MTR 01W0000046V01

MITRE TECHNICAL REPORT

TIDES

Instruction Manual for the Annotation of Temporal Expressions

June 2001

Lisa Ferro

Contact: lferro@mitre.org

MITRE
Washington C3 Center
McLean, Virginia

Table of Contents

List of Tables

1 Introduction

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 exact 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. We achieve this by utilizing (and slightly extending) the International Standard from the International Organization for Standardization (ISO), *ISO 8601: Information Interchange – Representation of Dates and Times*, Second Edition (1997), which can be obtained from ftp://ftp.qsl.net/pub/g1smd/8601v03.pdf. Before reading much farther, and certainly before embarking on any annotation, you should thoroughly familiarize yourself with the ISO document.

1.1 How This Manual is Organized

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.

2 Determining What Kinds of Expressions to Annotate

As indicated 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).

2.1 Markable Expressions

Markable expressions are the expressions that should be annotated. To be markable, the syntactic head of the expression must contain an appropriate ***lexical trigger***. Each lexical trigger is a word or numeric expression whose meaning conveys a temporal unit or concept. The following list of lexical triggers is partially representative of the variety of terms that are currently being treated as triggers. The table also includes examples of closely related temporal concepts that are not considered triggers.

Table 2-1 Lexical Triggers

Part of Speech	Lexical Triggers	Non-Triggers
Noun	day, minute, weekend, midnight, millenium, era, semester, summer, [the] future, [the] past, months	instant, jiffy
Proper name	(unique identifier for temporally-defined event): Monday, January, New Year's Eve, Washington's Birthday	
Specialized time pattern:	8:00, 12/2/00, 1994, 1960s	
Adjective	current, future, former, past, next, medieval, monthly	earlier, ahead, subsequent, frequent, later
Adverb	currently, lately, then, next, hourly, daily, monthly	immediately, instantly, forthwith, previously, beforehand, soon, sooner, shortly, eventually, once
Time noun/adverb	today, yesterday, now	
Number	3 (as in "He arrived at 3."), three	

These terms are triggers only when used in a temporal sense; e.g., "next" is not a trigger when used in a spatial sense ("the next person in line") instead of in a temporal sense ("the next day").

Pronouns (such as "that") and other expressions that corefer with a trigger-containing expression are also markable. See the section "Coreferring Temporal Expressions" on page 30.

Note that subordinating conjunctions such as *when* are not triggers ("Robins fly south when it gets cold."). Also note that the lists of adverb and adjective triggers are fairly conservative and do not include temporally related terms such as *eventually*, *frequent*, and *frequently*. Prepositions are never markable, including "before" and "after."

Note that some of the terms deemed non-triggers, such as "earlier," are permitted within the scope of a markable expression as a modifier, as in "earlier years"; the non-triggers are not markable on their own (in non-head positions), as in "an earlier argument."

Some expressions are markable but are not able to be normalized; that is, the expression does not refer to any particular point, set of points, or durations on a time line. These include idioms, non-specific holidays, and event-anchored expressions.

You should consider examples found elsewhere in the body of this document in order to better understand the breadth of the class of triggers that is intended.

2.2 Non-Markables

Below is a sampling of "negative cases;" i.e., examples of non-markable, but time-related expressions. Currently, the set of examples includes a number of single-word adverbs and adjectives, plus the short phrase "too often." This list is offered to help you in determining the boundary between markable and non-markable expressions.

2.2.1 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.

> The world watched the U.S.-led bombing of Iraq in 1991 and the *subsequent* freeing of Kuwait from Iraqi occupation.
>
> The truth will come out, *eventually*.
>
> Iraqi leader Saddam Hussein, *meanwhile*, sent a message to Russian President Boris Yeltsin.
>
> Across Europe, banks are converting their systems to offer euro-denominated bank accounts *immediately* for any customers who want them.
>
> But even on Thursday, there were signs of potential battles *ahead*.

