

7.0 REPRESENTING ATTITUDES AND TARGETS

Private states in language are often quite complex in terms of the attitudes they express and the targets of those attitudes. For example, consider the private state represented by the direct subjective phrase “are happy” in the following sentence.

(7.1) “I think people are happy because Chavez has fallen.”

In this sentence, the word “happy” expresses a positive attitude, specifically, the positive sentiment of the people toward the fall of Chavez. However, the private state attributed to the people in this sentence encompasses more than just a positive sentiment. There is a second attitude, a negative sentiment toward Chavez himself, which can be inferred from the phrase “happy because Chavez has fallen.”

Just as a private state may involve more than one type of attitude, an attitude may be directed toward more than one target. In sentence (7.2) there is a private state being expressed by Tsvangirai.

(7.2) Tsvangirai said the election result was a clear case of highway robbery by Mugabe, his government and his party, Zanu-PF.

The negative sentiment of this private state is expressed with the phrase “a clear case of highway robbery,” and it is directed toward two things: “the election results” and “Mugabe, his government and his party, Zanu-PF.”

In this chapter, I extend the original conceptual representation of private states (Chapter 3) to better model attitudes and their targets. In the original conceptualization, attitudes are represented with the *attitude type* attribute in direct subjective and expressive subjective element frames, and targets are represented with the *target* attribute in direct subjective

frames. A drawback to representing attitudes and targets in this way is that it does not allow for multiple attitudes and targets to be associated with a private state. In the new representation, attitudes and targets are conceptualized as annotation frames, with target frames linking to attitude frames and attitude frames linking to private state frames. This representation gives the flexibility needed to associate multiple attitudes and targets with a single private state.

The new representation also includes a new, more clearly-defined set of attitude types. What types of attitudes are useful for NLP is an open question and, at least to a certain extent, application dependent. Sentiment, which so far has received the most attention, is clearly important. However, as the contextual polarity annotations in Chapter 6 show, sentiment is far from the only type of attitude: 45% of the subjective expressions in the MPQA corpus express some type of attitude other than a sentiment. Sentence 7.1 above contains an example of a private state expressing a type of attitude other than sentiment. In the context of Sentence 7.1, the word “think” is being used to express an opinion about what is true according to its source. I developed the set of attitude types presented in this chapter with an eye toward what would be useful for NLP applications, in particular an application like question answering. I hypothesize that these attitude types can be reliably annotated, and that they will provide good coverage of the private states expressed in the MPQA Corpus.

With my extension to the conceptualization, I aim to improve one more aspect of the representation of private states: intensity. Judging the intensity of private states is challenging. In Chapter 3, I evaluated inter-annotator agreement for the various intensity judgments in the original conceptualization. Although inter-annotator agreement is acceptable for the *intensity* and *expression-intensity* of the combined set of direct subjective and objective speech event annotations, agreement for the intensity of expressive subjective elements is low. One way in which intensity judgments might be improved is to judge intensity with respect to attitude type, for example, to compare the intensity of a positive sentiment to other positive sentiments rather than very dissimilar types of attitudes such as speculations or intentions. Thus, in the new conceptualization, I define intensity explicitly according to the new set of attitude types.

7.1 CONCEPTUAL REPRESENTATION

In this section, I describe my extensions to the conceptual representation for attitude types and targets. I begin by introducing the new set of attitude types, and then describe the new attitude and target annotation frames and how they are integrated into the overall conceptual representation for private states. At the end of the section I give a graphical example to illustrate the new annotations.

7.1.1 Types of Attitude

When determining a set of attitude types, there are any number of possible distinctions that might be considered. The attitude types in the original conceptualization distinguish very generally between *positive* and *negative* attitudes, with other types of attitudes being lumped together into one category. For the new set of attitude types, my goal is to define more fine-grained distinctions. However, some distinctions may actually be too fine grained to be of use to an application. Would an application such as question answering benefit from being able to distinguish between a positive emotion and a positive evaluation? Or, would distinguishing between positive sentiments (which include both emotions and evaluations) and intentions be more helpful? Working with the annotators of the MPQA Corpus, looking at the private states already annotated, and keeping in mind what might be useful for an application like QA, I developed the set of attitude types listed in Table 7.1.

At the coarser level of distinction, there are six attitude types: sentiment, agreement, arguing, intention, speculation, and all other attitudes. Sentiment, agreement, arguing, and intention may be further broken down into positive and negative variants. Below I define and give examples of each of the attitude types and their targets. In each example, the span of text where the attitude is expressed is in bold, and the span of text that denotes the target of the attitude (if a target is given) is in angle brackets.

