MPQA Opinion Corpus

Theresa Wilson, Janyce Wiebe, and Claire Cardie

Abstract The MPQA Opinion Corpus is a collection of documents with expression-level, multi-attribute annotations of opinions, sentiments, and other private states. This chapter describes the MPQA annotation scheme and the development of the MPQA Corpus.

1 Introduction

Opinion mining and sentiment analysis has become one of the most active areas of NLP, with applications to a wide range of problems in political science, sociology, economics, and the humanities, as well as in many day-to-day affective computing settings [30] in healthcare, finance, public relations, and social media applications. The *MPQA Opinion Corpus* was one of the first corpora with detailed, expression-level opinion and sentiment annotations made available to the research community. This chapter describes the MPQA annotation scheme and the development the MPQA Corpus, which has served as training and evaluation evaluation data in many NLP projects since its release.

The MPQA annotation scheme is built on the fundamental concepts of private state and linguistic subjectivity. Quirk et al. [31] define *private state* as an internal mental or emotional state that is not open to objective verification. Thus, private states encompass not only sentiments, evaluations, opinions and emotions, but also (dis)belief, (un)certainty, speculation, (dis)agreement, and other inner states. In our

Theresa Wilson
Oberlin College, Oberlin, OH 44074, U.S.A. e-mail: taw@cs.oberlin.edu

Janyce Wiebe
University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, PA 15260, U.S.A. e-mail: wiebe@cs.pitt.edu

Claire Cardie
Cornell University, Ithaca, NY 14850, U.S.A. e-mail: cardie@cs.cornell.edu

work, *subjectivity* is defined as the expression of private states in language,[1] and a private state is defined as an *attitude* held by an *experiencer* (more specifically, a *source*) toward an optional *target* [42, 43]. Up to the early 2000s, we had annotated subjectivity only at the sentence level (see [46] for work carried out using this sentence-level corpus). Our goal in developing the MPQA annotation scheme was to delve further into subjectivity, and provide fine-grained annotations of private states and their components.

The motivation for this work was the need to provide tools for information analysts in government, commercial, and political domains, who want to automatically track attitudes and feelings in the news and on-line forums; such tools require analysis at a fine-grained level. The first version of the corpus was collected and annotated as part of the summer 2002 NRRC Workshop on Multi-Perspective Question Answering (MPQA) [45] (hence the name of the corpus).

In this case study, we present the current version of the MPQA annotation scheme. We start by describing the frame-based conceptualization and then review the steps involved in moving from the conceptualization to a fully annotated corpus. We briefly discuss some key challenges we faced in the developnment process. We end the chapter with an overview of inter-coder reliability studies and a short review of related work.

2 Expressing Private States: A Primer

In this work we focus on four main ways that private states are expressed: direct references to private states, private states expressed in speech events,[2] private states expressed indirectly using expressive subjective language, and private states expressed through actions. The sentences below give examples of each of these.

Example 1. direct reference to a private state
Democrats also have doubts about Miers' suitability for the high court.

Example 2. private state expressed in a speech event
Miers' nomination was criticized from people all over the political spectrum.

Example 3. private state expressed in a speech event using expressive subjective language
"She [Miers] will be a breath of fresh air for the Supreme Court," LaBoon said.

Example 4. private state expressed through action
As the long line of would-be voters marched in, those near the front of the queue began to spontaneously applaud those who were far behind them.

[1] This term has been borrowed and adapted from literary theory [5].

[2] We use the term speech event to refer to any event of speaking or writing.

Direct references to a private state ("have doubts" in 1) are the most straightforward way we see private states expressed in language. However, if we focused only on these direct references to private states, a huge number of private state expressions would be overlooked. We frequently find private states being conveyed in speech events. Mixture terms ("criticize" in 2) are used to indicate that a private state is expressed as part of a speech event, without needing to give the actual words. Often, though, it is in the way something is described or through a particular wording that a private state is expressed. This is the case with the speech event referred to by "said" in 3. Within the quoted speech, it is the phrase "breath of fresh air" that conveys the private state of the speaker. These indirect expressions of private states are called *expressive subjective elements* [5]. Private states may also be expressed through certain actions, such as booing, laughing, protesting, or applauding (4). References to *private state actions* [43] are common in third-person discourse, such as news and media reporting.

3 Conceptualization

The MPQA annotation scheme is conceptualized using a frame-style representation of private states and attributions. It contains six representational frames: two types of private state frames, a frame for objective speech events, and frames representing agents, attitudes, and targets. In earlier versions, attitudes and targets were represented as attributes on private state frames [47]. In [51], the conceptualization was revised, and thereafter attitudes and targets have been represented by their own frames. Figure 1 shows the most recent version of the conceptualization.

There are two attributes that we find on all or most of the annotation frames. The first of these is the *text anchor* attribute. As the name implies, text anchors point to the spans of text on which frames are anchored. Generally the anchor is the word or phrase that expresses the frame concept. The exception to this is for speech events that are *implicit*. Implicit speech events are speech events for which there is not a discourse parenthetical, such as, "she said." Every sentence in a document is an implicit speech event for the writer of the document. Direct quotations unaccompanied by discourse parentheticals are also implicit speech events. With implicit speech events there is no phrase referencing the speech event to serve as the text anchor. In these cases, the text anchor points to the sentence or quoted string that contains the text of the speech event, and the *implicit* attribute on the frame is set to true.

