the ravages' (marked by resp. annotators B and C for English) are considered to be the same expression.

Tables 5 and 6 show the number of matching polar expressions found for each of the annotator pairs.

To measure inter-annotator agreement for the identification of polar expressions, we calculate F-score ( $\beta = 1$ ) for each annotator pair by considering the polar expressions of the first annotator as the gold standard, and measuring precision and recall of the second annotator on that gold standard set of expressions. Recall and precision are calculated by dividing the number of matching polar expressions by the number of polar expressions identified by resp. the first and second annotator. The average F-scores for English and Dutch can be found in Table 7. Similar scores are obtained for both languages, viz. 0.67 for English and 0.66 for Dutch.

Because most of the existing sentiment annotation schemes focus on the detection of explicit expressions of sentiment, we also want the measure the extent to which implicit sentiment in particular can be identified in a reliable way. Table 7 therefore also contains separate inter-annotator agreement scores for the identification of explicit and implicit sentiment expressions, which indicate that for implicit expressions, agreement is lower. We believe this observation can be explained by

Annotator pair	No. of matching polar expressions
A–B	137
A–C	133
B-C	134

Table 5 Number of matching polar expressions identified in English

Table 6 Number of matching polar expressions identified in Dutch

Annotator pair	No. of matching polar expressions
A–B	138
A–C	117
В-С	129

Table 7 Inter-annotator agreement for polar expression identification

Language	Overall	Explicit PEs	Implicit PEs
English	0.67	0.70	0.64
Dutch	0.66	0.72	0.58

the fact that the detection of implicit sentiment requires more interpretation, i.e. the inference of a positive or negative impression from factual information.

For the set of polar expressions marked by both annotators of an annotator pair, we calculate inter-annotator agreement for the classification of these polar expressions by subjectivity type. In other words, we measure the extent to which annotators assign a polar expression to the same position on the subjectivity continuum. Since the starting point is the set of polar expressions annotated by both annotators, agreement for this task can be calculated by means of Krippendorff's alpha or Cohen's kappa. Agreement is measured by means of Krippendorff's weighted alpha measure, with squared distance used as a distance metric, since not all disagreements for this classification task should be treated equally. On the subjectivity continuum (0-1-2-3), the value 1, for instance, is closer to 0 than the value 3. Table 8 shows the agreement scores obtained for English and Dutch. Alpha scores of resp. 0.65 and 0.54 are obtained. According to Krippendorff (1980), this means that tentative conclusions can be drawn from the annotation for English, but not for Dutch. Following the conventions proposed by Landis and Koch (1977), substantial agreement is obtained for English, whereas the agreement score

Table 8 Inter-annotator agreement for polar expression classification by subjectivity type

Language	Alpha score
English	0.69 α
Dutch	0.59 α

achieved for Dutch indicates moderate agreement. In other words, labeling a polar expression as either objective (implicit) or subjective (explicit) is a difficult task.

After having determined the set of polar expressions identified by both annotators of an annotator pair, we measure the extent to which both annotators further analyze these polar expressions in the same way. Inter-annotator agreement scores for target, source, source expression and modifier annotation are calculated for the sets of polar expressions described in Tables 5 and 6.

# 4.3.2 Targets and sentiment polarity

For each of the polar expressions marked by both annotators of an annotator pair, we investigate whether the annotators have identified the same targets. Since a polar expression can have multiple targets, inter-annotator agreement is measured by calculating F-score ( $\beta = 1$ ). The targets marked by the first annotator are considered to be the gold standard. Precision and recall of the second annotator are then measured on that gold standard set of targets. Strict as well as relaxed inter-annotator agreement scores are calculated. To measure agreement in a strict manner, we look for matches between direct targets of the polar expressions. When calculating relaxed agreement scores, we also take into account entities linked to the targets of a polar expression by means of a coreference or feature relation. Table 9 shows the inter-annotator agreement scores for target identification. Strict agreement scores of 0.89 and 0.88 are obtained for resp. English and Dutch. These scores rise to over 0.90 in a relaxed scenario. This means that annotators achieve very high agreement when assigning targets to polar expressions they have both identified, even if we do not take into account coreference and feature relations.