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

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

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

2.2.2 Non-Markable Frequency Expressions

The following expressions reference reoccurring times. While some of these terms, such as "usually" and "always," indicate the presence of a markable set-denoting expression, they are not markable in themselves. (See also, "Set-Denoting Temporal Expressions" on page 21.)

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.

2.2.3 Non-Markable Proper Names

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

We infiltrated the terrorist group *Black September*.

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

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

3 Capturing the Meaning of Temporal Expressions

Now that we've defined what is markable, the next step is to show how the expressions should be marked. The objective of the annotation is to capture the meaning of the expressions in a consistent manner. 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. Some expressions are "fuzzy," where you have a general sense of the meaning, but cannot confidently posit a precise value. Later in this section we explain how to annotate these precise and fuzzy expressions, focusing initially on ***points*** of time and ***durations*** of time. Following that, there are subsections discussing modified expressions ("almost two o'clock") and ***frequencies***, which are expressions denoting sets of recurring times ("every Tuesday"). The penultimate subsection discusses non-specific expressions like "a Tuesday." The final sections illustrate the annotation of ***holidays*** and ***coreferring*** temporal expressions, including pronouns.

3.1 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 backslash). For example:

<TIMEX2>Halloween</TIMEX2>

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

Table 3-1 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"
SET	Identifies expressions denoting sets of times.	SET="YES"
PERIODICITY	Captures the period between regularly recurring times.	PERIODICITY="P1M"
GRANULARITY	Captures the unit of time denoted by each set member in a set of times.	GRANULARITY="G3D"
NON_SPECIFIC	Identifies non-specific expressions.	NON_SPECIFIC="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="EARLY">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.

3.2 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.

3.2.1 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 Decembe*r.
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 temporal 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>.

3.2.1.1 Decades, Centuries, and Millenia: Points not Accommodated by the ISO-Standard

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 the "millennium" component, the second one the "century" component, the third the "decade" component, and the fourth the "year" component. Thus, in our version, a well-formed value could consist of just one number (for example, "3" could represent "the next millennium"), two numbers (for example, "20" could represent "the 21st century"), or three (for example, "196" could 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>.

3.2.2 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="1999-07-15-T14:50">ten minutes to 3</TIMEX2>.

2/27/1998 08:14:00
<TIMEX2 VAL="1998-02-27-T08:14:00">2/27/1998 08:14:00</TIMEX2>
I returned to work at twelve o'clock January 3, 1984.[1]
I returned to work at <TIMEX2 VAL="1984-01-03-T12:00">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.[2]
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*.[3]
On <TIMEX2 VAL="1999-07-19">the nineteenth</TIMEX2> I am in class until <TIMEX2 VAL="1999-07-19T11">eleven in the morning</TIMEX2>.

[1] Assume for this example that we know that the time refers to noon.
[2] Assume for this example that we know we're talking about 1 a.m., but we don't know the date.
[3] This example is taken from a dialogue transcript.

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>.

3.2.3 Units of Weeks

The ISO standard also allows for specific weeks to be referenced and normalized:

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

3.2.4 Duration

An expression of duration describes an interval or period of time, indicating explicitly **how long** it 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

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.)

Here are some sample duration expressions:

The video is only *half an hour long*.[4]
The video is only <TIMEX2 VAL="PT30M">half an hour long</TIMEX2>.

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

The gestation period in humans is *nine months*.
The gestation period in humans is <TIMEX2 VAL="P9M">nine months</TIMEX2>.[5]

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

[4] “Only” is treated as modifying “is” rather than as modifying the time expression.
[5] “the gestation period in humans” is not a markable expression.

He will be in school for *another year*.[6]
He will be in school for <TIMEX2 VAL="P1Y">another year</TIMEX2>.

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

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

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

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

3.2.4.1 Decades, Centuries, and Millenia: Periods not Accommodated by the ISO-Standard

Whereas the ISO standard uses YMD to indicate "year, month, day" (respectively), it does not supply the means to capture some other common temporal periods. We have therefore added the following unit abbreviations that can be used in period expressions. (These units can also be used in GRANULARITY and PERIODICITY values, discussed in "Set Denoting Expressions").

