

# TIDES Temporal Annotation Guidelines

## Version 1.0.2

**June 2001**

Lisa Ferro  
Inderjeet Mani  
Beth Sundheim  
George Wilson

**Sponsor:** DARPA  
**Dept. No.:** W150

**Contract No.:** DAA-B07-99-C-C201  
**Project No.:** 0701D070-1D

**Contact:** [gwilson@mitre.org](mailto:gwilson@mitre.org)

The views, opinions and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

©2000 The MITRE Corporation

**MITRE**  
Washington C3 Center  
McLean, Virginia

## TABLE OF CONTENTS

<b>1. INTRODUCTION .....</b>	<b>1</b>
<b>2. BASIC DISTINCTIONS AND TERMINOLOGY.....</b>	<b>3</b>
<b>3. RELATED WORK.....</b>	<b>6</b>
<b>4. ANNOTATION SCHEME.....</b>	<b>7</b>
4.1 INTRODUCTION .....	7
4.2 SOFTWARE TOOLS FOR ANNOTATION .....	8
4.3 BASIC ANNOTATION FEATURES .....	8
4.3.1 VAL attribute.....	8
4.3.2 MOD attribute.....	10
4.3.3 SET attribute .....	10
4.3.4 PERIODICITY attribute.....	10
4.3.5 GRANULARITY attribute.....	11
4.3.6 NON_SPECIFIC attribute .....	11
4.3.7 COMMENT attribute .....	11
4.4 ANNOTATION GUIDELINES PROPER.....	12
4.4.1 Principles .....	12
4.4.2 What Portions of the Document to Tag.....	12
4.4.3 Tag Extent.....	12
4.4.4 Embedded tags.....	14
4.4.5 Selection of Output Format.....	17
4.4.6 Lexical Triggers .....	21
<b>REFERENCES .....</b>	<b>23</b>
<b>GLOSSARY .....</b>	<b>25</b>
<b>APPENDIX A: USAGE OF ISO 8601 .....</b>	<b>28</b>
A1 APPLICABLE ISO FORMATS.....	28
A2 PLACEHOLDER CHARACTER .....	28
A3 ADDITIONAL DIFFERENCES FROM THE ISO STANDARD.....	29
<b>APPENDIX B: TOKEN LISTS FOR VAL ATTRIBUTE .....</b>	<b>30</b>
B1 INTRODUCTION .....	30
B2 TOKENS THAT OCCUPY THE ENTIRE VALUE OF VAL.....	30
B3 TOKENS THAT OCCUPY PARTICULAR POSITIONS IN THE VALUE OF VAL .....	31
<b>APPENDIX C: TOKEN LISTS FOR MOD ATTRIBUTE.....</b>	<b>35</b>
C1 INTRODUCTION .....	35
C2 TOKENS FOR MOD.....	35
<b>APPENDIX D: TOKENIZATION RULES.....</b>	<b>38</b>
D1 PUNCTUATION MARKS AND SPECIAL CHARACTERS EXCLUDED.....	38
D1.1 Periods (and ellipses) and Commas.....	38
D1.2 Hyphens or Dashes .....	38

D1.3 Apostrophes.....	38
D1.4 Parentheses and Quotes.....	38
D2 PUNCTUATION MARKS AND SPECIAL CHARACTERS INCLUDED.....	38
D2.1 Periods and Commas .....	38
D2.2 Slashes.....	38
D2.3 Hyphens.....	38
D2.4 Apostrophes.....	38
D2.5 Other Special Cases.....	39
<b>APPENDIX E: EXAMPLES OF EXPRESSIONS AND ANNOTATIONS.....</b>	<b>40</b>
E1 INTRODUCTION.....	40
E2 BASIC, ANCHORED EXPRESSIONS.....	40
E2.1 Expressions of calendar dates, with single value for VAL.....	40
E2.2 Expressions of times of day (TODs) in input .....	41
E2.3 Expressions that combine date and TOD in input and output.....	42
E2.4 Expressions involving units of weeks.....	42
E2.5 Expressions that are anchored, but vague.....	43
E2.6 Expressions of duration.....	44
E3 EXAMPLES WHERE EXPRESSIONS APPEAR IN CLOSE PROXIMITY, BUT DO NOT RESULT IN A SINGLE TAG OR IN NESTED TAGS .....	45
E3.1 Expressions of explicitly defined ranges.....	45
E3.2 Expressions with elements separated by a conjunction.....	46
E4 ANAPHORIC EXPRESSIONS .....	46
E5 EXPRESSIONS OF FREQUENCY.....	47
E5.1 Expressions that convey regular recurrence .....	47
E5.2 Expressions that do not necessarily imply regular recurrence.....	48
E6 EXPRESSIONS THAT INCLUDE EXPLICIT ANCHORS FOR OFFSETS .....	49
E6.1 Anchors that are markable time expressions.....	49
E6.2 Anchors that are not markable time expressions.....	50
E7 NON-SPECIFIC TIME EXPRESSIONS .....	50
E7.1 Expressions with placeholders in high-order position in value of VAL in output .....	50
E7.2 Expressions with placeholder in non-initial position of value of VAL in output.....	51
E7.3 Expressions without a value for VAL in output .....	52
E8 MISCELLANEOUS.....	53
E8.1 Expressions whose value cannot be adequately captured in the representation.....	53
E8.2 Expressions whose value for VAL is null or underspecified in the absence of outside evidence.....	54
E8.3 Expressions involving comparatives whose semantics cannot be captured in the representation .....	55
<b>APPENDIX F: EXAMPLES OF NON-MARKABLE EXPRESSIONS .....</b>	<b>56</b>
F1 INTRODUCTION.....	56
F2 POINT .....	56
F3 DURATION .....	56
F4 FREQUENCY .....	56
F5 PROPER NAMES THAT CONTAIN OR COMPRISE A TIME EXPRESSION .....	56

# 1. Introduction

During the 1990's, the field of computational linguistics achieved several practical advances through the construction of annotated corpora for developing and testing various technologies. This document describes an initial set of guidelines for annotating time expressions with a canonicalized representation of the times they refer to. This work has been carried out in support of a number of research activities under DARPA's Translingual Information Detection, Extraction, and Summarization (TIDES) research program. The research which can most directly benefit includes question answering (e.g, answering "when" questions), event characterization and tracking, visualization of events on timelines, and production of biographical summaries. Other research related to information extraction may also benefit.

This is a draft document, which is expected to undergo considerable revision based on feedback from other researchers, before it is officially published and before any large-scale annotation efforts are initiated. The guidelines specify more details of semantic representation than the TIMEX recognition tasks used in recent DARPA-sponsored evaluations (MUC7 1998), but are similar in that they treat the temporal expressions as stand-alone targets for annotation/extraction. These guidelines are intended to support a variety of downstream applications in the performance of some useful tasks; they are not intended to represent all the varieties of temporal information conveyed in natural language communication (the latter is a hopelessly ambitious goal, in our view). The guidelines are aimed at two sets of users:

1. Human annotators about to embark on the annotation of temporal expressions in order to construct corpora consisting of temporally annotated data for use by the NL community.
2. System developers who are building tagging programs to extract temporal information from documents.

To simplify matters, we will use the term **tagger**<sup>1</sup> to designate both a human annotator and a computer system; the two user classes (**annotator** and **system**) will be distinguished where needed. The guidelines are not written in the form of an annotator's instruction manual; however, such a manual is also available, see (Ferro 2001) . The guidelines may also be accompanied by software to score a tagger's output (called a **hypothesis annotation**) against a **reference annotation**, i.e., corpus annotated according to the guidelines.

The representation of the meaning of time expressions in natural languages is a very challenging problem. To emerge with a useful product, the guidelines are limited in scope to the following:

- While the TIDES program is highly multilingual, the language being addressed in this annotation effort in FY2000 is English. However, we expect that the representational scheme used will be relatively language-independent. In future work in FY2001, we will examine other languages.

---

<sup>1</sup> Terms which have a definition in the Glossary are in boldface. Examples in running text are shown in italics.

- 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 guidelines specify which of the flagged expressions will be given a time value.
- The flagging of temporal expressions is restricted to those temporal expressions which contain a reserved time word, called a **lexical trigger**, when the trigger is used in a temporal sense. For example, although *when* is a time word which can be used in a temporal sense, it is not in our set of lexical triggers; thus *when the stock market closed* will not be flagged. Likewise, *Black September*, when referring to a political group, contains a reserved time word, but that word isn't being used in a (basic) temporal sense, so the expression isn't to be flagged. Our set of lexical triggers is described later; it includes words like *day, week, weekend, now, Monday, current, future*, etc.
- The tagger is required to represent identified values in the prescribed format.

## 2. Basic Distinctions and Terminology

We now introduce some additional terminology related to annotation of time expressions. While a short glossary of terms with their definitions is provided in the Glossary section, here we situate these definitions in terms of a brief introduction to the temporal distinctions made by natural languages.

There are a wide variety of time expressions in natural language, including times of day (called **TOD**) such as *3 pm*, durations such as *3 hours long*, and dates such as *June 22* or *two weeks ago*. Time expressions display a number of interesting properties related to their meaning:

- **Indexicality:** Indexical time expressions require knowing when the speaker is speaking to help determine the intended time value. Examples include *now*, *today*, *yesterday*, *tomorrow*, *next Tuesday*, *two weeks ago*, *20 minutes after the next hour*, etc.
- **Granularity:** The intended value of a time expression may be relatively more or less precise. *Two weeks ago* may refer to a specific week (the **granularity of the expression**), or a specific day (an example of **more specific granularity**). The most precise granularity is one where the time format is fully specified.
- **Context-dependence:** The intended value of a time expression may require an inference based on information elsewhere in the sentence (**local context dependence**), or information in the surrounding discourse (**global context dependence**). Thus, in a news article, given the sentence *The hostages were seized on June 4<sup>th</sup>*, the value of *June 4<sup>th</sup>* may be further specified to a specific year based on global context of other information in the article. On the other hand, in the sentence *In 1995, the months of June and July were exceedingly hot*, the value of *June* and *July* can be resolved to a specific year based on local context. Often, the context comes from a prior reference to that particular time, e.g., *On Tuesday...that day*, where *that day* is identified with *Tuesday*. This form of prior reference is called **coreference**. Coreference can involve identical reference, as in the above example, or the reference can be to some time related to the prior reference, e.g., *Tuesday...that afternoon*, or *Thursday...the following Tuesday*. A time expression may depend on a prior event rather than a prior time expression for its value to be fully resolved, e.g., *the day after our meeting*; this is called an **event-dependent** time expression.
- **Fuzzy boundaries:** *Saturday morning* and *Fall* are fuzzy in their intended value with respect to when the time period starts and ends. Likewise, in *now that we walk upright, we have lots of back problems* (Perry 1999), *now* references a fuzzy interval of time extending from some period when human beings merged to the time of the speaker. Similarly, *today's youth* is fuzzy as to what part of the modern era is being considered, while *the early 60's* is fuzzy as to which part of the 1960's is included; *about three years ago* is fuzzy as to how close to three years ago the value is. Our prescribed format for representing time values includes **tokens**, such as FA (for *Fall*), EARLY (for *early*, etc.), PRESENT\_REF (for *today*, *current*, etc.). Our complete set of tokens is described in

Appendices B and C. The intent here is that a given application may choose to assign specific values to these tokens if desired; the guidelines themselves don't dictate the specific values.

- **Ambiguity:** Like many other expressions, time expressions can be **ambiguous**, e.g., *last year* (when uttered in the year 2000) could mean the year 1999, or the previous fiscal year (which the U.S. government in its infinite wisdom defines to last from October 1, 1998 to September 30, 1999), while *the last year* can mean the final element in a series of years previously referred to, or to an interval of 12 months that is coming to an end. Disambiguating a time expression involves figuring out which of the possible values was intended. In some cases, the tagger may not be able to resolve the ambiguity, in which case the value is left unspecified.

As mentioned earlier, the time expressions that are flagged here are those which have lexical triggers. In terms of assigning values, a time expression can be **anchored**, so that either it, or its start and end, can be positioned on a time line. For example, in *He arrived on the scene more than a year ago*, the expression *more than a year ago* may be positioned at some granularity on a timeline offset from NOW. Here the expression granularity of *year* will be used in the time value. Or else, the expression may be **unanchored**, where it has no connection to a time line, as in *The video is only half an hour long*. In this latter case, the duration of half an hour is represented in the value.

The timeline we use is based on the Gregorian calendar, and derives from the ISO 8601 (1998) format for representing time values (ISO8601 1998). This format is of the general form “YYYYMM[WW]DDhhmmss” (with many ways of expressing the individual fields, and ways of combining forms). This means that the **granularity** of an expression can have the values Year, Month, Week, Day, Hour, Minute, and Second, as well as Millennium, Century and Decade. Our modified ISO format represents unknown values for fields with the symbol “X”; thus, we can represent *April* in *April was a great month* even if we can't tell which year was intended. See Appendix A for an explanation of our use of the ISO 8601 spec.

This ISO format captures two basic ways of representing time: as a **point**, e.g., *three p.m.*, and as an **interval**. An interval can in general represent a **range** expression, e.g., *from three p.m. to six p.m.*, or a **duration** expression *three hours long*. A range expression is represented as two separate temporal expressions, each represented as a point. An anchored duration, such as *three hours long* in *The video started at 3 pm. It was three hours long, and filled with blood and gore*, is represented as a pair of points. An unanchored duration is represented using a special format.

Arguments can be made in favor of both point-based and interval-based representations for natural language. In some representations, e.g., (Bennett and Partee 1972), (Dowty 1979) in the formal semantics literature, points are treated as primitive, with intervals being defined in terms of them. In other representations, e.g., (Allen 1983), intervals are taken to be primitive, with points treated as very short intervals. Our approach is point-based to the extent possible, though intervals are also represented.

