Guidelines for Annotating Temporal Information

Inderjeet Mani, George Wilson
The MITRE Corporation, W640
11493 Sunset Hills Road
Reston, Virginia 20190-5214, USA
+1-703-883-6149
imani@mitre.org

Lisa Ferro
The MITRE Corporation, K329
202 Burlington Road, Rte. 62
Bedford, MA 01730-1420, USA
+1-781-271-5875
lferro@mitre.org

Beth Sundheim
SPAWAR Systems Center, D44208
53140 Gatchell Road, Room 424B
Sand Diego, CA 92152-7420, USA
+1-619-553-4195
sundheim@spawar.navy.mil

ABSTRACT
This paper introduces a set of guidelines for annotating time expressions with a canonicalized representation of the times they refer to. Applications that can benefit from such an annotated corpus include information extraction (e.g., normalizing temporal references for database entry), question answering (answering "when" questions), summarization (temporally ordering information), machine translation (translating and normalizing temporal references), and information visualization (viewing event chronologies).

Keywords
Annotation, temporal information, semantics, ISO-8601.

1. INTRODUCTION
The processing of temporal information poses numerous challenges for NLP. Progress on these challenges may be accelerated through the use of corpus-based methods. This paper introduces a set of guidelines for annotating time expressions with a canonicalized representation of the times they refer to. Applications that can benefit from such an annotated corpus include information extraction (e.g., normalizing temporal references for database entry), question answering (answering "when" questions), summarization (temporally ordering information), machine translation (translating and normalizing temporal references), and information visualization (viewing event chronologies).

Our annotation scheme, described in detail in [Ferro et al. 2000], has several novel features:

- It goes well beyond the one used in the Message Understanding Conference [MUC7 1998], not only in terms of the range of expressions that are flagged, but, also, more importantly, in terms of representing and normalizing the time *values* that are communicated by the expressions.
- In addition to handling fully-specified time expressions [e.g., *September 3rd, 1997*), it also handles *context-dependent* expressions. This is significant because of the ubiquity of context-dependent time expressions; a recent corpus study [Mani and Wilson 2000] revealed that more than two-thirds of time expressions in print and broadcast news were context-dependent ones. The context can be local (within the same sentence), e.g., *In 1995, the months of June and July were devilishly hot*, or global (outside the sentence), e.g., *The hostages were beheaded that afternoon*. A subclass of these context-dependent expressions are 'indexical' expressions, which require knowing when the speaker is speaking to determine the intended time value, e.g., *now*, *today*, *yesterday*, *tomorrow*, *next Tuesday*, *two weeks ago*, etc.

Our scheme differs from the recent scheme of [Setzer and Gaizauskas 2000] in terms of our in-depth focus on representations for the values of specific classes of time expressions, and in the application of our scheme to a variety of different genres, including print news, broadcast news, and meeting scheduling dialogs.

The annotation scheme has been designed to meet the following criteria:

Simplicity with precision: We have tried to keep the scheme simple enough to be executed confidently by humans, and yet precise enough for use in various natural language processing tasks.

Naturalness: We assume that the annotation scheme should reflect those distinctions that a human could be expected to reliably annotate, rather than reflecting an artificially-defined smaller set of distinctions that automated systems might be expected to make. This means that some aspects of the annotation will be well beyond the reach of current systems.

Expressiveness: The guidelines require that one specify time values as fully as possible, within the bounds of what can be confidently inferred by annotators. The use of 'parameters' and the representation of 'granularity' (described below) are tools to help ensure this.

Reproducibility: In addition to leveraging the [ISO-8601 1997] format for representing time values, we have tried to ensure consistency among annotators by providing an example-based approach, with each guideline closely tied to specific examples.

While the representation accommodates both points and intervals, the guidelines are aimed at using the point representation to the extent possible, further helping enforce consistency.

The annotation process is decomposed into two steps: flagging a temporal expression in a document, and identifying the time value that the expression designates, or that the speaker intends for it to designate. The flagging of temporal expressions is restricted to those temporal expressions which contain a reserved time word used in a temporal sense, called a 'lexical trigger', which include words like *day*, *week*, *weekend*, *now*, *Monday*, *current*, *future*, etc.

2. SEMANTIC DISTINCTIONS

Three different kinds of time values are represented: points in time (answering the question "when?"), durations (answering "how long?"), and frequencies (answering "how often?").

Points in time are calendar dates and times-of-day, or a combination of both, e.g., *Monday 3 pm*, *Monday next week*, *a Friday*, *early Tuesday morning*, *the weekend*. These are all represented with values (the tag attribute VAL) in the ISO format, which allows for representation of date of the month, month of the year, day of the week, week of the year, and time of day, e.g.,

<TIMEX2 VAL="2000-11-29-T16:30">4:30 p.m. yesterday afternoon</TIMEX2>.

Durations also use the ISO format to represent a period of time. When only the period of time is known, the value is represented as a duration, e.g.,

<TIMEX2 VAL="P3D">a three-day</TIMEX2> visit.

Frequencies reference *sets* of time points rather than particular points. SET and GRANULARITY attributes are used for such expressions, with the PERIODICITY attribute being used for regularly recurring times, e.g., *<TIMEX2 VAL="XXXX-WXX-2" SET="YES" PERIODICITY="F1W" GRANULARITY="G1D">every Tuesday</TIMEX2>*. Here "F1W" means frequency of once a week, and the granularity "G1D" means the set members are counted in day-sized units.

