The Role of Predicates in Opinion Holder Extraction

Michael Wiegand and **Dietrich Klakow**
Spoken Language Systems
Saarland University
D-66123 Saarbrücken, Germany
{Michael.Wiegand|Dietrich.Klakow}@lsv.uni-saarland.de

Abstract

In this paper, we investigate the role of predicates in opinion holder extraction. We will examine the shape of these predicates, investigate what relationship they bear towards opinion holders, determine what resources are potentially useful for acquiring them, and point out limitations of an opinion holder extraction system based on these predicates. For this study, we will carry out an evaluation on a corpus annotated with opinion holders. Our insights are, in particular, important for situations in which no labeled training data are available and only rule-based methods can be applied.

1 Introduction

One of the most important tasks in sentiment analysis is opinion holder extraction in which the entities uttering an opinion, also known as *opinion holders*, need to be extracted from a natural language text. For example, the opinion holders in (1) and (2) are *the vet* and *Russia*, respectively.

1. The owner put down the animal, although the vet had **forbidden** him to do so.
2. Russia **favors** creation of "international instruments" to regulate emissions.

As this is an entity extraction problem it can be considered as a typical task in information extraction. Though there is much work on that subject, most work focuses on data-driven methods. Thus, to a great extent it fails to fully describe certain linguistic aspects of that task.

In this work, we will have a close look at the role of predicates involved in opinion holder extraction. Predictive predicates for this task are, for example, *forbidden* in (1) and *favors* in (2). Unlike previous work, we will examine predicates in isolation. This means we do not consider them as some feature used in a data-driven classifier but as a part of an unsupervised rule-based extraction system which almost exclusively relies on them.

Apart from carrying out a quantitative examination regarding the shape of these predicates and the relationship they bear towards opinion holders, our main contributions of this paper are the investigation of what lexical resources are potentially useful for acquiring predictive predicates and pointing out the limitations of opinion holder extraction based on these predicates.

Our insights are important for building opinion holder extraction systems, in particular, rule-based systems. In particular, we hope that our analysis will provide a realistic rule-based baseline for opinion holder extraction. We also believe that many observations from this paper carry over to languages other than English. For only few of them, there are some corpora annotated with opinion holder information available. For all the remaining languages, rule-based systems leveraging the insights of this paper could be an option for automatic analysis.

2 Related Work

There has been much research on supervised learning for opinion holder extraction. Choi et al. (2005) examine opinion holder extraction using CRFs with several manually defined linguistic features and automatically learnt surface patterns. Bethard et al. (2004) and Kim and Hovy (2006) explore the usefulness of semantic roles provided by FrameNet (Fillmore et al., 2003) for both opinion holder and opinion target extraction. The approaches of those two papers have mostly been evaluated on some artificial data sets. More recently, Wiegand and Klakow (2010) explored convolution kernels for opinion holder extraction.

Rule-based opinion holder extraction heavily relies on lexical cues. Bloom et al. (2007) use a list of manually compiled communication verbs and

Proceedings of the Workshop on Information Extraction and Knowledge Acquisition, pages 13–20,
Hissar, Bulgaria, 16 September 2011.

identify opinion holders as noun phrases having a specific grammatical relation towards those verbs. The rule-based classifiers we evaluate in this work stem from this basic idea. However, we extend this classifier, for example, by considering a more diverse set of predicates and different grammatical relations.

Another work closely related to this study is (Ruppenhofer et al., 2008) which presents a roadmap to both opinion holder and target extraction outlining diverse linguistic phenomena involved in these tasks. In this work, we focus on the role of predicates. Moreover, we also carry out a quantitative evaluation of those related phenomena. Unlike Ruppenhofer et al. (2008), we thus try to identify the most immediate problems of this task. By also considering resources in order to solve these problems we hope to be a helpful guide for practitioners building an opinion holder extraction system from scratch.

3 Data

As a labeled (test) corpus, we use the MPQA 2.0 corpus[1] which is a large text corpus containing fine-grained sentiment annotation. It (mainly) consists of news texts which can be considered as a primary domain for opinion holder extraction. Other popular domains for sentiment analysis, for example, product reviews contain much fewer opinion holders according to the pertaining data sets (Kessler et al., 2010; Toprak et al., 2010). Opinions uttered in those texts usually express the author's point of view. Therefore, the extraction of sources of opinions is of minor importance.

