Discourse Level Opinion Relations: An Annotation Study

Swapna Somasundaran
Dept. of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260
swapna@cs.pitt.edu

Josef Ruppenhofer
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, PA 15260
josefr@cs.pitt.edu

Janyce Wiebe
Dept. of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260
wiebe@cs.pitt.edu

Abstract

This work proposes *opinion frames* as a representation of discourse-level associations that arise from related opinion targets and which are common in task-oriented meeting dialogs. We define the opinion frames and explain their interpretation. Additionally we present an annotation scheme that realizes the opinion frames and via human annotation studies, we show that these can be reliably identified.

1 Introduction

There has been a great deal of research in recent years on opinions and subjectivity. Opinions have been investigated at the phrase, sentence, and document levels. However, little work has been carried out at the level of discourse.

Consider the following excerpt from a dialog about designing a remote control for a television (the opinion *targets* - what the opinions are about - are shown in *italics*).

(1) D:: And I thought **not too** *edgy and like a box*, **more** kind of *hand-held* **not** as *computery*, yeah, **more** *organic shape* I think. *Simple designs*, like the last one we just saw, **not too** *many buttons* . . .

Speaker *D* expresses an opinion in favor of a design that is simple and organic in shape, and against an alternative design which is not. Several individual opinions are expressed in this passage. The first is a negative opinion about the design being too edgy and box-like, the next is a positive opinion toward a hand-held design, followed by a negative opinion toward a computery shape, and so on. While we believe that recognizing individual expressions of opinions, their properties, and components is important, we believe that discourse interpretation is needed as well. It is by understanding the passage as a discourse that we see *edgy*, *like a box*, *computery*, and *many buttons* as descriptions of the type of design *D* does not prefer, and *hand-held*, *organic shape*, and *simple designs* as descriptions of the type he does. These descriptions are not in general synonyms/antonyms of one another; for example, there are hand-held "computery" devices and simple designs that are edgy. The unison/opposition among the descriptions is due to how they are used in the discourse.

This paper focuses on such relations between the targets of opinions in discourse. Specifically, we propose *opinion frames*, which consist of two opinions which are related by virtue of having united or opposed targets. We believe that recognizing opinion frames will provide more information for NLP applications than recognizing their individual components alone. Further, if there is uncertainty about any one of the components, we believe opinion frames are an effective representation incorporating discourse information to make an overall coherent interpretation (Hobbs, 1979; Hobbs, 1983).

To our knowledge, this is the first work to extend a manual annotation scheme to relate opinions in the discourse. In this paper, we present opinion frames, and motivate their usefulness through examples. Then we provide an annotation scheme for capturing these opinion frames. Finally we perform fine-grained annotation studies to measure the human reliability in recognizing of these opinion frames.

Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue, pages 129–137,
Columbus, June 2008.

Opinion frames are presented in Section 2, our annotation scheme is described in Section 3, the inter-annotator agreement studies are presented in Section 4, related work is discussed in Section 5, and conclusions are in Section 6.

2 Opinion Frames

2.1 Introduction

The components of opinion frames are individual opinions and the relationships between their targets.

We address two types of opinions, *sentiment* and *arguing*. Following (Wilson and Wiebe, 2005; Somasundaran et al., 2007), sentiment includes positive and negative evaluations, emotions, and judgments, while arguing includes arguing *for* or *against* something, and arguing *that* something should or should not be done. In our examples, the lexical anchors revealing the opinion type (as the words are interpreted in context) are indicated in **bold face**. In addition, the text span capturing the target of the opinion (again, as interpreted in context) is indicated in *italics*.

(2) D:: ... this kind of rubbery material, *it's* a **bit more bouncy**, like you said they get chucked around a lot. A **bit more durable** and *that* can also be **ergonomic** and *it* kind of feels **a bit different from all the other remote controls**.

Speaker *D* expresses his preference for the rubbery material for the remote. He reiterates his opinion with a number of positive evaluations like **bit more bouncy**, **bit more durable**, **ergonomic** and so on.

All opinions in this example are related to the others via opinion frames by virtue of having the same targets, i.e., the opinions are essentially about the same things (the rubbery material for the remote). For example, the opinions **ergonomic** and **a bit different from all the other remote controls** are related in a frame of type *SPSPsame*, meaning the first opinion is a *S(entiment)* with polarity *P(ositive)*; the second also is a *S(entiment)* with polarity *P(ositive)*; and the targets of the opinions are in a same (target) relation.

