Annotating Subjective Content in Meetings

Theresa Wilson

School of Informatics
University of Edinburgh
Edinburgh, U.K. EH9 8LW
`twilson@inf.ed.ac.uk`

Abstract

This paper presents an annotation scheme for marking subjective content in meetings, specifically the opinions and sentiments that participants express as part of their discussion. The scheme adapts concepts from the Multi-perspective Question Answering (MPQA) Annotation Scheme (Wiebe et al., 2005; Wilson, 2008), an annotation scheme for marking opinions and attributions in the news. The adaptations reflect the differences in multiparty conversation as compared to text, as well as the overall goals of our project.

1. Introduction

Applications such as meeting browsers and meeting assistants aim to identify, extract, and summarise *meeting content* — information about what happens and what is discussed in meetings. Some meeting content is primarily objective, for example, information about what topics are discussed (Hsueh and Moore, 2006) and who is assigned to work on a given task (Purver et al., 2006). However, another type of meeting content that is important is the *subjective content* of meetings, that is, the opinions and sentiments that the participants express during discussion in the meeting. Recognizing subjective content is important because, intuitively, it seems that such information would help with existing meeting-browser tasks, such as decision detection (Hsueh and Moore, 2007). But subjective content in and of itself is also interesting and important to extract and summarise. We would like to know not only what a particular decision was but who supported or opposed the decision. Imagine asking a meeting assistant not only to summarise the major ideas that were discussed but also the pros and cons expressed about those ideas.

To extract and summarise the subjective content of meetings, we first need to be able to identify when something subjective is being said and also to recognize the type of subjective content that is being expressed (e.g., positive or negative sentiment). We also need to be able to identify the *source* and the *target* of the subjectivity—who the subjectivity is attributed to and what it is about. Although it is likely that most of the time the speaker is expressing his or her own opinions, it is not unusual for the speaker to report someone else's opinion or to be speaking on behalf of the group. For example, in (1) below, the speaker is reporting the opinion of the company, and in (2), the speaker is reporting information from a user study about remote controls. In example (3), the speaker is reiterating an opinion that the group as a whole holds.

(1) The first one is that um uh the company's decided that teletext is outdated uh because of how popular the internet is.

(2) Um people uh additionally aren't aren't liking the appearance of their products

(3) Also we talked earlier about R_S_I and wanting to prevent um any sort of like Carpal Tunnely kind of thing

In this paper, we present an annotation scheme for marking subjective content in meetings. The annotation scheme adapts concepts from the Multi-perspective Question Answering (MPQA) Annotation Scheme (Wiebe et al., 2005; Wilson, 2008), an annotation scheme developed for marking opinions and attributions in news articles. The adaptations reflect the differences in multiparty conversation as compared to text, as well as our overall project goals, which are focused more on recognizing positive and negative sentiments than on other types of subjectivity.

In the sections that follow, we first give an overview of the MPQA Annotation Scheme, followed by a discussion of the things we needed to consider as we set about adapting the concepts in the MPQA Scheme to multiparty conversation and our project goals. After that, we present our new annotation scheme for marking subjective content in meetings and an inter-annotator agreement study to show that the annotations can be marked reliably. We discuss the interaction between the subjectivity annotations and the dialogue act tags in the annotated corpus. The paper ends with a discussion of related work and conclusions.

2. Overview of MPQA Annotation Scheme

The MPQA Annotation Scheme is centered around the concept of *private state* (Quirk et al., 1985). A private state is any internal mental or emotional state, including opinions, beliefs, sentiments, emotions, evaluations, uncertainties, and speculations, among others. In its most basic representation, a private state can be described based on its functional components: the state of an *experiencer* holding an *attitude* optionally toward a *target* (Wiebe, 1994).

