
TIDES

2005 Standard for the Annotation of Temporal Expressions

April 2005

Lisa Ferro, Laurie Gerber, Inderjeet Mani, Beth Sundheim, George Wilson

Contact: lferro@mitre.org, sundheim@spawar.navy.mil

Approved for Public Release; Distribution Unlimited. 03-1046

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.

MITRE

Table of Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Background..... | 1 |
| 1.2 | Related Work..... | 3 |
| 1.3 | Changes Since 2003 Version | 4 |
| 1.4 | How This Manual is Organized | 4 |
| 2 | Why Annotate Text?: A Brief Overview for Human Annotators | 5 |
| 3 | Determining What Kinds of Expressions to Annotate | 6 |
| 3.1 | Markable Expressions..... | 6 |
| 3.2 | Non-Markables | 7 |
| 3.2.1 | Non-Markable Parts of Speech: Prepositions and Subordinating Conjunctions | 8 |
| 3.2.2 | Non-Markable Point and Duration Expressions | 8 |
| 3.2.3 | Non-Markable Frequency Expressions..... | 9 |
| 3.2.4 | Non-Markable Proper Names | 10 |
| 4 | Capturing the Meaning of Temporal Expressions..... | 11 |
| 4.1 | Annotation Format..... | 11 |
| 4.2 | Precise Temporal Expressions | 12 |
| 4.2.1 | Calendar Dates..... | 12 |
| 4.2.1.1 | Decades, Centuries, Millennia, and BCE: Points not Accommodated by ISO | 13 |
| 4.2.2 | Times of Day | 15 |
| 4.2.2.1 | Time Zones | 16 |
| 4.2.3 | Units of Weeks (Week-Based Format)..... | 17 |
| 4.2.4 | Duration | 17 |
| 4.2.4.1 | Fractions and Decimals..... | 21 |
| 4.2.4.2 | Decades, Centuries, and Millennia: Periods not Accommodated by ISO | 21 |
| 4.2.5 | Choosing Between Point and Duration Format | 22 |
| 4.3 | Fuzzy Temporal Expressions..... | 22 |
| 4.3.1 | Indeterminate Precision | 22 |
| 4.3.2 | Past, Present, Future | 23 |
| 4.3.2.1 | Choosing a Reference Time..... | 26 |
| 4.3.3 | Seasons | 28 |
| 4.3.4 | Fiscal Years | 30 |
| 4.3.5 | Yearly Quarters and Halves..... | 30 |
| 4.3.6 | Weekends | 31 |
| 4.3.7 | Morning, Afternoon, and Night | 32 |
| 4.3.8 | Unspecified Components of Calendar Dates and Times of Day..... | 33 |
| 4.3.9 | Unspecified Plural Numbers..... | 34 |
| 4.3.10 | Combined Week-Based and Month-Based Expressions..... | 35 |
| 4.4 | Modified Temporal Expressions..... | 36 |
| 4.4.1 | Anchoring Approximate Offsets..... | 38 |
| 4.5 | Set-Denoting Temporal Expressions | 39 |
| 4.5.1 | Example Annotations of Sets of Regularly Recurring Times | 39 |
| 4.5.2 | Example Annotations of Sets of Irregularly Recurring Times..... | 41 |
| 4.6 | Non-Specific Temporal Expressions | 41 |
| 4.6.1 | Non-Specific Point Expressions | 43 |
| 4.6.2 | Non-Specific Durations | 44 |
| 4.6.3 | Non-Specific TOD Expressions..... | 44 |
| 4.6.4 | Non-Specific Set-Denoting Expressions..... | 45 |
| 4.6.5 | Idioms and Metaphors Containing Temporal Expressions | 45 |
| 4.7 | Pronouns and Elided Elements | 45 |
| 4.8 | Event-Anchored Temporal Expressions | 46 |
| 4.8.1 | Birthdays and Anniversaries..... | 49 |
| 4.9 | Culturally-Determined Expressions..... | 50 |
| 4.9.1 | Alternative Calendars | 50 |

| | | |
|----------|---|-----------|
| 4.9.2 | Other Expressions Open to Historical and Individual Interpretation..... | 51 |
| 4.9.3 | Holidays..... | 52 |
| 4.10 | Expressions Whose Values can Change | 52 |
| 4.11 | Metonymic Expressions | 53 |
| 4.12 | Expressions You Just Can't Figure Out..... | 54 |
| 4.12.1 | Use that Comment Field!..... | 54 |
| 5 | Determining the Extent of the Annotations..... | 56 |
| 5.1 | Lexical Criteria | 56 |
| 5.2 | Syntactic Criteria | 56 |
| 5.2.1 | Appositives | 58 |
| 5.2.2 | Range Expressions..... | 58 |
| 5.2.3 | Conjoined Expressions | 58 |
| 5.2.4 | Embedded Expressions | 59 |
| 5.2.4.1 | When to Create One Tag..... | 59 |
| 5.2.4.2 | When to Create Multiple Tags, with Embedding..... | 61 |
| 5.2.4.3 | When to Create Multiple Tags, without Embedding..... | 62 |
| 5.3 | Punctuation | 63 |
| | Glossary..... | 64 |
| | References | 66 |
| | Appendix A: An Example Annotated Document..... | 68 |
| | Appendix B: Future Directions and Open Issues for Set Expressions..... | 69 |
| | Index | 71 |

List of Tables

| | | |
|-----------|--|----|
| Table 3-1 | Sample Lexical Triggers and Non-Triggers | 6 |
| Table 4-1 | TIMEX2 Tag Attributes | 12 |
| Table 4-2 | Anchor Attributes..... | 17 |
| Table 4-3 | Present, Past, Future Tokens | 23 |
| Table 4-4 | Season Tokens..... | 28 |
| Table 4-5 | Fiscal Year Token | 30 |
| Table 4-6 | Yearly Quarters & Halves Tokens | 30 |
| Table 4-7 | Weekend Token | 31 |
| Table 4-8 | Part-of-Day Tokens | 32 |
| Table 4-9 | Modifier Tokens..... | 36 |

1 Introduction

During the 1990s, the field of computational linguistics achieved several practical advances through the construction of annotated corpora for developing and testing various technologies. This document describes a set of guidelines for annotating temporal expressions with a normalized representation of the times they denote. This work has been carried out in support of a number of research activities under the Defense Advanced Research Projects Agency (DARPA) Translingual Information Detection, Extraction, and Summarization (TIDES) research program (<http://www.darpa.mil/iao/TIDES.htm>) and the Automatic Content Extraction (ACE) program (<http://www.nist.gov/speech/tests/ace/index.htm>). Research which can directly benefit from this document includes question answering (e.g., answering “when” questions or time-dependent questions), relation and event characterization and tracking, visualization of events on timelines, and production of biographical summaries. Other research related to information extraction can also benefit.

This annotation standard specifies more details of semantic representation than the TIMEX recognition tasks used in past 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.

1.1 Background

This standard was originally developed during the year 2000 under the TIDES program, and was first documented in Ferro et al. (2000). The annotation scheme calls for a single XML tag, <TIMEX2>, to be applied to temporal expressions, so we often refer to the standard as “the TIMEX2 standard.” During the same time period that the TIMEX2 standard was being developed, Mani and Wilson (2000) created an automatic time tagger, TEMPEX, which generates TIMEX2 tags in text documents.

Based on the early version of the standard, annotation studies were conducted and reported in Mani et al. (2001). In the spring of 2001, a revised draft of the annotation guidelines was produced (Ferro et al. 2001) as well as a complementary instruction manual for human

annotators (Ferro 2001). From this, two annotated corpora were manually prepared from Spanish and English dialogues, which together form the TIDES Temporal Corpus (Gerber et al. 2002). This corpus includes the following:

- A parallel corpus of 95 Spanish dialogs, their English translations, and temporal annotations of all the dialogs and their translations (44,081 words of raw text in all). The raw Spanish dialogs are part of the Enthusiast corpus collected earlier at Carnegie-Mellon University under U.S. government funding.
 - 193 documents of the Linguistic Data Consortium's English TDT-2 corpus annotated with temporal annotations (171,535 words of raw text).

To allow for the easy manual application of the TIMEX2 standard, a special data entry window was created for MITRE's Alembic Workbench (AWB) annotation tool, which is available at no cost from <http://www.mitre.org/tech/alembic-workbench/>. The "TIMEX2 widget" for AWB was first described in Ferro et al. (2002).

More recently, MITRE's newest annotation tool, Callisto, has been configured with a TIMEX2 Task environment to allow data markup according to this annotation standard. This Java-based software runs in Windows and Unix, and is available at no cost from <http://callisto.mitre.org>.

Also in 2001, the Relation Detection and Characterization (RDC) task in the ACE program adopted the TIMEX2 standard for capturing timestamps of relations between entities. In support of the RDC evaluation, 180,000 words of newswire, newspaper, and transcribed broadcast news texts were annotated with TIMEX2 tags, using the annotation tool mentioned above. Some modifications and extensions were made to the standard in order to better support the characterization of ACE relations, as documented in Gerber et al. (2002).

In the summer of 2002 a corpus of reading comprehension tests (each test consisting of a short non-fiction text and five questions) were annotated with the TIMEX2 standard (Wellner 2003). The TIMEX2 standard is also currently employed by the MiTAP (MITRE Text and Audio Processing) system developed under TIDES (Damianos et al. 2002). MiTAP (<http://mitap.sdsu.edu/>) is a prototype system for monitoring infectious disease outbreaks and other global threats. Temporal expressions are one of the many critical types of information that are extracted, processed by the TEMPEX system according to the TIMEX2 standard.

It has become too costly to maintain two separate documents, namely, the annotation guidelines of Ferro et al. (2001) and the instruction manual of Ferro (2001). The 2003 guidelines therefore represented an update of the instruction manual with some material merged in from the original annotation guidelines. In addition, the 2003 version contained new examples and additional discussions which were deemed important by the ACE 2002 annotators. The 2003 version also specified additional extensions to the scheme itself to expand its coverage of temporal expressions.

[[Discuss Callisto]]

[[In 2004, TERN...

Iagr results

Chinese version]]

Significant changes to the standard since the 2003 documentation are summarized in section below in section 1.3.

1.2 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 1995), 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. In addition, the current scheme 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 standard has had to be extended in various ways.

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). 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.

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.

Another notable effort in temporal annotation is the project known as TERQAS (Time and Event Recognition for Question Answering Systems). This six-week workshop funded by ARDA developed the TimeML standard (Ingria and Pustejovsky, 2002), which provides an annotation scheme for identifying events mentioned in a text document and orienting them on a timeline.

The scheme is also designed to capture much of the lexical and syntactic information that is hypothesized to be necessary to compute or infer the order of events as they are described in natural language. The TimeML scheme includes a TIMEX3 tag, which represents a variation of the TIMEX2 scheme.

1.3 Changes Since 2003 Version

As discussed above, the upgrades for this version focus on capturing the policies that were followed during the TERN 2004 annotation, and on clarifying former policies. Very little has changed that will affect system or human output:

- Added a PM token to the Part-of-Day Tokens. See Table 4-8 Part-of-Day Tokens.
- Eliminated the NON_SPECIFIC attribute, though the expressions themselves are still taggable. See section 4.6. Note that this attribute was ignored for TERN 2004 as well, though it still lingered in the guidelines.
- Allow predicate nominals to be normalized. See 4.6.1 Non-Specific Point Expressions.
- Idiomatic and metaphorical temporal expressions should use the X placeholder in VAL. See section 4.6.5.

1.4 How This Manual is Organized

Immediately following this introduction is a brief overview for those readers who might be new to the task of manually annotating text in support of NLP research. Following that, the heart of the manual is organized into three main sections:

1. Determining What Kinds of Expressions to Annotate
2. Capturing the Meaning of Temporal Expressions
3. Determining the Extent of Annotations

The first section introduces you to the types of expressions that require annotation. The second section then provides details on how to systematically encode the meaning behind each temporal expression. To do this, we start with the simplest types of expressions and gradually work up to those requiring more complex representations. Finally, the third section provides guidance on determining the textual boundaries and constituency of each annotated expression. A glossary at the end provides definitions for terms found in ***bold italic*** throughout the manual, and an appendix contains a completely annotated sample document.

2 Why Annotate Text?: A Brief Overview for Human Annotators

As a human, you generally understand your native language with little conscious effort. To do this, you not only use your knowledge of your language, but also your knowledge of the world. To build a computer system that can approximate what a human does in this regard, it is helpful to give actual language samples to people and ask them to use annotations to record their understanding of the text. The resulting annotated text can then be used in a variety of ways to support the building of a computer system. This document will teach you how to annotate natural language text for this purpose.

Our goal in the current task is the interpretation of expressions that refer to time. Such expressions tell us *when* something happened, or *how long* something lasted, or *how often* something occurs. Such expressions also often require knowledge of the temporal context in order to truly understand them. As a human, you are always aware of your temporal location. You generally know what day it is, or if you're fortunate enough to be on vacation, you at least know what year and month it is, and have a rough idea of how far along in the month you are. So, when you refer to a particular time or day, you usually say or write things like "on October sixteenth" "on "Tuesday" or "next week" or "two days from now." Even newspaper writing, which is fairly formal, uses these context-dependent expressions in the body of a story. For all of these expressions, one needs to understand the temporal context in order to determine which particular date and time is being referred to. But the complexity of temporal expressions doesn't end there. Sometimes the narrative text isn't framed in the present – it may have been written in the past, or may even be describing some hypothetical future – but there are generally clues in the text to orient you, such as a date line on an old newspaper article. Sometimes you're not meant to determine the precise calendar date, particularly in fictional narration. And then there are various expressions referring to periods of time, such as "three months," where it is the duration and not the precise endpoints that are important to the text. We want computer systems to accurately interpret and distinguish all these temporal expressions.

A secondary objective in interpreting temporal expressions is more of a housekeeping task: the normalization of temporal expressions. We want computers to be able to share the information they obtain. If a programmer in France encodes "October sixteenth" as "1962.10.16" and one in the U.S. encodes it as "10/16/1962," it will appear as if two different dates are being referenced. Thus, the standards presented in this document will require that the same meaning is always encoded in the same way.

3 Determining What Kinds of Expressions to Annotate

As explained in the introduction, we are interested in temporal expressions. Such expressions can reference *calendar dates*, times of day (*TOD*), or *durations* (such as periods of hours, days, or even periods of centuries). Basically, if a phrase or word refers to some area on a timeline, we want to capture its meaning.

3.1 Markable Expressions

Markable expressions are the expressions that should be annotated. To be markable, the syntactic head of the expression must be an appropriate *lexical trigger*. Each lexical trigger is a word or numeric expression whose meaning conveys a temporal unit or concept, such as “day” or “monthly.” Furthermore, to be a trigger, the referent must be able to be oriented on a timeline, or at least oriented with relation to a time (past, present, future).

The following list contains a sampling of lexical triggers. The table also includes examples of closely related temporal concepts that are not considered triggers; these are discussed further in section 3.2.

Table 3-1 Sample Lexical Triggers and Non-Triggers

| Part of Speech | Lexical Triggers | Non-Triggers |
|----------------------------------|---|---|
| Noun | minute, afternoon, midnight, day, night, weekend, month, summer, season, quarter, year, decade, century, millennium, era, semester, [the] future, [the] past, time, period, point | instant, jiffy, episode, occasion, tenure, timetable, reign, light year, megawatt hour, lifetime, history |
| Proper name | (unique identifier for temporally-defined events:) Monday, January, New Year’s Eve, Washington’s Birthday, Solstice | |
| Specialized time patterns | 8:00, 12/2/00, 1994, 1960s | |
| Adjective | recent, former, current, future, past, daily, monthly, biannual, semiannual, daytime, daylong, onetime, ago, preseason, short-term, long-term | early, ahead, next, subsequent, frequent, perpetual, later, contemporary, simultaneous, preceding, previous, existing, modern |

| | | |
|---------------------------|--|---|
| Adverb¹ | currently, lately, hourly, daily, monthly, ago | earlier, immediately, instantly, forthwith, meanwhile, heretofore, previously, next, beforehand, following, later soon, sooner, shortly, eventually, occasionally, once, still, again, timely, whenever |
| Time noun/adverb | now, today, yesterday, tomorrow | |
| Number | 3 (as in “He arrived at 3.”), three, fifth (as in referring to “the fifth of June”), Sixties (as in referring to the decade “the Sixties”) | |

Included nouns are of course pronouns (such as “that,” “then,” and “it”) that can co-refer with a markable time expression. See the section “Pronouns and Elided Elements” on page **Error! Bookmark not defined.** for examples.

Note also the same word can have different meanings and even functions, so that the same term can be a lexical trigger in some contexts and not others. For example, the term “now” is used temporally and is a trigger in “We need to leave now.” But it is not a trigger in “Now, don’t blame me. I told the president of the risks involved.”

3.2 Non-Markables

In contrast to the triggers presented above, the non-triggers (i.e., the non-markables), although they are temporal in their semantics, are *as a class* less amenable to being pinned down to a timeline. For purely practical reasons, this ability to orient an expression in time is the basic distinction we rely on in constraining the scope of markable expressions. Of course, even the designated markables can be non-referring (and thus impossible to relate to a timeline), as in idiomatic phrases like “burn the midnight oil.”

Likewise, note that adjective non-triggers are permitted within the extent of a markable expression, as in “early years” or “the next day” or “hours earlier.” They are not markable on their own, as in “the early meeting.” Starting with the 2005 guidelines we are also allowing the extent to include non-triggers that are syntactic heads of markable expressions when the semantic head is a trigger word; e.g., partitives like “much of the week”. See section 5.2.4.1 on page 59.

¹ If the adjective form of a word is a trigger, then an adverbial form will also be a trigger.

3.2.1 Non-Markable Parts of Speech: Prepositions and Subordinating Conjunctions

Table 3-1 shows that some parts of speech contain both triggers and non-triggers. In contrast, subordinating conjunctions (which introduce clauses) and prepositions (which introduce noun phrases) are two parts of speech that are never triggers; that is, they never appear as the syntactic head of an annotated expression. The following table shows some examples of these parts of speech.