The specific target relations addressed in this paper are the relations of either being the same or being alternatives to one another. While these are not the only possible relations, they are not infrequent, and they commonly occur in task-oriented dialogs such as those in our data.

SPSPsame, SNSNsame, APAPsame, ANANsame, SPAPsame, APSPsame, SNANsame, ANSNsame, SPSNalt, SNSPalt, APANalt, ANAPalt, SPANalt, SNAPalt, APSNalt, ANSPalt
SPSNsame, SNSPsame, APANsame, ANAPsame, SPANsame, APSNsame, SNAPsame, ANSPsame, SPSPalt, SNSNalt, APAPalt, ANANalt, SPAPalt, SNANalt, APSPalt, ANSNalt

Table 1: Opinion Frames

With four opinion type - polarity pairs (*SN*, *SP*, *AN*, *AP*), for each of two opinion slots, and two possible target relations, we have 4 * 4 * 2 = 32 types of frame, listed in Table 1.

In the remainder of this section, we elaborate further on the *same* target relation (in 2.2) the *alternative* target relation (in 2.3) and explain a method by which these relationships can be propagated (in 2.4). Finally, we illustrate the usefulness of opinion frames in discourse interpretation (in 2.5).

2.2 *Same* Targets

Our notion of sameness for targets includes cases of anaphora and ellipses, lexically similar items, as well as less direct relations such as part-whole, subset, inferable, and instance-class.

Looking at the opinion frames for Example 2 in more detail, we separately list the opinions, followed by the relations between targets.

Opinion Span - target Span	Type
O1 **bit more bouncy** - *it's* [t1]	SP
O2 **bit more durable** - ellipsis [t2]	SP
O3 **ergonomic** - *that* [t3]	SP
O4 **a bit different from all the other remote** - *it* [t4]	SP

Target - target	Rel
t1 - t2	same
t1 - t3	same
t3 - t4	same

Ellipsis occurs with **bit more durable**. [t2] represents the (implicit) target of that opinion, and [t2] has a *same* relation to [t1], the target of the **bit more bouncy** opinion. (Note that the interpretation of the first target, [t1], would require anaphora resolution of its target span with a previous noun phrase, *rubbery material.*)

Let us now consider the following passage, in which a meeting participant analyzes two leading remotes on the market.[1]

(3) D:: These are two **leading** *remote controls* at the moment. You know *they*'re **grey**, *this one*'s got **loads of buttons**, it's **hard to tell** from here what *they* actually do, and *they* **don't look very exciting at all**.

Opinion Span - target Span	Rel
O1 **leading** - *remote controls* [t1]	SP
O2 **grey** - *they* [t2]	SN
O3 **loads of buttons** - *this one* [t3]	SN
O4 **hard to tell** - *they* [t4]	SN
O5 **don't look very exciting at all** - *they* [t5]	SN

Target - target	Rel
t1 - t2	same
t2 - t3	same
t3 - t4	same
t5 - t1	same

Target [t2] is the set of two leading remotes, and [t3], which is in a *same* relation with [t2], is one of those remotes. Target [t4], which is also in a *same* relation with [t3], is an aspect of that remote, namely its buttons. Thus, opinion O3 is directly about one of the remotes, and indirectly about the set of both remotes. Similarly, opinion O4 is directly about the buttons of one of the remotes, and indirectly about that remote itself.

2.3 *Alternative* Targets

The *alt(ernative)* target relation arises when multiple choices are available, and only one can be selected. For example, in the domain of TV remote controls, the set of all shapes are alternatives to one another, since a remote control may have only one shape at a time. In such scenarios, a positive opinion regarding one choice may imply a negative opinion toward the rest of the choices, and vice versa.

As an example, let us now consider the following passage (some intervening utterances have been removed for clarity).

(4) C:: ... shapes **should be** *curved*, so round shapes[2] **Nothing** *square-like*.
C:: ... So we **shouldn't have too** *square corners* and that kind of thing.
B:: Yeah okay. **Not** ***the old box look***.

[1]In the other examples in this paper, the source (holder) of the opinions is the speaker. The **leading** opinion in this example is an exception: its source is implicit; it is a consensus opinion that is not necessarily shared by the speaker (i.e., it is a *nested source* (Wiebe et al., 2005)).