Next, we assess whether the annotators agree on the polarity of the sentiment expressed towards the targets identified by both annotators of an annotator pair. First, a kappa score for polarity classification is calculcated on a coarse-grained level (positive-negative-other-unknown). Second, agreement is measured on a fine-grained level by means of Krippendorff's alpha metric, using squared distance as a distance metric. For this fine-grained evaluation of the annotations, the intensity of the sentiment expressed towards the targets was taken into account, which resulted in the following scale: -3 (high negative), -2 (medium negative), -1 (low negative), 0 (other or unknown), 1 (low positive), 2 (medium positive), 3 (high positive). The inter-annotator agreement scores for polarity classification on a coarse- and fine-grained level can be found in Table 10. According to Landis and Koch (1977), these scores indicate almost perfect agreement for both languages on a coarse-grained as well as on a fine-grained level. Following the conventions of Krippendorff (1980), good reliability is obtained.

Language	Strict	Relaxed
English	0.89	0.96
Dutch	0.88	0.92

 Table 9 Inter-annotator agreement for target identification

The round of a ground of polarity encontrained			
Language	Coarse-grained	Fine-grained	
English	0.87 к	0.93 α	
Dutch	0.83 κ	0.89 α	

Table 10 Inter-annotator agreement for polarity classification

# 4.3.3 Sources and source expressions

Since only one source is assigned to each polar expression, inter-annotator agreement for source identification is measured by calculating accuracy scores. The number of polar expressions for which two annotators have identified the same source is divided by the total number of polar expressions annotated by both annotators. A source can be explicitly marked in the text, or it can be specified by a source attribute of the polar expression (viz. *Author\_is\_source* or *Impersonal\_source*). Similar to target identification, inter-annotator agreement for source identification is measured in a strict and relaxed manner. As shown in Table 11, fairly high relaxed as well as strict scores are achieved for both English and Dutch.

For the set of polar expressions assigned the same source by both annotators of an annotator pair, we also measure inter-annotator agreement for source expression identification by calculating accuracy scores (Table 12). Like sources, source expressions can be explicitly marked in the text, or they can be specified by the attribute *Is\_also\_source\_expression* of the polar expression at hand. The average accuracy scores obtained for both languages are very high, indicating that annotators are able to reliably identify the source expression of a polar expression.

# 4.3.4 Modifiers

Finally, we measure inter-annotator agreement for the identification of modifiers. Since the polarity and intensity of the sentiment expressed by a polar expression can be modified by multiple lexical items, agreement is calculated using F-score ( $\beta = 1$ ). Table 13 shows the overall inter-annotator agreement scores for modifier

Language	Strict	Relaxed		
English	0.83	0.84		
Dutch	0.81	0.82		

Table 11 Inter-annotator agreement for source identification

<b>Table 12</b> Inter-annotator agreement for source expression identification	
Language	Accuracy
English	0.99
Dutch	0.95

Table 12 Inter-annotator agreement for source expression identification

Туре	English	Dutch
Overall	0.60	0.69
Negation	0.75	0.85
Intensifiers	0.61	0.65
Diminishers	0.63	0.33
Modality	0.61	0.49
Past	0.66	0.78
Future	0.55	0.68
Question	0.00	0.13
Conditional	0.89	1.00
Perspective	0.36	0.55
Specifying	0.55	0.67
Other	0.40	0.48