The second attribute that is found on all frames, with the exception of the attitude and target frames, is the *source* attribute.[3] This attribute is used to mark the experiencer of the private state or the speaker/writer of the speech event. Obviously, the writer of an article is a source, because he or she wrote the sentences that constitute the article. However, the writer may also write about other people's private states and speech events, leading to multiple sources in a single sentence. In example 1

[3] Although not included on attitude and target frames, the source of these annotations can be retrieved by following the attitude and target links back to the direct subjective frames.

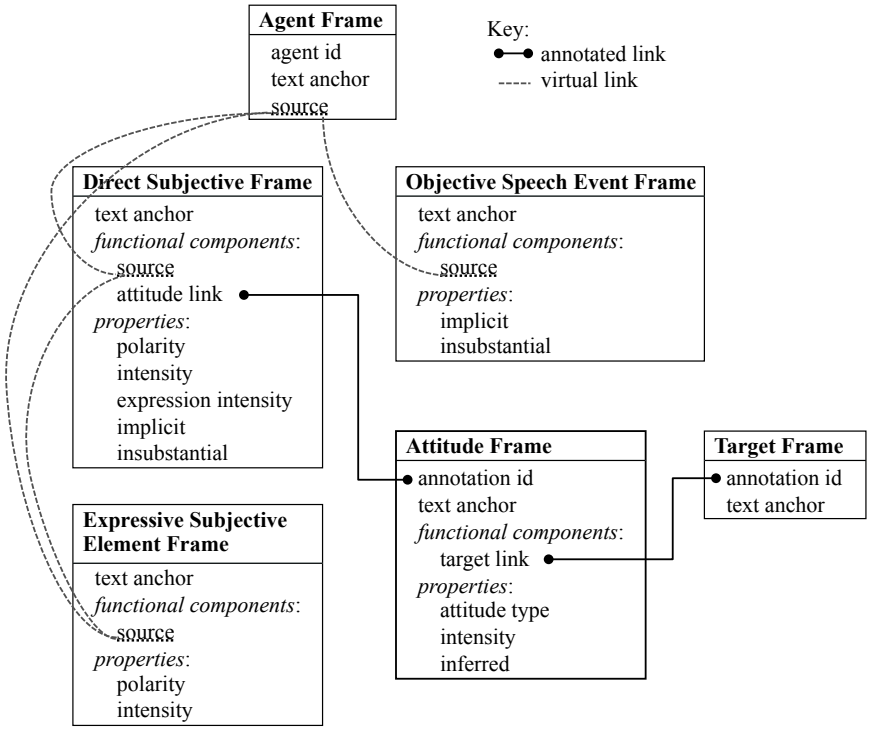

Fig. 1 Conceptual Representation

above, there are two sources: 1) the writer of the sentence, and 2) Democrats, the experiencer of the private state "have doubts."

A key aspect of sources is that they are nested to capture levels of attribution. In example 1, the Democrats do not directly state that they have doubts. Rather it is according to the writer that the Democrats have doubts about Miers' suitability for the Supreme Court. The full source of the private state expressed by "have doubts" is thus the **nested source**: ⟨*writer, Democrats*⟩. The nested source is composed of the agent IDs associated with each source.

3.1 Private State Frames

The two types of private state frames are **direct subjective frames** and **expressive subjective element frames** (ESE frames). Direct subjective frames are used for marking direct references to private states, speech events expressing private states, and private state actions. ESE frames are used for marking expressive subjective elements. Having the two types of private state frame allows us to distinguish between expressions that introduce another level of attribution and those that do not.

Direct references to private states, references to speech events (whether or not a private state is expressed), and references to private state actions typically introduce another level of attribution. That is, the private state, the speech event, or the action referenced by the expression is attributed to a different entity than the speaker/writer/experiencer of the speech event or private state in which it is scoped. The source of expressive subjective elements, on the other hand, remains the same in most cases.

Aside from the text anchor and source attributes, the two private state frames have several attributes that capture various properties of the private state and the text anchor. The *intensity* attribute is used to mark the overall intensity of the private state that is represented by the direct subjective or expressive subjective element frame. Intensity is rated on a four-point scale: *low, medium, high, extreme*. For direct subjective frames, there is an additional intensity rating: *expression intensity*. This attribute is used to mark the contribution to the overall intensity made just by the private state or speech event phrase. For example, *say* is often neutral, even if what is uttered is not neutral. The word *excoriate*, on the other hand, by itself implies a very strong private state. Values for expression intensity range from *neutral* to *extreme*.

Another attribute of both types of private state frames is *polarity*. The polarity attribute is used to indicate whether the private state or speech event phrase is expressing a sentiment, and if so, whether the sentiment is *positive, negative,* or *both* positive and negative.

In addition to the implicit and expression-intensity attributes, the direct subjective frame has two more attributes not found on the ESE frames: *attitude link* and *insubstantial*. The attitude link attribute is a list of one or more attitude frame IDs. This attribute functions to connect direct subjective frames and attitude frames.

The *insubstantial* attribute is used to mark direct subjective frames that are not substantial in the discourse. A private state or speech event may be insubstantial either because it is not real or because it is not significant in the discourse. Private states and speech events may not be real in the discourse for several reasons; an example of one is when the private state or speech event is hypothetical. Private states or speech events that are not significant are those that do not contain a significant portion of the contents of the private state or speech event.

3.2 Objective Speech Event Frames

The **objective speech event frame** is used to mark speech events that do not express private states. They capture when material is attributed to some source, but is being presented objectively, such as with the speech event in the following example:

Example 5. objective speech event

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

The objective speech event frame contains a subset of the attributes found in the direct subjective frame.

3.3 Agent Frames

The **agent frame** is used to mark noun phrases that refer to sources of private states and speech events. For example, agent frames would be created for "Democrats" above in 1, "LaBoon" in 3, and "White House spokesman Jim Dyke" in 5.

