Sentiment Analysis for Media Reputation Research

Samuel Läubli[φ] **Mario Schranz**[γ] **Urs Christen**[γ] **Manfred Klenner**[φ]

[φ]Institute of Computational Linguistics,
University of Zurich
Binzmühlestrasse 14
8050 Zürich
{laeubli,klenner}
@cl.uzh.ch

[γ]Center for Research on the Public
Sphere and Society, University of Zurich
Andreasstrasse 15
8050 Zürich
{mario.schranz,urs.christen}
@foeg.uzh.ch

Abstract

As a subtask of qualitative media reputation research, human annotators manually encode the polarity of actors in media products. Seeking to automate this process, we have implemented two baseline classifiers that categorize actors in newspaper articles under six and four polarity classes. Experiments have shown that our approach is not suitable for distinguishing between six fine-grained classes, which has turned out to be difficult for humans also. In contrast, we have obtained promising results for the four class model, through which we argue that automated sentiment analysis has a considerable potential in qualitative reputation research.

1 Introduction

While opinion mining techniques have been successfully implemented in large-scale appliances such as social media monitoring on the web (e.g., Godbole et al. (2007), Chen et al. (2012)), detailed studies in the field of reputation research still raise the need for human assessments. By focussing on a sub-task of qualitative media reputation analysis—identifying fine-grained actor polarities in full newspaper articles—we seek to examine whether automated sentiment analysis approaches can be used in traditional encoding workflows.

From a pragmatic perspective, supporting encoding processes has the potential of saving precious time for annotation experts, be it by preselecting texts for further examination or suggesting classifications for targeted variables, such as the centrality or polarity of the reputation objects in scope. For sentiment analysis research on the other hand, we see opportunities to evaluate selected methods in a real-world scenario. In our current work, we primarily seek to explore the feasibility of employing more fine-grained classes than the traditional trichotomic distinction between *positive*, *neutral*, and *negative* polarities in automated sentiment analysis.

First, we give an introduction to traditional media reputation analysis and detail the resources we use for our experiments. In section 3, we introduce a lightweight approach to sentiment composition, which we implemented in a prototype classifier for the polarity of actors in newspaper articles. The evaluation of our system is presented in section 4 and subsequently discussed in section 5. Finally, we conclude our report in section 6, also listing further work that is planned or currently pursued.

2 Background

From the very beginning, text classification has been a very active research direction in the area of sentiment analysis. However, the focus of attention was mostly on the classification of product or movie reviews (e.g., Hu and Liu (2004)). There are a few exceptions, e.g., work based on the MPQA corpus (Wilson et al., 2005), where newspapers are dealt with.

Most of the time, a three-partite classification

Proceedings of KONVENS 2012 (PATHOS 2012 workshop), Vienna, September 21, 2012

Class	# Events	in %
neutral	45'018	48.5
controversial	14'096	15.2
negative, explicit	13'251	14.3
negative, implicit	9'977	10.8
positive, explicit	6'244	6.7
positive, implicit	4'236	4.6
Total	92'822	100.0

Table 1: Class Distribution in the Media Sample Corpus

Actor	Category	# Events	in %
UBS	bank	43'440	46.8
Crédit Suisse	bank	30'662	33.0
ZKB	bank	5'897	6.4
Swisscom	telecom	5'070	5.5
Novartis	pharma	3'270	3.5
Roche	pharma	2'381	2.6
Cablecom	telecom	1'637	1.8
Sunrise	telecom	465	0.5
Total		92'822	100.0

Table 2: Actors in the Media Sample Corpus

is carried out: a text is either *positive*, *negative* or *neutral*. In contrast, we cope with four and even six classes, among which is a rather demanding class for controversial texts. Moreover, the classes of *negative* and *positive* are split into the more fine-grained distinction of *implicit* and *explicit*, respectively. This makes a challenging demand of our application scenario—media reputation analysis.

We are also in the tradition of Moilanen and Pulman (2007) and, more basically, Polanyi and Zaenen (2004), since we regard the notion of compositionality as crucial for the analysis of advanced texts. Instead of the elaborated syntax- and rule-based approach of Moilanen and Pulman (2007), our approach (described in section 3) is closer to the grammar independent one proposed by Choi and Cardie (2008).

In the remainder of this section, we give an introduction to our application domain and list resources we rely on for our present work.

