# Opinion Holder and Target Extraction for Verb-based Opinion Predicates – The Problem is Not Solved

**Michael Wiegand** and **Marc Schulder**
Spoken Language Systems
Saarland University
D-66123 Saarbrücken, Germany
michael.wiegand@lsv.uni-saarland.de
marc.schulder@lsv.uni-saarland.de

**Josef Ruppenhofer**
Dept. of Information Science
and Language Technology
Hildesheim University
D-31141 Hildesheim, Germany
ruppenho@uni-hildesheim.de

## Abstract

We offer a critical review of the current state of opinion role extraction involving opinion verbs. We argue that neither the currently available lexical resources nor the manually annotated text corpora are sufficient to appropriately study this task. We introduce a new corpus focusing on opinion roles of opinion verbs from the Subjectivity Lexicon and show potential benefits of this corpus. We also demonstrate that state-of-the-art classifiers perform rather poorly on this new dataset compared to the standard dataset for the task showing that there still remains significant research to be done.

## 1 Introduction

We present a critical review of previous research in opinion holder and target extraction. *Opinion holders (OH)* are the entities that express an opinion, while *opinion targets (OT)* are the entities or propositions at which sentiment is directed. The union of opinion holders and opinion targets are referred to as *opinion roles*.

In this work we focus on opinion roles evoked by verbs. We examine verbs since opinion role extraction is considered a lexical semantics task and for such tasks verbs are the central focus.

We argue for more lexical resources and corpora that are less biased by domain artifacts. The common practice for producing labeled corpora has so far mostly been extracting contiguous sentences from a particular domain and then labeling those sentences with regard to the entities that were intended to be extracted, i.e. opinion holders and/or opinion targets. In this paper we argue that certain important aspects of the task of opinion role extraction get overlooked if one exclusively considers those corpora that are currently available.

We particularly focus on the relationship between opinion roles and their syntactic argument realization. Previous work hardly addressed this issue since either little variation between opinion roles and their syntactic arguments was perceived on the corpora on which this task was examined, or there were other domain-specific properties that could be used in order to extract opinion roles correctly without the knowledge about opinion role realization.

Currently, there exists only one commonly accepted corpus for English containing manual annotation of both opinion holders and targets, i.e. the MPQA corpus (Deng and Wiebe, 2015). Apart from that, not a single lexical resource for that specific task is available. Moreover, there does not exist any publicly available tool that supports *both* opinion holder and target extraction. Typical applications, such as opinion summarization, however, require both components simultaneously (Stoyanov and Cardie, 2011). These facts indicate that there definitely needs to be more research on the task of opinion role extraction.

In order to stimulate more research in this direction, we present a verb-based corpus for opinion role extraction. The difference to previous datasets is that it has been sampled in such a way that all opinion verbs of a common sentiment lexicon are widely represented. Previous corpora have a bias towards those opinion expressions that are frequent in a particular domain. We demonstrate on two opinion holder extraction systems that performance on the new corpus massively drops compared to their performance on a standard dataset. This shows that current systems are not fit for open-domain classification.

## 2 Opinion Roles and Lexical Semantics

Conventional syntactic or semantic levels of representation do not capture sufficient information that allows a reliable prediction in what argument

positions an opinion role may be realized. This is illustrated by (1) and (2) which show that, even with the PropBank-like semantic roles (i.e. *agent*, *patient*[1]) assigned to the entities, one may not be able to discriminate between the opinion roles.

(1) [Peter]$_{agent}^{OH}$ **dislikes** [Mary]$_{patient}^{OT}$.
(2) [Peter]$_{agent}^{OT}$ **disappoints** [Mary]$_{patient}^{OH}$.

We assume that it is lexical information that decides in what argument position opinion roles are realized. That is, a verb, such as *dislike*, *believe* or *applaud*, belongs to a group with different linguistic properties than verbs, such as *disappoint*, *interest* or *frighten*. However, the realizations of opinion roles observed in (1) and (2) are not the only possibilities. In (3), there is no explicitly mentioned opinion holder while the target is the agent. Such cases are triggered by verbs, such as *gossip*, *blossom* or *decay*.

(3) [These people]$_{agent}^{OT}$ are **gossiping** a lot.