Temporal expressions in natural language also include **non-specific** expressions where no specific time is indicated. These expressions, which go somewhat beyond the purposes for which the ISO standard was developed, include as a subclass the generic expressions, like *April is*

*usually wet, Mondays suck!, or They report the traffic conditions 15 minutes after the hour.* These generic expressions are expressions which do not designate a specific time, stating instead a generalization or regularity of some kind associated with a class or property. There are also other non-specific expressions, e.g., indefinite non-specifics, like *He left on a sunny day in June.* These very subtle distinctions among non-specifics may be hard for a tagger to distinguish. Therefore, we require only that a tagger identify non-specific expressions, where either the value contains a placeholder, e.g., *a Tuesday in May,* or *April is usually wet,* or where there is no value but the granularity can be represented, e.g., *Election Day.*

### 3. Related Work

This work builds on the work of others. In particular, it can be viewed as picking up where the Message Understanding Conference (MUC7 1998) (Grishman and Sundheim 1996) left off, by extending the range of expressions which are flagged and by replacing the TYPE (DATE versus TIME) categorization attribute with a set of attributes to represent the actual time indicated by the expression. Our use of tokens is inspired by the work of (Pustejovsky and Busa 1998), though it does not follow their scheme, which was developed under the DARPA Tipster program in support of the event template-filling tasks defined for MUC-5. As indicated earlier, it takes full advantage of ISO 8601 in terms of the format used to represent time values. Since our annotation scheme takes into account characteristic properties of temporal expressions in natural language, namely indexicality, granularity, context-dependence, fuzziness of boundaries, and ambiguity, the ISO spec has had to be extended in various ways (see Appendix A for details).

Others have developed 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). More recently, (Setzer and Gaizauskas 2000) have independently developed an annotation scheme which represents both time values and finer-grained inter-event and event-time temporal relations. Our work focuses in depth on specific classes of time expressions, and is therefore much narrower in scope; it thus appears complementary to their effort.

Recently, the Stanford Knowledge Systems Lab (KSL-Time 1999) has defined a time ontology which treats time points (points on a real number line) as primitive. Intervals (which are sets of more than one point) have start and end time points which are located on the real number line. The location of a time point on the line can be specified under different time granularities, and precedence and equality based on time line location are defined for particular granularities, as well as across granularities when there isn't an overlap. Even though we do not, in our scheme, commit to all the representational distinctions made in the KSL ontology, we mention it because it will be obvious how our scheme can be mapped to this ontology. After all, making an annotation scheme compatible with an existing time ontology may help eventually support advanced inferential capabilities based on information extraction from text.

## 4. Annotation Scheme

### 4.1 Introduction

The annotation language used throughout is SGML. An SGML Data Type Definition (DTD) for these tags is being developed and will be made available.

As mentioned above, the annotation scheme incorporates a modified version of the ISO 8601 representation of time points and intervals. See Appendix A “Usage of ISO 8601” and Appendix B “Token Lists for VAL Attribute” for details of these modifications.

There are several basic principles which underly the annotation scheme:

- *Simplicity with precision:* Any annotation scheme should aim to be simple enough to be executed by humans, and yet precise enough for use in various natural language processing tasks. We have tried to meet both these requirements.
- *Human versus System:* We assume that the annotation scheme should reflect those distinctions that a human could be expected to reliably annotate, rather than just those things that a system could be expected to carry out. This means that some aspects of the annotation will be well beyond the reach of current tagging systems.
- *Detail:* The guidelines require that one specify values as fully as possible, within the bounds of what can be confidently inferred by annotators. In other words, the normalization of values shouldn't go beyond the meaning that could be inferred. The use of tokens and the representation of expression granularity are important tools to help ensure this.
- *Point-Based:* The guidelines are aimed at using the point representation to the extent possible, helping enforce consistency in tagging.
- *Reproducibility:* We have taken pains to design the annotation guidelines so that different annotators would find it easier to produce similar annotations of a given corpus. At this stage, we have not yet carried out an inter-annotator study to discover how well this goal is met. In addition to leveraging the ISO spec, we have tried to help ensure consistency by providing an example-based approach, with each guideline closely tied to specific examples to which the guideline pertains. We also indicate clearly which expressions are not to be flagged at all, i.e., those which lack lexical triggers. The guidelines also leverage a classification of time expressions based in part on the descriptive grammar of (Quirk and Greenbaum 1982).
- *Interpretation by systems:* We expect systems being developed based on these guidelines to be scored against corpora marked based on the same guidelines. Suggestions for scoring metrics are not included here. We also expect that there will be various stricter and looser scoring metrics, for example, metrics which provide partial credit in flagging and assigning values, or which exclude from the scoring tags which are hard for a system to tag. Also, a given application may want to further instantiate the values of tokens, and take correct token instantiation into account in the scoring.

It is worth stressing that this annotation scheme does not provide a compositional semantics specifying how the meanings of time expressions are combined with the meanings of other constituents. As mentioned earlier, a tagger can infer the value of time expressions based on local and global context. As a result, an indexical time expression like “Tuesday” will have a time value which represents a meaning derived from context, rather than its lexical meaning. Since the annotation scheme does not provide a syntactic analysis of each sentence, a meaning which arises from a time expression combining directly or indirectly with other constituents is represented inside the tag for the time expression rather than in a tag for a combined constituent. Thus, given “I always drink beer on Saturday”, “Saturday” would be given a value indicating that it denotes a set of times, though it is only by virtue of a constituent combining with “always” that this sense gets realized.

## 4.2 Software tools for annotation

- MITRE’s Alembic Workbench (Day et al. 1997) incorporates a suite of tools for the analysis of a corpus, along with the Alembic system to enable the automatic acquisition of domain-specific tagging heuristics. The Alembic Workbench makes it extremely easy to annotate textual data with fully customizable tagsets. A tagset to support these guidelines is also available. To download the (free) Workbench, go to: <http://www.mitre.org/technology/alembic-workbench/>. An on-line user manual in HTML is also available at the above address.
- MITRE’s TEMPEX time tagger is a highly preliminary software prototype used to tag a small subset of the time expressions described here, according to these guidelines. The system architecture, and performance measures of an earlier version are described in (Mani and Wilson 2000). Contact [gwilson@mitre.org](mailto:gwilson@mitre.org) to obtain a free copy.

## 4.3 Basic annotation features

The tag element used to represent time expressions is TIMEX2. This section outlines the tag attributes of the TIMEX2 tag elements. Appendices A, B and C give fuller information concerning the VAL and MOD attributes, and Appendix D provides tokenization rules. Appendix E “Examples of Expressions and Annotations” provides sentence-based examples of markable expressions, and Appendix F provides some examples of expressions that are time-related but not markable as TIMEX2 expressions.

### 4.3.1 VAL attribute

The VAL attribute is to be used for any expression that can be pinned down to a point (or interval) on a calendar/clock or that can be identified as an unanchored duration. Both local and global context can be used to determine the value, if any, of VAL. The granularity of VAL should reflect the granularity of the text expression (see *granularity of expression* in the Glossary).

The value of VAL is a normalized form of the contextually-determined interpretation of the time designated by the expression. Normalization is based on the formats defined by ISO 8601 (1998). Our representation extends and diverges from the standard in the ways that are described in Appendix A of this document.

A value may take one of the following basic forms:

1. A simple string based on the calendar/clock, e.g., “1999-10-01” for *October 1, 1999*, “1999-10-01-T09” for *9 a.m. Friday, October 1, 1999*. T is the ISO time-of-day designator (ISO 1997: 5.1.2).

Special notes:

- One or more components (especially the month and hour components) may appear in the form of two alpha characters, e.g., “1999-FA” for *Fall 1999*. See Appendix B.
  - It is also possible for the entire string to appear in alpha form as a special token indicating that the NL expression is essentially a lexicalized tense marker, e.g., “PRESENT\_REF” for *now*. See Appendix B.
  - A variant of the calendar/clock format is used for representing the value for expressions of weeks and days of weeks, if they cannot be confidently interpreted as expressions of months and days of months, e.g., “1999-W1” for *last week*, used in the context of a narrative anchored sometime in the second week of 1999. See Appendix A, “Usage of ISO 8601”.
2. A pair of simple strings based on the calendar/clock, with a slash separator. So far, this format has been used only rarely, as when a duration expression such *three-week [visit]* is supplemented through broader context with the actual dates of the visit (see information on duration format, below). The slash separator (sometimes called “solidus”) is an ISO notation for specifying time intervals (ISO 1997: 5.5).

Special notes:

- This interval form is used for representing anchored interpretations of duration expressions (see below).
  - It is not used for explicit range expressions, such as *[from] 1960 to 1969*; the NPs in such expressions are represented as two unrelated dates.
3. A simple string representing only a duration (e.g., “P3W” for *three-week*). P is the ISO notation for representing durations (ISO 1997: 5.5.3).

Special notes:

- If larger context provides both anchoring dates for the duration, the same text string will have two TIMEX2 tags, one representing the information as a duration, and another representing it as a calendar- or clock-based interval.

The determination of the appropriate form of the value may need to be made using evidence from leading prepositions or other terms that are outside the extent of the tagged temporal expression (see Section 4.4.3, “Tag Extent”). For example, a temporal expression introduced by the preposition *on* would probably be represented using the calendar-/clock-based format, while an expression introduced by *during* is a candidate for representation using the duration format.

A placeholder character, “X”, is used when higher-order portions of a value are unknown, or if the values of middle portions of a value are unknown, or if the numeric value of a duration is unknown. In a subset of cases of low precision of a time expression, the placeholder may appear in the right-hand portion of a value. See Appendix A for further explanation.

As mentioned in Appendix A, the calendar format optionally includes separator characters between ISO-defined components of the value. They are not necessary, but do improve readability.

See Appendix E for sentences containing example expressions that are represented using these three forms.

### 4.3.2 MOD attribute

The MOD attribute is used in conjunction with other attributes when the NL time expression includes a modifier that changes or clarifies the interpretation of the value of VAL in some way. In general, MOD captures the basic semantics of quantifier modifiers (e.g., *approximately*, *no more than*) and lexicalized aspect markers, e.g., *early*, *start [of]*.

MOD does not capture the semantics of leading prepositions or other terms that are outside the extent of the tagged temporal expression (see Section 4.4.3, “Tag Extent”).

Appendix C “Token Lists for MOD Attribute” contains the set list of possible values for MOD, along with a sampling of NL expressions that map to each value. Appendix E “Examples Of Expressions And Annotations” includes a number of examples that include MOD in their representation, such as “*about* three years ago,” “*more than* a decade ago,” “*the dawn of* 2000,” “the *early* 1960s”, “*late* last night.”

### 4.3.3 SET attribute

The SET attribute is used as part of the representation of expressions of sets of times, i.e., of times that recur regularly or irregularly. Examples of expressions of regular recurrence include “every Tuesday” and “each week.” Examples of expressions of irregular recurrence include “numerous Saturdays” and “some Thursdays.”

The unit of time denoted by the members of the set is represented using the GRANULARITY attribute (see below). The frequency of *regular* recurrence is expressed using the PERIODICITY attribute (see below). The number of regular or irregular recurrences is not captured in the representation; the SET attribute indicates only that there is at least one recurrence conveyed by the expression.

The only value for SET is “YES”. If the expression does not denote a set, the attribute is omitted from the annotation altogether.

See Appendix E5 “Expressions of Frequency” for examples of usage of SET, PERIODICITY and GRANULARITY.

### 4.3.4 PERIODICITY attribute

The PERIODICITY attribute applies only to sets of *regularly* recurring times. The recurrence can be expressed by words like “always” or “every” or “each” in the local context. Periodicity is used

only when a set of recurring times is further qualified by a day of the week (e.g., "every Friday") or month of the year ("each June"), or when an explicit periodicity is given, e.g., "daily", "monthly", "yearly".

The format of the value of the attribute is a string containing the character F in first position, followed by a number or the placeholder character X in second position, and terminated by an alpha character in final position. First position redundantly indicates that the expression conveys time *frequency* (periodicity); second position represents the interval value; and third position indicates the unit of the interval, where the character used to represent the unit is the same as would appear in expressions of duration. For example, both “every month” and “monthly” would have a periodicity of “F1M”.

#### **4.3.5 GRANULARITY attribute**

The GRANULARITY attribute applies to sets of regularly and irregularly recurring times. It represents the unit of time denoted by each set member, as indicated by the expression.

The format of the value of the attribute is similar to the format of the PERIODICITY attribute. It is a string containing the character G in first position, followed by a number or the placeholder character X in second position, and terminated by an alpha character in final position. First position redundantly indicates that the expression conveys the granularity of the set; second position represents the number of units at the expressed level of granularity; and third position indicates the granularity unit level itself, where the character used to represent the unit level is the same as would appear in expressions of duration. For example, “Fridays” would have a granularity of “G1D” (and, if the set is a set of regularly recurring times – e.g., if the expression is “every Friday” – a periodicity of “F1W”).

#### **4.3.6 NON\_SPECIFIC attribute**

The NON\_SPECIFIC attribute is used in the representation of several distinct types of time expressions:

- NPs used generically, as in “I love *December*.”
- Singular indefinite NPs that are used referentially, as in “The election took place on *a Tuesday*”.
- Other non-specific, non-generic uses of NPs, as in “I always vote on *Election Day*.”

The only value for NON\_SPECIFIC is “YES”. If the expression is specific rather than non-specific, the attribute is omitted from the annotation altogether.

See Appendix E7 for examples of non-specific expressions.

#### **4.3.7 COMMENT attribute**

The COMMENT attribute is intended to record annotator comments in reference annotations, but can also be used in hypothesis annotations. This attribute is not used in scoring. Comments should be as succinct and clear as possible. They must be enclosed in double quotes.

## 4.4 Annotation Guidelines Proper

### 4.4.1 Principles

There are two basic principles in applying temporal annotation tags. The first is as follows:

1. “If a human can determine a value for the temporal expression, it should be tagged.”