2.1 Media Reputation Analysis

The Center for Research on the Public Sphere and Society (*fög*) of the University of Zurich has analyzed the media reputation of Swiss companies since 1998. Media reputation is defined by Deephouse (2000, p. 1097) as "the overall evaluation of the firm presented in the media resulting from the stream of media stories about the firm". Reputation arises and decays wherever information about the trustworthiness of an actor circulates in arenas of public communications, be it in the traditional mass media or the new internet-based media. Measurement instruments to determine relevant reputation dynamics must therefore necessarily be based on an analysis of public communications.

The *fög* has conducted a quantitative-qualitative media content analysis (Eisenegger et al., 2010). It is aimed at determining the media reputation of the 39 largest Swiss companies on a daily basis in thirteen leading Swiss media. Accordingly, the most significant Swiss business sectors such as banking, insurance, pharmaceuticals, telecommunications, as well as manufacturing, food, and retail are part of this monitoring. The content analysis examines how frequently and strongly (centrality) the media report on specific companies—to which we refer as *actors*—and how they were evaluated (polarity). The recorded encodings (*positive*, *neutral*, *negative*, and more fine-grained sub-classes) allow the *fög* to build a Media Reputation Index (Eisenegger and Imhof, 2008). The content analysis is focussed on a sample of thirteen major Swiss opinion-forming media, covering both key print media as well as newscasts by public service broadcasters.

2.2 Resources

Our current work is primarily based on two central resources: a big sample of manually encoded newspaper articles, stemming from the *fög* content analysis, and an unweighted sentiment lexicon. Both are written in or for German, respectively.

2.2.1 Media Sample Corpus

We extracted a sample corpus from the *fög* database of encoded texts (see section 2.1). It comprises newspaper articles that each include at least one of the actors listed in table 2. All ar-

Class	# Positive	# Negative	# Neutral
adjectives	1’677	2’097	91
nouns	1’202	2’144	595
verbs	528	1’001	2
Total[a]	3’407	5’242	688

[a] Besides 9337 polar entries for adjectives, nouns, and verbs, the lexicon comprises 61 base forms with polarity functions (shifter, intensifier, diminisher).

Table 3: Polar Base Forms in the Polarity Lexicon (Clematide and Klenner, 2010)

ticles were published and encoded between 1998 and 2011.

Newspaper articles can contain more than one actor, hence we use the term *event* whenever we refer to an encoded occurrence of an actor in a text. The presented corpus consists of 85’817 articles that contain 92’822 events altogether; an article features 1.08 events on average (std. dev = 0.36). We point out that due to this fact we cannot just classify the polarity of a text in order to derive the polarity of its actor(s), which is a further challenge for automating the classification.

Despite the fine-grained annotations, there is no structural information stored for the articles in our sample corpus. Although titles, leads, and image captions could be particularly informative for classification, we regard this as a given constraint and leave according experiments to future work.

Used as a training set for our learning algorithm and a gold standard for our overall evaluation, the media sample corpus makes the cornerstone of our present study. Section 4 gives further details on how the corpus was used for training and evaluation.

2.2.2 Polarity Lexicon

As a further resource, we use a polarity lexicon compiled by Clematide and Klenner (2010). It contains 9’398 base forms (see table 3), each of which has either assigned a polarity—*positive* (+), *neutral* (=), *negative* (−)—or a polarity function—*shifter* (¬), *intensifier* (<), *diminisher* (>). As we pursue an unweighted approach for sentiment computation, we did not use intensifiers and diminishers, and neither did we use any weights associated with polar base forms.

In the following section, we describe how the sentiment lexicon is used for the purpose of sentiment composition, and we give an example of how the approach can be used to generate features for machine learning algorithms.

3 Method

While human experts can rely on rather loosely defined coding instructions and their world knowledge to classify an actor’s polarity, classification systems call for a set of well-defined features that capture characteristics of the samples to be classified. For example, the *fög* class definition of *positive, implicit*—“the reputation object is discussed in a positively connoted context”[1]—needs to be translated into computable features such as “a context is defined as a sentence unit”, “a context is positive if it only contains positive words” or “an actor is implicit to a context if it is not directly mentioned, but strongly related to the context’s subject”. Such rules are clearly error-prone; they are only heuristic, and in addition, they need to cope with noise caused by preceding system components such as a syntactic parser, thus raising the need for, e.g., a machine learning algorithm.

