Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering

Claire Cardie, Janyce Wiebe, Theresa Wilson, and Diane Litman

Dept. of Computer Science
Cornell University
Ithaca, NY 14850, USA
cardie@cs.cornell.edu

Dept. of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260, USA
wiebe,twilson,litman@cs.pitt.edu

Abstract

While much recent progress has been made in research on fact-based question answering (Voorhees and Dice, 2000; Voorhees, 2001), our work aims to extend question-answering research in a different direction — to handle multi-perspective question-answering tasks, i.e. question-answering tasks that require an ability to find and organize opinions in text. In particular, this paper proposes an approach to multi-perspective question answering that views the task as one of *opinion-oriented information extraction*. We first describe an annotation scheme developed for the low-level representation of opinions, and note the results of interannotator agreement studies using the opinion-based annotation framework. Next, we propose the use of opinion-oriented "scenario templates" to act as a summary representation of the opinions expressed in a document, a set of documents, or an arbitrary text segment. Finally, we outline an approach for the automatic construction of opinion-based summary representations and describe how they might be used to support a variety of multi-perspective question answering tasks.

1 Introduction

Current research in question answering focuses primarily on the development of methods for finding short answers to factual questions like the following:

- Where is the Susquehanna River?
- When was Derek Jeter born?
- Who composed the Moonlight Sonata?

While much progress has been made in recent years in fact-based question answering (Voorhees and Dice, 2000; Voorhees, 2001), our work aims to extend question-answering (QA) methods to handle *multi-perspective question-answering* tasks: we are interested in developing techniques to support the answering of opinion-based questions of the following sort:

- Was the most recent presidential election in Zimbabwe regarded as a fair election?
- What is the general opinion from the African press towards the recent presidential election in Zimbabwe?
- What was the world-wide reaction to the recently released 2001 annual U.S. report on human rights?

- Has there been any change in the official opinion from China towards the 2001 annual U.S. report on human rights since its release?

It should be clear from these questions that the ability to find and organize opinions in on-line text is of critical importance in building successful multi-perspective question-answering (MPQA) systems. In response, this position paper proposes an approach to multi-perspective question answering that views the task as one of *opinion-oriented information extraction*. Very generally, an information extraction system takes as input an unrestricted text and "summarizes" the text with respect to a prespecified topic or domain of interest: it finds useful information about the domain and encodes that information in a structured form, suitable for populating databases (Lehnert and Sundheim, 1991; Sundheim, 1992; Chinchor *et al.*, 1993). An information extraction system in in the domain of natural disasters, for example, might extract for each disaster event in the input stream the type of disaster, the date and time that it occurred, any objects damaged or destroyed, their estimated monetary value, and the number of people injured or killed as a result of the natural disaster. Event-based information extraction systems typically operate by identifying low-level "template relations" throughout the text (e.g. that a "tornado" occurred on a particular day ("today"), at a particular time ("11:15a.m.") and location ("just outside of Dallas"); that "the twister" killed "5 people" and injured "20"), and then merging the template relations into the "scenario template" that summarizes the entire event (Cardie, 1997).

In contrast, we hypothesize that an information extraction system for MPQA tasks should rely on a set of opinion-oriented template relations that identify and categorize each expression of opinion in a text along with its source (i.e. the agent expressing the opinion), its type (e.g. positive, negative, uncertain), and its strength (e.g. strong, weak). Once identified, these low-level annotations can then be combined to create an opinion-based scenario template — a summary representation of the opinions expressed in a document, a group of documents, or an arbitrary text span. This summary representation, in turn, acts as the primary knowledge source that supports a variety of MPQA tasks.

In the sections that follow, we first describe the annotation scheme developed for the low-level representation of opinions and present the results of interannotator agreement studies using the opinion-based annotation framework (Section 2). We then motivate the need for summary representations of opinion and propose a method for their automatic construction that combines techniques from information extraction with a discourse-based approach to "perspective segmentation" (Section 3). We conclude (Section 4) with a discussion of the potential uses of summary representations of opinions for a variety of related MPQA task scenarios.