Table 13 Inter-annotator agreement for modifier identification

Table 14 Inter-annotator agreement for modifier classification

Language	Карра
English	0.89 к
Dutch	0.83 κ

 Table 15
 Number of annotations in the full English and Dutch corpora

	English	Dutch
Polar expressions	3787	4323
Average number of polar expressions per word	0.09	0.09
Modifiers	3478	3337
Average number of modifiers per polar expression	0.92	0.77
Source expressions	758	304
Average number of source expressions per polar expression	0.20	0.07

identification as well as separate F-scores for each type of modifier. The wide spread of the results for the different modifier types and the differences across the two languages can be explained by the small size of the sub-corpus used for the interannotator agreement study. In this corpus, the absolute number of occurrences for several individual categories is limited; a small absolute number of disagreements for these modifier types can thus lead to a significant decrease in F-score. For instance, in the Dutch corpus, only 3, 0 and 2 question modifiers were identified by resp. annotators A, B and C. In the English corpus, annotator B marked 2 question modifiers, whereas no modifiers of this type were identified by annotators A and C. The latter results in an inter-annotator agreement score of 0.00 F for English.

To measure the extent to which modifiers are assigned to the same modifier category, kappa scores for modifier classification are calculated on the sets of modifiers identified by both annotators of an annotator pair. As can be seen in Table 14, kappa scores of 0.89 and 0.83 are obtained for resp. English and Dutch, indicating that good reliability is obtained according to the conventions of Krippendorff (1980). When considering the conventions proposed by Landis and Koch (1977), the conclusion can be drawn that almost perfect agreement is achieved for modifier classification in both languages.

Overall, we can conclude from the inter-annotator agreement study that our annotation scheme allows reliable identification of polar expressions, that polar expression type classification is somewhat ambiguous, and that targets, sources and modifiers are consistently annotated.

4.4 Corpus statistics

In this section, we present statistics on the entire annotated corpus, as described in Sect. 4.1 and Table 1. Calculations are based on the consensus annotations from two annotators. We should note that the size and the specific domain of the corpus do not allow us to draw general conclusions about the differences between English and Dutch with regards to the expression of sentiment, since the differences found



Fig. 35 Relative frequency of polar expressions by subjectivity type

in the corpus could be attributable to the writing style employed in this specific domain or in these specific newspapers.

The absolute numbers of annotations for English and Dutch are shown in Table 15. Relative to the number of words, both corpora have a similar amount of polar expressions. We can therefore assume that the corpora are comparable with regards to the amount of polarity that is encountered. In the English corpus, a greater number of polar expressions is modified, and more of them can be linked to a source by means of a source expression.

### 4.4.1 Polar expressions

The distribution of polar expressions on the subjectivity scale is presented in Fig. 35. It is immediately apparent that polarity is not only present in subjective expressions: for English, less than half of the polar expressions are annotated as



Fig. 36 Relative frequency of target relation intensity, by coarse polar expression type (*top* English, *bottom* Dutch)



#### Polarity of targets associated with polar expressions

Fig. 37 Relative frequency of polar orientation of target links associated with polar expressions



Fig. 38 Relative frequency of modification types



Fig. 39 Relative frequency of modification type, by coarse polar expression type (*top* English, *bottom* Dutch)

subjective (types 2 and 3), and for Dutch, fewer than 60 %. We can state with considerable confidence that, for a comprehensive view on polarity in text, both explicit and implicit expressions have to be accounted for. Failing to annotate implicit polarity would result in a substantial amount of the expressed polarity being missed, at least in our data.

We also find that the Dutch corpus is skewed more towards subjective expressions, whereas the English data presents more polar facts. Annotators score most polar expressions at the extremes of the subjectivity continuum.

# 4.4.2 Targets and sentiment polarity

It might be that explicit expressions are associated with more intense polarity than implicit ones. Figure 36 plots polar expression type against the intensity of its target relation. For English, there is no significant difference between types, so polarity stemming from objective expressions tends to be no less intense. For Dutch, on the other hand, subjective expressions trigger somewhat higher intensities. Overall, annotators most frequently link targets with medium intensity. The polar orientations of target links associated with polar expressions are shown in Fig. 37. The Dutch data contains more positive than negative links, the English data is balanced. Both languages present a small number of polar expressions to which targets are linked with opposing or other polarities.

# 4.4.3 Modifiers

Figure 38 presents the distribution of modification types. The distributions are similar between English and Dutch, with past, intensifiers and specifying being the most common modification types. In the Dutch data, other, negation and question types occur proportionally more frequently. For English, future modifiers are more frequent.

