Separating Actor-View from Speaker-View Opinion Expressions using Linguistic Features

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 examine different features and classifiers for the categorization of opinion words into actor and speaker view. To our knowledge, this is the first comprehensive work to address sentiment views on the word level taking into consideration opinion verbs, nouns and adjectives. We consider many high-level features requiring only few labeled training data. A detailed feature analysis produces linguistic insights into the nature of sentiment views. We also examine how far global constraints between different opinion words help to increase classification performance. Finally, we show that our (prior) word-level annotation correlates with contextual sentiment views.

1 Introduction

While there has been much research in sentiment analysis on the tasks of subjectivity detection and polarity classification, there has been less work on other types of categorizations that can be imposed upon subjective expressions.

In this paper, we focus on the *views* that an opinion expression evokes. By views, we understand the perspective of the holder of some opinion. We distinguish between the two most common types: expressions conveying sentiment of the entities participating in the event denoted by the opinion word, referred to as **actor views** (e.g. *disappointed* in (1) or *praised* in (2)), and expressions conveying sentiment of the speaker of the utterance, referred to as **speaker views** (e.g. *excelled* in (3) or *wasted* in (4)).

(1) Party members were **disappointed**$_{actor}$ at the election outcome.

(2) All representatives **praised**$_{actor}$ the final agreement.
(3) Sarah **excelled**$_{speaker}$ in virtually every subject.
(4) The government **wasted**$_{speaker}$ a lot of money.

The distinction between those categories is relevant for related tasks in sentiment analysis, most importantly, opinion holder and target extraction. This has already been demonstrated for verbs (Wiegand and Ruppenhofer, 2015). For example, even though the noun *Peter* has the same grammatical relation to the opinion verb in (5) & (6), in the former sentence it is a holder but in the latter it is a target. Similar cases can be observed for opinion nouns (7) & (8) and opinion adjectives (9) & (10). Only the knowledge of sentiment views helps us to assign opinion roles correctly.

(5) [Peter]$_{Holder}$ **criticizes**$_{actor}^{verb}$ Mary.
(6) [Peter]$_{Target}$ **cheated**$_{speaker}^{verb}$ in the exam.
(7) [Peter's]$_{Holder}$ **belief**$_{actor}^{noun}$ is that all people should be treated equally.
(8) [Peter's]$_{Target}$ **misbehaviour**$_{speaker}^{noun}$ is unbearable.
(9) [Peter]$_{Holder}$ is **disappointed**$_{actor}^{adj}$.
(10) [Peter]$_{Target}$ is **intelligent**$_{speaker}^{adj}$.

While the distinction of sentiment views is not new, we put a different emphasis on this task. Our focus is on the prior meaning that opinion words evoke. Hence we consider this as a **word-level** task. Every opinion word from a sentiment lexicon is to be categorized as conveying either an actor or a speaker view. Our aim is to find comprehensive methods to automatically categorize opinion words of various parts of speech (verbs, nouns, adjectives). The resulting lexical resources are indispensable for open-domain categorization. Previous work focused on contextual classification of sentiment views (Johans-

son and Moschitti, 2013). Wiegand and Ruppen-hofer (2015) showed that while prior lexical knowledge of sentiment views is effective in transferring opinion role extractors to other domains, this does not apply to contextual classifiers.

In this work, we focus on linguistic properties for predicting sentiment views. We examine in how far morphological information can be used. Distributional and syntactic information is also considered. In terms of lexical resources, we examine WordNet and FrameNet. We show that information from a sentiment lexicon can give some additional clues.

In order to combine the different features to predict the sentiment views evoked by opinion words we employ supervised classification. As a classifier, we use Markov Logic Networks (Richardson and Matthew, 2006) since they do not only allow us to define features for instances (i.e. opinion words) but also to formulate global constraints between different instances. The latter cannot be expressed by traditional classifiers (e.g. SVM). We examine two types of constraints: consistency between instances that are distributionally similar and consistency between morphologically related instances.

Finally, we also examine the relationship between prior lexical information (i.e. our approach) and contextual annotation in the MPQA corpus.

2 Related Work

The annotation scheme of the MPQA corpus (Wiebe et al., 2005) was the first work to address the distinction between different sentiment views. The two sentiment views are referred to as *direct subjectivity* (=actor view) and *expressive subjectivity* (=speaker view). In subsequent research, some approaches have been proposed to distinguish these two categories in the MPQA corpus. The most extensive work is Johansson and Moschitti (2013). Since MPQA provides annotation regarding sentiment in context, sentiment views are exclusively considered in *contextual* classification. The fact that it is the opinion words that convey those views, as we do in this paper, is not addressed. Unlike in this paper, the focus of Johansson and Moschitti (2013) is also on optimizing a machine-learning classifier, in particular to model the interaction between different subjective phrases within the same sentence.

	Actor View		Speaker View	
Part of Speech	**Freq**	**Perc**	**Freq**	**Perc**
adjective	223	8.9	2279	91.1
noun	487	29.1	1189	70.9
verb	618	52.6	557	47.4

Table 1: Distribution of the different sentiment views.

