Yes we can!? Annotating the senses of English modal verbs

Josef Ruppenhofer, Ines Rehbein

University of Hildesheim, Potsdam University
ruppenho@uni-hildesheim.de, irehbein@uni-potsdam.de

Abstract

This paper presents an annotation scheme for English modal verbs together with sense-annotated data from the news domain. We describe our annotation scheme and discuss problematic cases for modality annotation based on the inter-annotator agreement during the annotation. Furthermore, we present experiments on automatic sense tagging, showing that our annotations do provide a valuable training resource for NLP systems.

Keywords: modality, subjectivity, WSD

1. Introduction

Much recent work in NLP has been centrally concerned with modality and modal verbs. Most prominent is the area of sentiment analysis (Wiebe et al., 2005) and allied areas such as hedge detection (Medlock and Briscoe, 2007; Morante and Daelemans, 2009), committed belief tagging (Diab et al., 2009), factuality profiling (Saurí and Pustejovsky, 2009), and further areas such as information extraction, where the ability to distinguish between events portrayed as certain or real and others portrayed as uncertain or non-factual is important.

Merely detecting the presence of a modal expression is not always sufficient. For instance, if we want to mark speculative sentences, we might be interested only in the epistemic reading(1b) of sentence (1).

(1) My dad COULD have done it.

a. My dad had the ability/capacity to do it.

b. It's possible that my dad did it.

Polysemy is common among modals and many, if not all of them can express several major kinds of modal meaning. *Can*, for instance, seems to have at least 3 meanings. In its dynamic use, as seen in (1a), it refers to the ability of a person or animal to perform some action, or to the capacity of an instrument or similar. In its epistemic sense, shown in (1b), *can* says that the speaker considers some state of affairs (im)possible given the evidence they have. In its deontic use, shown in (2), *can* is used to talk about permission.[1]

(2) Deontic: You CAN come in now!

By contrast, the form *might* is very specialized: this auxiliary only seems to express epistemic modality.

(3) The director of the provincial Aids Action Unit ... suggested this week the figures MIGHT even be higher than McKerrow's.

[1] In our terminology, we follow (Palmer, 1986) for the most part.

In this work, we present sense-annotations by two annotators of English modal verbs on the 535 documents of the first MPQA corpus release (Wiebe et al., 2005). This choice is motivated by an interest in subjectivity analysis. The annotations in the MPQA corpus reflect the total contextualized understandings of larger spans–e.g. in example (3), the MPQA has not annotated *might* individually but only as part of the span *might even be higher than*. We want to complement the MPQA annotations with sense-annotations of individual modal instances, as part of a larger project of studying the extent to which the interpretation of larger spans can be derived from their parts. Exploring the extent to which subjective meanings are compositional also makes sense from a resource point of view given that the annotation of larger spans results in greater data sparsity. A side benefit of our work is that we provide a partial replication of the MPQA annotations, which mostly reflect the judgment of one annotator rather than an adjudicated gold standard. Finally, our annotations provide training data on modal senses for the news domain and will also be useful to corpus linguists. Our annotations are available for download at `www.uni-hildesheim.de/ruppenhofer/data`.

The remainder of our paper is structured as follows. In section 2., we situate our work in the research context. We go on to describe our annotation scheme and resulting corpus in section 3. and give numbers for inter-annotator agreement as well as a discussion of problematic cases for the annotation of English modal verbs. In section 4., we present experiments on automatic sense tagging, providing evidence that our annotations can be used as training data for NLP systems. In section 5., we compare our annotations to the original ones in the MPQA corpus and in section 6. we conclude.