2 Low-level Opinion Annotations

The goal of the low-level annotation scheme (Wiebe *et al.*, 2002)[1] is to identify opinions, evaluations, emotions, and speculations in language. A general covering term for such states, from Quirk, Greenbaum, Leech, and Svartviks' (1985) *A Comprehensive Grammar of the English Language*, is *private state*, "a state that is not open to objective observation or verification." You can observe evidence of someone else being happy, for example, but you cannot directly observe their happiness.

Annotating private states. The annotation scheme focuses on two main ways that private states are expressed in language:

[1]The annotation scheme is built on work in linguistics and literary theory on *subjectivity* in language. For references please see (Fludernik, 1993; Wiebe, 1994; Stein and Wright, 1995).

- explicit mentions of private states and speech events, and
- expressive subjective elements.

An example of an explicit mention of a private state is "fears" in (1):

> **(1)** "The US **fears** a spill-over," **said** Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities.

An example of an explicit mention of a speech event is "said" in (1). By *speech event*, we mean an event of speaking or writing. These are important for our task, because people often express their private states when they speak or write.

An example of an *expressive subjective element* (Banfield, 1982) is "full of absurdities" in

> **(2)** "The report is **full of absurdities**", Xirao-Nima said.

With expressive subjective elements, sarcasm, emotion, evaluation, etc., are expressed through the way something is described, or though particular wording. Expressive subjective elements are often used by people to express their frustration, anger, wonder, negative evaluation, mirth, etc., without explicitly stating that they are frustrated, angry, etc. Sarcasm and irony often involve expressive subjective elements. Other examples of expressive subjective elements are: "what a", "jerk", and "!" in "what a jerk!"; "so-called" as in "so-called expert"; "far be it from me", as in "Far be it from me to suggest otherwise"; "how" and "wonderful" in "How wonderful"; and "absolutely" and "radiant" in "She was absolutely radiant."

Nested sources. Another important aspect of a speech event or private state is its SOURCE. The source of a speech event is the speaker or writer. The source of a private state is the experiencer of that state, that is, the person whose opinion, evaluation, etc. is being expressed. Obviously, the writer of an article is a source, because he wrote the sentences composing the article. But the writer may also write about other people's private states and speech events, so there may be multiple sources in a sentence. For example, each of the following sentences has two sources: the writer (because he wrote the sentences), and Sue (because she is the source of a speech event in (3) and of private states in (4) and (5), namely thinking and being afraid).

> **(3)** Sue said, "The election was fair."

> **(4)** Sue thinks that the election was fair.

> **(5)** Sue is afraid to go outside.

Note, however, that we really don't **know** what Sue says, thinks, or feels. All we know is what the writer tells us. Thus, (3), for example, does not directly present Sue's speech event but rather Sue's speech event according to the writer. Thus, we have a natural *nesting of sources* in a sentence.

Consider the following passage, with nested sources indicated on the speech-event and private-state terms:

```
(6) <source=writer>‘‘The US <source=writer, Xirao-Nima, US>fears</>
a spill-over,’’  <source=writer, Xirao-Nima>said</> Xirao-Nima,
a professor of foreign affairs at the Central University
for Nationalities.</>
```

The nested source of the "said" speech event in (6) is <writer, Xirao-Nima>: Xirao-Nima's words are presented, according to the writer. The nested source of the "fear" private state is <writer, Xirao-Nima, US>: it is a private state of the US, according to Xirao-Nima, according to the writer.

Expressive subjective elements may also have nested sources. For example, the evaluation expressed by “full of” and “absurdities” in (2) is attributed to Xirao-Nima, by the writer.

Attributes. Another important component of the annotation scheme is the ONLYFACTIVE attribute, which is included in the SOURCE annotation for the writer, and for the SOURCE of each explicit private-state or speech-event. The ONLYFACTIVE attribute indicates whether, according to that source, only objective facts are being presented (ONLYFACTIVE=YES), or whether some emotion, evaluation, etc., is being expressed (ONLYFACTIVE=NO). Note that a value of NO for ONLYFACTIVE does not mean the person does not believe what is presented, but that, whether or not they do, some emotion, evaluation, etc., of theirs is being expressed.