Another type of opinion verb is presented in (4) and (5) where two viewpoints are evoked by the same verb in the same sentence. (4) denotes the sentiment view of *Peter* towards *Mary* while (5) represents the sentiment view of *Mary* towards *Peter* (i.e. *Peter* made *Mary* feel better).

(4) [Peter]$_{agent}^{OH}$ **consoles** [Mary]$_{patient}^{OT}$.
(5) [Peter]$_{agent}^{OT}$ **consoles** [Mary]$_{patient}^{OH}$.

These types of **selectional preferences** (1)-(5) have been observed before including the case of **multiple viewpoint evocation** (4)-(5), most prominently by Ruppenhofer et al. (2008). Yet little research on opinion role extraction has actually paid attention to this issue. One exception is Wiegand and Klakow (2012) who experiment with an induction approach to distinguish cases like (1) and (2). Nonetheless, datasets and lists of types of opinion verbs have not been publicly released.

The above analysis suggests more research on lexical resources is required. In the following, we show that existing resources are not suitable to provide the type of information we are looking for. As a reference of opinion verbs, we use the set of 1175 verbs contained in the **Subjectivity Lexicon** (Wilson et al., 2005). Our main assumption is that the opinion verbs from that lexicon can be considered a representative choice of all kinds of opinion expressions that exists in the English language.

---

[1]By *agent* and *patient*, we mean constituents labeled as $A0$ and $A1$ in PropBank (Kingsbury and Palmer, 2002).

## 3 On the Potential of Existing Lexical Resources

In §2, we demonstrated the need for acquiring more lexical knowledge about opinion verbs for open-domain opinion role extraction. This raises the question whether existing general-purpose resources could be exploited for this purpose. If one considers the plethora of different lexical resources developed for sentiment analysis, i.e. sentiment lexicons listing subjective expressions and their prior polarity (Wilson et al., 2005; Baccianella et al., 2010; Taboada et al., 2011), emotion lexicons (Mohammad and Turney, 2013) or connotation lexicons (Kang et al., 2014), one finds, however, that with respect to opinion role extraction there is a gap. What is missing is a lexicon that states for each opinion verb in which argument position an opinion role can be found.

### 3.1 Sparsity and Other Shortcomings of FrameNet

One resource that has previously been examined for this task is **FrameNet** (Baker et al., 1998). The idea is to identify in frames (which predominantly contain opinion expressions) those frame elements that typically contain either opinion holders or opinion targets. Once this mapping has been established, a FrameNet-parser, such as Semafor (Das et al., 2010), could be used to automatically recognize frame structures in natural language text. By consulting the mapping from frame elements to opinion roles, specific opinion roles could be extracted. Kim and Hovy (2006) followed this approach for a set of opinion verbs and adjectives. Thus, they were able to correctly resolve some problems which cannot be solved with the help of syntactic parsing or PropBank-like semantic roles, such as the role distinctions in (1) and (2). For instance, while the opinion holders in (6) and (7) map to the same frame element *EXPERIENCER*, the PropBank-like semantic roles differ. Unfortunately, the resulting mapping lists from that work are not publicly available.

(6) [Peter $_{EXPERIENCER}$]$_{agent}^{OH}$ **dislikes** [Mary]$_{patient}^{OT}$.
(7) [Peter]$_{agent}^{OT}$ **disappoints** [Mary $_{EXPERIENCER}$]$_{patient}^{OH}$.

Table 1 shows some statistics of our opinion verbs with regard to matched frames and frame elements. Considering that there are 615 different frame elements associated to the different frames[2]

---

[2]This count conflates frame elements of the same name that occur in different frames.

| | |
|---|---|
| # opinion verbs (from the Subjectivity Lexicon) | 1175 |
| # opinion verbs with at least one frame | 691 |
| # different frames associated with opinion verbs | 306 |
| # different frame elements associated with opinion verbs | 615 |

Table 1: Statistics of opinion verbs w.r.t frames and frame elements from FrameNet.

containing at least one of our opinion verbs, it becomes obvious that mapping opinion roles to frame elements is a challenging undertaking.

