Preliminary RDC Guidelines for Time Attributes

(Addendum to "Preliminary Guidelines for Relation Detection and Characterization (RDC)")

Version 1.2
1 December 2001

Beth Sundheim, SPAWAR Systems Center
(sundheim@spawar.navy.mil)[1]

1. Introduction

RDC records certain types of reported associations between pairs of entities. It also requires the reporting of the relation timestamps, i.e. of the time when a relation is reported as existing. For the first round of RDC, we take a conservative approach to associating times with relations, and require that the text evidence be found within the predication of a relation mention. Thus, we treat only Class A (explicit) relations, and only those that include explicit temporal evidence.

The text evidence of a timestamp may be in the form of a temporal phrase that is an adjunct to the predicate ("Bush's *Saturday* visit to Manhattan", "Rumsfeld became U.S. Secretary of Defense again *in 2001*"). A more general timestamp may be present in the form of a finite verb that heads the predication of the relation. A large proportion of Class A relation mentions will contain neither a time adjunct phrase nor a finite verb, and will therefore be assigned null temporal attribute values.

There may be more than one time at which a particular relationship is reported as having existed, i.e., there may be more than one timestamp on a relation. For example, a document may note that John Doe was LOCATED at Las Vegas June and August in 2001, or that Jane Doe had the ROLE of Technical Director for SPAWAR Systems Center in 1998 and 2000. The two different times referenced in each example are not evidence of different underlying relations; they are different attribute values for the same underlying relation.

2. Approach

Our approach to timestamping relations reflects our interest in extracting only highly reliable information and in posing a focused challenge to extraction systems. It also reflects our concern about achieving sufficient human interannotator agreement on the

---

[1] with significant contributions by Lisa Ferro (Mitre) and Laurie Gerber (Northrop Grumman IT)

task. The outlined approach requires recognition and interpretation of temporal expressions that appear in adjuncts modifying relation mentions, and recognition and contextual interpretation of finite verb forms that head the predication of relation mentions.

With respect to temporal expressions, the outlined approach applies previous work carried out under TIDES sponsorship. This work is documented in two MITRE technical reports dated June 2001: "TIDES Temporal Annotation Guidelines, Version 1.0.2" by Lisa Ferro et al. and "TIDES Instruction Manual for the Annotation of Temporal Expressions" by Lisa Ferro. These documents provide guidelines for the use of a time expression tag called TIMEX2.

ACE relations that can be pinned down to a single point or interval in time (on a clock, month-based calendar or week-based calendar) or to an explicit time duration will be timestamped in RDC in a way that builds from the TIMEX2 annotation. (See below re the RDC use of TIMEX2 annotations that do *not* have values.) In an effort to keep the ACE annotation effort relatively simple in the beginning, some of the more complex aspects of TIMEX2 annotation are not reflected in the target RDC output; in particular, the distinction between references to single dates/times versus references to *sets* of dates/times is not maintained in the RDC output, and explicit flagging of *non-specific* dates/times is not included in the RDC output.

3. RDC Temporal Attributes

3.1 General Information

Three temporal attributes are currently defined for RDC: TIMESTAMP_SOURCE, TIMESTAMP_VALUE and TIMESTAMP_ANCHOR. These represent the text form of a timestamp (in TIMESTAMP_SOURCE) and the corresponding normalization (in TIMESTAMP_VALUE and/or TIMESTAMP_ANCHOR), in accordance with the guidelines given in the following sections. These are attributes of a relation *mention*, not of a relation overall. In the future, we may consider defining temporal attributes for the overall relation as well; the purpose of such attributes would be to fuse the temporal evidence that is recorded at the mention level.

3.2 Guidelines for Deciding Whether to Generate One Timestamp, or Two

There are a few different situations where the annotator will encounter expressions that denote more than one time for a relation mention.

- Two TIMEX2s, two RDC timestamps: It is possible for there to be more than one timestamp on a particular relation mention, as when both a duration and point are given ("my last visit to Las Vegas was for *three hours* in *1968*").