Figures 39 and 40 show the correlation between the type of a modifier and the type and target link orientation of its associated polar expression, respectively.

For both languages, we can observe that specifying modifiers occur almost exclusively with objective polar expressions, whereas negation and modality modifiers are rarely associated with them.



Fig. 40 Relative frequency of modification type, by polar expression polarity orientation (*top* English, *bottom* Dutch)

Modifier types appear to be independent from polarity orientation, except for negation, which proportionally occurs only rarely with positive target links. This suggests that positive expressions are switched to become negative much more often than the other way around. In our corpus, the presence of a negation marker can therefore be considered indicative of an utterance with negative polarity.

# 5 Conclusions and future work

We have developed and described a new scheme for the fine-grained annotation of polar sentiment in text. It defines polar expressions, which may be subjective (private states i.e. explicit expressions of sentiment) or objective (polar facts i.e. implicit expressions of sentiment) in nature. These expressions are annotated on a sub-sentential level. Furthermore, the scheme allows annotation of sources, target attribution and polar expression modification. These entities can be linked to other entities with coreferential or feature relations. To our knowledge, none of the other existing sentiment annotation schemes combine the identification of explicit and implicit sentiment expressions with the fine-grained analysis of these expressions and their components below sentence level.

We tested the applicability of the annotation scheme on real-world textual data using a corpus of Dutch and English financial news. First, we performed an interannotator agreement study between three annotators, on a small subset of this corpus. The identification of polar expressions proved to be non-trivial, especially for implicit occurrences of sentiment (polar facts). However, agreement is around 0.66 F-score, and once annotators agree on the presence of a polar expression, further annotations can be done with moderate to very high agreement (viz. source, source expression and target attribution, and polarity classification). For modifiers, agreement is moderate for identification and high for classification. A particularly difficult task is the classification of polar expressions as objective or subjective.

Based on these findings, we can state that the fully annotated corpora are valuable resources for research on coarse- and fine-grained automatic sentiment analysis, with 44,645 words of English and 47,650 words of Dutch economic newswire text. The proportion of subjective polar expressions is below 60% for both languages. This validates the claim that implicit as well as explicit polar expressions need to be annotated for a comprehensive account of polarity.

Since overall, similar inter-annotator agreement scores were obtained for the English and Dutch corpora, we believe our annotation scheme can be reliably applied to different languages. In future research, we will investigate the suitability of our annotation scheme for domains other than financial newswire. The results of a preliminary inter-annotator agreement study conducted on a collection of Dutch tweets about the Belgian local elections of 2012 indicate that for a corpus of political Twitter messages, agreement scores similar to those measured for the financial newswire corpora can be obtained. This suggests that our sentiment annotation scheme is applicable to various domains and text genres.

In future work, we will also make use of the annotated corpora as training and evaluation datasets for the automatic detection of explicit and implicit expressions of positive and negative sentiment in text. Implicit sentiment is not only found in financial news, but can also occur in other text types. Balahur et al. (2011a, b), for example, capture implicit expressions of sentiment in ISEAR, a corpus in which student respondents describe situations in which they experienced certain emotions (e.g. anger, joy, etc.).

Since the created corpora do not only provide us with annotations of polar sentiment expressions, they can be used for more than sentiment detection and polarity classification. The annotations of other components such as sources, source expressions, targets and modifiers make the datasets a valuable resource for other sentiment analysis tasks as well. In the near future, we will, among others, focus on the task of target attribution for topic-dependent sentiment analysis. Detecting the target of sentiment expressions would allow us to extract only the sentiment relevant to a given subject of interest. This can be useful, for example, for the analysis of sentiment expressed about certain product features in product reviews (Hu and Liu 2004; Zhuang et al. 2006), or for the detection of sentiment related to specific companies, sectors or events in the financial domain (O'Hare et al. 2009).

