A Multilingual Approach to Annotating and Extracting Temporal Information[1]

George Wilson
Inderjeet Mani[2]
The MITRE Corporation,
W640
11493 Sunset Hills Road
Reston, VA 20190-5214
USA
gwilson@mitre.org
imani@mitre.org

Beth Sundheim
SPAWAR Systems Center,
D44208, 53140 Gatchell Rd.
San Diego, CA 92152-7420
USA
sundheim@spawar.navy.mil

Lisa Ferro
The MITRE Corporation,
K329, 202 Burlington Road
Bedford, MA 01730-1420
USA
lferro@mitre.org

Abstract

This paper introduces a set of guidelines for annotating time expressions with a canonicalized representation of the times they refer to, and describes methods for extracting such time expressions from multiple languages.

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, and describes methods for extracting such time expressions from multiple languages. Applications that can benefit 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, including the following:

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.

[1] This work has been funded by DARPA's Translingual Information Detection, Extraction, and Summarization (TIDES) research program, under contract number DAA-B07-99-C-C201 and ARPA Order H049.
[2] Also at the Department of Linguistics, Georgetown University, Washington, DC 20037.

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 (based on the presence of specific *lexical trigger* words), 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 Interlingual Representation

2.1 Introduction

Although the guidelines were developed with detailed examples drawn from English (along with English-specific tokenization rules and guidelines for determining tag extent), the semantic representation we use is intended for use across languages. This will permit the development of temporal taggers for different languages trained using a common annotation scheme.

It will also allow for new methods for evaluating machine translation of temporal expressions at the level of interpretation as well as at the surface level. As discussed in (Hirschman et al. 2000), time expressions generally fall into the class of so-called *named entities*, which includes proper names and various kinds of numerical expressions. The translation of named entities is less variable stylistically than the translation of general text, and once predictable variations due to differences in transliteration, etc. are accounted for, the alignment of the machine-translated expressions with a reference translation produced by a human can readily be accomplished. A variant of the word-error metric used to evaluate the output of automatic speech transcription can then be applied to produce an accuracy score. In the case of our current work on temporal expressions, it will also be possible to use the normalized time values to participate in the alignment and scoring.

2.2 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="2000-11-29T16:30">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</TIMEX> is usually wet*, and non-specific indefinites, like *<TIMEX2 VAL="1999-06-XX" NON_SPECIFIC="YES" GRANULARITY="G1D">a sunny day in <TIMEX2 VAL="199906">June</TIMEX2> </TIMEX2>*.

3 Reference Corpus

Based on the guidelines, we have arranged for 6 subjects to annotate an English reference corpus, consisting of 32,000 words of a telephone dialog corpus – English translations of the 'Enthusiast' corpus of Spanish meeting scheduling dialogs used at CMU and by (Wiebe et al. 1998), 35,000 words of New York Times newspaper text and 120,000 words of broadcast news (TDT2 1999). This corpus will soon be made available to the research community.

4 Time Tagger System

4.1 Architecture

The tagging program takes in a document which has been tokenized into words and sentences and tagged for part-of-speech. The program passes each sentence first to a module that flags time expressions, and then to another module (SC) that resolves self-contained (i.e., '*absolute*') time expressions. Absolute expressions are typically processed through a lookup table that translates them into a point or period that can be described by the ISO standard.

The program then takes the entire document and passes it to a discourse processing module (DP) which resolves context-dependent (i.e., '*relative*') time expressions (indexicals as well as other expressions). The DP module tracks transitions in temporal focus, using syntactic clues and various other knowledge sources.

The module uses a notion of *Reference Time* to help resolve context-dependent expressions. Here, the *Reference Time* is the time a context-dependent expression is relative to. The reference time (italicized here) must either be described (as in "a week from *Wednesday*") or implied (as in "three days ago [from *today*]"). In our work, the reference time is assigned the value of either the *Temporal Focus* or the document (creation) date. The *Temporal Focus* is the time currently being

talked about in the narrative. The initial reference time is the document date.

4.2 Assignment of Time Values

We now discuss the assigning of values to identified time expressions. Times which are fully specified are tagged with their value, e.g, "June 1999" as 1999-06 by the SC module. The DP module uses an ordered sequence of rules to handle the context-dependent expressions. These cover the following cases:

- *Explicit offsets from reference time*: indexicals like "yesterday", "today", "tomorrow", "this afternoon", etc., are ambiguous between a specific and a non-specific reading. The specific use (distinguished from the generic one by machine learned rules discussed in (Mani and Wilson 2000)) gets assigned a value based on an offset from the reference time, but the generic use does not. For example, if "fall" is immediately preceded by "last" or "next", then "fall" is seasonal (97.3% accurate rule). If "fall" is followed 2 words after by a year expression, then "fall" is seasonal (86.3% accurate).
- *Positional offsets from reference time:* Expressions like "next month", "last year" and "this coming Thursday" use *lexical markers* (underlined) to describe the direction and magnitude of the offset from the reference time.
- *Implicit offsets based on verb tense*: Expressions like "Thursday" in "the action taken Thursday", or bare month names like "February" are passed to rules that try to determine the direction of the offset from the reference time, and the magnitude of the offset. The tense of a neighboring verb is used to decide what direction to look to resolve the expression.
- *Further use of lexical markers*: Other expressions lacking a value are examined for the nearby presence of a few additional markers, such as "since" and "until", that suggest the direction of the offset.
- *Nearby Dates*: If a direction from the reference time has not been determined, some dates, like "Feb. 14", and other expressions that indicate a particular date, like "Valentine's Day", may still be untagged because the year has not been determined. If the year can be chosen in a way that makes the date in question less than a month from the reference date, that year is chosen. Dates more than a month away are not assigned values by this rule.

4.3 Time Tagging Performance

The system performance on a test set of 221 articles from the print and broadcast news section of the reference corpus (the test set had a total of 78,171 words) is shown in Table 1[3]. Note that if the human said the tag had no value, and the system decided it had a value, this is treated as an error. A baseline of just tagging values of absolute, fully specified expressions (e.g., "January 31st, 1999") is shown for comparison in parentheses.

Type	Human Found Correct	System Found	System Correct	F-measure
TIMEX2	728	719	696	96.2
VAL	728	719	602 (234)	83.2 (32.3)

Table 1: Performance of Time Tagger (English)

5 Multilingual Tagging

The development of a tagging program for other languages closely parallels the process for English and reuses some of the code. Each language has its own set of lexical trigger words that signal a temporal expression. Many of these, e.g. day, week, etc., are simply translations of English words.

Often, there will be some additional triggers with no corresponding word in English. For example, some languages contain a single lexical item that would translate in English as "the day after tomorrow". For each language, the triggers and lexical markers must be identified.

As in the case of English, the SC module for a new language handles the case of absolute expressions, with the DP module

[3] The evaluated version of the system does not adjust the Reference Time for subsequent sentences.

handling the relative ones. It appears that in most languages, in the absence of other context, relative expressions with an implied reference time are relative to the present. Thus, tools built for one language that compute offsets from a base reference time will carry over to other languages.

As an example, we will briefly describe the changes that were needed to develop a Spanish module, given our English one. Most of the work involved pairing the Spanish surface forms with the already existing computations, e.g. we already computed "yesterday" as meaning "one day back from the reference point". This had to be attached to the new surface form "ayer". Because not all computers generate the required character encodings, we allowed expressions both with and without diacritical marks, e.g., mañana and manana.

Besides the surface forms, there are a few differences in conventions that had to be accounted for. Times are mostly stated using a 24-hour clock. Dates are usually written in the European form day/month/year rather than the US-English convention of month/day/year.

A difficulty arises because of the use of multiple calendric systems. While the Gregorian calendar is widely used for business across the world, holidays and other social events are often represented in terms of other calendars. For example, the month of Ramadan is a regularly recurring event in the Islamic calendar, but shifts around in the Gregorian[4].

Here are some examples of tagging of parallel text from Spanish and English with a common representation.

<TIMEX2 VAL="2001-04-01">hoy</TIMEX2>
<TIMEX2 VAL="2001-04-01">today</TIMEX2>

<TIMEX2 VAL="1999-03-13">el trece de marzo de 1999</TIMEX2>
<TIMEX2 VAL="1999-03-13">the thirteenth of March, 1999</TIMEX2>

<TIMEX2 VAL="2001-W12">la semana pasada</TIMEX2>
<TIMEX2 VAL="2001-W12">last week</TIMEX2>

6 Related Work

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. 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, our annotation of the Enthusiast corpus being an example.

7 Conclusion

In the future, we hope to extend our English annotation guidelines into a set of *multilingual* annotation guidelines, which would include *language-specific* supplements specifying examples, tokenization rules, and rules for determining tag extents. To support development of such guidelines, we expect to develop large keyword-in-context concordances, and would like to use the time-tagger system as a tool in that effort. Our approach would be (1) to run the tagger over the desired text corpora; (2) to run the concordance creation utility over the annotated version of the same corpora, using not only TIMEX2 tags but also lexical trigger words as input criteria; and (3) to partition the output of the creation utility into entries that are tagged as temporal expressions and entries that are not so tagged. We can then review the untagged entries to discover classes of cases that are not yet covered by the tagger (and hence, possibly not yet covered by the guidelines), and we can review the tagged entries to discover any spuriously tagged cases that may correspond to guidelines that need to be tightened up.

We also expect to create and distribute multilingual corpora annotated according to these guidelines. Initial feedback from machine translation system grammar writers (Levin 2000) indicates that the guidelines were found to

[4] Our annotation guidelines state that a holiday name is markable but should receive a value only when that value can be inferred from the context of the text, rather than from cultural and world knowledge.

be useful in extending an existing interlingua for machine translation. For the existing English annotations, we are currently carrying out inter-annotator agreement studies of the work of the 6 annotators.