- Two TIMEX2s, one RDC timestamp: When the endpoints of a range in time are given for a relation mention ("my last visit to Las Vegas was *1-3 September*"),

they are captured as independent TIMEX2 expressions, but for RDC, they are to be combined into a single, complex timestamp.

- One TIMEX2, one RDC timestamp: When there is an expression of a set of times modifying a relation mention ("my visits to Las Vegas *every month in 2000*"), these are captured as a single timestamp for RDC, and the fact that the expression denotes a set of times is ignored (for the time being).

3.3 Three Types of Timestamps

Some relation timestamps will come from temporal adjuncts, while others will come from tensed predicates. One of three different types of information is to be generated. The three types and the conditions for generating them are listed below, in order of preference.

1. **Specific, absolute time:** When a relation mention includes a temporal adjunct with a TIMEX2 expression that has a VAL attribute, that expression is to be used as the source of the relation timestamp.

2. **General, relative time:** When condition 1 does not apply, a general, relative time value is to be produced from the relation mention predicate, if headed by a finite verb. The relative time value indicates whether the relation is reported as existing at, before or after the time of the report.

3. **General temporal category:** If the first two conditions don't apply, and if there is a TIMEX2 adjunct to the mention that has no VAL attribute, an RDC time value that simply indicates the nature of the expression is to be produced; this value is either POINT or DURATION, depending on the interpretation of the TIMEX2 expression in the context of the relation mention.

Here is an example of each type of timestamp, with the extracted information encoded in "pseudo-RDC" format. The specifications for the RDC attributes are found in section 4.

1. Generation of RDC time attribute is licensed because the relation mention includes an adjunct containing a TIMEX2 expression that is encoded with a VAL.

I was spending *two weeks* on vacation there *that summer*.
(Context tells us that "that summer" is the summer of 1968. Note that this relation mention has two timestamps.)
Relation: Located (I, there)
Timestamp1 SOURCE: "two weeks"
Timestamp1 VALUE: P2W
Timestamp1 ANCHOR: nil
Timestamp2 SOURCE: "that summer"
Timestamp2 VALUE: 1968-SU

Timestamp2 ANCHOR: nil

2. Generation of RDC time attribute is licensed because the predication of the relation mention is headed by a finite verb.

The PDC's premises *are* located at 683 Sotomayor Street, in the downtown area
of the city.
(This example is from an article dated Sep 28, 1988.)
Relation: Located (The PDC's premises, 683 Sotomayor Street...)
Timestamp SOURCE: "are"
Timestamp VALUE: nil
Timestamp ANCHOR: AS_OF 1988_09_28

3. Generation of RDC time attribute is licensed because the relation mention includes an adjunct containing a TIMEX2 expression that is encoded *without* a VAL.

The petition is addressed to Mr Augusto Pinochet Ugarte -- appointed president of
the Republic for *another period of government* -- and it points out some
serious human rights violations, which must be explained.
Relation: Role (Mr Augusto Pinochet Ugarte, the Republic)
Role: appointed president
Timestamp SOURCE: "another period of government"
Timestamp VALUE: DURATION
Timestamp_ANCHOR: nil

4. Specifications for Filling Time Attributes

4.1 TIMESTAMP_SOURCE

For timestamp types 1 and 3, the fill for TIMESTAMP_SOURCE is a text string whose extent is the same as that of the corresponding TIMEX2 expression. For type 2 timestamps, the fill is a text string whose extent is the same as that of the tensed verb form in the corresponding predicate.

- In the case of types 1 and 3, note that the extent of TIMEX2-tagged expressions excludes the leading preposition that heads many time adjunct phrases ("*on* Saturday", "*for* two weeks"), and the specifications for TIMESTAMP_SOURCE follow that same practice. The *semantics* of some such prepositions is captured in TIMESTAMP_ANCHOR.