2. Related work

Baker et al. (2010) developed a modality annotation scheme, a modality lexicon, and built two automatic taggers on the basis of the lexicon and the annotation scheme. Like our annotation scheme, Baker et al. (2010)'s scheme identifies three components of modal meaning: a modal expression, a source and a target. The modality lexicon was produced semi-automatically, starting from a hand-selected seed list of modal words and phrases. The lexicon contains

not only modal verbs but also predicates that exhibit what is called sub-lexical modality, namely, predicates such as *try* and *want*. The inventory of modal categories given to annotators was also somewhat different from ours. The authors basically work with eight types of factivitiy-related modality but split some of them into two categories for the purposes of annotation. For instance, the category Intention is referenced by the two annotation choices "H intends [to make P true/false]" and "H does not intend [to make P true/false]", where H stands for holder and P for the proposition at issue. The resulting 13 categories were presented to the annotators ordered by specificity and annotators were instructed to choose the first applicable one. For instance, one so-called "entailment grouping" that the ordering of categories reflects is {succeeds → tries → intends → is able → wants}. We compare our inventory to Baker et al. (2010)'s below in Section 3.2. Some key practical differences are that in our annotation we allow multiple labeling of a token, that is, explicit underspecification. While we did not introduce any dummy marking for cases where the implicit text producer is the source (or holder, in Baker et al. (2010)'s way of talking), we did label overt occurrences of sources. In the work of Baker et al. (2010), no holders were marked at all for lack of time, though the notion itself was acknowledged to be a key part of modality. Two final differences to note are that Baker et al. (2010)'s annotations are not hand-validated and, unlike the lexicon, they were not made publicly available.

Saurí (2008)'s work on automatically building factuality profiles also involves analyzing modal verbs (which are grouped with modal particles, along with adjectives and adverbs such as *certain* and *certainly*.) In FactBank, the associated lexical resource by Saurí and Pustejovsky (2009), 208 documents are manually annotated according to the notion of event as defined in TimeML (Pustejovsky et al., 2005). Within the FactBank annotations, different uses of modals are distinguishable only if they exhibit different factuality values.

Matsuyoshi et al. (2010) perform, among other things, annotation of modals on Japanese data as part of what they call an extended modality scheme with seven components of annotation: Source, Time, Conditional, Primary modality type, Actuality, Evaluation and Focus.

In corpus linguistic work, quantitative research on modals typically has not relied on sense-annotated instances. For instance, in a study of modals in the British National Corpus, Kennedy (2002) looks at the relative frequencies of modals and the differences in their frequencies across genres. Questions of semantics are investigated only indirectly, by making inferences about the kind of semantics that is associated with the different phrasal patterns in which modals participate. Annotated data of the kind we present will allow research to directly take different senses of modals into account.

3. Annotations

3.1. Format and Tool

All annotations were carried out using the Salto annotation tool on TigerXML-format data (Erk and Pado, 2004). The annotation consisted of labeling terminal and non-terminal nodes of an automatically predicted constituency parse[2] with annotation frames and roles. The interface is shown in Figure 1.

3.2. General Considerations

The linguistic analysis of modality is by no means settled. Von Fintel (2006) notes that "[i]n the descriptive literature on modality, there is taxonomic exuberance". In our annotations, we have so far targeted five different modal lemmas, *can/could*, *may/might*, *must*, *ought*, and *shall/should*. For these modals we assume the sense inventory specified in Table 1. [3]

	can	may	must	ought	shall
epistemic	+	+	+	+	+
deontic	+	+	+	+	+
dynamic	+	-	-	-	-
optative	-	+	-	-	-
concessive	-	+	-	-	-
conditional	-	-	-	-	+

Table 1: Senses of the modals

Compared to Baker et al. (2010)'s scheme, we have fewer categories. However, to a large degree this difference results from the fact that we work with modal verbs only. The English modal verbs do not cover all of the categories in Baker et al. (2010)'s scheme. There are, for instance, no modal verbs expressing Effort or Success in English. We also have not used a category similar to Baker et al. (2010)'s Intention because we have not tackled *will* in our work yet and because the few instances of *shall* in our data did not exemplify a volitional/intentional reading. Our epistemic uses of modals are covered by their Belief category. Unlike them, we have no annotations of (Firm)Belief for unmodalized sentences since we explicitly target modal verbs.

As shown by Table 1, there is overlap/similarity between modal meanings across items. *May* and *can* both express permission, i.e. they have a deontic use. Not too surprisingly, there is no complete interchangeability. For instance, as pointed out by Huddleston (2002, 175), while negation scopes over the possibility meaning of *can*, it is the reverse with *may*, as shown by (4–5).

(4) He CANnot have done it.
'It is not possible that he did it.'

(5) He MAY not have done it.
'It is possible that he didn't do it.'