7.1.1.1 Sentiments *Sentiments* are positive and negative emotions, evaluations, and stances. This is the same definition of sentiment that is used in Chapter 6. The target of a

Table 7.1: Set of attitude types

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

sentiment is what the sentiment is directed toward. Sentence 7.3 contains an example of a positive sentiment, and Sentence 7.4 contains an example of a negative sentiment.

Positive Sentiment:

(7.3) The Namibians went as far as to say ⟨Zimbabwe’s election system⟩ was “**water tight, without room for rigging**”.

Negative Sentiment:

(7.4) His disenfranchised supporters **were seething**.

7.1.1.2 Agreement Private states in which a person does or does not agree, concede, consent, or in general give assent to something fall into the category of *Agreement*. Agreement includes both agreeing with a statement or idea and agreeing to an action. The target for this attitude type is what is (or is not) being agreed to. Sentence 7.5 gives an example of positive agreement, and sentence 7.6 gives an example of a negative agreement. Sentence 7.7 has examples of both negative (“differed over”) and positive (“agreed”) agreement.

Positive Agreement:

(7.5) Republicans **concede** that ⟨at this point it could be his only option⟩.

Negative Agreement:

(7.6) Afghanistan is now under US bombardment for **refusing** ⟨to hand over the chief suspect in the Sept. 11 attacks on New York and Washington⟩.

(7.7) Japanese Prime Minister Junichiro Koizumi and visiting U.S. president George W. Bush **differed over** ⟨the Kyoto Protocol and how to prevent global warming⟩ but **agreed** ⟨to cooperate on that issue⟩.

7.1.1.3 Arguing Private states in which a person is arguing or expressing a belief about what is true or should be true in his or her view of the world are categorized as *Arguing*. Arguing attitudes include private states where the source is arguing for or against something.

Deciding on what spans to annotate for arguing attitudes (and speculation (Section 7.1.1.5), which is similar to arguing) and their targets actually turned out to be a challenging part of the annotation scheme development. In initial annotation rounds with another annotator, there was a great deal of inconsistency in what spans were marked for arguing attitudes and their targets, even though there was agreement that arguing is present. Eventually, I decided on the following strategy for marking arguing attitude and arguing target spans, because it seemed to produce the most consistent span annotations: mark the arguing attitude on the span of text expressing the argument or *what the argument is*, and mark *what the argument is about* as the target of the arguing attitude.

Sentences 7.8 and 7.9 contain examples of positive arguing attitudes, and sentence 7.10 and 7.11 contain negative arguing attitudes.

Positive Arguing:

(7.8) Iran **insists** ⟨its nuclear program⟩ **is purely for peaceful purposes.**

(7.9) Putin remarked that ⟨the events in Chechnia⟩ **“could be interpreted only in the context of the struggle against international terrorism.”**

Negative Arguing:

(7.10) Officials in Panama **denied that** ⟨Mr. Chavez or any of his family members⟩ **had asked for asylum.**

(7.11) “⟨It⟩ **is analogous to the US crackdown on terrorists in Afghanistan,**”
Ma said.

7.1.1.4 Intentions *Intentions* include aims, goals, plans, and other *overt* expressions of intention. Positive intentions are straightforward. Negative intentions are the opposite of positive intentions. They are the intentions that the source of the private state is described explicitly as **not** holding. The target of an intention is the thing that is (or is not) the aim,

goal, plan, or intention. Sentence 7.12 has an example of a positive intention, and the private state in sentence 7.13 is an example of a negative intention.

Positive Intention:

(7.12) The Republic of China government believes in the US **commitment** ⟨to separating its anti-terrorism campaign from the Taiwan Strait issue⟩, an official said Thursday.

Negative Intention:

(7.13) The Bush administration **has no plans** ⟨to ease sanctions against mainland China⟩.

7.1.1.5 Speculations Private states in which a person is speculating about what is or is not true, or what may or may not happen, are categorized as *Speculation*. Similar to arguing, the span of text marked for speculation is *what the speculation is*, and *what the speculation is about* is marked as the target of the speculation. Sentence 7.14 gives an example.

(7.14) ⟨The president⟩ is **likely to endorse the bill**.

7.1.1.6 Other Attitudes The hope is that most private states will fall into the set of attitudes described above. However, for those that do not there is this category. Private states that would be captured by this catch-all category are neutral emotions (emotions that don't seem clearly positive or negative), cognition, and general uncertainty. Sentences 7.15 and 7.16 give two examples of other attitudes.

(7.15) To **the surprise of** many, ⟨the dollar hit only 2.4 pesos and closed at 2.1⟩.

(7.16) “I'm **not sure** whether ⟨I should wait in line or sell to one of the street traders⟩,” said Fabian, a 36-year old attorney.

In sentence 7.15, it is not clear from the context whether the emotion surprise is positive or negative, so it is categorized as other. In sentence 7.16, Fabian, the source of the private state, is expressing his uncertainty.