Table 3-2 Sample Non-Triggers

| Part of Speech | Non-Triggers |
|---------------------------|---|
| Subordinating Conjunction | when, while, as, since, now [that], as long as, as soon as, sooner than, every time, any time |
| Preposition | at, on, in, for, over, throughout, during, before, after, since, until |

3.2.2 Non-Markable Point and Duration Expressions

Each of the expressions illustrated here only vaguely indicates a point in time (calendar date or time of day), or references some vague duration (interval) of time. Although many *point* and *duration* expressions are markable, the ones illustrated here are not.

Sequencing and ordering expressions are not markable:

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

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*.

But even on Thursday, there were signs of potential battles *ahead*.

These expressions are admittedly borderline in terms of markability. With the anchoring attributes introduced later in this version of the TIMEX2 scheme, it might well be feasible to capture some of the semantics of these terms. For example:

But even on <TIMEX2 VAL="1999-07-22">Thursday</TIMEX2>, there were signs of potential battles
<TIMEX2 VAL="FUTURE_REF" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-07-22">ahead</TIMEX2>.

There are many ordering/sequencing terms that would need to be more fully explored before they could be added to the list of markables.² Thus, although such terms are not “officially” markable, we encourage TIMEX2 users to explore this possibility and report their results to the NLP community.

Manner adverbs, which say how soon or how quickly something is done, are not markable:

Across Europe, banks are converting their systems to offer euro-denominated bank accounts *immediately* for any customers who want them.

I will be with you *momentarily*.

Non-quantifiable durations are not markable:

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.

Other examples in this class include *long*, *permanently*, and *temporarily*.

Negatives and references to non-existent times are not markable:

Five-time champion Rick Swenson wasted *no time* moving to the front of the pack in the Iditarod Trail Sled Dog Race.

No date has been set for another summit meeting.

“Time” when it means “Situation” or “Occasion” is not markable. Ironically, the trigger word “time” is one of the more difficult terms in terms of markability. In many cases it is not referring to a time interval. Instead, it means something more like “situation” or “occasion”, and so is not markable:

Conventional wisdom says that running mates don’t make a difference. Each side is hoping *this time* it will make a difference.

Whatever the Supreme Court decides *this time*, you say this case could come back before the U.S. Supreme Court again?

It is, however, often very hard to distinguish between markable and non-markable cases of “time.” Fortunately, although it lowers inter-annotator agreement and system scores a bit, neither the markable nor non-markable expressions are particularly useful for understanding the document as a whole, as they are usually not normalized.

3.2.3 Non-Markable Frequency Expressions

Bare frequencies (frequency expressions with no time period given) are not markable:³

It only happened *once*.

² Consider: ahead, next, subsequent, later, preceding, previous, previously, early, earlier, heretofore, so far, afterwards, before, after, beforehand, following, eventually, finally, [for] the first time, first, later, originally, previously, subsequently, then, already, yet, still, ever.

³ In contrast, frequencies whose semantics include a temporal unit, such as *daily*, are markable. See section 4.5.

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.

Other examples in this class include *three times*, *on five occasions*, *normally*, *usually*, *commonly*, *generally*, *constantly*, *often*, *repeatedly*, *sometimes*, *rarely*, and *never*.

3.2.4 Non-Markable Proper Names

Proper names that designate something other than a temporal entity but happen to correspond to lexical triggers are not markable. For example, the following list contains names of organizations and books. Since these are not temporal entities, they are not markable.

We infiltrated the terrorist group *Black September*.

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

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

Four Days In September is making the leap from the bookshelf to the movie screen.

However, triggers that are functioning as temporal modifiers within titles are markable. Examples include titles of conferences, "2001 National Conference for Agribusiness," and awards, "The Coach of the Year".

4 Capturing the Meaning of Temporal Expressions

This section illustrates how the semantics of markable expressions are captured in the annotations. As with all annotation efforts, it is important to capture the meaning of the expressions in a consistent manner, so a great deal of documentation is provided here.

The TIMEX2 standard is built upon the foundation of *ISO 8601: Information Interchange – Representation of Dates and Times*, Second Edition (1997), from the International Organization for Standardization (ISO). The document is available at <ftp://ftp.qsl.net/pub/g1smd/8601v03.pdf>. We assume the reader has complete familiarity with the ISO standard.

With temporal expressions, the “meaning” is very often a particular calendar date or *TOD* that you deduce from the context. We call these “precise” expressions and discuss these first. Next, we address “fuzzy” expressions, those where you have a general sense of the meaning, but cannot confidently posit a precise value. Following that, there are sections discussing modified expressions (“almost two o’clock”), *frequencies*, which are expressions denoting sets of recurring times (“every Tuesday”), and non-specific expressions like “a Tuesday.” The final subsections address a collection of cases that have proved problematic for annotators: event-anchored expressions, culturally determined expressions, expressions whose values can change, and metonymic expressions.

4.1 Annotation Format

Temporal expressions are annotated by inserting a special SGML (Standard Generalized Markup Language) tag around the text string. At the start of the expression, <TIMEX2> is inserted directly into the text, and at the end of the expression </TIMEX2> is inserted (the same tag, but with a forward slash). For example:

<TIMEX2>Halloween</TIMEX2>

In addition, the TIMEX2 tag may contain one or more attributes. There are six possible attributes:

Table 4-1 TIMEX2 Tag Attributes

| Attribute | Function | Example |
|------------|---|---------------------------|
| VAL | Contains a normalized form of the date/time. | VAL="1964-10-16" |
| MOD | Captures temporal modifiers. | MOD="APPROX" |
| ANCHOR_VAL | Contains a normalized form of an anchoring date/time. | ANCHOR_VAL="1964-10-16" |
| ANCHOR_DIR | Captures the relative direction/orientation between VAL and ANCHOR_VAL. | ANCHOR_DIR="BEFORE" |
| SET | Identifies expressions denoting sets of times. | SET="YES" |
| COMMENT | Contains any comments the annotator wants to add. | COMMENT="context garbled" |

Here's an example of what a finished tag can look like:

<TIMEX2 VAL="2000-10-31TNI" MOD="START">early last night</TIMEX2>

Each attribute has its own rules about what kind of values it can have. The attributes will be discussed and illustrated in greater depth below, as we introduce the types of expressions that make use of them.

4.2 Precise Temporal Expressions

A precise temporal expression is one in which you can confidently determine the *calendar date*, *TOD*, or *duration* that is intended. When annotating naturally occurring text and speech, you use the context of the document to make that determination. In all cases, you should specify values as fully as possible, but only within the bounds of what you can confidently infer.

Below we illustrate different kinds of unmodified expressions, each denoting a specific point or duration. Each receives a TIMEX2 tag with a VAL attribute, and the VAL attribute contains an ISO-compliant value, unless specified otherwise.

Because it would be unwieldy to provide the full context for all examples, you can assume, unless stated otherwise, that the reference date for all the examples in the remainder of this document is Thursday, July 15, 1999. In terms of weeks, that is week 28 of 1999.

4.2.1 Calendar Dates

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>.

A **range** expression is one with explicit begin and end points. Each point is annotated as a separate expression:

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>.

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>.

An **anchored point expression** can sometimes contain the expression it is anchored on. For example, in "*three years ago today*" the anchor is "*today*," and the larger expression "*three years ago today*" is considered anchored because it can be positioned on a timeline in relation to "*today*." Note that both the anchor and the larger expression are annotated:

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>.

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>.

There are two anchoring attributes, ANCHOR_VAL and ANCHOR_DIR, which you will use for certain anchored expressions. These will be discussed further in sections 4.2.4 and 4.3.2.

4.2.1.1 Decades, Centuries, Millennia, and BCE: Points not Accommodated by ISO

ISO treats "year" as an essentially fixed-width, four-place component (YYYY). In our approach, we do not require all four positions to be represented. We consider the first position to be 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" would represent "the 21st century"), or three (for example, "196" would represent "the 1960s").

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

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

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>.

Angkor Wat, the fabled *11th century* temple, is Cambodia's main tourist attraction.
Angkor Wat, the fabled <TIMEX2 VAL="10">11th century</TIMEX2> temple, is Cambodia's main tourist attraction.

We are entering what is popularly regarded as *the last year of this millennium*.
We are entering what is popularly regarded as <TIMEX2 VAL="1999">the last year of <TIMEX2 VAL="1">this millennium</TIMEX2></TIMEX2>.

Because the ISO standard assumes that years have four digits, years before the year 1000 (or after 1000 BCE) are written with leading 0s:

The story of Mu Lan, about a Chinese woman warrior, was written in *500 AD*.
The story of Mu Lan, about a Chinese woman warrior, was written in <TIMEX2 VAL="0500">500 AD</TIMEX2>.

Before Current Era (BCE) Dates. The ISO standard has no way to capture Before Current Era (BCE) dates, but along with others in the community we have adopted the use of a two-character token, BC, preceding the year portion of the ISO value. This allows us capture those dates of four digits or less. We also follow convention by pretending that there is a year 0 in the Judeo-Christian calendar, to make the arithmetic straightforward for computers.

Aristotle wrote his "Poetics" *2,300 years ago*.
Aristotle wrote his "Poetics" <TIMEX2 VAL="BC0301">2,300 years ago</TIMEX2>.⁴
The king lived 4,000 years ago.
The king lived <TIMEX2 VAL="BC2001">4,000 years ago</TIMEX2>.

Geologic Eras. In referring to the far distant past, scientific convention does not use the standard calendar and instead gives the time as an offset. We adopt the same two-letter abbreviations used by geologists, and place them in initial position in VAL.

KA, some number of thousand years ago
MA, some number of million years ago
GA, some number of billion years ago

For example:

The Jurassic Period began *210 million years ago*.
The Jurassic Period began <TIMEX2 VAL="MA210">210 million years ago</TIMEX2>.

⁴ Note that the granularity of VAL reflects the granularity of the lexical expression, as is expected (see the Granularity Rule in section 4.3.1 for more details). However, in both these BCE examples one wonders if the writer was using round numbers and did not intend them to be taken quite so literally. In such cases the human annotator is encouraged to use the COMMENT attribute to make note of ambiguity.

4.2.2 Times of Day

If a **TOD** is indicated in the expression, the contents of VAL will reflect the TOD in standard ISO format. If you can determine the calendar date from the context of the document, then the YYYY-MM-DD portion of the ISO value is specified as well. Otherwise, just the time of day portion can be specified.

The sponsor arrived at *ten minutes to 3*.

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

<DATE_TIME>10/08/1998 21:36:42.85</DATE_TIME>⁵

<DATE_TIME><TIMEX2 VAL="1998-10-08T21:36:42.85">10/08/1998
21:36:42.85</TIMEX2></DATE_TIME>

I returned to work at *twelve o'clock January 3, 1984*.⁶

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

Hickory dickory dock. The mouse ran up the clock. The clock struck *one*, the mouse ran down. Hickory dickory dock.⁷

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.

On the *nineteenth* I am in class until *eleven in the morning*.⁸

On <TIMEX2 VAL="1999-07-19">the nineteenth</TIMEX2> I am in class until <TIMEX2 VAL="1999-07-19T11:00">eleven in the morning</TIMEX2>.

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.

At <TIMEX2 VAL="1998-12-31T23:59">11:59 p.m.</TIMEX2>, Mayor Rudolph W. Giuliani sat on a stage at 45th Street and Broadway and pushed a button. <TIMEX2 VAL="1998-12-31T24:00:00">Sixty seconds later</TIMEX2>, the ball atop 1 Times Square completed its slow descent into retirement.⁹

Note that when the time referred to is at the top of the hour (*twelve o'clock*, *one*, *eleven in the morning*), the minute portion of VAL is given as :00 to ensure that it is not misinterpreted as what the ISO standard calls an expression of “reduced precision.”

For purposes of annotation consistency, we will always represent “midnight” as hour 24 of the earlier day:

I was up until *midnight*.

I was up until <TIMEX2 VAL="1999-07-14T24:00">midnight</TIMEX2>.

⁵ This example is taken from the transcript of a broadcast news story, where the header information typically contains a <DATE_TIME> element.

⁶ Assume for this example that we know that the time refers to noon.

⁷ Assume for this example that we know we're talking about 1 a.m., but we don't know the date.

⁸ This example is taken from a dialogue transcript.

⁹ The granularity of the first expression is in minutes, while the granularity of the second expression is in seconds.

4.2.2.1 Time Zones

In the preceding examples, no time zone was specified in the textual expressions, and thus no time zone was indicated in VAL. The general policy of this standard is that the time zone only needs to be indicated if the text overtly specifies it. The ISO standard is adhered to in the format of VAL when specifying a time zone. In the following example a “Z” indicates that the time is given in Universal Coordinated Time (UTC) or Greenwich Meridian Time (GMT).

April 11, 1996 11:13 GMT

<TIMEX2 VAL="1996-04-11T11:13Z">April 11, 1996 11:13 GMT</TIMEX2>

In this next example, “-05” indicates that the local time is five hours behind UTC. Time zone differences are limited to whole-hour increments.

January 21, 1994 08:29 Eastern Standard Time

<TIMEX2 VAL="1994-01-21T08:29-05">January 21, 1994 08:29 Eastern Standard Time</TIMEX2>

As seen in the following example, if an alternative time zone reference is given in parentheses following the local time, each expression is tagged separately.

September 12, 2001 Posted: 12:27 PM EDT (1627 GMT)

<TIMEX2 VAL="2001-09-12">September 12, 2001</TIMEX2> Posted: <TIMEX2 VAL="2001-09-12T12:27-04">12:27 PM EDT</TIMEX2> (<TIMEX2 VAL="2001-09-12T16:27Z">1627 GMT</TIMEX2>)

Occasionally, a time zone is overtly declared to have global scope over all time references in a document. In these cases, all the TIMEX2 tags contain the time zone designators:

Timeline of Terrorist Attacks --

8:45 a.m. (all times are EDT): A hijacked passenger jet crashes into the north tower of the World Trade Center.

9:03 a.m.: A second hijacked airliner crashes into the south tower of the World Trade Center and explodes.

...

Timeline of Terrorist Attacks --

<TIMEX2 VAL="2001-09-11T08:45-04">8:45 a.m.</TIMEX2> (all times are EDT): A hijacked passenger jet crashes into the north tower of the World Trade Center.

<TIMEX2 VAL="2001-09-11T09:03-04">9:03 a.m.</TIMEX2>: A second hijacked airliner crashes into the south tower of the World Trade Center and explodes.

...

If no time zone is specified globally or in the immediate expression (i.e., within the extent of the TIMEX2 tag), it does not need to be specified in VAL, even if a human annotator could conceivably look up and compute the time zone based on world knowledge or other information in the document. For example, in the following, “at 5 p.m.” is not given a time zone designator because the immediate expression has none. Only “1300 GMT” has a time zone indicated in VAL.

The unidentified attackers opened fire at 5:00 p.m. (1300 GMT).

The unidentified attackers opened fire <TIMEX2 VAL="1999-07-15T17:00">at 5:00 p.m.</TIMEX2>

(<TIMEX2 VAL="1999-07-15T13:00Z">1300 GMT</TIMEX2>).

We leave it up to users of the TIMEX2 scheme to use time zones designators in a wider range of cases if their application supports and requires it.

4.2.3 Units of Weeks (Week-Based Format)

The ISO standard also allows for specific weeks to be referenced and normalized. We use this notation when the *lexical trigger* is the word “week.”

He will visit Norway *next week*.

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

ISO also allows for a day of the week to be specified in the week-based format, e.g., “1999-W29-4” for Thursday of week 29 in 1999 (July 22, 1999, in other words), where Monday is day 1 and thus Thursday is day 4. However, remember this rule of thumb:

“Prefer Month-Based to Week-Based Format” Rule:

If an expression can be encoded equally in a month-based or a week-based format, use the month-based representation.

Thus, in an example like the following, although a day of the week is mentioned, the month-based format should be used rather than the week-based one:

Congress convenes *next Thursday*.

Congress convenes <TIMEX2 VAL="1999-07-22">next Thursday</TIMEX2>.

The week-based format is, however, useful in those cases where the day of the week is given in the linguistic expression, but there is not enough detail to use the month-based format. See section 4.3.8 Unspecified Components of Calendar Dates and Times of Day.

4.2.4 Duration

An expression of duration indicates a period of time, indicating **how long** something lasts, e.g., “*three hours long*.” If duration is expressed, VAL uses the format outlined in section 5.5.3.2 of the ISO standard, i.e.:

PnYnMnDTnHnMnS

PnW

Durations that refer to specific periods of time can be oriented or anchored with respect to other points or periods of time.¹⁰ The 2001 version of the TIMEX2 standard did not capture these relations at all; the 2002 version developed for the ACE program introduced ANCHOR attributes to anchor a limited set of durations. In the 2003 standard we expanded the ANCHOR attributes to capture a wider range of anchored durations.

The anchoring attributes and their values are shown in the following table:

Table 4-2 Anchor Attributes

| Attribute | Value | Example |
|-----------|-------|---------|
|-----------|-------|---------|

¹⁰ For durations that cannot be so anchored, see section 4.6.2 Non-Specific Durations.

| | | |
|------------|--|-------------------------|
| ANCHOR_DIR | WITHIN STARTING ENDING AS_OF ¹¹ BEFORE AFTER | ANCHOR_DIR="BEFORE" |
| ANCHOR_VAL | A normalization of the anchoring date/time in ISO format | ANCHOR_VAL="1999-07-15" |

The two anchoring attributes ANCHOR_VAL and ANCHOR_DIR are always used together.¹²

Keep in mind that the TIMEX2 annotation scheme is designed to aid in the representation of temporal relations expressed in a document. Without anchors, the normalization of duration expressions only conveys *how long* something happened, without any information about *when* this duration occurred. Thus, to maximize the information captured by the annotation, we encourage the use of anchor attributes wherever the human annotator feels confident in specifying them. The human annotator should consider the entire context of the document being annotated, and if an anchoring point can be confidently specified based on this information, the anchoring attributes should be included.

WITHIN, STARTING and ENDING. Here are some examples to illustrate the use and interpretation of the first three ANCHOR_DIR values; in each case, assume that it is clear from the context of the document that July 15, 1999 is the reference time (i.e., the time of the utterance).

The accounts are paid in full for *the six months ended March 31*.

The accounts are paid in full for <TIMEX2 VAL="P6M" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999-03-31">the six months ended <TIMEX2 VAL="1999-03-31">March 31</TIMEX2></TIMEX2>

He wrapped up a *three-hour* meeting with the Iraqi president in Baghdad *today*.