We use the definition of opinion holders as described in (Wiegand and Klakow, 2010), i.e. every source of a *private state* or a *subjective speech event* (Wiebe et al., 2003) is considered an opinion holder. This is a very strict definition and the scores produced in this work can only be put into relation to the numbers presented in (Wiegand and Klakow, 2010). The final corpus comprises approximately 11,000 sentences with more than 6,200 opinion holders.

4 Examination of Predicates

4.1 The Different Types of Predicates

Table 1 displays the distribution of the different predicate types. We divided them into three categories being: unigram predicates (*verb*[2], *noun* and *adj*), multiword expressions of common syntactic structures (i.e. *verb+object*, *verb+prepObject*, *have+object* and *phrasal verb*) and a category for everything else. The table shows that the unigram predicates are most frequent. Since they cover almost 90% of the opinion holder predicate instances, we will focus on these expressions in the following experiments.

4.2 The Different Types of Grammatical Relations

Table 2 shows the distribution of the most frequent grammatical relations between opinion holder and its related predicate listed separately for each unigram predicate type. We use the Stanford parser (Klein and Manning, 2003) for obtaining all syntactic information. The table displays the percentage of that grammatical relation *within* the particular predicate type when it is observed as a predicate of an opinion holder in our labeled data set (*Perc.*)[3], the property of being a fairly reliable relation for a semantic *agent* (*Agent*), and the precision of that grammatical relation in conjunction with that opinion holder predicate type for detecting opinion holders (*Precision*). As a gold-standard of opinion holder predicates we extracted all unigram predicates from our data set that co-occur at least twice with an actual opinion holder.[4]

One may wonder why we did not mark the relation *nsubj* for nouns as *Agent* while the relation is marked as such for the other parts of speech. We found that subjects of predicate nouns can very often be found in constructions like (3). Clearly, *this* is not an agent of *idea*.

3. This is really an unwise **idea**. *[nsubj(This,idea)]*

Table 2 shows that there are some specific grammatical relations that co-occur frequently with opinion holders. These relations are exactly those implying an agent. Moreover, these relations are also the ones with the highest precision.

This insight may suggest using semantic-role labeling (SRL) for this task. We deliberately stick to using syntactic parsing since most publicly available SRL-systems only consider verb

[1]www.cs.pitt.edu/mpqa/databaserelease

[2]Note that by *verb*, we always only refer to full verbs, i.e. auxiliary and modal verbs are excluded.

[3]Note that for verbs we display relations with a lower percentage (>1%) than for nouns or adjectives (>4%) since verb predicates occur much more often.

[4]Singletons may be fairly noisy which is why we omit them.

Predicate Type	Frequency	Percentage	Example
verb	4272	70.89	I **believe** that this is more than that.
noun	948	15.73	This includes a growing **reluctance** by foreign companies to invest in the region.
adj	201	3.34	Ordinary Venezuelans are even less **happy** with the local oligarchic elite.
verb+object	234	3.88	Some officials **voiced concern** that China could secure concessions on Taiwan.
verb+prepObject	58	0.96	The United States **stands on** the Israeli **side** in dealing with the Middle East situation.
have+object	40	0.66	The KMM **had** no **confidence** in the democratic system.
phrasal verb	34	0.56	Washington **turned down** that protocol six months ago.
else	239	3.97	*NA*

Table 1: Distribution of the different opinion predicates.

Type	Relation	Perc.	Agent	Precision	Example
verb	↓nsubj	80.59	✓	47.47	China had always firmly **opposed** the US Taiwan Affairs Act.
verb	↓by_obj	2.69	✓	29.89	The agreements signed in 1960 for Cyprus were **considered** as nonexistent by many countries.
verb	↑rcmod	2.55		10.03	It was the President who **banned** postal voting by all Zimbabweans outside their constituencies.
verb	↓nsubjpass	1.50		8.85	I am **shocked**.
verb	↓dobj	1.24		2.38	Washington **angered** Beijing last year.
verb	↑partmod	1.08		7.13	Mugabe has no moral excuse for shooting people **demanding** a new constitution.
noun	↓poss	45.04	✓	44.56	President Bush's **declaration** touched off questions around the globe.
noun	↓of_obj	10.75		19.06	Through the **protests** of local labor groups, foreign laborers' working rights were protected.
noun	↓nsubj	6.12		6.42	Chavez is a staunch **supporter** of oil production cuts.
adj	↓nsubj	75.12	✓	71.63	We are **grateful** for the strong support expressed by the international community.
adj	↑amod	4.98		6.48	Soldiers **loyal** to the sacked chief of army staff exchanged gunfire with presidential guard units.