One major shortcoming of the FrameNet-approach for opinion role extraction is that the current FrameNet (version 1.5) still severely suffers from a data-sparsity problem. For example, approximately 45% of the opinion verbs from the Subjectivity Lexicon are missing from FrameNet (Table 1). Even though there exist ways to expand the knowledge contained in FrameNet (Das and Smith, 2012), there are also conceptual problems with the current FrameNet-ontology (Ruppenhofer and Rehbein, 2012). Since FrameNet is a general-purpose resource, there is no guarantee that frame structures perfectly match selectional preferences of opinion roles. For instance, we found that there are many frames that contain opinion verbs with different selectional preferences. The frame SCRUTINY, for example, typically contains many verbs that take an opinion holder in agent position and an opinion target in patient position (e.g. *investigate* or *analyse*). However, it also contains different verbs, such as *pry*. *Prying* means *to be interested in someone's personal life in a way that is annoying or offensive* (Macmillan Dictionary). Given this definition, we must note that this verb also contains another opinion view (in addition to the one also conveyed by the other verbs in this frame – as exemplified by (8) and (9)), namely that of the speaker of the utterance (condemning the behaviour of the agent of *pry*). As a consequence, the agent of *pry* is also an opinion target while its respective opinion holder is the speaker of the utterance (10).

(8) [The police]$_{OH}$ **investigated** [her]$_{OT}$ thoroughly.

(9) [The press]$_{OH}$ continues to **pry** [into their affairs]$_{OT}$.

(10) [The press]$_{OT}$ continues to **pry** into their affairs. *(OH: speaker of the utterance)*

## 3.2 WordNet Lacking Syntactic Knowledge

At first glance, using **WordNet** (Miller et al., 1990) as a way to acquire knowledge for selectional preferences of opinion verbs seems a better alternative. This resource has a far greater lexical coverage than FrameNet (for example, the set of opinion verbs from the Subjectivity Lexi-con are all contained in WordNet). A straightforward solution for using that resource in the current task would be to group opinion verbs that share the same selectional preferences for opinion holders and targets with the help of the Word-Net ontology graph. One common way of doing so would be the application of some bootstrapping method in which one defines seed opinion verbs with distinct selectional preferences (for instance, one defines as one group opinion verbs that take agents as opinion holders, such as *dislike*, as another group verbs that take patients as opinion holders, such as *disappoint*, and so on) and propagate their labels to the remaining opinion verbs via the WordNet graph. Such bootstrapping on WordNet has been effectively used for the induction of sentiment lexicons (Esuli and Sebastiani, 2006; Rao and Ravichandran, 2009) or effect predicates (Choi and Wiebe, 2014). It relies on a good similarity metric in order to propagate the labels from labeled seed words to unlabeled words.

We experimented with the metrics in Word-Net::Similarity (Pedersen et al., 2004) and found that the opinion verbs most similar to a specified opinion verb do not necessarily share the same syntactic properties. For example, Table 2 lists the 12 opinion verbs most similar to *outrage* and *please*, which are typical opinion verbs that take an opinion holder in patient position and an opinion target in agent position.[3] (They would be plausible candidates for verb seeds for that verb category.) Unfortunately, among the list of similar verbs, we find many opinion verbs which have opinion holder and target in a different argument position, such as *hate* on the list for *outrage*:

(11) [Mary]$^{OT}_{agent}$ **outrages/appals/scandalizes/**... [Peter]$^{OH}_{patient}$.

(12) [Peter]$^{OH}_{agent}$ **hates/fears/loves/**... [Mary]$^{OT}_{agent}$.

From a semantic point of view, the similarities obtained look reasonable. *rage*, *hate* and *dread* bear a semantic resemblance to *outrage*. However, the syntactic properties, i.e. the selectional (argument) preferences, which are vital for opinion role extraction, differ from *outrage*. Word-Net is a primarily *semantic* resource (mainly with a view towards lexical relations rather than valence or argument structure), syntactic aspects that would be necessary in order to induce selectional preferences, are missing. Therefore, we suspect that, by itself, WordNet is not a useful resource for the extraction of opinion roles.

---

[3]We employ the metric by Wu and Palmer (1994).

| outrage: appall, scandalize, anger, <u>rage</u>, sicken, <u>temper</u>, <u>hate</u>, <u>fear</u>, <u>love</u>, alarm, <u>dread</u>, tingle |
| please: delight, enthral, enchant, gratify, <u>signify</u>, <u>obviate</u>, madden, blind, <u>avoid</u>, despair, disagree, crush |

Table 2: The 12 most similar verbs to *outrage* and *please* according to the WordNet::Similarity (underlined verbs do **not** share the selectional preference of the respective target verb).

| verbs | | adjectives | |
|---|---|---|---|
| tokens | types | tokens | types |
| 252 | 113 | 1467 | 302 |

Table 3: Comparison of distribution of opinion verbs and opinion adjectives in the Darmstadt Service Review Corpus (DSRC).

