MUC-7 Supplement to Time Treatment

These are working notes prepared by Elaine and extracted from e-mail correspondence between Lois and Beth concerning treatment of time expressions. It will cover three issues: (1) the structure of the proposed time template element, (2) the language of expression within it, and (3) the scope of its application. This serves as a supplement to “A Revised Template Description for Time (v3)” prepared by James Pustejovsky and Federica Busa for Tipster Phase I and MUC-5 and the Supplement to Time Treatment Used for MUC-5 Labor Negotiations scenario (time-guidelines-supplement-12apr95) prepared by Beth for MUC-6.

1. Structure

For MUC-5, the time object looked like this:

```
<TIME>:=
        AFTER:[DDMMYY]
        BEFORE:[DDMMYY]
        DURING:[DDMMYY]
        ON:     [DDMMYY]
```

where DDMMYY could include more descriptive expressions of time such as temporal modifiers (early = EA, end of = EO, etc.) and cyclic descriptions (fall = FA, 1st quarter = 1Q, 1st fiscal quarter = 1F). (This is based on “A Revised Template Description for Time (v3)” prepared by James Pustejovsky and Federica Busa.; see also Section 2 of this document.)

However, this structure caused undue scoring harshness by requiring that the time expression be in the right slot in order to be scored as correct. This was modified in the MUC-6 Dry Run to move the temporal relationship into the time slot as a modifier of the date expression:

DATE_TALK_1:“relationship date-expression”

For the Dry Run, the relationship candidates were changed slightly, as described in Beth's modifications:

> “1. AFTER, BEFORE, and ON may be used with date expressions at any level of granularity; use BETWEEN only when a date range cannot otherwise be expressed accurately.”

For MUC-7, we revise the definition for TIME to accommodate a standardization of the time object as an interval of time. All dates will be expressed as occurring within some interval. The year will be represented as a 4-digit year in the normalized form, rather than as a 2-digit year, as was done previously, so as to avoid the Year 2000 issue.

```
<TIME>:=
        START:[DDMMYYYY]
```

END: [DDMMYYYY]
DESCRIPTOR:“description within text”

The specification of Time Relations such as AFTER, BEFORE, ON, DURING, BETWEEN will no longer be required since those relations are inherent in the interpretation of the time interval.

Including the descriptor phrase within the time element allows us to punt on the normalization of some expressions that are too vague to calculate or that are beyond what we consider to be our scope. In these cases, the descriptive string itself can be extracted. For example, “every other day” could be represented as:
TIME: DESCRIPTOR: “every other day”.

The following conventions for FILL RULES will be followed:

1. “in the next 2 days” means that the START value is today, and the END value is today+2. The START and END values represent the longest possible duration of the interval, e.g., START 290687 starts just after midnight on the 29th and END 010787 ends at midnight on the 1st. The times need not be specified in the date expression

2. Weeks will not be calculated. It is usually UNCLEAR, what is meant by expressions like: next week, this week, last week, in two weeks.

3. UNCLEAR will be used if the START or END points are not specified fully or cannot be inferred from other time information. Using the UNCLEAR value will mitigate the problems that may occur if we require a full calculation of the interval. Time expressions are often vague. It would be dangerous to introduce more precision than is warranted. This will be used for such expressions as:

4. Interpretation of the template element may require calculation from the reference time of the message. Specific date information for relative time will be in no more detail than the time expression itself. For example, if the date of the article is 12/16/87,

“three months ago”, would be expressed as follows:
TIME: START: 091987
END: 091987
“3 days ago” would have the form (if DD= 12/16/87):
TIME: START: 13121987
END: 13121987
“Three years ago” would have the form if no additional date information were provided in the article.
TIME: START: 1984
END: 1984
On the other hand, something like “during December” could be fully specified, as could “during 1984":
TIME: START: 01121987
END: 31121987

If we're extracting information to be stored in a database, we might want to avoid this type of information. It would depend on the application, and the opinions of the database experts, I guess.

5. We will break with our convention of requiring most specific information in template elements, in that we will <u>not</u> extract time information on “time of day” for MUC-7, although it should be included in the DESCRIPTOR when available.[1] For example, “at 10:00am on December 15" would have the following form:

```
TIME:      START: 15121987
           END:   15121987
           DESCRIPTOR: “at 10:00am on December 15"
```

2. Language

The language for describing the temporal expressions in MUC-5 and the MUC-6 Dry Run is sufficient for MUC-7, assuming that the domain is not a military one, with Zulu time. The Pustejovsky and Busa document fully specifies the time expressions. (“A Revised Template Description for Time (v3)”). The only change in the DATE-EXP definition is that YY is now YYYY and there is an option available for UNCLEAR. We leave the abbreviations such as FY, EO, etc. for when it becomes impossible to derive a normalized form.

```
<TIME> :=
      START:       DATE-EXP
      END:         DATE-EXP
      DESCRIPTOR: "DESCRIPTOR" ^
      OBJ_STATUS: { OPTIONAL }
      COMMENT:    "COMMENT"

DATE-EXP ::([[01-31]]|{EA, MD, LT, EO, BO})[[01-12]][[1000-9999]YYYY]
                  | {EA, MD, LT, EO, BO}
                    {FA, WI, SP, SU, 1Q, 2Q, 3Q, 4Q, 1F, 2F, 3F, 4F, FY}
                    [[1000-9999]YYYY]
                  | {EA, MD, LT, EO, BO, FA, WI, SP, SU, 1Q, 2Q, 3Q, 4Q,
                       1F, 2F, 3F, 4F, FY}
                    [[1000-9999]YYYY]
                  | [[01-12]][[1000-9999]YYYY]
                  | [[1000-9999]]
                  | UNCLEAR
```