The annotation scheme also addresses several semantic problems characteristic of temporal expressions:

Fuzzy boundaries. Expressions like *Saturday morning* and *Fall* are fuzzy in their intended value with respect to when the time period starts and ends; *the early 60's* is fuzzy as to which part of the 1960's is included. Our format for representing time values includes parameters such as FA (for *Fall*), EARLY (for *early*, etc.), PRESENT_REF (for *today*, *current*, etc.), among others. For example, we have *<TIMEX2 VAL="1990-SU">Summer of 1990</TIMEX2>*. Fuzziness in modifiers is also represented, e.g., *<TIMEX2 VAL="1990" MOD="BEFORE">more than a decade ago</TIMEX2>*. The intent here is that a given application may choose to assign specific values to these parameters if desired; the guidelines themselves don't dictate the specific values.

Non-Specificity. Our scheme directs the annotator to represent the values, where possible, of temporal expressions that do not indicate a specific time. These non-specific expressions include generics, which state a generalization or regularity of some kind, e.g., *<TIMEX2 VAL="XXXX-04" NON_SPECIFIC="YES">April</TIMEX2> is usually wet*, and non-specific indefinites, like *<TIMEX2 VAL="1999-06-XX" NON_SPECIFIC="YES" GRANULARITY="G1D">a sunny day in <TIMEX2 VAL="1999-06">June</TIMEX2></TIMEX2>*.

3. USEFULNESS

Based on the guidelines, we have annotated a small reference corpus, consisting of 35,000 words of newspaper text and 78,000 words of broadcast news [TDT2 1999]. Portions of this corpus were used to train and evaluate a time tagger with a reported F-measure of .83 [Mani and Wilson 2000]; the corpus has also been used to order events for summarization.

Others have used temporal annotation schemes for the much more constrained domain of meeting scheduling, e.g., [Wiebe et al. 1998], [Alexandersson et al. 1997], [Busemann et al. 1997]; our scheme has been applied to such domains as well. In particular, we have begun annotation of the 'Enthusiast' corpus of meeting scheduling dialogs used at CMU and by [Wiebe et al. 1998]. Only minor revisions to the guidelines' rules for tag extent have so far been required for these dialogs.

This annotation scheme is also being leveraged in the Automatic Content Extraction (ACE) program of the U.S. Department of Defense, whose focus is on extraction of time-dependent relations between pairs of 'entities' (persons, organizations, etc.).

Finally, initial feedback from Machine Translation system grammar writers [Levin, personal communication] indicates that the guidelines were found to be useful in extending an existing interlingua for machine translation.

4. CONCLUSION

The annotation scheme we have developed appears applicable to a wide variety of different genres of text. The semantic representation used is also highly language-independent. In Spring 2001, we will be embarking on a large-scale annotation effort using a merged corpus consisting of Enthusiast data as well as additional TDT2 data (inter-annotator agreement will also be measured then). An initial annotation exercise carried out on a sample of this merged corpus by 20 linguistics students using our guidelines has been encouraging, with 12 of the students following the guidelines in a satisfactory manner. In the future, we expect to extend this scheme to multilingual corpora.

5. ACKNOWLEDGMENTS

Our thanks to Lynn Carlson (Department of Defense), Lori Levin (Carnegie Mellon University), and Janyce Wiebe (University of Pittsburgh) for providing the Enthusiast corpus to us.

6. REFERENCES

[1] Alexandersson, J., Riethinger, N. and Maier, E. *Insights into the Dialogue Processing of VERBMOBIL*. Proceedings of the Fifth Conference on Applied Natural Language Processing, 1997, 33-40.

[2] Busemann, S., Decleck, T., Diagne, A. K., Dini, L., Klein, J. and Schmeier, S. *Natural Language Dialogue Service for Appointment Scheduling Agents*. Proceedings of the Fifth Conference on Applied Natural Language Processing, 1997, 25-32.

[3] Ferro, L., Mani, I., Sundheim, B., and Wilson, G. TIDES Temporal Annotation Guidelines. Draft Version 1.0. MITRE Technical Report MTR 00W0000094, October 2000.

[4] ISO-8601 ftp://ftp.qsl.net/pub/g1smd/8601v03.pdf 1997.

[5] Mani, I. and Wilson, G. *Robust Temporal Processing of News*, Proceedings of the ACL'2000 Conference, 3-6 October 2000, Hong Kong.

[6] MUC-7. Proceedings of the Seventh Message Understanding Conference, DARPA. 1998.

[7] Setzer, A. and Gaizauskas, R. *Annotating Events and Temporal Information in Newswire Texts*. Proceedings of the Second International Conference On Language Resources And Evaluation (LREC-2000), Athens, Greece, 31 May- 2 June 2000.

[8] TDT2 http://morph.ldc.upenn.edu/Catalog/LDC99T37.html 1999

[9] Wiebe, J. M., O'Hara, T. P., Ohrstrom-Sandgren, T. and McKeever, K. J. *An Empirical Approach to Temporal Reference Resolution*. Journal of Artificial Intelligence Research, 9, 1998, pp. 247-293.