[2]In the context of the dialogs, the annotators read the "so round shapes" as a summary statement. Had the "so" been interpreted as Arguing, the round shapes would have been annotated as a target (and linked to *curved*).

Opinion Span - target Span	Rel
O1 **should be** - *curved* [t1]	AP
O2 **Nothing** - *square-like* [t2]	AN
O3 **shouldn't have** - *square corners* [t3]	AN
O4 **too** - *square corners* [t3]	SN
O5 **Not** - *the old box look* [t4]	AN
O6 **the old box look** - *the old box look* [t4]	SN

Target - target	Rel
t1 -t2	alternatives
t2 - t3	same
t3 - t4	same

There is an *alt* relation between, for example, [t1] and [t2]. Thus, we have an opinion frame between *O1* and *O2*, whose type is *APANalt*. From this frame, we understand that a positive opinion is expressed toward something and a negative opinion is expressed toward its alternative.

2.4 Link Transitivity

When individual targets are linked, they form a chain-like structure. Due to this, a connecting path may exist between targets that were not directly linked by the human annotators. This path may be traversed to create links between new pairs of targets - which in turn results in new opinion frame relations. For instance, in Example 4, the frame with direct relation is *O1O2 APANalt*. By following the *alt* link from [t1] to [t2] and the *same* link from [t2] to [t3], we have an *alt* link between [t1] and [t3], and the additional frames *O1O3 APANalt* and *O1O4 APSNalt*. Repeating this process would finally link speaker *C's* opinion O1 with *B's* opinion O6, yielding a *APSNalt* frame.

2.5 Interpretation

This section illustrates two motivations for opinion frames: they may unearth additional information over and above the individual opinions stated in the text, and they may contribute toward arriving at a coherent interpretation (Hobbs, 1979; Hobbs, 1983) of the opinions in the discourse.

Through opinion frames, opinions regarding something not explicitly mentioned in the local context and not even lexically related can become relevant, providing more information about someone's opinions. This is particularly interesting when *alt* relations are involved, as opinions towards one alternative imply opinions of opposite polarity toward the remaining options. For instance in Example 4 above, if we consider only the explicitly stated opinions, there is only one (positive) opinion about the curved shape, namely *O1*. However, the speaker expresses several other opinions which reinforce his positivity toward the curved shape. These are in fact opinion frames in which the other opinion has the opposite polarity as *O1* and the target relation is *alt* (for example frames such as *O1O3 APANalt* and *O1O4 APSNalt*).

In the dialog, notice that speaker *B* agrees with *C* and exhibits his own reinforcing opinions. These would be similarly linked via targets resulting in frames like *O1O6 APSNalt*.

Turning to our second point, arriving at a coherent interpretation obviously involves disambiguation. Suppose that some aspect of an individual opinion, such as polarity, is unclear. If the discourse suggests certain opinion frames, this may in turn resolve the underlying ambiguity. For instance in Example 2, we see that out of context, the polarities of **bouncy** and **different from other remotes** are unclear (bounciness and being different may be negative attributes for another type of object). However, the polarities of two of the opinions are clear (**durable** and **ergonomic**). There is evidence in this passage of discourse continuity and *same* relations, such as the pronouns, the lack of contrastive cue phrases, and so on. This evidence suggests that the speaker expresses similar opinions throughout the passage, making the opinion frame *SPSPsame* more likely throughout. Recognizing the frames would resolve the polarity ambiguities of **bouncy** and **different**.

Example 2 is characterized by opinion frames in which the opinions reinforce one other. Interestingly, interplays among different opinion types may show the same type of reinforcement. As we analyzed above, Example 4 is characterized by mixtures of opinion types, polarities, and target relations. However, the opinions are still unified in the intention to argue for a particular type of shape. There is evidence in this passage suggesting reinforcing frames: the negations are applied to targets that are alternative to the desired option, and the passage is without contrastive discourse cues. If we are able to recognize the best overall set of opinion frames for the passage, the polarity ambiguities will be resolved.

On the other hand, evidence for non-reinforcing opinions would suggest other frames, potentially resulting in different interpretations of polarity and relations among targets. Such non-reinforcing associations between opinions and often occur when the speaker is ambivalent or weighing pros and cons. Table 1 lists the frames that occur in reinforcing scenarios in the top row, and the frames that occur in non-reinforcing scenarios in the bottom row.