7.1.2 Attitude Frames

When considering the representation of the new attitude annotations, the major question to address, aside from what set of attitude types to use, is to which spans of text should the new

attitudes be anchored? Arguing and speculation attitude types created their own challenge, and the spans to mark for these attitudes are described above. Sentence 7.17 illustrates a more general problem.

(7.17) The MDC leader said systematic cheating, spoiling tactics, rigid new laws, and shear obstruction – as well as political violence and intimidation – were just some of the irregularities practised by the authorities in the run-up to, and during the poll.

In this sentence, there are five private state frames attributed to the MDC leader: a direct subjective frame anchored to “said,” and four expressive subjective element frames anchored respectively to “systematic cheating . . . obstruction,” “as well as,” “violence and intimidation,” and “just some of the irregularities.” One option is to create an attitude frame for each of the private state frames. However, this would be very redundant, both in the expressions that would be annotated and in the sentiment annotations that would result. A better solution is to annotate the span of text that expresses the attitude of the overall private state represented by the direct subjective frame. Specifically, for each direct subjective frame, first the attitude type(s) being expressed by the source of the direct subjective frame are determined by considering the text anchor of the frame and everything within the scope of the annotation attributed to the source. Then, for each attitude type identified, an attitude frame is created and anchored to whatever span of text completely captures the attitude type. In sentence 7.17, this results in just one attitude frame being created to represent the negative attitude of the MDC leader. The anchor for this attitude frame begins with “systematic cheating” and ends with “irregularities.”

To tie the attitude frames back to the direct subjective frame, each attitude annotation is given a unique, alphanumeric identifier, and a new *attitude link* attribute is created in the direct subjective frame. The value of the *attitude link* attribute is a list of one or more attitude frame identifiers.

Figure 7.1 gives the attributes for the attitude frame. The *id* attribute is the unique identifier used to link the attitude frame back to its corresponding direct subjective frame. The *text anchor* attribute points to the span of text on which the attitude is marked. The *attitude type* attribute is one of the attitude types described in the previous section. The *target link* attribute is used to link the attitude frame to its targets. For the rare cases in

Figure 7.1: Attitude frame

- **id:** a unique, alphanumeric ID for identifying the attitude annotation. The ID is used to link the attitude annotation to the private state that it is a part of.
- **text anchor:** a pointer to the span of text that captures the attitude being expressed.
- **attitude type:** type of attitude being expressed (Table 7.1).
- **target link:** list of one or more target frame IDs, or the string *none*.
- **intensity:** *low, low-medium, medium, medium-high, high, high-extreme*.
- **properties:**
 - **inferred:** true, if the attitude is inferred. (The *inferred* property will be described in more detail in Section 7.1.5.)
 - **sarcastic:** true, if the attitude is realized through sarcasm. (This attribute is discussed in more detail in Section 7.1.5.)
 - **repetition:** true, if the attitude is realized through the repetition of words, phrases, or syntax. (This attribute is discussed in more detail in Section 7.1.5.)
 - **contrast:** true, if the attitude is realized only through contrast with another attitude. (This attribute is discussed in more detail in Section 7.1.5.)

which an attitude has no clear target, the *target link* attribute is assigned the string *none*. Otherwise, it is a list of one or more target frame IDs (described in Section 7.1.3). The *intensity* attribute captures the intensity of the attitude type being expressed. Table 7.2 defines how the intensity for each attitude type should be evaluated. In addition to defining intensity more explicitly and with respect to the different attitude types, the values for intensity are more fine-grained than those used for the various intensity attributes in the original conceptualization. The hope is that this as well will help to improve inter-annotator agreement for intensity judgments. The remaining properties of the attitude frame are described later in Section 7.1.5.

7.1.3 Target Frames

Once an attitude frame has been created, the span of text representing the target of the attitude is identified and a target frame is created and anchored to that text span. Like the attitude frames, each target frame is given a unique, alphanumeric identifier. These identifiers and the *target link* attributes on the attitude frames are used to tie the attitude

Table 7.2: Measures of intensity for different attitude types

Attitude Type	Measure of Intensity	Example
Positive Sentiment	degree of positiveness	<i>like < love</i>
Negative Sentiment	degree of negativeness	<i>criticize < excoriate</i>
Positive Agreement	degree of agreement	<i>mostly agree < agree</i>
Negative Agreement	degree of disagreement	<i>mostly disagree < completely disagree</i>
Positive Arguing	degree of certainty/strength of belief	<i>critical < absolutely critical</i>
Negative Arguing	degree of certainty/strength of belief	<i>should not < really should not</i>
Positive Intention	degree of determination	<i>promise < promise with all my heart</i>
Negative intention	degree of determination	<i>no intention < absolutely no intention</i>
Speculation	degree of likelihood	<i>might win < really might win</i>

and target frames together. The target frame is given in Figure 7.2.