Table 2: Distribution of the different grammatical relations (percentage measured within predicate type).

predicates. Given our statistical analysis in Table 1, however, we would exclude a large portion of predicates, i.e. nouns and adjectives, if we used the output of a standard SRL-system.

It is interesting to note that there are also verbs occurring in argument positions that are definitely not agentive, i.e. *↓dobj* and *↓nsubjpass*. We inspected these cases in order to find out whether there is a set of verbs that systematically realizes opinion holders in non-agentive positions. Table 3 lists those verbs we found in our data set. 87.5% of them are also part of the so-called *amuse verbs*, a subset of transitive *psych-verbs* whose object is an experiencer and their subject is the cause of the psychological state (Levin, 1993). The subject, i.e. the cause (this does not even have to be a person), is unlikely to be the opinion holder, whereas the experiencer is often observed to denote such an entity.

4.3 The Different Resources for Opinion Holder Predicates

In this section, we want to examine in how far existing resources contain predicates that usually co-occur with opinion holders. The resources we consider are different in their design and serve diverse purposes. Only one has been specifically designed for opinion holder extraction. For the remaining resources, there may be some modification necessary, for example, by selecting a subset. As we want to examine these resources for an unsupervised (open-domain) rule-based method, these modifications should be pretty simple, fast to implement, and not require extensive knowledge about our particular data set.

4.3.1 Communication Verbs from Appraisal Lexicon (AL)

The communication verbs from Appraisal Lexicon (*AL*) are the only lexicon that has been designed for opinion holder extraction (Bloom et al., 2007). With 260 entries, it is the smallest resource in this paper. Little is known about the creation of this resource (e.g. whether the resource has been optimized for some domain) except that several verb classes from Levin (1993) have been considered.

alienate	concern	exasperate	lure	rile
anger	cow	frighten	obsess	scare
annoy	disappoint	frustrate	offend	shock
astonish	discourage	humiliate	persuade	stunt
baffle	disgust	infuriate	please	suit
bias	disturb	intimidate	rankle	surprise
bother	embarrass	irk	relieve	tear
captivate	enrage	irritate	remind	worry

Table 3: Predicates taking opinion holders in a non-agentive argument position.

appoint	conjecture	admire	correspond	say
characterize	see	marvel	assessment	want
dub	sight	complain	transfer of message	long
declare	judgment	advise	manner of speaking	tell

Table 4: Levin's verb classes taking opinion holders in agentive argument position (only *amuse verbs* take opinion holder in non-agentive positions).

4.3.2 Subjectivity Lexicon (SL)

The Subjectivity Lexicon (*SL*) from the MPQA-project (Wilson et al., 2005) is one of the most commonly used sentiment lexicons. The lexicon contains 8222 subjective expressions from different parts of speech. For our experiments we will only consider its verbs, nouns and adjectives.

This lexicon has been used for various subtasks in sentiment analysis, primarily subjectivity detection and polarity classification (Wilson et al., 2005). It has also been used for opinion holder extraction (Choi et al., 2005; Wiegand and Klakow, 2010) though the lexicon does not contain any annotation specifically designed for this task which is why each entry is considered *some clue* for an opinion in a sentence. In this work, we will even assume each entry to be a predicate predictive for opinion holder extraction.

Since this resource is fairly large, we also consider the subset SL_{strong} consisting of (fairly unambiguous) *strong-subjective* expressions.

4.3.3 Levin's Verb Classes (Levin)

Even though *AL* already considers verb classes from Levin (1993), we constructed a separate subset from that resource for this study. The reason for this is that we found that there are many relevant verbs (e.g. *agree*, *deem* or *disapprove*) not contained in *AL* but that are part of Levin's lexicon. Our selection method was ad-hoc but we did not tune this resource for any particular data set, i.e we included every verb class in our lexicon of which the majority were verbs we would associate *a priori* with opinion holders.

Another important aspect of Levin's work (as already mentioned in §4.2) is that it allows a distinction of verbs taking opinion holders in agentive argument positions and verbs taking them in other positions. We identified *amuse verbs* to be precisely the latter class. (Note that *AL* completely excludes this class.) Admittedly, other resources, such as FrameNet, also encode that distinction. Unfortunately, using FrameNet for an unsupervised classifier would be more difficult. We would need to choose from 1049 (partially overlapping) frames.[5] In Levin's lexicon, we only needed to choose from 193 classes. The final selection is shown in Table 4.