The derived value (VAL) can be a specific date and/or TOD, but an expression is also deemed taggable if the VAL can be expressed with one of the VAL tokens like FUTURE\_REF. The expression is also taggable if only a portion of VAL is known, as in certain **non-specific time expressions** like *a Tuesday*.

A second important principle is:

2. “VAL must be based on evidence internal to the document that is being annotated.”

Thus, there are tagged temporal expressions to which no VAL is assigned. An expression containing one of the key lexical triggers like *day* or *week* is a TIMEX2 expression even if VAL is not recoverable (see section 4.4.6 for a complete listing of lexical triggers). Also, although tagged as TIMEX2, we do not assign VALs to holidays, as these are based on world knowledge, and usually culturally dependent as well. See E.8.2 “Expressions whose value for VAL is null or underspecified in the absence of outside evidence” for further examples. (The intent here is that a given application may choose to assign specific values to holidays; but these guidelines won’t dictate the values.)

As mentioned earlier, the temporal annotation is not attempting to flag every expression relating to time or temporal sequence. Thus, many temporal modifiers like *lately*, *frequently*, and *subsequent* do not get annotated at all.

Finally, proper names are considered “atomic” for tagging purposes and should not contain TIMEX2 tags. For example, the terrorist group name *Black September* and the George Orwell novel entitled *1984* should not contain TIMEX2 tags.

### 4.4.2 What Portions of the Document to Tag

When applied to news stories such those found in the TDT2 (TDT2 1999) data, temporal annotation tags should never occur within SGML tags themselves. The basic portions that can be tagged are the headline and dateline of the story (if present), and the body of the story itself. In the TDT2 data, the following SGML tags delineate the bounds of markable portions of text:

CNN, ABC, PRI, VOA: <DATE\_TIME> , <TEXT>

NYT, APW: <DATE\_TIME> , <BODY> , <HEADLINE> , <TEXT>

### 4.4.3 Tag Extent

**Lexical criteria.** When tagged, the full extent of the tag must be one of the following grammatical categories:

- noun (including proper noun): “*today*”, “*Thursday*”
- noun phrase (NP): “*the morning*” “*Friday night*” “*the last two years*”

- adjective: *current*
- adverb: *recently*
- adjective/adverb phrase: “*half an hour long*” “*two weeks ago*” “*nearly half-hour*”

The full extent cannot be a prepositional phrase (i.e., the expression cannot start with a preposition) or a clause of any type (for example, the expression cannot start with a subordinating conjunction). Thus, the following are disallowed extents: “*before Thursday*,” “*in the morning*,” “*after the strike ended on Thursday*,” “*over the last 2 years*.”

**Syntactic criteria.** The full extent of the tagged expressions includes premodifiers of the time expression, including determiners, those premodifiers that qualify as MOD attributes, and also premodifiers with no corresponding MOD token. For example, all of the following represent markable temporal expressions:

*that cold day*  
*the next day*  
*late last night*  
*next summer*  
*recent decades*  
*numerous Saturdays*  
*more than a month*  
*no less than 60 days*  
*just a year ago*  
*its own future*

The full extent also includes all postmodifiers of the time expression, including prepositional phrases and dependent clauses. For example, all of the following represent markable temporal expressions:

*five days after he came back*  
*the future of our peoples*  
*nearly four decades of experience*  
*months of renewed hostility*  
*a historic day for the European enterprise*  
*the second-best quarter ever*

Note that the inclusion of pre- and post modifiers applies only to the modifiers of the time expression. The time expression may be included within a larger phrase whose head is not part of the time expression; this head **and the non-temporal modifiers** of this head are not included within the scope of the TIMEX2 tag. For example:

His partner is a 22-year-old *future* accountant who holds much promise.  
 His partner is a 22-year-old <TIMEX2 VAL="FUTURE\_REF">future</TIMEX2>  
 accountant who holds much promise

[The string “*accountant who holds much promise*” is not included in the TIMEX2 tag, and neither is the non-temporal premodifier of that string, namely “*a 22-year-old*”.]

There were doughnuts at the 8:00 meeting *this morning*.

There were doughnuts at the <TIMEX2 VAL="1999-07-15-T08:00">8:00</TIMEX2>  
meeting <TIMEX2 VAL="1999-07-15-TMO">this morning</TIMEX2>.

[Neither the determiner “*the*” before “8:00” nor the noun phrase “*this meeting*” after it is included within the scope of either TIMEX2 tag.]

Will you come to the *Saturday* picnic in the park?

Will you come to the <TIMEX2 VAL="1999-07-17">Saturday</TIMEX2> picnic in the  
park?

[The string “*picnic in the park*” is not included within the scope of the TIMEX2 tag, and neither is the non-temporal premodifier of that string, namely the determiner “*the*”.]

**Handling of punctuation and special characters.** See Appendix D, Tokenization Rules, concerning the exclusion or inclusion of punctuation and special characters in tag extent.

**Appositives.** An appositive to a temporal expression is excluded from the expression’s tag. If the appositive phrase itself contains a temporal expression, it is tagged separately.

I’m a creature of the 1960s, *the days of free love*.

I’m a creature of <TIMEX2 VAL="196">the 1960s</TIMEX2>, <TIMEX2 VAL="196">the  
days of free love</TIMEX2>.

**Range Expressions and Conjoined Expressions.** If the text of a temporal range expression has explicit begin and end points, then it is considered a range expression, and the points are tagged separately, even if they share modifiers. For example:

She served as Canada’s ambassador to the U.N. from 1992 through 1995.

She served as Canada’s ambassador to the U.N. from <TIMEX2  
VAL="1992">1992</TIMEX2> through <TIMEX2 VAL="1995">1995</TIMEX2>.

The prime minister's visit is to run *August 6-8*.

The prime minister's visit is to run <TIMEX2 VAL="1999-08-06">August  
6</TIMEX2>-<TIMEX2 VAL="1999-08-08">8</TIMEX2>.

Phrases involving conjunction or disjunction of time expressions are handled similarly. For example, “between *now* and *Monday morning*”, “*six months* or *a year from now*”. See E3 for more such examples.

#### **4.4.4 Embedded tags**

This section contains guidelines for determining whether a text segment that contains more than one hierarchically-related or offset-related temporal element should be assigned one tag or two tags. In the cases where two tags are expected, these guidelines also define the conditions for generating embedded versus non-embedded tags.

**When to Create One Tag.** If there is no intervening token between temporal terms that express values for a single unit of time, e.g., *twelve o’clock midnight* (where both *twelve o’clock* and *midnight* express values for a clock time (TOD) unit) or that express values for units that are hierarchically related, e.g., *Friday evening* (where the day unit is larger than the part-of-day unit)

or *November 1943* (where the month unit is smaller than the year unit), the expression is treated as an indivisible syntactic unit.<sup>2</sup> Thus, the following expressions represent the full extent of single TIMEX2 tags:

*twelve o'clock midnight*  
*Friday evening*  
*8:00 p.m. Friday*  
*Tuesday the 18th*  
*November 1943*  
*Fall 1998*

Note that the familiar “month day, year” format contains a comma, but is considered a single TIMEX2 expression. A comma between TOD and Calendar Date expressions is also ignored. Thus, the following are single TIMEX2 expressions:

*twelve o'clock January 3, 1984*  
*9 a.m. Friday, October 1, 1999*

While prepositions introduce syntactically embedded phrases and therefore typically introduce an embedded TIMEX2 tag, there are three notable exceptions:

1. The preposition “of”. Thus, the following are treated as single expressions:

*the second of December*  
*October of 1963*  
*summer of 1964*  
*the morning of January 31*  
*ten of two*

2. Prepositions like “to, till, after” etc. used in expressing TODs. Thus, the following are treated as single expressions:

*ten minutes to three*  
*five till eight*  
*twenty after twelve*  
*half past noon*

3. The preposition “in” in *time* expressions. Thus, the following is treated as a single expression:

*eleven in the morning*

---

<sup>2</sup> This guideline does not apply to cases where contiguous time expressions play different syntactic roles in the sentence. For example, in the sentence, “We celebrated *New Year's Eve Thursday* in Times Square.”, the expression “New Year's Eve” is the logical object of the predicate “celebrated”, i.e., it expresses *what* was celebrated, while “Thursday” is a temporal adjunct, expressing *when* the celebration took place.

**When to Create Multiple Tags, with Embedding.** There are two situations in which the syntactic dependence of a phrase that contains a temporal expression on another such phrase is reflected directly in the embedding of one TIMEX2 tag in another.

1. If a text segment is a temporal expression that includes an explicit offset (i.e., if it explicitly expresses a temporal sequence), and if the anchoring phrase is itself a time expression, two tags are created, and the tag on the anchoring phrase is contained within the extent of the tag of the complete phrase. The value of the complete phrase is computed in relation to the value of the anchoring phrase.

I'm leaving on vacation *two weeks from next Tuesday*.

I'm leaving on vacation <TIMEX2 VAL="1999-08-03">two weeks from <TIMEX2 VAL="1999-07-20">next Tuesday</TIMEX2></TIMEX2>.

[We assume here that the discourse has already been anchored at a date in mid-July 1999.]

A major earthquake struck Los Angeles *three years ago today*.

A major earthquake struck Los Angeles <TIMEX2 VAL="1993-07-15">three years ago <TIMEX2 VAL="1996-07-15">today</TIMEX2></TIMEX2>.

[Note that an elided *from* is assumed in the analysis of the syntax that is reflected by this representation.]

2. In a possessive construction, if both the possessive phrase and the phrase that it modifies are temporal expressions, then two tags are created, and the tag on the possessive phrase is contained within the extent of the tag of the complete phrase.

*This year's summer* was unusually hot.

<TIMEX2 VAL="1999-SU"><TIMEX2 VAL="1999">This year</TIMEX2>'s summer</TIMEX2> was unusually hot.

**When to Create Multiple Tags, without Embedding.** In cases other than those described above, temporal phrases that appear in close proximity are tagged as independent phrases. (Note that this guideline also covers appositives, range expressions, and conjoined expressions, which are discussed in relation to tag extent in 4.4.3.) Although they are tagged independently in terms of the extent, there is a dependency in terms of the value. As shown in the following examples, the expression with finer granularity "inherits" the value of the coarser-grained expression. This inheritance happens regardless of the relative ordering of the two expressions.

I tutored an English student *some Thursdays* in 1998.

I tutored an English student <TIMEX2 VAL="1998-WXX-4" SET="YES" GRANULARITY="G1D">some Thursdays</TIMEX2> in <TIMEX2 VAL="1998">1998</TIMEX2>.

The concert is at 8:00 *p.m.* on *Friday*.

The concert is at <TIMEX2 VAL="1999-07-16T20:00">8:00 p.m.</TIMEX2> on <TIMEX2 VAL="1999-07-16">Friday</TIMEX2>.

The concert is *Friday* at 8:00 *p.m.*

The concert is <TIMEX2 VAL="1999-07-16">Friday</TIMEX2> at <TIMEX2 VAL="1999-07-16T20:00">8:00 p.m.</TIMEX2>

Are you busy *Tuesday* after *12 PM*?

Are you busy <TIMEX2 VAL="1999-07-20">Tuesday</TIMEX2> after <TIMEX2 VAL="1999-07-20T12">12 PM</TIMEX2>?

#### 4.4.5 Selection of Output Format

The purpose of this section is to relate the types of expressions that taggers are likely to encounter to the output formats available for VAL in the TIMEX2 tag.

The taggable temporal expressions can be divided into three basic types:

- Points in time. These answer the question “when?”
- Duration. These answer the question “how long?”
- Frequency. These answer the question “how often?”

The remainder of this section will present prototypical examples of such expressions, and how they are annotated.

#### Points in Time

Points in time are Calendar Dates and TODs, or a combination of both (e.g., “Monday” “3 pm Monday” “next week” “current” “a Friday” “early Tuesday morning” “the weekend”).

Such expressions use one of the following types of formats for VAL:

Format	Further information and examples
VAL=“YYYY-MM-DDThh:mm:ss”	E2 “Basic, anchored expressions”
VAL=“YYYY-WOY-D”	E2.4 “Expressions involving units of weeks”
VAL=“ <i>token</i> ”	B1.1 “Tokens that replace the entire value of VAL” and E2.5 “Expressions that are anchored, but vague”
VAL=“YYYY- <i>*token*</i> ”	B1.2 “Tokens that replace particular positions in the value of VAL”
VAL=“YYYY-MM-DDT <i>*token*</i> ”	B1.2 “Tokens that replace particular positions in the value of VAL”
VAL=“WOY- <i>*token*</i> ”	B1.2 “Tokens that replace particular positions in the value of VAL”

Note: *\*token\** indicates that the token may be preceded and/or followed by other elements.

Factors that will influence the exact realization of the above formats are that truncation is allowed and any numeric position can have an X as a placeholder rather than a numeral (see Appendix A). Also, attributes in addition to VAL may be present (MOD, SET, PERIODICITY, GRANULARITY, NON\_SPECIFIC), depending on the particular context and form of the expression being tagged. Here are some annotated examples:

<TIMEX2 VAL="1999-07-12">Monday</TIMEX2>  
<TIMEX2 VAL="1999-07-12T15:00">3 p.m. Monday</TIMEX2>  
<TIMEX2 VAL="1999-W29">next week</TIMEX2>  
<TIMEX2 VAL="PRESENT\_REF">current</TIMEX2>  
<TIMEX2 VAL="1999-WXX-5" NON-SPECIFIC="YES">a Friday in <TIMEX2 VAL="1999">1999</TIMEX2></TIMEX2>

<TIMEX2 VAL="1999-FA">Fall 1999</TIMEX2>  
<TIMEX2 VAL="1999-07-13TMO" MOD="EARLY">early Tuesday morning</TIMEX2>  
<TIMEX2 VAL="1999-W28-WE">this weekend</TIMEX2>

Note again that all these examples could answer the question “When did/will this happen?”