Differences also exist in certain formulaic uses where the expression of permission by *may* is acceptable but the use of *can* is not or much less so.

(6) I would like, if I may/*?can, to take you on a strange journey.

[2]The data was parsed using the Berkeley parser (Petrov and Klein, 2007)

[3]When distributing the corpus, we will split up *shall* and *should* as they are not fully parallel. *Shall* does not occur in conditionals with auxiliary inversion, for instance.

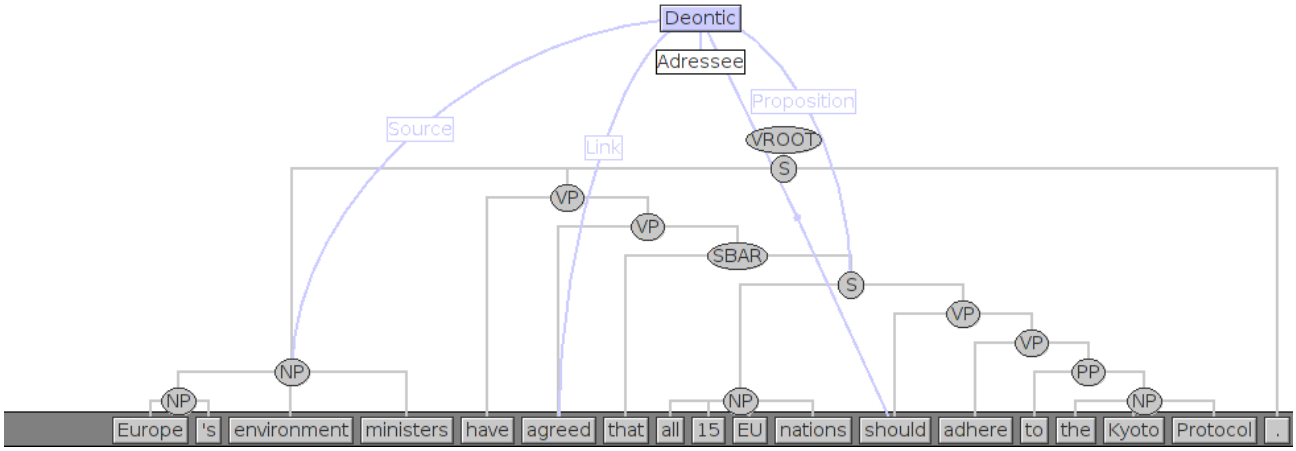

Figure 1: Screenshot of the Salto annotation tool

Further, as already suggested by the previous example, some of the modals have special uses. For instance, the optative and concessive categories are used to account for special uses of the modal *may*:

(7) Optative: Long MAY she live!

(8) Concessive: But, fool though he MAY be, he is powerful.

In the optative use, *may* expresses a wish rather than speculating about a state of affairs or giving permission. The concessive use is a bit like the epistemic use, except that it is clear from the context that the speaker takes the relevant state of affairs as a given rather than considering it a possibility. The conditional use of *should* occurs in *if*-clauses and inversion constructions such as *Should anyone call, please take a message*. Of course, many modals are part of idioms. For instance, *can't* is part of the negative polarity item *can't stand*.

In addition to distinguishing the senses of modals, we annotate further aspects of modality. PROPOSITION is the content that the modal verb modalizes. Typically, it corresponds to a sentence or clause. If the sentence node includes discourse connectives (e.g. *however*) or adverbials that scope higher than the modal, such as *quite frankly* in (9), then we leave these out of the PROPOSITION.

(9) Quite frankly, [he MAY be the better player PROPOSITION].

Although the modal expressions that we annotated do occur nested with each other e.g. in Southern dialects of American English, no embeddings such as *might could* were found in our data. Since we also did not cover modal expressions such as *able* for this work, we have no embeddings in our data at all. Nevertheless, the scheme already covers how such cases would be treated: they would be handled by simply embedding one modal in the other's Proposition. For instance, in *He might be able to help you*, the dynamic (ability) meaning of *able* would occur inside the proposition modalized by *might*.