- In the case of type 2, the string fill documents the fact that the mention is headed by a finite verb, which is evidence that a time reference can reliably be generated. From the point of view of the end user of an RDC database, this string serves only as a link to the general text location where the evidence for the normalized form can be found. It is not intended to serve the user as direct and sole justification for that normalized form.

4.2 TIMESTAMP_VALUE

The fill for TIMESTAMP_VALUE is derived from TIMEX2 annotations. Thus, it is non-null for timestamp types 1 and 3, and it is null for type 2.

- For type 1 timestamps, TIMESTAMP_VALUE always contains the value of the TIMEX2 VAL attribute. If there is also a MOD attribute on the TIMEX2 expression, TIMESTAMP_VALUE will have a two-part fill in which the value of MOD is prepended to the value of VAL.

- Type 3 timestamps are based on TIMEX2 expressions that do not have an assigned VAL attribute. TIMESTAMP_VALUE in these cases is always a one-part attribute, with a set-fill value of either POINT or DURATION, depending on the interpretation of the TIMEX2 expression in the context of a relation mention.

4.3 TIMESTAMP_ANCHOR

The fill for TIMESTAMP_ANCHOR provides a reference (anchoring) time and a directionality of offset from that time. The attribute takes a two-part fill, with a set-fill to indicate directionality as the first part and a value in TIMEX2 VAL (ISO) format to represent the anchor as the second part. The directionality set-fill options are AS_OF, BEFORE and AFTER.

This attribute is non-null for all type 2 timestamps and for some type 1 timestamps.

- For type 1 timestamps, TIMESTAMP_ANCHOR is filled when anchoring information is provided either by corresponding attributes within a TIMEX2 annotation itself or by the context of the TIMEX2 expression in a relation mention.
    - ➢ It is provided by TIMEX2 annotations that include the attributes ANCHOR_VAL and ANCHOR_DIR.[3] To form the fill for TIMESTAMP_ANCHOR, the directionality value provided by ANCHOR_DIR is prepended to the anchoring value provided by ANCHOR_VAL.
        - Examples: "the past five days"
                     "at least the next year"
                     "now"
    - ➢ It is provided contextually when the TIMEX2 expression appears in a temporal prepositional phrase in the relation mention that designates a begin or end point. Only prepositions that convey a BEFORE or AFTER relationship are considered. The interpretation of the temporal adjunct as a whole in the context of the relation mention provides the directionality

---

[3] These will be documented in a future version of the TIMEX2 guidelines.

value. The anchoring value is the same as the TIMEX2 value of the VAL attribute.

        Examples: "before 5 pm"
                              "until 5 pm"
                              "from 5 pm on"

- For type 2 timestamps, the fill for TIMESTAMP_ANCHOR consists of a normalized form of the anchoring date/time (usually the document date/time) in TIMEX2 VAL (ISO) format, prepended by one of the three directionality values, according to the general time reference of the predicate of the relation mention.

- As mentioned earlier, the string that fills TIMESTAMP_SOURCE for type 2 timestamps is not intended to provide a complete explanation for the normalized form. Here are a few examples in which a present tense verb is not reflected as AS_OF in the normalized form, or a future tense form is not reflected as AFTER.

    (1) Annan *has* returned to New York City.
        Relation: Located (Annan, New York City)
        Timestamp SOURCE: "has"
        Timestamp VALUE: nil
        Timestamp ANCHOR: BEFORE 1988-09-28
    (2) Annan *is* returning to New York City.
        Relation: Located (Annan, New York City)
        Timestamp SOURCE: "is"
        Timestamp VALUE: nil
        Timestamp ANCHOR: AFTER 1988-09-28
    (3) He *is* expected to return to New York City in a good mood.
        Relation: Located (Annan, New York City)
        Timestamp SOURCE: "is"
        Timestamp VALUE: nil
        Timestamp ANCHOR: AFTER 1988-09-28
    (4) Annan *will* not return from Baghad until there is a truce.
        Relation: Located (Annan, Baghdad)
        Timestamp SOURCE: "will"
        Timestamp VALUE: nil
        Timestamp ANCHOR: AS_OF 1988-09-28