Decade: E
Century: C
Millennium: L

For example:

The recession lasted *a decade*.
The recession lasted <TIMEX2 VAL="P1E">a decade</TIMEX2>.

Peace reigned for *two millennia*.
Peace reigned for <TIMEX2 VAL="P2L">two millennia</TIMEX2>.

3.2.5 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:

[6] "Another" is regularized as "1."

Point in Time:	He was happy *five days ago*. He was depressed <TIMEX2 VAL="1999-07-10">five days ago</TIMEX2>.
Duration:	He was happy for *five days*. He was depressed for <TIMEX2 VAL="P5D">five days</TIMEX2>.

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 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*."[7]

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.

3.3 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.

3.3.1 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.

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

"Expression Granularity" Rule:

Use <u>only</u> the trigger word itself to determine the precision of the VAL.

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

[7] The annotation of set-denoting expressions, and this particular example, are illustrated further in "Example Annotations of Sets of Irregularly Recurring Times" on page 24.

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

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

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

3.3.2 Past, Present, Future

Many temporal expressions only vaguely refer 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 value 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 3-2 Present, Past, Future Tokens

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

The mapping between expression and token is based on the context of the time of the document rather than the context of when you are reading it. 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 potentially in our past.

When these tokens are used, they are the only element in the VAL attribute. Here are some examples using these tokens:

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

We can trust *today*’s youth to do the right thing about the environment.
We can trust <TIMEX2 VAL="PRESENT_REF">today</TIMEX2>’s youth to do the right thing about the environment.

Binge drinking is the *current* plague of college campuses.
Binge drinking is the <TIMEX2 VAL="PRESENT_REF">current</TIMEX2> plague of college campuses.

Europe will be stronger than in *the past*.
Europe will be stronger than in <TIMEX2 VAL="PAST_REF">the past</TIMEX2>.

The prospects for resuscitating the peace process have appeared grim *lately*.
The prospects for resuscitating the peace process have appeared grim <TIMEX2 VAL="PAST_REF">lately</TIMEX2>.

We may be moving back to the *medieval* notion of romance.
We may be moving back to the <TIMEX2 VAL="PAST_REF">medieval</TIMEX2> notion of romance.

His partner is a 22-year-old *future* accountant.
His partner is a 22-year-old <TIMEX2 VAL="FUTURE_REF">future</TIMEX2> accountant.

Duisenberg is a *former* Dutch central banker.
Duisenberg is a <TIMEX2 VAL="PAST_REF">former</TIMEX2> Dutch central banker.

We need to discuss *the future* of our peoples.
We need to discuss <TIMEX2 VAL=”FUTURE_REF”>the future of our peoples</TIMEX2>.

We are expecting a reply *two days from now*.
We are expecting a reply <TIMEX2 VAL="1999-07-17">two days from <TIMEX2 VAL="PRESENT_REF">now</TIMEX2></TIMEX2>.[8]

3.3.3 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, ***tokens*** rather than precise numerical values are used in the ISO month position of VAL. Note that we also use these season tokens for academic terms.

Table 3-3 Season Tokens

Token	Position	Expressions
SP	month	spring
SU	month	summer
FA	month	fall autumn fall semester fall term
WI	month	winter

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>.

[8] Although "now" is only given a PRESENT_REF value, the larger anchored expression is specified to the granularity of the day. Each tag's VAL is determined independently of the VAL of any embedding or embedded tag.

3.3.4 Yearly Quarters and Halves

Tokens can also be used for expressions denoting quarters or halves of years. This obviates the need for the annotator to determine, for example, if a calendar or fiscal year is intended. The token goes in the ISO month position of VAL.

Table 3-4 Yearly Quarter & Halves Tokens

Token	Position	Expressions
Q1	month	1st quarter
Q2	month	2nd quarter
Q3	month	3rd quarter
Q4	month	4th quarter
H1	month	1st half (of year)
H2	month	2nd half (of year)

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>.