Expressions which indicate points in time but for which no VAL is recoverable may be at least partially normalized using just one or more other attributes, depending on each expression’s form and usage. See E7.3 “Expressions without a value for VAL in output.”

Today is a *sunny day*.

<TIMEX2 VAL="1999-07-15">Today</TIMEX2> is <TIMEX2  
NON\_SPECIFIC="YES">a sunny day</TIMEX2>.

I always vote on *Election Day*.

I always vote on <TIMEX2 SET="YES" GRANULARITY="G1D"  
NON\_SPECIFIC="YES">Election Day</TIMEX2>

Other expressions for which no VAL is recoverable may be annotated without any attributes at all. In particular, see E6.2 “Anchors that are not markable time expressions.”

We saw him *five days after he came back*.

We saw him <TIMEX2>five days after he came back</TIMEX2>.

## Durations

Expressions of duration, like “three-day” or “almost half an hour” answer the question “how long.” The VAL uses the format outlined in section 5.5.3.2 of the ISO standard, i.e.:

PnYnMnDTnHnMnS

PnW

The video is only half an hour long.

The video is only <TIMEX2 VAL="PT30M">half an hour long</TIMEX2>.

The United States proposed that Israel withdraw from 13.1 percent of the West Bank over 12 weeks.

The United States proposed that Israel withdraw from 13.1 percent of the West Bank over <TIMEX2 VAL="P12W">12 weeks</TIMEX2>.

The gestation period in humans is nine months.

The gestation period in humans is <TIMEX2 VAL="P9M">nine months</TIMEX2>.<sup>3</sup>

He will make a visit to Norway next week.

He will make a <TIMEX2 VAL="P3D">three-day</TIMEX2> visit to Norway <TIMEX2 VAL="1999-W29">next week</TIMEX2>.

He will be in school for another year.

He will be in school for <TIMEX2 VAL="P1Y">another year</TIMEX2>.

She is part of the most visible and influential presence that women have had in the *52-year* history of the United Nations.

She is part of the most visible and influential presence that women have had in the <TIMEX2 VAL="P52Y">52-year</TIMEX2> history of the United Nations.

Note that we do not attempt to orient a duration expression within other temporal units. Thus, in the following example, the annotation on the expression “three-hour” does not capture which day the period occurred in, even though the context makes it clear which day it was.

He wrapped up a <TIMEX2 VAL="PT3H">three-hour</TIMEX2> meeting with the Iraqi president in Baghdad <TIMEX2 VAL="1999-07-15">today</TIMEX2>.

Our annotation does not make use of any of the ISO formats that indicate a specific end and/or start (i.e., those using the “solidus” (/) separator and described in 5.5.3.1, 5.5.3.3, and 5.5.3.4 of the ISO standard.)

Truncation of the above formats is allowed and the variable X can be used in place of a numeral (see Appendix A). In addition, other attributes may be present, such as MOD.

almost half an hour

<TIMEX2 VAL="PT30M" MOD="LESS\_THAN ">almost half an hour</TIMEX2>

Note that whether something is considered a duration or a point in time can depend largely on the context. Almost identical expressions can be tagged differently if the context implies different meanings. For example, compare the following two examples:

Point in Time:

He became depressed *five days following his graduation.*

He became depressed <TIMEX2>five days following his graduation</TIMEX2>.

Duration:

---

<sup>3</sup> “the gestation period in humans” is not a markable expression.

He was depressed for *the five days following his graduation*.  
He was depressed for <TIMEX2 VAL="P5D">the five days following his graduation</TIMEX2>.

## Frequency

Expressions of frequency answer the question “how often?”. Because they reference sets of time points rather than particular points, the SET attribute is the required attribute for such expressions. The PERIODICITY attribute is used for recurring times. VAL may be present if the set of times is anchored on a timeline.

Angry consumers grew tired of unpredictable *monthly* phone bills.  
Angry consumers grew tired of unpredictable <TIMEX2 SET="YES" PERIODICITY="F1M" GRANULARITY="G1M">monthly</TIMEX2> phone bills.  
*Last summer*, I went to the beach on *numerous Saturdays*.  
<TIMEX2 VAL="1998-SU">Last summer</TIMEX2>, I went to the beach on <TIMEX2 VAL="1998-WXX-6" SET="YES" GRANULARITY="G1D">numerous Saturdays</TIMEX2>

See E5 “Expressions of frequency” for more examples.

## 4.4.6 Lexical Triggers

### 4.4.6.1 Indicators of markable and non-markable expressions

We have so far not been explicit about the terms that trigger the tagging of a time expression. These terms are triggers for TIMEX2 tags only when used in a temporal sense, e.g., “next” is not a trigger when used in a spatial sense instead of in a temporal sense. The lists that follow are only partially representative of the variety of terms that are currently being treated as triggers.

- **Noun:** day, minute, weekend, midnight, millenium, era, semester, summer, [the] future, [the] past, months
- **Proper name (unique identifier for temporally-defined event):** *Monday, January, New Year’s Eve, Washington’s birthday*
- **Specialized time pattern:** *8:00, 12/2/00, 1994, 1960s*
- **Adjective:** *current, future, past, next, medieval, monthly*
- **Adverb:** *currently, then, next, monthly*
- **Time noun/adverb:** *today, yesterday, now*
- **Number:** *3* (as in “*He arrived at 3.*”), *three*

Note that subordinating conjunctions such as *when* are not triggers. Also note that the lists of adverb and adjective triggers are fairly conservative and do not include temporally related terms such as *eventually, lately, and frequent*.

Readers should consider examples found elsewhere in the body of this document and in

appendices E (for positive examples of triggers) and F (for negative examples) in order to better understand the breadth of the class of triggers that is intended. In the future, it is expected that a coherent characterization of the nature of the triggers will be developed.

#### **4.4.6.1 Idioms that contain trigger words**

Many time words and many time phrases that contain time words can be used in an idiomatic sense. For example, “*the last minute*”, “*the eleventh hour*”, “the order of *the day*”, “*midnight oil*” etc. The temporal expressions in such idioms are to be tagged, but are not to be assigned a value.

#### **4.4.6.2 Proper names that contain trigger words**

See Appendix F5 “Proper Names that Contain or Comprise a Time Expression”. These are not to be tagged at all.

#### **4.4.6.3 Names of holidays**

We globally refer to names of festivals, holidays and other occasions of religious observance, remembrance of famous massacres, etc. as “holidays”. Some of these expressions, like “Shrove Tuesday” and “Thanksgiving Day” contain trigger words. Others, like “Thanksgiving”, “Christmas”, and “Diwali”, do not.

A tagger is allowed to tag any holiday it wants (sorry, there is NO fixed list of holidays!), but is to assign it a value only when that value can be inferred from the local and global context of the text, rather than from cultural and world knowledge. For example, given “Christmas is celebrated in December”, the value of December is assigned to Christmas, but given only “Christmas left me poor”, “Christmas” is to be tagged without a value.

## References

- J. Alexandersson, N. Riethinger, and E. Maier. *Insights into the Dialogue Processing of VERBMOBIL*. Proceedings of the Fifth Conference on Applied Natural Language Processing, 1997, 33-40.
- J. F. Allen. *Maintaining Knowledge About Temporal Intervals*. Communications of the ACM, Volume 26, Number 11, 1983.
- M. Bennett and B. H. Partee. *Towards the Logic of Tense and Aspect in English*, Indiana University Linguistics Club, 1972.
- S. Busemann, T. Decleek, 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.
- D. Day, J. Aberdeen, L. Hirschman, R. Kozierok, P. Robinson and M. Vilain, Mixed-Initiative Development of Language Processing Systems, Fifth Conference on Applied Natural Language Processing, 1997. <http://focofa.mitre.org/~awb/anlp97-paper/ANLP97-bigger.html>
- D. Dowty. "Word Meaning and Montague Grammar", D. Reidel, Boston, 1979.
- L. Ferro, L. Hirschman, N. Chinchor, and B. Sundheim. "Relative Temporal Expressions as Named Entities." Technical Note, MITRE Corporation.
- L. Ferro. "Instruction Manual for the Annotation of Temporal Expressions". MITRE Technical Report MTR 01W0000046V01, June 2001.
- R. Grishman, B. Sundheim. Message Understanding Conference - 6: A Brief History. Proceedings of the 16th International Conference on Computational Linguistics, Copenhagen, Denmark, 1996. <http://www.cs.nyu.edu/cs/projects/proteus/publication/index.html>
- C. H. Hwang. A Logical Approach to Narrative Understanding. Ph.D. Dissertation, Department of Computer Science, U. of Alberta, 1992.
- ISO-8601 <ftp://ftp.qsl.net/pub/g1smd/8601v03.pdf> 1997.
- KSL-Time 1999. <http://www.ksl.Stanford.EDU/ontologies/time/> 1999.
- I. Mani and G. Wilson. *Robust Temporal Processing of News*, Proceedings of the ACL'2000 Conference, 3-6 October 2000, Hong Kong.
- M. Moens and M. Steedman. *Temporal Ontology and Temporal Reference*. Computational Linguistics, 14, 2, 1988, pp. 15-28.
- MUC-7. Proceedings of the Seventh Message Understanding Conference, DARPA. 1998.
- R. J. Passonneau. *A Computational Model of the Semantics of Tense and Aspect*. Computational Linguistics, 14, 2, 1988, pp. 44-60.

- J. Pustejovsky and F. Busa. A Revised Template Description for Time (v3).
- R. Quirk and S. Greenbaum. A Concise Grammar of Contemporary English, Harcourt, Brace, Jovanovich, 1973.
- H. Reichenbach. Elements of Symbolic Logic. London, Macmillan. 1947.
- 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
- B. Webber. *Tense as Discourse Anaphor*. Computational Linguistics, 14, 2, 1988, pp. 61-73.
- 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.

## Glossary

**Ambiguity.** A time expression which can have more than one meaning is called ambiguous. For example, “last year” (when uttered in the year 2000) could mean the year 1999, or the previous fiscal year (which the U.S. government defines to last from October 1, 1998 to September 30, 1999). Often, the ambiguity can be resolved, based on information in the local or global context. The guidelines call for resolving the ambiguity if possible, otherwise, to leave the VAL attribute unspecified.

**Anchored time expression.** An anchored time expression is such that either it, or its start or end, can be positioned on a time line. For example, in “*He arrived on the scene more than a year ago*”, the expression “*more than a year ago*” may be positioned at some granularity on a timeline offset from NOW.

**Annotator.** The human being who carries out the annotation, i.e., the markup of a corpus of texts according to these guidelines.

**Calendar date.** This is a representation of the day of the month, the month, and the year, as expressed in ISO spec (i.e., with the general form being CCYYMMDD).

**Context-dependence.** The dependence of the meaning of a time expression on information outside the time expression. The annotator can take context into account in resolving time expressions. See local and global context-dependence.

**Coreference.** A relationship of prior reference which is used to resolve the value of a time expression. For example, given “*Tuesday.....that day*”, where “*that day*” is identified with “*Tuesday*”, “*that day*” exhibits a relation of coreference with “*Tuesday*”. Coreference can involve identical reference, as in the above example, or the reference can be to some time related to the prior reference, e.g., “*Tuesday...that afternoon*”. The guidelines represent the values of time expressions which are coreferent with a prior reference.

**Duration.** An expression which describes an interval of time, indicating explicitly how long it lasts, e.g., “*three hours long*”. Both anchored and unanchored durations are represented in the scheme.

**Event-dependent time expression.** A time expression which is dependent on a prior event for its value to be fully resolved, e.g., “*the day after our meeting*”. Such expressions are flagged, with the granularity recorded, but they are not given a time line value.

**Fuzzy boundary.** A time expression’s value has a fuzzy boundary if the end points of the value aren’t precise. For example, “*Saturday morning*” and “*Fall*” are fuzzy with respect to when the time period starts and ends. Tokens are used to help represent fuzzy boundaries.

**Global context-dependence.** The dependence of the meaning of a time expression on information outside the time expression and the sentence containing it. The global context is confined here to the previous discourse, i.e., the sentences of the text seen so far. Thus, in a news article, given the sentence “*The hostages were beheaded on June 4th*”, the value of “*June 4<sup>th</sup>*” may be further specified to a specific year based on global context of the prior information in the article.

**Granularity.** An explicit representation for how precise a time expression is. Thus, “*Today/yesterday/next Tuesday, March 2<sup>nd</sup>*” refers to a specific time at day granularity, whereas “*next month/nine months ago*” refers to a specific time at month granularity. Granularity is represented with a separate attribute in the annotation scheme. See also **More-specific granularity**.

**Granularity of expression.** The granularity made explicit in the expression. Under granularity of expression, “*Two weeks ago*” refers to a specific week. The guidelines require that the value of the expression be represented at the level of precision that corresponds to the granularity of the expression itself.

**Holiday.** We use this term to cover names of festivals, holidays and other occasions of religious observance, remembrance of famous massacres, etc.

**Hypothesis annotation.** An annotation of a document or corpus that is used to compare against a reference annotation. Usually, the reference annotation is provided by one or more humans, and the hypothesis annotation by a program.

**Idiom.** A time expression where the literal meaning of the expression or its parts is not being used. Examples include “the last minute” in “Saddam will not cave in at the last minute”, and “*the day*”, in “*Beer-drinking is the order of the day*”. Idioms aren’t assigned a value.

**Indexicality.** Indexical time expressions require knowing when the speaker is speaking to determine the intended time value. Examples include “*now*”, “*today*”, “*yesterday*”, “*tomorrow*”, “*next Tuesday*”, “*two weeks ago*”, “*20 minutes after the next hour*”, etc. The annotation scheme emphasizes the resolution of the time values in such indexicals.