We implemented a lightweight sentiment analysis approach in our prototype, which is explained in the following section.

3.1 Lightweight Lexical Sentiment Composition

Moilanen and Pulman (2007) have proposed a model for calculating global polarities of syntactic phrases via their subordinate constituents. The model is based on sentiment lexica that assign prior polarities to leaf constituents, which are subsequently propagated or reversed by applying a considerable number of rules for weighting, filtering, and conflict resolution.

Aiming at robustness and simplicity, we propose a sentiment composition approach that is entirely based on an unweighted sentiment lexicon and head-dependencies of candidate words (tokens).

[1] “Das Reputationsobjekt wird in einem positiv konnotierten Kontext thematisiert.” (*fög*, 2011)

3.1.1 Rules

Let $t_{1..n}$ be tokens of a candidate word sequence T. An initial $polarity\_marker$ function assigns prior polarity tags m to all $t \in T$ that are contained in the lexicon:

$$polarity\_marker : t \rightarrow t_{m \in \{+,=,-,\neg\}}$$

The model consists of two simple rules that operate on accordingly tagged and dependency-parsed word sequences:

(A) Each shifter $t_\neg$ can invert its parent $t_{+/-}$ exactly once.

(B) Each polar token $t_{+/-}$ can change its parent $t_{+/=/-}$ exactly once:

- (i) child $t_+ +$ parent $t_= \rightarrow$ parent t_+
- (ii) child $t_- +$ parent $t_= \rightarrow$ parent t_-
- (iii) child $t_- +$ parent $t_+ \rightarrow$ parent t_-

Rules (A) and (B) are applied repeatedly until convergence, that is, until no marked token t can be altered any further. Composition rules (i)–(iii) are taken from the pattern-matching approach described in (Klenner et al., 2009). Note that we do not list relationships where none of the tokens need to be altered, e.g., $t_+ + t_+ \rightarrow t_+$.

3.1.2 Example

To give an example, we look at a sentence occurring in our sample corpus of encoded newspaper articles (see sections 2.1 and 2.2.1):

> Particularly if no convincing managerial concept is at hand, which the UBS with brutal openness admitted to be the case this week.[2]

Applying the initial steps of our feature extraction pipeline—dependency parsing and polarity marking—yields the following structure for the first part of the above input sentence:

... no convincing ... concept ...
$\neg$ $+$ $=$

[2] German original: "Vor allem dann, wenn kein überzeugendes betriebswirtschaftliches Konzept auf dem Tisch liegt, was die UBS diese Woche mit brutaler Offenheit zugab."

Next, rules (A) and (B) (see section 3.1.1) are applied in turn. As rule (A) cannot be applied in the first iteration (shifters can only invert polar tokens $t_{+/-}$), (B) is fired first:

B

... no convincing ... concept ...
$\neg$ $(+)$ $\Rightarrow +$

In the next iteration, (A) can be applied to "kein$_\neg$" as its *regens* is now polar:

A
B

... no convincing ... concept ...
$(\neg)$ $(+)$ $\Rightarrow -$

As rules (A) and (B) can only be applied once for each child-head token relationship, the final state is reached for this polar chunk.

3.2 Features for Machine Learning

We form *n*-grams from polar dependency chunks in order to use our sentiment composition approach for machine learning. To avoid data sparseness we limit *n* to 2; longer sequences are split into multiple bigrams. In short, our features are constructed as follows:

(i) *zerogram*: the polarity of the head token, e.g., `NEG`

(ii) *unigram*: the head token and its polarity, e.g., `NEG_concept`

(iii) *bigram*: (ii) plus the child token, e.g., `NEG_concept_convincing`

To take actor proximity into account, we append `-S` to a feature whenever an actor is present in the same sentence, e.g. `NEG_concept_convincing-S`. As there is no structural information such as title or image caption sections available for the texts in our sample corpus, further contextual information cannot be considered at this point.

We use polar *n*-grams as feature names (dimensions) and corresponding absolute counts as feature values. For example, encountering "no convincing concept" in an article would raise the value of `NEG_concept_convincing-S` by 1 (default value: 0). In order to give more weight to long polar chunks, all lower-order *n*-gram counts are also increased when adding a uni- or bigram.