Given the interval structure of the template element, it is now possible to fully describe the interval. For example,
DURING 87would become

```
START:01011987
END:   12311987
DESC: “in 1987"
```

This representation would be applicable to all time expressions and would be more accurate than our previous models. The following are examples from Beth's time guidelines supplement of April 1995 for the MUC-6 Dry Run. These have been converted to the proposed form:

"yesterday" [where DD is 03/27/87"]

MUC6: ON 260387
MUC7: START:26031987
END: 26031987
DESC: "yesterday"

"next month" [where DD is "06/29/87"]

MUC6: ON 0787
MUC7: START:01071987
END: 31071987
DESC: "next month"

"in the fall" [where DD is "07/03/87"]

MUC6: ON FA87
MUC7: START:FA1987
END: FA1987

---

1. Making each time expression an interval allows for greater granularity of expression, allowing an extension to accommodate time of day as expressed in Coordinated Universal Time (UTC), previously known as Greenwich Meridian Time (GMT)UTC, also known as Zulu Time, would be useful in some military applications. Capturing the time interval would also allow downstream systems to do temporal reasoning, once the expressions are normalized sufficiently.For example, the following is an unclassified example provided for the War Breaker Message Handler prototype:

"THE 133RD ARMOR DIVISION CONDUCTED A FTX ON 8 JULY 93. THIS EXERCISE WAS HIGHLIGHTED BY A RIVER CROSSING CONDUCTED BY THE DIV'S 28TH TANK BATTALION AT 0900Z. ALSO, AT 1000Z, THE 3RD CO. 54TH ARMOR BN, SUPPORTED BY THE 12TH CO.18TH INF BN, CONDUCTED AN ASSAULT AGAINST NOTIONAL FORCES LOCATED AT...."

ftx: START:0000Z080793
END: 2400Z080793
DESC: "ON 8 JULY 93"

river crossing:START:0900Z080793
END: 2400Z080793
DESC: "AT 0900Z"

assault: START:1000Z080793
END: 2400Z080793
DESC: "AT 1000Z"

Note: This interpretation assumes the river crossing and assault are completed within the day's ftx activities.

DESC: “in the fall”

“in the next two days” [where DD is “06/29/87”]
MUC6: BEFORE 010787
MUC7: START:290687
END: 010787
DESC: “in the next two days”

“by the end of the year” [where DD is “06/29/87”]
MUC6: BEFORE EO87

MUC7: START:29061987
END: 31121987
DESC: “by the end of the year”

“sometime in April” [where DD is “12/16/87”]
MUC6: ON 0487
MUC7: START:01041987
END: 30041987
DESC: “sometime in April”

“sometime since April” [where DD is “12/16/87”]
MUC6: AFTER 0487 [between april and now]
MUC7: START: 01051987
END: 16121987
DESC: “sometime since April”

For full database functionality, a more complete expression of normalization would probably be necessary; however, that could be left for later MUCs.

3. Scope

The scope of the task will be expanded from that in MUC-6 dry run scenario. Specifically, calendar information about dates of days of the week will be used so that the time intervals can be described more fully.

The following is an excerpt also from Beth's time guidelines supplement for the MUC-6 Dry Run. It addresses limiting the scope of the task.

“2. If (a) a calendar (i.e., knowledge of more than just how many days
there are in a month) would have to be used to determine exact date or
(b) the date expression is too vague to be mapped directly onto a
normalized date, use “BEFORE|AFTER <date-derived-fm-DD-date>” as fill,
even though this normalization will lose a lot of the meaning:

a. Cases that would require a calendar

For MUC-7, we will use calendar information concerning days of the week (Sunday-Saturday) and their dates.

"[last] Saturday" [where DD is "04/27/87"]
MUC6: BEFORE 270487
MUC7: START: 25041987
END: 25041987
DESC: last Saturday

"the 30-day cooling off period ends 12:01 EDT Saturday" [where the date fill is supposed to represent the *start* of the
cooling-off period, and where DD is "04/08/87"]. The interval would end on the Saturday after 4/08/97 and begin thirty days before that.
MUC6: BEFORE 080487
MUC7: START: 060387
END: 110497

b. Vague date expressions.

We will use the same conventions as for MUC6, but will allow the value UNCLEAR as START or END if a specific date is not derivable from information in the article.

"[will continue...] this week" [where DD is "07/15/87"]
MUC6: AFTER 150787
MUC7: START 15071987
END: UNCLEAR
DESCRIPTOR: "this week"

"two weeks ago" [where the date fill is supposed to represent the
interval starting two weeks ago, and where DD is "08/28/87"]
MUC6: BEFORE 280887"
MUC7: START: UNCLEAR
END: 28021987

We will leave for future MUCs such issues as:

(a) how to represent temporal/aspectual information available from the verb form (in English), including information about the completion of an action or a process. This includes perfective and progressive information. For example, for MUC7, we will only use temporal information found in the prepositional phrase in "A river crossing has been underway since March 15, where DD=04/15/87":
MUC7: START: 15031987R
END: 15041987

(b) how to represent intermittent actions. e.g. "They cross the river every morning." If there is a specific reference point, for example “every January 25th” only the most recent will be specified:

MUC7: START: 25011987
END: 25011987

If the expression was “every January 28th since 1977", it would be:

TIME: START: 28011977
END: 28011987

(c) how to represent expressions that are not anchored to the dateline of the story.An example from the joint venture domain would be “profits for the first year.” Any type of expression that is not anchored to the dateline of the story, but must depend on a reference to some other event time, is of a higher level of difficulty.