

Annotating Temporal Information – From Theory to Practice

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

Laurie Gerber
Northrop Grumman
4065 Hancock Street
San Diego, California 92110, USA
+1-619-200-8344
lgerber@gerbersite.com

Beth Sundheim
SPAWAR Systems Center, 244207
53560 Hull Street
San Diego, CA 92152-5001, USA
+1-619-553-4145
sundheim@spawar.navy.mil

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

ABSTRACT

This paper expands on and updates a report on a set of guidelines for annotating time expressions [3]. Recent developments in the annotation guidelines, development of corpora, and annotation software are described, along with results on inter-annotator agreement in applying the tagging scheme.

Keywords

Annotation, annotation tools, temporal information, ISO-8601

1. INTRODUCTION

In [3][4][5], the authors outlined a set of annotation guidelines for creating normalized representations of temporal expressions in free text. This scheme, called TIMEX2, was originally developed under the DARPA TIDES Program. This paper reports on progress over the past year in three areas: guidelines, corpora, and annotation software. The developments reported here occurred both in preparation for a new tagging effort, and through the initial stages of applying the guidelines to two different genres of text.

2. BACKGROUND

In addition to the goal of marking the textual extent of a temporal expression, the TIMEX2 scheme has the goal of normalizing the VALUE of the expressions for use in applications such as question-answering, machine translation and summarization. The scheme also includes representations for sets of times and various kinds of vague temporal expressions. Where possible, the information recorded in the tags adheres to *ISO 8601: Information Interchange – Representation of Dates and Times* (1997) [6].

3. RECENT DEVELOPMENTS

Extended guidelines. The annotation scheme laid out in [5] has been extended to provide greater specificity in characterizing when events and relations hold true. The tagging conventions now allow certain expressions to be more precisely anchored on a timeline by normalizing them in relation to the narrative time or the timestamp of the document. One advantage of this extension is that the representation can now anchor a duration on a timeline. For example, the expression “the past three weeks” formerly had a normalized value of “P3W” (meaning “a period of 3 weeks”). It will now also contain the anchor time, ANCHOR_VAL=“2002-W02” (if the phrase were uttered during the second week of January 2002), as well as the relative direction of the period with the respect to the anchor, namely ANCHOR_DIR=“BEFORE”.

Fuzzy expressions like “past,” “present,” and “future” are similarly pinned down. For example, <TIMEX2 VAL=“PAST_REF” ANCHOR_VAL=“2002-01-07” ANCHOR_DIR=“BEFORE”>the past</TIMEX2> Formerly, this expression would only have VAL=“PAST_REF”.

The complete set of tag attributes is shown in Table 1.

New corpora. Two new annotated corpora have been prepared and a third is in preparation. The TIDES Temporal Corpora include the following:

- (1) a parallel corpus of 95 Spanish dialogs, their English translations, and temporal annotations of all the dialogs and their translations (44,081 words of raw text in all). This corpus is freely available at [9]. The raw Spanish dialogs are part of the Enthusiast corpus collected earlier at Carnegie-Mellon University under U.S. government funding.
- (2) 193 documents of the English TDT-2 corpus [8] with temporal annotations (171,535 words of raw text). This

corpus will be released in cooperation with the Linguistic Data Consortium.

There are 8243 TIMEX2 tags in these corpora.

An additional corpus is being annotated with TIMEX2 tags under the auspices of the Automatic Content Extraction (ACE) program's Relation Detection and Characterization (RDC) task. The corpus consists of 180,000 words of newswire, newspaper, and transcribed broadcast news texts. Information on that corpus is available at [1].

Table 1: Tag Attributes

Attribute	Function	Example
VAL	Contains a normalized form of the date/time.	VAL= "1964-10-16"
MOD	Captures temporal modifiers.	MOD= "APPROX"
SET	Identifies expressions denoting sets of times.	SET= "YES"
PERIODICITY	Captures the period between regularly recurring times.	PERIODICITY = "P1M"
GRANULARITY	Captures the unit of time denoted by each set member in a set of times.	GRANULARITY = "G3D"
ANCHOR_VAL	Contains a normalized form of an anchoring date/time.	ANCHOR_VAL= "1964-10-16"
ANCHOR_DIR	Captures the direction of offset from ANCHOR_VAL.	ANCHOR_DIR= "BEFORE"
NON_SPECIFIC	Identifies non-specific expressions.	NON_SPECIFIC= "YES"
COMMENT	Contains any comments the annotator wants to add.	COMMENT= "context garbled"

Temporal annotation software. A Temporal Attributes Entry window has been added to the Alembic Workbench text annotation tool, an open source product built and made available by The MITRE Corporation [2]. The special window, shown in Figure 1, is designed to increase annotator accuracy, consistency, and productivity over manual annotation. The window is currently being used in the preparation of the ACE RDC corpus mentioned

above. The data entry window is fully compliant with [5] but also flexible enough for any scheme that uses the ISO-8601 format.