7.1.4 Example

To help to illustrate the new attitude and target frames and how they fit in with the original conceptual representation, Figure 7.3 gives the various direct subjective, attitude, and target frames for sentence 7.18 and shows how they are all linked together.

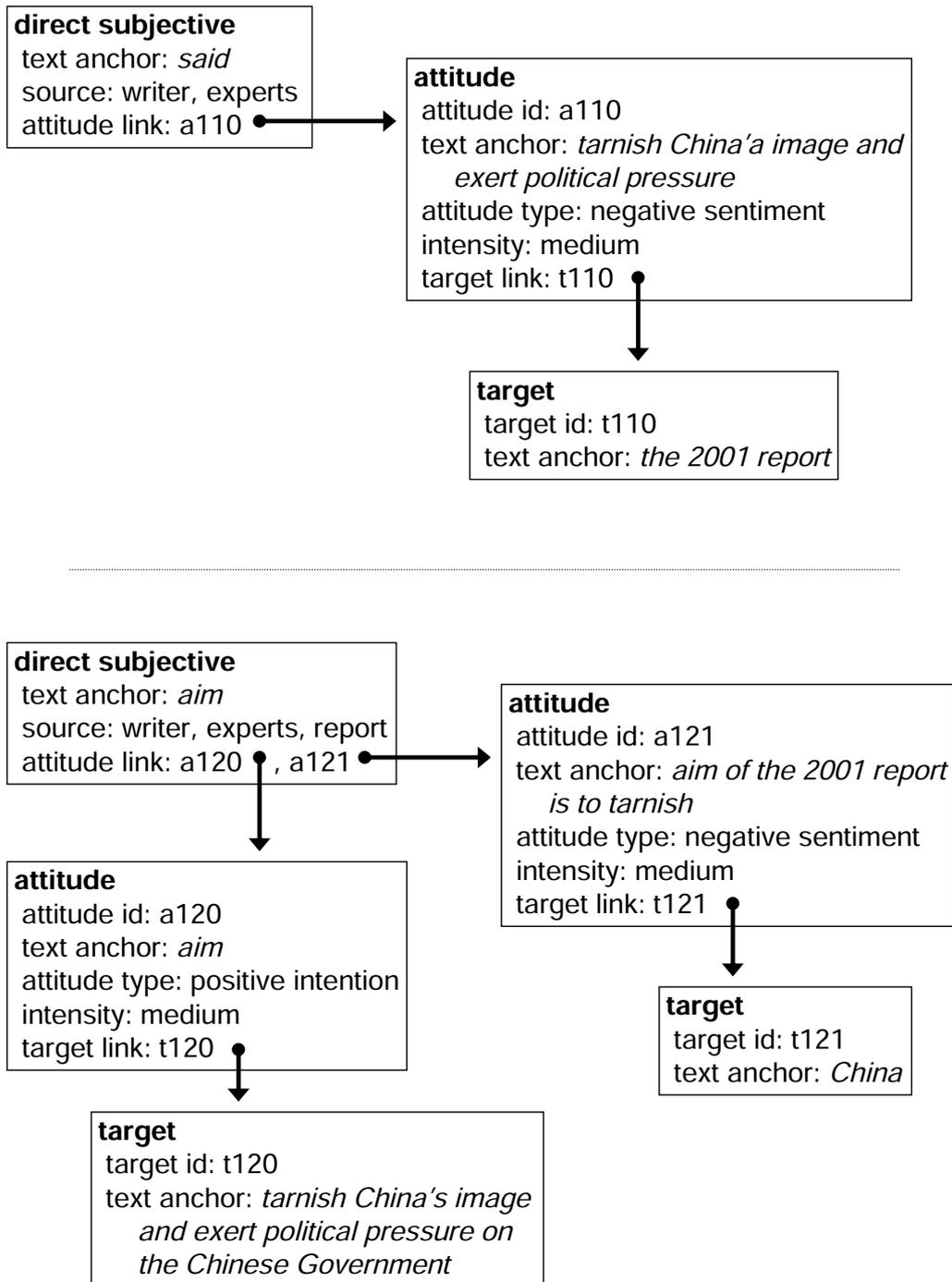
(7.18) Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government, human rights experts said at the seminar held by the China Society for Study of Human Rights (CSSHR) on Friday.

There are two direct subjective frames in sentence 7.18. The private state represented by the direct subjective frame for “said” has one attitude with one target. The attitude is a negative sentiment expressed by the phrase “tarnish China’s image and exert political

Figure 7.2: Target frame

- **id:** a unique alphanumeric ID for identifying the target annotation. The ID is used to link the target to the attitude frame.
- **text anchor:** a pointer to the span of text that denotes the target.