In our annotation, we omit complementizers from PROPOSITIONs, as shown by (10) but we do include relative pronouns, as shown by (11). SOURCE marks the referents whose point of view on the PROPOSITION is presented. LINKs are predicates of communication or cognition that present the SOURCE of a PROPOSITION. We do not constrain the syntactic relations between LINK and SOURCE, as shown in (12). E.g. if SOURCE is expressed by an NP that depends on a prepositional LINK, we still only annotate the NP inside the PP that refers to the LINK referent. Similarly, in (13), the LINK predicate serves as a depicitve modifier of the SOURCE rather than as a finite clausal predicate taking the SOURCE as a dependent.

(10) [Europe's environment ministers SOURCE] have [agreed LINK] that [all 15 EU nations SHOULD adhere to the Kyoto Protocol PROPOSITION].

(11) But how could Blair have " a Commonwealth report " when no one else had received such a report [that SHOULD be channeled through the Secretariat to all members PROPOSITION] , [asked LINK] [one of the other 53 Commonwealth leaders at the retreat SOURCE] ?

(12) [Under LINK] [the law SOURCE], [a new president SHOULD be installed within three months PROPOSITION].

(13) [Several governments SOURCE] joined the chorus , [saying LINK] [the prisoners SHOULD be granted prisoner of war status under the Geneva Convention PROPOSITION].

3.3. Detailed guidelines

3.3.1. Must

As with *ought to*, we distinguish two senses for *must*.

- The epistemic use is concerned with the speaker being compelled to come to a particular conclusion given her state of knowledge.

  (14) He MUST be home now. It's past 7 o'clock.

- The deontic use is about obligations that are imposed by some source on an agent, who may (15) or may not (16) be expressed as part of the modalized proposition.

  (15) You MUST go home now.

  (16) The door MUST remain closed at all times.

Among cases of epistemic modality, we do not distinguish between cases of what Huddleston et al. (2002) call subjective as opposed to objective inferences. While (17) is not a logically compelling inference, (18) follows from mathematical laws.

(17) The light is on. He MUST be home.

(18) John is 35 and Peter is only a year or two older than John so he MUST be under 40 still.

Within the category of deontic modality, which is concerned with what the world should be like according to some source, we do not make any sub-distinctions based on the nature of the force that impinges on the actor. Thus, example (19) in which an external, relatively impersonal force is the source of the compulsion, is treated no different than example (20), where the pull on the actor comes from within his own psyche.

(19) Dogs MUST be leashed here. A city ordinance requires it.

(20) I really MUST call him. He will be worried.

Further, we do not make any distinction between reports of obligations (21) and occasions where the obligation is being imposed by the speech act (22). Example (15) above is actually ambiguous between those two cases, unlike the contextualized uses in (21) and (22).

(21) Mom says you MUST go home now. It's past 10 pm.

(22) You MUST go home now. I want you gone.

3.3.2. Should

As with ought to we distinguish among two main senses.

- The epistemic use is concerned with the speaker being compelled to come to a particular conclusion given her state of knowledge.

  (23) He SHOULD be home now. It's past 7 o'clock.

- The deontic use is about imposing an obligation or reporting the existence of an obligation.

  (24) You SHOULD go home now.

Both in its epistemic and its deontic uses, *should* is typically weaker than *must*.
In addition to the above two uses, we recognize a special use in conditional constructions, one type of which also involves subject-auxiliary inversion:

(25) SHOULD you see him, please tell him to call me.

(26) If you SHOULD see him, please tell him to call me.

Apart from the above uses, there exist additional minor idiomatic uses. These are exemplified below in (27)-(30) with the labels that Huddleston et al. (2002) use. We have not encountered these in our data. Note that we would simply treat (28)-(30) as deontic cases.

(27) It was odd that she should be so rigid. (emotive)

(28) It is essential that he should be apprehended. (mandative)

(29) Her mother worked for Mr. Morse and so she could never bring her friends to their rooms lest she should annoy the Morses. (adversative)

(30) Her mother worked for Mr. Morse and so she could never bring her friends to their rooms in order that she should not annoy the Morses. (purposive)

3.3.3. Ought to

We distinguish among two senses.

- The epistemic use is concerned with the speaker being compelled to come to a particular conclusion given her state of knowledge.

  (31) He OUGHT TO be home now. It's past 7 o'clock.