The annotation scheme presented in (Wiebe et al., 2005) is a detailed, expression-level representation of private states and attributions that adapts and expands the more basic functional-component representation. The annotations in the scheme are represented as frames, with slots in the frames representing various attributes and properties. The initial MPQA scheme contains four annotation

frames: *direct subjective frames*, *expressive subjective element frames*, *objective speech event frames*, and *agent frames*. In (Wilson, 2008), the MPQA scheme is extended to include two new types of annotation frames: *attitude frames* and *target frames*.

The direct subjective frame and the expressive subjective element frame are both used for representing private states, but they capture distinct ways that private states are expressed. Direct subjective frames are used to mark expressions that explicitly refer to private states and expressions that refer to speech events[1] in which a private state is expressed. The phrase "have doubts" in (4) is an example of an expression that explicitly refers to a private state. In (5), the phrase "was criticized" refers to a speech event in which a private state is being expressed, as does the phrase "said" in (6). The word "criticized" conveys that a negative evaluation was expressed by many people, even though their exact words are not given. With "said" in (6), it is the quoted speech that conveys the private state of the speaker, specifically the phrase "a breath of fresh air." Expressive subjective element frames are used to mark expressions that indirectly express private states, through the way something is described or through a particular wording. The phrase "a breath of fresh air" is an example of an expressive subjective element, as is the phrase "missed opportunity of historic proportions" in (7).

(4) Democrats also have doubts about Miers' suitability for the high court.
(5) Miers' nomination was criticized from people all over the political spectrum.
(6) "She [Miers] will be a breath of fresh air for the Supreme Court," LaBoon said.
(7) This the nomination of Miers is a missed opportunity of historic proportions.

Although private states are often expressed during speech events, not all speech events express private states. The objective speech event frame in the MPQA scheme is used to mark speech event phrases that refer to these objective speech events. In sentence (8), an objective speech event is marked on the word "said."

(8) White House spokesman Jim Dyke said Miers' confirmation hearings are set to begin Nov. 7.

The agent frame in the scheme is used to mark noun phrases that refer to sources of private states and speech events. The source of a private state is the experiencer of the private state, and the source of a speech event is its speaker or writer. In (4) above, "Democrats" would be marked as an agent, as would "people all over the political spectrum" in (5) and "LaBoon" in (6).

All of the above annotation frames contain various attributes used to further characterize each expression that is annotated. Both private state frames, for example, include attributes for capturing the intensity of the private state being expressed and the polarity of the expression that

Table 1: Attitude Types in the MPQA Scheme

Sentiment	Agreement
Positive Sentiment	Positive Agreement
Negative Sentiment	Negative Agreement
Arguing	**Intention**
Positive Arguing	Positive Intention
Negative Arguing	Negative Intention
Speculation	**Other Attitude**

is marked. One attribute that is included in all the annotation frames is the *nested source* attribute, which represents a key part of the MPQA annotation scheme. We describe this attribute below; details on the other frame attributes can be found in (Wiebe et al., 2005).

As previously mentioned, the source of a private state is the experiencer of the private state, and the source of a speech event is its speaker or writer. However, in textual discourse such as the news, there are frequently *levels of attribution*. For example, in (4) above, it is according to the writer of the sentence that the Democrats have doubts. Similarly, in (5) is it according to the writer that people are criticising the nomination. The *nested source* attribute captures these levels of attribution. In sentence (4), both the direct subjective frame ("have doubts") and the agent frame ("Democrats") are marked with the attribute $nestedsource = \langle writer, democrats \rangle$, where $writer$ and $democrats$ are unique identifiers that represent those agents in the discourse. Similarly, in (6) the expressive subjective element frame ("breath of fresh air"), the direct subjective frame ("said"), and the agent frame ("LaBoon") are all marked with the attribute $nestedsource = \langle writer, laboon \rangle$. In the example sentences above, there are no more than two levels of attribution; sentence (7) only has one level for the writer of the sentence. However, in the news domain, it is not uncommon to find three or even more levels of attribution.

The last two types of annotation frames in the MPQA scheme are the attitude frame and the target frame (Wilson, 2008). The attitude frames are linked to direct subjective frames. The purpose of an attitude frames is to capture the attitude being expressed overall by the private state to which it is linked. Similarly, target frames are linked to attitude frames; they are used to capture the target of the attitudes to which they are linked. The types of attitudes that are included in the attitude frame representation are listed in Table 1.

