Annotating Opinions in German Political News

Hong Li, Xiwen Cheng, Kristina Adson, Tal Kirshboim, Feiyu Xu

German Research Center for Artificial Intelligence (DFKI) GmbH
Saarbrücken, Germany
{lihong, xiwen.cheng, kristina.adson, feiyu}@dfki.de

Abstract

This paper presents an approach to construction of an annotated corpus for German political news for the opinion mining task. The annotated corpus has been applied to learn relation extraction rules for extraction of opinion holders, opinion content and classification of polarities. An adapted annotated schema has been developed on top of the state-of-the-art research. Furthermore, a general tool for annotating relations has been utilized for the annotation task. An evaluation of the inter-annotator agreement has been conducted. The rule learning is realized with the help of the minimally supervised machine learning framework DARE.

Keywords: opinion mining, annotation, rule learning

1. Introduction

Opinion mining is the task to automatically identify and extract opinions from free texts. It consists of several subtasks such as sentimental resource acquisition (Esuli and Sebastiani, 2006; Remus et al., 2010), objective and subjective text classification (Wiebe et al., 2004), opinion express identification (Riloff and Wiebe, 2003), opinion target extraction (Hu and Liu, 2004; Cheng and Xu, 2008), opinion holder recognition (Kim and Hovy, 2006; Choi et al., 2005) and sentiment analysis (Pang et al., 2002; Turney and Littman, 2003) as well as contenxt-dependent sentiment analysis (Popescu and Etzioni, 2005).

Reliable annotated data is the foundation of investigation of the research phenomena and data-driven approaches. In previous research, the task of opinion annotation has been conducted on newspaper articles (Wiebe et al., 2005; Wilson and Wiebe, 2005; Wilson, 2008), on meeting dialogs (Somasundaran et al., 2008) and on user-generated product reviews (Toprak et al., 2010). But most of the corpora are in English. Only few work has considered other languages, for example (Schulz et al., 2010) describes their work on annotated product reviews in German and Spanish.

Our goal is to create an expression level annotated corpus of political news articles from German popular newspapers. In comparison to product reviewers, journalists often describe their opinions less explicitly. However "the overall tone of a written message affects the reader just as one's tone of voice affects the listener in everyday exchanges" (Ober, 1995).

Example 1a) and 1b) are the headlines of two different news articles that tell the same event, namely, the German chancellor Merkel has received the medal of freedom from the president Obama. The two authors used two different verbs: 1a) with the verb *ehren* (engl. honor) and 1b) with the verb *überreicht* (engl. hands over). The first one is strongly positive, while the second one is more or less neutral. Even if the tones of the two verbs are different, both sentences mention a positive event.

Example 1

a) US-Präsident Obama wird die Kanzlerin mit einer Auszeichnung **ehren**.
(*Engl.:* The US-president Obama will **honor** the chancellor an award.)

b) US-Präsident Barack Obama **überreicht** der Kanzlerin die "Medal of Freedom".
(*Engl.:* US-president Barack Obama **hands over** the "Medal of Freedom" to the chancellor.)

The remainder of paper is structured as follows. Section 2. presents two pieces of related work: one shows a foundation schema for our task and another shares some experience with respect to German for opinion mining. Section 3. introduces the annotation schema for words and phrases revealing the authors' attitudes in detail and Section 4. describes the tool and shows its GUI elements and functions, while Section 5. presents the annotation result and discusses our observations. Section 6. provides information about inter-annotator agreement and in Section 7. we show how the annotated frames can be utilized for relation-based opinion extraction task. Section 8. summarizes the task and discusses the future research plan.

2. Related Work

Concerning the expression level annotation, the most relevant and well-known work is the Multi-Perspective Question Answering (MPQA) corpus (Wiebe et al., 2005; Wilson and Wiebe, 2005; Wilson, 2008). This work presents a schema for annotating expressions of opinions, sentiments, and other private states in newspaper articles. *Private state* describes mental and emotional states and cannot be directly observed or verified (Quirk et al., 1985). The annotation schema focuses on the functional components of *private states* (i.e. *experiencers* holding *attitudes*, optionally toward *targets*) and consists of 3 main types of frames — *explicit mentions of private states, speech events expressing private states* and *expressive subjective*

elements — with attributes (i.e. *text anchor, target, source*) and properties (e.g. *intensity, attitude type*).