Figure 7.3: Private state, attitude, and target frames for sentence 7.18



pressure.” The target of the negative sentiment is the 2001 report.

There are two attitudes for the private state represented by the direct subjective frame for “aim.” The first attitude is a positive intention with the target tarnishing China’s image and exerting political pressure on the Chinese Government. The second attitude is the negative sentiment that is conveyed by the phrase “aim of the 2001 report is to tarnish.” Having an aim to tarnish indicates a negative sentiment. The target of the negative sentiment is China.

7.1.5 Additional Characteristics of Attitudes

The attitude frame has four additional properties that are used to mark particular characteristics of attitudes when they are relevant. The first of these properties is for marking when an attitude is *inferred*. The remaining properties represent characteristics of how attitudes are sometimes expressed. I include these properties in the attitude frame because I feel they may be useful in developing automatic systems for recognizing different types of attitudes.

7.1.5.1 Inferred Attitudes Most attitudes are directly evoked by the words and phrases that are used to express a private state. However, sometimes attitudes are *inferred*. For example, in the sentence *I think people are happy because Chavez has fallen* (sentence 7.1 above), the negative sentiment of the people toward Chavez is an inferred attitude. The most prominent attitude of the private state attributed to the people is a positive sentiment toward Chavez’s fall, but the negative sentiment toward Chavez is only a short inference away.

One problem with marking inferred attitudes is that it is very easy to start “digging too deep” and inferring any number of very subtle attitudes. To cut down on the possibilities for this, annotators are instructed to mark only inferred attitudes that have people or other entities as their targets.

7.1.5.2 Characteristics of How Attitudes Are Expressed The properties that represent various ways attitudes may be expressed are *sarcastic*, *repetition*, and *contrast*.

The *sarcastic* property is for marking attitudes expressed using sarcasm. In general, I

believe this property will be of interest for NLP applications working with opinions. Detecting sarcasm may also help a system learn to distinguish between positive and negative sentiments. The sarcasm in sentence 7.19 below makes the word “Great” an expression of negative rather than positive sentiment.

(7.19) “Great, keep on buying dollars so there’ll be more and more poor people in the country,” shouted one.

The *repetition* property is used when an attitude and its intensity are expressed at least in part using the repetition of a word or phrase within a sentence or within several consecutive sentences. In sentence 7.20, the repetition of the phrase “a window” contributes a great deal to the intensity of the positive sentiment expressed by Taiwan.

(7.20) Taiwan’s WTO access has given Taiwan a window to the world, a window to the century and a window of opportunity . . .

The *contrasted* property is used to mark positive and negative attitudes where the type of attitude is only evident because the attitude is contrasted with an attitude of the opposite polarity. For example, consider the attitudes for the Italian senator and the United States in sentence 7.21.

(7.21) The Italian senator’s words are in sharp contrast to what was contained in the so-called China human rights report compiled by the United States, which blindly accuses China of restricting religious freedom in Tibet.

The negative sentiment of the United States toward China is clearly indicated by the phrase “blindly accuses China of restricting religious freedom in Tibet.” However, the positive sentiment of the Italian senator is understood only because it is contrasted with the negative sentiment of the US. The attitude for the Italian senator would thus be marked with the *contrast* property.

7.2 AGREEMENT STUDIES

In this section, I test the general hypothesis that the extensions to the conceptual representation of private states presented in the previous section can be reliably annotated. Given documents already annotated with private state frames according to the original conceptual representation, I first evaluate whether annotators agree about the attitudes for these private states. I then turn to the question of whether judging intensity according to attitude type gives an improvement in intensity agreement. Finally, I evaluate how well annotators agree in their target frame annotations.

I conducted two inter-annotator agreement studies. In the first study, another annotator and I independently annotated 13 documents with 325 sentences and 409 direct subjective annotations. Two months later, during which we at times discussed our annotations, we annotated another 11 documents with 211 sentences and 207 direct subjective annotations. All intensity and contextual polarity attributes were removed from the existing private state annotations in these documents before each study began.

7.2.1 Agreement for Attitude Frames and Attitude Types

Measuring agreement for attitudes requires first identifying and aligning the sets of attitudes marked by both annotators. If each annotator marked only one attitude for a given direct subjective frame d , the process of matching up their attitude annotations is straightforward. Let d_a and d_b be the attitude frames marked by the two annotators for d . If the text anchors of d_a and d_b overlap, then d_a is said to **match** d_b , and the two attitude frames are included in the set of attitudes marked by both annotators.

