Introduction to Special Issue on Advances in Question Answering

James Pustejovsky and Janyce Wiebe

1. Introduction

This special issue on advanced topics in question answering is devoted to two critically important areas of annotation and reasoning in support of this task: recognizing opinions, speculations, and emotions in text; and identifying temporally sensitive expressions in text. This issue contains four articles addressing "these" two problems: *Annotating Expressions of Opinions and Emotions in Language* , by Janyce Wiebe, Theresa Wilson, and Claire Cardie; *Temporal and Event Information in Natural Language Text*, by James Pustejovsky, Robert Knippen, Jessica Littman, and Roser Sauri; *Temporal Closure in an Annotation Environment*, by Marc Verhagen; and *The Role of Inference in the Temporal annotation and Analysis of Text*, by Andrea Setzer, Robert Gaizauskas, and Mark Hepple. In this introduction, we situate this work in the larger context of question answering systems, and describe briefly the content of each article.

Research in the area of question answering has recently become one of the fastest growing and most challenging topics in computational linguistics and information access. It has been the topic of numerous workshops, seminars, and conferences over the past few years. Some of the recent activities in question answering include a TREC QA track, two phases of ARDA-funded research into question answering, i.e., AQUAINT, and three NRRC summer workshops in 2002 and 2003. It is the results of these latter workshops that are reported on, in part, in this special issue. These workshops focused on the temporal and multiple perspective aspects of question answering, and the difficulties of annotating complex relational information in natural language texts.

2. Annotating Opinions and Emotions

The purpose of the MPQA workshop (*Multiple Perspectives in Question Answering*) was to address a form of question answering that does not focus on finding facts, but rather on finding what people think, as evidenced by what is said in news reports around the world. This involves

the identification and organization of opinions in natural language text, to support information analysis of the following sort:

a. Given a particular topic, event, or issue, find a range of opinions being expressed about it in the world press.

b. Once opinions have been found, cluster them and their sources in different ways, including attitude, basis for beliefs, etc.

c. Construct perspective profiles of various groups and sources, and track attitudes over time.

Manual annotation of opinions is the focus of the first article in this issue, "Annotating Expressions of Opinions and Emotions in Language", by Janyce Wiebe, Theresa Wilson, and Claire Cardie. The high-level goal is to investigate the use of opinion and emotion in language (*subjective language*) through a corpus annotation study. They propose an annotation scheme that identifies key components and properties of opinions and other attitudes. Unlike many previous coarse-grained classifications of attitudes and opinions, Wiebe et al. propose a fine-grained annotation of the text, at the word and phrase-level. They introduce the concept of a *private state frame*, which represents the source of the state, the target content, and properties of the state including intensity, significance, and type of attitude.

The annotation scheme has been used to hand annotate a corpus of news articles, providing training and testing data for developing systems to extract opinions from natural language texts. While the article summarizes work using the corpus to develop automatic systems, the article focuses on the manual annotations. Numerous examples are given illustrating the broad range of the scheme. The annotation procedure is presented, and the results of inter-annotator agreement studies are given. The nature of the annotation scheme is such that a tremendous range of words and constituents are marked by the annotators; the annotators are not given fixed word lists to consider, but are asked to mark whatever expressions they interpret as expressing private states in the contexts in which they appear. The contextual nature of the annotations makes the annotated data valuable for studying ambiguities that arise with subjective language. The article provides a sampling of corpus-based observations that attest to the variety and ambiguity of language used to express opinions and emotions.

The annotated corpus is freely available. The corpus promises to be useful to researchers working in corpus-based explorations of subjective language and to encourage NLP researchers to experiment with subjective language in their applications.

3. Temporal and Event Expressions in Language

The purpose of the TERQAS workshop (*Temporal and Event Recognition for Question Answering Systems*) was to address the problem of how to answer temporally-based questions about the events and entities in text, specifically news articles.

For example, questions such as those shown below are currently not supported by question answering systems.

a. Is Gates currently CEO of Microsoft?

b. When did Iraq pull out of Kuwait during the first Gulf War?

c. Did the Enron merger with Dynegy take place?

What characterizes these questions as beyond the scope of existing systems is the following: they refer, respectively, to the temporal aspects of the properties of the entities being questioned, the relative ordering of events in the world, and events that are mentioned in news articles, but which have never occurred. The three articles in this issue on temporal annotation address four basic problems in event and temporal expression identification:

a. Time stamping of events (identifying an event and anchoring it in time);

b. Ordering events with respect to one another (lexical versus discourse properties of ordering);

c. Reasoning with contextually underspecified temporal expressions (temporal functions such as last week and two weeks before);

d. Reasoning about the persistence of events (how long does an event or the outcome of an event last).

The article "Temporal and Event Information in Natural Language Text", by James Pustejovsky, Robert Knippen, Jessica Littman, and Roser Sauri, discusses the contribution of corpus analysis towards the design and scope of a specification language for event predicates and temporal expressions in natural language text. The first requirement is an expressive language in which the kind of event and time information we are concernedwith can be made explicit. Next they consider more specifically the kinds of temporal information that might be needed for answering questions and how this information might be represented for use by a QA system. Since the construction of a knowledge base

for QA will involve marking up a document collection with some annotation language, the question addressed here is what such a markup language has to be like in order to use it to annotate documents for temporally sensitive question answering. They then go on to describe this language, TimeML, which is designed to account for the major features of temporal and event expressions in natural language. The study and development of an expressive language for events and their anchoring is a necessary prerequisite for evaluating different algorithms that compute the closure over events in a discourse. A gold standard corpus of 300 annotated articles, TIMEBANK, is described briefly as well.

The final two articles deal with the role of temporal closure in the annotation of event and temporal expressions within a document. Closure is the operation of applying axioms associated with a temporal model to the relations that have been annotated over a text. For example, transitivity of the ordering of *before* is one such axiom. This creates new temporal relations between the events and times in the document that were not explicitly marked up by an annotator or algorithm. Marc Verhagen's article, "Temporal Closure in an Annotation Environment", looks at the problems of embedding a temporal closure algorithm within a temporal annotation environment. A temporal closure component helps to create an annotation that is complete and consistent, but the effort needed to completely close a document can be quite high, and in most cases, tedious. Verhagen discusses a user-assisted mode of adding relations, where the user is asked to fill in temporal relations and the machine continues to add facts after each user-added relation, according to the closure axioms. He shows how this approach makes it possible to achieve a nearly complete annotation, where closure will derive about 95% of the temporal relations.

Finally, in their article "The Role of Inference in the Temporal annotation and Analysis of Text", Andrea Setzer, Robert Gaizauskas, and Mark Hepple also address the role that temporal closure plays in deriving complete and consistent temporal annotations of a text. First they discuss approaches to temporal annotation that have been taken in the literature, and then further motivate the need for a closed temporal representation of a document. No deep inferencing, they argue, can be performed over the events or times associated with a text without creating the hidden relations that are inherent in the text, but only surface after the application of a closure algorithm. They then address the problem of comparing diverse temporal annotations of the same text. This is much more difficult than comparing, for example, two annotations of part-of-speech tagging or named entity extent tagging; this is due to the *derived annotations* that are generated by closure,

making any comparison of temporal relations in a document a difficult task. They demonstrate that two articles cannot be compared without examining their full temporal content, which involves applying temporal closure over the entire document, relative to the events and temporal expressions in the text. Once this has been done, however, an inter-annotator scoring can be performed over the two annotations.

We believe that the articles in this issue will open up discussion in these two areas relating to question answering. We hope that they stimulate further work to support the goal of constructing intelligent question answering systems.