Consider sentence (6). The ONLYFACTIVE attribute for the <writer> is YES: at the <writer>’s level, it is objectively presented that somebody said something. The ONLYFACTIVE attribute for <writer, Xirao-Nima>’s speech event is also YES: it is simply presented as a fact (according to the writer, according to Xirao-Nima) that the US fears something. The ONLYFACTIVE attribute for <writer, Xirao-Nima, US>’s fear state is NO, since fearing is a private state.

There are other aspects of the proposed low-level annotation scheme, such as identifying TYPE of private state (e.g., POSITIVE or NEGATIVE); STRENGTH of private state; and annotators’ CERTAINTY in their judgments. These are described in detail in manual annotation guidelines (Wiebe *et al.*, 2002).

2.1 An Example

We end this section with a discussion of the low-level opinion annotations for a short sample of translated text from a Chinese newspaper:

> “It is heresy,” said Cao. “The ‘shouters’ claim they are bigger than Jesus.”

First consider the SOURCE=<WRITER> level, which encompasses both sentences in the short document. At the writer’s level, the text simply indicates that someone said something. As a result, the annotation for the writer includes only the ONLYFACTIVE attribute with a value of YES.

Now consider the text attributed to SOURCE = <WRITER, CAO>. There is an explicit speech event attributed to Cao via “said”. The ONLYFACTIVE attribute is NO, and the TYPE of private state is a NEGATIVE evaluation: there is negative evaluation expressed by Cao toward a claim of the shouters (according to the writer). There are expressive subjective elements attributed to <WRITER, CAO>: “heresy” and “bigger than” (in this context, “bigger than” is sarcastic, recalling phrases such as “bigger than the beatles”).

Finally, we have the nested SOURCE = <WRITER, CAO, THE SHOUTERS> in which the shouters’ claim or belief is presented (according to the writer, according to Cao). The ONLYFACTIVE attribute here is NO, lexically signaled by “claim”.

Preliminary annotation studies have been performed using the above low-level opinion annotation scheme with promising results. For the ONLYFACTIVE attribute, two trained annotators achieved a Kappa (Cohen, 1960) value of 0.80 for the 82% of the judgments for which both were certain using a corpus of 114 annotated documents. We are currently performing an extended interannotator agreement study.

In addition, we have performed several preliminary experiments to reproduce a subset of the proposed low-level opinion annotations: identification and categorization of opinion-based speech acts and expressive subjective elements. We trained naive Bayes and k-nearest-neighbor classifiers for these tasks using simple lexical features and syntactic features from a partial parser. Ten-fold cross-validation results showed improvements in F-measure for both algorithms over a simple (knowledge-based) baseline system (58.0 and 66.4, respectively vs. 56.7 for the baseline).

3 Summary Representation of Opinions

We expect the opinion-based annotation scheme described above to support a wide variety of end-to-end applications in multi-perspective question answering. For any particular MPQA application, however, we anticipate the need to go beyond the low-level annotations and have begun to investigate the creation of **summary representations of opinions** that would provide concise, and ultimately user-tailored summaries of the opinions expressed in an article, in a set of articles, or in any arbitrary segment of text.

As explained in the introduction, we propose to view these summary representations as information extraction (IE) scenario templates. Rather than traditional event-oriented IE scenario templates, ours will instead summarize opinion-oriented scenarios to accommodate the MPQA task. Nevertheless, we postulate that methods from information extraction will be adequate for the automatic creation of opinion-based summary representations. More specifically, machine learning methods can be applied to acquire syntactico-semantic extraction patterns for the identification of the low-level opinion annotations; and, once identified, the low-level expressions of opinion will then be merged to create the opinion-based scenario template summary representation. Cardie (1997) provides a survey of machine learning techniques for information extraction, which have become increasingly employed in state-of-the art IE systems (MUC-6, 1995; MUC-7, 1998).