Acknowledgments This research was conducted in the framework of SentiFM (http://www. lt3.ugent.be/en/projects/sentifm/) (Sentiment mining for Financial Markets), a project funded by a Ph.D. grant of the Agency for Innovation by Science and Technology (IWT), and the HOF project SubTLe (http://www.lt3.ugent.be/en/projects/subtle/) (Subjectivity Tagging and Learning).

# References

- Abdul-Mageed, M., & Diab, M. T. (2011). Subjectivity and sentiment annotation of modern standard Arabic newswire. In *Proceedings of the 5th linguistic annotation workshop (LAW V)* (pp. 110–118). Portland, Oregon, USA.
- Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. Computational Linguistics, 34(4), 555–596.
- Asher, N., Benamara, F., & Mathieu, Y. Y. (2008). Categorizing opinion in discourse. In Proceedings of the 18th European conference on artificial intelligence (ECAI 2008) (pp. 835–836). Patras, Greece.
- Balahur, A., Hermida, J. M., & Montoyo, A. (2011a). Detecting emotions in social affective situations using the emotinet knowledge base. In *Proceedings of the 8th international conference on advances* in neural networks, Part III (ISNN 2011), volume 6677 of lecture notes in computer science (pp. 611–620). Springer.
- Balahur, A., Hermida, J. M., Montoyo, A., & Muñoz, R. (2011b). EmotiNet: A knowledge base for emotion detection in text built on the appraisal theories. In *Proceedings of the 19th conference on* applications of natural language to information systems (NLDB 2011), volume 6716 of lecture notes in computer science (pp. 27–39). Springer.
- Banfield, A. (1982). Unspeakable sentences. Boston: Routledge and Kegan Paul.
- Bermingham, A., & Smeaton, A. F. (2009). A study of inter-annotator agreement for opinion retrieval. In Proceedings of the 32nd annual international ACM SIGIR conference on research and development in information retrieval (SIGIR 2009) (pp. 784–785). Boston, Massachusetts, USA.
- Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., & Jurafsky, D. (2004). Automatic extraction of opinion propositions and their holders. In *Proceedings of the AAAI spring symposium on exploring attitude and affect in text: Theories and applications* (pp. 20–27). Palo Alto, California, USA.
- Boldrini, E., Balahur, A., Martínez-Barco, P., & Montoyo, A. (2009). EmotiBlog: An annotation scheme for emotion detection and analysis in non-traditional textual genres. In *Proceedings of the 2009 international conference on data mining (DMIN 2009)* (pp. 491–497). Las Vegas, Nevada, USA.