4.4 Summary Table of Attribute Specifications

A summary of the fill specification of each type of fill is shown in the table below, with examples in the lower half of each row. Some examples in sentence context are provided in the appendices, which are essentially working files that can be used in part as the basis for critiquing the current guidelines.[4]

---

[4] The appendices are currently contained in a separate file named "RDC-time-guides-appendices-v1". Warning: This file has not been updated to match the current version of the guidelines.

Type of Timestamp	TIMESTAMP_SOURCE	TIMESTAMP_VALUE	TIMESTAMP_ANCHOR
Type 1: Specific, absolute time	Text string: TIMEX2 expression	*Either* a one-part, normalized fill: value of TIMEX2 VAL *Or* a two-part fill: set-fill value of TIMEX2 MOD plus value of TIMEX2 VAL	Two-part fill: set fill (AS_OF, BEFORE or AFTER) plus normalized date/time
	September 26 of this year	2001-09-26	NIL
	2:40 p.m. Sept. 26, 2001	2001-09-26T14:40	NIL
	next week	2001-W40	NIL
	about two years ago	APPROX 1999	NIL
	two-week	P2W	NIL
	the past two weeks	P2W	BEFORE 2001-W39[5]
	less than a year ago	AFTER 2000	BEFORE 2001[6]
	currently	PRESENT_REF	AS_OF 2001-09-26T14:40[7]
	[before] *today*	2001-09-26	BEFORE 2001-09-26[8]
Type 2: General, relative time	Text string: finite verb	Not applicable	Two-part fill: set fill (AS_OF, BEFORE or AFTER) plus normalized date/time

---

[5] The week (in this case "W39") represents the current week. The interpretation of "the past two weeks" is "the two weeks before some point that's within the range of the current week". The information for TIMESTAMP_ANCHOR comes from the TIMEX2 annotation.

[6] The year (in this case "2001") represents the current year. The interpretation of "less than a year ago" is "a time that's within the range of 2000-2001". The representation is not powerful enough to enforce a 12-month constraint on the length of the interval. The information for TIMESTAMP_ANCHOR comes from the TIMEX2 annotation.

[7] The anchor value is the same as the reference time (document date/time or narrative date/time), and the granularity of the meaning of "currently" is the same as the granularity of the reference time. The information for TIMESTAMP_ANCHOR is obtained by converting the fill that's in TIMESTAMP_VALUE.

[8] See previous footnote. Note that the interpretation of "before today" is not accurate given the current representation. The interpretation here would be "before some point that's within the range of the current day", rather than completing excluding the current day. The inaccuracy of the current representation will be addressed in a later version of these guidelines. The information for TIMESTAMP_ANCHOR comes from the leading preposition "before", which is outside the extent of the TIMEX2 expression.

	is, as in "He is in Baghdad" (& where doc-time is given down to the second)	NIL	AS_OF 2001-09-26T14:40:30
	was, as in "He was in Baghdad" (& where doc-time is given down to the minute)	NIL	BEFORE 2001-09-26T14:40[9]
	will, as in "He will be in Baghdad" (& where doc-time is given down to the day)	NIL	AFTER 2001-09-26[10]
Type 3: General temporal category	Text string: TIMEX2 expression	Set fill: POINT or DURATION	Always NIL
	the appropriate time (as in "his appearance in Baghdad at the appropriate time")	POINT	NIL
	another period of government (as in "Pinochet, appointed president of the Republic for another period of government, ...")	DURATION	NIL

---

[9] See footnote above re inacurracy of interpretation.

[10] Ditto.

[11] If there are two TIMEX2s and they represent the endpoints of a range, convert those TIMEX2 VALs into a range representation.

[12] Ignore this spurious footnote.