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

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

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

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

¹¹ See section 4.3.2 for illustrations of the AS_OF value, which is used only with the PRESENT_REF token described in that section.

¹² 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).

She is part of the most visible and influential presence that women have had in the <TIMEX2 VAL="P52Y" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999">52-year</TIMEX2> history of the United Nations.

Police said the *31-year-old* Britain died *Thursday*.

Police said the <TIMEX2 VAL="P31Y" ANCHOR_DIR="ENDING" ANCHOR_VAL="1990">31-year-old</TIMEX2> Britain died <TIMEX2 VAL="1999-07-15">Thursday</TIMEX2>¹³

In *the past four years*, I have written you in confidence about the various leaks that have occurred from your office.

In <TIMEX2 VAL="P4Y" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999">the past four years</TIMEX2>, I have written you in confidence about the various leaks that have occurred from your office.

The president attended the meeting for *the fifth straight day in a row*.

The president attended the meeting for <TIMEX2 VAL="P5D" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999-07-15">the fifth straight day in a row</TIMEX2>.

Italy's request will be made within *the next three days* via the foreign ministry.

Italy's request will be made within <TIMEX2 VAL="P3D" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999-07-15">the next three days</TIMEX2> via the foreign ministry.

He will be in school for *another year*.¹⁴

He will be in school for <TIMEX2 VAL="P1Y" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999-07-15">another year</TIMEX2>.

There has been a lot of rain *the past three weeks*.

There has been a lot of rain <TIMEX2 VAL="P3W" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999-W28">the past three weeks</TIMEX2>.

Pete Sampras continues to edge closer to being the world's best for the sixth straight year.

Pete Sampras continues to edge closer to being the world's best for <TIMEX2 VAL="P6Y" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999">the sixth-straight year</TIMEX2>.

Granularity of ANCHOR_VAL. In the first example given above, the anchoring date is explicitly March 31, so ANCHOR_VAL is specified to the 'day' granularity. Even when the anchoring date is not given explicitly, it is often clear from context. In other cases, like *52-year* and *the past four years*, it is unclear if the speaker intended the timeframe to extend from the day of the utterance or the year of the utterance. In this type of situation, the granularity of ANCHOR_VAL should be specified at the same granularity of the duration expression, in this case, in years. Also, when specifying an anchor time for durations with a STARTING or ENDING value, the ANCHOR_VAL unit should be contained within the duration unit. For example, in the following, the anchoring date of 1997-01 is within the duration (it is the first month of the duration):

The job rate increased during *the first nine months of 1997*.

The job rate increased during the <TIMEX2 VAL="P9M" ANCHOR_DIR="STARTING"

ANCHOR_VAL="1997-01">the first nine months of <TIMEX2 VAL="1997">1997</TIMEX2></TIMEX2>

Choosing between STARTING and ENDING. Generally the expression itself will dictate whether to anchor a duration at the start or the end. (e.g., *the last three years* uses ENDING, *the*

¹³ Note that an expression of age is only markable when it has a noun phrase headed by a trigger word such as "year." Other statements of age such as "Mary Smith, 61, was appointed CEO" are not markable.

¹⁴ "Another" and "an" are regularized as "1."

next three years uses STARTING). At times, the context of the document will provide both STARTING and ENDING points; all other things being equal, the STARTING value is preferred:

On *September 1, 1985* a team of deep-sea divers found the Titanic, which sank in *1912* on its first sea voyage. For *73 years*, no one had known exactly where the ship sank.
On <TIMEX2 VAL="1985-09-01">September 1, 1985</TIMEX2> a team of deep-sea divers found the Titanic, which sank in <TIMEX2 VAL="1912">1912</TIMEX2> on its first sea voyage. For <TIMEX2 VAL="P73Y" ANCHOR_DIR="STARTING" ANCHOR_VAL="1912">73 years</TIMEX2>, no one had known exactly where the ship sank.

Under no circumstances is ANCHOR_DIR to contain more than one value, nor should there be more than one ANCHOR_DIR attribute.

If the period is ongoing, or will never end, the STARTING value is preferred. The first case below is an example of an ongoing duration, as indicated by the present perfect form of the verb, “has not seen.” The second example is a case of a duration that will never end; that is, it will always be some number of months since the crises hit. Thus, in both these cases the annotator must count backwards from the reference time to compute the ANCHOR_VAL, and use the STARTING value in ANCHOR_DIR.

Abdulah is a father of three who has not seen his children for *six months*.
Abdulah is a father of three who has not seen his children for <TIMEX2 VAL="P6M" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999-01">six months</TIMEX2>.

In the two months since the crisis hit, the ruble has nosedived.
In <TIMEX2 VAL="P2M" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999-05">the two months since the crisis hit</TIMEX2>, the ruble has nosedived.

BEFORE and AFTER. For these next two examples illustrating the BEFORE and AFTER values, assume that the context does not provide any clue as to when the durations began or ended, but notice that the tense of the verb indicates that they are after and before the pretend reference date of July 15, 1999, respectively.

I will be on vacation in Scotland for *three weeks*.
I will be on vacation in Scotland for <TIMEX2 VAL="P3W" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-07-15">three weeks</TIMEX2>.

In the preceding example, the notation captures the meaning of period of three weeks occurring sometime after the reference time; in other words, a period of three weeks sometime in the future. In the following example, the notation captures the meaning that the period of three weeks occurred sometime in the past with regard to the reference date.

I was on vacation in Scotland for *three weeks*.
I was on vacation in Scotland for <TIMEX2 VAL="P3W" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15">three weeks</TIMEX2>.

Thus, the BEFORE and AFTER values are used when the end or begin points of a duration cannot be anchored precisely, but it is clear from context whether the span occurred before or after the reference time.

4.2.4.1 Fractions and Decimals

If periods of time are not given in whole numbers, it is often arithmetically equivalent to use decimal notation, e.g. “P6.5Y” for “six and a half years” or convert it to whole time units, e.g., “P6Y6M.” Both decimal and non-decimal VALs are thus allowed, but for the sake of consistency, the format should follow the linguistic expression wherever possible.

Following the federal sentencing guidelines, the judge gave him the maximum of *10 years and 1 month*.
Following the federal sentencing guidelines, the judge gave him the maximum of <TIMEX2 VAL="P10Y1M" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999">10 years and 1 month</TIMEX2>.

He spent *six and a half years* in prison.
He spent <TIMEX2 VAL="P6.5Y" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15">six and a half years</TIMEX2> in prison.

The National Hockey League imposed a freeze on trading over *the next two and a half weeks* as many of the National Hockey League's players compete at the Winter Olympics.
The National Hockey League imposed a freeze on trading over <TIMEX2 VAL="P2.5W" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999-07-15">the next two and a half weeks</TIMEX2> as many of the National Hockey League's players compete at the Winter Olympics.

The last example makes clear the benefit of allowing decimal notation. What, after all, is two and half weeks? Two weeks and three days? Two week and four days? Because this is often unclear, it best to take the expression literally and use the decimal notation in these cases.

4.2.4.2 Decades, Centuries, and Millennia: Periods not Accommodated by ISO

The ISO standard does not supply the means to capture periods of decades, centuries, and millennia. We have therefore added the following unit abbreviations:

Decade: DE
Century: CE
Millennium: ML

For the first two examples below, assume that it is not clear from the document context what the granularity of the anchor should be, so we follow the fall-back rule of using the granularity of the expression to determine the granularity of ANCHOR_VAL.

The recession lasted *a decade*.
The recession lasted <TIMEX2 VAL="P1DE" ANCHOR_DIR="ENDING" ANCHOR_VAL="199">a decade</TIMEX2>.

Peace reigned for *two millennia*.
Peace reigned for <TIMEX2 VAL="P2ML" ANCHOR_DIR="ENDING" ANCHOR_VAL="1">two millennia</TIMEX2>.

Charles reaches his *half-century* mark *Saturday*.
Charles reaches his <TIMEX2 VAL="P.5CE" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999-07-17">half-century</TIMEX2> mark <TIMEX2 VAL="1999-07-17">Saturday</TIMEX2>.

4.2.5 Choosing Between Point and Duration Format

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 was happy *five days ago*.
 He was happy <TIMEX2 VAL="1999-07-10">five days ago</TIMEX2>.

Duration: He was depressed for *five days*.
 He was depressed for <TIMEX2 VAL="P5D">five days</TIMEX2>.¹⁵

Thus, our word of caution here is to pay close attention to the context of the temporal expressions. It is often the context, and not the expression itself, which dictates the format of the annotation.

There are other instances in which an expression might be interpreted either way, as in the following example:

"I am busy *the hours that you are not*."¹⁶

This could be interpreted as set of points (in this case a set of hours) or a period of hours (a duration). Our rule of thumb is as follows:

"Prefer Point" Rule:

**If both a point and a duration interpretation is possible,
the point representation is preferred over the duration representation.**

4.3 Fuzzy Temporal Expressions

Fuzzy temporal expressions are markable expressions that are vague or have imprecise boundaries. In this section we discuss a number of strategies for coping with such imprecision while still capturing as much of the meaning of the expression as possible. These strategies include rules of interpretation, the use of *tokens* in VAL, and the use of the X placeholder in VAL. Each is explained in more detail below.

4.3.1 Indeterminate Precision

Unlike "a year ago today," which we saw earlier, expressions like "a year ago" are imprecise because they lack an explicit anchor. The writer could mean "a year ago today," (which would be 1998-07-15 if today were 1999-07-15) or just "last year" (1998). "In a week" could mean precisely seven days from now or any time in the following week.

¹⁵ This expression may or may not be anchorable, depending on the larger context, which is not provided here.

¹⁶ The annotation of set-denoting expressions, and this particular example, are illustrated further in "Example Annotations of Sets of Irregularly Recurring Times" on page 41.

To deal with these expressions, we invoke a rule of interpretation:

“Expression Granularity” Rule:

When no explicit anchor exists in the document context, use only the head noun to determine the precision of the VAL.

For example, if the string specifies “year,” only the YYYY portion of VAL is specified.¹⁷

Ann took office *a year ago*.

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

The Defense Minister referred to U.N. resolution 425 in a speech *six weeks ago*.

The Defense Minister referred to U.N. resolution 425 in a speech <TIMEX2 VAL="1999-W23">six weeks ago</TIMEX2>.

In *a year*, everyone’s salaries will be reviewed again.

In <TIMEX2 VAL="2000">a year</TIMEX2>, everyone’s salaries will be reviewed again.

Note however, that “decade” can often be interpreted as “10 years” and “century” as “100 years. See footnote 29 on page 37.

4.3.2 Past, Present, Future

Many temporal expressions refer in general terms to the past, the present, or the future. To handle these expressions, we have introduced an extension to the ISO standard that allows us to use alphabetical *tokens* that occupy the entire span of VAL. There are currently three such tokens, shown in the following table with some example expressions. The table also provides samples of non-markable expressions.

Table 4-3 Present, Past, Future Tokens

| Token | Sample Markable Expressions | Sample Non-Markable Expressions |
|-------------|---|---|
| PRESENT_REF | now [with the “nowadays” meaning] today [with the “nowadays” meaning] 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] (in) a few/couple days/months/weeks/years | ahead after soon, sooner shortly |

¹⁷ If the annotator feels that this rule results in a VAL that is less or more specific than the writer intended, this can be indicated in the COMMENT attribute discussed in section 4.12.

| | | |
|----------|---|---|
| | long-term, short-term | later eventually subsequent |
| PAST_REF | past yesterday [unless further resolvable to a particular date] former onetime lately recently long ago, a while ago, some time ago (<i>time unit</i> +) ago, e.g, a few days/months/weeks/years ago once ¹⁸ | before previously earlier beforehand once |

As with Table 3-1, it is important to note that some expressions in the “Non-Markable” column can be included in the extent of a markable, particularly “later” and “earlier”.

The mapping between expression and token is based on the reference time given in the document, rather than the time at which you are reading the document. For example, the day July 15, 1999 is in our past (unless this document was magically transported into a time before that). If you encounter a newspaper article dated July 15, 1999, which says, “Now is a good time to buy stock,” the “Now” is marked PRESENT_REF, even though that “now” is in our past.

In order to further anchor these time periods, the ANCHOR attributes introduced earlier for anchored durations (section 4.2.4) are used to orient the past/present/future to some reference time. Further examples are given next for each token.

For PRESENT_REF expressions, it is usually possible to give the general timeframe under which “present” should be interpreted. To do so, the tag is given an ANCHOR_VAL with the utterance’s reference time and an ANCHOR_DIR with a value of AS_OF. If the value for the reference time is highly specific, as when it comes from the DATE_TIME stamp on a transcribed broadcast news article (e.g., <DATE_TIME>10/08/1998 21:36:42.85</DATE_TIME>), **the granularity of ANCHOR_VAL** will have the same specificity, even if it seems to the annotator that such a degree of detail wasn’t intended by the speaker. In the following examples, we assume the reference time is July 15, 1999. Thus, the tag in the first example below means “the present time as it was interpreted as of July 15, 1999.”

Now there are hundreds of schools with thousands and thousands of dancers.
<TIMEX2 VAL="PRESENT_REF" ANCHOR_DIR="AS_OF" ANCHOR_VAL="1999-07-15">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.

¹⁸ “once” is markable in “Huang once ran Lippo’s U.S. operations” but is not markable in “Once the country develops economic infrastructure...”, “We visited Prague once” nor “They responded at once”. See section 3.2.

We can trust <TIMEX2 VAL="PRESENT_REF" ANCHOR_DIR="AS_OF" ANCHOR_VAL="1999-07-15">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" ANCHOR_DIR="AS_OF" ANCHOR_VAL="1999-07-15">current</TIMEX2> plague of college campuses.

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" ANCHOR_DIR="AS_OF" ANCHOR_VAL="1999-07-15">now</TIMEX2></TIMEX2>.

Hotel and resort operators in Malaysia are experiencing tough times *these days*.

Hotel and resort operators in Malaysia are experiencing tough times <TIMEX2 VAL="PRESENT_REF" ANCHOR_DIR="AS_OF" ANCHOR_VAL="1999-07-15">these days</TIMEX2>.

The stay of execution keeps the Bells out of the long-distance market for *the time being*.

The stay of execution keeps the Bells out of the long-distance market for <TIMEX2 VAL="PRESENT_REF" ANCHOR_DIR="AS_OF" ANCHOR_VAL="1999-07-15">the time being</TIMEX2>.

For PAST_REF expressions, ANCHOR_DIR uses the BEFORE value to indicate the general timeframe under which the notion of “past” should be interpreted. With one exception, explained below, **the granularity of ANCHOR_VAL** is based on the reference time, as described earlier for the PRESENT_REF token. The following examples illustrate several uses of PAST_REF. For example, in the first one the TIMEX2 tag conveys that “Europe will be stronger than in *the time before July 15, 1999*.”

Europe will be stronger than in *the past*.

Europe will be stronger than in <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15">the past</TIMEX2>.

The prospects for resuscitating the peace process have appeared grim *lately*.¹⁹

The prospects for resuscitating the peace process have appeared grim <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15">lately</TIMEX2>.

They knew about his *recent* order to conduct a search for the body.

They knew about his <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15">recent</TIMEX2> order to conduct a search for the body.

Duisenberg is a *former* Dutch central banker.

Duisenberg is a <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15">former</TIMEX2> Dutch central banker.²⁰

We also use PAST_REF (and FUTURE_REF) tokens for vague offsets from the present, e.g. *some days ago*, *a few months ago*, *a couple of years ago*. This is the one case where **the granularity of the ANCHOR_VAL** should be based on the granularity of the lexical expression. What distinguishes these expressions is that their meaning incorporates measurable

¹⁹ We do not attempt to distinguish between the past and the recent past, or the future and the near future. We also interpret “lately” to mean “before now,” not “as of now.”

²⁰ We do not attempt to capture the full aspectual information of this expression, in which Duisenberg was a banker *before a time in the past*, nor that Duisenberg will continue to be a “former banker” into the future.

time units, e.g., day, month, year, etc.²¹ We can exploit this degree of specificity while at the same time accommodating the vagueness of the offset (*some, few, couple*). In the following example, although “several weeks ago” refers to some unspecified point in the past, it is definitely not a point within the current week. We can capture this by using the current week (not day or hour) as the anchor:

The incident marked the most serious confrontation between Israeli troops and Palestinians since pro-Iraq demonstrations began in the West Bank *several weeks ago*.

The incident marked the most serious confrontation between Israeli troops and Palestinians since pro-Iraq demonstrations began in the West Bank <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-W28">several weeks ago</TIMEX2>.

Expressions of type FUTURE_REF use the ANCHOR_DIR value AFTER to capture what was considered “the future” at the time of the utterance:

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

His partner is a <TIMEX2 VAL="P22Y" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999">22-year-old</TIMEX2> <TIMEX2 VAL="FUTURE_REF" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-07-15">future</TIMEX2> accountant.

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

We need to discuss <TIMEX2 VAL="FUTURE_REF" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-07-15">the future of our peoples</TIMEX2>.

The following examples illustrate those FUTURE_REF cases where **the granularity of ANCHOR_VAL** should come from the lexical expression because the head word is a measurable time unit:

Madeleine Albright warned Milosevic that if he doesn't withdraw his ground forces from Kosovo, he'll face air strikes in a *matter of days*.

Madeleine Albright warned Milosevic that if he doesn't withdraw his ground forces from Kosovo, he'll face air strikes in <TIMEX2 VAL="FUTURE_REF" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-07-15">a matter of days</TIMEX2>.

The appeals court will hear arguments in the case in a *few months*.

The appeals court will hear arguments in the case in <TIMEX2 VAL="FUTURE_REF" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-07">a few months</TIMEX2>.

The dianthus bloomed on *Thursday*, and then a *few days later* my son drove over them.

The dianthus bloomed on <TIMEX2 VAL="1999">Thursday</TIMEX2>, and then <TIMEX2 VAL="FUTURE_REF" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-07-15">a few days later</TIMEX2> my son drove over them.

4.3.2.1 Choosing a Reference Time

The preceding examples were relatively context-free, so the default reference time of July 15, 1999 was used for illustration purposes. However, it is possible for a document to have multiple reference times. For example, a newspaper story written on July 15 can report on an interview that was conducted the previous week, so that all the interviewee's utterances have the earlier

²¹ An expression like “some time ago” does not belong to this group, because of the vague noun “time.”

reference time. If a document contains a publication time and writing time (as found in bylines), the writing time is the more logical reference time.