Some of the lexical resources we examine, i.e. WordNet (§4.1) and FrameNet (§4.2), have also been employed in Breck et al. (2007) who, like Johansson and Moschitti (2013), also deal with contextual (sentiment) classification. However, the authors do not examine in how far these individual resources separate speaker and actor views.

Maks and Vossen (2012b) link sentiment views to opinion words as part of a lexicon model for sentiment analysis. Maks and Vossen (2012a) also examine a corpus-driven method to induce opinion words for the different sentiment views. The authors, however, conclude that their approach, which sees news articles as a source for actor views and news comments as a source for speaker views, is not sufficiently effective.

The work most closely related to our research is Wiegand and Ruppenhofer (2015). Opinion words are categorized according to their sentiment view. Our work substantially goes beyond that previous research: Firstly, Wiegand and Ruppenhofer (2015) only consider distributional similarity for inducing opinion views. In this work, we consider various linguistic features and also compare this with distributional information. Secondly, Wiegand and Ruppenhofer (2015) only consider opinion verbs, while we also consider opinion nouns and opinion adjectives.

Wiegand and Ruppenhofer (2015) distinguish between two types of actor views, *agent views* and *patient views*. The former take their opinion holder as an agent and their target as a patient (typical verbs are *criticize*, *love*, *believe*), while the latter align their roles inversely (typical verbs are *disappoint*, *please*, *interest*). Since this distinction between actor views does not exist among nouns or adjectives, we consider one merged (actor-view) category for all three parts of speech in this paper.

3 Data

We manually annotated all verbs, nouns and adjectives contained in the Subjectivity Lexicon (Wilson et al., 2005) for view type. The dataset comprises 2502 adjectives, 1676 nouns and 1175 verbs. Since our new dataset[1] is an extension of the dataset from Wiegand and Ruppenhofer (2015), we adhere to the annotation process proposed in that paper. That is, the basis of the annotation were online dictionaries (e.g. *Macmillan Dictionary*) which provide both a word definition and example sentences. Each word is either labeled as primarily conveying an actor or a speaker view. (Our categorization is binary.) On a subset of 250 words for each part of speech, we computed an interannotation agreement (Cohen's κ) of 61.9, 71.9 and 60.1 for verbs, nouns and adjectives, respectively. This agreement can be considered substantial (Landis and Koch, 1977). Table 1 shows the distribution of the different sentiment views among the different parts of speech.

The expressions comprising our gold standard do not represent anywhere near the full set of English subjective words with these parts of speech. Otherwise, an automatic categorization would not be necessary in the presence of our gold standard. The classification approach that we propose in this paper, which works well with few labeled training data, would also be helpful for categorizing sentiment views on much larger sets of subjective expressions.

4 Feature Design

4.1 WordNet

WordNet (Miller et al., 1990) is the largest lexical ontology for the English language. It is organized in synsets. However, we want to assign categories to words. Due to the lack of robust word sense disambiguation, in order to use this resource, we consider the union of synsets in which a word with the same part of speech to be categorized is contained.

4.1.1 Gloss Information (GLOSS)

One common way to harness WordNet is by taking into account its glosses. A gloss represents some

explanatory text for each synset, usually some definition of the concept. We use the words from those glosses as features in a supervised classifier. We assume that opinion words conveying the same sentiment view also contain similar glosses.

Glosses are a special type of feature. It is basically a bag-of-words feature set, i.e. a *low-level* feature set, which is known to be sparse yet effective when sufficient training data are used. All the other features presented in this paper are *high-level* features, i.e. more frequently occurring features already being effective if only few labeled data are used. Glosses are one of the most frequently used features for lexicon induction tasks in sentiment analysis (Esuli and Sebastiani, 2005; Andreevskaia and Bergler, 2006; Gyamfi et al., 2009; Choi and Wiebe, 2014; Kang et al., 2014). We will consider them as a baseline, showing that our proposed high-level features are more suitable for our task.

4.1.2 Lexicographer Files (LEX)

Lexicographer files organize the synset inventory of WordNet into a coarse-grained set of semantic categories. In total, there are 45 categories for the three parts of speech we consider.[2] The advantage of such a coarse-grained inventory is that it should require only few labeled training data in supervised classification.

4.2 FrameNet (FN)

FrameNet (Baker et al., 1998) is a semantic resource that has been found useful for subtasks of sentiment analysis related to ours, i.e. opinion holder/target extraction (Bethard et al., 2004; Kim and Hovy, 2006). It includes a large set of more than $1,200$ semantic frames that comprise words with similar semantic behaviour. As a feature we use the frame-membership of the opinion words, assuming that different frames are associated with different sentiment views. We use FrameNet version 1.5.

4.3 Subcategorization Frames (SUB)

Subcategorization frames could also be predictive. For example, actor views demand the presence of an

---

[2]For adjectives there exist only two categories. These are too general for our task. Instead we use the lexicographer files of all nouns and verbs occurring in the glosses of those adjectives.