To date, the MPQA Annotation scheme has been used to annotate a corpus of 535 news articles (about 10,000 sentences) [2]. The MPQA annotations have been used in sentence-level subjectivity classification, phrase-level subjectivity and sentiment recognition, and source identification.

---

[1] A speech event is considered any event of speaking or writing.

3. Adapting the MPQA Scheme to Multi-party Conversation

As we set about adapting the MPQA Scheme for marking subjective content in meetings, there were several questions that we needed to consider. First, what type of subjective content would be most valuable to mark? To answer this question, we considered the goals of the end application we were envisioning, an automatic meeting assistant. Ideally, a meeting assistant would be able to extract and summarise information such as who supported or opposed a particular decision and what were the pros and cons behind a certain idea. To extract this kind of information the system will need to be able to identify positive and negative opinions, evaluations, and emotions, as well as agreements and disagreements. Although other types of subjectivity may also be informative, those listed above are the most important for our purposes, and this should be reflected in the new annotation scheme.

Also relevant to the question of what types of subjectivity we should mark is whether there are *new* types of subjective content that appear in conversational discourse that are not encountered in discourse such as the news. The answer to this is *questions*. In multiparty conversation such as we find in meetings, participants often ask each other questions. Questions obviously can be used to elicit objective information, such as in (9), but they also can be used to elicit subjective information, as in (10).

(9) What shape are the buttons?

(10) What do you think about the design?

Example (10) doesn't express a private state for the asker of the question, but the expected response to the question will be subjective. Because of this, it will be helpful for the end system to be able to identify subjective questions: Identifying subjective questions may in turn help to identify the subjective utterances that follow. Also, subjective questions use much of the same subjective terminology (e.g., "think") as subjective statements, and thus would be a potential source of noise if they were excluded from consideration.

The next question to consider was whether there was a need to represent sources of private states in the new subjective content annotations. In conversation, most of the opinions and sentiments that are expressed will likely be those of the current speaker. However, as previously mentioned, there are certainly instances where the speaker reports opinions that are attributed to someone else. Thus, the new annotation scheme at least needs to be able to differentiate between opinions belonging to the speaker and opinions being reported by the speaker. Is there a need to represent the nesting of sources, as is done in the MPQA Corpus? For spoken conversation, we think this is not important. Rarely will there me more than two levels of sources. If a source is marked that is not the speaker, the nesting will be implicit. The final question that we considered was what granularity of subjectivity annotation would be most appropriate. The MPQA annotations are expression-level annotations. Would expression-level annotations be appropriate

Subjective Utterances
positive subjective
negative subjective
positive and negative subjective
uncertainty
other subjective
subjective fragment
Objective Polar Utterances
positive objective
negative objective
Subjective Questions
positive subjective question
negative subjective question
general subjective question

Table 2: AMIDA Subjectivity Annotation Types

or would larger units such as turns be a better choice to annotate? The more fine-grained the annotations are, the better the subjective content is pinpointed. However, the more fine-grained and detailed the annotations are, the more time consuming they are to produce. Would it be important or even feasible to mark the spans that refer to the sources and targets of opinions? Or, should source and target information just be captured as attributes on the subjectivity annotations? After exploring the meeting data and considering different levels of annotation, we decided on *utterance-level* annotations. For these annotations, *utterance* is defined loosely. An *utterance* may be a single phrase or expression, but whenever possible it is a sentence or proposition with references to the source and target of the subjectivity included in the span that is marked. Sources and targets are then marked as attributes of the subjectivity annotations.

4. AMIDA Scheme for Annotating Subjective Content in Meetings

In this section we describe the final scheme we developed for marking subjective content in multiparty conversation. Table 2 lists the annotation types that are included in the scheme. There are three main categories of annotations, *subjective utterances*, *objective polar utterances*, and *subjective questions*. We describe each of these in more detail below.