Aside from the *text anchor* and *source* attributes, the agent frame has one additional atribute: *agent id*. An agent ID is an alpha-numeric identifer that serves to uniquely identify a particular agent within the document. It is added to the agent frame marking the first informative (e.g., non-pronomial) reference to the agent. The IDs are then used in the source attributes for the agent, direct subjective, ESE, and objective speech frames to capture the levels of attribution (i.e., nested sources).

3.4 Attitude and Target Frames

The **attitude frame** and the **target frame** provide a representation for the attitudes that compose private states and the targets of those attitudes. Each attitude frame and each target frame is assigned a unique ID. The attitude frame IDs are used to link attitudes to direct subjective frames, and the target frame IDs link targets to attitudes. Every direct subjective frame will link to one or more attitude frames. Every attitude frame will link to zero or more target frames.

Attitude frames also have attributes for representing the *attitude type*, the *intensity* of the attitude, and whether the attitude is *inferred*. Listed below is the set of attitude types marked in the corpus:

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

The inferred attribute is used for marking attitudes that are not *syntactically* the most prominent, yet their presence is more or less unambigous. Consider the private state attributed to "people" in the following sentence.

Example 6. "I think people are happy that Chavez has fallen."

There are two attitudes being expressed within the span, "happy that Chavez has fallen." Syntactically, the most prominent attitude is the positive sentiment towards the fall of Chavez. However, if one is happy about a political fall, it is a very short step to infer that the happiness is rooted in a negative sentiment toward the one falling. In such cases as these, frames are created for both attitudes, and the one that is less syntactically prominent is marked as inferred.

3.5 Example

To help illustrate the different annotation frames and how they mesh together, this section steps through the annotations created for the following sentence:

Example 7. Its aim of the 2001 report is to tarnish China's image and exert political pressure on the Chinese Government, human rights experts said at the seminar held by the China Society for Study of Human Rights (CSSHR) on Friday.

The first thing to note is that there are three levels of attribution in the sentence: a) the entire sentence attributed to the writer, b) the indirect quotation attributed to the human rights experts by the writer, and c) the intention indicated by the direct subjective expression "aim" attributed to the 2001 report by the human rights experts according to the writer. The annotation frames corresponding to each of these levels of attribution are given in Figures 2a, 2b, and 2c, respectively.

First, consider the writer. Although the sentence is *subjective*—there are private states expressed within the sentence—the writer is not the direct source of any of these private states. The part of the sentence directly attributed to the writer, that the human rights experts said something at a seminar on Friday, is objective. Therefore, we create an objective speech event frame for the writer. Because there is no actual speech expression on which to anchor the frame, the *implicit* attribute on the frame is set to true, and the frame is anchored on the sentence.

The next level of attribution (Figure 2b) is what the human rights experts say, according to the writer. Within the indirect quotation, we have the experts attributing an intention to the 2001 report. Merely attributing a private state to another entity is not sufficient evidence to conclude that the ones doing the attributing (in this case, the experts) are themselves expressing a private state. However, when we consider the full context of what the experts said, we find further evidence that indeed they are expressing a private state. Two ESE frames are marked in the sentence on "tarnish" and "exert political pressure." Both have a negative polarity. The phrase "exert political pressure" by itself is fairly mild, and in a different context might not even be considered subjective. The word, "tarnish," on the other hand, is more strongly negative and unambiguously subjective in this context. By accusing another entity of intentionally trying to harm and coerce, which is what the experts are doing, part of what is being communicating is a negative sentiment. Thus, we create a direct subjective frame anchored to "said" for the source, $< writer, experts >$. We also create an attitude frame anchored to the phrase "tarnish ... pressure" with type negative sentiment, and a target frame anchored to "the 2001 report." The target is linked to the attitude via the target id. Likewise, the attitude frame is linked to the direct subjective frame via the attitude id.

There is one more frame that is created for this level of attribution: an agent frame anchored to "human rights experts." This is the first reference to this particular set of entities. Therefore, the agent frame is given an id, which is then used to refer to this particular set of experts in the annotations within this and possibly later sentences.

The third level of attribution (Figure 2c) is the intention attributed to the 2001 report by the human rights experts, according to the writer. We do create an agent