Type	Affixes Used
Sentiment	*-able, dis-, mis-, over-, under-, -(i)sm*
Neutral	adj → noun: *-cy, -ity, -ness*; adj/noun → verb: *-ize*; verb → adj: *-ed, -ing*; verb → noun: *-ion, -ing*

Table 2: Affixes used as features.

explicit entity that utters some opinion, i.e. the opinion holder. For a speaker view, this entity remains implicit. This should be reflected in the argument valence of the respective opinion words. We employ the subcategorization frames encoded in *COMLEX* (Grishman et al., 1994) for verbs and adjectives, and *NOMLEX* (Macleod et al., 1998) for nouns.

4.4 Morphological Information (MORPH)

As morphological information, we consider derivational affixes. Table 2 lists our choice of prefixes and suffixes. We only included affixes that occurred at least 10 times in our dataset.

We distinguish between sentiment and neutral affixes. The *sentiment affixes* are affixes which, due to their meaning, suggest a sentiment view. For example, *mis-* as in *misinterpret* indicates that the speaker believes that a given interpretation is incorrect. *-able* as in *admirable* has the meaning of *capable of* which corresponds to an evaluation of the speaker. We could only find sentiment affixes for speaker views.

The *neutral affixes* that we use specify which kinds of bases they can combine with. For example, the noun suffix *-ness* as in *foolishness* indicates that the word originates from an adjective (i.e. *foolish*). Even though this knowledge is syntactic, it may give us some clue as to what sentiment view an opinion word conveys. Table 1 shows that adjectives predominantly carry speaker views. Therefore, a noun ending in *-ness* (thus originating from an adjective) may be similarly likely to convey a speaker view.

4.5 Context Patterns (PATT)

Wiegand and Ruppenhofer (2015) proposed patterns for actor-view and speaker-view <u>verbs</u>. For actor views (PATT_actor), they rely on *prototypical opinion holders (protoOHs)*, i.e. common nouns, such as *opponents* or *critics*, that act like opinion holders (Wiegand and Klakow, 2011). If a verb often co-occurs with an opinion holder – Wiegand and Ruppenhofer (2015) take protoOHs as a proxy – then this is a good indicator of being an actor view

(speaker views, per definition, do not have any opinion holder as their dependent). ProtoOHs can similarly be used to extract actor-view nouns and adjectives. For speaker views (PATT_speaker), Wiegand and Ruppenhofer introduced *reproach patterns*, e.g. *blamed for X* as in (11). These patterns can also be applied to nouns (12) but not to adjectives. For the latter, we did not find any pattern. The patterns were applied to the North American News Text Corpus (LDC95T21).

(11) The UN was **blamed for** *misinterpreting$_{verb}$* climate data.

(12) The UN was **blamed for** the *misinterpretation$_{noun}$* of climate data.

4.6 Polarity Information (POLAR)

We also investigate in how far polarity information correlates with sentiment views. This information is obtained from the Subjectivity Lexicon (Wilson et al., 2005). Each opinion word is assigned a polarity type, i.e. *positive*, *negative* or *neutral*.

5 Markov Logic Networks and Global Constraints

Markov Logic Networks (MLNs) are a supervised classifier combining first-order logic with probabilities. MLNs are a set of pairs (F_i, w_i) where F_i is a first-order logic formula and w_i a real valued weight associated with F_i. They build a template for constructing a Markov network given a set of constants C. The probability distribution that is estimated is a log-linear model

$$P(X = x) = \frac{1}{Z} exp \left(\sum_{i=1}^{k} w_i n_i(x) \right) \qquad (1)$$

where $n_i(x)$ is the number of groundings in F_i in x and Z is a normalization constant. As an implementation, we use *thebeast* (Riedel, 2008).

We employ MLNs since they allow us (in addition to including ordinary features, i.e. §4.1-§4.6) to formulate constraints holding between individual instances. Such *global constraints* have been effectively exploited with MLNs in related tasks, such as semantic-role labeling (Meza-Ruiz and Riedel, 2009), anaphora resolution (Hou et al., 2013), question answering (Khot et al., 2015) and discourse-based sentiment analysis (Zirn et al., 2011). We formulate three such constraints. Two of them are

Abbrev.	Constraint as Logic Formula
w2v	$\forall x[\forall y[\forall z[\forall u[[Opin.Word(x) \land Opin.Word(y) \land Word2Vec\text{-}Similar(x,y) \land ViewOf(z,x) \land ViewOf(u,y)] \rightarrow (z == u)]]]]$
lin	$\forall x[\forall y[\forall z[\forall u[[Opin.Word(x) \land Opin.Word(y) \land DekangLin\text{-}Similar(x,y) \land ViewOf(z,x) \land ViewOf(u,y)] \rightarrow (z == u)]]]]$
morph	$\forall x[\forall y[\forall z[\forall u[[Opin.Word(x) \land Opin.Word(y) \land MorphRelated(x,y) \land ViewOf(z,x) \land ViewOf(u,y)] \rightarrow (z == u)]]]]$