As far as we know, there is only one German corpus available which is annotated with opinion relevant information for user-generated reviews (Schulz et al., 2010). In addition to direct and indirect opinions, their annotation schema captures additional information such as comparisons, suggestions and recommendations of products and their features. Furthermore, it includes both explicit and implicit product features as opinion targets and handles pronouns and even product features that are not mentioned in the texts.

3. Annotation Schema

The goal of our annotation task is to create a news article corpus annotated with opinions about German politicians. The application of this work will be political trend/opinion monitoring. The annotated corpus will be used as input for the further machine learning approach to automatic extraction of linguistic patterns which indicate the relations between elements of an opinion: source of the opinion, target of the opinion and the polarity of the sentiment. Therefore, our focus is to identify the tones of voice via the words used by the journalists but not to cover all the private states. Inspired by MPQA, our annotation frame has four elements (see Table 1) and three properties (see Table 2). The four elements are *Text anchor*, *Target*, *Source* and *Auxiliary*. *Text anchor* is the text span that expresses attitudes, which might be idioms, phrases or words. *Target* is about what or whom the attitude is. *Source* is who is holding the attitude, except of the author. *Auxiliary* is a word that influences sentiment and can be negations, intensifiers or diminisher.

Element	**Property**
Target	
Source	
Text anchor	isaIdiom, isaPhrase, isaWord isaCompoundNoun
Auxiliary	isaNegation, isaIntensifier, isaDiminisher

Table 1: Opinion frame elements

In comparison to English, the German language has some special features, such as particle verbs (in German: *trennbare Verben*) as shown in Examples 2 a) and 2 b), and complex compound nouns (see Example 3). Example 2 shows the two different word orders for the same idiom *jemanden an der Nase herumführen*. The verb "führen" can be put separately from other parts of the idiom. This idiom means *to muck around with somebody* and has a negative attitude. The bold text parts are annotated by us as *Text anchor* with a property of *isaIdiom*.

Example 2
a). Sie haben es satt, sich erneut von Politikern **an der Nase herumführen** zu lassen.
(*Engl.:* They are fed up with being **mucked around** with through the politicians.)
b). Sie **führte** mich **an der Nase herum**.
(*Engl.:* She **mucked around** with me.)

In German, compound nouns are widely used. In particular, journalists prefer to generate new compound words. Example 3 shows a new created compound noun *toothpaste smile*, which is used to describe the smile of a politician in a sarcastic way. This phenomenon is annotated with a special value *isaCompound* for a *Text anchor*.

Example 3 Wulff hat erst am Abend sein Zahnpastalächeln wiedergefunden.
(*Engl.:* Wulff hasn't recovered his toothpaste smile until the evening.)

We have also defined properties of a frame. The three properties in Table 2 are *Attitude* with the value *negative*, *positive* and *sarcastic*, *Type* with *context-dependent* and *context-independent*, and *Intensity* with *low*, *medium* and *high*. For instance, the annotation of Example 3 is listed below:

Target	Wulff
Text anchor	hat ... Zahnpastalächeln wiedergefunden (*Engl.:* has found .. toothpaste smile again)
Attitude	negative
Type	context-dependent
Intensity	medium

The sentiment of *Text anchor* is dependent on the context that Mr. Wulff is a politician, who is supposed to give professional smiles all the time. Under this circumstance, the value of the property *Type* is *content-dependent* which suggests that the sentiment of the frame is dependent on the context or even the domain. In contrast, in Example 4, *Text anchor* is *nervt* meaning *annoy* and its sentiment is always negative, independent on the fillers of *Target* and *Source*.

Property	**Value**
Attitude	negative, positive, neutral, sarcastic
Type	context-dependent, context-independent
Intensity	low, medium, high

Table 2: Opinion frame properties

Example 4 Deshalb nervt die FDP die Union.
(*Engl.:* Because of that, the FDP annoys the Union.)

For each frame, *Text anchor*, *Target* and *Attitude* are obligatory and the rest of elements and properties are optional.