**Interval.** A representation for a time expression in terms of start and end points (which we call a “range” expression), or in terms of how long something took or lasted (a “duration”), without knowing the start and end points. The ISO spec accommodates intervals, as do we.

**Lexical trigger.** A designated time word whose absence means the expression is not to be flagged. Presence of the time word, when used in a temporal sense, means the expression is to be tagged. Our set of lexical triggers is described in Section 4.6.5; it includes words like “*day*”, “*week*”, “*weekend*”, “*now*”, “*Monday*”, “*current*”, “*future*”, etc. Since “*when*” is not a lexical trigger, an expression like “*when the stock market crashed*,” is not to be flagged. Although “*September*” is a lexical trigger, “*Black September*” is not to be flagged since the phrase is not being used primarily as an expression of time.

**Local context-dependence.** The dependence of the meaning of a time expression on information outside the time expression, but within the sentence containing it. Given “*In 1995, the months of June and July were devilishly hot.*”, the value of “*June*” and “*July*” can be resolved to a specific year based on information in the sentence.

**More specific granularity.** Granularity more specific than explicitly expressed in the time expression. “*Two weeks ago*” explicitly expresses week granularity, but may be intended to mean a specific day (an example of more specific granularity). The most precise granularity is one where the time format is fully specified. When the guidelines require that the annotator express granularity, it means granularity of the expression, rather than more specific granularity.

**Non-specific time expression.** A temporal expression where no specific time is indicated. These include generics (see above), as well as indefinites, like “*He left on a sunny day in June*”. We do not require that a tagger subcategorize such expressions, but only that it identify them, where either the VAL attribute contains a placeholder, e.g., “*a Tuesday in May*”, or “*April is usually wet*”, or where there is no VAL attribute and little else to characterize the expression, as in the case of non-referential, non-generic usage such as “*Saddam will not cave in at the last minute*”.

**Point.** A point representing a time such as a TOD or a calendar date. For example, the meaning of “*six pm today*” may be represented with the ISO value 1999-07-15T18:00. Points do not have to be completely specified on the ISO time-line; for example, a time of day may be specified without knowing the year.

**Range.** A time expression which explicitly provides a start time and an end time, e.g., “*from Monday to Thursday*”, “*from three p.m. to six p.m.*”. Rather than representing the phrase as a whole as a single anchored duration, we represent the two time NPs within the phrase as separate, unrelated points in time.

**Reference annotation.** An annotation of a document or corpus that is used as ground truth to compare against a hypothesis annotation. Usually, the reference annotation is provided by one or more humans, and the hypothesis annotation by a program.

**System.** A computer program system carrying out an annotation of a text.

**Tagger.** A human or system carrying out an annotation of a text.

**Tagging.** The process of carrying out an annotation, involving a human, a system, or both.

**TOD.** Time of Day, as expressed in 5.3 of the ISO spec (e.g., hhhmss). The clock used here is a 24-hour one.

**Token.** A representational device for capturing time expressions with fuzzy boundaries. The actual value of the boundary isn’t committed to in the representation. Examples include FA (for “Fall”), EARLY (for “early”, etc.), PRESENT\_REF (for “today”, “current”, etc.). Our complete set of tokens is described in Appendices B and C. The intent here is that a given application may choose to assign specific values to these tokens if desired; the guidelines themselves don’t dictate the specific values.

**WOY.** Week of the Year, as expressed in 5.2.3 of the ISO spec.

**Unanchored time expression.** A time expression which has no connection to a time line, as in “The video is only half an hour long.” In this latter case, the duration of half an hour is represented in the value.

# Appendix A: Usage of ISO 8601

## A1 Applicable ISO formats

Natural-language expressions of dates and of times of day (TOD) that can be completely or partially resolved to a particular value are represented via the VAL attribute using a modified version of the ISO 8601 standard.

For calendar dates, the format found in 5.2.1.1 of the ISO standard is used. For expressions of calendar weeks that are not resolved as calendar dates, the format found in 5.2.3.1 is used (but see notes about combination format and about placeholders below). For expressions of TOD, the format found in 5.3.1.1 is used. For expressions that are resolved to a combination of date and TOD, the format found in 5.4.1 is used. Note that the ISO standard allows “midnight” to be represented as either hour 00 or hour 24; for purposes of annotation consistency, we will always represent it as hour 24 of the earlier day.

For periods of time, two subsections of the standard apply. Simple expressions of duration do not permit reliable inference of a specific start or end point; for such expressions, the format in 5.5.3.2 is used. We extend this standard format by including symbols for additional units of duration: L for millenium, C for century, E for decade. For expressions that are represented as a range between start and end points (see section 4.4.5 of the guidelines), the format in 5.5.3.1 of the ISO standard is used.

The ISO standard includes both “basic” and “extended” versions of most of the formats. The extended formats include a separator character between each pair of components in the representation. Hyphens act as separators between date components, and colons as separators between TOD components. The extended format improves readability, and may be used if desired in reference and hypothesis annotations.

## A2 Placeholder character

In order to allow representation of a wide variety of NL expressions of differing degrees of precision, we have taken some liberties with the ISO standard. In addition to permitting omission of values in the calendar date and TOD representations starting from the left (truncation) and right (reduced precision), we also allow for gaps in the middle.

To make the interpretation of the resulting values unambiguous, we require the use of a placeholder character, X, for each unfilled position in the value of a component. The placeholder is always required in cases of truncated and gapped components. For example, if only the two-place year portion (decade and year) of a date value is unknown, two placeholder characters, XX, must appear in the appropriate position in the representation. These characters are also used in the representation of unknown week values. For example, our representation for ‘all Fridays in 1999’ would use the value 1999WXX5 (or 1999-WXX-5), plus other attributes to specify the set.

In the case of reduced precision, where the unfilled position(s) are on the right-hand end of the value, no placeholder is used except in special cases. The placeholder character is only used when the NL expression explicitly identifies the intended degree of precision and, at the same time, does *not* identify the value for the component that is at that degree of precision. For example, the sentence, “He left on a sunny day in June”, explicitly identifies *day* as the intended degree of precision, and it indicates that the speaker has a particular day in mind, but does not identify *which* day the speaker has in mind. The

representation of ‘a sunny day in June’ would therefore be 199906XX (or 1999-06-XX), plus an attribute to indicate the non-specific nature of the expression, along with a granularity value of “G1D”.

In the representation of durations, the X character is also used, but as a placeholder for an unspecified plural number value of a component rather than as a placeholder for a component as a whole. For example, PXY represents a duration of ‘X’ years and would appear in the representation for “years” in a sentence such as, “He has been going to school for years.”

Note that an omitted value or a placeholder value indicates an unknown or unspecified value. In contrast to the usage described in the ISO standard, there is no implication of default to “current” time values in such cases.

### **A3 Additional differences from the ISO standard**

There are complex NL expressions that contain both week-based and month-based subexpressions. To handle such cases, we have extended the ISO standard to permit the value to be expressed as a single string. For example, “They were sitting on the porch *one Friday night in fall 1998*” would be assigned the value of 1998-FA-WXX-5-TNI (plus an attribute to indicate the non-specific nature of the reference). See Appendix B re the tokens “FA” and “NI”.

Also note that we diverge slightly from the standard in what we regard as a “component” of a value. ISO treats “year” as an essentially indivisible, four-place component (YYYY). In contrast with that, we consider the first position the “millennium” component, the second one the “century” component, the third the “decade” component, and the fourth the “year” component. Thus, in our version, a well-formed value could consist of just one number (for example, “3” could represent “the next millennium”), two numbers (for example, “20” could represent “the 21<sup>st</sup> century”), or three (for example, “196” could represent “the 1960s”).

The ISO standard is not intended to accommodate inherently imprecise expressions of time such as “summer”, “night”, and “future”. To accommodate representation of such concepts, we have augmented the formats specified in the standard. See Appendix B.

# Appendix B: Token Lists for VAL Attribute

## B1 Introduction

Some inherently imprecise temporal expressions and subexpressions are represented in the target output as alpha character tokens in the value for the VAL attribute.

Below are two complete lists of the alpha character tokens that can appear in the value for the VAL attribute, along with examples of the NL expressions that map to each token. The first list identifies tokens that replace the entire value of VAL. The second list identifies tokens that replace particular positions in the value.

The list of tokens will be augmented if annotation of corpora reveals a need for additional tokens. Developers should exploit annotated corpora to identify the full set of pertinent NL expressions that have been represented using the documented tokens.

## B2 Tokens that occupy the entire value of VAL

These tokens augment the options for the representation of the value of time points. They are not relevant to the representation of durations.

There is currently no allowance in the encoding for capturing different reference times for these values. For example, given the value FUTURE\_REF, it is not possible to tell if the reference time for that value is the publication date of the document or if it is some narrative-internal reference time.

Token	Markable Expressions	Non-Markable Expressions
PRESENT_REF	now today [unless further resolvable to a particular date] current, currently present, presently nowadays (at) this (point in) time (at) the present time/moment	immediately instantly forthwith
FUTURE_REF	future tomorrow [unless further resolvable to a particular date]	ahead after soon, sooner shortly later eventually subsequent
PAST_REF	past yesterday [unless further resolvable to a particular date]	before previously earlier

	former lately long ago medieval	beforehand once
--	--	--------------------

*Now* there are hundreds of schools with thousands and thousands of dancers.  
 <TIMEX2 VAL="PRESENT\_REF">Now</TIMEX2> there are hundreds of schools with thousands and thousands of dancers.

We can trust *today*'s youth to do the right thing about the environment.

We can trust <TIMEX2 VAL="PRESENT\_REF">today</TIMEX2>'s youth to do the right thing about the environment.

We may be moving back to the *medieval* notion of romance.

We may be moving back to the <TIMEX2 VAL="PAST\_REF">medieval</TIMEX2> notion of romance.

His partner is a 22-year-old *future* accountant.

His partner is a 22-year-old <TIMEX2 VAL="FUTURE\_REF">future</TIMEX2> accountant.

Duisenberg is a *former* Dutch central banker.

Duisenberg is a <TIMEX2 VAL="PAST\_REF">former</TIMEX2> Dutch central banker.

We need to discuss *the future* of our peoples.

We need to discuss <TIMEX2 VAL="FUTURE\_REF">the future of our peoples</TIMEX2>.

### **B3 Tokens that occupy particular positions in the value of VAL**

These tokens augment the options for the representation of the value of both time points and durations. In the case of durations, the two-character token substitutes for the corresponding one-character designator defined in the ISO standard.

Each of the tokens consists of two alpha characters. The ones that occupy the month position do not apply to the week-based format; the WE token, on the other hand, applies *only* to the week-based format.

<b>Token</b>	<b>Position</b>	<b>Expressions</b>
AF	hour	afternoon
DT	hour	morning + afternoon (basically, daytime or working hours)
EV	hour	evening
FA	month	fall autumn fall semester fall term
H1	month	1st half (of year)
H2	month	2nd half (of year)

MI	hour	mid-day
MO	hour	morning
NI	hour	night
Q1	month	1 <sup>st</sup> quarter
Q2	month	2 <sup>nd</sup> quarter
Q3	month	3 <sup>rd</sup> quarter
Q4	month	4 <sup>th</sup> quarter
SP	month	spring
SU	month	summer
WE	day	weekend
WI	month	winter

There were doughnuts at the 8:00 meeting *this morning*.

There were doughnuts at the <TIMEX2 VAL="1999-07-15-T08:00">8:00</TIMEX2> meeting  
<TIMEX2 VAL="1999-07-15-TMO">this morning</TIMEX2>.

The bug will get fixed between *now* and *Monday morning*.

The bug will get fixed between <TIMEX2 VAL="PRESENT\_REF">now</TIMEX2> and <TIMEX2  
VAL="1999-07-19-TMO">Monday morning</TIMEX2>

Note that these tokens are only used if the precise time of day is not present in the expression. For example, “eleven in the morning” is simply given a TOD value of “T11” and “morning” is not annotated separately.

A particular kind of vagueness arises with the expression “*last night*.” This could mean the night of the previous day, or the night portion of the current 24 hour period, or both. As a convention, “last night” always gets the date of “yesterday”.

For example:

The talks ended *last night*.

The talks ended <TIMEX2 VAL="1999-07-14-TNI"> last night</TIMEX2>.

# Appendix C: Token Lists for MOD Attribute

## C1 Introduction

Some modifiers of time expressions convey vague quantification over the time referred to by the time expression as a whole, and others provide aspectual information about the time expression as a whole. To the extent possible within the confines of a simple representation scheme, the value for the MOD attribute captures the semantics of these sorts of modifiers.

## C2 Tokens for MOD

Below is a complete list of the alpha character tokens that can appear as the value for the MOD attribute, along with examples of NL expressions that map to each token.

The list of tokens will be augmented if annotation of corpora reveals a need for additional tokens. Developers should exploit annotated corpora to identify the full set of pertinent NL expressions that have been represented using the documented tokens.

	Token	Sample Expressions
<b>Points</b>	BEFORE	more than (“more than a decade ago”)
	AFTER	less than (“less than a year ago”)
	ON_OR_BEFORE	no less than (“no less than a year ago”)
	ON_OR_AFTER	no more than (“no more than a year ago”)
<b>Durations</b>	LESS_THAN	less than (e.g., “less than 2 hours long”) nearly (e.g., “nearly four decades of experience”)
	MORE_THAN	more than (e.g., “more than 5 minutes”)
	EQUAL_OR_LESS	no more than (e.g., “...will be open no more than 10 days”)
	EQUAL_OR_MORE	at least (e.g., “...will be open at least 10 days”)
<b>Points and Durations</b>	START	early (e.g., “the early 1960s”) dawn (e.g., “the dawn of 2000”) start (e.g., “the start of the quarter”) beginning
	MID	middle (e.g., “the middle of the month”) mid- (e.g., “mid-February”)
	END	end late
	APPROX	about (e.g., “about three years ago”) around

Sen. Alton Waldon, who served briefly in Congress *more than a decade ago*, is now retired.