4.1. Subjective Utterances

A subjective utterance is a span of words (or possibly sounds) where a *private state* is being expressed, either through choice of words or prosody. There are six types of subjective utterances in the scheme: *positive subjective*, *negative subjective*, *positive and negative subjective*, *other subjective*, *uncertainty*, and *subjective fragment*.

The *positive subjective* category includes agreement, positive sentiments (emotions, evaluations, and judgments), positive suggestions, arguing for something, beliefs from which positive sentiments can be inferred, and positive responses to subjective questions. The *negative subjective* category encompasses private states that are the opposite of

those in the positive subjective category, specifically, disagreement, negative sentiments, negative suggestions, arguing against something, beliefs from which negative sentiments can be inferred, and negative responses to subjective questions. Example (11) contains two positive subjective utterances and one negative subjective utterance. Each annotation is indicated by a pair of angle brackets.

> (11) Um ⟨**POS-SUBJ** it's very easy to use⟩. Um ⟨**NEG-SUBJ** but unfortunately it does lack the advanced functions⟩ ⟨**POS-SUBJ** which I I quite like having on the controls⟩.

The *positive and negative subjective* category is for marking cases of positive and negative subjectivity that are so closely interconnected that it is difficult or impossible to separate the two. For example, some subjective words or phrases inherently evoke both a positive and negative sentiment. One example is the word *bittersweet*. Other examples come from when the grammatical structure makes it difficult to separate the positive and negative subjectivity into two utterances that clearly capture both the positive and the negative. This is the case with example (12) below. The positive suggestion is to use the remote only to control the television. The negative suggestion is to not use the remote to control the VCR, DVD player, etc.

> (12) Um ⟨**POSNEG-SUBJ** they've also suggested that we um we only use the remote control to control the television, not the VCR, DVD or anything else⟩.

Other than positive and negative subjectivity, the only other type of subjectivity that we are specifically identifying is uncertainty. We included this category with the idea that being able to identify when individuals are uncertain might be useful for recognizing when there are problems or when things remain undecided. Below is an example of an uncertain utterance.

> (13) Um ⟨**UNCERT** I'm not entirely sure what the corporate colour is⟩.

Contrary to its name, the *other subjective* category is not used to mark all other expressions of private states that do not fall into the above categories. Rather, it is used to mark neutral emotions or evaluations, as well as beliefs, arguments, and speculations that are not indirectly conveying a positive or negative sentiment. An example of an emotion that is not always positive or negative is surprise. Example (14) is an utterance that would also be marked as other subjective. In this example, a belief or opinion is being express that it is neither positive or negative.

> (14) ⟨**OTHER-SUBJ** I think one factor would be production cost⟩.

The *subjective fragment* category is used to mark short fragments that are clearly subjective, either because of subjective words or phrases or because of the prosody used when expressing the fragment, but there is not enough being said to determine one of the other subjective categories or what the subjectivity is about. There is a subjective fragment marked in (15).

> (15) ⟨**SUBJ-FRAG** Looks kind of⟩ Yeah .

4.2. Objective Polar Utterances

Objective polar utterances are statements or phrases that describe positive or negative factual information about something without conveying a private state. The sentence *The camera broke the first time I used it* gives an example of negative factual information; generally, something breaking the first time it is used is not good. An example of a sentence with positive factual information is *The camera lasted for several years past its warranty*. Utterances that express positive factual information are marked with the *positive objective* category. Those that express negative factual information are marked with the *negative objective* category.

4.3. Subjective Questions

Subjective questions are questions in which the speaker is eliciting the private state of someone else. In other words, the speaker is asking about what someone else thinks, feels, wants, likes, etc., and the speaker is expecting a response in which the other person expresses what he or she thinks, feels, wants, or likes. A subjective question may be a yes/no question, as in example (16) below, or it may be a more open-ended question, as in example (17).

> (16) Do you like the large buttons?
> (17) What do you think about the large buttons?