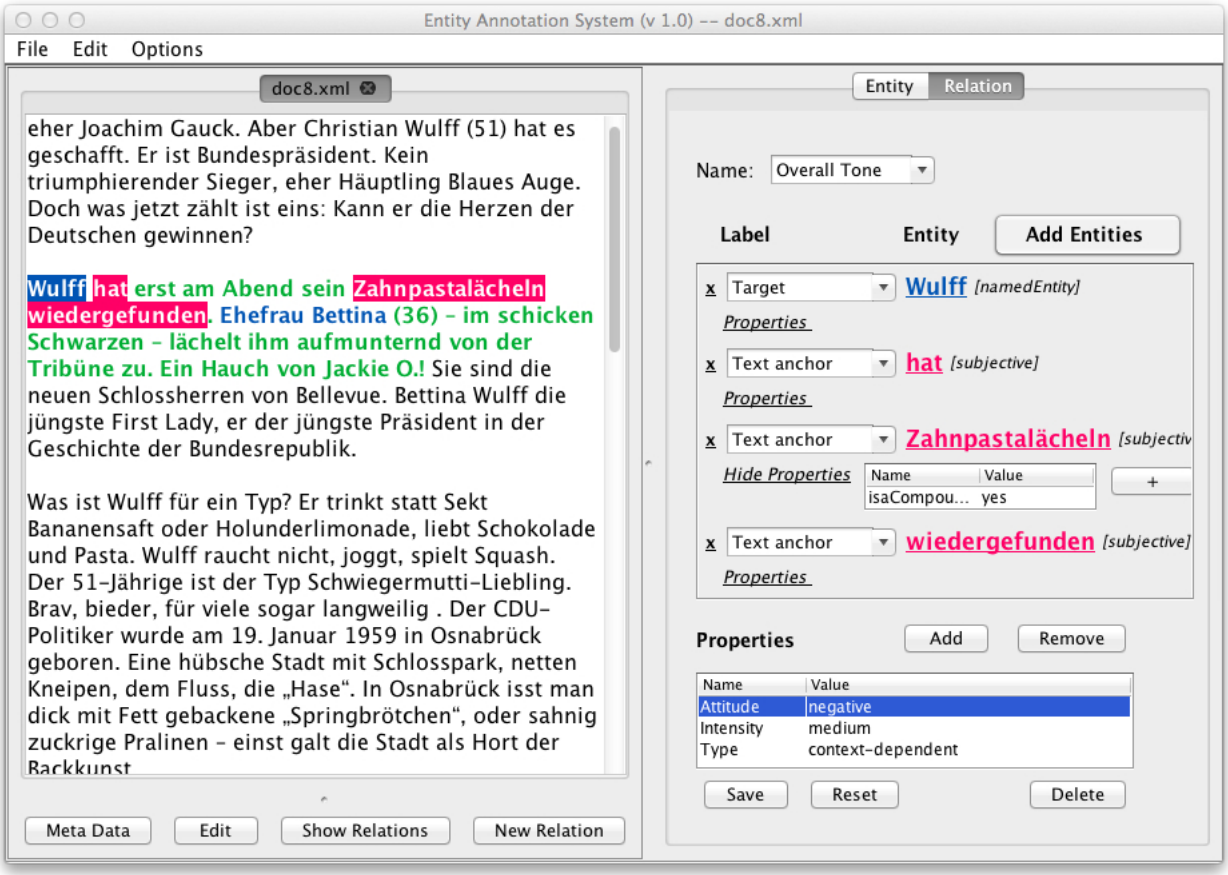

Figure 1: Annotation in Recon for *Wulff hat erst am Abend sein Zahnpastalächeln wiedergefunden.*

4. Annotation Tool: Recon

Recon is a general annotation tool for annotating relations among textual elements and semantic concepts. In comparison to the Gate annotation tool (Cunningham et al., 2002), Recon allows users to annotate not only individual text spans but also relations among them.

Recon provides a java-based graphical user interface (GUI) (see Figure 1). The GUI contains two major parts. The left part presents the text document for annotation and visualizes the annotated results. Within this panel, users can mark text spans and add annotations to them: 1) assign entity type, 2) assign semantic role to them if they are involved in a relation. Beneath the text panel, there are four functions available:

- *Meta data*, which lets users enter information about the document, such as *topic* and *publication date*;
- *Edit*, which makes the text area editable;
- *Show relations*, which show all the annotated relations in the current text;
- *New relation*, which lets users to add a new relation.

The right part of the GUI has two areas. The upper area is for

- definition of a new entity or a new relation with its semantic arguments;
- edition of the definition of an existing relation;
- annotation of a relation instance and assigning semantic roles to the selected text spans or selected entities.