The following discussion of three additional examples illustrates the need for careful reading of the context in determining a reference time.

In the following example from a news article about the 150th anniversary of *The Communist Manifesto*, the context makes it clear that the anchoring time for “the future” is 1848 rather than the year that the news article was written, which was 1998:

In 1848 Marx predicted that economics and not politics would be the driving force in *the future*.
In <TIMEX2 VAL="1848">1848</TIMEX2>Marx predicted that economics and not politics would be the driving force in <TIMEX2 VAL="FUTURE_REF" ANCHOR_DIR="AFTER" ANCHOR_VAL="1848">the future</TIMEX2>.

In the next example, we have three potential anchors for “long ago” and “the past”: the day the article was written, the day they finished their trip, and the day they started their trip. In such cases human annotators are free to use their world knowledge to determine the correct anchor. Although this will likely lead to some interannotator inconsistencies, the differences in the actual ANCHOR_VAL value are not likely to be too great. We do encourage annotators to settle on a single reference time for a passage if there is one that will work for the interpretation of all the time expressions contained within it. For this example we chose the day of departure as the anchoring date, on the grounds that the temporal modifiers are related to activities (“crossed the sea...long ago”) and feelings (“wanted”) that existed before they left on their trip.

(SOUTH SEA ISLANDS, *August 1, 1947*) – Six sailors finished a 5,000-mile sailboat trip *yesterday*. In a ship called the Kon Tiki, the men crossed the sea just like people did *long ago*. ... The men left Peru on *April 28*. They wanted to show how, in *the past*, people traveled miles across the seas.
(SOUTH SEA ISLANDS, <TIMEX2 VAL="1947-08-01">August 1, 1947</TIMEX2>) – Six sailors finished a 5,000-mile sailboat trip <TIMEX2 VAL="1947-07-31">yesterday</TIMEX2>. In a ship called the Kon Tiki, the men crossed the sea just like people did <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1947-04-28">long ago</TIMEX2>. The men left Peru on <TIMEX2 VAL="1947-04-28">April 28</TIMEX2>. They wanted to show how, in <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1947-04-28">the past</TIMEX2>, people traveled miles across the seas.²²

Here is another example where the annotator must use world knowledge to determine an anchoring date for the expression “one time.” In annotating this example, we felt that the statues were safe before the dam was built. There was no date provided for the dam building, but we assumed that the workers would not have started moving the statues five years ago (1961) if the statues had been considered safe then, so the anchoring point was set to “before 1961.”

(EGYPT, 1966) – A team of workers has spent *five years* moving heavy statues. At *one time* the statues were safe, carved in rock next to the Nile River. But then a big dam was built to hold back the water and keep the Nile from flooding. Unfortunately, the resulting lake would also cover the statues.

²² Anchoring the distant past on the present is a bit odd, but because there’s no agreed-upon notion of what counts as the distant past, it is difficult to formulate a rule for handling it differently; To maintain consistency, we use ANCHOR values for the distant past, but some users may wish not to.

(EGYPT, <TIMEX2 VAL="1966">1966</TIMEX2>) – A team of workers has spent <TIMEX2 VAL="P5Y" ANCHOR_DIR="ENDING" ANCHOR_VAL="1966">five years</TIMEX2> moving heavy statues. At <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1961">one time</TIMEX2> the statues were safe, carved in rock next to the Nile River. But then a big dam was built to hold back the water and keep the Nile from flooding. Unfortunately, the resulting lake would also cover the statues.

Sometimes there is not enough information in the document to specify an anchor time. This can happen when the granularity of the expression is finer than the granularity of the known reference time. In the following example from a transcribed broadcast news show, the document doesn't indicate what time the reporter uttered this statement, only what day. In these cases ANCHOR_VAL can use the X placeholder, which is discussed in section 4.3.8.

A couple of minutes ago, we reported that the big tobacco companies were in Congress *today*, pushing hard for a national settlement of all the lawsuits against them.
 <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15TXX:XX">A couple of minutes ago</TIMEX2>, we reported that the big tobacco companies were in Congress <TIMEX2 VAL="1999-07-15">today</TIMEX2>, pushing hard for a national settlement of all the lawsuits against them.

4.3.3 Seasons

Seasons of the year have different meanings to different people. Some might interpret “winter” as the cold part of the year, while others interpret it more literally, from winter solstice to spring equinox. Thus, when the tagged lexical expression refers to a season, **tokens** rather than precise numerical values are used in the ISO month position of VAL for **point** expressions. For **duration** expressions, the token is used in the same manner as the ISO “PnW” format for periods of weeks. Note that we also use these season tokens for academic terms; i.e., “semester” in this context is a **lexical trigger**. Other academic terms that are less readily mapped to seasons are treated as culturally-determined expressions, discussed in section 4.8.1.

Table 4-4 Season Tokens

| Token | Position in VAL | Sample Expressions |
|-------|--------------------------------|--|
| SP | point: month duration: PnSP | spring springtime |
| SU | point: month duration: PnSU | summer summertime |
| FA | point: month duration: PnFA | fall autumn fall semester fall term |
| WI | point: month duration: PnWI | winter wintertime |

The following examples illustrate point expressions:

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>.

In the Midwest, after *an unusually mild winter*, they're digging out from a fierce snowstorm.

In the Midwest, after <TIMEX2 VAL="1999-WI">an unusually mild winter</TIMEX2>, they're digging out from a fierce snowstorm.

Winters are tricky because they span the new year boundary. When no year is specified (as in the preceding example), the default year for winter (WI) is the year covering January-March of that winter. That is, for the winter spanning 1998-1999, use 1999 as the default year, in the absence of evidence to the contrary in the document itself. However, using 1998-WI is correct for “winter” in something like “We had a winter snowstorm on December 12, 1998.”

In *duration* expressions, the season tokens can be combined with anchoring attributes. It would defeat the purpose of using tokens if we required the anchoring attributes to reflect the precise starting or ending dates of the season. For a reference to a single spring, fall, or summer, the WITHIN token can be used:

We had heavy rains all spring.

We had heavy rains <TIMEX2 VAL="P1SP" ANCHOR_DIR="WITHIN" ANCHOR_VAL="1999">all spring</TIMEX2>.

For references to a single winter, the annotation must take into account the fact that winter spans the year boundary. Thus, in the following example, the value for ANCHOR_DIR is “STARTING,” indicating the winter of 1999-2000.

That raises the prospect that the position could go unfilled *all winter*.

That raises the prospect that the position could go unfilled <TIMEX2 VAL="P1WI" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999">all winter</TIMEX2>.

References to more than one season can be anchored with the current reference time using the BEFORE/AFTER values:

Confrontations among Catholic protesters, Protestant marchers and the police have provoked widespread violence in Northern Ireland for *the past three summers*.

Confrontations among Catholic protesters, Protestant marchers and the police have provoked widespread violence in Northern Ireland for <TIMEX2 VAL="P3SU" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1998-03-22">the past three summers</TIMEX2>.²³

4.3.4 Fiscal Years

Fiscal years are another time unit that is very much open to interpretation, since they vary freely from organization to organization. We use a two-character token pre-pended to the ISO year position to indicate that a fiscal year is being referred to.

Table 4-5 Fiscal Year Token

| Token | Position in VAL | Sample Expressions |
|-------|-------------------------------------|-----------------------------|
| FY | point: pre-pended duration: PnFY | fiscal fiscal year FY |

The aid was expected to total about \$37 million in *fiscal 1998*.

The aid was expected to total about \$37 million in <TIMEX2 VAL="FY1998">fiscal 1998</TIMEX2>.

Their profits rose throughout *FY98*.

Their profits rose throughout <TIMEX2 VAL="P1FY" ANCHOR_DIR="WITHIN" ANCHOR_VAL="FY1998">FY98</TIMEX2>

See the next section for additional examples of fiscal years.

4.3.5 Yearly Quarters and Halves

Tokens can also be used for expressions denoting quarters or halves of years.

Table 4-6 Yearly Quarters & Halves Tokens

| Token | Position in VAL | Sample Expressions |
|-------|--------------------------------|---|
| Q1 | point: month duration: PnQ1 | 1 st quarter |
| Q2 | point: month duration: PnQ2 | 2 nd quarter |
| Q3 | point: month duration: PnQ3 | 3 rd quarter |
| Q4 | point: month duration: PnQ4 | 4 th quarter |
| QX | point: month duration: PnQX | (within) that quarter (for) three quarters |

²³ To keep the example genuine, we are using the real reference time for this example, which was March 22, 1998.

| | | | |
|----|---------------------|---------------|--|
| H1 | point: duration: | month PnH1 | 1st half (of year) |
| H2 | point: duration: | month PnH2 | 2nd half (of year) |
| HX | point: duration: | month PnHX | (within) a half-year (for) one half |

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, <TIMEX2 VAL="1998-Q4">its second-best quarter ever</TIMEX2>

Their profits rose *this year* during *the first three quarters*.

Their profits rose <TIMEX2 VAL="1999">this year</TIMEX2> during <TIMEX2 VAL="P3QX" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999">the first three quarters</TIMEX2>. ²⁴

Service industries maintained strong executive demand throughout *the first quarter of 1998*.

Service industries maintained strong executive demand throughout <TIMEX2 VAL="P1Q1" ANCHOR_DIR="STARTING" ANCHOR_VAL="1998">the first quarter of 1998</TIMEX2>.

The fiscal year and quarter/half tokens can be combined if the context makes it clear that the quarters are part of a fiscal rather than a calendar year:

Their expenses increased during *the first quarter of FY1998*.

Their expenses increased during <TIMEX2 VAL="FY1998-Q1">the first quarter of <TIMEX2 VAL="FY1998">FY1998</TIMEX2></TIMEX2>.

4.3.6 Weekends

To capture the meaning behind the expression “weekend,” first determine which week is intended, and then place a *WE token* in the day position of the ISO value. Thus, “1999-W28-WE” means “the weekend of Week 28 of 1999.” For point expressions, the WE token is only used in the week-based ISO format. For *duration* expressions, the anchoring attributes can be used to identify the week.

Table 4-7 Weekend Token

| Token | Position in VAL | Sample Expression |
|-------|---|-------------------|
| WE | point: day in week-based format duration: PnWE | weekend |

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.

²⁴ Note that the annotation does not capture the meaning of “first.” Also, the reason this is normalized as a duration is because it is a plural expression, unlike the previous example, which is a point.

The senators will be working through *the weekend*.

The senators will be working through <TIMEX2 VAL="P1WE" ANCHOR_DIR="WITHIN" ANCHOR_VAL="1999-W28">the weekend</TIMEX2>.

The seminar was taught over *two weekends*.

The seminar was taught over <TIMEX2 VAL="P2WE" ANCHOR_DIR="WITHIN" ANCHOR_VAL="1999">two weekends</TIMEX2>.

For holiday weekends that extend beyond Saturday and Sunday, the weekend token can still be used:

Many Americans were traveling *that long holiday weekend*.

Many Americans were traveling <TIMEX2 VAL="1999-W21-WE">that long holiday weekend</TIMEX2>.

4.3.7 Morning, Afternoon, and Night

Periods of the day such as “morning” are also subject to individual interpretation, so we use *tokens* in place of the hour position of the ISO value in point expressions.

Table 4-8 Part-of-Day Tokens

| Token | Position in VAL | Sample Expressions |
|-------|-------------------------------|---|
| MO | point: hour duration: PnMO | morning |
| MI | point: hour duration: PnMI | mid-day |
| AF | point: hour duration: PnAF | afternoon |
| EV | point: hour duration: PnEV | evening |
| NI | point: hour duration: PnNI | night |
| PM | point: hour duration: PnPM | PM |
| DT | point: hour duration: PnDT | morning + afternoon (basically, daytime or working hours) |

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

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

He explained that he was flying off to Bosnia in *the morning* and wanted to go out with a bang.²⁵

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

²⁵ This example appeared in a narrative about New Year’s Eve 1998.

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

The bug will get fixed between <TIMEX2 VAL="PRESENT_REF" ANCHOR_VAL="1999-07-15" ANCHOR_DIR="AS_OF">now</TIMEX2> and <TIMEX2 VAL="1999-07-19TMO">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 “T11:00” and “morning” is not annotated separately.

Like “winter,” nighttime is tricky because it spans a new day boundary. Thus, “*last night*” could mean the latter part of the previous day or early part of the current day. To prevent the annotator from having to discern the writer’s intention, we invoke the following rule if the context is not explicit about the day:

“*Last Night* is Yesterday” Rule

When in doubt, “*last night*” gets the date of yesterday

For example:

The talks ended *last night*.

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

If the context is explicit about which portion of the night is being referred to, the above rule can be ignored:

The talks ended *last night* at 1 a.m.

The talks ended <TIMEX2 VAL="1999-07-15-TNI">last night</TIMEX2> at <TIMEX2 VAL="1999-07-15T01:00">1 a.m.</TIMEX2>

Here are some duration examples using the TOD tokens:

Madeleine Albright has been sort of shuttling back and forth uh through *the day* and *the night* speaking to both sides.

Madeleine Albright has been sort of shuttling back and forth uh through <TIMEX2 VAL="P1DT" ANCHOR_DIR="WITHIN" ANCHOR_VAL="1999-07-15">the day</TIMEX2> and <TIMEX2 VAL="P1NI" ANCHOR_DIR="WITHIN" ANCHOR_VAL="1999-07-14">the night</TIMEX2> speaking to both sides.

First lady Hillary Rodham Clinton, on a *day* trip to Baltimore to speak at a college, found herself pressed for a response on the uproar.

First lady Hillary Rodham Clinton, on a <TIMEX2 VAL="P1DT" ANCHOR_DIR="WITHIN" ANCHOR_VAL="1999-07-15">day</TIMEX2> trip to Baltimore to speak at a college, found herself pressed for a response on the uproar.

4.3.8 Unspecified Components of Calendar Dates and Times of Day

The ISO standard permits 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 if the context does not allow the values to be specified. 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.

The building was erected in '63.²⁶

The building was erected in <TIMEX2 VAL="XX63">'63</TIMEX2>.

Sunday's paper was thrown out.²⁷

<TIMEX2 VAL="XXXX-WXX-7">Sunday</TIMEX2>'s paper was thrown out</TIMEX2>.

Fox said he believes that Lewinsky visited the Oval Office, on a *Saturday afternoon* in either *September, October* or *November 1995*.

Fox said he believes that Lewinsky visited the Oval Office, on <TIMEX2 VAL="1995-WXX-6TAF">a Saturday afternoon</TIMEX2> in either <TIMEX2 VAL="1995-09">September</TIMEX2>, <TIMEX2 VAL="1995-10">October </TIMEX2>or <TIMEX2 VAL="1995-11">November 1995</TIMEX2>.

The X placeholder also tends to be useful in temporal expressions denoting sets of times (see “Set-Denoting Temporal Expressions” on page 39) and non-specific expressions (see “Non-Specific Temporal Expressions” on page 41).

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

Keep in mind that the entire context of the document should be used in specifying VAL. In the following example, from a document dated February 19, 1998, VAL can be fully specified, so no X placeholders are needed:

He was referring to a *Feb. 14* stunt in a streetcar dubbed Valentine, in which a justice of the peace married 20 young couples.

He was referring to a <TIMEX2 VAL="1998-02-14">Feb. 14</TIMEX2> stunt in a streetcar dubbed Valentine, in which a justice of the peace married 20 young couples.

4.3.9 Unspecified Plural Numbers

The X character is also used in the representation of durations as an indicator of an unspecified plural number. For example, PXY is interpreted as “a period of X years,” where the value of X is greater than one. Note that as with fully specified durations, these expressions can often be anchored.

After *months of renewed hostility*, the fighting has suddenly ceased.

After <TIMEX2 VAL="PXM" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999-07">months of renewed hostility</TIMEX2>, the fighting has suddenly ceased.

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

The conventions around dating fell away in <TIMEX2 VAL="PXDE" ANCHOR_DIR="BEFORE" ANCHOR_VAL="199">recent decades</TIMEX2>.

²⁶ Assume for this example that the context is not known, so it is unclear which century and millennium are intended. If instead, the century and millennium were known from context, the VAL would be “1963.”

²⁷ Assume for this example that you cannot tell which Sunday is intended, but the context makes it clear that it is a specific Sunday.

Activity will be intensifying over *the coming months*.

Activity will be intensifying over <TIMEX2 VAL="PXM" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-07">the coming months</TIMEX2>.

The people of Northern Ireland are going to have to address these concerns in *the coming weeks and months*.

The people of Northern Ireland are going to have to address these concerns in <TIMEX2 VAL="PXW" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-W28">the coming weeks</TIMEX2> and <TIMEX2 VAL="PXM" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999-07">months</TIMEX2>.

Chinese military thinkers are divided on whether or not they see the Americans as rivals in *the years ahead*.

Chinese military thinkers are divided on whether or not they see the Americans as rivals in <TIMEX2 VAL="PXY" ANCHOR_DIR="AFTER" ANCHOR_VAL="1999">the years ahead</TIMEX2>.

For *millennia*, dancing was a form of social glue.

For <TIMEX2 VAL="PXML" ANCHOR_DIR="ENDING" ANCHOR_VAL="1">millennia</TIMEX2>, dancing was a form of social glue.²⁸

Similarly, vague quantifiers such as “a couple” and “a few” are represented with the X placeholder:

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" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999">the past few years</TIMEX2>.

I was away for *a couple of years*.

I was away for <TIMEX2 VAL="PXY" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999">a couple of years</TIMEX2>.

The X placeholder can be used with the geologic tokens KA, MA, and GA to represent “thousands of years ago” “millions of years ago” and “billions of years ago”:

The triceratops from *millions of years ago* was chewing branches.

The triceratops from <TIMEX2 VAL="MAX">millions of years ago</TIMEX2> was chewing branches.

4.3.10 Combined Week-Based and Month-Based Expressions

There are complex temporal 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-5TNI (plus an attribute to indicate the non-specific nature of the reference).

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

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

²⁸ Recall that “ML” stands for “Millennium” in our extension to the ISO standard.