In our annotation scheme, we distinguish between three types of subjective questions: *positive subjective questions*, *negative subjective questions*, and *general subjective questions*. A positive subjective question is a question that is specifically trying to elicit the positive private state of someone else, and a negative subjective question is trying to elicit a negative private state. For example, (16) above is a positive subjective question. It can easily be changed into a negative subjective question: *Do you hate/not like the large buttons?* Positive subjective and negative subjective questions do not need to be yes/no questions. For example, the question *What do you like about the buttons?* would also be a positive subjective question because it is specifically trying to elicit an answer that is expressing a positive private state. Subjective questions that do not ask specifically about the positive or negative private state of someone else are *general subjective questions*. Question (16) above is an example of a general subjective question.

4.4. Sources and Targets

In our new annotation scheme, sources and targets are represented conceptually as enumerated attributes of subjective utterances and objective polar utterances. Table 3 lists the categories of sources and targets that can be marked. Sources and targets are marked only on the subjective utterances and the objective polar utterances.

The source category *speaker* is used if the current speaker is the source of the subjective utterance or the objective polar utterance. The source for the annotations in examples (18) and (19) is *speaker*. In (18), it is implicit that the source is the speaker of the utterance; in (19) that the speaker is the source is explicitly indicated by the pronoun "I".

Sources	Targets
speaker	meeting
speaker speaking for group	meeting project
other meeting participant	remote design
specific external entity	remote design project
general external entity	previous statement/idea
	speaker-self
	other

Table 3: AMIDA Source Types and Target Types

(18) ⟨**NEG-SUBJ SOURCE=SPEAKER** Finding them is really a pain, you know⟩.

(19) Um ⟨**UNCERT SOURCE=SPEAKER** I'm not entirely sure what the corporate colour is⟩.

The source of a subjective utterance or an objective polar utterance may also be some entity external to the meeting. If this is a specific entity, for example, the company or the speaker's parents, the source is marked as a *specific external entity*. In example (20), the source of the negative subjective utterance is the company, which is a *specific external entity*. An external entity may also be general, for example, "people" or "the man on the street." If this is the case, as in example (21), the source is marked as a *general external entity*.

(20) The first one is that um uh ⟨**NEG-SUBJ SOURCE=SPECIFIC-EXTERNAL** the company's decided that teletext is outdated⟩ uh because of how popular the internet is.

(21) Um ⟨**NEG-SUBJ SOURCE=GENERAL-EXTERNAL** people uh additionally aren't aren't liking the appearance of their products⟩

If the source of the subjective utterance or objective polar utterance is one or more of the meeting participants other than the speaker, then the source is marked *other meeting participant*. If the source of the subjective utterances or objective polar utterance is the speaker plus one or more of the other meeting participants (e.g., "Tom and I", "we") then the source is marked *speaker speaking for group*.

There are seven possible target categories that can be marked on a subjective utterance or an objective polar utterance (Table 3). The *meeting* target category is used when the subjectivity is specifically about the activity or the progress of the meeting itself. For example, a strong suggestion that the participants move onto the next topic would have *meeting* as its target. The target of the subjective utterance below is *meeting*.

(22) Shall we sh well ⟨**POS-SUBJ SOURCE=SPEAKER TARGET=MEETING** we'll stick to kind of your area for now⟩.

The meetings that we developed this annotation scheme for are scenario meetings from the AMI corpus (Carletta et al., 2005a). For these meetings, participants play the role of a design team and hold a series of meeting in an instrumented meeting room. Occasionally, participants make a subjective comment related to some aspect of the overall meeting project, as in (23) below. For these utterances, the target is marked as *meeting project*.

(23) ⟨**NEG-SUBJ SOURCE=SPEAKER TARGET=MEETING-PROJ** Unfortunately we're not allowed to talk outside the meeting room⟩.

The *remote design* and *remote design project* target types are task related categories. In the AMI scenario meetings, the participants are designing a remote control. If the subjectivity is specifically being expressed about some aspect of the remote design, then *remote design* is the target. For subjectivity that is expressed about things related to the remote design project other than the actual design of the remote, e.g., marketing strategy or the project budget, the target is *remote design project*.