3.3.5 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." The WE token is only used in the week-based ISO format.

Table 3-5 Weekend Token

Token	Position	Expression
WE	day	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.

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

3.3.6 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:

Table 3-6 Part-of-Day Tokens

Token	Position	Expressions
MO	hour	morning
MI	hour	mid-day
AF	hour	afternoon
EV	hour	evening
NI	hour	night
DT	hour	morning + afternoon (basically, daytime or working hours)

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

He explained that he was flying off to Bosnia in *the morning* and wanted to go out with a bang.[9]
He explained that he was flying off to Bosnia in <TIMEX2 VAL="1999-01-01-TMO">the morning</TIMEX2> and wanted to go out with a bang.

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

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

A different kind of vagueness arises with the expression "*last night*." This could mean the night of the previous day, or the night portion of the current 24 hour period, or both. To prevent the annotator from having to discern the writer's intention, we invoke the following rule for all cases:

***"Last Night* is Yesterday" Rule**

***"last night"* always gets the date of yesterday**

[9] This example appeared in a narrative about New Year's Eve 1998.

For example:

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

3.3.7 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*.[10]
The building was erected in <TIMEX2 VAL="XX63">'63</TIMEX>.

Sunday's paper was thrown out.[11]
<TIMEX2 VAL="XXXX-WXX-7">Sunday</TIMEX>'s paper was thrown out</TIMEX>.

The X placeholder is also tends to be useful in temporal expressions denoting sets of times (see"Set-Denoting Temporal Expressions" on page21) and non-specific expressions (see "Non-Specific Temporal Expressions" on page 25).

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

3.3.8 Unspecified Plural Numbers in Durations

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.

After *months of renewed hostility*, the fighting has suddenly ceased.
After <TIMEX2 VAL="PXM">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="PXE">recent decades</TIMEX2>.[12]

[10] 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."
[11] Assume for this example that you cannot tell which Sunday is intended, but the context makes it clear that it is a specific Sunday.

For *millenniums*, dancing was a form of social glue.
For <TIMEX2 VAL="PXL">millenniums</TIMEX2>, dancing was a form of social glue.[13]

She has been going to the gym to play basketball for *the past few years*.
She has been going to the gym to play basketball for <TIMEX2 VAL="PXY">the past few years</TIMEX2>.[14]

3.4 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 32). 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.

To capture the semantics of modifiers, we use the MOD attribute. 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.

[12] Recall that "E" stands for "Decade" in our extension to the ISO standard. Also, the representation does not capture meaning of "recent" in any way. Also, this expression is marked as a period rather than a point because the context indicates that the phrase "in recent decades" is intended to mean "during (over the course of) recent decades."
[13] Recall that "L" stands for "Millennium" in our extension to the ISO standard.
[14] The representation does not capture any meaning for "past few" (nor for its near-synonym, "recent").

Table 3-7 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
	APPROX	about (e.g., “about three years ago”) around

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

The teacher has *nearly four decades of experience*.
The teacher has <TIMEX2 VAL="P4E" MOD="LESS_THAN">nearly four decades of experience</TIMEX2>.

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

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

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

Britain is staying outside the currency union for *at least the next year* or *two*.
Britain is staying outside the currency union for <TIMEX2 VAL="P1Y" MOD="EQUAL_OR_MORE">at least the next year</TIMEX2> or <TIMEX2 VAL="P2Y" MOD="EQUAL_OR_MORE">two</TIMEX2>.[15]

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>.

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

3.5 Set-Denoting Temporal Expressions

In this section we discuss temporal expressions that indicate sets of times, such as "every hour," "Octobers" or "numerous Saturdays." Such expressions are typically plural or contain a quantifier like "each" or "every." Whereas points of time tell when something happen, and durations tell how long something lasted, set-denoting expressions tell how often something happened.

With sets of times, there are three additional pieces of information the annotation seeks to encode (other than VAL):

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

Set: The fact that the expression is a set-denoting one.

Granularity: The unit of time denoted by each set member (e.g., day, week, month).