3 Annotation Scheme

Our annotation scheme began with the definition and basics of the opinion annotation from previous work (Wilson and Wiebe, 2005; Somasundaran et al., 2007). We then add to it the attributes and components that are necessary to make an Opinion Frame.

First, the text span that reveals the opinion expression is identified. Then, the text spans corresponding to the targets are marked, if there exist any (we also allow span-less targets). Then, the type and polarity of the opinion in the context of the discourse is marked. Finally the targets that are related (again in the context of the discourse) are linked. Specifically, the components that form the Annotation of the frame are as follows:

Opinion Span: This is a span of text that reveals the opinion.

Type: This attribute specifies the opinion type as either *Arguing* or *Sentiment*.

Polarity: This attribute identifies the valence of an opinion and can be one of: *positive, negative, neutral, both, unknown.*

Target Span: This is a span of text that captures what an opinion is about. This can be a proposition or an entity.

Target Link: This is an attribute of a target and records all the targets in the discourse that the target is related to.

Link Type: The link between two targets is specified by this attribute as either *same* or *alternative*.

In addition to these definitions, our annotation manual has guidelines detailing how to deal with grammatical issues, disfluencies, etc. Appendix A illustrates how this annotation scheme is applied to the utterances of Example 4.

Links between targets can be followed in either direction to construct chains. In this work, we consider target relations to be commutative, i.e., Link(t1,t2) => Link(t2,t1). When a newly annotated target is similar (or opposed) to a set of targets already participating in *same* relations, then the *same* (or *alt*) link is made only to one of them - the one that looks most natural. This is often the one that is closest.

4 Annotation Studies

Construction of an opinion frame is a stepwise process where first the text spans revealing the opinions and their targets are selected, the opinion text spans are classified by type and polarity and finally the targets are linked via one of the possible relations. We split our annotation process into these 3 intuitive stages and use an evaluation that is most applicable for the task at that stage.

Two annotators (both co-authors on the paper) underwent training at each stage, and the annotation manual was revised after each round of training. In order to prevent errors incurred at earlier stages from affecting the evaluation of later stages, the annotators produced a consensus version at the end of each stage, and used that consensus annotation as the starting point for the next annotation stage. In producing these consensus files, one annotator first annotated a document, and the other annotator reviewed the annotations, making changes if needed. This prevented any discussion between the annotators from influencing the tagging task of the next stage.

In the following subsections, we first introduce the data and then present our results for annotation studies for each stage, ending with discussion.

4.1 Data

The data used in this work is the AMI meeting corpus (Carletta et al., 2005) which contains multimodal recordings of group meetings. We annotated meetings from the scenario based meetings, where four participants collaborate to design a new TV remote control in a series of four meetings. The meetings represent different project phases, namely project kick-off, functional design, conceptual design, and detailed design. Each meeting has rich transcription and segment (turn/utterance) information for each speaker. Each utterance consists of one or more sentences. At each agreement stage we used approximately 250 utterances from a meeting for evaluation. The annotators also used the audio and video recordings in the annotation of meetings.

Gold	Exact	Lenient	Subset
ANN-1	53	89	87
ANN-2	44	76	74

Table 2: Inter-Annotator agreement on Opinion Spans

4.2 Opinion Spans and Target Spans

In this step, the annotators selected text spans and labeled them as *opinion* or *target* We calculated our agreement for text span retrieval similar to Wiebe et al. (2005). This agreement metric corresponds to the Precision metric in information retrieval, where annotations from one annotator are considered the gold standard, and the other annotator's annotations are evaluated against it.

Table 2 shows the inter-annotator agreement (in percentages). For the first row, the annotations produced by Annotator-1 (ANN-1) are taken as the gold standard and, for the second row, the annotations from annotator-2 form the gold standard. The "Exact" column reports the agreement when two text spans have to match exactly to be considered correct. The "Lenient" column shows the results if an overlap relation between the two annotators' retrieved spans is also considered to be a hit. Wiebe et al. (2005) use this approach to measure agreement for a (somewhat) similar task of subjectivity span retrieval in the news corpus. Our agreement numbers for this column is comparable to theirs. Finally, the third column, "Subset", shows the agreement for a more strict constraint, namely, that one of the spans must be a subset of the other to be considered a match. Two opinion spans that satisfy this relation are ensured to share all the opinion words of the smaller span.