Another common target for subjectivity is a *previous statement or idea*. This will typically be the target category used when the subjectivity category is agreement or disagreement, but other types of subjectivity may also be directed toward previous statements or ideas raised earlier.

When the target of the subjective utterance or objective polar utterance is the speaker of the utterance, e.g., the speaker saying something negative about his or herself, then the target is *speaker-self*, as in (24).

(24) ⟨**NEG-SUBJ SOURCE=SPEAKER TARGET=SPEAKER-SELF** I'm kind of pathetic with things like this⟩.

4.5. Overlapping Annotations

It is possible for the annotation types described above to overlap or nest. For example, if the speaker expresses a private state about someone else's private state this will lead to nested annotations. This is the case for (15). The speaker is expressing an opinion ("I think a recurring theme here is …"). In addition, the speaker is also reporting the opinion of the company ("the company wants …something that's fashionable and sleek and trendy").

(25) ⟨**OTHER-SUBJ SOURCE=SPEAKER TARGET=REMOTE DESIGN** I think a recurring theme here is ⟨**POS-SUBJ SOURCE=SPECIFIC EXTERNAL TARGET=REMOTE DESIGN** the company wants it to be [disfmarker] wants us to make something that's fashionable and sleek and trendy ⟩ ⟩

Other sorts of nesting can occur as well, for example, the speaker expressing his or her own private state through prosody while asking a subjective question.

5. Agreement Study

To evaluate whether the annotations in the above scheme can be annotated reliably, two annotators independently annotated two meetings from the AMI corpus using the NITE

	Kappa	% Agreement
Subjective Utterances (excluding fragments)	0.56	79
Positive Subjective	0.58	84
Negative Subjective	0.62	92
Positive Subjective + Positive Objective	0.58	83
Negative Subjective + Negative Objective	0.68	93
Subjective Question	0.56	95

Table 4: Interannotator agreement for the AMIDA subjectivity annotations

XML Toolkit (NXT) (Carletta et al., 2005b). Although annotations are marked on the meeting transcript, annotators were instructed to listen to the meeting audio and to view the meeting videos as part of the annotation process.

When performing the subjective content annotations, annotators were not given a fixed set of units or unit boundaries to annotate. Instead, the annotators chose the spans that they felt best captured the concepts described in the scheme. Because of this, the sets of annotations that the two annotators marked are different, which makes evaluating agreement less straightforward. One possible method for calculating agreement is to measure the precision and recall of one annotator's tags with respect to the other's. Although this method is informative, it does not take into account the agreement that is expected by chance. Instead, we chose to measure agreement based on the dialogue act segments already marked in the corpus. We found that only a small percentage of the subjectivity annotations marked by each annotator (13% for annotator A, and 27% for annotator B) actually crossed dialogue act segment boundaries. Between the two meetings in the study, there are a total of 1,889 dialogue act segments.

Because it is possible for a dialogue act segment to contain more than one subjectivity annotation, we measured agreement for each annotation type separately. Table 4 shows the agreement for key annotation types measured in terms of Kappa (Cohen, 1960) and percent agreement. Agreement for whether a segment contains a subjective utterance is 0.56 kappa. The annotators have similar agreement for positive subjective utterances and subjective questions. Interestingly, agreement for whether a segment contains a negative subjective utterance is higher, 0.62 kappa, suggesting that negative subjectivity is easier to recognise, or at least less ambiguous, than positive subjectivity. Hypothesising that some of the disagreement might be due to confusion between the positive/negative subjective categories and the positive/negative objective categories, we also calculated agreement after conflating the two positive categories and the two negative categories. Although this did not lead to improved agreement for recognizing the combined positive categories, it did improve agreement for the combined negative categories, indicating that there is some confusion between negative subjective and negative objective utterances.