## 4 Text Corpora for Fine-Grained Sentiment Analysis

The previous section suggested that none of those existing lexical resources yield the type of information that is required for opinion role extraction. We now also look at available text corpora and examine whether they reflect opinion verbs in such a way that the problem of opinion role extraction can be appropriately evaluated on them. We start by looking at the review domain.

### 4.1 Why the review domain is not suitable for studying opinion role extraction for verbs

There has been a lot of research on the review domain, which also means that there are several datasets from different domains allowing cross-domain sentiment analysis. However, for more in-depth opinion role extraction evoked by verb predicates, these types of texts seem to be less suitable – despite the plethora of previous publications on opinion target extraction (Hu and Liu, 2004; Jakob and Gurevych, 2010; Liu et al., 2013b; Liu et al., 2013a; Liu et al., 2014). We identified the following reasons for that:

Firstly, the subtask of opinion holder extraction is not really relevant on this text type. Product reviews typically reflect the author's views on a particular product. Therefore, the overwhelming majority of explicitly mentioned opinion holders

| agent of verb | patient of verb | no (direct) relationship |
|---|---|---|
| 21.8 | 44.5 | 33.8 |

Table 4: Proportion of relationships between opinion targets and opinion verbs in the Darmstadt Service Review Corpus (DSRC).

refer to the author of the pertaining review.

Secondly, opinion roles evoked by opinion verbs are less frequent. We extracted all sentences with opinion targets from the *Darmstadt Service Review Corpus (DSRC)* (Toprak et al., 2010)[4] and counted the parts of speech of the corresponding opinion expressions. Table 3 compares the frequency of opinion adjectives and verbs. It shows that adjectives are much more frequent than verbs.

Thirdly, the review domain is typically focused on products, e.g. movies, books, electronic devices etc. This also means that only specific semantic types are eligible for opinion holders and targets, e.g. persons are less likely to be opinion targets. Therefore, much of the research in opinion target extraction relies on entity priors. By that we mean that (supervised) classifiers learn weights for specific entities (typically nouns or noun phrases) of how likely they represent *a priori* an opinion target (Zhuang et al., 2006; Qiu et al., 2011; Liu et al., 2013b; Liu et al., 2014). For example, in the movie domain *Psycho* is very likely to be an opinion target as will be *iPhone* in the electronics domain. However, as such features do not transfer to other domains, they distract research efforts from the universally applicable feature of selectional preferences. Table 4, for example, shows the proportion of different relationships between opinion targets and opinion verbs on DSRC. It shows that there is a considerable number of targets in both agent position (14) and patient position (13) & (15). So, it is not trivial to detect opinion targets here. However, if one looks at typical sentences that fall into these two classes, one finds that entity priors and a few other heuristics would help to solve this extraction problem.

For example, all a supervised classifier would need to learn is that the personal pronoun *I* can never be an opinion target (13) – in the review domain it is typically an opinion holder. (This is a typical entity prior that can be learned.) Otherwise, agents are preferred opinion targets (14) but if the agent is not realized, we simply tag the patient (15). We found that these simple heuristics would manage to correctly identify more than 70% of opinion targets on DSRC (being a dependent of some opinion verb). Under these circumstances, one does not need to know that *recommend* and *stink* have different selectional prefer-

---
[4]We chose this corpus as a typical representative corpus for sentiment analysis in the review domain.

|  | MPQA | VERB |
|---|---|---|
| # sentences | 15753 | 1073 |
| avg. # mentions of the same opinion verb | 6.3 | 1.1 |
| avg. # (explicit) opinion holders per sentence | 0.1 | 0.7 |
| holders in agent position [in %] | 77.4 | 45.8 |
| holders in patient position [in %] | 3.1 | 13.3 |
| missing explicit holder [in %] | 19.6 | 41.0 |
| multiple viewpoint evocation [in %] | 2.6 | 41.0 |

Table 5: Statistics of MPQA and VERB.

ences on opinion targets.