The numbers indicate that, while the annotators

Gold	Exact	Lenient	Subset
ANN-1	54	73	71
ANN-2	54	75	74

Table 3: Inter-Annotator agreement on Target Spans

Gold	Exact	Lenient	Subset
ANN-1	74	87	87
ANN-2	76	90	90

Table 4: Inter-Annotator agreement on Targets with Perfect Opinion spans

do not often retrieve the exact same span, they reliably retrieve approximate spans. Interestingly, the agreement numbers between Lenient and Subset columns are close. This implies that, in the cases of inexact matches, the spans retrieved by the two annotators are still close. They agree on the opinion words and differ mostly on the inclusion of function words (e.g. articles) and observation of syntactic boundaries.

In similar fashion, Table 3 gives the inter-annotator agreement for target span retrieval. Additionally, Table 4 shows the inter-annotator agreement for target span retrieval when opinions that do not have an exact match are filtered out. That is, Table 4 shows results only for targets of the opinions on which the annotators perfectly agree. As targets are annotated with respect to the opinions, this second evaluation removes any effects of disagreements in the opinion detection task. As seen in Table 4, this improves the inter-coder agreement.

4.3 Opinion Type and Polarity

In this step, the annotators began with the consensus opinion span and target span annotations. We hypothesized that given the opinion expression, determining whether it is Arguing or Sentiment would not be difficult. Similarly, we hypothesized that target information would make the polarity labeling task clearer.

As every opinion instance is tagged with a type and polarity, we use Accuracy and Cohen's Kappa (κ) metric (Cohen, 1960). The κ metric measures the inter-annotator agreement above chance agreement. The results, in Table 5, show that κ both for type and polarity tagging is very high. This confirms our hypothesis that Sentiment and Arguing can be reliably distinguished once the opinion spans are known. Our polarity detection task shows an improvement in κ over a similar polarity assignment task by Wilson et al. (2005) for the news corpus (κ of 0.72). We believe this improvement can partly be attributed to the target information available to our annotators.

	Type Tagging	Polarity Tagging
Accuracy	97.8%	98.5%
κ	0.95	0.952

Table 5: Inter-Annotator agreement on Opinion Types and Polarity

4.4 Target Linking

As an intuitive first step in evaluating target linking, we treat target links in the discourse similarly to anaphoric chains and apply methods developed for co-reference resolution (Passonneau, 2004) for our evaluation. Passonneau's method is based on Krippendorf's α metric (Krippendorff, 2004) and allows for partial matches between anaphoric chains. In addition to this, we evaluate links identified by both annotators for the type (*same* / *alternative*) labeling task with the help of the κ metric.

Passonneau (2004) reports that in her co-reference task on spoken monologs, α varies with the difficulty of the corpus (from 0.46 to 0.74). This is true in our case too. Table 6 shows our agreement for the four types of meetings in the AMI corpus: the kickoff meeting (a), the functional design (b), the conceptual design (c) and the detailed design (d).

Of the meetings, the kickoff meeting (a) we use has relatively clear discussions. The conceptual design meeting (c) is the toughest, as as participants are expressing opinions about a hypothetical (desirable) remote. In our detailed design meeting (d), there are two final designs being evaluated. On analyzing the chains from the two annotators, we discovered that one annotator had maintained two separate chains for the two remotes as there is no explicit linguistic indication (within the 250 utterances) that these two are alternatives. The second annotator, on the other hand, used the knowledge that the goal of the meeting is to design a single TV remote to link them as alternatives. Thus by changing just two links in the second annotator's file to account for this, our α for this meeting went up from 0.52

Meeting:	a	b	c	d
Target linking (α)	0.79	0.74	0.59	0.52
Relation Labeling (κ)	1	1	0.91	1

Table 6: Inter-Annotator agreement on Target relation identification

to 0.70. We plan to further explore other evaluation methodologies that account for severity of differences in linking and are more relevant for our task. Nonetheless, the resulting numbers indicate that there is sufficient information in the discourse to provide for reliable linking of targets.

The high κ for the relation type identification shows that once the presence of a link is detected, it is not difficult to determine if the targets are similar or alternatives to each other.