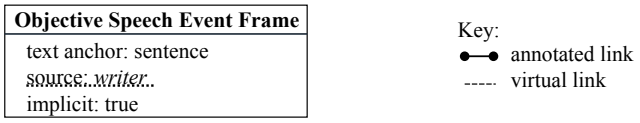

Fig. 2a Frames attributed to source <writer>

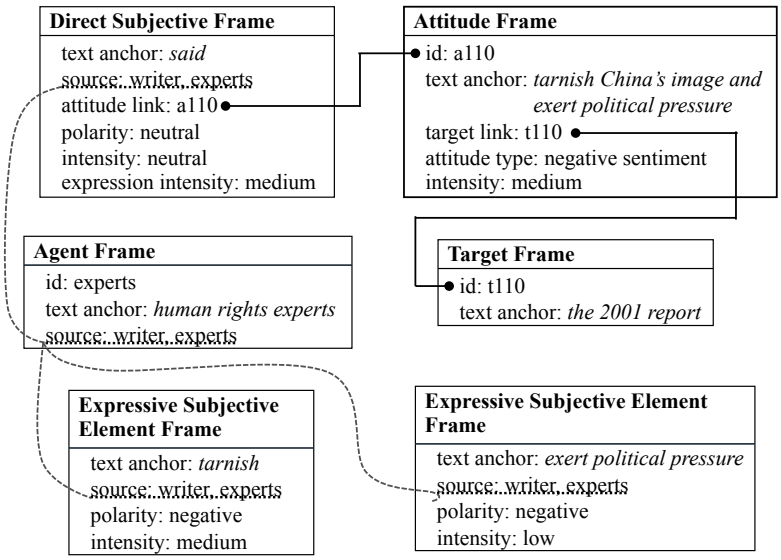

Fig. 2b Frames attributed to source <writer, expert>

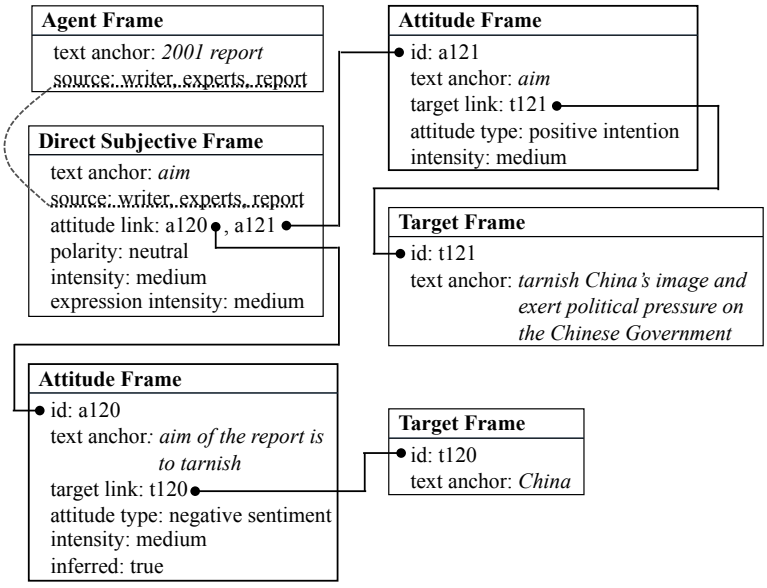

Fig. 2c Frames attributed to source <writer, expert, report>

frame for the report. However, we do not add an id attribute, as one was already created for the report in an eariler agent frame.

To capture the intention attributed to the report, we start by creating a direct subjective frame anchored to "aim." From there, we create two attitude frames and link them both to the direct subjective frame. The first agent frame captures the positive intention of the report. The target of this attitude is also marked and linked back to the attitude frame via the target id. From this positive intention, we can also infer a negative sentiment. This negative sentiment is represented by the second additude frame anchored to "aim of the report is to tarnish." The target of this second attitude is China.

4 Corpus Annotation

To move from a conceptual representation to a fully annotated corpus involves a number of steps, from choosing an annotation tool to converting finished annotations into their final, physical representation. This section, describes these steps for the MPQA Corpus.

4.1 Data Selection

The documents in the MPQA Opinion Corpus were drawn from a much larger collection of English and English-translated news articles, dating from June 2001 to May 2002. Innitially, documents were chosen for annotation from international news topics highly likely to provoke controversy and opinions (e.g., the annual U.S. State Department Human Rights Report and the contested 2002 presidential election in Zimbabwe). These documents did indeed prove to be rich in subjective language. However, having plentiful examples of objective language to counterbalance was also important. Thus, as annotation progressed, randomly selected documents and documents on more objective topics were also included.

The MPQA Corpus contained 535 annotated documents in its initial release, and an additional 157 documents were included in the most recent release. The new documents come from Xbank (85 Wall Street Journal texts), the ULA (48 texts from the American National Corpus), and the ULA-LU (24 texts from the ULA language understanding subcorpus).

4.2 Annotator Training

Training a new annotator always began with having the annotator read the coding manual [44], which was later supplanted by [47] as the terminology evolved. After

reading the manual, training would proceed in two stages. First, the annotator would focus on learning the conceptual representation. Then, after the annotator had a firm grasp of the concepts, he or she would learn to use the annotation tool.

To learn the core concepts in the MPQA scheme, the new annotator would label a document using pencil and paper, compare his or her annotations to the gold standard annotations for the document, and then discuss the document and annotations with the trainer or a more senior annotator. The annotator would repeat this process until he or she had completed four to six training documents. The training documents were not trivial. They were news articles drawn from the same collection of documents that the annotator would be annotating. When the annotation scheme was first being developed, these documents were studied and discussed in detail until consensus annotations were agreed upon.

Once the new annotator could apply the MPQA scheme consistently on paper, he or she would learn to perform the annotations using the annotation tool. The annotator was given an instruction manual that documented exactly how to annotate the MPQA scheme using the annotation tool, as well as a self-paced tutorial that walked the annotator through the process of annotating several short documents. Training would wrap up with additional practice using the tool to annotate never-before-seen documents.

Completing the above training required about 40 hours. From this point, the annotator would annotate independently, although he or she was encouraged to ask questions as needed. The annotations for completed documents would continue to be spot-checked with feedback given as necessary.

4.3 Annotation Tool

The MPQA Corpus was annotated using GATE[4] [10], primarily version 1.2. GATE is open source software that has grown over the years to encompass a vast array of computational tools for research and development in human language technology. In 2002, the range of tools provided by GATE was more modest, but among the functionality it did provide was a pipeline for basic text preprocessing (i.e., tokenization, sentence splitting, and part-of-speech tagging) and an annotation framework. We chose GATE over the other annotation tools available at that time for its ease of use and its ability to store annotations in a stand-off format using byte references.