(13) I **recommend** [this site]$^{OT}_{patient}$ to anyone.

(14) [Their programs]$^{OT}_{agent}$ **stink**.

(15) **Avoid** [this institution]$^{OT}_{patient}$ if you are a Canadian student!

These heuristics may work on review datasets, but they become misleading when used in a cross-domain setting, since their predictiveness may be confined to specific domains. For example, in a novel written in the first person, the mere occurrence of *I* is not telling. No mention of *I* in Sentence (16) (taken from *Gulliver's Travels*) represents an opinion holder.

(16) When [I] left Mr. Bates, [I] went down to my father: where [...] [I] got forty pounds, and a promise of thirty pounds a year to maintain me at Leyden: there [I] studied physic two years and seven months [...]

## 4.2 Is the news domain any better?

While we think that the review domain is less suitable for opinion role extraction, the conditions we find on news corpora seem more promising. Typically, news corpora tend to be multi-topic. As a consequence, opinion targets can be of different semantic types. Persons can function both as opinion holders and targets. In other words, corpus artifacts like the ones mentioned in §4.1 are less likely to be helpful in solving the task. The fact that the only corpus with a significant amount of both opinion holders and targets annotated, namely *MPQA 3.0* (Deng and Wiebe, 2015), consists of news text, further lends itself to the usage of that domain. Moreover, we do not have a bias towards adjectives. On the MPQA corpus, for example, we actually found that there are 10% more opinion verb mentions than opinion adjective mentions. This analysis may suggest that the existing MPQA corpus would be suitable for our studies. Yet in the next sections, we show why for the study of opinion roles of opinion verbs, it is advisable to consider yet another corpus.

## 5 Our New Opinion Verb Corpus

With our new corpus for fine-grained analysis, we mainly pursue three goals that, as discussed above, are not sufficiently met by previous resources:

1. Our corpus is designed for the evaluation of opinion role extraction systems focusing on mentions of opinion verbs.
2. It should widely represent various types of selectional preferences.
3. It should appropriately represent multiple viewpoint evocation.

Our new corpus was sampled from the *North American News Text Corpus* (LDC95T21). The dataset comprising 1073 sentences contains 753 opinion holders, 745 opinion targets and 499 opinion targets of a speaker view (e.g. as in (3)). We sampled in such a way that all opinion verbs from the Subjectivity Lexicon were contained (**Goal 1**). To compare: In the MPQA corpus, almost every second opinion verb is unattested.

In order to demonstrate that our new corpus is a more suitable resource in order to study selectional preferences (**Goal 2**) and multiple viewpoint evocation (**Goal 3**), we prepared some statistics regarding mentions of opinion verbs and their properties in the MPQA corpus and our corpus (denoted by VERB). Due to the unavailability of MPQA 3.0, we had to use MPQA 2.0, whose annotation with regard to opinion targets is incomplete. We therefore compare opinion verbs only with regard to their opinion holders. However, given the strong interrelations between opinion holders and targets (Yang and Cardie, 2013), we think that if it is shown that our corpus better represents the versatility of opinion holders, this should (almost) equally also apply for opinion targets.

Table 5 examines the types of argument positions in which an opinion holder is realized. We distinguish between three different roles (already informally introduced in §2): the holder is in **agent position** (example: *dislike*), the holder is in **patient position** (example: *disappoint*) or the holder is **not an argument at all** (example: *gossip*). The latter are cases in which the speaker (or some nested source) is the opinion holder. Table 5 also shows the proportion of verbs with multiple viewpoint evocation and the average frequency of individual opinion verbs. The table clearly shows that on MPQA opinion verbs selecting opinion holders in an agent position are predominant. We think that this is just an artifact of having a corpus of contiguous sentences whereby frequent verbs predominate. VERB, like MPQA, originates from the news domain. The only difference is that it has been sampled so that all opinion verbs of the Subjectivity Lexicon are equally represented (and not only the frequent ones). A look at our new corpus, which represents the set of opinion

verbs of the Subjectivity Lexicon, shows that other types of opinion verbs are actually underrepresented in MPQA. The same can be said about multiple viewpoint evocation. (The number for this latter phenomenon is surprisingly high. We found that the reason of this is that there are many verbs that follow the pattern of *pry* (9)-(10), i.e. conveying both a view of its agent and another view of the speaker, such as *idealize*, *moan*, *overemphasize*, *patronize*, *snub*, *swindle* or *trivialize*.)