When one annotator or both mark more than one attitude for d , the process of matching up the attitude frames is a bit more complicated, involving both the text anchors and the attitude types of the attitude frames. No attitude marked by annotator a is allowed to match with more than one attitude marked by annotator b , and vice versa. Thus, if annotator a marked two attitudes on d , d_{a1} and d_{a2} , and annotator b only marked one attitude, d_{b1} , and if both d_{a1} and d_{a2} overlap with d_{b1} , only one of a 's annotations can be matched with d_{b1} . When

Table 7.3: Inter-annotator agreement: Attitudes

	$ A $	$ B $	$recall(a b)$	$recall(b a)$	F-measure
Study 1	515	549	0.91	0.86	0.88
Study 2	247	283	0.95	0.83	0.89

choosing which attitude of a to match to d_{b_1} , preference is given first to whichever attitude has the same attitude type as d_{b_1} , and then to whichever attitude of a has a text anchor with the larger overlap with the text anchor of d_{b_1} . There are direct subjective annotations in which both annotators mark multiple attitudes that overlap. These are more complicated, but the matches are resolved in a similar way.

To evaluate how well the two annotators agree on identifying the same set of attitude annotations, I use recall and F-measure, just as I did for measuring inter-annotator agreement for private state and speech event text anchors in Chapter 3. As before, recall is calculated with respect to each annotator. The recall of a with respect to b is

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

and the recall of b with respect to a is

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

Table 7.3 gives the agreement for the attitude frames marked by the annotators in the two studies. The first two columns in the table show the number of attitudes marked by each annotator in the two studies, followed by their respective recalls. The last column is F-measure. In study 1, the two annotators agree on a total of 470 attitude frames, and in study 2 they agree on 235 attitude frames. This results in an average F-measure of 0.885 over the two studies, which indicates that the annotators largely agree on the attitude frames that they marked.

Now that the set of attitudes marked by both annotators has been identified and the attitude annotations have been match up, I can use Cohen’s Kappa (κ) to evaluate how well

Table 7.4: Inter-annotator agreement: Attitude-types

	Fine		Conflated	
	κ	%	κ	%
Study 1	0.79	83%	0.78	86%
Study 2	0.81	85%	0.77	86%

the annotators agree on the attitude types of the attitudes that they marked. As Table 7.4 shows, agreement for attitudes types is high, with κ values near 0.80. The first two columns in the table give κ and percent agreement values for the finer grained set of attitude types. The last two columns give κ and percent agreement for the more general set of attitude types, in which the positive and negative variants are conflated (e.g., positive sentiment and negative sentiment are conflated into one category).

One interesting finding from these studies is that very few of the disagreements come from the annotators agreeing about the general type, but disagreeing about the polarity. There are only two positive sentiment/negative sentiment disagreements, and one positive arguing/negative arguing disagreement in study 2. In study 1, only 12 (15%) of the disagreements are between positive and negative attitudes of the same general category. By far the majority of disagreements in attitude-type judgments are between different general categories. Table 7.5 gives the contingency table showing these disagreements for study 1.

7.2.2 Agreement for Attitude Intensity

To evaluate the inter-annotator agreement for the intensity of attitudes, I again use the sets of matched attitude frames identified by both annotators. As I did in Chapter 3, I use Krippendorff’s α to calculate agreement for intensity. Like κ , Krippendorff’s α takes into account chance agreement, but unlike κ , it can be used to calculate agreement for ordinal judgments.

With α , a distance metric is used to weight disagreements. When measuring agreement for the different intensity judgments that are part of the original conceptualization, I use the

Table 7.5: Confusion matrix for conflated attitude-type agreement for Study 1

		B					Total	
		Sentiment	Agreement	Arguing	Intention	Speculation		Other
A	Sentiment	203	3	13	2	0	13	234
	Agreement	0	5	2	0	0	2	9
	Arguing	6	1	145	0	0	4	156
	Intention	1	1	0	17	0	3	22
	Speculation	1	0	0	0	6	2	9
	Other	4	2	3	3	0	28	40
	Total	215	12	163	22	6	52	470

scale [0,1,2,3], where 0 represents neutral and 3 represents high. The scale that I use for the intensity of attitudes is more fine grained, but it can still be matched to the original scale by mapping the *low-medium* and *medium-high* ratings to mid-points (1.5 and 2.5), and by merging the *high* and *high-extreme* ratings. With this numeric scale of intensity, I can use the square of the difference between any two disagreements as the distance metric. Thus, the distance weight is 0.25 for any disagreement that differs by one half (e.g., low-medium and medium), the distance weight is 1 for any disagreement that differs by 1 (e.g., low and medium), the weight is 4 for any disagreement that differs by two (e.g., low and high).

The α -agreement scores for attitude intensity for the two agreement studies are 0.65 and 0.61. These values are not high. Krippendorff recommends a value of at least 0.67 in order to draw tentative conclusions about reliability.