- The deontic use is about imposing an obligation or reporting the existence of an obligation.

  (32) You OUGHT TO go home now.

3.3.4. May/might

We distinguish among the following senses:

- The epistemic use is concerned with the speaker's possible conclusions given her state of knowledge. Unlike *must* which is used to express deontic and epistemic necessity, *may* and *might* concern deontic and epistemic possibility.

  (33) He MAY be home now. It's past 7 o'clock.

  (34) These animals MAY be dangerous.

- The deontic use is about giving permission.

  (35) Yes, you MAY come in now.

- The concessive use is in principle also compatible with the epistemic one. We use it in those cases where it's clear from the context that the speaker actually thinks the proposition holds rather than it merely possibly holding.

(36) He MAY be a professor, but he is still a fool.

- The optative is used to express a wish.

(37) MAY you live a 100 years!

3.3.5. Can

We distinguish three senses of *can*, as exemplified in (38–45).

- The dynamic use concerns ability or potential for involvement in events or behavior.

(38) My father can RUN real fast.

(39) The wind CAN still get in.

(40) These animals CAN be aggressive

(41) Garlic blossoms CAN be white or pink.

- The deontic use is about giving permission.

(42) You CAN come in now!

(43) We CAN borrow up to six books at a time.

- The epistemic use concerns the possibility for the speaker to come to certain conclusions.

(44) He says it was Jill but it CAN'T have been.

(45) That CAN'T be her - she's twenty five, that woman is at least 45.

The divisions we make follow those of Huddleston et al. (2002).

The uses of *can* are often divided up differently. For instance, in the grammar of Quirk et al. (1985), the sense division for *can* recognizes the following categories:

- Ability (able to, capable of , know how to)
- Possibility (it is possible for x to)
- Permission (may, be allowed to)

Ability would apply only to (38) and Possibility would cover (39–41). (Quirk et al., 1985)'s Permission category corresponds to our deontic category. Our epistemic cases would be collapsed into their Possibility sense.

The main reason for the non-epistemic analysis by Quirk et al. (1985) is the following. For instance, in a sentence such as (40) there is no speculation about whether or not the animals in question are or are not dangerous, as in (34) above. Instead there is a simple assertion that they have the disposition or propensity to be dangerous on occasion, or that some of them are dangerous. In line with that, (Quirk et al., 1985) use a paraphrase test of *possible-to* to assess the Possibility uses of *can*, rather than a paraphrase test of the form *possible-that*.

(46) It is possible for these animals to be dangerous. (cf. 40)

(47) He says it was Jill but it is not possible for her to have done it. (cf. 44)

Overall, *can* is not as widely usable in the clear epistemic cases where *may* can occur. Example (48) is very odd compared to (44–45).

(48) ??She CAN be home now. It's past 7 o'clock.

It is only in negated uses such as (44) and (45) that *can* looks plausibly like an epistemic modal that assesses the factuality of past or present states of affairs.

At an earlier stage of this research we tried to make a four-way distinction between Ability, Possibility, Epistemic and Deontic uses, which would have constituted a combination of Huddleston et al. (2002) and Quirk et al. (1985). However, we but found that agreement was very poor and so we decided to stick with Huddleston et al. (2002)'s distinctions for the purposes of the present work. Nonetheless, as will be shown and discussed more below, *can* was the most difficult modal to get agreement on.

Finally, we note that *could* can have all the meanings of *can*, although in some cases they are only possible in contexts of shifted tense (49-54). A notable difference between *can* and *could* is that *could* seems to readily allow for an epistemic reading in assertive contexts such as (54).

(49) My father COULD run really fast when he was younger.

(50) Water COULD still get in.

(51) I knocked and she said I COULD come in.

(52) Back then, Poinsettias COULD only be red or yellow but now they have created blue ones, too.

(53) I said that COULDN'T be her - she was twenty five, while that woman was at least 45.

(54) The NHL star hinted he COULD be in the lineup.