4.4 Modified Temporal Expressions

In this section we discuss temporal expressions that are quantified or modified in some way. For example, “1999” is an unmodified expression, but “late 1999” is modified. In general, we want the annotation to capture the basic semantics of quantifier modifiers (e.g., *approximately*, *no more than*) and lexicalized aspect markers, e.g., *early*, *start [of]*.

We do not want to capture the semantics of leading prepositions or other terms that are outside the extent of the tagged temporal expression (see the section “Determining the Extent of the Annotations” starting on page 56). For example, the expression “before Tuesday” is not considered a modified expression for our purposes because “before,” as a preposition, is not included within the extent of the TIMEX2 tag. Although it is important to NLP applications to capture the semantics of such prepositional phrases, we view this as a different layer of annotation that would overlay the TIMEX2 annotations, so that the referent of the noun phrase “Tuesday” is retained in the markup. For example, an “event timestamp” tag could make reference to a TIMEX2 tag:

```
<EVENT_TS DIR="BEFORE" TID="1">before <TIMEX2 VAL="1999-07-15"
ID="1">Tuesday</TIMEX2></EVENT_TS>
```

The ACE Relation Detection and Characterization

(<http://www.nist.gov/speech/tests/ace/index.htm>) task used an approach similar to this. See also Setzer (2001) and Ingia and Pustejovsky (2002) for the treatment of prepositional phrases as temporal modifiers.

In the TIMEX2 standard we use the MOD attribute to capture the semantics of modifiers within the scope of the TIMEX2 expression. There are values for MOD that apply only to points in time, values that apply only to durations, and values that can apply to either.

Table 4-9 Modifier Tokens

| | Token | Sample Expressions |
|------------------|---------------|--|
| Points | 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”) |
| Durations | 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”) |

| | | |
|-----------------------------|--------|--|
| Points and Durations | 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 (e.g., "late afternoon") |
| | APPROX | about (e.g., "about three years ago") around (e.g., "around three o'clock") approximately (e.g., "approximately one hour") |

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" ANCHOR_DIR="AS_OF" ANCHOR_VAL="1999-07-15">now</TIMEX2> retired.²⁹

The king lived *more than 4,000 years ago*.

The king lived <TIMEX2 VAL="BC2001" MOD="BEFORE">more than 4,000 years ago</TIMEX2>.

The teacher has *nearly four decades of experience*.

The teacher has <TIMEX2 VAL="P4DE" MOD="LESS_THAN" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15">nearly four decades of experience</TIMEX2>.³⁰

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

She has been at work for <TIMEX2 VAL="P1M" MOD="MORE_THAN" ANCHOR_DIR="ENDING" ANCHOR_VAL="1999-07-15">more than a month</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" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999-07-16">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" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999">at least the next year</TIMEX2> or <TIMEX2 VAL="P2Y" MOD="EQUAL_OR_MORE" ANCHOR_DIR="STARTING" ANCHOR_VAL="1999">two</TIMEX2>.³¹

²⁹ This markup appears to violate the "Expression Granularity" Rule, since it is normalized to the granularity of a year, not a decade. But annotators typically agree that decade often means the same thing as "10 years," so that normalizing to the year granularity is acceptable. The same is true of "century," when interpreted as "100 years."

³⁰ Note that LESS_THAN, not APPROX, is used for "nearly."

³¹ The MOD is shown as applying to each half of the disjunction, although that is somewhat misleading semantically.

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>.

The trend began in *the early 1960s*.

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

Marx was an important social thinker of *the mid nineteenth century*.

Marx was an important social thinker of <TIMEX2 VAL="18" MOD="MID">the mid nineteenth century</TIMEX2>.

The talks ended *late last night*.

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

The restaurant opened *about three years ago*.

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

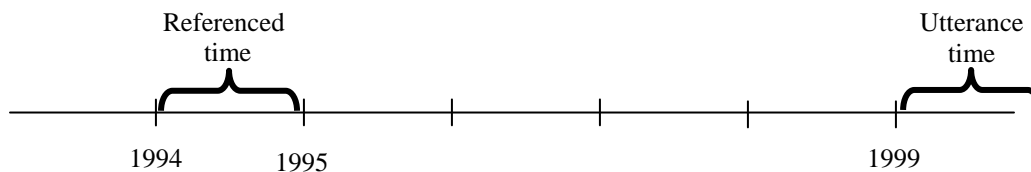
Aristotle wrote his "Poetics" *some 2,300 years ago*.

Aristotle wrote his "Poetics" <TIMEX2 VAL="BC0301" MOD="APPROX">some 2,300 years ago</TIMEX2>.

4.4.1 Anchoring Approximate Offsets

In an expression such as “nearly five years ago,” if the utterance time is sometime in 1999, we can reasonably assume that the time being referred to is after the start of 1994, but before the start of 1995, as diagrammed below:

“*nearly five years ago*”



In cases such as these, the tag uses a combination of VAL, MOD, and anchoring attributes to anchor the vague offset. In the above example, the VAL and MOD attributes are used to capture the notion of “after 1994” and the anchoring attributes are used to capture the notion of “before 1995.”

Nearly five years ago, the plan was to rebuild the team around the new draftees.

<TIMEX2 VAL="1994" MOD="AFTER" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1995">Nearly five years ago</TIMEX2>, the plan was to rebuild the team around the new draftees.

Here are some more examples:

The British mine removal expert was kidnapped by Khmer Rouge guerrillas *just over two years ago*.

The British mine removal expert was kidnapped by Khmer Rouge guerrillas <TIMEX2 VAL="1997" MOD="BEFORE" ANCHOR_DIR="AFTER" ANCHOR_VAL="1996">just over two years ago</TIMEX2>.

The investigation concluded *almost a year and a half ago*.

The investigation concluded <TIMEX2 VAL="1998-01" MOD="AFTER" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15">almost a year and a half ago</TIMEX2>.

4.5 Set-Denoting Temporal Expressions

In this section we discuss temporal expressions that indicate sets of times, such as “every hour,” “Octobers” or “numerous Saturdays.” Whereas points of time tell when something happen, and durations tell how long something lasted, set-denoting expressions tell how often something happened.

Previous versions of the TIMEX2 standard attempted to capture the semantics of such expressions with two attributes, GRANULARITY and PERIODICITY. Experience showed that these attributes were insufficient for the task and were confusing to annotators. They have therefore been eliminated and efforts are currently underway to develop a new model for set expressions. The proposed model is discussed in Appendix B. In the meantime, the only attribute remaining for these expressions is SET, which is set to “YES” in all set expressions. If the expression does not denote a set, the attribute is omitted from the annotation altogether.

In addition, we have expanded the use of the X placeholder to encode the granularity of the linguistic expression. Previously, we allowed the X placeholder in regularly recurring times like “Fridays,” where VAL was “XXXX-WXX-5.” We now require the X placeholder to be used wherever the lexical trigger contains a temporal unit that maps to a portion of the ISO format. Thus, “monthly” receives a VAL of “XXXX-XX” to denote the month granularity of the lexical expression.

4.5.1 Example Annotations of Sets of Regularly Recurring Times

Sets of regularly recurring times can be expressed by words like “always” or “every” or “each” in the local context or by the plural form of *triggers*.

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

They watched Millionaire on TV <TIMEX2 SET="YES" VAL="1999-WXX-2">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.³²

There were, as there have been <TIMEX2 SET="YES" VAL="XXXX-12-31">every December 31</TIMEX2> for <TIMEX2 VAL="PXDE" ANCHOR_DIR="ENDING" ANCHOR_VAL="199">decades</TIMEX2>, thousands of people in Times Square.

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

We spend <TIMEX2 SET="YES" VAL="P3D" ANCHOR_DIR="STARTING" ANCHOR_VAL="XXXX-XX-01">the first three days of <TIMEX2 VAL="XXXX-XX" SET="YES">every month</TIMEX2></TIMEX2> writing status reports.

On *Friday nights*, the ballroom dance club was overwhelmed.³³

³² The phrase “for decades” is not considered a modifier of the preceding time expression, so there are two independent TIMEX2 annotations.

³³ Assume for this example that 1998 is implied.

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

Two years ago, the dance club drew about 100 students *each week*.³⁴
<TIMEX2 VAL="1997">Two years ago</TIMEX2>, the dance club drew about 100 students <TIMEX2 SET="YES" VAL="1997-WXX">each week</TIMEX2>.

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

Beware of “**everyday**” (sometimes also written “every day”) when it means “common” or “typical, as in “Italians are masters at transforming simple, everyday ingredients into dishes that are quick, healthy and satisfying.” This word sense is not markable.

Temporal adverbs like the following are also considered set expressions:

The family has visited the cemetery *monthly* since his death.
The family has visited the cemetery <TIMEX2 SET="YES" VAL="XXXX-XX">monthly</TIMEX2> since his death.

This picnic has been an *annual* event in our family.
This picnic has been an <TIMEX2 SET="YES" VAL="XXXX">annual</TIMEX2> event in our family.

The Colombian government provided *semiannual* certification during the recent regime.
The Colombian government provided <TIMEX2 SET="YES" VAL="XXXX-HX">semiannual</TIMEX2> certification during the recent regime.

Set expressions also occur in frequency (rate) expressions like the following:

The Orient Express traveled at speeds of 50 miles *per hour*.
The Orient Express traveled at speeds of 50 miles <TIMEX2 SET="YES" VAL="PT1H">per hour</TIMEX2>.

Singer and Howe predict that they will be selling 1/2 million machines *a year*.
Singer and Howe say that they will be selling 1/2 million machines <TIMEX2 SET="YES" VAL="P1Y">a year</TIMEX2>.

For temporal adverbs and frequency expressions like those illustrated above, it is often possible to discern the timeframe in which these sets occur, based on the document’s context. While it may be possible to use the anchoring attributes to anchor such sets (see Appendix B), we are not doing so at this time. **Do not use ANCHOR attributes with SET expressions.**

In frequencies expressed with adverbial numbers (e.g., “twice a day”, “once a year”), the adverbial number is included in the extent of the tag, but its semantics are ignored. Only the granularity of the temporal head is captured. Again, Appendix B suggests ways this might be handled in the future.

Myles Lorten jogs around the museum *twice a day*.
Myles Lorten jogs around the museum <TIMEX2 SET="YES" VAL="XXXX-XX-XX">twice a day</TIMEX2>.

³⁴ “Each” is represented the same as “every” (“each and every”).

We can capture durations occurring within members of a set, but currently there is no annotation to indicate that the two are connected:

Doctors recommend walking for *one hour each day*.
Doctors recommend walking for <TIMEX2 VAL="PT1H">one hour</TIMEX2> <TIMEX2 SET="YES" VAL="XXXX-XX-XX">each day</TIMEX2>.

4.5.2 Example Annotations of Sets of Irregularly Recurring Times

Irregularly recurring times are also indicated with plural triggers, and vague quantifiers like some, several, few etc. In our current tagging scheme, the tag attributes themselves do not capture the distinction between regular and irregularly recurring times.

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

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

Some winters, he was too sick to go to school.
<TIMEX2 VAL="XXXX-WI" SET="YES">Some winters</TIMEX2>, he was too sick to go to school.

Because on *Tuesday* I am busy *the hours that you are not*.
Because on <TIMEX2 VAL="1999-07-20">TUESDAY</TIMEX2> I am busy <TIMEX2 VAL="1999-07-20TXX">the hours that you are not</TIMEX2>

Note that we are currently not capturing the semantics of some SET modifiers, such as “almost” in the following example.

Military contacts have ballooned to *almost weekly* visits by Chinese officers to the U.S.
Military contacts have ballooned to <TIMEX2 SET="YES" VAL="XXXX-WXX">almost weekly</TIMEX2> visits by Chinese officers to the U.S.

4.6 Non-Specific Temporal Expressions

This section discusses temporal expressions that do not reference a specific time. These non-specific temporal expressions, which go somewhat beyond the purposes for which the ISO standard was developed, fall into two distinct categories:

- Generic: **Lexical triggers** used generically, as in “I love *December*” or “I like *the winter*” or “*Winters* are cold here” or “The average movie lasts *two hours*.” Such expressions specify a whole class of temporal entities rather than referring to a specific time. (Note that the example sentences are stating generalizations.)
- Indefinite: Singular lexical triggers that are used referentially but are indefinite, as in “The election took place on *a Tuesday*.” (Note the indefinite article “a” in this example.)

The 2003 and earlier versions of the TIMEX2 standard had a NON_SPECIFIC attribute to identify such expressions. However, an analysis of the TIMEX2 tags in the ACE 2002 corpus revealed that this attribute is completely redundant. All of the semantically non-specific expressions fall into one of three categories:

1. Calendar or week-based expressions that have X placeholders. The presence of X placeholders tells us it is non-specific.
2. Duration expressions that have no ANCHOR attributes. The lack of anchoring attributes tells us it is a non-specific duration.
3. Expressions that receive no VAL, such as sports seasons and holidays (see section 4.8 on page 50). Such expressions are already impoverished by not having a VAL, so it is not important to specify whether they are non-specific or not.

Nevertheless, we've left this section in the guidelines to provide examples of how such expressions are normalized, particularly with respect to VAL. Because of the nature of non-specific expressions, the value of VAL is necessarily underspecified in many cases. When possible, the X placeholder should be used to indicate the granularity of the lexical expression (whether it's a day, year, etc.). In some cases the temporal expression itself is not a distinct unit, so no VAL at all is possible. The following examples illustrate the range of possibilities for VAL:

The trip to Romania takes *eight hours*.

The trip to Romania takes <TIMEX2 VAL="PT8H">eight hours</TIMEX2>.

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

They report the traffic conditions <TIMEX2 VAL="TXX:15">15 minutes after the hour</TIMEX2>.³⁵

The ideal situation for *January to March in a decent year* for hotel owners is an occupancy rate of 70-75% with good room rates.

The ideal situation for <TIMEX2 VAL="XXXX-01">January</TIMEX2> to <TIMEX2 VAL="XXXX-03">March </TIMEX2>in <TIMEX2 VAL="XXXX">a decent year</TIMEX2> for hotel owners is an occupancy rate of 70-75% with good room rates.

Each year, thousands of people come to Sea World to see Kandu the killer whale do tricks.

<TIMEX2 VAL="XXXX" SET="YES">Each year</TIMEX2>, thousands of people come to Sea World to see Kandu the killer whale do tricks.

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.³⁶

We will give you as many answers as we can, as soon as we can, at <TIMEX2>the appropriate time</TIMEX2>, consistent with our obligation to also cooperate with the investigations.

In the following sections you will see additional examples of the various ways VAL is encoded in non-specific expressions.

³⁵ Note that if the entire YYYY-MM-DD portion is unspecified, it should be eliminated rather than filled in with X placeholders.

³⁶ This expression is taken to be essentially synonymous with “*some appropriate time*”, i.e., it's not really definite.

4.6.1 Non-Specific Point Expressions

The following sentence illustrates a **plural noun** used generically:

Winters are cold here.
<TIMEX2 VAL="XXXX-WI">Winters</TIMEX2> are cold here.

The following examples illustrate **definite noun phrases** (and other types) being used generically:

If you were living on Mars, in *the morning* you would wake up to pink skies. At *night*, you would see blue sunsets.
If you were living on Mars, in <TIMEX2 VAL="TMO">the morning</TIMEX2> you would wake up to pink skies. At <TIMEX2 VAL="TNI">night</TIMEX2>, you would see blue sunsets.
An old folk tale says that groundhogs can tell how long *the winter* will last. The tale says to watch the groundhogs on *February 2*. If they don't see their shadows when they emerge from their dens, they will stay outside, and that means *spring* will start soon.
An old folk tale says that groundhogs can tell how long <TIMEX2 VAL="XXXX-WI">the winter</TIMEX2> will last. The tale says to watch the groundhogs on <TIMEX2 VAL="XXXX-02-02">February 2</TIMEX2>. If they don't see their shadows when they emerge from their dens, they will stay outside, and that means <TIMEX2 VAL="XXXX-SP">spring</TIMEX2> will soon start.

Named temporal units can be used generically as well:

I love *December*.
I love <TIMEX2 VAL="XXXX-12">December</TIMEX2>.

Indefinite noun phrases can be non-specific even when used referentially:

The last election took place on *a Tuesday*.
The last election took place on <TIMEX2 VAL="1998-WXX-2">a Tuesday</TIMEX2>.
He left on *a sunny day in June*.
He left on <TIMEX2 VAL="1999-06-XX">a sunny day in <TIMEX2 VAL="1999-06">June</TIMEX2></TIMEX2>.

However, if the context of the document makes it clear what the interpretation of VAL would be, it should be fully specified:

He left on *a sunny day in June*. In fact, it must have been *the 20th*, because that's when we were supposed to get married.
He left on <TIMEX2 VAL="1999-06-20">a sunny day in <TIMEX2 VAL="1999-06">June</TIMEX2></TIMEX2>. In fact, it must have been <TIMEX2 VAL="1999-06-20">the 20th</TIMEX2>, because that's when we were supposed to get married.

In addition, in predicate nominal constructions like the following, the indefinite expression should be normalized. This is a change from the 2003 standard. We have found it very difficult to prevent annotators from entering a specific VAL in such cases, so we have changed the guidelines to conform to what appears to be the natural interpretation of such expressions. It is also consistent with the ACE standard for annotating entity mentions in such positions.

Today is a sunny day.

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

4.6.2 Non-Specific Durations

Non-specific durations frequently occur in sentences that are stating generalizations. The following duration expressions are non-specific. They do not use the X placeholder in VAL because the length of time is known, but since no specific period is being referred to, it cannot be anchored.

In an assembly line, each worker does only one task *all day*.

In an assembly line, each worker does only one task <TIMEX2 VAL="P1DT">all day</TIMEX2>.

The gestation period in humans is *nine months*.

<TIMEX2>The gestation period in humans</TIMEX2> is <TIMEX2 VAL="P9M">nine months</TIMEX2>.³⁷

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>.

When there is no overt quantifier for the duration expression, the X placeholder is used in VAL:

New-born quintuplets used to live only *hours* or *days*.

New-born quintuplets used to live only <TIMEX2 VAL="PXH">hours</TIMEX2> or <TIMEX2 VAL="PXD">days</TIMEX2>.

Within *a few days* or *weeks* of being bitten by an infected mosquito, the person comes down with a high fever and tires easily.