In the subsections below, we first provide a concrete example of an MPQA summary representation, i.e. scenario template, for a portion of one article in the MPQA collection (section 3.1). We then describe a number of issues that must be addressed in the automatic creation of summary representations. We concentrate, in particular, on the role that "perspective segmentation" might play in that process (section 3.2). *Note that we have no empirical results to date for the automatic creation of summary representations; this portion of the paper represents work in progress.*

3.1 An Example

An MPQA scenario template is meant to encode a summary of the opinions expressed throughout one or more texts or text spans. They are "summaries" in that they merge and make inferences from the lower-level MPQA annotations that have been identified in the text.

As an example, consider the text in Figure 1, which is the first ten sentences of one document (#20.20.10-3414) from the Human Rights portion of an MPQA collection developed as part of an 8-week ARDA-sponsored summer research workshop on multi-perspective question answering held at Mitre this past summer of 2002. Given the first sentence of the document,

> The Annual Human Rights Report of the US State Department has been strongly criticized and condemned by many countries,

our MPQA system should produce the following low-level opinion annotations:

SOURCE=<WRITER>: ONLYFACTIVE=YES

SOURCE=<WRITER>: EXPRESSIVE-SUBJ, STRENGTH=MEDIUM.

In particular, from the writer's perspective, the sentence can be classified as ONLYFACTIVE=YES. In addition, the lexical cue "strongly" indicates some (MEDIUM) amount of EXPRESSIVE SUBJECTIVITY.

The Annual Human Rights Report of the US State Department has been strongly criticized and condemned by many countries. Though the report has been made public for 10 days, its contents, which are inaccurate and lacking good will, continue to be commented on by the world media.

Many countries in Asia, Europe, Africa, and Latin America have rejected the content of the US Human Rights Report, calling it a brazen distortion of the situation, a wrongful and illegitimate move, and an interference in the internal affairs of other countries.

Recently, the Information Office of the Chinese People's Congress released a report on human rights in the United States in 2001, criticizing violations of human rights there. The report quoting data from the Christian Science Monitor, points out that the murder rate in the United States is 5.5 per 100,000 people. In the United States, torture and pressure to confess crime is common. Many people have been sentenced to death for crime they did not commit as a result of an unjust legal system. More than 12 million children are living below the poverty line. According to the report, one American woman is beaten every 15 seconds. Evidence show that human rights violations in the United States have been ignored for many years.

Figure 1: MPQA Sample Text. First ten sentences from document #20.20.10-3414 from the Human Rights portion of the MPQA collection.

<writer>: onlyfactive <writer>: expr-subj (medium)

<writer>: neg-attitude (medium) → <report> <writer>: neg-attitude (medium)
<writer>: neg-attitude (medium)

<writer>: onlyfactive <writer, many-countries>: neg-attitude (medium)
→ <report> <writer, many-countries>: extreme
<writer, many-countries>: neg-attitude (high, high, medium)

<writer>: onlyfactive
<writer, info-office>: neg-attitude (medium) → <US>
<writer>: onlyfactive <writer, chinarep>: onlyfactive
<writer>: ?neg-attitude
(medium) → <US> <writer>: expr-subj (low) <writer>: neg-attitude (low)
→ <US> <writer>: expr-subj (low) <writer>: neg-attitude (medium)
<writer>: onlyfactive
<writer>: onlyfactive
<writer>: neg-attitude (low) → <US> <writer>: expr-subj (low)

Figure 2: Set of Lower-Level MPQA Annotations for the Text Sample from Document #20.20.10-3414.