Periodicity: The period of time between each set member, if the period is regular.

For example, the plural expression "Fridays" denotes a set of individual Fridays. Also, Fridays are days (not hours or months), and they reoccur once a week. Thus, the expression "Fridays" would be annotated as a set with a granularity of "day" and a periodicity of "week."

To encode this information, we use three attributes in the TIMEX2 tag:

SET: The set attribute has only one possible value, which is "YES." If the expression does not denote a set, the attribute is omitted from the annotation altogether.

GRANULARITY: The GRANULARITY attribute follows the form of the ISO period format. Thus, where ISO indicates a period of one day as "P1D," we represent a granularity of one day as "G1D." Note that in addition to the ISO YMD and HMS characters, the third position can contain our extensions to the ISO: E (for decade), C (for century) and L (for millennium), and any of our **tokens**, such as NI (for night).

PERIODICITY: The PERIODICITY attribute uses the same format as the GRANULARITY attribute. However, since "P" is already used by the ISO period standard, the leading character is F (for Frequency). Thus, a periodicity of one week is encoded as "F1W."

3.5.1 VAL in Set-Denoting Expressions

The VAL attribute in set-denoting expressions should encode whatever time information is indicated in the tagged words. VAL is typically underspecified in set-denoting expressions. Consider "Fridays" again. From just this word, we can't tell which particular Fridays are intended, and even in an expression like "Fridays in 1999," we don't want to require that the annotators specify the exact VAL for each of the 52 Fridays in that year. Thus, the VAL in set-denoting expressions often makes use of the X placeholder introduced earlier ("Unspecified Components of Calendar Dates and Times of Day"). The plain expression "Fridays" would receive a VAL="XXXX-WXX-5," which is interpreted as "day 5 (Friday) of any week of any year." The SET="YES" attribute then ensures that the VAL is interpreted as a set of such things, and not a single instance. The word "Fridays" in the more specified context "Fridays in 1999" would receive a VAL="1999-WXX-5," which is interpreted as "day 5 (Friday) of any week of the year 1999." Again, the SET="YES" attribute then guarantees that the VAL is interpreted as a set of Fridays, and not just a single instance.

3.5.2 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. All set-denoting expressions will be annotated with a SET and a GRANULARITY attribute. Sets of regularly recurring times like those shown here will also have the PERIODICITY attribute:

They watched Millionaire on TV *every Tuesday in 1999*.
They watched Millionaire on TV <TIMEX2 VAL="1999-WXX-2" SET="YES" GRANULARITY=”G1D” PERIODICITY="F1W" >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.[16]
There were, as there have been <TIMEX2 VAL="XXXX-12-31" SET="YES" GRANULARITY=”G1D” PERIODICITY="F1Y" >every December 31</TIMEX> for <TIMEX2 VAL="PXE">decades</TIMEX2>, thousands of people in Times Square.

We spend the first three days of every month writing status reports.
We spend <TIMEX2 VAL="P3D" SET="YES" GRANULARITY=”G3D” PERIODICITY="F1M" >the first three days of <TIMEX2 SET="YES" GRANULARITY=”G1M” PERIODICITY="F1M">every month</TIMEX2></TIMEX2> writing status reports.[17]

On *Friday nights*, the ballroom dance club was overwhelmed.[18]
On <TIMEX2 VAL="1998-WXX-5-TNI” SET="YES" GRANULARITY=”G1NI” PERIODICITY="F1W">Friday nights</TIMEX2>, the ballroom dance club was overwhelmed.

Note that in the following two examples, the VAL shows a larger granularity than the expression itself; e.g., “each week” is only has the year portion of VAL specified. The alternative would be to use the X placeholder in the right-hand portions of VAL (e.g., “1997-WXX” for “each week.”). However, we only allow the X placeholder to be used in this way in cases where the speaker had a single specific time (week, day, etc.) in mind (see page 27).