Within <TIMEX2 VAL="PXD">a few days</TIMEX2> or <TIMEX2 VAL="PXW">weeks</TIMEX2> of being bitten by an infected mosquito, the person comes down with a high fever and tires easily.

During *the day*, Noah teaches school.

During <TIMEX2 VAL="PXDT">the day</TIMEX2>, Noah teaches school.

4.6.3 Non-Specific TOD Expressions

In the case of non-specific TOD expressions, the ISO standard allows the calendar date portion to be missing, so the X placeholder is not needed in VAL:

The clock strikes 12 at *noon* and *midnight*.

The clock strikes 12 at <TIMEX2 VAL="T12:00">noon</TIMEX2> and <TIMEX2 VAL="T24:00">midnight</TIMEX2>.³⁸

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

That emergency clinic is open from <TIMEX2 VAL="T19:00">7 p.m.</TIMEX2> to <TIMEX2 VAL="T07:00">7 a.m.</TIMEX2>

³⁷ “the gestation period in humans” is not given a VAL. See section 4.10 Expressions Whose Values can C.

³⁸ Granularity of VAL is indicated by the terms “noon” and “midnight.”

4.6.4 Non-Specific Set-Denoting Expressions

Set-denoting expressions can be used in generic contexts. Recall that expressions like adverbs like “always” and “usually” are indicators of SET expressions, highlighted in bold below.

I **always** vote on *Election Day*.

I **always** vote on <TIMEX2 VAL="XXXX-XX-XX" SET="YES">Election Day</TIMEX2>.

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

We **always** watch the game on <TIMEX2 VAL="XXXX-WXX-7" SET="YES">Super Bowl Sunday</TIMEX2>.

April is **usually** wet.

<TIMEX2 VAL="XXXX-04" SET="YES">April</TIMEX2> is **usually** wet.

On *some nights*, one of Mars’ moons rises twice.

On <TIMEX2 VAL="TNI" SET="YES">some nights</TIMEX2>, one of Mars’ moons rises twice.

Yellow fever kills thousands of people around the world *each year*.

Yellow fever kills thousands of people around the world <TIMEX2 VAL="XXXX" SET="YES">each year</TIMEX2>.

Consumers will grow tired of unpredictable *monthly* phone bills.

Consumers will grow tired of unpredictable <TIMEX2 VAL="XXXX-XX" SET="YES">monthly</TIMEX2> phone bills.

4.6.5 Idioms and Metaphors Containing Temporal Expressions

Many time words and phrases can be used in an idiomatic sense, as in “the last minute,” “the eleventh hour,” “the order of the day,” “at the end of the day,” and “midnight oil.” The temporal expressions in such idioms are markable, but VAL should only contain the X placeholder to identify the granularity of the expression. E.g., in the following example, “the day” is not given the VAL from “tomorrow.”

Beer-drinking will indeed be the order of *the day*, assuming the contract is signed *tomorrow*.

Beer-drinking will indeed be the order of <TIMEX2 VAL="XXXX-XX-XX">the day</TIMEX2>, assuming the contract is signed <TIMEX2 VAL="1999-07-16">tomorrow</TIMEX2>.

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

Saddam will not cave in at <TIMEX2 VAL="TXX:XX">the last minute</TIMEX2>.

Greetings and similar oft-said phrases are also considered idiomatic:

Good morning!

<TIMEX2 VAL="TMO">Good morning</TIMEX2>!

4.7 Pronouns and Elided Elements

Demonstrative pronouns can refer to times and dates mentioned elsewhere in the document. In the following example, the demonstrative pronoun “that [day]” is coreferential with “Tuesday.” The demonstrative determiner “That” in “That night” is referring to “The day after that”, i.e., to Wednesday.

The contractor submitted a proposal on *Tuesday*. *The day after that*, the contract was awarded. *That night*, they had a big party.

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="1999-07-14TNI">That night</TIMEX2>, they had a big party.

The pronoun “then,” meaning “at that time,” can refer to a specific time or date mentioned elsewhere in the document:

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.

When the trigger term in a temporal expression has been elided, the expression can receive the same VAL as an expression elsewhere in the document that it co-refers with. In the following example, “his last...” can be interpreted as an elided form of “his last year...” and is thus markable:

Schneider said *this year* would be *his last with the symphony*.

Schneider said <TIMEX2 VAL="1999">this year</TIMEX2> would be <TIMEX2 VAL="1999">his last with the symphony</TIMEX2>.

While all the preceding illustrated markable anaphors, recall that relative pronouns such as “which” and “when” are not given separate TIMEX2 tags. See section 3.2 Non-Markables. In addition, we do not consider “It” referential in cases like the following:

It has been *a strange couple of days*.

It has been <TIMEX2 VAL="P2D" ANCHOR_VAL="1999-07-15" ANCHOR_DIR="ENDING">a strange couple of days</TIMEX2>.

It is *time for the news*.

It is <TIMEX2 VAL="1999-07-15T13:00">time for the news</TIMEX2>.

4.8 Event-Anchored Temporal Expressions

An event-anchored time expression is one in which, in order for its value to be fully resolved, one must know the time of an event. For example, to determine the value of “the day” in “*the day after our meeting*,” one needs to know when the meeting occurred. In prior versions of the TIMEX2 standard, expressions like “*the day after our meeting*” were markable but received no VAL, even in those cases where the annotator knew from world knowledge or could deduce from the context the time of the event. In the 2003 version of the guidelines we began allowing VAL to be specified if it can be determined based on information available within the document. We still prohibit world knowledge from being used to determine VAL.³⁹ We are also still

³⁹ Of course, as in all situations where one is determining the value of VAL, we permit the use of a calendar. This is the only form of “world knowledge” permitted.

limiting the annotations to *lexical triggers*; therefore, just “*the day after our meeting*” and NOT “our meeting” will receive a TIMEX2 tag.

This first set of examples contains temporal *point* expressions that are anchored on actual, specific events, but because the time of the event is not given in the document, VAL does not specify a point on a timeline. X placeholders can be used to capture the granularity of the head noun when it corresponds to a time unit within the ISO format.

I remember *the day that Roosevelt died*.

I remember <TIMEX2 VAL="XXXX-XX-XX">the day that Roosevelt died</TIMEX2>.

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

The firefighters came home <TIMEX2 VAL="XXXX-XX-XX">three days after the fire</TIMEX2>.

He returned some gifts *five days after Christmas*.

He returned some gifts <TIMEX2 VAL="XXXX-XX-XX">five days after
<TIMEX2>Christmas</TIMEX2></TIMEX2>.⁴⁰

Jerry had been brought in *a few weeks earlier* to try to rebalance the story.

Jerry had been brought in <TIMEX2 VAL="XXXX-WXX">a few weeks earlier</TIMEX2> to try to rebalance the story.

By *the time it ended*, it should have been called slushboard, not snowboard.

By <TIMEX2>the time it ended</TIMEX2>, it should have been called slushboard, not snowboard.

Selangor, meanwhile, is expected to have more than 10,000 hotel rooms by *the time the Commonwealth Games take place*.

Selangor, meanwhile, is expected to have more than 10,000 hotel rooms by <TIMEX2>the time the Commonwealth Games take place</TIMEX2>.

Many Russians yearn for *the days when Moscow was the world's other superpower*.

Many Russians yearn for <TIMEX2>the days when Moscow was the world's other superpower</TIMEX2>.

We have been extending our hands since *the days of Israeli Prime Minister Yitzhak Rabin*, and our hands are still slapped.

We have been extending our hands since <TIMEX2>the days of Israeli Prime Minister Yitzhak Rabin</TIMEX2>, and our hands are still slapped.

In *the distant era known as Watergate*, it was Johnny Carson who ruled.

In <TIMEX2>the distant era known as Watergate</TIMEX2>, it was Johnny Carson who ruled.

In theory, the past/present/future tokens and the ANCHOR attributes could be used in the previous examples in order to orient the expressions with respect to some general time frame. For example, in all cases where the anchoring events are described in the past tense, one could specify VAL="PAST_REF" in the TIMEX2 tag and provide an anchoring time. We note this here because although the official TIMEX2 standard does not require this level of detail, users of the TIMEX2 standard are free to use the token and anchor values in this way if it would be useful in their domain.

⁴⁰ Although “Christmas” is markable, it doesn’t provide an anchoring date value because it does not have a VAL itself. See section 4.9.3 about holidays.

In the following example, although the information available about the time of the murder is not specific enough to provide an exact value in minutes, a value at a broader granularity can be given, along with X placeholders to capture the granularity of the lexical trigger:

Vasily Naumov died of multiple gunshot wounds at the wheel of his BMW on *Tuesday night*. ... Police immediately announced an all-points alert and *40 minutes after the killing* arrested two suspects in a car. Vasily Naumov died of multiple gunshot wounds at the wheel of his BMW on <TIMEX2 VAL="1999-07-13TNI">Tuesday night</TIMEX2>. ... Police immediately announced an all-points alert and <TIMEX2 VAL="1999-07-13TXX:XX">40 minutes after the killing</TIMEX2> arrested two suspects in a car.⁴¹

In the following *duration* expressions, the period of the duration can be specified in VAL, but because the time of the event is not given in the document, no ANCHOR attributes are given:⁴²

In *the 30 years since Neil Armstrong walked on the Moon*, interest in the U.S. space program has wavered. In <TIMEX2 VAL="P30Y">the 30 years since Neil Armstrong walked on the Moon</TIMEX2>, interest in the U.S. space program has wavered.

Wendy Willits, Arkansas' freshman guard, nailed five 3-pointers in *the first nine minutes of the game*. Wendy Willits, Arkansas' freshman guard, nailed five 3-pointers in <TIMEX2 VAL="PT9M">the first nine minutes of the game</TIMEX2>.

The Idaho congresswoman publicly fed false militia-spawned rumors of malevolent government helicopters in *the months before the Oklahoma City bombing*.

The Idaho congresswoman publicly fed false militia-spawned rumors of malevolent government helicopters in <TIMEX2 VAL="PXM">the months before the Oklahoma City bombing</TIMEX2>.

It couldn't have been this bad even in *the last 10 minutes of the Roman Empire*.

It couldn't have been this bad even in <TIMEX2 VAL="PT10M">the last 10 minutes of the Roman Empire</TIMEX2>.

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

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

In statements of generalizations, an underspecified VAL can often be given:

He's the kind of guy who can go out and party *the night before* and come *the next day* and pitch a no hitter, a perfect game.

He's the kind of guy who can go out and party <TIMEX2 VAL="TNI">the night before</TIMEX2> and come <TIMEX2 VAL="XXXX-XX-XX">the next day</TIMEX2> and pitch a no hitter, a perfect game.

Parents of premature babies must spend *the first few weeks of their babies' lives* with doctors and nurses.

Parents of premature babies must spend <TIMEX2 VAL="PXW">the first few weeks of their babies' lives</TIMEX2> with doctors and nurses.

In the following generalization, no VAL is possible because the lexical trigger itself ("time") is too vague:

We usually hold a meeting on the plane....So you are using *the time that you're in the air* to conduct business.

⁴¹ Given that the anchoring event occurred at night, it's possible that a new day had begun 40 minutes later. We ignore this possibility in specifying the day.

⁴² As with point expressions that refer to events, individual users of the TIMEX2 scheme are free to anchor the event durations with respect to a document reference time, if they would find that useful, but this is not part of the "official" standard.

We usually hold a meeting on the plane....So you are using <TIMEX2>the time that you're in the air</TIMEX2> to conduct business.

The following examples contain event-anchored expressions in which the time of the anchoring event can be determined based on the information in the document.

Guided by that outlook, Netanyahu's cabinet staked out its position in a series of actions before his meeting *Tuesday* with President Clinton. ... Palestinian leader Yasser Arafat, scheduled to meet with Clinton *two days after Netanyahu*, has reacted in generally subdued tones.

Guided by that outlook, Netanyahu's cabinet staked out its position in a series of actions before his meeting <TIMEX2 VAL="1999-07-20">Tuesday</TIMEX2> with President Clinton. ... Palestinian leader Yasser Arafat, scheduled to meet with Clinton <TIMEX2 VAL="1999-07-22">two days after Netanyahu</TIMEX2>, has reacted in generally subdued tones.

06/05/1998 20:15:00 Judge Herve Stephan, is trying to determine whether nine photographers and a motorcyclist arrested at the scene of the accident *last Aug. 31* may have somehow caused the crash....The French authorities determined that the chauffeur was legally drunk at *the time of the accident*.

<TIMEX2 VAL="1998-06-05T20:15:00">06/05/1998 20:15:00</TIMEX2>

Judge Herve Stephan, is trying to determine whether nine photographers and a motorcyclist arrested at the scene of the accident <TIMEX2 VAL="1997-08-31">last Aug. 31</TIMEX2> may have somehow caused the crash....The French authorities determined that the chauffeur was legally drunk at <TIMEX2 VAL="1997-08-31">the time of the accident</TIMEX2>.

Dr. Sheppard was convicted of murdering his wife and sent to prison in *1954*. But Sheppard won the right to a new trial, and *12 years later*, another jury found him not guilty.

Dr. Sheppard was convicted of murdering his wife and sent to prison in <TIMEX2 VAL="1954">1954</TIMEX2>. But Sheppard won the right to a new trial, and <TIMEX2 VAL="1966">12 years later</TIMEX2>, another jury found him not guilty.

By the time he arrived at the New York Hilton to be interrogated by the news media at 4 p.m., he was dragging. By <TIMEX2 VAL="1998-07-14T16:00">the time he arrived at the New York Hilton to be interrogated by the news media at <TIMEX2 VAL="1998-07-14T16:00">4 p.m.</TIMEX2></TIMEX2>, he was dragging.

Now is the time to recognize the possibilities which lie before us in the taking up and developing of this part of our forefathers' vision.

<TIMEX2 VAL="PRESENT_REF" ANCHOR_VAL="1936-11"

ANCHOR_DIR="AS_OF">Now</TIMEX2> is <TIMEX2 VAL="PRESENT_REF" ANCHOR_VAL="1936-11" ANCHOR_DIR="AS_OF">the time to recognize the possibilities which lie before us in the taking up and developing of this part of our forefathers' vision</TIMEX2>.⁴³

4.8.1 Birthdays and Anniversaries

Birthdays and anniversaries are necessarily anchored on some event. The terms “birthday” and anniversary” are *lexical triggers*, so these expressions are markable. Like other event-anchored expressions, they receive a VAL only if there is sufficient information in the document for it to be determined. In the first two examples, there is insufficient information to specify a VAL attribute:

But the heir to the throne was a hit nonetheless *Friday* at a *50th birthday* party staged in his honor.

⁴³ This quote is from an article by Eleanor Roosevelt in *Parent's Magazine* of September 1936.

But the heir to the throne was a hit nonetheless <TIMEX2 VAL="1999-07-09">Friday </TIMEX2> at a <TIMEX2>50th birthday</TIMEX2> party staged in his honor.⁴⁴

As *the tenth anniversary of the Pan Am bombing* draws near, a policy shift is occurring.

As <TIMEX2>the tenth anniversary of the Pan Am bombing</TIMEX2> draws near, a policy shift is occurring.

In these next two examples, the context allows VAL to be specified, as is often the case in news stories.

The discovery of the problem came on *my birthday, December twenty-ninth*.

The discovery of the problem came on <TIMEX2 VAL="1998-12-29">my birthday, <TIMEX2 VAL="1998-12-29">December twenty-ninth</TIMEX2></TIMEX2>.

02/19/1998 Two Russians and a Frenchman left the Mir and endured a rough landing on the snow-covered plains of Central Asia on *Thursday*. But no one was hurt as the crew completed a successful mission on *the 12th anniversary of the space station's launch*.

<TIMEX2 VAL="1998-02-19">02/19/1998</TIMEX2> Two Russians and a Frenchman left the Mir and endured a rough landing on the snow-covered plains of Central Asia on <TIMEX2 VAL="1998-02-19">Thursday</TIMEX2>. But no one was hurt as the crew completed a successful mission on <TIMEX2 VAL="1998-02-19">the 12th anniversary of the space station's launch</TIMEX2>.

4.9 Culturally-Determined Expressions

This section deals with temporal expressions whose interpretation typically requires cultural or domain-specific knowledge. In general, these expressions are markable but receive no VAL. Specific end-users of the TIMEX2 tagging scheme may then choose to specify VAL where appropriate for their applications. For further discussion and examples, refer to the following sections.

4.9.1 Alternative Calendars

The ISO standard, upon which this tagging scheme is based, is designed to represent dates in the Gregorian calendar. This is admittedly a cultural bias, because there are other calendars in use throughout the world, such as the Chinese calendar or the Hindu calendar. Then there are those temporal units whose meanings are only understood with relation to a particular field of endeavor, such as academia (e.g., “the school year”), sports (e.g., “baseball season”), television (e.g., “prime time”), and political terms in office (e.g., “fourth term”).⁴⁵

To the speakers who use these expressions, they might correlate to some coordinates on the Gregorian calendar, but outside the group of users, the meaning is not well understood.

Therefore, when these expressions have a *lexical trigger* as their head, they are markable, but receive no VAL.

⁴⁴ Assume for the purposes of this example that it's not clear from the context if the party actually took place on his birthday.

⁴⁵ Fiscal years are also an form of alternative calendar. These are identified with the FY token, discussed in section 4.3.4.

On the Chinese calendar, *this is the year 4688, which is the year of the horse.*

On the Chinese calendar, <TIMEX2>this</TIMEX2> is <TIMEX2>the year 4688, which is <TIMEX2>the year of the horse</TIMEX2></TIMEX2>.

The Braves' starting pitching, the best and deepest in the major leagues, dominates in *the regular season* because no team can match the starting five.

The Braves' starting pitching, the best and deepest in the major leagues, dominates in <TIMEX2 NON-SPECIFIC="YES">the regular season</TIMEX2> because no team can match the starting five.

The Atlanta Braves are not playing in the World Series for the third time in their *seven-season* run as a division champion.

The Atlanta Braves are not playing in the World Series for the third time in their <TIMEX2>seven-season</TIMEX2> run as a division champion.

One of the old timers who had covered the White House since *Franklin Roosevelt's first term* surprised us when he said that if he was ever marooned on a desert island, he would like Goldwater as a companion.