3.4. Agreement

We measured inter-annotator agreement between the two annotors, determining average percentage agreement and Cohen's kappa. As suggested by Table 2, for most modals the distribution of senses was highly skewed, resulting in high percentage agreement on the many clear cases but making it difficult to achieve a high kappa value. We omit *ought to* from the table because there are only 4 instances, on all of which the annotators agreed. Recall that we cannot compare our results to Baker et al. (2010)'s work since the annotation there was done by a single annotator.

	items	kappa	%-agreement
may/might	195	0.621	0.89
must	183	0.848	0.98
shall/should	182	0.602	0.96
can	598	0.614	0.77

Table 2: Kappa and average percentage agreement per modal

In what follows we show confusion matrices for two of the individual predicates. Table 3 shows that for *can*, there was

very little confusion between deontic and epistemic readings. The majority of cases involved uncertainty about the boundaries between the dynamic and the deontic senses and between the dynamic and the epistemic senses.

	Cond.	Deont.	Dynam.	Epist.
Conditional	0	0	1	0
Deontic	0	78	**32**	4
Dynamic	0	**46**	271	**15**
Epistemic	0	5	**33**	113

Table 3: Confusion matrix for *can*

Example (55) shows an example where one annotator chose the dynamic category and the other the deontic one. On the deontic reading, there is a sort of prohibition against considering Al Qaeda a state party based on the consideration that they are a terrorist group. On the dynamic reading, the speaker reports on the failure of efforts to think of Al Qaeda in a way that makes them appear as a state party.

(55) ' It will not change their material life on a day-to-day basis : they will continue to be treated well because that 's what the United States does , " said Fleischer , who noted neither group would be granted prisoner of war status . And , " Al Qaeda is an international terrorist group and CANNOT be considered a state party to the Geneva Convention . Its members therefore are not covered , " by the accord , said the spokesman .

Example (56) exemplifies a case of disagreement between an epistemic and a dynamic reading. In that example, there is a sort of scope-ambiguity between the epistemic marker and the existential reading of *coups*. On the first reading, with the epistemic operator outscoping the existential quantification, it is possible that any coups that occur will destabilize the region. On the second reading, with the existential quantification outsocping the epistemic operator, we say about any coups that might occur that they would have the power to destabilize the region. Similarly, a sentence such as "The article claims that a sun storm could cause 20 x more economic damage than Katrina" would be said not to present a speculation on what might be an ongoing event but to report on the generic power of a sun storm to do damage.

(56) Venezuela is this hemisphere 's second-oldest democracy . At a time when democracy is losing currency in many countries , particularly Argentina , coups COULD be potentially destabilizing to the region . " We are happy to collectively have overcome the era of coup d'etats in the region , and when events in Venezuela took on the appearance of a military coup there was a reaction by everyone , " said Brazil 's President Fernando Henrique Cardoso .

In the case of *may*, too, there was no difficulty in distinguishing epistemic and deontic uses (Table 4). The main problem was the distinction between the epistemic category and the special concessive sense. An example of a disagreement is given in (57), where there was uncertainty whether the text author conceded what was presented in the previous sentence: that some irregular forces captured in battle need not be considered POWs.

	Concess.	Deont.	Optat.	Epist.
Concessive	10	1	0	**10**
Deontic	0	6	0	0
Optative	0	0	1	0
Dynamic	0	0	0	1
Epistemic	**4**	2	0	160

Table 4: Confusion matrix for *may*

(57) The Geneva Convention does contemplate that some irregular forces captured in battle need not be considered POWs. That MAY well apply to members of al-Qaeda , a free-floating band of terrorists. But not all of those at Gitmo are al-Qaeda men.

One of the annotators used about one and half as many concessive labels as the other, indicating a possible personal bias and/or a lack of clarity in the guidelines. The epistemic and concessive categories are, however, closely related and indeed candidates for merging, as pointed out in section 3.2.

About *shall* not much of interest can be said since there was such a strong majority sense in the form of the deontic uses, which accounted for 169 of the 182 instances. We therefore dispense with displaying a confusion matrix. The same goes for *must*, which had 171 agreed-upon deontic uses among the 183 instances.

4. Automatic Sense Tagging

To provide proof of concept and show that our sense categories are meaningful and can be learned by an automatic system, we implemented a simple modality tagger and trained it on the annotated data.