Two years ago, the dance club drew about 100 students *each week*.[19]
<TIMEX2 VAL="1997">Two years ago</TIMEX2>, the dance club drew about 100 students <TIMEX2 VAL="1997" SET="YES" GRANULARITY=”G1W” PERIODICITY="F1W" >each week</TIMEX2>.

[16] The phrase “for decades” is not considered a modifier of the preceding time expression, so there are two independent TIMEX2 annotations.
[17] Note that the GRANULARITY here is “G3D,” which corresponds to the VAL.
[18] Assume for this example that 1998 is implied.
[19] “Each” is represented the same as “every” (“each and every”).

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

Note that the following two examples have no VAL attribute, as there is no straightforward method of encoding these meanings in VAL.

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

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

3.5.3 Example Annotations of Sets of Irregularly Recurring Times

Unlike regularly recurring times, sets of *irregularly* recurring times, such as those illustrated below, will not contain the PERIODICITY attribute.

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

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

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=”1999-07-20” SET=”YES” GRANULARITY=”G1H”>the hours that you are not</TIMEX2>

Note that two set-denoting expressions in the following examples have no VAL attribute, as there is no straightforward method of encoding these meanings in VAL.

It was easier for the average reveler to see and hear what was going on this year than in *recent years*.[20]

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

It was easier for the average reveler to see and hear what was going on <TIMEX2 VAL="1999">this year</TIMEX2> than in <TIMEX2 SET="YES" GRANULARITY="G1Y">recent years</TIMEX2>.[21]

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

3.6 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 several distinct categories:

- Generic: ***Lexical triggers*** used generically, as in "I love *December*." Such expressions do not designate a specific time, stating instead a generalization or regularity associated with a class or property.
- 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.)
- Other non-specific, non-generic uses of trigger terms, as in "I always vote on *Election Day*."

These very subtle distinctions among non-specifics may be hard to distinguish. Therefore, we require only that non-specifics be identified by adding a NON_SPECIFIC="YES" attribute to the TIMEX2 tag. If the expression is specific rather than non-specific, the attribute is omitted from the annotation altogether, as in all the previous examples in this document.

3.6.1 VAL in Non-Specific Expressions

The value of VAL is necessarily underspecified in non-specific expressions. Often, no VAL attribute is needed. In other cases, VAL contains the X placeholder.

3.6.1.1 No VAL

In the case of non-referential, non-generic usage there is no need for any VAL attribute. Note that in the first two examples below, the indefinite expressions are not treated as referential; e.g., "a sunny day" does not refer to, and is not given the same value as, "today."

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

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

[21] Note that there is no VAL for "recent years", since we do not allow "recent" to be normalized as PAST_REF.

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

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.[22]
We will give you as many answers as we can, as soon as we can, at <TIMEX2 NON_SPECIFIC="YES">the appropriate time</TIMEX2>, consistent with our obligation to also cooperate with the investigations.

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

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

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

A day where the investigation of the new allegations against President Clinton have taken some very interesting and startling turns.[23]
<TIMEX2 NON_SPECIFIC="YES">A day where the investigation of the new allegations against President Clinton have taken some very interesting and startling turns</TIMEX2>.

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," and "midnight oil." The temporal expressions in such idioms are markable but do not receive a value and are designated NON-SPECIFIC.

Beer-drinking will indeed be the order of *the day*, assuming the contract is signed.
Beer-drinking will indeed be the order of <TIMEX2 NON_SPECIFIC="YES">the day</TIMEX>, assuming the contract is signed.

Saddam will not cave in at *the last minute*.
Saddam will not cave in at <TIMEX2 NON_SPECIFIC="YES">the last minute</TIMEX2>.

Note that even set-denoting expressions can be non-specific:

I always vote on *Election Day*.

[22] This expression is taken to be essentially synonymous with "*some* appropriate time", i.e., it's not really definite.
[23] This is the first sentence of a news story.

I always vote on <TIMEX2 SET=”YES” GRANULARITY="G1D" NON_SPECIFIC=”YES”>Election Day</TIMEX2>.[24]

Meanwhile, *each passing day* brings new evidence of collusion.
Meanwhile, <TIMEX2 SET=”YES” GRANULARITY=”G1D” PERIODICITY=”F1D” NON_SPECIFIC=”YES”>each passing day</TIMEX2> brings new evidence of collusion.