- Boldrini, E., Balahur, A., Martínez-Barco, P., & Montoyo, A. (2012). Using EmotiBlog to annotate and analyse subjectivity in the new textual genres. *Data Mining and Knowledge Discovery*, 25(3), 603–634.
- Breck, E., Choi, Y., & Cardie, C. (2007). Identifying expressions of opinion in context. In *Proceedings of the 20th international joint conference on artificial intelligence (IJCAI-2007)* (pp. 2683–2688). Hyderabad, India.
- Choi, Y., Breck, E., & Cardie, C. (2006). Joint extraction of entities and relations for opinion recognition. In Proceedings of the 2006 conference on empirical methods in natural language processing (EMNLP 2006) (pp. 431–439). Sydney, Australia.
- Choi, Y., & Cardie, C. (2008). Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the 2008 conference on empirical methods in natural language processing (EMNLP 2008)* (pp. 793–801). Honolulu, Hawaii, USA.
- Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. In *Proceedings of the conference on human language* technology and empirical methods in natural language processing (HLT-EMNLP 2005) (pp. 355–362). Vancouver, British Columbia, Canada.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.
- Dabrowski, M., Acton, T., Jarzebowski, P., & O'Riain, S. (2010). Improving customer decisions using product reviews—CROM—Car Review Opinion Miner. In Proceedings of the 6th international conference on Web information systems and technologies, Volume 1 (WEBIST 2010) (pp. 354–357). Valencia, Spain.
- Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web (WWW 2003)* (pp. 519–528). Budapest, Hungary.
- Deng, L., Choi, Y., & Wiebe, J. (2013). Benefactive/malefactive event and writer attitude annotation. In Proceedings of the 51st annual meeting of the association for computational linguistics (Vol. 2: Short Papers) (pp. 120–125). Sofia, Bulgaria.
- Desmet, B., & Hoste, V. (2014). Recognising suicidal messages in Dutch social media. In Proceedings of the 9th international conference on language resources and evaluation (LREC 2014) (pp. 830–835). Reykjavik, Iceland.
- Devitt, A., & Ahmad, K. (2007). Sentiment polarity identification in financial news: A cohesion-based approach. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 984–991). Prague, Czech Republic.
- Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In Proceedings of the 2008 international conference on Web search and data mining (WSDM 2008) (pp. 231–240). Palo Alto, California, USA.
- Drury, B. & Almeida, J. J. (2011). Identification of fine-grained feature-based event and sentiment phrases from business news stories. In *Proceedings of the international conference on Web intelligence, mining and semantics (WIMS 2011)*. Sogndal, Norway.
- Esuli, A., & Sebastiani, F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th conference on language resources and evaluation (LREC 2006)* (pp. 417–422). Genoa, Italy.
- Feng, S., Kang, J. S., Kuznetsova, P., & Choi, Y. (2013). Connotation Lexicon: A dash of sentiment beneath the surface meaning. In *Proceedings of the 51st annual meeting of the association for computational linguistics (Vol. 1: Long Papers)* (pp. 1774–1784). Sofia, Bulgaria.
- Ferguson, P., O'Hare, N., Davy, M., Bermingham, A., Tattersall, S., Sheridan, P., et al. (2009). Exploring the use of paragraph-level annotations for sentiment analysis of financial blogs. In *Proceedings of the 1st workshop on opinion mining and sentiment analysis (WOMSA 2009)* (pp. 42–52). Seville, Spain.
- Girju, R. (2003). Automatic detection of causal relations for question answering. In Proceedings of the ACL 2003 workshop on multilingual summarization and question answering, Vol. 12 (MultiSumQA 2003) (pp. 76–83). Sapporo, Japan.
- Halliday, M. (1994). An introduction to functional grammar. London: Edward Arnold.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In Proceedings of the 10th ACM SIGKDD international conference on knowledge discovery and data mining (KDD 2004) (pp. 168–177). Seattle, Washington, USA.