4.3.4 WordNet - Lexicographer files (WN-LF)

WordNet[6] is possibly the largest and most popular general-purpose lexico-semantic ontology for the English language. Most work using this resource focuses on the relationship between the different synsets, i.e. the groups of synonymous words that represent the nodes in the ontology graph. Due to the high number of these synsets, we found it very difficult to select an appropriate subset predictive for opinion holder extraction. This is why we tried to harness another form of word grouping that this resource provides. The *lexicographer files* (*WN-LF*) seem to operate on a more suitable level of granularity. The entire ontology (i.e. the set of synsets) is divided in 44 of such *files* where each file represents a very general semantic word class (e.g. *noun.food* or *verb.motion*). We consider the files *noun.cognition*, *noun.communication*, *verb.cognition* and *verb.communication*. Due to the coarse-grained nature of the *WN-LF*, the resulting set of words contains 10151 words (7684 nouns and 2467 verbs).

Table 5 summarizes the properties of the different resources. Due to the high number of nouns in *WN-LF*, we will evaluate this lexicon both with and without nouns. For all resources only containing verbs, we also use Nomlex (Macleod et al., 1998) to find corresponding noun predicates, e.g.

[5] according to: `http://framenet.icsi.berkeley.edu/`

[6] `wordnet.princeton.edu`

Resource	Abbrev.	Size	Parts of Speech	Description
Subjectivity Lexicon	SL	8222	verbs, nouns and adjs	resource built for sentiment analysis in general
Subjectivity Lexicon - strong	SL_{strong}	5569	verbs, nouns and adjs	subset of SL with exprs. having a strong subjective connotation
Communication Verbs	AL	260 (354)	verbs	resource built for opinion holder extraction
Levin's Verb Classes	Levin	715 (869)	verbs	general purpose resource
WordNet Lexicographer Files	WN-LF	10151	verbs and nouns	general purpose resource
		2467 (2948)	only verbs (from WN)	

Table 5: Properties of the different resources (numbers in brackets denote the size of a resource including nouns obtained by Nomlex extension).

believe (verb) → *belief* (noun), as we already established in Table 1 that nouns play a significant part in the recognition of opinion holders.

4.4 Comparison of Resources

Table 6 displays the performance of the different resources when used in a simple rule-based opinion holder classifier. It classifies a noun phrase (NP) as an opinion holder when the NP is an agent (according to the unambiguous grammatical relations from Table 2)[7] of an entry in a particular lexicon. Only for the *amuse verbs* in *Levin*, we consider the other grammatical relations ↓*nsubjpass* and ↓*dobj*.

The different resources produce quite different results. Surprisingly, *SL* is the lowest performing resource even though it has been used in previous work (Choi et al., 2005; Wiegand and Klakow, 2010). Though the recall increases by adding nouns and adjectives to verbs, the precision notably drops. For the subset SL_{strong} the precision drops slightly less so that the F-Score always increases when the other parts of speech are added to the verbs. Overall, SL_{strong} has a much higher precision than *SL* and its F-Score (considering all parts of speech) is on a par with *SL* even though it is a significantly smaller word list (see Table 5). *SL* is a resource primarily built for subjectivity and polarity classification and these results suggest that the lexical items to imply opinion holders are only partially overlapping with those clues.

Though *AL* and *Levin* are considerably smaller than *SL*, they perform better. Moreover, *Levin* is considerably better than *AL*. In both cases, the extension by noun predicates using Nomlex results in a marginal yet consistent improvement. Unfortunately, the usage of the *amuse verbs* does not produce a notable improvement. We mostly ascribe it to the fact that those verbs occurred only very infrequently (i.e. either once or twice in the entire data set).

[7]By that we mean those relations marked with *Agent*.

Resource	Subtype	Prec	Rec	F1
SL	verb	42.25	27.54	33.34
	verb+noun	38.20	32.20	34.94
	verb+noun+adj	34.30	35.39	34.84
SL_{strong}	verb	**59.80**	20.17	30.17
	verb+noun	56.01	22.92	32.53
	verb+noun+adj	52.71	25.19	34.09
AL	plain	41.88	32.65	36.69
	+Nomlex	41.66	34.32	37.64
Levin	noAmuse	42.59	44.59	43.57
	withAmuse	42.12	45.74	43.86
	withAmuse+Nomlex	41.51	47.74	44.41
WN-LF	verb	33.49	65.44	44.31
	verb+noun	30.19	**71.33**	42.42
	verb+Nomlex	32.97	68.73	**44.56**

Table 6: Performance of the different resources on opinion holder extraction.