Table 3: Global constraints enforcing sentiment view consistency as incorporated in MLNs.

based on the two most effective types of word similarities from Wiegand and Ruppenhofer (2015). The first word similarity measures the cosine of word vectors representing opinion words produced by *Word2Vec*-embeddings (Mikolov et al., 2013). The second word similarity is represented by the metric of Lin (1998), which exploits the rich set of dependency-relation labels in the context of distributional similarity.[3] The third type of consistency considers morphological relatedness by which we understand two words deriving from two different parts of speech but belonging to the same lexical root and therefore carrying similar meaning (e.g. *happiness.noun* and *happy.adj*). We obtain that type of relatedness from WordNet (Miller et al., 1990).

Table 3 lists our constraints. They state that if for two opinion words some similarity or morphological relatedness holds, then these words should convey the same sentiment view. For the two types of word-similarity consistencies we considered the top 3 most similar words for each noun, and the top 5 most similar words for each verb and adjective. These values were determined empirically. For the generation of word vectors, we used 200 dimensions along the default configuration of *Word2Vec*. Word similarity and word vectors were generated from the North American News Text Corpus.

6 Experiments

For our evaluation of supervised classification, we focus on a setting in which only few labeled training data are available. We sampled from our gold standard 20% of the labeled training data. The remaining 80% are used as test data. This process was repeated five times. We report performance averaged over these five (test) samples. We focus on small training sizes since we think that for the given lexicon induction task, we should pursue an approach that requires little human annotation. Moreover, we

show that our approach yields good results despite the absence of large amounts of training data.

6.1 High-Precision Features

Before we evaluate supervised classification, we look for each part of speech at the 10 features with the highest precision (for each of the two views) as displayed in Table 4. This provides a good overview of the quality of different features. Since we do not have an equal class distribution, we also list a baseline-precision that always predicts the sentiment view under consideration. Since this is just an exploratory experiment, we measure precision on the entire dataset. We exclude the WordNet glosses (§4.1.1) from our analysis as we found individual words from glosses too difficult to interpret.

Table 4 shows that features from all feature groups (§4.1-§4.6) achieve a high precision. Subcategorization features (§4.3) are very predictive for verbs conveying actor views. The frame types that are predictive mostly have in common that one of their arguments is some proposition (13)-(17). This is also true for adjectives (18).

(13) PP-HOW-TO-INF: They agree$_{actor}$ [with him]$_{PP}$ [how to solve the problem]$_{HOW\text{-}TO\text{-}INF}$.

(14) NP-TOBE: They believe$_{actor}$ [him]$_{NP}$ [to be honest]$_{TOBE}$.

(15) NP-ADJP-PRED: They consider$_{actor}$ [him]$_{NP}$ [foolish]$_{ADJP\text{-}PRED}$.

(16) NP-TO-INF: He allowed$_{actor}$ [her]$_{NP}$ [to go]$_{TO\text{-}INF}$.

(17) S: They thought$_{actor}$ [he was always late]$_S$.

(18) THAT-S-ADJ: They were aware$_{actor}$ [that he was sick]$_{THAT\text{-}S\text{-}ADJ}$.

FrameNet-frames (§4.2) achieve high precision; but the only frame with good coverage is *Stimulus-focus* for adjectives conveying a speaker view.

There are fewer lexicographer files (§4.1.2) than FrameNet-frames in Table 4, but some of them have high coverage, most notably *LEX_person* for speaker-view nouns and *LEX_feeling* for actor-view nouns. Given the strength of *LEX_person*, we conclude that most opinion nouns denoting persons tend to be speaker views (e.g. *idiot* or *loser*). There are also several predictive lexicographer files whose label seems fairly unintuitive, e.g. *LEX_weather* for

---

[3]WordNet is not a good option for measuring similarity, as many synonyms of the same synset have different sentiment views, e.g. *comment.noun* (actor) vs. *gossip.noun* (speaker).