The lower part is responsible for definition and annotation of properties of a relation or a semantic concept. Users are allowed here to define and edit features and values. On the bottom, there are three buttons available for *save*, *reset* or *delete* the edited content. In sum, Recon is a very easy and convenient tool for defining entities and their relations and assigning features with values to them.

5. Single-annotator Annotation

We have annotated 108 documents from German political news and have extracted frames from 714 sentences (see Table 3). Two experiments are conducted with respect to the effort spent by the annotators: *thorough* and *context-independent* annotation. *Thorough annotation* means that we annotate both context-dependent and context-independent frames. *Context-independent* annotation means that the mention of the opinions and their sentiments is explicit and no knowledge of politics is needed for the interpretation. One third of the corpus has obtained a thorough annotation, while the rest with a context-independent annotation.

In 36 (1/3 of the corpus) documents, we acquired 15 positive and 95 negative context-independent frames

and 12 positive and 57 negative context-dependent ones from 331 sentences. The numbers of these two kinds of frames are similar but the annotation of context-dependent frames requires three times more effort than the context-independent frames, since the annotators have to be well-informed about the details of the politics. In the rest of 72 documents, the annotator acquired 21 positive and 147 negative context-independent frames from 383 sentences. 86.2% of overall frames are negative and 13.8% positive. The result shows the strong bias in German political news, namely, the overwhelming tones of voices are negative.

		Context-independent		Context-dependent	
Doc	Sen	Pos	Neg	Pos	Neg
36	331	15	95	12	57
72	383	21	147	–	–

Table 3: Number of frames extracted from the annotation result

In the annotation result, 45.6% of frames have words as their text anchors, while 54.4% frames have phrases as their text anchors. Among frames containing words as their text anchors, 58.8% inherent sentiment from verbs, 26.5% from nouns (around 50% are compound nouns) and the other 14.7% from adjectives. Specifically, we acquired 268 words, among which 51 are positive, such as *Erfolg (Engl.: effort)* and 199 are negative, such as *Krieg (Engl.: war)*, *Konflikt (Engl.: conflict)* and *Vorwurf (Engl.: complain)*. The number of annotated words with different intensity is listed in Table 4. The acquired negation words are *nichts*, *weder ... noch*, *kein*, *nie*, *nicht* and *niemand*. These extracted phrases and words can be considered as opinion instances and used as seeds for bootstrapping opinion learning procedures.

	low	medium	high
Negative	32	146	21
Positive	5	38	8

Table 4: Number of sentimental words with different intensity in the annotation result

We compared our result with SentiWS (Remus et al., 2010) - an automatically generated German sentiment dictionary with 3468 words. The overlap between our result and the SentiWS is 118 words and the coverage in our corpus is 44%. Among 118 words, there are 9 conflicts and the precision is 92.4%. For example, *kritisch* (*engli.: critical*) has a positive tone of voice score value of *0.0040* and a negative tone of voice score value with *-0.203* in SentiWS. But *kritisch* is marked as strongly negative in our result. *Konflikt* (*engli.: conflict*) appears in our corpus 33 times and has a strong negative tone of voice value in our result, while only weakly negative tone of voice score value *-0.0048* in SentiWS. The result shows that SentiWS has a small coverage in the political domain.

6. Inter-annotator Agreement

To measure the agreement between the annotators, we use the metrics described in (Wiebe et al., 2005) with A and B denoting the two annotators respectively:

$$arg(A||B) = \frac{\text{\# of frames annotated by } A \text{ and } B}{\text{\# of frames annotated by } A} \quad (1)$$

The results are shown in Figure 2(a). Figure 2(b) shows the number of annotated frames. The annotators identified 315 frames as total. 186 frames of them are consistent. In 6 cases, both annotators have identified the same text anchors but assign different polarities. We call these cases *conflict*. In the other 123 cases, either one annotator does not assign any annotation or the two annotators select completely different text anchors. There are no overlaps in their annotations.

$\boldsymbol{arg}(A\|\|B)$	$\boldsymbol{arg}(B\|\|A)$	**F1**
0.59	0.97	0.73

(a) The value of inter-annotator agreement

consistent	**inconsistent**	
	conflict	**no overlap**
186	6	123

(b) Consistency and inconsistency in the two annotation results

Figure 2: Inter-annotator agreement measure and the annotated frames by annotators A and B

According to Figure 2(a), the value of of $arg(A||B)$ is much lower than $arg(B||A)$ because annotator A identified much more frames than B. Example 5 shows examples of the *no-overlap* cases, in which annotator A marked the frame as slightly negative while B did not recognize these sentences as subjective.