WN-LF performs slightly better than *Levin*. Adding the large set of nouns is not effective. The set of verbs augmented by corresponding noun predicates obtained by Nomlex produces better results. The large F-Score of *WN-LF* is only due to a high recall. The precision is comparably low. For this task, another set of predicates maintaining a higher precision is clearly preferable.

4.5 Combination of the Resources

In this section, we combine the different resources (by that we mean taking the union of different resources). For each resource, we use the best performing configuration from the previous evaluation. Table 7 shows the performance of different combinations. As testing all combinations would be beyond the scope of this work, we mainly focus on combinations not using *WN-LF*.

Resource(s)	Prec	Rec	F1
WN-LF *(baseline)*	32.97	68.73	44.56
SL+AL	37.52	52.80	43.68
SL_{strong}+AL	**42.91**	49.69	46.05
SL+Levin	34.56	62.95	44.62
SL_{strong}+Levin	40.97	57.43	**47.82**
AL+Levin	41.49	47.80	44.42
SL+AL+Levin	34.56	62.95	44.62
SL_{strong}+AL+Levin	40.96	57.43	47.82
SL+AL+Levin+WN-LF	29.47	**78.28**	42.82
SL_{strong}+AL+Levin+WN-LF	32.19	75.32	45.10
OraclePRED	46.44	67.83	55.13
OraclePRED*	47.04	68.62	55.82

Table 7: Performance of combined resources.

We seek a classifier with a higher precision than that achieved by *WN-LF*. Combining *WN-LF* with other resources would only result in another increase in recall.

We also want to have an estimate of an upper bound of this method. *OraclePRED* uses all predicates that occur as a predicate of an opinion holder on our test set at least twice. We only consider predicates which have been observed in prototypical agentive positions. *OraclePRED** also uses the knowledge of the predicates from the data set but is not restricted to agentive patterns. That is, we store for each predicate the grammatical relation(s) with which it has been observed (e.g. *oppose*+↓*nsubj* or *anger*+↓*dobj*); we only consider the frequent grammatical relationships from Table 2. Thus, like semantic role labeling, we should be more flexible than a classifier that exclusively considers opinion holders to be in an agentive argument position of some predicate.

Table 7 shows that a combination of resources is indeed beneficial. SL_{strong} and *Levin* produce a higher F-Score than *WN-LF* by preserving a considerably higher precision. Adding *AL* to this set has no effect on performance, since the few predicates of *AL* are already in the union of SL_{strong}*+Levin*. Comparing the performance of the different configurations with *OraclePRED*, we can conclude that the resources that are considered are not exactly modeling opinion holder predicates but a combination of them does it to a large extent. Looking at the false negatives that the best configuration produced (note that we will discuss the issue of false positives in §4.6), we could not really make out a particular group of verbs that this classifier systematically excluded. As far as *Levin* is concerned, however, we assume that the fact that this typology only considers 3000 verbs *in total* also means that many infrequent verbs, such as *ratify* or *lobby*, have simply been excluded from consideration even though their behavior would enable an assignment to one of the existing verb classes.

The performance of *OraclePRED* also shows that opinion holder extraction is a really difficult task as this upper bound is fairly low in absolute numbers. The oracle using the grammatical relations (*OraclePRED**) improves performance only slightly. This is consistent with our experiments using *amuse verbs*.

4.6 Ambiguity of Predicates

In this section we evaluate individual predicates that occur very frequently and also state in which resources these expressions can be found. Table 8 shows that these predicates behave quite differently. The verb *say* is by far the most predictive individual predicate though this is mainly due to its high recall. Other verbs, such as *want*, *believe* or *think*, have a considerably lower recall but their precision is almost twice as high. In terms of coverage, *WN-LF* is the only resource that contains all expressions. This is consistent with its high recall that was measured in previous experiments. On the other hand, $SL_{(strong)}$ only contains a subset of these expressions but the expressions are mostly those with a very high precision.