There are two aspects to consider when evaluating how easy a tool is to use. The first is how straightforward it is to implement the annotation scheme in the tool. We were able to implement the MPQA scheme in the XML format used by GATE in a fairly easily, without losing much of the conceptual representation. The second consideration is how easy the tool is to use for annotation. If the annotation tool is overly cumbersome or difficult to use, it will increases training time and get in the way rather than facilitate annotation. GATE's annotation interface was

[4] https://gate.ac.uk/

very straightfoward. For an annotator familiar with the conceptual representation, introducing them to GATE required little time, and becoming proficient in using the tool for MPQA annotations could be accomplished in an afternoon.

4.4 Annotation Process

To prepare a document for annotation, it was first passed through a tokenizer, sentence splitter, and part-of-speech tagger. The resulting automatic annotations were saved, along with the document text, in a GATE XML file with off-set annotations. We then ran a tool to automatically add a number of default MPQA annotations to the XML file. Implicit, objective-speech event frames for the writer were added at the beginning of each sentence. These could later be changed by the annotator to direct-subjective frames if the annotator determined that the writer was expressing a private state. Several zero-span frames were also added. These included an agent frame, which was used to assign the writer a source ID, and several temporary, zero-span annotation frames. We discovered that GATE would not visibly list the full range of possible annotation types unless an annotation already existed for each type.[5] The temporary annotations were a work around, so all possible annotation types were always displayed to the annotator.

Once preprocessing was complete, the document was assigned to an annotator. The annotator started by correcting any sentence-splitting errors produced during preprocessing. If left uncorrected, such errrors severely affected the resulting annotations, particularly those for the writer of the document. After annotating a document, the annotator would run one or more checkers. These checkers identified errors such as missing frame attributes and orphaned attitude frames. After correcting errors and performing a final check, the annotator would upload the XML file with the annotations to a local dropbox.

The last step in the annotation process was to convert the XML annotations into flat-text files. These flat-text annotation files used a tab-delimited format that was easily read and took up much less space than the XML files.[6] The automatic annotations created during preprocessing were extracted and saved in the location for automatic annotations. The MPQA frames created by the annotator were extracted and saved in a location reserved only for manual annotations.

[5] Annotating all types was always possible, but having them visibly listed and clickable from the start made the task more straightforward.

[6] See the documentation accompanying the MPQA Corpus release for specifics on the MPQA annotation file format and the directory structure for the corpus.

5 Representational Challenges

All annotation projects encounter representational challenges. In this section we discuss some of main challenges we faced, and how the decisions made in response to these challenges affected the annotation scheme.

5.1 Linking Annotations

Before moving from the conceptual representation to the annotation tool, we planned to link agents frames representing sources to their respective private state and speech event frames. However, when implementing the annotation scheme in GATE, we were unable to determine a method for easily linking together annotations.

The solution we came up with was to include the source attribute on every agent, direct subjective, speech event, and ESE frame. Although it is an imperfect linking solution, for many sentences the source attributes do function as virtual links between frames, as can be seen in the detailed example given in Figure 3.5.

5.2 Insubstantial Private States

An early decision we needed to make was how to handle *irrealis* references to private states and speech events. They can be found in hypothetical and conditional statements (*If only he believed* ...), following negations (*The president did not say* ...), and in exaggerations (*Everyone in the world thinks* ...), and they are *not real* in the discourse.

Within the scope of the larger project, irrealis private states were not ones that would be extracted by a question answering system. However, the words and phrases that directly refer to these private states and speech events are the same. The difference comes from the context in which the expressions are used. Excluding these private state and speech event expressions would introduce noise and make the task of learning how private states are expressed even more challenging.

In the end we decided to annotate frames for irrealis private states and speech events, but to mark them as *insubstantial*.[7] In this way, they could be included for experiments that focused on learning subjective language, but excluded for later work on opinion extraction.

[7] Insubstantial private state and speech events also include those that are not significant in the discourse.

5.3 Attitudes and Targets

Although attitudes and targets are key components of private states, it was not always clear how to represent them in the larger conceptualization. We experimented with treating them as attributes on direct subjective frames. We also tried using the agent frame to mark references to entities that were targets.

As annotation continued, we observed that a single direct reference to a private state might encompass more than one type of attitude, as in example 6. Similarly, we found attitudes directed toward multiple targets. Treating attitudes and targets as mere attributes of private states could not capture the complexity that we were seeing in the data. Targets were also proving to be extremely diverse, and limiting them to just entities was becoming very unsatisfactory.

To address these challenges, we chose to change the conceptual representation to give attitudes and targets their own frames. This meant that each attitude and each target annotation would have its own text anchor. To tie all the private state components together, attitude and target frames were given ID attributes, and link attributes were added to direct subjective frames and attitude frames. Allowing the links to be a list of IDs ensured that we could now represent private states with multiple attitudes and attitudes with multiple targets. It is this representation of attitudes and targets that was presented in Section 3.

6 Evaluation

We wrap up our presentation of the MPQA annotations with an overview of the inter-coder reliability for three key aspects of the scheme: identification of text anchors for ESE frames, identification of text anchors for the combined set of direct subjective frames and objective speech event frames (referred to collectively as *explicit* frames), and distinguishing between direct subjective frames and objective speech event frames. For full details of the annotation study that produced these results, and for studies evaluating other aspects of the MPQA scheme, see [47, 51].

To obtain the results reported below, three annotators (A, M, and S) independently annotated 13 documents with a total of 210 sentences. The articles are from a variety of topics and were selected so that 1/3 of the sentences are from news articles reporting on objective topics, 1/3 of the sentences are from news articles reporting on opinionated topics, and 1/3 of the sentences are from editorials.