To sum up, a text consisting of nothing but the sample sentence from section 3.1.2 would result in the following features:

Feature Name	Value
NEG_concept_convincing-S	1
NEG_concept_no-S	1
NEG_concept-S	2
NEG_convincing-S	1
NEG_no-S	1
NEG_openness_brutal-S	1
NEG_openness-S	1
NEG_brutal-S	1
NEG-S	8

4 Evaluation

In order to explore the feasibility of automating media reputation analysis processes, we have implemented a prototype system that classifies the polarity of actors in newspaper articles. In this section, we present the results of our corresponding evaluations.

4.1 Prototype Implementation

We have set up a feature extraction pipeline that processes digital newspaper articles. After handing an article's full text to a dependency parser (Sennrich et al., 2009), the parser output is converted into CG format (Constraint Grammar; (VISL-group, 2008)) for subsequent enrichment by the *polarity marker*, a compiled VISL constraint grammar that adds prior polarities and polarity functions to single words. This corresponds to the $polarity\_marker$ function explained in section 3.1.1. Next, the marked CG serialization is handled by the polarity composition component (also VISL-based), and finally, all accordingly derived polarity chunks are converted into a set of features suitable for machine learning algorithms (see section 3.2).

For our prototype implementation, we abstracted all actors from the input texts for training and evaluation, i.e., all occurrences of actor names such as "UBS" or "Crédit Suisse" were replaced by an arbitrary token ("ACTOR"). In this way, we ensure that our classifiers can evaluate *any* actor in principle, given their name and an optional list of synonyms. This optimizes flexibility in real-world scenarios, but may lower recall in cases where actor synonyms are not recognized in the replacement process.

Although we did not focus our efforts on optimizing quantitative performance, our prototype pipeline runs reasonably fast with the dependency parser being the only "bottleneck" in speed. On average, parsing an article of the media sample corpus took 4.9 seconds[3] in our experiments, while passing it through the remainder of the pipeline (polarity marking, composition, feature extraction, and classification) took another 0.4 seconds.

4.2 Evaluation Method

As mentioned in section 2.2.1, our gold standard consists of 92'822 events in 85'817 newspaper articles (see table 1). It was produced by *various* annotators over the last 14 years. For the purpose of our evaluation, we asked a *single* expert to re-encode a random sample of 200 articles (220 events). Re-encoding took place in the expert's usual working environment. We used a dedicated web application to collect all annotations and corresponding time stamps.

In this way, the evaluation task was the same for the human expert and our system: reproducing the original gold standard annotations for the random sample of texts. Since there is no inter-annotator agreement known for our gold standard (i.e., the media sample corpus), the performance of the human annotator on the 200 texts may indicate how hard the classification task at hand actually is.

4.3 Experiment 1: 6 Classes

The current revision of the *fög* coding manual (2011) lists six polarity classes (see table 1). Using all but the separated evaluation articles of our sample corpus, we trained a *Naive Bayes* classi-

[3]All times were measured using a standard Unix server (24x2.3GHz CPU, 128GB RAM). The prototype pipeline runs comparably fast on simple workstation computers.

	Human$_6$			System$_6$		
Accuracy:	59.5			51.9		
Class	P	R	F	P	R	F
neutral	86.3	67.0	75.4	61.9	86.0	72.0
controversial	33.9	71.4	46.0	26.3	20.0	22.7
negative, exp.	61.9	78.8	69.3	37.0	30.8	33.6
negative, imp.	40.0	11.1	17.4	25.5	10.7	15.1
positive, exp.	40.0	57.9	50.0	24.6	12.0	16.2
positive, imp.	33.3	15.8	21.4	8.0	0.5	0.9

	Human$_4$			System$_4$		
Accuracy:	66.8			57.4		
Class	P	R	F	P	R	F
neutral	86.3	67.0	75.4	62.8	81.9	71.1
controversial	33.9	71.4	46.0	25.3	25.0	25.2
negative	78.7	72.5	75.5	68.6	52.9	59.8
positive	58.3	61.8	55.3	42.9	21.1	28.3

Table 4: Evaluation of Human and System Classification Accuracy on a Test Set of 200 Full Newspaper Articles

fier[4] that assigns these class labels to all actors in candidate texts, based on our feature pipeline described in section 4.1.