4.1. Experimental setup

In our experiments we used a 10-fold cross-validation setup where, for each fold, we split the annotated data for each of the modal verbs in a training set (90%) and a test set (10%). Then we trained a maximum entropy classifier[4] on features extracted from the training set and predicted the senses for each modal verb in the test set. We extracted three different feature types: target-specific features, context features and syntactic path features. Table 5 gives a short description of each feature type.

4.2. Results

Table 6 presents results for the automatic annotation of modal verb senses. To facilitate comparability, the first two rows show the interannotator agreement (IAA) for the human annotators (κ and % agreement). Row 3 displays the most frequent sense baseline (the results one would get

[4]Downloadable from http://incubator.apache.org/opennlp.

Target-specific features:	
i	word form, lemma and POS of the target modal
ii	syntactic category of parent/grandparent/ grandgrandparent node
iii	word, lemma and POS of the leftmost child node of parent/grandparent/grandgrandparent
Context features:	
i	bag of words/lemmas/POS context to the left and to the right of the target
ii	word/lemma/POS of token in sentence-initial position
Path features:	
i	syntactic categories along the path from target to root node in the syntax tree
ii	combinations (bigrams/trigrams) of the root path features (above)
iii	syntactic categories along the path to the leftmost child node for all non-terminal nodes to the right of the modal (siblings right)
iv	word form/lemma/POS for the leftmost child node
v	path to the leftmost child node for all non-terminal nodes to the left of the modal (siblings left)
vi	word form/lemma/POS for the leftmost child

Table 5: Feature types used for automatic sense tagging

when always predicting the most frequent sense for a particular modal verb). The baseline is already quite high for the modals *must* and *shall/should*, which both have the deontic reading as their dominant sense in our data. Such a skewed distribution makes it hard for machine learning methods to beat the baseline as there is only little training data for the less frequent senses.

Rows 4-6 show the impact of each of the individual feature types on classification accuracy. Rows 7-10 present results for using all three, target-specific features, context and path features, for varying context sizes. Here we see that a small context window is sufficient while larger context sizes seem to hurt performance.

Rows 11-13 show results for combinations of only two feature types. For *can* we obtain best results (68.70%) when using target-specific features and context features only, while including path features leads to a decrease in performance of around 2%. For *may/might* we see a different picture. Here the path features seem to encode more important information. Using only target-specific features and path features yields an improvement of around 2% over using a combination of all three feature types. This clearly shows the importance of tuning the features to the specific target word. For *shall/should* we managed to achieve a small improvement when using only a subset of the path features (row 14). For *must*, however, we were not able to improve on the baseline.

These results are quite promising, considering that we put only little effort in developing the system. The main goal of our experiments was to show that our sense inventory represent a meaningful categorisation and that the annotated data can be used as a resource for the development of a modality tagger. There is room for improvement with regard to the feature set as well as to the algorithmic side of the system, but both is beyond the scope of this paper. Future work might explore the use of external knowledge sources like WordNet or FrameNet for incorporating semantic information in the feature set. To beat the strong baseline for *must*, it is necessary to address the class imbalance problem.

5. Comparison to the original MPQA annotations

Comparing our annotations to that in the MPQA, we find that the two are very close in coverage: about 80% of the modal instances that we labeled are also covered by a span in MPQA.[5] One significant exception is *can*, a third of whose instances are not part of a subjective span in the MPQA. Table 7 illustrates the point mentioned above that many instances of modals are only annotated as part of a larger span. As shown by the total, only about 30.8% of all modal instances are in a span by themselves. The remaining cases are the ones where it will be interesting to see how the word sense of the modal contributes to the interpretation of the span.

	Total occurrences			only modal in span		
	ESE	DSE	OSE	ESE	DSE	OSE
can	125	26	6	7	0	0
cannot	69	9	2	12	0	0
could	153	23	2	36	0	0
couldn't	0	0	0	0	0	0
may	104	5	1	40	0	0
might	53	2	1	28	0	0
must	163	10	1	58	0	0
shall	11	0	0	1	0	0
should	241	14	0	99	1	0
ought to	4	0	0	3	0	0
Total	**923**	89	13	**284**	1	0

Table 7: Modals and spans in the MPQA

6. Conclusions

We have presented sense annotations of English modal verbs in the MPQA corpus. These annotations can be used in multiple ways. They can serve as data for corpus linguistic studies or as training data for word sense disambiguation. Importantly, they can be a starting point for analyzing whether the interpretation of the larger spans of text that are marked in the MPQA corpus can be derived compositionally.