6. Subjectivity Annotations and Dialogue Acts

To date, 20 scenario meetings (almost 12 hours) from the AMI corpus have been annotated with the scheme presented in Section 4.. Out of the 19,071 dialogue act seg-

ments in the meetings[3], 42% were annotated with (i.e., overlap with) one or more of the subjective content annotations. Table 5 shows the percentage of total dialogue act segments and the percentage of annotated segments with different annotation types. Interestingly, the largest category of annotations by far is the *positive subjective* category. This is due in part to agreement being included in the positive category. However, it also seems to to be a property of the type of discourse. In the news, for example, negative subjectivity is more prevalent. Evidently the opposite is true for meeting data, at least for the types of meetings that were collected for the AMI corpus.

One last thing we want to explore is the intersection between dialogue act tags and the subjective content annotations. Dialogue act tags capture the intention of the speaker. Although dialogue act coding schemes vary, some schemes include labels specifically for marking when the intention of the speaker is to express something subjective. This is also true for the AMI dialogue act coding scheme. The AMI dialogue act tagset has a total of 15 labels. Included in tagset are the *Suggest*, *Assessment*, *Be Positive*, and *Be Negative* labels. The *Suggest* and *Assessment* labels are used for marking suggestions and assessments, things that we expect would also be captured by the new subjectivity annotations. The *Be Positive* and *Be Negative* labels are used to mark utterances in which the speaker's intention is to make an individual or the group feel more or less happy. We would also expect this sort of information to be captured by the new annotations. This raises the questions of how the subjective dialogue act tags and the new subjectivity annotations overlap, and how much new information is being captured by the subjectivity annotations.

To answer this question, we looked at how many segments for each of the different dialogue act tags overlapped with subjective utterance annotations and subjective question annotations. This information is presented in Table 6. Three types of dialogue act tags are excluded from the analysis: *stall*, *fragment*, and *other*. Looking at the table, we see that the majority of dialogue acts that we expected to be subjective based on their dialogue act tags are indeed subjective. However, none of the majorities are overwhelming: There are still quite a few dialogue act segments that we would expect to be subjective that are not. For example, around 35% of suggestions are not marked as subjective, and close to 45% of assessment are not marked as subjective. On the other hand, over 40% of informs are marked as sub-

---

[3]Segments that contained only vocal sounds and segments from the "Drawing exercise" portion of meetings were excluded.

	% Total DA Segments	% Annotated DA Segments
Subjective Utterances	36.6	87.6
Subjective Question	3.5	8.4
Subjective Utterances + Questions	39.3	94.2
Positive Subjective	22.0	52.6
Negative Subjective	8.5	20.4
Uncertainty	4.0	9.7
Other Subjective	3.3	7.9
Objective Polar Utterances	2.6	6.3

Table 5: Percentage of total dialogue act (DA) segments and percentage of annotated segments with the different annotation types.

	% DAs overlapping with	
	Subj Utter	Subj Quest
inform	41.8	0.8
elicit inform	28.6	13.3
suggest	61.8	3.6
offer	12.9	0.8
elicit offer/suggest	12.9	47.9
assess	56.3	0.5
elicit assess	27.8	56.0
comment understand	13.8	1.2
elicit comment understand	27.3	18.2
be positive	26.0	0.9
be negative	67.9	0.0
backchannel	10.6	0.0

Table 6: Percentage of each dialogue act (DA) type that overlap with subjective utterance and subjective question annotations.

jective. These results reinforce the idea that the expression of private states is distinct from the expression of speaker intention.