- Ikeda, D., Takamura, H., Ratinov, L.-A., & Okumura, M. (2008). Learning to shift the polarity of words for sentiment classification. In *Proceedings of the 3rd international joint conference on natural language processing (IJCNLP 2008)* (pp. 296–303). Hyderabad, India.
- Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2), 110–125.
- Kessler, J. S., Eckert, M., Clark, L., & Nicolov, N. (2010). The ICWSM 2010 JDPA sentiment corpus for the automotive domain. In 4th international AAAI conference on weblogs and social media data workshop challenge (ICWSM-DWC 2010). Washington, DC, USA.
- Kim, S.-M., & Hovy, E. (2005). Automatic detection of opinion bearing words and sentences. In Companion volume to the proceedings of the 2nd international joint conference on natural language processing (IJCNLP 2005) (pp. 61–66). Jeju Island, Korea.
- Kim, S.-M., & Hovy, E. (2006). Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the workshop on sentiment and subjectivity in text (SST 2006)* (pp. 1–8). Sydney, Australia.
- Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the OMG!. In Proceedings of the 5th international AAAI conference on weblogs and social media (ICWSM 2011) (pp. 538–541) Barcelona, Spain.
- Krippendorff, K. (1970). Estimating the reliability, systematic error and random error of interval data. Educational and Psychological Measurement, 30, 61–70.
- Krippendorff, K. (1980). Content analysis: An introduction to its methodology, chapter 12. Beverly Hills: Sage.
- Krippendorff, K. (2004). Content analysis: An introduction to its methodology (2nd ed.). Thousand Oaks: Sage.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Li, S., Lee, S. Y. M., Chen, Y., Huang, C.-R., & Zhou, G. (2010). Sentiment classification and polarity shifting. In *Proceedings of the 23rd international conference on computational linguistics (Coling* 2010) (pp. 635–643). Beijing, China.
- Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies. San Rafael: Morgan & Claypool.
- Macdonald, C., Ounis, I., & Soboroff, I. (2007). Overview of the TREC-2007 blog Track. In Proceedings of the 16th text REtrieval conference (TREC 2007) (pp. 31–43). Gaithersburg, Maryland, USA.
- Martin, J. R., & White, P. R. (2005). *The language of evaluation: Appraisal in English*. Hampshire/New York: Palgrave Macmillan.
- Musat, C., & Trausan-Matu, S. (2010). The impact of valence shifters on mining implicit economic opinions. In Proceedings of the 14th international conference on artificial intelligence: Methodology, systems, and applications (AIMSA 2010), volume 6304 of Lecture Notes in Computer Science (pp. 131–140). Springer.
- Nakov, P., Rosenthal, S., Kozareva, Z., Stoyanov, V., Ritter, A., & Wilson, T. (2013). SemEval-2013 Task 2: Sentiment analysis in Twitter. In *Proceedings of the 7th international workshop on semantic* evaluation (SemEval 2013) (pp. 312–320). Atlanta, Georgia, USA.
- Nigam, K., & Hurst, M. (2004). Towards a robust metric of opinion. In *Proceedings of the AAAI spring symposium on exploring attitude and affect in text: Theories and applications* (pp. 598–603). Palo Alto, California, USA.
- O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C., et al. (2009). Topicdependent sentiment analysis of financial blogs. In *Proceedings of the 1st international CIKM workshop on topic-sentiment analysis for mass opinion measurement (TSA 2009)* (pp. 9–16). Hong Kong, China.
- Ounis, I., de Rijke, M., Macdonald, C., Mishne, G., & Soboroff, I. (2006). Overview of the TREC-2006 blog track. In *Proceedings of the 15th text REtrieval conference (TREC 2006)* (pp. 17–31). Gaithersburg, Maryland, USA.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on empirical methods in natural language processing, volume 10 (EMNLP 2002)* (pp. 79–86). Philadelphia, Pennsylvania, USA.