Table 4 (left portion) shows the results for the six class experiment. For each class, we list precision (P), recall (R) and balanced f-score (F). Overall classification accuracy is indicated at the top. We used 200 full newspaper articles containing 220 events for this experiment, i.e., the same events were labelled by *fög* experts (Human$_6$) as well as our classifier (System$_6$). On average, classification took 56.3 seconds per event (Human$_6$) and 4.8 seconds (System$_6$) respectively.

4.4 Experiment 2: 4 Classes

In a second experiment, we folded *implicit* and *explicit* into one class each for *positive* and *negative* (see table 4, right portion). This addresses the low scores that were obtained especially for *implicit* polarities in both human and system classification. All other conditions were left unchanged.

The findings of our experiments are discussed in the following section.

5 Discussion

Automatically assigning fine-grained classes to actors turned out to be an all but trivial task in our wide domain. This is reflected in the evaluation results of our six class model (System$_6$), which performs considerably less accurate than a human annotator. The system assigns too much probability to *neutral* events—an obvious drawback of using a *Naive Bayes* classifier, which elevates the most frequent class in the gold standard because of its high *a priori* probability—resulting in high recall for *neutral*, but lowering precision for this class and recall for all other classes. With f-scores below 35% for all non-neutral classes, classifying actors by use of our proposed feature pipeline is clearly not promising.

Still, our first experiment sheds light on how fragile it is to classify in a fine-grained mode even for experienced annotators. For Human$_6$, four out of six classes feature f-scores of 50% or lower, hinting that there are ambiguous cases that are difficult to resolve for humans also. This could indicate that "soft" (continuous) boundaries are more suitable than clearly delimitable (nominal) class boundaries when assessing the polarity of an actor in a wide context.

In our second experiment, we have folded the *positive* and *negative* classes. Removing the somewhat "blurry" distinction between *implicit* and *explicit* classes had a positive effect on both human and system classification accuracy, especially in terms of precision. Although System$_4$ still assigns too much probability to *neutral*, other classes do clearly benefit from the folding. Most remarkably, the f-score of *negative* has nearly reached 60%, which is remarkably higher than the sum of the *negative, implicit* and *negative, explicit* f-scores in System$_6$ ($\Delta = 11.1\%$). The same holds for *positive* ($\Delta = 11.2\%$), although on a much lower level.

We hypothesize that additional improvements could be gained from including structural information of newspaper articles in the gold standard. As mentioned in earlier sections, such annotations were not available in our sample corpus. How-

[4] Despite preliminary experiments with a number of other learning algorithms using WEKA (Hall et al., 2010), we decided to opt for fast iteration cycles and hence relied on the NLTK framework (Loper and Bird, 2002), which allowed for rapid prototyping.

ever, sentimental ascriptions could be particularly relevant for an actor if they appear in titles, leads or image captions of a newspaper article.

Apart from that, a thorough assessment of more powerful learning algorithms is indispensable for future iterations on our prototype system. This should be accompanied by including additional carefully thought out features, such as polarity class ratios or, as outlined in the previous section, structural information. Also, we will have to separately evaluate our polarity composition component (as illustrated in section 3.1.1) in order to consider possible extensions.

As for polarity class granularity, it is remarkable that *controversial* cases were particularly hard to identify, even in the four class model. The *fög* coding manual says that an event is controversial if "the reputation object is discussed controversially; positive and negative ascriptions are equally balanced."[5] Our classifier had no means of assessing the balance of negative and positive ascriptions when trained on our feature set described in section 3.2—there were of course counts for `POS` and `NEG` zerograms, but due to the *Naive Bayes* independence assumption, no positive-negative ratio could be obtained.

From a pragmatic point of view, one could argue that the system's moderate accuracy could partially be outweighed by quantitative considerations. As mentioned in section 4.1, the pipeline processes an article in 5.3 seconds on average, and in the nature of things, classifier decisions are fully reproducible. Viewed in this light, it does not seem unreasonable to use automated systems for work that is rather tedious for human experts, such as pre-selecting texts or encoding articles where the polarity of an actor is perfectly obvious.

6 Conclusion

In our exploratory work, we sought to assess the feasibility of automating a fine-grained classification process for media reputation analysis. We have trained two prototype classifiers relying on a lightweight approach to lexical sentiment composition, which we evaluated on a set of 200 previously annotated newspaper articles. It clearly turned out that our approach is not suitable for handling a six-class polarity model. However, we gained substantial improvement from folding *implicit* and *explicit* ascriptions into a four-class-model, through which we argue that automated approaches have a promising potential in qualitative media reputation analysis.