7. Related Work

In the past few years, there has been some work on recognizing subjective content in multiparty conversations, including the annotation of meeting data. However, none of annotation schemes capture the level of detail provided by our annotation scheme. Wrede and Shriberg (2003) have worked on recognizing meeting hotspots, which are places in meetings in which the participants are highly involved in the discussion. For that work, they developed a scheme for annotating spurts in terms of speaker involvement. Hillard et al. (2003) developed an annotation scheme for marking agreements and disagreements in multiparty conversation. Although agreements and disagreements are an important type of subjective content, they are far from the only type of subjective content that we will need to identify. Somasundaran et al. (2007) developed a scheme for annotating sentiment and arguing expressions in meetings. Although their scheme lacks some of the types of information that we include, such as the sources and targets of opinions, we feel that our two annotation schemes are complementary. Their annotations capture subjective expressions, which is a level of detail that we do not aim for in our an-

notations. Laskowski and Burger (2006) propose an annotation scheme for marking *emotionally relevant behavior* in meetings. Although their annotation scheme is fairly rich in the types of subjective content captured, as with the other schemes above, it lacks information about sources and targets. Reidsma et al. (2006) annotate emotions and mental states in meetings; however, their annotations are not geared toward extracting subjective content.

8. Conclusions

In this paper we presented an annotation scheme for marking subjective content in meetings. The annotation scheme identifies positive and negative subjectivity and different types of subjective questions, as well as uncertainty and positive and negative objective polar statements. The scheme also identifies the sources and targets of the annotations that are marked, distinguishing between subjectivity attributed to the speaker of an utterance and subjectivity attributed to someone else that the speaker is only reporting. Through an inter-annotator agreement study, we showed that the key concepts in the annotation scheme can be reliably annotated.

We also investigated the interaction between the new subjectivity annotations and dialogue act tags, which capture speaker intention. Although there is some overlap between the two types of information, the results of this study show that the expression of subjectivity is not merely a subtype of speaker intention, but rather that the two concepts are distinct. There is much more subjectivity being expressed in multiparty conversation than is captured by subjective dialogue act tags. Furthermore, utterances that we expect to be subjective based on their dialogue act tag, e.g., suggestions and assessments, in actually are not always expressing subjectivity.

9. References

J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, I. McCowan, W. Post, D. Reidsma, and P. Wellner. 2005a. The AMI meeting corpus. In *Proceedings of the Measuring Behavior Symposium on "Annotating and Measuring Meeting Behavior"*.

J. Carletta, S. Evert, U. Heid, and J. Kilgour. 2005b. The NITE XML Toolkit: Data Model and Query Lan-

guage. *Language Resources and Evaluation Journal*, 39(4):313–334.

J. Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.

Dustin Hillard, Mari Ostendorf, and Elizabeth Shriberg. 2003. Detection of agreement vs. disagreement in meetings: Training with unlabeled data. In *Proceedings of HLT-NAACL 2003*.

Pei-Yun Hsueh and Johanna Moore. 2006. Automatic topic segmentation and lablelling in multiparty dialogue. In *Proceedings of IEEE/ACM Workshop on Spoken Language Technology*.

Pei-Yun Hsueh and Johanna Moore. 2007. What decisions have you made: Automatic decision detection in conversational speech. In *Proceedings of NAACL-HLT 2007*.

K. Laskowski and S. Burger. 2006. Annotation and analysis of emotionally relevant behavior in the ISL meeting corpus. In *Proceedings of LREC 2006*.

M. Purver, P. Ehlen, and J. Niekrasz. 2006. Detecting action items in multi-party meetings: Annotation and initial experiments. In *Proceedings of MLMI 2006*.

Randolph Quirk, Sidney Greenbaum, Geoffry Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language.* Longman, New York.

Dennis Reidsma, Dirk Heylen, and Roeland Ordelman. 2006. Annotating emotion in meetings. In *Proceedings of LREC 2006*.

Swapna Somasundaran, Josef Ruppenhofer, and Janyce Wiebe. 2007. Detecting arguing and sentiment in meetings. In *Proceedings of SIGdial 2007*.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation (formerly Computers and the Humanities)*, 39(2/3):164–210.

Janyce Wiebe. 1994. Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287.

T. Wilson. 2008. *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*. Ph.D. thesis, University of Pittsburgh.

Britta Wrede and Elizabeth Shriberg. 2003. Spotting "hot spots" in meetings: Human judgments and prosodic cues. In *Proceedings of EUROSPEECH 2003*.