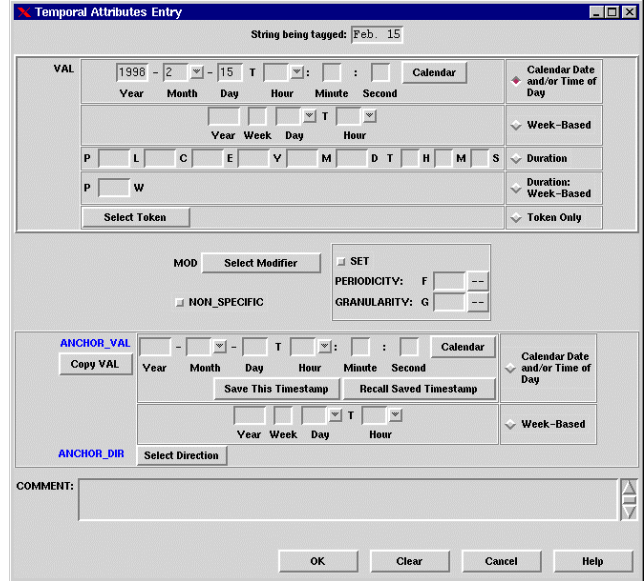


Figure 1: Alembic Workbench

Temporal Attributes Entry interface

As an example of its versatility, the data entry window is currently being used for two types of temporal tagging. The first marks explicit temporal entities referred to as "TIMEX2 markable." These include specific or fuzzy calendar and clock time references, as well as durations and sets.

The second use is in annotating the time attributes of relations. A relation here means a certain type of association between a pair of entities, such as PERSON AT LOCATION. The time attributes encode when a relation occurred or was true, based on temporal evidence in the text. For this second purpose, evidence may appear in the form of TIMEX2 markable expressions, expressions that contain TIMEX2 markable expressions (leading prepositions "in 2002" are not part of TIMEX2 markable expressions), and finite verb forms. The time attributes assigned to relations are in some cases simpler than the full TIMEX2 annotation; for example, relations are not tagged with the SET attribute. Data annotated with both TIMEX2 tags and time attributes on relations are being developed in support of the ACE program's RDC task.

The window's structured input fields and drop-down menus allow for rapid and accurate entry of tag attributes. The "Copy VAL" and "Save This Timestamp" buttons allow the annotator to reuse previously entered values. There is a calendar widget [10], shown in Figure 2, that allows the user to click directly on a day and have its value automatically converted to ISO format.

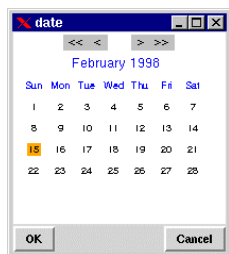


Figure 2: Temporal attributes calendar

A built-in error checking procedure also helps improve compliance with the annotation standard. The Alembic Workbench software is available for Windows (DOS), Unix, and Linux.

4. ROBUSTNESS

One of the stated goals for the system introduced in [3] was “reproducibility”: the system itself should promote consistency among annotators. To that end, a preliminary assessment was conducted on the TIDES Temporal Corpora mentioned earlier. The documents were annotated by six human annotators, using guidelines and instructions developed under the TIDES program [4][5]. The F-measures, showing the average inter-annotator agreement across five annotators (covering 182 TDT2 documents in all), are shown in Table 2.

Table 2: F-measures for inter-annotator accuracy for human annotators (TDT2)

	Human
TIMEX2	0.73 (17,604)
TIMEX2:GRANULARITY	0.35 (286)
TIMEX2:MOD	0.65 (393)
TIMEX2:NON_SPECIFIC	0.26 (160)
TIMEX2:PERIODICITY	0.90 (346)
TIMEX2:SET	0.67 (326)
TIMEX2:VAL	0.83 (13,037)

The two most important results are shown in the first and the last rows. The row labeled TIMEX2 represents performance on simply marking the expressions as time expressions. The scoring is strict; if the two annotators under comparison have not assigned identical extents to the expression, the scorer will generate at least one (and frequently two) errors. The average F-measure across all pairings of human annotators is 0.73. The last row, TIMEX2:VAL, represents performance on assigning normalized values to those expressions on whose extent the two humans under comparison had agreed. The average agreement across all scoring runs was 0.83. In each row, the number of instances which were compared are shown in parentheses.

The other rows show the results for secondary attributes of TIMEX2 tags. On the whole, we expected there to be less agreement on these attributes than on VAL and on TIMEX2 extent identification. However, it is worth noting that the number

of instances of these other attributes is much fewer compared to VAL and TIMEX2 extent.