6.1 Measuring Agreement for Text Anchors

Our first step in measuring agreement was to verify that annotators did indeed agree on which expressions should be marked. To illustrate this agreement problem, consider the words and phrases identified by annotators A and M in example 8. Text

anchors for direct subjective frames are in bold; text anchors for expressive subjective elements are underlined.

Example 8.
A: We **applauded** this move because it was not only just, but it made us **begin to feel** that we, as Arabs, were an integral part of Israeli society.
M: We **applauded** this move because it was not only just, but it made us **begin to feel** that we, as Arabs, were an integral part of Israeli society.

In this sentence, the two annotators mostly agree on which expressions to annotate. Both annotators agree that "applauded" and "begin to feel" express private states and that "not only just" is an expressive subjective element. However, in addition to these text anchors, annotator M also marked the words "because" and "but" as expressive subjective elements. The annotators also do not completely agree about the extent of the expressive subjective element beginning with "integral."

The annotations from 8 illustrate two issues that need to be considered when measuring agreement for text anchors. First, how should agreement be defined for cases when annotators identify the same expression in the text, but differ in their marking of the expression boundaries? The second question to address is which statistic is appropriate for measuring agreement between annotation sets that disagree with respect to the presence or absence of individual annotations.

Regarding the first issue, there was no attempt to define rules for boundary agreement in the annotation scheme or instructions, nor was boundary agreement stressed during training. For the purposes of this research, we believed that it was most important for annotators to identify the same general expression, and that boundary agreement was secondary. Thus, when measuring agreement for text anchors, we consider overlapping text anchors to be matches.

The second issue is that annotators will identify different sets of expressions as part of this task, and thus Cohen's Kappa (κ) [8] is not an appropriate metric for evaluation. In 8, the set of expressive subjective elements identified by annotator A is {"not only just", "integral"}. The set of expressive subjective elements identified by annotator M is {"because", "not only just", "but", "integral part"}. Cohen's κ is appropriate for tasks in which the annotators tag the same set of objects, for example, sense tags applied to a set of word instances. In contrast, measuring agreement for text anchors requires evaluating the intersection between the sets of expressions identified by the annotators. An appropriate evaluation metric for this is F-measure. When evaluating the performance of a system, F-measure is the harmonic mean of precision and recall. When evaluating two sets of annotations from different annotators, precision and recall can be calculated with either annotator standing in for the system, which in practice makes precision and recall interchangeable. If A and B are the sets of anchors annotated by annotators a and b, respectively, then the recall of a with respect to b ($rec(a\|b)$) is as follows:

$$rec(a\|b) = \frac{|A \text{ matching } B|}{|A|}$$

In the 210 sentences in the annotation study, the annotators A, M, and S respectively marked 311, 352 and 249 ESE frames. Table 1, columns 3-5, show the pairwise agreement for these sets of annotations. For example, M agrees with 76% of the expressive subjective elements marked by A, and A agrees with 72% of the expressive subjective elements marked by M.

Table 1 Inter-annotator agreement for text anchors

		ESE Frames			Explicit Frames		
a	b	$rec(a\|\|b)$	$rec(b\|\|a)$	F	$rec(a\|\|b)$	$rec(b\|\|a)$	F
A	M	0.76	0.72	0.74	0.75	0.91	0.82
A	S	0.68	0.81	0.74	0.80	0.85	0.82
M	S	0.59	0.74	0.66	0.86	0.75	0.80
		average		0.71	average		0.81

We measure text-anchor agreement for the combined set of objective speech and direct subjective frames (*explicit* frames), excluding *implicit* frames for the writer of the document. The three annotators, A, M, and S, respectively identified 338, 285, and 315 explicit frames in the data. Table 1, columns 6-8, show the agreement for these sets of annotations. The average F-measure for the text anchors of explicit frames is 0.81, which is 10 points higher than for ESE frames, indicating that speech event and direct subjective frames are more straightforward to identify.

6.2 Agreement Distinguishing between Objective Speech Event and Direct Subjective Frames

Next we focus on inter-rater agreement for judgments that reflect whether or not an opinion, emotion, or other private state is being expressed. We measure agreement for these judgments by considering how well the annotators agree in distinguishing between objective speech event frames and direct subjective frames.

Consider the following example:

Example 9. [*implicit*]"Those digging graves for others, get engraved themselves', he [Abdullah] said while citing the example of Afghanistan.

The underlined words are the text anchors with explicit frames marked by both annotators[8]. Both annotators agree that there is an objective speech event frame for the writer. Likewise they agree that "said" is a direct subjective frame for Abdullah. They disagree, however, as to whether an objective speech event or a direct subjective frame should be marked for text anchor "citing."

To measure agreement for distinguishing between objective speech and direct subjective frames, we first match up the explicit frames identified by both annotators (i.e., based on overlapping text anchors), this time including frames that are

[8] The underlined "implicit" represents the text anchor for frames for the writer of the sentence.

implicit. We then measure how well the annotators agree on the frame type for the annotations in that set using Cohen's κ. Pairwise κ scores for distinguishing between objective speech and direct subjective frames range from 0.74 to 0.84, with an average pairwise κ of 0.81. Under Krippendorff's scale [21], this allows for definite conclusions about the reliability of the annotations.

7 Related Work

The conceptual representation of private states that forms the core of the MPQA annotation scheme grew out of an earlier model developed for tracking point of view in narrative [42, 43]. That model in turn was based on work in literary theory and linguistics [11, 39, 22, 23, 6, 9, 13, 5]. The nested levels of attribution in the conceptual representation were inspired by work on propositional attudes and belief spaces in artificial intelligence [49, 4, 32] and linguistics [13, 12].