We should wonder what impact this bias of opinion role realizations has on building classifiers. If one just focuses on MPQA, then always considering opinion holders in agent position will mean being right in almost $80\%$ of the cases. Similarly, there is no need to consider multiple viewpoint evocation. So, this explains why previous research paid little attention to these issues.

## 6 Details on Annotation

We followed the annotation scheme of Ruppenhofer et al. (2014). It is based on SalsaTigerXML (Erk and Padó, 2004), an annotation scheme originally devised for representing FrameNet-like semantic roles. On a sample of 200 sentences, we measured an interannotation agreement of Cohen's $\kappa = 0.69$ for opinion holders and $\kappa = 0.63$ for opinion targets. The corpus is going to be made publicly available to the research community.

## 7 Some Baselines

We now empirically prove that further research on opinion role extraction is needed. For this proof, we consider the two previously discussed corpora, MPQA and VERB. MPQA is chosen as a training set.[5] It is also the largest corpus. We want to show that despite its size, open-domain opinion role extraction requires some information that is still not contained in that corpus. Almost every second opinion verb from the Subjectivity Lexicon is not contained in that corpus.

In this evaluation, we only consider opinion holders. One reason for this is that opinion holders are less controversial to annotate (this also usually results in a higher interannotation agreement (§6)). Another reason is that there is no publicly available extraction system that covers targets.

For our experiments, we use the sequence labeler from Johansson and Moschitti (2013), *Mul-*

---

[5]The split-up of training and test set on the MPQA corpus follows the specification of Johansson and Moschitti (2013).

| Classifier | MPQA (train+test) | VERB (test) |
|---|---|---|
| MultiRel | 72.54 | 44.80 |
| CK | 62.98 | 43.88 |

Table 6: F-scores of opinion-holder classifiers on the MPQA corpus and the new VERB corpus.

*tiRel*. We chose this classifier since it is currently the most sophisticated system for opinion holder extraction and it is publicly available. *MultiRel* incorporates relational features taking into account interactions between multiple opinion cues. In addition to *MultiRel*, we also consider convolution kernels (*CK*) from Wiegand and Klakow (2012). We include that classifier since it achieved overall better performance than the traditional CRFs on a wide set of experiments (Wiegand and Klakow, 2012) including on cross-domain settings.

In the evaluation, we only consider the opinion holders of our opinion verbs. Recall that we are only interested in the study of opinion roles associated with opinion verbs.

Table 6 shows the results. *MultiRel* produces the best performance on MPQA, but on VERB suffers from a similar domain-mismatch as *CK*. This drop in performance is not only due to the fact that many opinion verbs do not occur in MPQA, but also because the selectional preferences of these uncovered verbs differ from the majority observed in MPQA (Table 5).

## 8 Conclusion

We have argued for more research regarding opinion role extraction involving opinion verbs. We showed that with existing corpora, certain problems, such as the differences in selectional preferences among opinion verbs cannot be properly addressed. One cause for this is that corpora available contain opinion verbs with predominantly one selectional preference. Another is that the corpora have certain characteristics that happen to allow inferring opinion roles for specific text types in the corpus (e.g. entity priors in reviews) but which are not transferable to other text types. In order to study the issue of opinion role realization more thoroughly, we have created a small dataset of sentences in which the opinion roles of opinion verbs from the Subjectivity Lexicon have been annotated. With two state-of-the-art classifiers trained on the large MPQA corpus, we could only produce comparatively poor results on opinion role extractions. This shows that further research on that research task is required.

## References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 2200–2204, Valletta.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics (COLING/ACL)*, pages 86–90, Montréal, Quebec, Canada.

Yoonjung Choi and Janyce Wiebe. 2014. +/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1181–1191, Doha, Qatar.

Dipanjan Das and Noah A. Smith. 2012. Graph-Based Lexicon Expansion with Sparsity-Inducing Penalties. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 677–687, Montréal, Quebec, Canada.

Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A. Smith. 2010. Probabilistic Frame-Semantic Parsing. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 948–956, Los Angeles, CA, USA.

Lingjia Deng and Janyce Wiebe. 2015. MPQA 3.0: An Entity/Event-Level Sentiment Corpus. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 1323–1328, Denver, CO, USA.