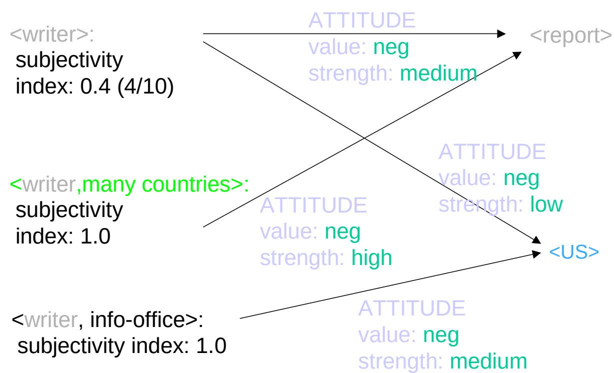

Figure 3: MPQA Summary Representation for the Text Sample from Document #20.20.10-3414.

A similar analysis of the remainder of the text fragment would produce the low-level annotations of Figure 2. It should be clear that the representation of opinions at this level is difficult for humans to absorb. It does, however, directly support the creation of a summary representation that provides the gist of the opinions expressed in the text. One possible MPQA summary representation for the sample text is shown in Figure 3. The summary makes it clear that there are three primary opinion-expressing SOURCES in the text — the <WRITER>; <WRITER, MANY COUNTRIES> (in Asia, Europe, Latin America, and Africa); and <WRITER, THE CHINESE INFORMATION OFFICE>. Furthermore, these sources expressed the following opinions:

- the writer expressed a negative attitude (of medium strength) towards the human rights report;
- the writer also expressed a mildly negative attitude towards the United States;
- according to the writer, many countries (in Asia, Europe, Latin America, and Africa) expressed a strongly negative attitude towards the human rights report; and
- according to the writer, the Chinese information office expressed a negative attitude (of medium strength) towards the United States.

Although expressed in graphical form in Figure 3, the summary representation is equivalent to the relational scenario template representation used in traditional IE systems.

As noted above, portions of the summary representation require making inferences across related groups of lower-level annotations. Associating a strength (low, medium, high) with each negative attitude is one such example. The *subjectivity index* associated with each nested agent is another example of the kind of summary statistic that one could generate from the lower-level annotations. It indicates, for example, that 4 out of 10 sentences of the writer include subjective language; and that all "utterances" associated with many countries and the Chinese information office include subjective content.

Like the lower-level MPQA annotations, the summary representation for a document, a set of documents, or one or more text fragments can be encoded as in-line annotations. This would allow for querying directly by the end-user.

Finally, there are many user-specified options for the level at which the MPQA summary representation could be generated. For example, the user might want summaries that focus only on particular agents, particular classes of agents, particular attitude types or attitude strengths. The user might also want to specify a particular level of nested source to include, e.g. create the summary from the point of view of only on the most nested sources.

3.2 Automatic Creation of Summary Representations of Opinion

In the paragraphs below, we discuss some of the issues involved in the automatic creation of MPQA summary representations.

Perfect lower-level annotations. Even given a complete and accurate set of the lower-level opinion annotations described in Section 2, building a summary representation will still be non-trivial. In particular, the MPQA system would need an accurate noun phrase coreference subsystem to identify the varying terms and phrases used to refer to each opinion SOURCE and each object towards which some opinion was expressed. Although our summer workshop annotated a subset of our MPQA corpus with coreference annotations, we did this only for SOURCES of opinions; there are no plans to include similar coreference annotations for objects: identifying the object of an *attitude-towards* relation is often very difficult for human readers to determine and it is often not explicitly expressed in the text.

Another likely source of error in creating summary representations from perfect low-level annotations would be cases where the text includes conflicting opinions from the same source.

Imperfect lower-level annotations. The situation becomes much harder, of course, when the MPQA summary representation is to be built on top of automatically generated lower-level annotations, which are likely to be incomplete and inaccurate. The situation will be akin to the information extraction task of "merging" extracted template relations into a scenario template. The noun phrase coreference system will continue to be important in this situation — it will need to provide accurate links between coreferent sources as well as coreferent objects.

On the other hand, since our goal is to derive a summary representation, we may be able to use redundancy in the expression of opinions to discard incorrect low-level annotations. For example, if the lower-level annotations associated with one source indicate that eight out of ten expressions from the source are negative evaluations of object X, then the system might be able to discard the remaining two positive evaluation of X as errors.