One of the old timers who had covered the White House since <TIMEX2>Franklin Roosevelt's first term</TIMEX2> surprised us when he said that if he was ever marooned on a desert island, he would like Goldwater as a companion.

Security forces have held back from intervening, despite a government ban on demonstrations and large gatherings during *the election period*.

Security forces have held back from intervening, despite a government ban on demonstrations and large gatherings during <TIMEX2>the election period</TIMEX2>.

She was not allowed to play video games *last school year*, and her grades improved dramatically.

She was not allowed to play video games <TIMEX2>last school year</TIMEX2>, and her grades improved dramatically.⁴⁶

And they're what young designers around the world say are *the season's* best buys.

And they're what young designers around the world say are <TIMEX2>the season</TIMEX2>'s best buys.⁴⁷

Note that in the following example an embedded expression using the Gregorian calendar is fully annotated, but the expression referring to the baseball season is given no VAL:

The 1994 baseball season was cut short by the strike.

<TIMEX2>The <TIMEX2 VAL="1994">1994</TIMEX2> baseball season</TIMEX2> was cut short by the strike.

4.9.2 Other Expressions Open to Historical and Individual Interpretation

Some expressions' meanings are understood in some fuzzy sense by the general population and not limited to specific fields of endeavor. However, the general rule is that no VAL is to be specified if they are culturally or historically defined, because there would be a high degree of disagreement over the exact value of VAL.

But he was among the few workers in the area who resisted gawking at the *lunchtime* spectacle.

But he was among the few workers in the area who resisted gawking at the <TIMEX2>lunchtime</TIMEX2> spectacle.

⁴⁶ Although "school year" is not given a VAL, "semester" can use a season token in VAL in certain circumstances. See section 4.3.3.

⁴⁷ This is "season" as the fashion industry defines it.

Malaysia is making an aggressive move into *the information age*.

Malaysia is making an aggressive move into <TIMEX2>the information age</TIMEX2>.

But 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*.

But 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>.

Nellie Campobello was an artist from *Mexico's golden age*.

Nellie Campobello was an artist from <TIMEX2>Mexico's golden age</TIMEX2>.

4.9.3 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 *lexical trigger* words. Others, like “Thanksgiving,” “Christmas,” and “Diwali,” do not. A holiday name is markable (sorry, there is NO fixed list of holidays!), but should receive a value only when that value can be inferred from the context of the text, rather than from cultural and world knowledge. This rule is to make the annotation standard applicable to the broadest range of users possible. Specific end-uses of the TIMEX2 scheme are free to assign values to holidays if it is feasible and necessary in their particular application.

In the following, the X placeholder in VAL is used to indicate the granularity of the lexical expression:

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 VAL="XXXX-XX-XX">New Year's Day</TIMEX2> hangover.

Some “holidays” are generic set-denoting expressions with some information that can be captured in VAL. Here, we know the day is a Sunday:

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

We always watch the game on <TIMEX2 VAL="XXXX-WXX-7" SET="YES">Super Bowl Sunday</TIMEX2>.

Holidays can also be definite and referential, but should not receive a VAL if one is not provided by the context.

My cousins came to visit on *Christmas*.

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

4.10 Expressions Whose Values can Change

Recall that the VAL attribute contains a normalized form of the date/time that is referenced by the tagged expression. In the following example, one might be inclined to say that “the gestation period for humans” refers to a period of “nine months.”

The gestation period for humans is normally nine months.

However, in the following examples, it becomes clearer that the value associated with such expressions is not fixed:

The gestation period for horses is 11 to 12 months.

With the new trains, the time it takes to travel between Washington D.C. and Boston has been reduced from nine hours to six hours.

Analogous to these examples are those dealt with in the coreference annotation task definition (Hirschman and Chinchor 1997) from the Sixth and Seventh Message Understanding Conferences (MUC-6 and MUC-7):

Henry Higgins, who was formerly sales director of Sudsy Soaps, became president of Dreamy Detergents. Sudsy Soaps named Eliza Dolittle as sales director effective last week.

The stock price fell from \$4.02 to \$3.85.

In the first example, the position “sales director of Sudsy Soaps” was associated with Henry Higgins at one time, but is now associated with Eliza Dolittle. In the second, the value of the stock price changes over time. Following the suggestion made by van Deemter & Kibble (2001), we treat such expressions as non-referring concepts, so that the values associated with them are viewed as *properties* of the concepts rather than referents. In the case of temporal concepts (period, time, etc.), the expression should receive a TIMEX2 tag. However, because we take the point of view that they are non-referring, they do not receive a VAL. A different type of annotation scheme, one capturing properties of entities and concepts, would perform the association between the expression and the properties. Note that the properties themselves are straightforward TIMEX2 expressions.

The gestation period for horses is 11 to 12 months.

<TIMEX2>The gestation period for horses</TIMEX2> is <TIMEX2 VAL="P11M">11</TIMEX2> to <TIMEX2 VAL="P12M">12 months</TIMEX2>.

With the new trains, *the time it takes to travel between Washington D.C. and Boston* has been reduced from *nine hours* to *six*.

With the new trains, <TIMEX2>the time it takes to travel between Washington D.C. and Boston</TIMEX2> has been reduced from <TIMEX2 VAL="PT9H">nine hours</TIMEX2> to <TIMEX2 VAL="PT6H">six</TIMEX2>.

4.11 Metonymic Expressions

People commonly use a reference to one entity to refer to another, e.g., “The ham sandwich is waiting for his check” or “Vietnam changed American’s attitudes towards war.” In the first example, the speaker is really referring to the person who ordered the ham sandwich, and in the second example, the speaker is referring to the conflict in Vietnam and not the country. These are examples of *metonymy*, and they occur with temporal expressions as well. Most notably, the expressions “September 11,” “9/11” etc. have come to refer to the terrorist attacks that occurred on 9/11, and not simply to the date itself. Thus, just as in the previous examples where a

sandwich stands for a person and a location stands for an event, in this case a date stands for an event. Because we anticipate that downstream information processing will want to link the date to the thing it stands for, all cases of temporal metonyms should be given a TIMEX2 tag, with a VAL if it can be determined based on information in the document. To be consistent with other event-anchored expressions, annotators may not use world-knowledge to determine VAL.

But *September 11* was possibly one of the most pivotal events in world history.

But <TIMEX2 VAL="XXXX-09-11">September 11</TIMEX2> was possibly one of the most pivotal events in world history.

They continue the investigation into who was behind 9/11.

They continue the investigation into who was behind <TIMEX2 VAL="XXXX-09-11">9/11</TIMEX2>.

Another common example of date-event metonymy from recent history is the killing of demonstrators on June 4, 1989 in Beijing's Tiananmen Square:

But the United States and the whole of the West have again and again raised the *June 4* problem and the human rights problem of China.

But the United States and the whole of the West have again and again raised the <TIMEX2 VAL="XXXX-06-04">June 4</TIMEX2> problem and the human rights problem of China.

4.12 Expressions You Just Can't Figure Out

Most of the examples in this document have come from real-world texts such as annotated news articles and broadcast news transcripts, and for the most part they accurately reflect what is to be found in real-world documents. However, human annotators will find that uncertainties can arise. Writers and speakers will occasionally provide conflicting information that leads to different interpretations of temporal expressions, and at other times writers and speakers simply don't provide enough information to permit interpretation. Often a precise expression will be used, but the reader may be in doubt as to how precisely the writer or speaker really intended the expression to be taken. In all these cases, the need to fill in the VAL attribute in a consistent manner can be at odds with the real-world sloppiness of many texts, and it would require reading the original author's mind in order to ensure that VAL is accurate.

4.12.1 Use that Comment Field!

Our guidance to the annotators in unclear cases is "do the best you can." In all cases, follow the rules of thumb intended to maximize consistency among annotators. But where no rules apply, use your best judgment. In addition, it is very useful to users of annotated corpora and the developers of the TIMEX2 standard when annotators use the COMMENT attribute to highlight problems, uncertainties, lack of coverage etc. For example:

The king lived *4,000 years ago*.

The king lived <TIMEX2 VAL="BC2001" COMMENT="I doubt the writer really meant 4000 exactly">4,000 years ago</TIMEX2>.

If certain problems occur systematically, annotators should choose a single consistent comment to flag all such cases, so that they can later be extracted and analyzed in more detail. For

example, in the TERN 2004 corpus we encountered some mistakes in the transcription and other garbled utterances that made it difficult to normalize some temporal expressions. Where that occurred, we put “T-ERROR” (for “transcription error”, though they may not have all been due to transcription problems) in the comment field.

5 Determining the Extent of the Annotations

The previous sections introduced the annotation format and focused on how to capture the meaning of temporal expressions. Up until now we've left unstated the rules for determining where each temporal expression begins and ends. We refer to this as the *extent* of the expression.

5.1 Lexical Criteria

When tagged, the full extent of the tag must be one of the following grammatical categories:

- noun (including proper noun): e.g., “today” “Thursday”
- noun phrase (NP): e.g., “the morning” “Friday night” “the last two years”
- adjective: e.g., current
- adverb: e.g., recently
- adjective/adverb phrase: e.g., “half an hour long” “two weeks ago” “nearly half-hour”

The full extent cannot be a prepositional phrase (i.e., in English the expression cannot start with a preposition) or a clause of any type (for example, in English the expression cannot start with a subordinating conjunction). Thus, from the following phrases we get the following annotated expressions:

| Context | Extent of TIMEX2 Expression |
|--|-----------------------------|
| “...before Thursday...” | Thursday |
| “...in the morning...” | the morning |
| “...after the strike ended on Thursday...” | Thursday |
| “...over the last 2 years...” | the last 2 years |

Note that expressions like “around,” “about,” “at least” and “up to” (but not “up until”) function more as modifiers than prepositions, and thus should be included in the extent of the tag.

5.2 Syntactic Criteria

The full extent of the tagged expressions includes all **premodifiers** of the time expression, among which are determiners, possessive noun phrases, and all nominal/adjectival premodifiers, including not only those premodifiers that qualify as MOD attributes, but also those premodifiers

with no corresponding MOD token. For example, all of the following represent the full extent of markable temporal expressions:

| | |
|--|---|
| <i>that cold day</i> <i>the next day</i> <i>late last night</i> <i>earlier that year</i> <i>next summer</i> <i>recent decades</i> <i>numerous Saturdays</i> <i>more than a month</i> <i>no less than 60 days</i> | <i>just a year ago</i> <i>only one hour long</i> <i>its own future</i> <i>the country's future</i> <i>just a year ago</i> <i>only one hour long</i> <i>five years old</i> <i>a few weeks later</i> <i>hours earlier</i> |
|--|---|

Note also that in adjective phrases like “a year ago” “five years old” and “one hour long,” both the adjective and the preceding noun phrase are included within the extent.

The full extent also includes all **postmodifiers** of the time expression, among which are adverbs, 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
three decades ago

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, in the following, the string “order to conduct a search for the body” is not included in the TIMEX2 tag, and neither is the non-temporal premodifier of that string, namely “his.”

They knew about his *recent* order to conduct a search for the body.

They knew about his <TIMEX2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="1999-07-15">recent</TIMEX2> order to conduct a search for the body.

In this next example, neither the determiner “the” before “8:00” nor the noun “meeting” after it is included within the scope of either TIMEX2 tag.

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

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

In this last example, 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.”

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?

5.2.1 Appositives

An appositive to a temporal expression is considered a post-modifier and is thus included within the extent. If the appositive phrase itself contains a temporal expression, it is given a separate tag as well.

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

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

5.2.2 Range 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

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>.

The class is *3-6 pm today*.

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

Dinner is from *five* to *six pm tomorrow*.

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

It usually takes five to six years to complete a project of this scope.

It usually takes <TIMEX2 VAL="P5Y">five</TIMEX2> to <TIMEX2 VAL="P6Y">six years</TIMEX2> to complete a project of this scope.

5.2.3 Conjoined Expressions

Phrases involving conjunction or disjunction of time expressions are handled similarly to ranges.

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

The bug will get fixed between <TIMEX2 VAL="PRESENT_REF" ANCHOR_VAL="1999-07-15" ANCHOR_DIR="AS_OF">now</TIMEX2> and <TIMEX2 VAL="1999-07-19TMO">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" ANCHOR_VAL="1999-07-15" ANCHOR_DIR="AS_OF">now</TIMEX2></TIMEX2>.

Many cases of conjoined expressions involve elision, but two tags should still be created:

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" ANCHOR_VAL="1999" ANCHOR_DIR="AFTER">at least the next year</TIMEX2> or <TIMEX2 VAL="P2Y" MOD="EQUAL_OR_MORE" ANCHOR_VAL="1999" ANCHOR_DIR="AFTER">two</TIMEX2>.⁴⁸

Clinton said he was not concerned about the possibility of a recession in the short and mid-term.
 Clinton said he was not concerned about the possibility of a recession in <TIMEX2 VAL="FUTURE_REF" ANCHOR_VAL="2000-09-15" ANCHOR_DIR="AFTER">the short</TIMEX2> and <TIMEX2 VAL="FUTURE_REF" ANCHOR_VAL="2000-09-15" ANCHOR_DIR="AFTER">mid-term</TIMEX2>.

However, in cases like the following, create one tag and use the EQUAL_OR_MORE value in MOD.

The two Cambodian prime ministers requested that the UN Secretary General extend the stay of his representative in Cambodia by six months or more.
 The two Cambodian prime ministers requested that the UN Secretary General extend the stay of his representative in Cambodia by <TIMEX2 VAL="P6M" MOD="EQUAL_OR_MORE">six months or more</TIMEX2>.

5.2.4 Embedded Expressions

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.

5.2.4.1 When to Create One Tag

An expression is treated as an indivisible syntactic unit if (a) there is no intervening element between temporal terms and (b) the two terms express values for units that are hierarchically related. For example:

Friday evening (where the day unit is larger than and contains the part-of-day unit)

November 1943 (where the month unit is smaller than and contained within the year unit)

Thus, each of the following expressions represents the full extent of a single TIMEX2 tag:

Friday evening
8:00 p.m. Friday
Tuesday the 18th

⁴⁸ The MOD is shown as applying to each half of the disjunction, although that is somewhat misleading semantically.

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

Prepositions introduce syntactically embedded phrases and therefore typically introduce an embedded TIMEX2 tag (see section 5.2.4.2, below). However, there are six notable exceptions in which you should ignore the preposition and create a single tag:

1. The preposition “of.” Thus, each of the following is treated as a single expression:

the second of December

October of 1963

summer of 1964

the morning of January 31

ten of two

2. Partitives and groups should be tagged as a single extent. These would also be covered by the previous rule, but these fall into a distinct subclass because they do not use a time word to identify the subpart of the larger temporal unit:

much of the week

most of the year

all of the year

the rest of the year

near the end of the year

a group of years

a string of quarters

3. Prepositions like “to, till, after” etc. used in expressing **TODs**. Thus, the following is each treated as a single expression:

ten minutes to three

five till eight

twenty after twelve

half past noon

4. The preposition “in” in **TODs**. Thus, the following is treated as a single expression:

eleven in the morning

5. Prepositions preceded by the adjectives such as “early,” “earlier,” “late,” and “later,” and for which there are noun phrase paraphrases. The preposition is serving very little role here, so it is ignored.

earlier in the year (cf. “earlier that year”)

late on Friday (cf. “late Friday”)

later on Friday (cf. “later on Friday”)

6. Prepositions in rate expressions. These expressions can contain quantifiers, articles, or prepositions, so we treat them all equally. They are all SET expressions.

(eight glasses of water) *per day* (cf. “eight glasses of water every day,” “...each day,” “...a day”)

5.2.4.2 When to Create Multiple Tags, with Embedding

There are three situations in which a TIMEX2 tag is embedded in the extent of another:

1. **Time-Anchored Expressions.** If a temporal expression includes an explicit anchor (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>.

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>.⁴⁹

2. **Possessive Constructions.** 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.

3. **Pre-Modifier and Head are Both Triggers.** When a collocation consists of two triggers that are not hierarchically related temporally, treat it as a compound noun, where the larger term is given one tag, and the pre-modifier(s) is/are given another.

her September 2 birthday

<TIMEX2 VAL="1999-09-02">her <TIMEX2 VAL="1999-09-02">September 2</TIMEX2> birthday</TIMEX2>

the holiday season

<TIMEX2>the <TIMEX2>holiday</TIMEX2> season</TIMEX2>

the June 2004 period

<TIMEX2 VAL="2004-06">the <TIMEX2 VAL="2004-06">June 2004</TIMEX2> period</TIMEX2>

a December 12 deadline

⁴⁹ Note that an elided “from” is assumed in the analysis of the syntax that is reflected by this representation.

<TIMEX2 VAL="1999-12-12">a <TIMEX2 VAL="1999-12-12">December 12</TIMEX2>
deadline</TIMEX2>

the April-June period

<TIMEX2 VAL="P3M">the <TIMEX2 VAL="XXXX-04">April</TIMEX2>-<TIMEX2 VAL="XXXX-06">June</TIMEX2> period</TIMEX2>

She worked *a 14-hour day*.

She worked <TIMEX2 VAL="P1DT">a <TIMEX2 VAL="PT14H">14 hour</TIMEX2>
day</TIMEX2>.

There are two **exceptions** to the above rule, and those are the expressions *twelve o'clock midnight*, and *twelve o'clock noon*, where in each expression both terms express values for the same **TOD** unit. However, this is a unique collocation, intended to disambiguate twelve noon from twelve midnight. Thus, these should be tagged as a single extent.

5.2.4.3 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 range expressions and conjoined expressions, which were discussed above.) 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.

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="1999-07-20T12:00">12 PM</TIMEX2>?

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

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

Note that we analyze the above expressions as being syntactically independent; i.e., as [NP] [P [NP]], not [NP [P [NP]]]. In many cases you can check for syntactic independence by seeing if the meaning is maintained when the constituents are transposed. Sometimes a minor change in the connecting words is required, but if the overall meaning is not changed, this indicates syntactic independence. This test works for three of the previous examples:

8:00 p.m. on Friday → Friday at 8:00 p.m.