Katrin Erk and Sebastian Padó. 2004. A powerful and versatile xml format for representing role-semantic annotation. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 799–802, Lisbon, Portugal.

Andrea Esuli and Fabrizio Sebastiani. 2006. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 417–422, Genova, Italy.

Minqing Hu and Bing Liu. 2004. Mining and Summarizing Customer Reviews. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 168–177, Seattle, WA, USA.

Niklas Jakob and Iryna Gurevych. 2010. Extracting Opinion Targets in a Single- and Cross-Domain Setting with Conditional Random Fields. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1035–1045, Boston, MA, USA.

Richard Johansson and Alessandro Moschitti. 2013. Relational Features in Fine-Grained Opinion Analysis. *Computational Linguistics*, 39(3):473–509.

Jun Seok Kang, Song Feng, Leman Akoglu, and Yejin Choi. 2014. ConnotationWordNet: Learning Connotation over the Word+Sense Network. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1544–1554, Baltimore, MD, USA.

Soo-Min Kim and Eduard Hovy. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text*, pages 1–8, Sydney, Australia.

Paul Kingsbury and Martha Palmer. 2002. From TreeBank to PropBank. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 1989–1993, Las Palmas, Spain.

Kang Liu, Liheng Xu, Yang Liu, and Jun Zhao. 2013a. Opinion Target Extraction Using Partially-Supervised Word Alignment Model. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2134–2140, Beijing China.

Kang Liu, Liheng Xu, and Jun Zhao. 2013b. Syntactic Patterns versus Word Alignment: Extracting Opinion Targets from Online Reviews. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1754–1763, Sofia, Bulgaria.

Kang Liu, Liheng Xu, and Jun Zhao. 2014. Extracting Opinion Targets and Opinion Words from Online Reviews with Graph Co-ranking. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 314–324, Baltimore, MD, USA.

George Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3:235–244.

Saif Mohammad and Peter Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 39(3):555–590.

154

Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. 2004. WordNet::Similarity – Measuring the Relatedness of Concepts. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL–Demonstration Papers (HLT/NAACL)*, pages 38–41, Boston, MA, USA.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion Word Expansion and Target Extraction through Double Propagation. *Computational Linguistics*, 37(1):9–27, March.

Delip Rao and Deepak Ravichandran. 2009. Semi-Supervised Polarity Lexicon Induction. In *Proceedings of the Conference on European Chapter of the Association for Computational Linguistics (EACL)*, pages 675–682, Athens, Greece.

Josef Ruppenhofer and Ines Rehbein. 2012. Semantic frames as an anchor representation for sentiment analysis. In *Proceedings of the Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, pages 104–109, Jeju, Korea.

Josef Ruppenhofer, Swapna Somasundaran, and Janyce Wiebe. 2008. Finding the Source and Targets of Subjective Expressions. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 2781–2788, Marrakech, Morocco.

Josef Ruppenhofer, Julia Maria Struß, Jonathan Sonntag, and Stefan Gindl. 2014. IGGSA-STEPS: Shared Task on Source and Target Extraction from Political Speeches. *Journal for Language Technology and Computational Linguistics*, 29(1):33–46.

Veselin Stoyanov and Claire Cardie. 2011. Automatically Creating General-Purpose Opinion Summaries from Text. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, pages 202–209, Hissar, Bulgaria.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2):267 – 307.

Cigdem Toprak, Niklas Jakob, and Iryna Gurevych. 2010. Sentence and Expression Level Annotation of Opinions in User-Generated Discourse. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 575–584, Uppsala, Sweden.

Michael Wiegand and Dietrich Klakow. 2012. Generalization Methods for In-Domain and Cross-Domain Opinion Holder Extraction. In *Proceedings of the Conference on European Chapter of the Association for Computational Linguistics (EACL)*, pages 325–335, Avignon, France.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 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)*, pages 347–354, Vancouver, BC, Canada.

Zhibiao Wu and Martha Palmer. 1994. Verb semantics and lexical selection. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 133–138, Las Cruces, NM, USA.

Bishan Yang and Claire Cardie. 2013. Joint Inference for Fine-grained Opinion Extraction. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1640–1649, Sofia, Bulgaria.

Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie Review Mining and Summarization. In *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, pages 43–50, Arlington, VA, USA.