Sen. Alton Waldon, who served briefly in Congress <TIMEX2 VAL="1989" MOD="BEFORE">more than a decade ago</TIMEX2>, is <TIMEX2 VAL="PRESENT\_REF">now</TIMEX2> retired.

The teacher has *nearly four decades of experience*.

The teacher has <TIMEX2 VAL="P4E" MOD="LESS\_THAN">nearly four decades of experience</TIMEX2>.

She has been at work for *more than a month*.

She has been at work for <TIMEX2 VAL="P1M" MOD="MORE\_THAN">more than a month</TIMEX2>.

Iraq proposes to open its presidential sites for *no more than 60 days*.  
Iraq proposes to open its presidential sites for <TIMEX2 VAL="P60D"  
MOD="EQUAL\_OR\_LESS">no more than 60 days</TIMEX2>.

Annan is due in Baghdad on *Friday* after a stop in Paris. He said he will stay in Iraq for *no more than two days* before returning to brief the Security Council.  
Annan is due in Baghdad on <TIMEX2 VAL="1999-07-16">Friday</TIMEX2> after a stop in Paris. He said he will stay in Iraq for <TIMEX2 VAL="P2D"  
MOD="EQUAL\_OR\_LESS">no more than two days</TIMEX2> before returning to brief the Security Council.

Britain is staying outside the currency union for *at least the next year or two*.  
Britain is staying outside the currency union for <TIMEX2 VAL="P1Y"  
MOD="EQUAL\_OR\_MORE">at least the next year</TIMEX2> or <TIMEX2 VAL="P2Y"  
MOD="EQUAL\_OR\_MORE">two</TIMEX2>.<sup>4</sup>

There is certain to be heightened excitement at *the dawn of 2000*.  
There is certain to be heightened excitement at <TIMEX2 VAL="2000"  
MOD="START">the dawn of 2000</TIMEX2>.

---

<sup>4</sup> The MOD is shown as applying to each half of the disjunction, although that is somewhat misleading semantically.

## Appendix D: Tokenization Rules

In the examples given in this appendix, square brackets are used to indicate where the open and close TIMEX2 tags are expected to be located.

### D1 Punctuation Marks and Special Characters Excluded

Punctuation marks and white space falling outside the extent of a taggable TIMEX2 tag should remain outside the tags. For example:

#### D1.1 Periods (and ellipses) and Commas

Ann took office [*a year ago*].  
He stopped by [*Monday*], [*Tuesday*], [*Wednesday*]...

#### D1.2 Hyphens or Dashes

A [*Y2K*]-induced hysteria  
[*December 7, 1941*] – That day will long be remembered in Hawaii.  
The class is [*3*]-[*6 pm today*].

#### D1.3 Apostrophes

We can trust [*today*]'s youth to do the right thing about the environment.

#### D1.4 Parentheses and Quotes

Can you meet [*tomorrow*] ([*Friday*])?  
He wrote “I’ll see you at the race in [*September*]” in his letter to her.

### D2 Punctuation Marks and Special Characters Included

Punctuation marks and white space falling within the extent of a taggable TIMEX2 tag should be included in the scope of the tag. For example:

#### D2.1 Periods and Commas

A [*Sept.*] snowfall is unusual in these parts.  
[*9 a.m. Friday, October 1, 1999*]

#### D2.2 Slashes

[*10/13/89*]

#### D2.3 Hyphens

He wrapped up a [*three-hour*] meeting with the Iraqi president in Baghdad today.  
He explained that he was flying off to Bosnia in [*the morning*] and wanted to go out with a bang.

Note: When a hyphen is used at the end of a line to separate a single word into two parts, the expression is treated as a single token.

#### D2.4 Apostrophes

That happened back in [*'87*].

## **D2.5 Other Special Cases**

I'm a creature of [*the 1960s*]

Note: The plural marker “s” is included within the scope of the tag.

I'm a creature of [*the 1960's*]

Note: The plural marker “'s” is included within the scope of the tag.

# Appendix E: Examples of Expressions and Annotations

## E1 Introduction

This appendix contains a categorized list of example sentences and their target annotation for temporal expressions. Some sentences repeat examples contained in the body of the guidelines.

This list focuses primarily on the conversion of appropriately identified expressions into the normalized values of the various tag attributes. In addition, there are some examples of time expressions that do not get normalized values, examples of expressions that do not qualify as time expressions, and examples in which time expressions appear in close proximity to each other but are tagged as separate expressions. Italics are used to indicate the expression(s) of interest for the particular section in which the example sentence appears. In cases where an example sentence contains multiple time expressions, the use of italics is intended to make it obvious to the reader which of the expressions is/are especially pertinent to that section.

For the sake of readability, the VAL notation makes use of the hyphen separator character.

Unless otherwise indicated, the following reference date for indexical expressions is assumed: Thursday, July 15, 1999. In terms of weeks, that is week 28 of 1999.

## E2 Basic, anchored expressions

### E2.1 Expressions of calendar dates, with single value for VAL

The two collaborated closely during the *1994* crisis over Haiti.

The two collaborated closely during the <TIMEX2 VAL="1994">1994</TIMEX2> crisis over Haiti.

After an emergency meeting in *November*, relations began to improve.

After an emergency meeting in <TIMEX2 VAL="1998-11">November</TIMEX2>, relations began to improve.

I was sick *yesterday*.

I was sick <TIMEX2 VAL="1999-07-14">yesterday</TIMEX2>.

The bombing took place on *the second of December*.

The bombing took place on <TIMEX2 VAL="1998-12-02">the second of December</TIMEX2>.

*Five years ago* there were dozens of schools with hundreds of students doing ballroom dance.

<TIMEX2 VAL="1994">Five years ago</TIMEX2> there were dozens of schools with hundreds of students doing ballroom dance.

Ann took office *a year ago*.

Ann took office <TIMEX2 VAL="1998">a year ago</TIMEX2>.

New funding arrived in *Fall 1998*.

New funding arrived in <TIMEX2 VAL="1998-FA">Fall 1998</TIMEX2>.

There was a big wildfire in Santa Barbara in *summer of 1964*.

There was a big wildfire in Santa Barbara in <TIMEX2 VAL="1964-SU">summer of 1964</TIMEX2>.

There was a big wildfire in *1964*, in *the summer*.

There was a big wildfire in <TIMEX2 VAL="1964">1964</TIMEX2>, in <TIMEX2 VAL="1964-SU">the summer</TIMEX2>.

A Chinese gymnast was paralyzed in the Goodwill Games *last summer*.

A Chinese gymnast was paralyzed in the Goodwill Games <TIMEX2 VAL="1998-SU">last summer</TIMEX2>.

Students were contacting their president with a desperate plea in *the fall semester*.

Students were contacting their president with a desperate plea in <TIMEX2 VAL="1998-FA">the fall semester</TIMEX2>.

%% Note reduction of "fall semester" into the token "FA". And, since "fall" is represented as being contained in one calendar year, so is "fall semester".

Dancing deteriorated in *the 1960s* into group chaos.

Dancing deteriorated in <TIMEX2 VAL="196">the 1960s</TIMEX2> into group chaos.

The U.N. Secretary-General departs *this weekend* for Baghdad.

The U.N. Secretary-General departs <TIMEX2 VAL="1999-W28-WE">this weekend</TIMEX2> for Baghdad.

The senators will be working through *the weekend*.

The senators will be working through <TIMEX2 VAL="1999-W28-WE">the weekend</TIMEX2>.

NATO is debating how the Atlantic security partnership should define its strategic interests for *the next century*.

NATO is debating how the Atlantic security partnership should define its strategic interests for <TIMEX2 VAL="20">the next century</TIMEX2>.

## **E2.2 Expressions of times of day (TODs) in input**

### **E2.2.1 TODs in input that result in combined date and TOD in output**

The sponsor arrived at *ten minutes to 3*.

The sponsor arrived at <TIMEX2 VAL="1999-07-15-T14:50">ten minutes to 3</TIMEX2>.

The talks ended *last night*.

The talks ended <TIMEX2 VAL="1999-07-14-TNI"> last night</TIMEX2>.  
%% Note: "last night" gets the date of yesterday.

On *the nineteenth* I am in class until *eleven in the morning*.

%% This example is taken from a dialogue transcript.

On <TIMEX2 VAL="1999-07-19">the nineteenth</TIMEX2> I am in class until <TIMEX2 VAL="1999-07-19T11">eleven in the morning</TIMEX2>.

%% Note: "eleven in the morning" is treated as one expression (cf. section 4.4.4), and it gets the date portion of its value from the expression, "the nineteenth".

### E2.2.2 TODs in input that result only in TOD in output

Hickory dickory dock. The mouse ran up the clock. The clock struck *one*, the mouse ran down.

Hickory dickory dock.

%% Assume for this example that we know we're talking about 1 a.m., but we don't know the date.

Hickory dickory dock. The mouse ran up the clock. The clock struck <TIMEX2 VAL="T01:00">one</TIMEX2>, the mouse ran down. Hickory dickory dock.

### E2.3 Expressions that combine date and TOD in input and output

2/27/1998 08:14:00

%% This example is from a document header.

<TIMEX2 VAL="1998-02-27-T08:14:00">2/27/1998 08:14:00</TIMEX2>

I returned to work at *twelve o'clock January 3, 1984*.

%% Assume for this example that we know that the time refers to noon.

I returned to work at <TIMEX2 VAL="1984-01-03-T12:00">twelve o'clock January 3, 1984</TIMEX2>.

She stayed home *Friday evening*.

%% Assume for this example that the most recent Friday evening is implied.

She stayed home <TIMEX2 VAL="1999-07-09-TEV">Friday evening</TIMEX2>.

She received an important phone call at *8:00 Friday night*.

%% Assume for this example that the most recent Friday night is implied.

She received an important phone call at <TIMEX2 VAL="1999-07-09-T20:00">8:00 Friday night</TIMEX2>.

The mail arrived at *mid-day yesterday*.

The mail arrived at <TIMEX2 VAL="1999-07-14-TMI">mid-day yesterday</TIMEX2>.

### E2.4 Expressions involving units of weeks

He will visit Norway *next week*.

He will visit Norway <TIMEX2 VAL="1999-W29">next week</TIMEX2>.

She took over in his absence *two weeks ago* to deliver a sharp protest to the Congolese ambassador.

She took over in his absence <TIMEX2 VAL="1999-W26">two weeks ago</TIMEX2> to deliver a sharp protest to the Congolese ambassador.

### **E2.5 Expressions that are anchored, but vague**

Now there are hundreds of schools with thousands and thousands of dancers.

<TIMEX2 VAL="PRESENT\_REF">Now</TIMEX2> there are hundreds of schools with thousands and thousands of dancers.

We can trust *today's* youth to do the right thing about the environment.

We can trust <TIMEX2 VAL="PRESENT\_REF">today</TIMEX2>'s youth to do the right thing about the environment.

Binge drinking is the *current* plague of college campuses.

Binge drinking is the <TIMEX2 VAL="PRESENT\_REF">current</TIMEX2> plague of college campuses.

Europe will be stronger than in *the past*.

Europe will be stronger than in <TIMEX2 VAL="PAST\_REF">the past</TIMEX2>.

NATO may be changing a military destiny *once* based on geography to a defense of common values.

NATO may be changing a military destiny <TIMEX2 VAL="PAST\_REF">once</TIMEX2> based on geography to a defense of common values.

We may be moving back to the *medieval* notion of romance.

We may be moving back to the <TIMEX2 VAL="PAST\_REF">medieval</TIMEX2> notion of romance.

His partner is a 22-year-old *future* accountant.

His partner is a 22-year-old <TIMEX2 VAL="FUTURE\_REF">future</TIMEX2> accountant.

We need to discuss *the future of our peoples*.

We need to discuss <TIMEX2 VAL="FUTURE\_REF">the future of our peoples</TIMEX2>.

The restaurant opened *about three years ago*.

The restaurant opened <TIMEX2 VAL="1996" MOD="APPROX">about three years ago</TIMEX2>.

Sen. Alton Waldon, who served briefly in Congress *more than a decade ago*, is *now* retired.

Sen. Alton Waldon, who served briefly in Congress <TIMEX2 VAL="1989" MOD="BEFORE">more than a decade ago</TIMEX2>, is <TIMEX2 VAL="PRESENT\_REF">now</TIMEX2> retired.

There is certain to be heightened excitement at *the dawn of 2000*.

There is certain to be heightened excitement at <TIMEX2 VAL="2000" MOD="EARLY">the dawn of 2000</TIMEX2>.

The trend began in *the early 1960s*.

The trend began in <TIMEX2 VAL="196" MOD="EARLY">the early 1960s</TIMEX2>.

The talks ended *late last night*.

The talks ended <TIMEX2 VAL="1999-07-14-TNI" MOD="LATE">late last night</TIMEX2>.

## E2.6 Expressions of duration

The video is only *half an hour long*.

The video is only <TIMEX2 VAL="PT30M">half an hour long</TIMEX2>.

%% “Only” is treated as modifying “is” rather than as modifying the time expression.

The United States proposed that Israel withdraw from 13.1 percent of the West Bank over *12 weeks*.

The United States proposed that Israel withdraw from 13.1 percent of the West Bank over <TIMEX2 VAL="P12W">12 weeks</TIMEX2>.

The gestation period in humans is *nine months*.

The gestation period in humans is <TIMEX2 VAL="P9M">nine months</TIMEX2>.

%% If we were to follow the MUC coreference guidelines, we would also tag “the gestation period in humans” as a TIMEX2, but we have decided to diverge from those guidelines in cases like this.

He will make a *three-day* visit to Norway next week.