Friday at 8:00 p.m → 8:00 p.m. on Friday

Tuesday after 12 PM → after 12 PM on Tuesday

In other cases, you can try deleting the second expression to see if the first expression can stand on its own in the context of the sentence. If so, it is considered independent. This works for the example “I tutored an English student some Thursdays in 1998.” → “I tutored an English student some Thursdays.” But note that it does not work for an example from section 5.2.4.2, “I’m leaving on vacation two weeks from next Tuesday” → *I’m leaving on vacation two weeks.” Thus, the latter example is not tagged as two independent phrases.

5.3 Punctuation

Typically, punctuation on the boundaries of the alphanumeric expression is excluded from the extent of the tag, including hyphens, dashes, periods etc. One known exception is punctuation which marks an abbreviated form of a temporal expression.

Chicago '68 was more than just another in a series of antiwar protests.

Chicago <TIMEX2 VAL=“1968”>'68</TIMEX2> was more than just another in a series of antiwar protests.

Glossary

Anchored point 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.

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

Coreference. Two expressions are said to be “coreferring” if they refer to the same thing. In temporal expressions, coreference occurs when expressions refer to the same time value. For example, given “*Tuesday...that day*,” where “*that day*” is referring to the same day as “*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*.”

Duration. An expression which describes an interval of time, indicating explicitly how long it lasts, e.g., “*three hours long*.” Also known as a “period,” particularly within the ISO 8601 document.

Event-anchored time expression. A time expression which is dependent on an event for its value to be fully resolved, e.g., “*the day after our meeting*.”

Extent. The lexical span of a **markable** TIMEX2 expression, or, in other words, where it begins and ends in the text. For those who are concerned with electronic date representation, the extent is delimited by “byte offsets” within the electronic form of the text.

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

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 VAL.

Lexical trigger. A designated time word whose presence means the expression is **markable**. Our set of lexical triggers includes words like “day,” “week,” “weekend,” “now,” “Monday,” “current,” and “future.” Since “when” is not a lexical trigger, an expression like “*when the stock market crashed*” is not to be marked. Although “September” is a lexical trigger, “Black September” is not to be marked since the phrase is not being used primarily as an expression of time.

Markable. An expression that should receive a TIMEX2 tag. An expression can be markable but still receive no VAL.

Non-specific time expression. A temporal expression where no specific time is indicated. These include generics, as well as indefinites, like “*He left on a sunny day in June.*”

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

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

TOD. Time of Day, as expressed in 5.3 of the ISO standard (e.g., hhmmss). 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”), START (for “early” etc.), PRESENT_REF (for “today,” “current,” etc.).

References

- Alembic Workbench, available for download at <http://www.mitre.org/tech/alembic-workbench/>
- “Annotation Guidelines for Relation Detection and Characterization (RDC),” available at <http://www.nist.gov/speech/tests/ace/index.htm>.
- Alexandersson, J, Riethinger, N. and Maier, E. (1997). “Insights into the Dialogue Processing of VERBMOBIL.” *Proceedings of the Fifth Conference on Applied Natural Language Processing*, 33-40.
- Busemann, S., Decleck, T., Diagne, A. K., Dini, L., Klein, J. and Schmeier, S. (1997). “Natural Language Dialogue Service for Appointment Scheduling Agents.” *Proceedings of the Fifth Conference on Applied Natural Language Processing*, 25-32.
- Callisto Configurable Annotation Tool. Available for download at <http://callisto.mitre.org>.
- Damianos, L., Ponte, J., Wohlever, S., Reeder, F., Day, D., Wilson, G., Hirschman, L. (2002). “MiTAP for Bio-Security: A Case Study.” *AI Magazine*, Volume 23, No. 4, 13-29.
- Ferro, L. (2001). “Instruction Manual for the Annotation of Temporal Expressions.” MITRE Technical Report MTR 01W0000046. McLean, Virginia: The MITRE Corporation.
- Ferro, L., Mani, I., Sundheim, B. and Wilson, G. (2001). “TIDES Temporal Annotation Guidelines - Version 1.0.2.” MITRE Technical Report MTR 01W0000041. McLean, Virginia: The MITRE Corporation. June 2001.
- Ferro, L., Mani, I., Sundheim, B. and Wilson, G. (2000). “TIDES Temporal Annotation Guidelines – Draft Version 1.0”. MITRE Technical Report MTR 00W0000094. McLean, Virginia: The MITRE Corporation.
- Gerber, L., Ferro, L., Mani, I., Sundheim, B., Wilson, G., and Kozierok, R (2002). “Annotating Temporal Information: From Theory to Practice.” *Proceedings of the 2002 Conference on Human Language Technology*. San Diego, CA, 2002.
- Ingria, R and Pustejovsky, J. (2002). “TimeML Specification 1.0.” Available at <http://www.time2002.org>.
- “KSL-Time Ontology” (1999). Available at <http://www.ksl.Stanford.EDU/ontologies/time/>.
- Mani, I. and Wilson, G. (2000). “Robust Temporal Processing of News”. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL'2000)*, 69-76.
- Mani, I., Wilson, G., Sundheim, B., and Ferro, L. (2001). “Guidelines for Annotating Temporal Information”. In *Proceedings of HLT 2001, First International Conference on Human Language Technology Research*, J. Allan, ed., Morgan Kaufmann, San Francisco, 2001.

- Pustejovsky, J. and Busa F. (1995). "A Revised Template Description for Time (v3)." Available at http://www.cs.nyu.edu/cs/faculty/grishman/time-guidelines.v3_1.html
- TDT-2. Topic Detection and Tracking Program
<http://morph.ldc.upenn.edu/Catalog/LDC99T37.html> 1999.
- Setzer, A. (2001). *Temporal Information in Newswire Articles: an Annotation Scheme and Corpus Study*, Unpublished PhD thesis, University of Sheffield.
- Setzer, A and Gaizauskas, R. (2000). "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.
- van Deemter, K. and Kibble, R. (2001). "On Coreferring: Coreference in MUC and Related Annotation Schemes." *Computational Linguistics*, Vol. 26, No. 4., 631-637.
- Wellner, B., Ferro, L, Greiff, W, and Hirschman, L. (2003). "Evaluating Language Understanding Through Reading Comprehension Tests," to appear *Natural Language Engineering*.
- Wiebe, J. M., O'Hara, T. P., Ohrstrom-Sandgren, T., and McKeever, K. J. (1998). "An Empirical Approach to Temporal Reference Resolution." *Journal of Artificial Intelligence Research*, 9, 247-293.

Appendix A: An Example Annotated Document

<TIMEX2 VAL="1623-07-XX">Day of Thanks</TIMEX2> To Be Held

(PLYMOUTH, MASSACHUSETTS, <TIMEX2 VAL="1623-07">July, 1623</TIMEX2>) - <TIMEX2 VAL="1623-07">This month</TIMEX2>, <TIMEX2 VAL="1623-07-XX">a day of thanks</TIMEX2> will be held. On <TIMEX2 VAL="1623-07-XX">that day</TIMEX2>, people will eat a great feast. They will be thankful for the rains that made a good crop.

<TIMEX2 VAL="1621-FA-XX">The first day of thanks</TIMEX2> was held in <TIMEX2 VAL="1621">1621</TIMEX2>. The people, called Pilgrims, gave thanks for their lives. They had come to this new land from England in <TIMEX2 VAL="1620">1620</TIMEX2>. <TIMEX2 VAL="1621-WI">The first winter</TIMEX2> was <TIMEX2 VAL="1621-WI">a long, hard one</TIMEX2>. Many men, women, and children died.

By <TIMEX2 VAL="1621-SP">spring</TIMEX2>, only 55 Pilgrims were left. They were very sad. An Indian named Squanto came to help them. He gave them seeds and showed them how to grow a food called corn.

By <TIMEX2 VAL="1621-FA">fall</TIMEX2>, the Pilgrims had enough food for <TIMEX2 VAL="1622-WI">the winter</TIMEX2>. They thought it would be nice to have a party. They asked Squanto to come and bring his friends.

For <TIMEX2 VAL="P3D" ANCHOR_DIR="WITHIN" ANCHOR_VAL="1621-FA">three days</TIMEX2>, the Pilgrims cooked deer, turkeys, ducks, and fish. They baked bread. Squanto showed them how to make pudding from corn.

Then the Pilgrims and the Indians ate and played games. "This is fun," they said. "We'll hold <TIMEX2 VAL="XXXX-XX-XX">a day of thanks</TIMEX2> <TIMEX2 SET="YES" VAL="XXXX">each year</TIMEX2>."

1. Who helped the Pilgrims?
2. What did they eat on <TIMEX2 VAL="1621-FA-XX">the first day of thanks</TIMEX2>?
3. When was <TIMEX2 VAL="1621-FA-XX">the first day of thanks</TIMEX2> held?
4. Where did the Pilgrims live before coming to America?
5. Why were the Pilgrims sad?

Source text: Copyright 1989 Remedia Publications

Appendix B: Future Directions and Open Issues for Set Expressions

As discussed in section 4.5, we are modifying the manner in which set expressions are to be encoded. We have eliminated the GRANULARITY and PERIODICITY attributes, and are exploring a different approach to capturing the semantics. However, this approach needs more research before it can be made part of the official guidelines. We therefore present here what our current model is and describe some open issues that we are investigating.⁵⁰

Our basic approach is to specify two templates for set expressions, one template being embedded in other:

1. number of temporal units, e.g., “every day”
2. frequency per number of temporal units, e.g., “once every day”

To handle the first type, we need a QUANT attribute for “number of.” VAL, ANCHOR_DIR, and ANCHOR_VAL can be used to encode the temporal units. For the second type, we also need a FREQ attribute. The following table illustrates how these attributes could be used to capture a variety of set expressions. (Keep in mind that the SET=“YES” attribute would be present in all cases as well.)

| Frequency | (per) | Number of | Temporal Units | | | |
|-----------|-------|-----------------|----------------|------------|------------|------------------------------------|
| | | | Unit | Anchor | | Linguistic Expression |
| FREQ | | QUANT | VAL | ANCHOR_DIR | ANCHOR_VAL | |
| | | EVERY | TXX | | | hourly |
| | | LESS_THAN_EVERY | XXXX-WXX | | | almost weekly |
| | | EVERY | XXXX-07 | | | every July |
| | | EVERY | 1999-WXX-4 | | | every Thursday in 1999 |
| | | 3 | XXXX-SU | BEFORE | 1999 | the past three summers |
| | | X | 1998-WXX-6 | WITHIN | 1998-SU | numerous Saturdays last summer |
| | | X | 1999-WXX-4 | | | occasional Thursdays in 1999 |
| | | X | 1999-07-20TXX | | | the hours that you are not |
| | | X | XXXX | BEFORE | 1999 | recent years |
| 1X | | | XXXX-WXX-4 | | | once every Thursday |
| 2X | | EVERY | XXXX-XX | | | twice a month, twice monthly |
| 3D | | EVERY | XXXX-XX | | | three days every month |
| XD | | EVERY | XXXX-WXX | | | several days a week |
| 2X | | | 1999 | | | two times in 1999 |
| 3X | | | P5Y | ENDING | 1998 | three times in the past five years |

⁵⁰ The proposal described here has benefited greatly from conversations with the TimeML (timeml.org) group.

What remains to be done is to test this approach on fair amount of real-world data. In applying this approach to the handful of examples in the existing guidelines, we've encountered the following issues:

- Except for the last example in the table, we have used the point format in VAL, because it easily allows for greater specificity. For example, it provides a straightforward way to indicate “every July.” But particularly for the expressions that contain a `FREQ` value, VAL is much more period-like. Should the period format be used in VAL in these cases? In all SET expressions? Then how would we capture “two times in 1999”?
- Should “hourly” be considered equivalent with “one time (once) every hour,” and thus have a `FREQ` value? Or is that reading too much into the semantics? Our current inclination is that the `TIMEX2` tag should reflect what is in the surface form of the lexical expression, and not infer a `FREQ` if it isn't overtly stated. In contrast, expressions like *semiweekly* (twice a week) or *biweekly* (once every two weeks) should contain a `FREQ` attribute, because the frequency is contained in the word morphology (*semi-*, *bi-*).
- How do we encode complex quantifiers like “every other day,” “every third day,” “every two weeks” etc.?⁵¹ Expanding the set of possible values for `QUANT` is a possible solution (e.g., `EVERY_OTHER`), but the cardinality is theoretically infinite.
- How should we encode modified set expressions? To handle expressions like “almost weekly,” we are exploring the possibility of enriching the allowable values for `QUANT` to include things like `LESS_THAN EVERY` or `MORE_THAN EVERY`. But in expressions like “at least once a week” the modifier applies to `FREQ`, not `QUANT`. It may be necessary to add a `MOD_FREQ` attribute.
- How should we encode anchored frequencies, as in “the first three days of each month”? We cannot use the anchoring attributes to indicate “first three days,” because the anchoring attributes apply to VAL, not `FREQ`. It may be necessary to add anchoring attributes for `FREQ`.

⁵¹ This interesting puzzle was point out to us by James Pustejovsky, personal communication.

Index

- [the] *time*, 25, 48, 49
- 9/11, 54
- about*, 56
- academic years, 50
- adjectives, 6, 56, 57
- adverbs, 7, 56
- AF, 32
- after*, 9
- AFTER
 - ANCHOR_DIR attribute, 18, 20
 - MOD attribute, 36
- afternoon*, 32
- afterwards*, 9
- ago*, 13, 22
- ahead*, 8, 9
- already*, 9
- always*, 39, 45
- ANCHOR_DIR, 12, 18
- ANCHOR_VAL, 12, 18
 - granularity in duration expressions, 19
 - granularity in FUTURE_REF expressions, 26
 - granularity in PAST_REF expressions, 25
 - granularity in PRESENT_REF expressions, 24
- anchored point expressions, 13, *See also*
 - ANCHOR_VAL and ANCHOR_DIR
- and*. *See* conjunctions
- anniversary*, 50
- annual*, 40
- another*, 19
- appositives, 58
- APPROX, 37
- approximately*, 37
- around*, 56
- AS_OF, 18
- aspect, 36
- at least*, 56
- attributes, 11
- autumn*, 28
- BCE, 14
- before*, 9
- BEFORE
 - ANCHOR_DIR attribute, 18, 20
 - MOD attribute, 36
- Before Current Era (BCE or BC) dates, 14
- beforehand*, 9
- billions of years ago*, 35
- birthday*, 49
- century*, 13, 21
- Chinese calendar, 50
- COMMENT, 12, 54
- conjunctions, 58
- couple*, 25, 26, 35
- daytime*, 32
- decade*, 13, 21
- definite article. *See* determiners
- determiners, 56
- disjunctions, 37, 58
- DT, 32
- each*, 39
- early*, 37
- early, earlier*, 9
 - extent rules*, 61
- elision, 46
- END, 37
- ENDING, 18
- EQUAL_OR_LESS, 36
- EQUAL_OR_MORE, 36
- EV, 32
- evening*, 32
- event-anchored expressions, 46
- eventually*, 8, 9
- ever*, 9
- every*, 39
- everyday*, 40
- FA, 28
- fall, 28
- few*, 25, 26, 35
- finally*, 9
- first*, 9
- following*, 9
- frequency expressions
 - bare (no time period given), 9
 - within time periods. *See* set expressions
- frequent, frequently*, 10
- FUTURE_REF, 23
- FY, 30
- generic, 41, 52
- genitives. *See* possessive constructions
- geological eras, 14
- GMT, 16
- GRANULARITY, 39
- Greenwich Meridian Time, 16
- H1,H2,HX, 31
- half* (of a year), 30
- heretofore*, 9
- Hindu calendar, 50
- holidays, 52
- idioms, 45
- immediately*, 9
- in*
 - in TOD expressions, 60
- indefinite noun phrases, 41, 43
- ISO standard, 11
- it*, 46
- Jurassic Period*. *See* geological eras
- last night*, 33

late, 37
late, later
 extent rules, 61
later, 9
 LESS_THAN, 36
 lexical trigger, 6, 41
long, 17
longstanding, 9
meanwhile, 8
 MI, 32
 MID, 37
mid-day, 32
midnight, 15, 44
millennium, 13, 21
millions of years ago, 35
 MO, 32
 MOD, 12, 36–39, 56
 modifiers, 36
momentarily, 9
monthly, 40
 MORE_THAN, 36
morning, 32
 named days. *See* holidays
 negation, 9
next, 9
 NI, 32
night, 32
noon, 44
 normalization, 5
 nouns, 6, 56
now, 7
 numbers, 7
o'clock, 15
of, 60
often, 10
 ON_OR_AFTER, 36
 ON_OR_BEFORE, 36
once, 9, 24
only, 57
or. *See* disjunctions
originally, 9
 PAST_REF, 24
per, 61
 period of time, 17
 PERIODICITY, 39
 plurals, 34
 political terms in office, 50
 possessive constructions, 61
 possessive pronoun, 56
 postmodifiers, 57
preceding, 9
 premodifiers, 56
 prepositions, 8, 36, 56, 60
 PRESENT_REF, 23
previous, previously, 9
prime time, 50
 pronoun, 45
 pronouns, 7, 56
 proper names, 6, 10, 43
 Q1-Q4, 30
 quantifiers, 36
quarter, 30
 range expressions, 13, 58
 school years. *See* academic years
semester, 28
semiannual, 40
September 11, 54
 SET, 12, 39–41
 set expressions, 39–41, 69–70
 non-specific, 45
 SGML, 11
since. *See* subordinating conjunctions
so far, 9
some, 26, 41
 SP, 28
 sports seasons, 50
spring, 28
 START, 37
 STARTING, 18
still, 9
 SU, 28
 subordinating conjunctions, 8
subsequent, 8
subsequent, subsequently, 9
summer, 28
 tag attributes. *See* attributes
the nth (straight) day in a row, 19
then, 9
thousands of years ago, 35
till, 60
time, 9
 TIMEX2 tag, 11
to
 in TOD expressions, 60
 token, 22, 23, 31, 32
 trigger. *See* lexical trigger
twelve o'clock midnight, 62
twelve o'clock noon, 62
 Universal Coordinated Time, 16
up to, 56
usually, 45
 UTC, 16
 VAL, 12
weekend, 31
 weeks, 17, 35
when. *See* subordinating conjunctions
while. *See* subordinating conjunctions
 WI, 28
winter, 28, 29
 WITHIN, 18
 X placeholder
 for unspecified components of VAL, 33

for unspecified plural numbers in durations,
34, 35
in set expressions, 39

year-old, n-year-old, 19, 57
yet, 9