- Polanyi, L., & Zaenen, A. (2004). Contextual valence shifters. In *Proceedings of the AAAI spring symposium on exploring attitude and affect in text: Theories and applications* (pp. 106–111). Palo Alto, California, USA.
- Pontiki, M., Papageorgiou, H., Galanis, D., Androutsopoulos, I., Pavlopoulos, J., & Manandhar, S. (2014). SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)* (pp. 27–35). Dublin, Ireland.
- Popescu, A.-M., & Etzioni, O. (2007). Natural language processing and text mining, chapter Extracting product features and opinions from reviews. London: Springer.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). A comprehensive grammar of the English language. London: Longman.
- Read, J., & Carroll, J. (2012). Annotating expressions of appraisal in English. Language Resources and Evaluation, 46, 421–447.
- Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In Proceedings of the 2003 conference on empirical methods in natural language processing (EMNLP 2003) (pp. 105–112). Sapporo, Japan.
- Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., & Harabagiu, S. M. (2012). EmpaTweet: Annotating and detecting emotions on Twitter. In *Proceedings of the 8th international conference on language resources and evaluation (LREC 2012)* (pp. 3806–3813). Istanbul, Turkey.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729.
- Seki, Y., Evans, D. K., Ku, L.-W., Chen, H.-H., Kando, N., & Lin, C.-Y. (2007). Overview of opinion analysis pilot task at NTCIR-6. In Proceedings of the workshop meeting of the National Institute of Informatics (NII) test collection for information retrieval systems (NTCIR-6) (pp. 265–278). Tokyo, Japan.
- Seki, Y., Evans, D. K., Ku, L.-W., Sun, L., Chen, H.-H., & Kando, N. (2008). Overview of multilingual opinion analysis task at NTCIR-7. In Proceedings of the 7th NTCIR workshop meeting on evaluation of information access technologies (NTCIR-7) (pp. 185–203). Tokyo, Japan.
- Somasundaran, S., Ruppenhofer, J., & Wiebe, J. (2008). Discourse level opinion relations: An annotation study. In *Proceedings of the 9th SIGdial workshop on discourse and dialogue (SIGdial 2008)* (pp. 129–137). Columbus, Ohio, USA.
- Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., & Tsujii, J. (2012). brat: A Web-based tool for NLP-assisted text annotation. In *Proceedings of the demonstrations at the 13th conference of the European chapter of the association for computational linguistics (EACL 2012)* (pp. 102–107). Avignon, France.
- Stoyanov, V., & Cardie, C. (2008). Annotating topics of opinions. In Proceedings of the 6th international conference on language resources and evaluation (LREC 2008) (pp. 3213–3217). Marrakech, Morocco.
- Strapparava, C., & Mihalcea, R. (2007). SemEval-2007 task 14: Affective text. In Proceedings of the 4th international workshop on semantic evaluations (SemEval 2007) (pp. 70–74). Prague, Czech Republic.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307.
- Toprak, C., Jakob, N., & Gurevych, I. (2010). Sentence and expression level annotation of opinions in user-generated discourse. In *Proceedings of the 48th annual meeting of the association for computational linguistics (ACL 2010)* (pp. 575–584). Uppsala, Sweden.
- Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting of the association for computational linguistics (ACL 2002)* (pp. 417–424). Philadelphia, Pennsylvania, USA.
- Van de Kauter, M., Coorman, G., Lefever, E., Desmet, B., Macken, L., & Hoste, V. (2013). LeTs Preprocess: The multilingual LT3 linguistic preprocessing toolkit. *Computational Linguistics in the Netherlands Journal*, 3, 103–120.
- Van de Kauter, M., Desmet, B., & Hoste, V. (2014). Guidelines for the fine-grained analysis of polar expressions, version 2.0. Technical report LT3 14–02, LT3, Language and Translation Technology Team—Ghent University.

van Rijsbergen, C. (1979). Information retrieval. London: Butterworths.

Wiebe, J., Bruce, R. F., & O'Hara, T. P. (1999). Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the association for computational linguistics (ACL 1999)* (pp. 246–253). College Park, Maryland, USA.

- Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, 39(2–3), 165–210.
- Wiegand, M., Balahur, A., Roth, B., Klakow, D., & Montoyo, A. (2010). A survey on the role of negation in sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing (NeSp-NLP 2010)* (pp. 60–68). Uppsala, Sweden.
- Wilson, T. (2008). Annotating subjective content in meetings. In Proceedings of the 6th international conference on language resources and evaluation (LREC 2008) (pp. 2738–2745). Marrakech, Morocco.
- Wilson, T., & Wiebe, J. (2005). Annotating attributions and private states. In Proceedings of the workshop on frontiers in corpus annotations II: Pie in the sky (CorpusAnno '05) (pp. 53–60). Ann Arbor, Michigan, USA.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the conference on human language technology and empirical methods in natural language processing (HLT-EMNLP 2005) (pp. 347–354). Vancouver, British Columbia, Canada.
- Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on empirical methods in natural language processing (EMNLP 2003)* (pp. 129–136). Sapporo, Japan.
- Zabin, J., & Jefferies, A. (2008). Social media monitoring and analysis: Generating consumer insights from online conversation. Aberdeen Group Benchmark Report, Aberdeen Group.
- Zhang, L., & Liu, B. (2011). Identifying noun product features that imply opinions. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies (HLT 2011) (pp. 575–580). Portland, Oregon, USA.
- Zhuang, L., Jing, F., & Zhu, X.-Y. (2006). Movie review mining and summarization. In Proceedings of the 15th ACM international conference on information and knowledge management (CIKM 2006) (pp. 43–50). Arlington, Virginia, USA.