I hypothesized that defining intensity according to attitude type and using a finer-grained intensity scale would result in better intensity agreement as compared to how intensity was judged in the original conceptualization. In the agreement study reported in Chapter 3, average pairwise α -agreement for the intensity of expressive subjective elements is 0.46, with the highest pairwise agreement being 0.52. I did not previously report agreement for the intensity of direct subjective frames.¹ However, by identifying the set of direct subjective

¹Agreement for the intensity and expression-intensity of the combined set of direct subjective and objective speech event frames was reported, but these are not directly comparable.

frames marked by each annotator pair in the Chapter 3 study, I calculate that the average pairwise α for direct subjective intensity is 0.44, with the highest pairwise agreement being 0.56.

At first glance, α -agreement for attitude intensity is higher than these earlier intensity agreement scores. However, is the agreement higher because of how it was defined (according to attitude type), or because of the finer-grained intensity ratings? I experimented with different strategies for conflating the mid-point intensity ratings (e.g., low-medium) so that the rating scale for the attitude intensity would be exactly the same as the intensity scale used in the previous study. With low-medium merged with medium and medium-high merged with high, α -agreement drops to 0.59 for study 1 and 0.55 for study 2. These agreement scores are still higher than the α -agreement for the best annotator pair agreements in the earlier study, but not by much. Judging intensity according to attitude type may be helpful, but the current annotation study does not provide very strong evidence in support of that hypothesis.

7.2.3 Agreement for Targets

In this section, I evaluate the inter-annotator agreement for the targets of the attitude frames identified by both annotators. Recall that the *target-link* attribute marked on every attitude frame either has the value *none*, or it is a list of one or more target frame ids. For the purpose of measuring target agreement, I treat the *none* value as a special type of target. If two matching attitude frames both have targets that are *none*, then the targets of those attitudes are also a match. To calculate whether two targets (other than *none*) of two matching attitude frames do themselves match, I look only at whether the text anchors for the targets overlap. If the text anchors overlap, then the two targets match; otherwise, they do not match.

Unlike with attitudes, I do allow a target marked by one annotator to match with more than one target marked by the other annotator. This is fairly uncommon, but it does happen, for example, when one annotator chooses to break one large target span into two different targets. This is what happened with the targets marked for the attitude anchored

Table 7.6: Inter-annotator agreement: Targets

	Attitudes	$ A $	$ B $	$recall(a b)$	$recall(b a)$	F-measure
Study 1	470	479	503	0.85	0.85	0.85
Study 2	235	244	247	0.86	0.86	0.86

on “criticized” in Sentence 7.22. The targets marked for each annotator are in angle-brackets.

(7.22)

A: US Senate Majority leader Tom Daschle **criticized** on Monday ⟨President George W. Bush for his remarks that described Iran, Iraq and the Democratic People’s Republic of Korea (DPRK) as ”axis of evil”⟩.

B: US Senate Majority leader Tom Daschle **criticized** on Monday ⟨President George W. Bush⟩ for ⟨his remarks that described Iran, Iraq and the Democratic People’s Republic of Korea (DPRK) as ”axis of evil”⟩.

Because the annotators in essence are still capturing the same entities for the target of the attitude, I decided it was appropriate to allow multiple target matches.

Table 7.6 gives the results for target agreement. As for the attitude frames, target agreement is measured using F-measure and recall. The first column of the table lists the number of matching attitudes for each study. It is the targets marked on these attitudes that I consider when calculating target agreement. The next two columns give the number of targets marked on these attitudes by annotators a and b , followed by the recall for each annotator. Target agreement is very similar for both studies, around 0.85. Although lower than agreement for the attitude frames, the study shows that annotators largely agree on what are the targets of attitudes.

7.3 OBSERVATIONS

About two-thirds of the documents in the MPQA Corpus version 2.0 have been annotated with a layer of attitude and target annotations. In this section, I briefly explore what has

Table 7.7: Distribution of attitude types for attitude frames and direct subjective frames

Attitude Type	% of Attitude Frames	% of Direct Subjective Frames
Positive Sentiment	16.7	20.0
Negative Sentiment	32.8	38.0
Positive Agreement	1.9	2.4
Negative Agreement	1.7	2.1
Positive Arguing	25.6	30.1
Negative Arguing	6.4	7.7
Positive Intention	5.2	6.5
Negative Intention	0.5	0.6
Speculation	2.3	3.0
Other Attitude	7.8	9.8

actually been annotated in terms of various distributions of the attitude annotations in a set of 284 documents. I call this set of documents the **attitude dataset**. I use the attitude dataset in the next chapter in my experiments in automatic attitude recognition.

There are 4,499 sentences in the attitude dataset. Of these sentences, 2,829 (63%) are subjective (i.e., they contain at least one direct subjective frame), with a total of 4,538 direct subjective frames and 5,739 attitude annotations. This means that, on average, there are 1.6 direct subjective frames and 2 attitude frames in every subjective sentence. The majority of direct subjective frames, 80%, are linked to just one attitude frame. 18% of direct subjective frames are linked to two attitudes, and a very small 2% are linked to three attitudes. There is one direct subjective frame linked to four attitudes.