He will make a <TIMEX2 VAL="P3D">three-day</TIMEX2> visit to Norway <TIMEX2 VAL="1999-W29">next week</TIMEX2>.

He will be in school for *another year*.

He will be in school for <TIMEX2 VAL="P1Y">another year</TIMEX2>.

%% “Another” is regularized as “1”.

The conventions around dating fell away in *recent decades*.

The conventions around dating fell away in <TIMEX2 VAL="PXE">recent decades</TIMEX2>.

%% Representation does not capture meaning of “recent” in any way. The phrase “in recent decades” is interpreted to mean “during (over the course of) recent decades.”

She has been at work for *more than a month*.

She has been at work for <TIMEX2 VAL="P1M" MOD="MORE\_THAN">more than a month</TIMEX2>.

The teacher has *nearly four decades of experience*.

The teacher has <TIMEX2 VAL="P4E" MOD="LESS\_THAN">nearly four decades of experience</TIMEX2>.

He wrapped up a *three-hour* meeting with the Iraqi president in Baghdad today.  
He wrapped up a <TIMEX2 VAL="PT3H">three-hour</TIMEX2> meeting with the Iraqi president in Baghdad <TIMEX2 VAL="1999-07-15">today</TIMEX2>.

After *months of renewed hostility*, the fighting has suddenly ceased.  
After <TIMEX2 VAL="PXM">months of renewed hostility</TIMEX2>, the fighting has suddenly ceased.

She is part of the most visible and influential presence that women have had in the *52-year* history of the United Nations.  
She is part of the most visible and influential presence that women have had in the <TIMEX2 VAL="P52Y">52-year</TIMEX2> history of the United Nations.

For *millenniums*, dancing was a form of social glue.  
For <TIMEX2 VAL="PXL">millenniums</TIMEX2>, dancing was a form of social glue.

What has helped most was my *three years* at the Defense Department.  
What has helped most was <TIMEX2 VAL="P3Y">my three years at the Defense Department</TIMEX2>.

Iraq proposes to open its presidential sites for *no more than 60 days*.  
Iraq proposes to open its presidential sites for <TIMEX2 VAL="P60D" MOD="EQUAL\_OR\_LESS">no more than 60 days</TIMEX2>.

Annan is due in Baghdad on Friday after a stop in Paris. He said he will stay in Iraq for *no more than two days* before returning to brief the Security Council.  
Annan is due in Baghdad on Friday after a stop in Paris. He said he will stay in Iraq for <TIMEX2 VAL="P2D" MOD="EQUAL\_OR\_LESS">no more than two days</TIMEX2> before returning to brief the Security Council.

%% Representation of this expression takes local context only into account. There is not enough information from global context about the endpoints of the intended interval to justify adding a second layer of tagging that would contain the interval of the visit (we are told only the beginning day, not the end day, and it would be risky (possibly contrary to the speaker's intention) to compute the latter from the former.

## **E3 Examples where expressions appear in close proximity, but do not result in a single tag or in nested tags**

### **E3.1 Expressions of explicitly defined ranges**

The class is *3-6 pm today*.  
The class is <TIMEX2 VAL="1999-07-15-T15">3</TIMEX2>-<TIMEX2 VAL="1999-07-15-T18">6 pm today</TIMEX2>.

The prime minister's visit is to run *August 6-8*.

The prime minister's visit is to run <TIMEX2 VAL="1999-08-06">August 6</TIMEX2>-  
<TIMEX2 VAL="1999-08-08">8</TIMEX2>.

Dinner is from *five to six pm tomorrow*.

Dinner is from <TIMEX2 VAL="1999-07-16-T17">five</TIMEX2> to <TIMEX2 VAL="1999-07-16-T18">six pm tomorrow</TIMEX2>.

She served as Canada's ambassador to the U.N. from *1992 through 1995*.

She served as Canada's ambassador to the U.N. from <TIMEX2 VAL="1992">1992</TIMEX2>  
through <TIMEX2 VAL="1995">1995</TIMEX2>.

### E3.2 Expressions with elements separated by a conjunction

The bug will get fixed between *now and Monday morning*.

The bug will get fixed between <TIMEX2 VAL="PRESENT\_REF">now</TIMEX2> and  
<TIMEX2 VAL="1999-07-19-TMO">Monday morning</TIMEX2>.

Saddam might play the whole game again *six months or a year from now*.

Saddam might play the whole game again <TIMEX2 VAL="2000-01">six months</TIMEX2>  
or <TIMEX2 VAL="2000">a year from <TIMEX2  
VAL="PRESENT\_REF">now</TIMEX2></TIMEX2>.

%% The granularity of each value corresponds to the granularity of the individual expression.

Britain is staying outside the currency union for *at least the next year or two*.

Britain is staying outside the currency union for <TIMEX2 VAL="P1Y"  
MOD="EQUAL\_OR\_MORE">at least the next year</TIMEX2> or <TIMEX2  
VAL="P2Y" MOD="EQUAL\_OR\_MORE">two</TIMEX2>.

%% The MOD is shown as applying to each half of the disjunction, although that is somewhat misleading semantically.

## E4 Anaphoric expressions

%% These examples show various ways in which expressions corefer. In most cases, the coreference relation is not one of complete identity of reference.

I'm a creature of the 1960s, *the days of free love*.

%% This is a case of apposition (syntactically-determined coreference).

I'm a creature of <TIMEX2 VAL="196">the 1960s</TIMEX2>, <TIMEX2 VAL="196">the  
days of free love</TIMEX2>.

The contractor submitted a proposal on Tuesday. The day after *that*, the contract was awarded.

*That night*, they had a big party.

%% Note: "that [day]" is coreferential with "Tuesday". The demonstrative "That" in "That night" is referring to "The day after that", i.e., to Wednesday.

The contractor submitted a proposal on <TIMEX2 VAL="1999-07-13">Tuesday</TIMEX2>.  
<TIMEX2 VAL="1999-07-14">The day after <TIMEX2 VAL="1999-07-13">that</TIMEX2> </TIMEX2>, the contract was awarded. <TIMEX2 VAL="1999-07-14-TNI">That night</TIMEX2>, they had a big party.

Duisenberg was named president of the European Central Bank last May. He was favored by Helmut Kohl, who was *then* chancellor of Germany.

Duisenberg was named president of the European Central Bank <TIMEX2 VAL="1999-05">last May</TIMEX2>. He was favored by Helmut Kohl, who was <TIMEX2 VAL="1999-05">then</TIMEX2> chancellor of Germany.

At 11:59 p.m., Mayor Rudolph W. Giuliani sat on a stage at 45th Street and Broadway and pushed a button. *Sixty seconds later*, the ball atop 1 Times Square completed its slow descent into retirement.

%% Assume for this example that the most recent New Year's Eve is implied.

At <TIMEX2 VAL="1998-12-31-T23:59">11:59 p.m.</TIMEX2>, Mayor Rudolph W. Giuliani sat on a stage at 45th Street and Broadway and pushed a button. <TIMEX2 VAL="1998-12-31-T24:00:00">Sixty seconds later</TIMEX2>, the ball atop 1 Times Square completed its slow descent into retirement.

%% The granularity of the first expression is in minutes, while the granularity of the second expression is in seconds.

He explained that he was flying off to Bosnia in *the morning* and wanted to go out with a bang.

%% Note: This example appeared in a narrative about New Year's Eve 1998.

He explained that he was flying off to Bosnia in <TIMEX2 VAL="1999-01-01-TMO">the morning</TIMEX2> and wanted to go out with a bang.

The Dow rose 17 percent during the 4th quarter. The S&P's 500-stock index gained nearly 21 percent, its *second-best quarter ever*.

The Dow rose 17 percent during <TIMEX2 VAL="1998-Q4">the 4th quarter</TIMEX2>. The S&P's 500-stock index gained nearly 21 percent, its <TIMEX2 VAL="1998-Q4">second-best quarter ever</TIMEX2>.

Secretary of State Madeleine K. Albright attended a meeting of NATO foreign ministers here two months ago. [...30 lines later] Albright addressed this perceptual problem at the *December* meeting.

Secretary of State Madeleine K. Albright attended a meeting of NATO foreign ministers here <TIMEX2 VAL="1998-12">two months ago</TIMEX2>. [...30 lines later] Albright addressed this perceptual problem at the <TIMEX2 VAL="1998-12">December</TIMEX2> meeting.

## E5 Expressions of frequency

%% These expressions are distinguished in the representation by having SET="YES".

### E5.1 Expressions that convey regular recurrence

%% These expressions are distinguished in the representation by having the PERIODICITY attribute and the GRANULARITY attribute.

They watched Millionaire on TV *every Tuesday in 1999*.

They watched Millionaire on TV <TIMEX2 VAL="1999-WXX-2" SET="YES" PERIODICITY="F1W" GRANULARITY="G1D">every Tuesday in <TIMEX2 VAL="1999">1999</TIMEX2></TIMEX2>.

There were, as there have been *every December 31 for decades*, thousands of people in Times Square.

There were, as there have been <TIMEX2 VAL="XXXX-12-31" SET="YES" PERIODICITY="F1Y" GRANULARITY="G1D">every December 31</TIMEX2> for <TIMEX2 VAL="PXE">decades</TIMEX2>, thousands of people in Times Square.

%% The phrase “for decades” is not considered a modifier of the preceding time expression, so there are two independent TIMEX2 annotations.. A system could detect this subtlety fairly readily on semantic grounds by noticing that a point expression is immediately followed by a duration expression.

Two years ago, the dance club drew about 100 students *each week*.

<TIMEX2 VAL="1997">Two years ago</TIMEX2>, the dance club drew about 100 students <TIMEX2 VAL="1997" SET="YES" PERIODICITY="F1W" GRANULARITY="G1W">each week</TIMEX2>.

%% “Each” is represented the same as “every” (“each and every”).

They reviewed their stock portfolio *the first day of each month in 1999*.

They reviewed their stock portfolio <TIMEX2 VAL="1999-XX-01" SET="YES" PERIODICITY="F1M" GRANULARITY="G1D">the first day of <TIMEX2 VAL="1999" SET="YES" PERIODICITY="F1M" GRANULARITY="G1M">each month in <TIMEX2 VAL="1999">1999</TIMEX2></TIMEX2></TIMEX2>.

%% “Each” is represented the same as “every” (“each and every”).

We spend *the first three days of every month* writing status reports.

We spend <TIMEX2 VAL="XXXX-XX-01/XXXX-XX-03" SET="YES" PERIODICITY="F1M" GRANULARITY="G3D">the first three days of <TIMEX2 SET="YES" PERIODICITY="F1M" GRANULARITY="G1M">every month</TIMEX2></TIMEX2> writing status reports.

Angry consumers grew tired of unpredictable *monthly* phone bills.

Angry consumers grew tired of unpredictable <TIMEX2 SET="YES" PERIODICITY="F1M" GRANULARITY="G1M">monthly</TIMEX2> phone bills.

She attends the *annual* stockholder meetings.

She attends the <TIMEX2 SET="YES" PERIODICITY="F1Y" GRANULARITY="G1Y">annual</TIMEX2> stockholder meetings.

## E5.2 Expressions that do not necessarily imply regular recurrence

%% These expressions do not receive the PERIODICITY attribute, but they do receive the GRANULARITY attribute.

It was easier for the average reveler to see and hear what was going on this year than in *recent years*.

%% In this example, “recent” does not necessarily imply “all recent”. Also, “in recent years” is not interpreted as a duration, but rather as a set of points (a set of New Year’s Eves, in the broader context in which this example was seen).

It was easier for the average reveler to see and hear what was going on <TIMEX2 VAL="1999">this year</TIMEX2> than in <TIMEX2 SET="YES" GRANULARITY="G1Y">recent years</TIMEX2>.

%% Note that there is no VAL for “recent years”, since we do not allow “recent” to be normalized as PAST\_REF.

On *Friday nights*, the ballroom dance club was overwhelmed.

%% Assume for this example that 1998 is implied.

On <TIMEX2 VAL="1998-WXX-5-TNI" SET="YES" GRANULARITY="G1NI">Friday nights</TIMEX2>, the ballroom dance club was overwhelmed.

Last summer, I went to the beach on *numerous Saturdays*.

<TIMEX2 VAL="1998-SU">Last summer</TIMEX2>, I went to the beach on <TIMEX2 VAL="1998-WXX-6" SET="YES" GRANULARITY="G1D">numerous Saturdays</TIMEX2>.

I tutored an English student *some Thursdays in 1998*.

I tutored an English student <TIMEX2 VAL="1998-WXX-4" SET="YES" GRANULARITY="G1D">some Thursdays in <TIMEX2 VAL="1998">1998</TIMEX2></TIMEX2>.

## E6 Expressions that include explicit anchors for offsets

### E6.1 Anchors that are markable time expressions

A major earthquake struck Los Angeles *three years ago today*.

A major earthquake struck Los Angeles <TIMEX2 VAL="1996-07-15">three years ago <TIMEX2 VAL="1999-07-15">today</TIMEX2></TIMEX2>.

They had lunch together *two weeks ago today*.

They had lunch together <TIMEX2 VAL="1999-07-01">two weeks ago <TIMEX2 VAL="1999-07-15">today</TIMEX2></TIMEX2>.

%% Note that calendar date format is used instead of week format, since “today” provides a precise anchor for computing the date of the full expression (and justifies using granularity of *day* rather than *week*).

I’m leaving on vacation *two weeks from next Tuesday*.

I’m leaving on vacation <TIMEX2 VAL="1999-08-03">two weeks from <TIMEX2 VAL="1999-07-20">next Tuesday</TIMEX2></TIMEX2>.

We are expecting a reply *two days from now*.

We are expecting a reply <TIMEX2 VAL="1999-07-17">two days from <TIMEX2 VAL="PRESENT\_REF">now</TIMEX2></TIMEX2>.

## E6.2 Anchors that are not markable time expressions

We saw him *five days after he came back*.