4.5 Discussion

Our agreement studies help to identify the aspects of opinion frames that are straightforward, and those that need complex reasoning. Our results indicate that while the labeling tasks such as opinion type, opinion polarity and target relation type are relatively reliable for humans, retrieval of opinions spans, target spans and target links is more difficult.

A common cause of annotation disagreement is different interpretation of the utterance, particularly in the presence of disfluencies and restarts. For example consider the following utterance where a participant is evaluating the drawing of another participant on the white board.

(5) *It's a baby shark* , **it looks to me**, . . .

One annotator interpreted this "it looks to me" as an arguing for the belief that it was indeed a drawing of a baby shark *(positive Arguing)*. The second annotator on the other hand looked at it as a *neutral* viewpoint/evaluation (*Sentiment*) being expressed regarding the drawing. Thus even though both annotators felt an opinion is being expressed, they differed on its type and polarity.

There are some opinions that are inherently on the borderline of Sentiment and Arguing. For example, consider the following utterance where there is an appeal to importance:

(6) **Also important** for you all is um the *production cost must be maximal twelve Euro and fifty cents*.

Here, "also important" might be taken as an assessment of the high value of adhering to the budget (relative to other constraints), or simply as an argument for adhering to the budget.

One potential source of problems to the target-linking process consists of cases where the same item becomes involved in more than one opposition. For instance, in the example below, speaker *D* initially sets up an alternative between speech recognition and buttons as a possible interface for navigation. But later, speaker *A* re-frames the choice as between having speech recognition only and having both options. Connecting up all references to speech recognition as a target respects the co-reference but it also results in incorrect conclusions: the speech recognition is an alternative to having both speech recognition and buttons.

(7) A:: One thing is **interesting** is talking about *speech recognition* in a remote control...
D:: ... So that we don't need any button on the remote control it would be all based on speech.
A:: ... I think *that* **would not work so well**. **You wanna** have *both options*.

5 Related Work

Evidence from the surrounding context has been used previously to determine if the current sentence should be subjective/objective (Riloff et al., 2003; Pang and Lee, 2004)) and adjacency pair information has been used to predict congressional votes (Thomas et al., 2006). However, these methods do not explicitly model the relations between opinions. Additionally, in our scheme opinions that are not in the immediate context may be allowed to influence the interpretation of a given opinion via target chains.

Polanyi and Zaenen (2006), in their discussion on contextual valence shifters, have also observed the phenomena described in this work - namely that a central topic may be divided into subtopics in order to perform evaluations, and that discourse structure can influence the overall interpretation of valence.

Snyder and Barzilay (2007) combine an agreement model based on contrastive RST relations with a local *aspect* (or target) model to make a more informed overall decision for sentiment classification. The contrastive cue indicates a change in the sentiment polarity. In our scheme, their aspects would be related as *same* and their high contrast relations would result in frames such as *SPSNsame*, *SNSP-same*. Additionally, our frame relations would link sentiments across non-adjacent clauses, and make connections via *alt* target relations.

Considering the discourse relation annotations in the PDTB (Prasad et al., 2006), there can be alignment between discourse relations (like contrast) and our opinion frames when the frames represent dominant relations between two clauses. However, when the relation between opinions is not the most prominent one between two clauses, the discourse relation may not align with the opinion frames. And when an opinion frame is between two opinions in the same clause, there would be no discourse relation counterpart at all. Further, opinion frames assume particular intentions that are not necessary for the establishment of ostensibly similar discourse relations. For example, we may not impose an opinion frame even if there are contrastive cues. (Please refer to Appendix B for examples)

With regard to meetings, the most closely related work includes the dialog-related annotation schemes for various available corpora of conversation (Dhillon et al. (2003) for ICSI MRDA; Carletta et al. (2005) for AMI) As shown by Somasundaran et al. (2007), dialog structure information and opinions are in fact complementary. We believe that, like discourse relations, dialog information will additionally help in arriving at an overall coherent interpretation.

6 Conclusion and Future work

This is the first work that extends an opinion annotation scheme to relate opinions via target relations. We first introduced the idea of opinion frames as a representation capturing discourse level relations that arise from related opinion targets and which are common in task-oriented dialogs such as our data. We built an annotation scheme that would capture these relationships. Finally, we performed extensive inter-annotator agreement studies in order to find the reliability of human judgment in recognizing frame components. Our results and analysis provide insights into the complexities involved in recognizing discourse level relations between opinions.