When the MPQA annotation scheme was developed, few annotation schemes had been proposed for marking opinions and affect in text. Of these, the most similar conceptually is Appraisal Theory [27, 41], which emerged from the field of systemic functional linguistics [15, 26]. Appraisal Theory provides a framework for analyzing evaluation and stance in discourse. The framework is composed of the following concepts:[9] Affect, Judgement, Appreciation, Engagement, and Amplification. Affect, Judgement, and Appreciation represent different types of positive and negative attitudes. Engagement distinguishes various types of "intersubjective positioning" such as attribution and expectation. Amplification considers the force and focus of the attitudes being expressed.

More recently, Kessler and Nicolov [18] created the *JD Power and Associates (JDPA) Sentiment Corpus*. The data are blog posts about the automotive domain and about digital cameras. The annotations are structural sentiment annotations, which include mentions, co-reference, meronymy, sentiment expressions, and modifiers of sentiment expressions including neutralizers, negators, and intensifiers. For more details of the JDPA Corpus, see Chapter XX.

Since the release of the MPQA Corpus, there has been other work annotating subjectivity in context. Earlier work tends to focus on sentence-level subjectivity and/or sentiment annotations (e.g., [55, 20]), but other corpora with fine-grained subjectivity annotations inspired by the MPQA scheme have also been developed. Included in these are the NTCIR-7 MOAT dataset [34], which has sentence and sub-sentence opinion annotations in English, Japanese, and Chinese, the subjective content annotations in the AMIDA Meeting Corpus [50], and the Darmstadt Service Review Corpus [38]. MPQA-style annotations also have been performed in Italian [3], Korean [35], German [7], and Arabic [1] .

In addition to setting the standard for fine-grained opinion and sentiment annotations, the MPQA Corpus has also served as data for many NLP experiments

[9] Called *systems* in systemic functional linguistics.

published by the co-authors of this chapter (e.g., [54, 2, 52]) and many others. The following are some recent examples: [24, 29, 16, 28, 14, 53, 48, 40, 25, 17, 37, 19].

Finally, other work has built on the MPQA Corpus by adding annotations to support opinion question answering [36] and by annotating modal expressions in subsets of the corpus [33].

Acknowledgements This work was supported in part by the National Science Foundation under grants IIS-0208798 and IIS-0208028, the Advanced Research and Development Activity (ARDA)'s Advanced Question Answering for Intelligence (AQUAINT) Program, and by the Northeast Regional Research Center (NRRC) which is sponsored by the Advanced Research and Development Activity (ARDA), a U.S. Government entity which sponsors and promotes research of import to the Intelligence Community which includes but is not limited to the CIA, DIA, NSA, NIMA, and NRO.

References

[1] Abdul-Mageed M, Diab MT, Korayem M (2011) Subjectivity and sentiment analysis of Modern Standard Arabic. In: Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)

[2] Akkaya C, Wiebe J, Conrad A, Mihalcea R (2011) Improving the impact of subjectivity word sense disambiguation on contextual opinion analysis. In: Proc. of the 15th Conference on Computational Natural Language Learning

[3] Andrea Esuli FS, Urciuoli I (2008) Annotating expressions of opinion and emotion in the Italian Content Annotation Bank. In: Proc. of the 6th International Language Resources and Evaluation

[4] Asher N (1986) Belief in discourse representation theory. Journal of Philosophical Logic 15:127–189

[5] Banfield A (1982) Unspeakable Sentences. Routledge and Kegan Paul, Boston

[6] Chatman S (1978) Story and Discourse: Narrative Structure in Fiction and Film. Cornell University Press, Ithaca, New York

[7] Clematide S, Gindl S, Klenner M, Petrakis S, Remus R, Ruppenhofer J, Waltinger U, Wiegand M (2012) Mlsa a multi-layered reference corpus for german sentiment analysis. In: Proc. of the 8th International Conference on Language Resources and Evaluation

[8] Cohen J (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:37–46

[9] Cohn D (1978) Transparent Minds: Narrative Modes for Representing Consciousness in Fiction. Princeton University Press, Princeton, NJ

[10] Cunningham H, Maynard D, Bontcheva K, Tablan V (2002) GATE: A framework and graphical development environment for robust NLP tools and applications. In: Proc. of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania

[11] Doležel L (1973) Narrative Modes in Czech Literature. University of Toronto Press, Toronto, Canada
[12] Fauconnier G (1985) Mental Spaces: Aspects of Meaning Construction in Natural Language. Cambridge, Massachusetts: MIT Press
[13] Fodor JD (1979) The Linguistic Description of Opaque Contexts. Outstanding dissertations in linguistics 13, Garland, New York & London
[14] Ghosh S, Tonelli S, Johansson R (2013) Mining fine-grained opinion expressions with shallow parsing. In: Proc. of the International Conference Recent Advances in Natural Language Processing
[15] Halliday M (1985/1994) An Introduction to Functional Grammar. London: Edward Arnold
[16] Hermann KM, Blunsom P (2013) The role of syntax in vector space models of compositional semantics. In: Proc. of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
[17] Johansson R, Moschitti A (2013) Relational features in fine-grained opinion analysis. Computational Linguistics 39(3):473–509
[18] Kessler JS, Eckert M, Clark L, Nicolov N (2010) The 2010 ICWSM JDPA Sentment Corpus for the automotive domain. In: 4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC 2010)
[19] Kim J, Li JJ, Lee JH (2010) Evaluating multilanguage-comparability of subjectivity analysis systems. In: Proc. of the 48th Annual Meeting of the Association for Computational Linguistics
[20] Kim SM, Hovy E (2004) Determining the sentiment of opinions. In: Proc. of the 20th International Conference on Computational Linguistics (COLING 2004)
[21] Krippendorff K (1980) Content Analysis: An Introduction to its Methodology. Sage Publications, Beverly Hills
[22] Kuroda SY (1973) Where epistemology, style and grammar meet: A case study from the Japanese. In: Kiparsky P, Anderson S (eds) A Festschrift for Morris Halle, Holt, Rinehart & Winston, New York, NY, pp 377–391
[23] Kuroda SY (1976) Reflections on the foundations of narrative theory—from a linguistic point of view. In: van Dijk T (ed) Pragmatics of Language and Literature, North-Holland, Amsterdam, pp 107–140
[24] Lan M, Xu Y, Niu Z (2013) Leveraging synthetic discourse data via multi-task learning for implicit discourse relation recognition. In: Proc. of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
[25] Lin C, He Y, Everson R (2011) Sentence subjectivity detection with weakly-supervised learning. In: Proc. of 5th International Joint Conference on Natural Language Processing
[26] Martin J (1992) English Text: System and Structure. Philadelphia/Amsterdam: John Benjamins