We saw him <TIMEX2>five days after he came back</TIMEX2>.

The firefighters came home *three days after the fire*.

The firefighters came home <TIMEX2>three days after the fire</TIMEX2>.

The company he had invested in went bankrupt within *minutes after the stock market closed for the day*.

The company he had invested in went bankrupt within <TIMEX2>minutes after the stock market closed for the day</TIMEX2>.

We were still talking about work *three hours after the meeting broke up*.

We were still talking about work <TIMEX2>three hours after the meeting broke up</TIMEX2>.

He realized his mistake *three days after Sunday's paper was thrown out*.

He realized his mistake <TIMEX2>three days after <TIMEX2 VAL="WXX-7">Sunday</TIMEX2>'s paper was thrown out</TIMEX2>.

He returned some gifts *five days after Christmas*.

He returned some gifts <TIMEX2>five days after  
<TIMEX2>Christmas</TIMEX2></TIMEX2>.

%% Although "Christmas" is a TIMEX2, it doesn't provide an anchoring date value for the offset.

He was depressed for *the five days following his graduation*.

He was depressed for <TIMEX2 VAL="P5D">the five days following his  
graduation</TIMEX2>.

He was very tired in *the hours after his inauguration*.

He was very tired in <TIMEX2 VAL="PTXH">the hours after his inauguration</TIMEX2>.

## E7 Non-specific time expressions

%% These are distinguished by NON\_SPECIFIC="YES".

### E7.1 Expressions with placeholders in high-order position in value of VAL in output

%% Examples include generics

They report the traffic conditions *15 minutes after the hour*.

They report the traffic conditions <TIMEX2 VAL="TXX:15" NON\_SPECIFIC="YES">15  
minutes after the hour</TIMEX2>.

%% Note that there is no embedded tag for "the hour".

The clock strikes 12 at *noon and midnight*.

The clock strikes 12 at <TIMEX2 VAL="T12:00" NON\_SPECIFIC="YES">noon</TIMEX2>  
and <TIMEX2 VAL="T24:00">midnight</TIMEX2>.

%% Placeholders for date components of value are implied rather than explicit. Granularity of VAL is

indicated by the terms “noon” and “midnight”.

That emergency clinic is open from 7 *p.m.* to 7 *a.m.*

That emergency clinic is open from <TIMEX2 VAL="T19" NON\_SPECIFIC="YES">7  
p.m.</TIMEX2> to <TIMEX2 VAL="T07" NON\_SPECIFIC="YES">7 a.m.</TIMEX2>  
%% Placeholders for date components of value are implied rather than explicit.

*April* is usually wet.

<TIMEX2 VAL="XXXX-04" SET="YES" NON\_SPECIFIC="YES">April</TIMEX2> is  
usually wet.  
%% The set is indicated by “usually”.

We always watch the game on *Super Bowl Sunday*.

We always watch the game on <TIMEX2 VAL="XXXX-WXX-7" SET="YES"  
GRANULARITY="G1D" NON\_SPECIFIC="YES">Super Bowl Sunday</TIMEX2>.  
%% The set is indicated by “always”.

I love *December*.

I love <TIMEX2 VAL="XXXX-12" NON\_SPECIFIC="YES">December</TIMEX2>.

*Christmas* is celebrated in *December*.

<TIMEX2 VAL="XXXX-12" NON\_SPECIFIC="YES">Christmas</TIMEX2> is celebrated in  
<TIMEX2 VAL="XXXX-12">December</TIMEX2>.  
%% Christmas is pinned down to December because of the sentence context, not because of the annotator’s  
world knowledge.

## E7.2 Expressions with placeholder in non-initial position of value of VAL in output

%% Examples include indefinites that are used referentially

The election took place on a *Tuesday*.

%% Assume for this example that 1998 is implied.

The election took place on <TIMEX2 VAL="1998-WXX-2" NON\_SPECIFIC="YES">a  
Tuesday</TIMEX2>.

They were sitting on the porch *early one Friday night in fall 1998*.

They were sitting on the porch <TIMEX2 VAL="1998-FA-WXX-5-TNI" MOD="EARLY"  
NON\_SPECIFIC="YES">early one Friday night in <TIMEX2 VAL="1998-FA">fall  
1998</TIMEX2></TIMEX2>.

%% This examples requires the special format that combines week-based and month-based values.

He left on a *sunny day in June*.

%% Assume for this example that 1999 is implied.

He left on <TIMEX2 VAL="1999-06-XX" NON\_SPECIFIC="YES"  
GRANULARITY="G1D">a sunny day in <TIMEX2 VAL="1999-  
06">June</TIMEX2></TIMEX2>.

%% This is a case where the placeholder appears in the right-hand position of the value. It means that  
although the value for the day is not given, the degree of precision of the expression is explicit (it is  
precise to the level of a day) and there is some particular day that the speaker has in mind.

### E7.3 Expressions without a value for VAL in output

%% Examples include non-referential, non-generic NPs

Today is *a historic day for the European enterprise*.

<TIMEX2 VAL="1999-07-15">Today</TIMEX2> is <TIMEX2 NON\_SPECIFIC="YES">a historic day for the European enterprise</TIMEX2>.

Today is *a sunny day*.

<TIMEX2 VAL="1999-07-15">Today</TIMEX2> is <TIMEX2 NON\_SPECIFIC="YES">a sunny day</TIMEX2>.

Saddam will not cave in at *the last minute*.

Saddam will not cave in at <TIMEX2 NON\_SPECIFIC="YES">the last minute</TIMEX2>.

%% This is another idiom.

we will give you as many answers as we can, as soon as we can, at *the appropriate time*, consistent with our obligation to also cooperate with the investigations.

%% This example is from a voice transcript.

we will give you as many answers as we can, as soon as we can, at <TIMEX2 NON\_SPECIFIC="YES">the appropriate time</TIMEX2>, consistent with our obligation to also cooperate with the investigations.

%% This expression is taken to be essentially synonymous with “*some appropriate time*”, i.e., it’s not really definite.

He took office a year ago with a pledge to reshape the organization to cope with the challenges of *a new century*.

He took office <TIMEX2 VAL="1998">a year ago</TIMEX2> with a pledge to reshape the organization to cope with the challenges of <TIMEX2 NON\_SPECIFIC="YES">a new century</TIMEX2>.

Bars were packed with people intent on taking part in the ritual of the *New Year's Day* hangover.

Bars were packed with people intent on taking part in the ritual of the <TIMEX2 NON\_SPECIFIC="YES">New Year's Day</TIMEX2> hangover.

In *an era of high joblessness and meager defense budgets*, European governments appear reluctant to devote greater resources to coping with distant or murky threats.

In <TIMEX2 NON\_SPECIFIC="YES">an era of high joblessness and meager defense budgets</TIMEX2>, European governments appear reluctant to devote greater resources to coping with distant or murky threats.

*A day where the investigation of the new allegations against President Clinton have taken some very interesting and startling turns.*

%% This is the first sentence of a news story.

<TIMEX2 NON\_SPECIFIC="YES">A day where the investigation of the new allegations against President Clinton have taken some very interesting and startling turns</TIMEX2>.

I always vote on *Election Day*.

I always vote on <TIMEX2 SET="YES" GRANULARITY="G1D"  
NON\_SPECIFIC="YES">Election Day</TIMEX2>.

%% The set is indicated by "always". Note that the sentence does not provide any value for PERIODICITY,  
so that attribute is absent.

Meanwhile, *each passing day* brings new evidence of collusion.

Meanwhile, <TIMEX2 SET="YES" PERIODICITY="F1D" NON\_SPECIFIC="YES"  
GRANULARITY="G1D">each passing day</TIMEX2> brings new evidence of collusion.

He said that in *recent days* he has been on the phone to Paris to speak with Perez de Cuellar.

He said that in <TIMEX2 SET="YES" GRANULARITY="G1D">recent days</TIMEX2> he  
has been on the phone to Paris to speak with Perez de Cuellar.

## E8 Miscellaneous

### E8.1 Expressions whose value cannot be adequately captured in the representation

The foreign minister responded in *less than one hour*.

%% This could represent an offset or a duration; in the larger context of the article in which this sentence  
appeared, the offset reading is clearly intended.

The foreign minister responded in <TIMEX2>less than one hour</TIMEX2>.

The almanac gives advice that applies only to *months with an 'r' in them*.

The almanac gives advice that applies only to <TIMEX2 SET="YES"  
GRANULARITY="G1M">months with an 'r' in them</TIMEX2>.

%% We cannot adequately express the membership of this set.

He was a DARPA program manager in the '60s and again in the '80s. *Those were the most  
exciting years of his life*.

He was a DARPA program manager in <TIMEX2 VAL="196">the '60s</TIMEX2> and again  
in <TIMEX2 VAL="198">the '80s</TIMEX2>. <TIMEX2 SET="YES"  
GRANULARITY="G1Y">Those</TIMEX2> were <TIMEX2 SET="YES"  
GRANULARITY="G1Y">the most exciting years of his life</TIMEX2>.

%% We cannot express the disjunction of decades in the VAL for the pronoun "Those" (and for the  
predicate nominal phrase).

She has been going to the gym to play basketball for *the past few years*.

She has been going to the gym to play basketball for <TIMEX2 VAL="PXY">the past few  
years</TIMEX2>.

%% Representation does not capture any meaning for "past few" (nor for its near-synonym, "recent").

## E8.2 Expressions whose value for VAL is null or underspecified in the absence of outside evidence

The relative success of the NATO-led peacekeeping mission in Bosnia has emboldened the United States to promote NATO initiatives for the Balkans, the Middle East and Africa that would have been unthinkable in *Cold War days*.

The relative success of the NATO-led peacekeeping mission in Bosnia has emboldened the United States to promote NATO initiatives for the Balkans, the Middle East and Africa that would have been unthinkable in <TIMEX2>Cold War days</TIMEX2>.

%% In annotator's judgment, this expression is idiomatic, and is therefore not represented as a set (of days).

Secretary of State Madeleine K. Albright declared that banishing the weapons of mass destruction of rogue states should become the new "unifying threat" that binds Europe and the United States in *the post-Cold War era*.

Secretary of State Madeleine K. Albright declared that banishing the weapons of mass destruction of rogue states should become the new "unifying threat" that binds Europe and the United States in <TIMEX2>the post-Cold War era</TIMEX2>.

Albright cited weapons proliferation as "the overriding security interest of *our time*."

Albright cited weapons proliferation as "the overriding security interest of <TIMEX2>our time</TIMEX2>."

In *the era of AIDS*, this is a phenomenon worthy of a close look.

In <TIMEX2>the era of AIDS</TIMEX2>, this is a phenomenon worthy of a close look.

My cousins came to visit on *Christmas Day*.

My cousins came to visit on <TIMEX2>Christmas Day</TIMEX2>.

We celebrated *Washington's birthday* by going shopping.

We celebrated <TIMEX2>Washington's birthday</TIMEX2> by going shopping.

We elected a new president on *the day of our last meeting*.

We elected a new president on <TIMEX2>the day of our last meeting</TIMEX2>.

The mood was somber during *the final day of Clinton's winter vacation*.

The mood was somber during <TIMEX2>the final day of Clinton's winter vacation</TIMEX2>.

Casualties were high during *the last year of WWII*.

Casualties were high during <TIMEX2>the last year of WWII</TIMEX2>.

We will have barbecued turkey on *Thanksgiving this year*.

We will have barbecued turkey on <TIMEX2 VAL="1999">Thanksgiving this year</TIMEX2>.

On *Super Bowl Sunday, 1999*, my in-laws arrived for an extended visit.

On <TIMEX2 VAL="1999-WXX-7">Super Bowl Sunday, 1999</TIMEX2>, my in-laws arrived for an extended visit.

### **E8.3 Expressions involving comparatives whose semantics cannot be captured in the representation**

He is expected to meet with the deputy prime minister *later this afternoon*.

He is expected to meet with the deputy prime minister <TIMEX2 VAL="1999-07-15-TAF">later this afternoon</TIMEX2>.

He went shopping for new shoes *earlier in the week*.

He went shopping for new shoes <TIMEX2 VAL="1999-W28">earlier in the week</TIMEX2>.

# Appendix F: Examples of Non-Markable Expressions

## F1 Introduction

In the course of developing the guidelines, we encountered a number of “negative cases”, i.e., examples of non-markable, time-related expressions. Currently, the set of examples includes a number of single-word adverbs and adjectives, plus the short phrase “too often”. These are offered to annotators and system developers for their information in determining the boundary between markable and non-markable expressions.

## F2 Point

The world watched the U.S.-led bombing of Iraq in 1991 and the *subsequent* freeing of Kuwait from Iraqi occupation.

The truth will come out, *eventually*.

Iraqi leader Saddam Hussein, *meanwhile*, sent a message to Russian President Boris Yeltsin.

Duisenberg is a former Dutch central banker.

## F3 Duration

The prospects for resuscitating the peace process have appeared grim *lately*.

American appeals to stretch alliance interests beyond NATO's territorial domain reflect a *longstanding* argument between the United States and Europe over "out of area" activities.

A recent study by the Western European Union, a defense-oriented affiliate of the European Union, found that its 10 member nations were so feeble in projecting military power that they could not sustain *long-term* deployment of more than one division or three brigades.

## F4 Frequency

He has been in *frequent* touch with the Iraqi government.

She has dealt gracefully with the *frequently* tedious tasks that have been assigned.

*Too often*, the U.S. takes the heat for dealing with significant issues.

April is *usually* wet.

We *always* watch the game on Super Bowl Sunday.

## F5 Proper names that contain or comprise a time expression

%%These proper names name entities (rather than days, months, etc.), and are not to be tagged as time expressions at all.

We infiltrated the terrorist group *Black September*.

I had to read George Orwell's *1984* in high school.

Have you ever heard of the *21st Century Party*?