Speaker View								
adj *(always-predict-this-view prec: 91.1)*			**noun** *(always-predict-this-view prec: 70.9)*			**verb** *(always-predict-this-view prec: 47.4)*		
Feature	Prec	Freq	Feature	Prec	Freq	Feature	Prec	Freq
SUB_EXTRAP-FOR-TO-INF	100.0	104	FN_Killing	100.0	11	MORPH_mis-	92.3	13
FN_Desirability	100.0	48	LEX_animal	91.7	24	LEX_weather	84.2	19
FN_Expertise	100.0	31	FN_Catastrophe	90.9	11	FN_Prevarication	81.8	11
FN_Morality-evaluation	100.0	30	LEX_body	90.9	11	MORPH_over-	78.6	14
FN_Candidness	100.0	20	MORPH_-cy	90.6	32	FN_Killing	76.9	13
FN_Usefulness	100.0	19	MORPH_-ity	90.2	132	FN_Self-motion	75.0	16
FN_Praiseworthiness	100.0	15	LEX_substance	87.0	23	MORPH_-ize	73.5	200
FN_Stimulus-focus	99.1	113	LEX_food	84.2	19	LEX_change	64.8	270
LEX_plant	98.7	79	LEX_person	83.5	267	PATT_speaker	62.7	252
MORPH_-able	98.3	172				FN_Communication-noise	62.5	16
Actor View								
adj *(always-predict-this-view prec: 8.9)*			**noun** *(always-predict-this-view prec: 29.1)*			**verb** *(always-predict-this-view prec: 52.6)*		
Feature	Prec	Freq	Feature	Prec	Freq	Feature	Prec	Freq
FN_Experiencer-focus	81.8	11	FN_Emotion-directed	95.1	41	FN_Experiencer-focus	100.0	21
SUB_THAT-S-ADJ	75.9	54	FN_Experiencer-focus	92.9	14	SUB_PP-HOW-TO-INF	100.0	13
FN_Emotion-directed	74.6	67	PATT_actor	83.0	53	SUB_NP-TOBE	100.0	12
SUB_ADJ-TO-INF	38.7	31	FN_Judgment	73.3	15	SUB_NP-ADVP-PRED	100.0	11
MORPH_-ed	29.2	288	LEX_feeling	65.7	268	SUB_NP-ADJP-PRED	94.4	18
PATT_actor	26.3	335	FN_Medical-conditions	61.5	13	FN_Judgment-direct-address	92.8	14
MORPH_dis-	25.4	71	FN_Hostile-encounter	58.3	12	SUB_NP-TO-NP	92.3	13
LEX_weather	22.2	18	SUB_NOM-INTRANS-RECIP	57.1	14	SUB_NP-TO-INF	91.7	12
LEX_feeling	18.1	695	POLAR_neutral	54.8	93	SUB_NP-ING-OC	90.9	11
SUB_ADJ-PP	17.2	548	LEX_relation	53.8	26	SUB_S	90.4	52

Table 4: High-precision features (*minimal frequency* > 10).

speaker-view verbs or *LEX_animal* for speaker-view nouns. These are not errors, however. They actually concern words that convey opinions in metaphorical usage. For instance, *cloud* (a typical weather verb) conveys a speaker view if it is used metaphorically as in *The stroke clouded memories of her youth.* Nouns denoting animals, such as *bull* and *dragon*, convey a speaker view if they are meant to describe a human being (*She is a real dragon!*). Other noun classes follow this pattern, e.g. *body (parts)* with terms such as *backbone* or *bum*.

Simple morphological features (§4.4) also seem to be meaningful. In particular, the noun suffix *-ity* (occurring 132 times in our set of opinion nouns) is indicative of speaker views. The relevant nouns are derived from adjectives (Table 2) and the set of adjectives predominantly conveys speaker views (Table 1).

Even plain polarity information (§4.6) has some significance. *Neutral* sentiment verbs often convey an actor view, such as *opinion*, *utterance* or *view*.

The fact that the pattern-feature (§4.5) also appears on the list of actor-view nouns and adjectives suggests that it is not only effective for verbs as shown in Wiegand and Ruppenhofer (2015) but also for nouns and adjectives.

Classifier(s)	Description
graph	graph-based induction approach as proposed in Wiegand and Ruppenhofer (2015)
mln_{local}	Markov Logic Networks with only local features, i.e. features from §4
svm	Support Vector Machines using exactly the same features as mln_{local}
mln_{w2v}, mln_{lin}, mln_{morph}	Markov Logic Networks with global constraints from Table 3
mln+graph	Markov Logic Networks that uses the output of *graph* as a further feature

Table 5: Description of the different classifiers.

Finally, we performed an ablation experiment in which we trained a classifier with all of these features in MLNs and compared it to another classifier in which each of the feature groups (*POLAR*, *LEX*, *MORPH* etc.) was removed, one by one. We computed statistical significance (t-test), testing whether the classifier trained on a feature set in which one feature group was removed performs significantly worse than a classifier with all features. We found that, at a significance level $p < 0.05$, this is always the case, with the exception of *LEX* (here, the significance level is $p = 0.0552$). This is proof that features from most feature groups contain information that is to some extent complementary.

6.2 Classification

Table 5 lists the different types of classifiers we consider. As one baseline, we consider the graph-based approach *graph* from Wiegand and Ruppenhofer (2015) which starts with the seeds gained by the surface patterns (§4.5)[4] and then runs label propagation (Talukdar et al., 2008) based on a distributional similarity graph (using the metric by Lin (1998)). *graph* is the only classifier not depending on manually labeled training data. So far, it has only been examined on verbs. As a further baseline, we consider our features from §4 on an SVM. (We use SVM^{light} (Joachims, 1999).) It should be considered as a state-of-the-art classifier that, unlike *mln*, cannot incorporate global constraints (Table 3).