Cross-document coreference. In contrast to the TREC-style QA task, effective MPQA will require collation of information across documents since the summary representation may span multiple documents. For example, if a user wants to know the range of perspectives on topic X, then the system will need to perform cross-document coreference w.r.t. topic X as well as w.r.t. the various agents that express views on the topic.

Segmentation and perspective coherence. Accurate grouping of the low-level annotations — e.g., according to source, topic, negative or positive attitude — will be critical in deriving usable summary representations of opinion. For this task, we believe that it will be important to develop text segmentation methods based on a new notion of "perspective coherence". In the natural language processing literature, the term *segmentation* refers to breaking up a document into smaller chunks — or segments — that are locally coherent. Depending on factors such as corpus type and application need, different notions of coherence have been proposed as the basis of segmentation.[2]

In the area of information retrieval, for example, text segmentation has usually been based on semantic coherence. Segmentation is performed by placing segment boundaries at points of semantic discontinuity, which in turn are computed using measures such as lexical cohesion (Morris and Hirst, 1991; Hearst, 1997). In the area of discourse analysis, segmentation has instead been based on notions of informational (Hobbs, 1979; Mann and Thompson, 1988), and / or intentional coherence (Grosz and Sidner, 1986; Passonneau and Litman, 1993). Determining which sentences are informationally coherent has often been computed using formal methods of inference (e.g. abduction), or using discourse-level linguistic cohesive devices such as discourse markers and referring expressions. Although intentional coherence can likewise be computed using inference and / or linguistic clues, it is typically based on a goal-oriented view of natural language processing: sentences are coherent when they can be related to the same purpose.

To incorporate text segmentation to derive MPQA summary representations, we propose extending our opinion annotation scheme to denote "perspective segments" that will identify sentence spans expressing coherent perspectives. As with other notions of segmentation, perspective segmentation will likely involve merging and performing shallow inferences across sentences.

To motivate this idea, consider an example segment produced during an informal manual clustering study we performed as part of the summer workshop. For this study, workshop participants were asked to label opinions, where each opinion could be described by a single sentence, or by a segment consisting of a sentence span.

The excerpt in Figure 4 illustrates a sample segmentation from our coding exercise. In particular, four out of seven coders placed sentences 3-8 through 3-10 in the same segment; a fifth coder placed the beginning of this segment one sentence earlier. The (deep) annotations, which were produced separately from the clustering study, are also shown.

First, the segment consisting of sentences 3-8 through 3-10 seems to illustrate one potential way in which perspective coherence can be defined in terms of the sentence-level annotations: merge sentences into a segment when a single source (e.g. <w, us report>) is explicitly stating a sequence of opinions (e.g. OnlyFactive=no). Note that segment boundaries thus occur where the previous and following sentences are discontinuous with respect to this type of coherence. In our example, the sources in the sentences before and after the segment (that is, sentences 3-6 and 3-11) are different from the sources within the segment. As with other types of segmentation, linguistic phenomena such as the use of "his" in 3-11 to refer to Mugabe (who was most recently mentioned outside the segment) lends further support to such a segmentation analysis. While this example shows one way of abstracting over properties given our current sentence-level annotations, we believe that other abstractions will also be useful.

Note that sentence 3-7 provides an interesting borderline case, as one coder also included this sentence in the segment. First, there was no no explicit mention of a private state or speech act. Second, sentences 3-8 through 3-10 can be seen as providing evidence for the expressive-subjective element "chaos." We hypothesize that the treatment of sentences whose content can be related by particular types of "informational relationships" (as discussed above) might impact perspective segmentation. For example, a more sophisticated notion of perspective coherence might be to cluster evidence together (as with sentences 3-8 to 3-10), then include it with the sentence(s) expressing the opinion that the evidence supports (sentence 3-7).