The individual examination of highly frequent predicates shows that a problem inherent in opinion holder extraction based on predicates is the lacking precision of predicates. In general, we do not think that the false positives produced by our best configuration are due to the fact that there are many predicates on the list which are wrong in general. Omitting a verb with a low precision, such as *say* or *call*, is not an option as it would always heavily degrade recall.

5 Other Clues for Opinion Holder Extraction

In this section, we want to put opinion holder predicates into relation to other clues for opinion holder extraction. We consider two types of clues that can be automatically computed. Both aim at improving precision when added to the clue based on opinion holder predicates since this clue

Pred	In Resources	Prec	Rec	F1
say	AL,Levin,WN-LF	42.52	**13.62**	**20.64**
want	Levin,SL$_{(strong)}$,WN-LF	**83.12**	2.04	3.99
call	Levin,WN-LF	53.66	1.41	2.74
believe	AL,Levin,SL$_{(strong)}$,WN-LF	79.46	1.42	2.79
support	AL,Levin,SL,WN-LF	71.08	0.94	1.86
think	AL,Levin,SL$_{(strong)}$,WN-LF	79.78	1.13	2.24
tell	Levin,WN-LF	35.68	1.21	2.35
know	AL,Levin,SL$_{(strong)}$,WN-LF	66.33	1.04	2.04
agree	Levin,SL$_{(strong)}$,WN-LF	63.64	0.89	1.76
decide	SL,WN-LF	69.81	0.59	1.17

Table 8: Individual performance of the 10 most frequent opinion holder predicates.

already provides a comparatively high recall.

The clue *PERSON* checks whether the candidate opinion holder is a person. For some ambiguous predicates, such as *critical*, this would allow a correct disambiguation, i.e. *Dr. Ren* in (4) would be classified as an opinion holder while *the cross-strait balance of military power* in (5) would not.

4. Dr. Ren was **critical** of the government's decision.
5. In his view, the cross-strait balance of military power is **critical** to the ROC's national security.

For this clue, we employ Stanford named-entity recognizer (Finkel et al., 2005) for detecting proper nouns and WordNet for recognizing common nouns denoting persons.

The second clue *SUBJ* detects subjective evidence in a sentence. The heuristics applied should filter false positives, such as (6).

6. "We do not have special devices for inspecting large automobiles and cargoes", Nazarov **said**.

If an opinion holder has been found according to our standard procedure using opinion holder predicates, some additional property must hold so that the classifier predicts an opinion holder. Either the candidate opinion holder phrase contains a subjective expression (7), some subjective expression modifies the predicate (8), or the proposition that is introduced by the opinion holder predicate[8] contains at least one subjective expression (9).

7. *Angry*$_{Subj}$ residents **looked** to Tsvangirai to confront the government.
8. Thousands **waited** *angrily*$_{Subj}$ to cast their votes.
9. Mr. Mugabe's associates **said** [it was a "*bad*$_{Subj}$ decision"]$_{proposition}$.

[8] We identify these propositions as the yield of an SBAR complement of the opinion holder predicate.

The subjective expressions are again obtained by using the Subjectivity Lexicon (Wilson et al., 2005). Since in our previous experiments the subset of strong subjective expressions turned out to be effective, we examine another clue *SUBJ*$_{strong}$ which just focuses on this subset.

As we assume this kind of subjectivity detection to be very error prone, we also want to consider a related upper bound. This upper bound *allSPEECH* addresses the most frequently found reason for misclassifying an opinion holder on the basis of predicates, namely failing to distinguish between the underlying *objective* and *subjective* speech events. (We will focus on only this error source in this work, since the other error sources are much more infrequent and diverse. Their discussion would be beyond the scope of this paper.) We previously measured a fairly low precision of predicates denoting speech events, such as *say* or *tell*. This is due to the fact that these predicates may not only be involved in subjective speech events, such as (9), but may also introduce objective speech events, such as (6), that typically involve no opinion holder. Our upper bound *allSPEECH* undoes the distinction between different speech events in the gold standard (i.e. it always considers a source of a speech event as an opinion holder). Thus, we simulate how opinion holder extraction would work if this distinction could be perfectly automatically achieved.

Table 9 displays the results of various combinations. For the opinion holder predicates, we consider the best combination of resources from our previous experiments in §4.5 (*PRED*) and the upper bound of predicates (*OraclePRED**). The table shows that adding *PERSON* to *PRED* results in an improved F-Score. The addition of *SUBJ* increases precision while recall drops. *allSPEECH*, on the other hand, causes a boost in performance. Even though the combination of the two upper bounds *OraclePRED** and *allSPEECH* together with the *PERSON* filter would largely increase performance, the total F-Score of 65% shows that it would not completely solve this task.