Table 6 shows the results. Both *graph* and *svm* are significantly outperformed. *graph* performs better on verbs (in terms of F-score) than on nouns and adjectives. It is also for these parts of speech that the global constraints *w2v* and *lin* notably improve the performance of *mln*. Global constraints have a lesser impact on verbs. However, a combination of global constraints is effective, as well as a combination of *graph* and *mln*. The best overall results are obtained by the combination of *mln* with global constraints and *graph*. These results suggest that our new features (including global constraints) are useful and complementary to previous work, i.e. *graph*.

Figure 1 compares the feature derived from WordNet glosses (§4.1.1), a standard feature for lexicon induction, with the remaining features we use on a learning curve. This feature performs poorly if only few labeled training data are used. Our proposed feature set is consistently better. The combination of glosses and our proposed features is only helpful if many labeled training instances are used ($> 60\%$).

6.3 Prior Labels and Context Labels

So far, we have considered sentiment views as prior information of words. Now we relate those labels to sentiment views annotated in context. For that, we consider the view annotation in the MPQA corpus.

Table 7 shows that prior labels of opinion words largely coincide with the respective context labels.

Figure 1: WordNet gloss feature vs. the remaining features *(results averaged over all three parts of speech).*

	context					
	verbs		**nouns**		**adjectives**	
prior	actor	speaker	actor	speaker	actor	speaker
actor	**2926**	834	**1671**	802	**269**	81
speaker	119	**810**	151	**2150**	354	**3489**

Table 7: Comparison of prior labels and context labels.

This proves that it is a valid approach to compile lexicons with sentiment views, which can subsequently be used in contextual sentiment-view classification.

However, in Table 7, we still observe mismatches between prior and contextual labels. This mostly concerns actor-view words in speaker-view contexts. We examine this mismatch more closely on nouns *(highlighted in gray)* where this confusion is greatest. In MPQA, most subjective expressions that are annotated are sequences of tokens rather than individual words. We found that the largest set of disagreements derives from the nature of MPQA's contextual annotation. The annotators were asked to label spans that expressed opinions that are salient in the document context. Often these are larger spans composed of multiple smaller subjective expressions. The component expressions were not kept track of because the opinion expressed by the larger span was more salient on the document level.

For example, the annotation of the subjective phrase *this must be a **warning*** as a speaker view (containing the actor noun *warning*), in our opinion is primarily triggered by the epistemic modal verb *must*, which signals that the speaker feels compelled to come to the conclusion that *this is a warn-*

---

[4]Since the surface patterns for speaker views from Wiegand and Ruppenhofer (2015) cannot be applied to adjectives (§4.5), we instead used the effective suffix feature MORPH_-*able* (Table 4) for generating speaker-view seeds of opinion adjectives.
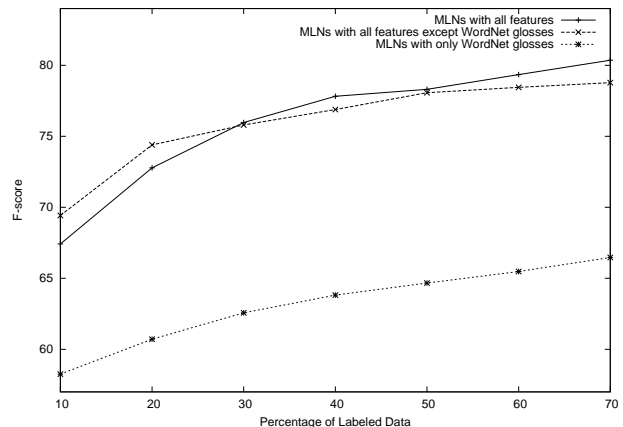
		graph	svm	mln_{local}	mln_{morph}	mln_{w2v}	mln_{lin}	$mln_{morph+lin}$	$mln_{morph+w2v+lin}$	graph	
										$+mln_{local}$	$+mln_{morph+w2v+lin}$
adj	F1	53.1	54.2	69.0°	69.0°	71.1°	72.3°•◇	73.0°•◇†	70.8°	70.4°•◇	**73.6**°•◇
	Acc	91.4°◇	91.6•	89.8	89.7	92.8°•◇	93.3°•◇†	93.3°•◇†	93.3°•◇†	90.6	**93.6**°•◇†
noun	F1	66.1	69.6	72.4°	73.3°•	74.7°•◇‡	73.9°•	75.3°•◇‡	75.4°•◇‡	73.9°•	**76.9**°•◇†‡
	Acc	70.7	78.2	77.7	79.0•	80.2°•◇	79.9°•◇	81.1°•◇†‡	81.6°•◇†‡	78.7°	**82.2**°•◇†‡
verb	F1	71.0°•◇†‡	69.6	69.1	69.6	69.6	69.8	70.2•	70.4•	71.8°•◇†‡	**72.7**°•◇†‡*
	Acc	71.1°•◇†‡	69.7	69.2	69.6	69.7	69.9	70.3•	70.5•	71.9°•◇†‡	**72.8**°•◇†‡*

statistical significance testing (paired t-test, significance level $p < 0.05$) °: better than *svm*; •: better than mln_{local}; ◇: better than mln_{morph}; †: better than mln_{w2v}; ‡: better than mln_{lin}; *: better than *graph (measured for verbs only)*