References

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

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

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

L. Hirschman, F. Reeder, J. Burger, and K. Miller, *Name Translation as a Machine Translation Evaluation Task*. Proceedings of LREC'2000.

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

L. Levin. Personal Communication.

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

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

http://www.itl.nist.gov/iad/894.02/related_projects/muc/

A. Setzer and R. Gaizauskas. *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.

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

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

Appendix 1: Annotated Corpus: Enthusiast Dialog Example (one utterance)

Transcript of Spanish source:
EL LUNES DIECISIETE IMAGINO QUE QUIERE DECIR EL DIECISIETE TENGO UN SEMINARIO DESDE LAS DIEZ HASTA LAS CINCO
Annotated English translation:
<TIMEX2 VAL="2000-05-17">MONDAY THE SEVENTEENTH</TIMEX2> I IMAGINE YOU MEAN <TIMEX2 VAL="2000-05-17">THE SEVENTEENTH</TIMEX2> I HAVE A SEMINAR FROM <TIMEX2 VAL="2000-05-17T10">TEN </TIMEX2> UNTIL <TIMEX2 VAL="2000-05-17T17">FIVE
</TIMEX2>
Note: Elements of range expressions are tagged separately. The VAL includes date as well as time because of the larger context. The annotator has confidently inferred that the seminar is during the daytime, and has coded the time portion of the VAL accordingly.

Appendix 2: Annotated Corpus: New York Times Article (excerpt)

Dominique Strauss-Kahn, France's finance minister, said: "<TIMEX2 VAL="1999-01-01">Today</TIMEX2> is clearly <TIMEX2 NON_SPECIFIC="YES">a historic day for the European enterprise</TIMEX2>. Europe will be strong, stronger than in <TIMEX2 VAL="PAST_REF">the past</TIMEX2>, because it will speak with a single monetary voice."

But even on <TIMEX2 VAL="1998-12-31">Thursday </TIMEX2>, there were signs of potential battles ahead.

One hint came from Duisenberg, a former Dutch central banker who was named president of the European Central Bank only after a bitter political fight <TIMEX2 VAL="1998-05">last May</TIMEX2> between France and Germany. Duisenberg, a conservative on monetary policy, was favored by Helmut Kohl, who was <TIMEX2 VAL="1998-05">then</TIMEX2> chancellor of Germany. But President Jacques

Chirac of France insisted on the head of the Bank of France, Jean-Claude Trichet.

Germany and France eventually cut a deal under which Duisenberg would become president of the new European bank, but "voluntarily" agree to step down well ahead of <TIMEX2 VAL=”P8Y” MOD=”END”>the end of his eight-year term</TIMEX2>.

Many information needs have a temporal dimension as expressed by a temporal phrase contained in the userâ€™s query. Existing retrieval models, however, often do not provide satisfying results for such temporal information needs, as the following example demonstrates. Consider a sports journalist, interested in FIFA World Cup tournaments dur-ing the 1990s, who issues the query fa world cup 1990s.Â Referring to Figure 1(a), A Language Modeling Approach for Temporal Information Needs. 17.Â It annotates a given input document using the TimeML [13] markup language. Building on TARSQIâ€™s out-put, we extracted range temporal expressions such as â€œfrom 1999 until 2002â€, which TARSQI does not yet support. Temporal Information Extraction Extracting Events and Temporal Expressions. A Literature Survey. Naman Gupta 133050012.Â Heideltime HeidelTime (StrÂ¨otgen and Gertz, 2010) is a multilingual, cross-domain temporal tagger developed at the Database Systems Research Group at Heidelberg University. It extracts temporal expressions from documents and normalizes them according to the TIMEX3 annotation standard. HeidelTime distinguishes between news-style documents and narrative-style documents (e.g., Wikipedia articles) in all languages.Â It can be used to annotate documents with temporal information. It is a deterministic rule-based system designed for extensibility. @inproceedings{Wilson2005AMA, title={A Multilingual Approach to Annotating and Extracting Temporal Information}, author={George Wilson and I. Mani and B. Sundheim and L. Ferro}, booktitle={The Language of Time - A Reader}, year={2005} }. George Wilson, I. Mani, +1 author L. Ferro. Published in.Â This paper introduces a set of guidelines for annotating time expressions with a canonicalized representation of the times they refer to, and describes methods for extracting such time expressions from multiple languages. View on ACL. dl.acm.org. Temporal Interpretation, Discourse Relations, and Commonsense Entailment News Stories as Narratives Tense as Discourse Anaphor Tense Interpretation in the Context of Narrative An Empirical Approach to Temporal Reference Resolution Tense Trees as the â€˜Fine Structureâ€™ of Discourse Algorithms for Analysing the Temporal Structure of Discourse Part IV Temporal Annotation Introduction to Part IV A Multilingual Approach to Annotating and Extracting Temporal Information The. Annotation of Temporal Information in Natural Language Sentences Assigning Time-Stamps to Event-Clauses From Temporal...