In terms of aiding the creation of higher-level summary representations of opinions, an MPQA system might use segmentation information to ignore the presence of factive sentences

[2]While many theories of segmentation are hierarchical and involving structuring the segments, for ease of explanation, we will focus here on the simpler case of linear segmentation.

3-6 Mugabe described the opposition as ”donkey being controlled by the British,” the former colonial power. (source=<writer, Mugabe>, OnlyFactive=no)

SEGMENT BEGIN (1 CODER)

3-7 The fledgling MDC won 57 of 120 elected seats in June 2000 parliamentary elections as Mugabe’s popularity plunged amid economic devastation and chaos. (source=<writer>, OnlyFactive=no>)

SEGMENT BEGIN (4 CODERS)

3-8 The U.S. State Department released a human rights report on Zimbabwe Monday that accused the government of extra-judicial killings, undermining the independence of the judiciary and waging a “systematic campaign of violence targeting supporters and potential supporters of the opposition.” (source=<writer, US report>, OnlyFactive=no>)

3-9 Security forces tortured opponents, ruling party militants abducted people, and police arrested opposition supporters who were themselves the victims of crimes. **3-10** Freedom of the press and freedom of assembly were also severely restricted, the report said. (source=<writer, US report>, OnlyFactive=no)

SEGMENT END (5 CODERS)

3-11 In his speech on Monday, Mugabe thanked African leaders for refusing to buckle to pressure to suspend Zimbabwe from the Commonwealth of Britain and its former territories at a summit of the 54-nation grouping in Australia. (source=<writer, Mugabe>, OnlyFactive=no) (source=<writer, Mugabe, African leaders>, OnlyFactive=no)

Figure 4: An Example Document Excerpt with Human **Segmentations** and Annotations.

that are providing evidence for an opinion when trying to merge a sequence of opinionated sentences into a larger segment. An informal analysis of our data suggests that when evidence is treated the same way by coders, segment boundary agreement is about 60%. Alternatively, the MPQA might simply restrict its creation of summary representations only to segments that either fully or partly convey subjective information.

4 Uses for Summary Representations of Opinion in MPQA

Although a primary use of summary representations is to provide a short, optionally tailored summary of the opinions expressed in a specific text(s) or text fragment, we anticipate other uses for the MPQA summary representations.

Direct querying. When the summary representation is stored as a set of document annotations, it can be directly queried directly by the end-user using XML "grep" utilities.

Collective perspectives. The summary representations can be used to describe the collective perspective w.r.t. some issue or object presented in an individual article, or across a set of articles.

User-specified views. The summary representations can be tailored to match (some types of) user-specified constraints, e.g. to describe the perspective of a particular writer, individual, government, or news service w.r.t. a particular issue or object in an individual article, across a set of articles.

Perspective profiles. The MPQA summary representation would be the basis for creating a perspective "profile" for specific sources / agents, groups, news sources, etc. The profiles, in turn, would serve as the basis for detecting changes in the opinion of agents, groups, countries, etc. over time.

Debugging. Because the summary representation is more readable than the lower level annotations, summary representations can be used to aid debugging of the lower-level annotations on which they were based. This is the case whether the lower-level annotations were manually generated or automatically generated.

Gold Standard ”answer keys”. Creating the "gold standard" by which to evaluate most empirical NLP tasks is generally an intensely time-consuming endeavor. Consider, for example, the amount of effort required to create the scenario template "answer keys" for MUC-style information extraction evaluations. Once the gold standard for the lower-level annotations has been created for a collection, however, it might be possible to largely automate the creation of gold standards for various MPQA summary representations. These can then be used to evaluate summary representations created on top of automatically generated lower-level annotations.

Closer to true MPQA. The MPQA summary representations should let us get closer to true question-answering for multi-perspective questions. To handle TREC-style, short-answer questions, for example, a standard QA system strategy is to first map each natural language question into a question type (e.g. a ”who” question, a ”where” question, a ”why” question) so that the appropriate class of answer (e.g. a person, a place) can be located in the collection.