While the 95 annotated dialog documents all had multiple human annotations, only 51 of the multiple-annotated ones have been converted so far from Word (the Alembic Workbench Temporal Attributes Entry was not yet available for the above annotation effort). While the agreement on dialog TIMEX2 extents is high (F-measure .84), the agreement on dialog VALs is much lower (F-measure of .51) than on TDT2 VALs. The reasons for this will be investigated, but we suspect they may include the relatively smaller number of dialog examples, as well as different assumptions about the dialog reference date (e.g., XXXX versus 1999).

It is expected that the Alembic Workbench Temporal Attributes Entry window will improve robustness. The window relieves the annotator of much of the mental burden required to adhere to the tag format. It also eliminates typographical errors.

5. FUTURE WORK

As the guidelines are refined through work on the ACE corpus, it will be important to address specific difficulties that humans could experience in applying the guidelines. A recent study on temporal annotation [7] notes that 30-40% of annotator errors may be attributed to ambiguity of the particular guidelines used or to mistakes in applying them. The issues that need to be resolved in our guidelines include:

1. *More consistent criteria for distinguishing markables from non-markables:* currently, modifiers like “former” are markable, while “previous” is not markable; expressions in idioms like “a man ahead of his time” are markable, while “in time” and “good evening” aren’t. Expressions headed by category words (“time”, “period”, etc.) are often vague or idiomatic and hard to normalize. They are currently markable depending on context.
2. *Improved representation for sets:* currently, sets of times can be represented as periods when they are without gaps (e.g., “the last 3 weeks” but not “the last 3 summers”), or as sets of regularly or irregularly occurring points (e.g., “every Thursday” or “Thursdays when the weather is good”), but sets that are bounded or quantified (e.g., “three summers in the 1990’s”) aren’t adequately represented.
3. *Coreference and Granularity:* the guidelines do not address the question of whether times expressed at different granularities should be considered strictly coreferential. Consider, as an example, “kidnapped almost two years ago...abducted in March 1996...captured on March 26, 1996”. It is unclear which, if any, of those expressions should be used to provide a value for a subsequent anaphor such as “then” (“Pol Pot, then the guerrilla group's leader”).
4. *More explicitness about fuzziness:* Fuzzy expressions like “six weeks ago” are normalized by counting back to six weeks prior to the current ISO week (likewise, “last night” is given the day of “yesterday”). A tag VAL like “2002-02” may arise from a precise date expression like “January 2002” or a fuzzy expression like “two months ago”. In the fuzzy cases, it would be useful to record the fact of the fuzziness somewhere in the tag.
5. *ISO-related limitations:* the representation doesn’t accommodate dates prior to the first century A.D., and fails

to normalize expressions based on alternative calendar systems.

Recently, the Advanced Research and Development Activity (ARDA) of the U.S. government, as part of the AQUAINT (Advanced Question and Answering for Intelligence) program, has launched a six-month research project called TERQAS (Time and Event Recognition for Question Answering Systems). The goal of this project is to outline a temporal annotation language called TenseML (for "Tense Markup Language"). The TIMEX2 annotation scheme is being leveraged as a starting point for that effort.

6. REFERENCES

- [1] Automatic Content Extraction (ACE) program URL: <http://www.nist.gov/speech/tests/ace/index.htm>
- [2] Alembic Workbench, available for download at <http://www.mitre.org/technology/alembic-workbench/>
- [3] Mani, I., Wilson, G., Sundheim, B., and Ferro, L. (2001). "Guidelines for Annotating Temporal Information ". In *Proceedings of HLT 2001, First International Conference on Human Language Technology Research*, J. Allan, ed., Morgan Kaufmann, San Francisco, 2001.
- [4] Ferro, L. (2001). "Instruction Manual for the Annotation of Temporal Expressions." MITRE Technical Report MTR 01W0000046. McLean, Virginia: The MITRE Corporation.
- [5] Ferro, L., Mani, I., Sundheim, B. and Wilson G. (2001). "TIDES Temporal Annotation Guidelines - Version 1.0.2". MITRE Technical Report MTR 01W0000041. McLean, Virginia: The MITRE Corporation.
- [6] ISO-8601 <ftp://ftp.qsl.net/pub/g1smd/8601v03.pdf> 1997.
- [7] Setzer, A. and Gaizauskas, R (2001). "A Pilot Study on Annotating Temporal Relations in Text." In *Proceedings of ACL Workshop on Temporal and Spatial Information Processing*. Toulouse, France. July 7, 2001.
- [8] TDT-2. Topic Detection and Tracking Program <http://morph ldc.upenn.edu/Catalog/LDC99T37.html> 1999.
- [9] Tides Temporal Corpus, Spanish and English dialogs. <http://www.nist.gov/speech/tests/ace/phase2/resource/index.htm>. See *tides-temporal-corpus.tar*.
- [10] The calendar widget illustrated in Figure 2 was written by Richard Suchenwirth and made available at <http://mini.net/tcl/1816.html>.