6 Discussion

If we compare our best fully automatic result, i.e. *PRED+PERSON* with 49.90% (Table 9) with that of data-driven methods using the same corpus and task definition, for example Wiegand and Klakow (2010), who obtain an F-Score of almost 63%, one

Clues	Prec	Rec	F1
PRED	40.97	**57.43**	47.82
PRED+PERSON	48.67	51.21	**49.90**
PRED+SUBJ	48.04	32.77	38.97
PRED+SUBJ$_{strong}$	48.89	23.32	31.58
PRED+PERSON+SUBJ	57.84	29.87	39.39
PRED+PERSON+SUBJ$_{strong}$	**60.13**	21.24	31.39
PRED+allSPEECH	53.79	58.33	55.97
PRED+PERSON+allSPEECH	64.00	53.27	58.14
OraclePRED*+allSPEECH	60.21	67.92	63.83
OraclePRED*+PERSON+allSPEECH	69.67	61.59	65.38

Table 9: Performance of opinion holder predicates in conjunction with other clues.

still notices a considerable gap. Of course, this particular data-driven method should be regarded as an upper bound since it uses a very large labeled training set (§3) and even incorporates some lexical resources for feature engineering we almost exclusively rely on in our rule-based classifier (i.e. AL and SL). This also shows that it is really hard to build a rule-based classifier for opinion holder extraction.

7 Conclusion

In this paper, we examined the importance of predicates from diverse resources for the extraction of opinion holders. We found that strong subjective expressions from the Subjectivity Lexicon combined with a subset of Levin's verb classes contain very predictive words. A classifier extracting noun phrases in an unambiguous agentive position of these predicates results in an opinion holder classifier with both reasonable recall and precision but our upper bound shows that there is still room for improvement. Opinion holders in non-agentive positions are so infrequent in our test set that their consideration is less critical. The classifier based on opinion holder predicates can only be improved by restricting holder candidates to persons. Further filters ensuring subjectivity are too restrictive and thus cause a large decrease in recall. Our exploratory experiments show, however, that some additional improvement in opinion holder extraction could be achieved if subjective speech events could be better separated from objective ones.

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (Software-Cluster) under grant no. "01IC10S01" and the Cluster of Excellence for Multimodal Computing and Interaction. The authors would like to thank Josef Ruppenhofer for interesting discussions.

References

S. Bethard, H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky. 2004. Extracting Opinion Propositions and Opinion Holders using Syntactic and Lexical Cues. In *Computing Attitude and Affect in Text: Theory and Applications*.

K. Bloom, S. Stein, and S. Argamon. 2007. Appraisal Extraction for News Opinion Analysis at NTCIR-6. In *NTCIR-6*.

Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. 2005. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. In *HLT/EMNLP*.

C. J. Fillmore, C. R. Johnson, and M. R. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16:235 – 250.

J. R. Finkel, T. Grenager, and C. Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *ACL*.

J. Kessler, M. Eckert, L. Clarke, and N. Nicolov. 2010. The ICWSM JDPA 2010 Sentiment Corpus for the Automotive Domain. In *ICWSM-DCW*.

S. Kim and E. Hovy. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *ACL Workshop on Sentiment and Subjectivity in Text*.

D. Klein and C. D. Manning. 2003. Accurate Unlexicalized Parsing. In *ACL*.

B. Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.

C. Macleod, R. Grishman, A. Meyers, L. Barrett, and R. Reeves. 1998. NOMLEX: A Lexicon of Nominalizations. In *EURALEX*.

J. Ruppenhofer, S. Somasundaran, and J. Wiebe. 2008. Finding the Source and Targets of Subjective Expressions. In *LREC*.

C. Toprak, N. Jakob, and I. Gurevych. 2010. Sentence and Expression Level Annotation of Opinions in User-Generated Discourse. In *ACL*.

J. Wiebe, T. Wilson, and C. Cardie. 2003. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 1:2.

M. Wiegand and D. Klakow. 2010. Convolution Kernels for Opinion Holder Extraction. In *HLT/NAACL*.

T. Wilson, J. Wiebe, and P. Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *HLT/EMNLP*.