Example 5

a) Schon jetzt haben die ersten Zahlungen von BP an Alabama, Florida und Mississippi den **Beigeschmack** von Wahlkreisgeschenken
(Engl.: The compensation damages of BP to Alabama, Florida, and Mississippi already have a **flavor/aftertaste** of election bribes.*)*

b) Es ist Zeit für Kofi Annan und seine UNO, **Rückgrat zu zeigen**.
(Engl.: It's time for Kofi Annan and his UN to **take on the challenges/stand tall**.*)*

Moreover, we asked another annotator C to recheck the 6 conflict cases. 2 of his results are the same as A's and 2 share with the results of B.

Given the following sentence,

Example 6

Nur einmal liess Blair etwas von **Reue** erkennen...
(Engl.: **Just once**, Blair showed a bit of **regret** *)*

the three annotators identified different text anchors with different tones:

annoator	frame
A, C	Ton: negative, Text Anchor: Nur einmal (*Engl.*: just one)
B	Ton: positive, Text Anchor: Reue (*Engl.*: regret)

Although it is avoidable that different people sense different sentiment from different parts of the texts, the inter-annotator agreement 0.73 shows that the annotators agree with most of the annotated frames, proving that the annotated results could be used for further applications. Moreover, for these 108 documents whose average lengths are 517 words each, the average annotation time is 50 hours per person. This result encourages us to use this way to prepare more annotated corpus as a solid foundation for further research.

7. Relation-based Opinion Extraction

The annotated results can be further applied to automatic extraction/learning of sentiment terms or relation-based opinion extraction rules. We have used the annotated result as the input for the relation-extraction system DARE[1] (Xu, 2007; Xu et al., 2007) to learn opinion extraction rules. DARE is a minimally supervised machine-learning system for relation extraction. The DARE system can learn rules for arbitrary relations or events automatically, given linguistically annotated documents (i. e., dependency parses) and some initial examples of the target relations.

Before the rule learning process, we preprocessed them with the German MaltParser[2] (Nivre et al., 2007) for dependency parsing. From each annotated frame, DARE learns extraction rules identifying the source, target and the sentiment polarity of the opinion. Using the 278 context independent frames, we have learned 419 DARE rules. These rules are classified into two categories:

1. **Context-dependent extraction rules**, which recognize the opinions of the sentences depending on the sentimental words in the context. Such rules derive the tones of the extracted opinions from the polarity of the subjective words contained in sentimental dictionaries, such as SentiWS. Example 7 shows two rules *"geben für" (Engl: give ... for)* and *"bezeichnen als" (Engl: denote as)*. These two phrases themselves couldn't convey any sentiment. But when they are used together with other sentimental words or phrases, such as "Vorwurf" (*Engl*: accusation) or "verantwortungslos" (*Engl*: irresponsible), the sentiment of the sentence will be the same as the tone of these subjective terms. *MO*, *PNK*, *OA*, *OB* in the rules are dependency labels tagged by the MaltParser.

   Example 7

   a) "geben" $\begin{cases} \textit{MO: ("für")} \{\textit{PNK: (}\texttt{Target}\textit{)}\} \\ \textit{OA: (}\texttt{Subjective Term}\textit{)} \end{cases}$
   ("geben" Engl.: "give"; "für" Engl.: "for")

   b) "bezeichnen" $\begin{cases} \textit{OB: (}\texttt{Target}\textit{)} \\ \textit{MO: ("als")} \\ \quad \{\textit{NK: (}\texttt{Subjective Term}\textit{)}\} \end{cases}$
   ("bezeichnen" Engl.: "denote"; "als" Engl.: "as")

2. **Context-independent extraction rules**, which can convey the sentiment by themselves. Example 8 shows two dependent rules "`Source` *kritisieren (Engl: criticize)* `Target`" and "`Source` *angreifen (Engl: attack)* `Target`", which both convey negative sentiment towards their targets.