Table 6: Comparison of different classifiers (*for training, 20% of the labeled data were sampled; the test data are the remaining 80%; this procedure is repeated 5 times; results represent averages over the 5 test samples*).

ing. The actor view of *warning* is not invalidated by this: it is just backgrounded relative to the speaker view introduced by the modal verb, which, going in parallel with its greater prominence, is also the syntactic governor of the verb phrase *be a warning*, of which the actor view *warning* is part. Our evaluation scheme might thus detect a match between our prior annotation for the modal *must* and the MPQA's larger phrase. But since the less prominent actor view was not picked up by the MPQA annotators, our prior annotation has no counterpart. Copula constructions similarly represent instances, where the speaker performs a speech act (e.g. a warning) by using the copular construction (e.g. *This is a warning*). Here, the speaker is identical to the actor of the warning. Practically, it makes no difference whether we call such a case speaker or actor view, as long as we can recognize that the actor is the speaker.

In order to show that the annotation of MPQA focuses on the more salient opinions and thus sentiment views as conveyed by less prominent expressions are not considered (and largely account for the mismatches in Table 7), we designed a supervised classifier whose features indicate whether a mention of an actor-view expression in a subjective phrase is salient. The features are displayed in Table 8.

The key salience features regarding speaker views, i.e. *modal* and *copular*, were already discussed above. Features indicating the salience of the actor-view word address the subcategorization frame of the word. Cases in which there is a person as some subcategorized argument (*personArg*) often imply an opinion holder. The presence of an (explicit) opinion holder indicates an actor view. A proposition as argument of an opinion word is typically the proposition of some opinion holder

(*propArg*) and not of the speaker of the utterance.

For this experiment we take the detection of subjective phrases as given. (Only the information regarding contextual sentiment views is withheld.) This allows us to define features that explicitly look into the entire text span constituting the subjective phrase in which each opinion word is contained. The two length features (*shortPhrase* and *longPhrase*) make use of this information. If a phrase is long, chances are high that there are other more salient opinion words contained than the one under consideration. In short subjective phrases, the presence of other salient words in it is unlikely. This is supported by the fact that the average length of subjective phrases with a speaker view (in which an actor-view opinion noun occurs) are 5.4 tokens while actor-view phrases (that include an actor-view noun) only have an average length of 2.3 tokens.

The majority features (*majActor* and *majSpeaker*) also exploit the information of the entire subjective phrase. We argue that the sentiment view of the phrase is likely to coincide with the view of the majority of the opinion words contained in that phrase.

Table 9 shows how these features separate the mentions of an actor-view opinion noun into contextual actor views and speaker views. We report classification using an SVM (10-fold cross-validation). With only those few features, we largely outperform the baseline always classifying an instance as an actor view (i.e. the majority class). Table 10 displays the precision of each individual feature, supporting that these features are effective. These experiments show that there is indeed a systematic relationship between salience and contextual sentiment views.

Abbreviation	Features Indicating Contextual ACTOR View
personArg	a person is argument of opinion word; persons may indicate opinion holders; (explicit) opinion holders indicate absence of speaker view (***his*** *warning of a catastrophe*)
propArg	a proposition is argument of the opinion word (*warning **that this fish is not fit to eat***); propositions are typically arguments of actor-views words
lightVerb	opinion word is governed by light verb (*they **issued/gave** a warning*), light verbs indicate the presence of an actor outside of the maximal phrase of a subjective noun
shortPhrase	opinion word is part of short subjective phrase (< 3 tokens); short phrases make embedding of another more salient (speaker-view) word unlikely
majActor	majority of other opinion words in subjective phrase are actor-view words
Abbreviation	**Features Indicating Contextual SPEAKER View**
copula	opinion word is part of copula construction (*this **is a warning***) – see discussion in §6.3
modal	opinion word is in modal scope (*this **must** be a warning*) – see discussion in §6.3
emphasis	opinion word is accompanied by emphatic cue, e.g. *!*, quotation (*they gave him a "**warning**"*), (rhetoric) question; emphases typically originate from speaker
precededByAs	preceded by *as* (*this was regarded* [***as** an urgent warning*]$_{as\text{-}phrase}$), *as*-phrase typically occurs as an argument of categorization predicates *regard*, *view*, *see*, *consider* etc. – with these predicates an *as*-phrase often conveys a speaker view, especially since the predicates often have no explicit holder
longPhrase	opinion word is part of long subjective phrase (> 4 tokens); long phrases make embedding of another more salient (speaker-view) word likely
majSpeaker	majority of other opinion words in subjective phrase are speaker-view words

Table 8: Salience features for detecting contextual views (given a mention of a prior actor-view opinion noun).

Majority Class.		Proposed Features			
Acc	65.9	F1	39.7	Acc	78.7	F1	73.5

Table 9: Contextual classification of opinion nouns with a prior actor view (using feature set from Table 8).