The MPQA summary representation acts as a question-answering template, defining the multi-perspective question types could be answered by our system.

5 Conclusions

This paper proposed an information-extraction approach to finding and organizing opinions in naturally occurring text as a means for supporting multi-perspective question answering. We first presented a low-level annotation scheme for representing opinion-based "template relations", i.e. localized and individual expressions of opinions. We next proposed one possible summary representation for concisely encoding the collection of opinions expressed in a document, a set of documents, or an arbitrary text segment. These are akin to the scenario templates produced by event-based information extraction systems. We hypothesize that information extraction techniques can be used to identify the low-level opinion annotations automatically, and that existing methods from information extraction for merging template relations can be extended via perspective segmentation for use in multi-perspective question answering. We concluded with a brief discussion of how opinion-based summary representations might be used to support a variety of multi-perspective question answering tasks.

References

(Banfield, 1982) Banfield, A. 1982. *Unspeakable Sentences*. Routledge and Kegan Paul, Boston.

(Cardie, 1997) Cardie, C. 1997. Empirical Methods in Information Extraction. *AI Magazine* 18(4):65–79.

(Chinchor *et al.*, 1993) Chinchor, N.; Hirschman, L.; and Lewis, D. 1993. Evaluating Message Understanding Systems: An Analysis of the Third Message Understanding Conference (MUC-3). *Computational Linguistics* 19(3):409–449.

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

(Fludernik, 1993) Fludernik, M. 1993. *The Fictions of Language and the Languages of Fiction*. Routledge, London.

(Grosz and Sidner, 1986) Grosz, B. and Sidner, C. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3):175–204.

(Hearst, 1997) Hearst, M. 1997. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics* 23(1):33–64.

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

(Lehnert and Sundheim, 1991) Lehnert, W. and Sundheim, B. 1991. A performance evaluation of text analysis technologies. *AI Magazine* 12(3):81–94.

(Mann and Thompson, 1988) Mann, W.C. and Thompson, S.A. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text* 8(3):243–281.

(Morris and Hirst, 1991) Morris, Jane and Hirst, Graeme 1991. Lexial cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics* 17(1):21–48.

(MUC-6, 1995) MUC-6, 1995. *Proceedings of the Sixth Message Understanding Conference (MUC-6).* Morgan Kaufmann, San Francisco, CA.

(MUC-7, 1998) MUC-7, 1998. *Proceedings of the Seventh Message Understanding Conference (MUC-7).* Morgan Kaufmann, San Francisco, CA.

(Passonneau and Litman, 1993) Passonneau, R. and Litman, D. 1993. Intention-based segmentation: Human reliability and correlation with linguistic clues. In *Proc. 31st Annual Meeting of the Association for Computational Linguistics (ACL-93)*. Association for Computational Linguistics. 148–155.

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

(Stein and Wright, 1995) Stein, D. and Wright, S., editors. *Subjectivity and Subjectivisation*. Cambridge University Press, Cambridge.

(Sundheim, 1992) Sundheim, B. 1992. Overview of the Fourth Message Understanding Evaluation and Conference. In *Proceedings of the Fourth Message Understanding Conference (MUC-4)*, San Mateo, CA. Morgan Kaufmann. 3–21.

(Voorhees and Dice, 2000) Voorhees, Ellen and Dice, D.M. 2000. The TREC-8 Question Answering Track Evaluation. In Voorhees, E. and Harman, D.K., editors, *Proceedings of the Eighth Text REtrieval Conference TREC 8*. NIST Special Publication 500-246.

(Voorhees, 2001) Voorhees, Ellen 2001. Overview of the TREC-9 Question Answering Track. In Voorhees, E., editor, *Proceedings of the Ninth Text REtrieval Conference TREC 9*. NIST Special Publication 500-249.

(Wiebe *et al.*, 2002) Wiebe, J.; Wilson, T.; Bruce, R.; Bell, M.; and Martin, M. 2002. Learning subjective language. Computer science technical report tr-02-100, University of Pittsburgh.

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