   Example 8

   a) "kritisieren", (Tone: negative)
   $\{\textit{SUBJ: (}\texttt{Source}\textit{)}, \textit{OB: (}\texttt{Target}\textit{)}\}$
   ("kritisieren" Engl.: "criticize")

   b) "angreifen", (Tone: negative)
   $\{\textit{SUBJ: (}\texttt{Source}\textit{)}, \textit{OB: (}\texttt{Target}\textit{)}\}$
   ("angreifen" Engl.: "attack")

8. Conclusion and Future Work

The annotation result is a valuable resource for the opinion mining research for German language, in particular, the German political news. The distinction between context-dependent and context-independent frames is important for the estimation of the need of world and domain knowledge for a running system. Considering the efficiency and cost, we could start with context-independent annotation and extend the domain knowledge with the acquired results step by step. Moreover, the high inter-annotator agreement for content-independent annotation shows that the annotation schema is promising to be applied for further research. We have used the annotated results in the relation extraction system DARE to automatically extract opinions and the first experimental result is encouraging. The next step is to train more annotators for the task in order to acquire more sophisticated frames from corpus and to conduct further experiments in opinion extraction.

Acknowledgements

This research work was conducted in the research projects Alexandria and Alexandria for Media of the THESEUS research program and in its use case Alexandria (funded by the German Federal Ministry of Economy and Technology, contract no. 01MQ07016) and the research project Deependance (funded by the German Federal Ministry of Education and Research, contract no. 01IW11003). We thank our project partner Neofonie GmbH[3] for providing us the news corpus and the linguistic annotations (tokenization, named entity recognition and dependency structures).

[1] http://dare.dfki.de/

[2] http://maltparser.org/

[3] http://www.neofonie.de/

9. References

Xiwen Cheng and Feiyu Xu. 2008. Fine-grained opinion topic and polarity identification. In *Proceedings of the 7th International Conference on Language Resources and Evaluation LREC*, page 2710–2714, Marrekech, Morocco. European Language Resources Association (ELRA).

Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Identifying sources of opinions with conditional random fields and extraction patterns. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, page 355–362, Morristown, NJ, USA.

H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. 2002. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proc. of ACL'02*.

Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, page 417–422, Genova, Italy.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *KDD'04: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 168–177, Seattle, Washington.

Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text at the joint COLING-ACL Conference*, page 1–8, Sydney, Australia.

J. Nivre, J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, and E. Marsi. 2007. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(02):95–135.

Scott Ober. 1995. Contemporary business communication. 2nd edition. Boston: Houghton Mifflin.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, page 79–86, Morristown, NJ, USA. Association for Computational Linguistics.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processingg*, page 339–346, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. A comprehensive gram- mar of the english language. Longman, New York.

Robert Remus, Uwe Quasthoff, and Gerhard Heyer. 2010. Sentiws-a publicly available germanlanguage resource for sentiment analysis. In *Proceedings of the 7th International Conference on Language Resources and Evaluation LREC*, pages 1168–1171. European Language Resources Association (ELRA).

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *EMNLP-03: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, page 105–112.

Julia Maria Schulz, Christa Womser-Hacker, and Thomas Mandl. 2010. Multilingual corpus development for opinion mining. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Swapna Somasundaran, Josef Ruppenhofer, and Janyce Wiebe. 2008. Discourse level opinion relations: An annotation study. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, page 129–137. Association for Computational Linguistics, June.

Cigdem Toprak, Niklas Jakob, and Iryna Gurevych. 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*, pages 575 – 584, Uppsala, Sweden, July. Association for Computational Linguistics (ACL).

Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. volume 21(4), page 315–346.

Janyce M. Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. volume 30(3), page 277–308.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. volume 39(2/3), page 165–210.

Theresa Wilson and Janyce Wiebe. 2005. Annotating attributions and private states. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, page 53–60, Ann Arbor, Michigan.

Theresa Ann Wilson. 2008. Fine-grained subjectivity and sentiment analysis: Recognizing the intensity, polarity, and attitudes of private states. In *PhD Thesis*, University of Pittsburgh.

Feiyu Xu, Hans Uszkoreit, and Hong Li. 2007. A seed-driven bottom-up machine learning framework for extracting relations of various complexity. In *Proceedings of ACL 2007*, Prague, Czech Republic, 6.

Feiyu Xu. 2007. *Bootstrapping Relation Extraction from Semantic Seeds*. Phd-thesis, Saarland University.