In future work, we will consider structural information of newspaper articles for classification, as well as thoroughly examine the impact of more sophisticated machine learning algorithms on classification accuracy. Currently, we are training and evaluating an additional classifier for a three-partite polarity model, as a distinction between *positive*, *neutral* and *negative* is still most important in high-level aggregations such as the *fög* Media Reputation Index (Eisenegger and Imhof, 2008). We consider our present work as a motivating first step towards automating qualitative reputation analysis processes, calling for further collaboration in the intersection between sentiment analysis- and media reputation research.

[5] "Das Reputationsobjekt wird kontrovers diskutiert; Positiv- und Negativzuschreibungen halten sich die Waage." (*fög*, 2011)

Acknowledgments

We would like to thank Simon Clematide and Michael Amsler for motivating the use of VISL in our feature extraction pipeline. Also, we are grateful to Corina Tobler for translating the examples provided in section 3.

References

Chen, Lu, Wenbo Wang, Meenakshi Nagarajan, Shaojun Wang, and Amit P. Sheth. 2012. Extracting Diverse Sentiment Expressions with Target-Dependent Polarity from Twitter. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM) 2012*, pages 50–57, Dublin.

Choi, Yejin, and Claire Cardie. 2008. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2008*, pages 793–801, Honolulu.

Clematide, Simon, and Manfred Klenner. 2010. Evaluation and extension of a polarity lexicon for German. In *Proceedings of the First Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA) 2010*, pages 7–13, Lisbon.

Codebuch Akteursanalyse. rev. 2011. Center for Research on the Public Sphere and Society (*fög*), University of Zurich. Internal coding manual for media reputation analysis.

Deephouse, David L. 2000. Media reputation as a strategic resource: An integration of mass communication and resource-based theories. *Journal of Management*, 26(6):1091–1112.

Eisenegger, Mark, and Kurt Imhof. 2008. The True, The Good and the Beautiful: Reputation Management in the Media Society. In Ansagr Zerfass, Betteke van Ruler, and Sriramesh Krishnamurthy, editors, *Public Relations Research: European and International Perspectives and Innovation*. VS Verlag für Sozialwissenschaften, Wiesbaden, pages 125–146.

Eisenegger, Mark, Mario Schranz, and Jörg Schneider. 2010. Corporate Reputation and the News Media in Switzerland. In: Craig E. Carroll, editor, *Corporate reputation and the news media. Agenda-setting within business news coverage in developed, emerging, and frontier markets*. Routledge (Communication series), New York, S. 207–220.

Godbole, Namrata, Manjunath Srinivasaiah, and Steven Skiena. 2007. Large-Scale Sentiment Analysis for News and Blogs. In *Proceedings of the First International AAAI Conference on Weblogs and Social Media (ICWSM) 2007*, Boulder.

Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1):10–18.

Hu, Minqing, and Bing Liu. 2004. Mining Opinion Features in Customer Reviews. *Proceedings of the 19th national conference on Artifical intelligence (AAAI) 2004*, pages 755–760, San Jose.

Klenner, Manfred, Angela Fahrni, and Stefanos Petrakis. 2009. PolArt: A Robust Tool for Sentiment Analysis. In *Proceedings of NODALIDA 2009*, pages 235–238, Odense.

Loper, Edward, and Steven Bird. 2002. NLTK: The Natural Language Toolkit. In *Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, pages 63–70, Philadelphia.

Moilanen, Karo, and Stephen Pulman. 2007. Sentiment Composition. In *Proceedings of Recent Advances in Natural Language Processing (RANLP) 2007*, pages 378–382, Borovets.

Polanyi, Livia, and Annie Zaenen. 2004. Contextual valence shifters. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*. Springer Verlag (The Information Retrieval Series), Dordrecht, pages 1–9.

Sennrich, Rico, Gerold Schneider, Martin Volk, and Martin Warin. 2009. A New Hybrid Dependency Parser for German. In *Proceedings of the Biennial GSCL Conference 2009*, pages 115–124, Tübingen.

VISL-group. 2008. http://beta.visl.sdu.dk/constraint_grammar.html. Institute of Language and Communication (ISK), University of Southern Denmark.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP) 2005*, pages 347–354, Vancouver.