Acknowledgments

This research was supported in part by the Department of Homeland Security under grant N000140710152.

References

J. Carletta, S. Ashby, and et al. 2005. The AMI Meetings Corpus. In *Proceedings of Measuring Behavior Symposium on "Annotating and measuring Meeting Behavior"*.

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

R. Dhillon, S. Bhagat, H. Carvey, and E. Shriberg. 2003. Meeting recorder project: Dialog act labeling guide. Technical report, ICSI Tech Report TR-04-002.

J. Hobbs. 1979. Coherence and coreference. *Cognitive Science*, 3:67–90.

J. Hobbs, 1983. *Why is Discourse Coherent?*, pages 29–70. Buske Verlag.

K. Krippendorff. 2004. *Content Analysis: An Introduction to Its Methodology, 2nd Edition*. Sage Publications, Thousand Oaks, California.

B. Pang and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *ACl 2004*.

R. J. Passonneau. 2004. Computing reliability for coreference annotation. In *LREC*.

L. Polanyi and A. Zaenen, 2006. *Contextual Valence Shifters*, chapter 1. Computing Attitude and Affect in Text: Theory and Applications. Springer.

R. Prasad, N. Dinesh, A. Lee, A. Joshi, and B. Webber. 2006. Annotating attribution in the Penn Discourse TreeBank. In *Workshop on Sentiment and Subjectivity in Text*. ACL.

E. Riloff, J. Wiebe, and T. Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *CoNLL 2003*.

B. Snyder and R. Barzilay. 2007. Multiple aspect ranking using the good grief algorithm. In *HLT 2007: NAACL*.

S. Somasundaran, J. Ruppenhofer, and J. Wiebe. 2007. Detecting arguing and sentiment in meetings. In *SIGdial Workshop on Discourse and Dialogue 2007*.

M. Thomas, B. Pang, and L. Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *EMNLP 2006*.

J. Wiebe, T. Wilson, and C Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, pages 164–210.

T. Wilson and J. Wiebe. 2005. Annotating attributions and private states. In *Proceedings of ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*.

T. Wilson, J. Wiebe, and P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT-EMNLP 2005*.

A Annotation Example

C:: . . . shapes **should be** *curved*, so round shapes. **Nothing** *square-like*.
C:: . . . So we **shouldn't have too** *square corners* and that kind of thing.
B:: Yeah okay. **Not** ***the old box look***.

Span	Attributes
O1 **should be**	type=Arguing; Polarity=pos; target=t1
t1 *curved*	Link,type=(t2,alt)
O2 **Nothing**	type=Arguing; Polarity=neg; target=t2
t2 *square-like*	Link,type=(t1,alt),(t3,same)
O3 **shouldn't have**	type=Arguing; Polarity=neg; target=t3
O4 **too**	type=Sentiment; Polarity=neg; target=t3
t3 *square corners*	Link,type=(t2,same),(t4,same)
O5 **Not**	type=Arguing; Polarity=neg; target=t4
t4 *the old box look*	Link,type=(t3,same)
O6 **the old box look**	type=Sentiment; Polarity=neg; target=t4

B Comparison between Opinion Frames and Discourse Relations

Opinion frames can align with discourse relations between clauses only when the frames represent the dominant relation between two clauses (1); but not when the opinions occur in the same clause (2); or when the relation between opinions is not the most prominent (3); or when two distinct targets are neither same nor alternatives (4).

(1) **Non-reinforcing opinion frame (SNSPsame); Contrast discourse relation**
D :: And so what I have found and after a lot of work actually I draw for you this *schema* that can be maybe **too technical** for you but is **very important** for me you know.

(2) **Reinforcing opinion frame (SPSPsame); no discourse relation**
Thirty four percent said it takes too long to learn to use a remote control, they **want** something that's *easier to use* straight away, *more intuitive* perhaps.

(3) **Reinforcing opinion frame (SPSPsame); Reason discourse relation**
She even likes my manga, actually the quote is: "I **like** *it*, because you **like** *it*, honey." (source: web)

(4) **Unrelated opinions; Contrast discourse relation**
A :: Yeah, what I have to say about means. *The smart board* is **okay**. *Digital pen* is **horrible**. I dunno if you use it. But if you want to download it to your computer, it's doesn't work. No.