7. Acknowledgements

Part of this work was supported by a grant from the DFG awarded to SFB 632:Information structure at Potsdam University.

[5] Due to the different coverage, the numbers we present for our annotated instances in Table 2 do not match those presented in Table 7 for the MPQA corpus.

		can	may/might	must	shall/should
1	IAA (κ)	61.40	62.10	84.80	60.20
2	IAA (% agreement)	77.00	89.00	98.00	96.00
3	baseline (most freq.)	**52.17**	**82.86**	**93.50**	**90.32**
4	target	66.23	83.81	91.00	90.97
5	context	58.70	80.95	93.50	90.97
6	path	57.25	85.24	93.50	90.97
7	context=3	66.52	83.81	93.50	90.65
8	context=5	66.67	83.81	93.50	90.65
9	context=7	66.38	82.38	93.50	90.65
10	context=10	65.80	83.33	93.50	90.32
11	context=3, target	**68.70**	81.43	93.50	91.29
12	context=3, path	59.42	83.33	93.50	90.65
13	context=0, target, path	66.09	**85.71**	93.00	91.29
14	context=0, siblings left/right	66.67	84.29	93.50	90.32
15	context=0, root/siblings left	64.78	85.24	91.00	**91.61**
16	context=0, root/siblings right	66.23	**85.71**	93.00	91.29

Table 6: Accuracies for automatic prediction of modal verb senses for different feature settings

8. References

K. Baker, M. Bloodgood, B. Dorr, N. Filardo, L. Levin, and C. Piatko. 2010. A modality lexicon and its use in automatic tagging. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, pages 1402–1407.

M. Diab, L. Levin, T. Mitamura, O. Rambow, V. Prabhakaran, and W. Guo. 2009. Committed belief annotation and tagging. In *ACL-IJCNLP '09: Proceedings of the Third Linguistic Annotation Workshop*, pages 68–73, Morristown, NJ, USA. Association for Computational Linguistics.

K. Erk and S. Pado. 2004. A powerful and versatile xml format for representing role-semantic annotation. In *Proceedings of LREC 2004*.

R. Huddleston, G.K. Pullum, et al. 2002. *The Cambridge Grammar of English*. Cambridge: Cambridge University Press.

G. Kennedy. 2002. Variation in the distribution of modal verbs in the British National Corpus. *Using corpora to explore linguistic variation*, 9:73.

S. Matsuyoshi, M. Eguchi, C. Sao, K. Murakami, K. Inui, and Y. Matsumoto. 2010. Annotating event mentions in text with modality, focus, and source information. *Proc. of the 7th International Language Resources and Evaluation (LREC'10)*, pages 1456–1463.

B. Medlock and T. Briscoe. 2007. Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of ACL-07, the 45th Annual Meeting of the Association of Computational Linguistics*, pages 992–999, Prague, CZ, June. Association for Computational Linguistics.

R. Morante and W. Daelemans. 2009. Learning the scope of hedge cues in biomedical texts. In *BioNLP '09: Proceedings of the Workshop on BioNLP*, pages 28–36, Morristown, NJ, USA. Association for Computational Linguistics.

F. R. Palmer. 1986. *Mood and modality*. Cambridge University, Cambridge, England.

S. Petrov and D. Klein. 2007. Improved inference for unlexicalized parsing. In *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2007)*.

J. Pustejovsky, B. Ingria, R. Sauri, J. Castano, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, and I. Mani. 2005. The specification language timeml. *The language of time: A reader*, pages 545–557.

R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language.* Longman, New York.

R. Saurí and J. Pustejovsky. 2009. Factbank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227–268.

R. Saurí. 2008. *A Factuality Profiler for Eventualities in Text.* Ph.d., Brandeis University.

K. Von Fintel. 2006. Modality and language. *Encyclopedia of Philosophy, second edn, MacMillan, Detroit. URL: http://mit. edu/fintel/www/modality. pdf*.

J. Wiebe, T. Wilson, and C. Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 1(2):0–0.