Table 7.7 shows two distributions of attitude types. The first column gives the distribution of attitude types for all the attitude frames marked in the dataset. The second column gives the distribution of the direct subjective frames with respect to the types of attitudes they are linked to. Because a direct subjective frame can be linked to more than one type of attitude, these percentages will not sum to 100.

As Table 7.7 shows, sentiments and arguing attitudes make up the largest number of attitude types. Almost 50% of the attitude annotations are sentiments, and nearly one-third are arguing. Interestingly, of the remaining attitude types, other attitude is the category

with the next highest number of attitude. However, fewer than 10% of the attitudes are marked as other, showing that the set of attitudes proposed in this chapter do have fairly good coverage of the types of private states expressed in the news.

7.4 RELATED WORK

Research into types of attitudes and models of emotion has been the focus of work in linguistics and psychology for many years. In psychology, for example, there is a long tradition of using hand-compiled emotion lexicons in experiments to help develop or support various models of emotion. One line of research (e.g., Osgood et al. (1957), Heise (1965), Russell (1980), and Watson and Tellegen (1985)) uses factor analysis to determine dimensions for characterizing emotions. Dimensions corresponding to polarity and intensity are two that are consistently identified. Other researchers (e.g., de Rivera (1977), Ortony et al. (1987), and Johnson-Laird and Oatley (1989)) develop taxonomies of emotions. My goals in developing the set of attitude types presented in this chapter and the goals of these works in psychology and linguistics are quite different. I am not interested in building models or taxonomies of emotion, but rather in identifying types of attitude that would be useful to recognize for improving NLP systems.

Appraisal Theory (Martin, 2000; White, 2002) is again the work most similar to the conceptual representation that I presented in this chapter. Appraisal Theory provides a framework for analyzing evaluation and stance in discourse, in context and below the level of the sentence. The three main concepts (systems) in the Appraisal framework correspond to different types of attitudes: *Affect*, which focuses on emotional responses and dispositions, *Judgement*, which is concerned with evaluating human behavior, and *Appreciation*, which is used for evaluating products and processes. For all of these concepts, Appraisal Theory distinguishes between positive and negative variants and the degree of force (intensity). Appraisal Theory also has a system for *Engagement*, which is used to capture hedging, modality, and evidentiality, among other types of subjectivity.

Although there is an overlap in the types of attitude represented by Appraisal Theory

and the set of attitude types that I present in this chapter, the two representations are not the same. I do not distinguish between affect and the different types of evaluations. Instead, these all fall under the general *sentiment* category in my representation. In addition, the set of attitudes that I propose includes several types of attitude that are not represented at all in the Appraisal framework, such as agreement and intention. Appraisal Theory also does not include a representation for the target of attitudes, and there is no notion that a single span of text can express more than one type of attitude.

Other text corpora have been developed with annotations of positive and negative sentiments (Yu and Hatzivassiloglou, 2003; Bethard et al., 2004; Kim and Hovy, 2004; Hu and Liu, 2004). In contrast to the below-the-sentence attitude annotations presented in this chapter, the corpora developed by Yu and Hatzivassiloglou (2003), Bethard et al. (2004), and Kim and Hovy (2004), only provide sentence-level annotations. The corpus developed by Hu and Liu (2004) is a bit different from the others. As I do, they annotate targets, specifically products and product features in review data. However they do not then mark the spans of text that express positive and negative sentiments about the targets. Instead, sentiment is annotated as an attribute of the target annotations. It simply captures whether in the sentence there is a positive or negative sentiment toward the target.

7.5 CONCLUSIONS

In this chapter, I extend the original conceptual representation for private states to better model attitudes and their targets. The extension includes a new, more clearly defined set of attitude types and new annotation frames for attitudes and targets. This new representation gives the flexibility needed to associate multiple attitudes and targets with a single private state. Also in the new conceptualization, I redefine intensity explicitly in terms of the new set of attitude types.

I hypothesized that the new scheme for attitudes and targets could be reliably annotated. To test this hypothesis, I conduct two different inter-annotator agreement studies with another annotator. F-measure agreement for attitude frames is 0.885, and agreement

for targets is on average 0.855. Inter-annotator agreement for attitude types is measured in terms of Cohen's κ over the sets of attitude frames marked by both annotators. Average κ -agreement for attitude type annotations across the two studies is high: 0.80. The results of these studies support the hypothesis that attitudes and targets can be reliably annotated.

I also had hypothesized that defining intensity in terms of the attitude types would lead to more consistent intensity annotations between annotators. Unfortunately, the results of the agreement studies do not provide evidence in support of this hypothesis.