Actor View	shortPhrase: 88.5; lightVerb: 85.8; personArg: 81.0; propArg: 74.6; majActor: 72.9
Speaker View	precededByAs: 90.9; majSpeaker: 90.0; modal: 79.6; longPhrase: 73.3; copula: 61.9; emphasis: 57.5

Table 10: Precision of features from Table 8.

7   Conclusion

We examined different types of features and classifiers for the categorization of sentiment views that opinion words convey. We found that many features are effective for this task. A detailed feature analysis provided linguistic insights into the nature of sentiment views. As a classifier, MLNs performed best. This classifier has the advantage that global constraints can be incorporated, which raises classification performance on nouns and adjectives. Our approach outperforms a previously proposed graph-based approach evaluated on opinion verbs. We also demonstrated that prior sentiment views correlate with contextual sentiment views on MPQA.

References

Alina Andreevskaia and Sabine Bergler. 2006. Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses. In *Proceedings of the Conference on European Chapter of the Association for Computational Linguistics (EACL)*, pages 209–216, Trento, Italy.

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.

Steven Bethard, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Dan Jurafsky. 2004. Extracting Opinion Propositions and Opinion Holders using Syntactic and Lexical Cues. In *Computing Attitude and Affect in Text: Theory and Applications*. Springer-Verlag.

Eric Breck, Yejin Choi, and Claire Cardie. 2007. Identifying Expressions of Opinion in Context. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2683–2688, Hyderabad, India.

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.

Andrea Esuli and Fabrizio Sebastiani. 2005. Determining the semantic orientation of terms through gloss classification. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, pages 617–624, Bremen, Germany.

Ralph Grishman, Catherine McKeown, and Adam Meyers. 1994. COMLEX Syntax: Building a Computational Lexicon. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 268–272, Kyoto, Japan.

Yaw Gyamfi, Janyce Wiebe, Rada Mihalcea, and Cem Akkaya. 2009. Integrating Knowledge for Subjectivity Sense Labeling. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 10–18, Boulder, CO, USA.

Yufang Hou, Katja Markert, and Michael Strube. 2013. Global Inference for Bridging Anaphora Resolution. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 907–917, Atlanta, GA, USA.

Thorsten Joachims. 1999. Making Large-Scale SVM Learning Practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 169–184. MIT Press.

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.

Tushar Khot, Niranjan Balasubramanian, Eric Gribkoff, Ashish Sabharwal, Peter Clark, and Oren Etzioni. 2015. Exploring Markov Logic Networks for Question Answering. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 685–694, Lisbon, Portugal.

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.

J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.

Dekang Lin. 1998. Automatic Retrieval and Clustering of Similar Words. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and International Conference on Computational Linguistics (ACL/COLING)*, pages 768–774, Montreal, Quebec, Canada.

Catherine Macleod, Ralph Grishman, Adam Meyers, Leslie Barrett, and Ruth Reeves. 1998. NOMLEX: A Lexicon of Nominalizations. In *Proceedings of EU-RALEX*, pages 187–193, Liège, Belgium.

Isa Maks and Piek Vossen. 2012a. Building a fine-grained subjectivity lexicon from a web corpus. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 3070–3076, Istanbul, Turkey.

Isa Maks and Piek Vossen. 2012b. A lexicon model for deep sentiment analysis and opinion mining applications. *Decision Support Systems*, 53:680–688.

Ivan Meza-Ruiz and Sebastian Riedel. 2009. Jointly Identifying Predicates, Arguments and Senses using Markov Logic. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 155–163, Boulder, CO, USA.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of Workshop at the International Conference on Learning Representations (ICLR)*, Scottsdale, AZ, 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.

Matthew Richardson and Pedro Matthew. 2006. Markov Logic Networks. *Machine Learning*, 62(1–2):107–136.

Sebastian Riedel. 2008. Improving the Accuracy and Efficiency of MAP Inference for Markov Logic. In *Proceedings of the Annual Conference on Uncertainty in AI (UAI)*, pages 468–475, Helsinki, Finland.

Partha Pratim Talukdar, Joseph Reisinger, Marius Pasca, Deepak Ravichandran, Rahul Bhagat, and Fernando Pereira. 2008. Weakly-Supervised Acquisition of Labeled Class Instances using Graph Random Walks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 582–590, Honolulu, HI, USA.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39(2/3):164–210.

Michael Wiegand and Dietrich Klakow. 2011. Prototypical Opinion Holders: What We can Learn from

Experts and Analysts. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, pages 282–288, Hissar, Bulgaria.

Michael Wiegand and Josef Ruppenhofer. 2015. Opinion Holder and Target Extraction based on the Induction of Verbal Categories. In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*, pages 215–225, Beijing, China.

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.

Cäcilia Zirn, Mathias Niepert, Heiner Stuckenschmidt, and Michael Strube. 2011. Fine-Grained Sentiment Analysis with Structural Features. In *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, pages 336–344, Chiang Mai, Thailand.