[27] Martin J (2000) Beyond exchange: APPRAISAL systems in English. In: Hunston S, Thompson G (eds) Evaluation in Text: Authorial stance and the construction of discourse, Oxford: Oxford University Press, pp 142–175
[28] Meng X, Wei F, Liu X, Zhou M, Xu G, Wang H (2012) Cross-lingual mixture model for sentiment classification. In: Proc. of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
[29] Mohtarami M, Lan M, Tan CL (2013) Probabilistic sense sentiment similarity through hidden emotions. In: Proc. of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
[30] Picard R (1997) Affective Computing. MIT Press
[31] Quirk R, Greenbaum S, Leech G, Svartvik J (1985) A Comprehensive Grammar of the English Language. Longman, New York
[32] Rapaport W (1986) Logical foundations for belief representation. Cognitive Science 10:371–422
[33] Ruppenhofer J, Rehbein I (2012) Yes we can!? Annotating English modal verbs. In: Proc. of the 8th International Conference on Language Resources and Evaluation
[34] Seki Y, Evans DK, Ku LW, Sun L, Chen HH, Kando N (2008) Overview of multilingual opinion analysis task at NTCIR-7. In: Proc. of NTCIR-7)
[35] Shin H, Kim M, Jang H, Cattle A (2012) Annotation scheme for constructing sentiment corpus in Korean. In: Proc. of the 26th Pacific Asia Conference on Language, Information, and Computation
[36] Stoyanov V, Cardie C, Wiebe J (2005) Multi-perspective question answering using the OpQA corpus. In: Proc. of the Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)
[37] Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. Computational Linguistics 37(2):1–308
[38] Toprak C, Jakob N, Gurevych I (2010) Sentence and expression level annotation of opinions in user-generated discourse. In: Proc. of the 48th Annual Meeting of the Association for Computational Linguistics
[39] Uspensky B (1973) A Poetics of Composition. University of California Press, Berkeley, CA
[40] Wang S, Manning C (2012) Baselines and bigrams: Simple, good sentiment and topic classification. In: Proc. of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)
[41] White P (2002) Appraisal: The language of attitudinal evaluation and intersubjective stance. In: Verschueren, Ostman, blommaert, Bulcaen (eds) The Handbook of Pragmatics, Amsterdam/Philadelphia: John Benjamins Publishing Company, pp 1–27
[42] Wiebe J (1990) Recognizing subjective sentences: A computational investigation of narrative text. PhD thesis, State University of New York at Buffalo
[43] Wiebe J (1994) Tracking point of view in narrative. Computational Linguistics 20(2):233–287

[44] Wiebe J (2002) Instructions for annotating opinions in newspaper articles. Department of Computer Science Technical Report TR-02-101, University of Pittsburgh
[45] Wiebe J, Breck E, Buckley C, Cardie C, Davis P, Fraser B, Litman D, Pierce D, Riloff E, Wilson T, Day D, Maybury M (2003) Recognizing and organizing opinions expressed in the world press. In: Working Notes of the AAAI Spring Symposium in New Directions in Question Answering, Palo Alto, California
[46] Wiebe J, Wilson T, Bruce R, Bell M, Martin M (2004) Learning subjective language. Computational Linguistics 30(3):277–308
[47] Wiebe J, Wilson T, Cardie C (2005) Annotating expressions of opinions and emotions in language. Language Resources and Evaluation (formerly Computers and the Humanities) 39(2/3):164–210
[48] Wiegand M, Klakow D (2012) Generalization methods for in-domain and cross-domain opinion holder extraction. In: Proc. of the 13th Conference of the European Chapter of the Association for Computational Linguistics
[49] Wilks Y, Bien J (1983) Beliefs, points of view and multiple environments. Cognitive Science 7:95–119
[50] Wilson T (2008) Annotating subjective content in meetings. In: Proc. of the 6th Language Resources and Evaluations Conference
[51] Wilson T (2008) Fine-grained subjectivity and sentiment analysis: Recognizing the intensity, polarity, and attitudes of private states. PhD thesis, Intelligent Systems Program, University of Pittsburgh
[52] Wilson T, Wiebe J, Hoffmann P (2009) Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Linguistics 35(3):399–433
[53] Xiao M, Guo Y (2012) Multi-view AdaBoost for multilingual subjectivity analysis. In: Proc. of the 24th International Conference on Computational Linguistics
[54] Yang B, Cardie C (2013) Joint inference for fine-grained opinion extraction. In: Proc. of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
[55] Yu H, Hatzivassiloglou V (2003) Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: Proc. of the Conference on Empirical Methods in Natural Language Processing