3.6.1.2 VALs Containing the X Placeholder

Non-specific expressions make frequent use of the X placeholder. For example, the ISO represents “December” as month 12 of a year. Thus, with the use of the X placeholder we can represent a generic usage of “December” as “XXXX-12.”

They report the traffic conditions *15 minutes after the hour*.
They report the traffic conditions <TIMEX2 VAL=”TXX:15” NON_SPECIFIC=”YES”>15 minutes after the hour</TIMEX2>.

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

Christmas is celebrated in *December*.
<TIMEX2 VAL="XXXX-12" NON_SPECIFIC=”YES”>Christmas</TIMEX2> is celebrated in <TIMEX2 VAL="XXXX-12">December</TIMEX2>.[25]

Note that even set-denoting expressions can be non-specific and use the X placeholder:

We always watch the game on *Super Bowl Sunday*.
We always watch the game on <TIMEX2 VAL="XXXX-WXX-7" SET=”YES” GRANULARITY="G1D" NON_SPECIFIC=”YES”>Super Bowl Sunday</TIMEX2>.[26]

April is usually wet.
<TIMEX2 VAL="XXXX-04" SET=”YES” NON_SPECIFIC=”YES”>April</TIMEX2> is usually wet.[27]

The following sentence illustrates an indefinite expression used referentially:

The election took place on *a Tuesday*.[28]
The election took place on <TIMEX2 VAL="1998-WXX-2" NON_SPECIFIC="YES">a Tuesday</TIMEX2>.

In the case of reduced precision, where the unfilled position(s) are on the **right-hand** end of the value, no placeholder is used except in special cases. The placeholder character is only used when the temporal expression explicitly identifies the intended degree of precision and, at the

[24] The set is indicated by “always.”
[25] Christmas is pinned down to December because of the sentence context, not because of the annotator’s world knowledge.
[26] The set is indicated by “always.”
[27] The set is indicated by “usually.”
[28] Assume for this example that 1998 is implied.

same time, does *not* identify the value for the component that is at that degree of precision. For example, the sentence, "He left on a sunny day in June," explicitly identifies *day* as the intended degree of precision, and it indicates that the speaker has a particular day in mind, but does not identify *which* day the speaker has in mind. The representation of 'a sunny day in June' would therefore be 1999-06-XX, plus an attribute to indicate the non-specific nature of the expression.

He left on a sunny day in June.[29]
He left on <TIMEX2 VAL="1999-06-XX" NON_SPECIFIC="YES" GRANULARITY="G1D">a sunny day in <TIMEX2 VAL="1999-06">June</TIMEX2></TIMEX2>.

3.6.1.3 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.

The clock strikes 12 at *noon* and *midnight*.
The clock strikes 12 at <TIMEX2 VAL="T12:00" NON_SPECIFIC="YES">noon</TIMEX2> and <TIMEX2 VAL="T24:00">midnight</TIMEX2>.[30]

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

3.6.1.4 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-5-TNI (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-5-TNI" MOD="EARLY" NON_SPECIFIC="YES">early one Friday night in <TIMEX2 VAL="1998-FA">fall 1998</TIMEX2></TIMEX2>.

3.7 Event-Anchored Temporal Expressions

An event-anchored time expression is anchored on a prior event for its value to be fully resolved, e.g., "*the day after our meeting*." Such expressions are markable but receive no VAL, even in those cases where the annotator knows from world knowledge or can deduce from the context the time of the event.

[29] Assume for this example that 1999 is implied.
[30] Granularity of VAL is indicated by the terms "noon" and "midnight."

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

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

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

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

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

He returned some gifts *five days after Christmas*.
He returned some gifts <TIMEX2>five days after <TIMEX2>Christmas</TIMEX2></TIMEX2>.[31]

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

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

3.8 Holidays

We globally refer to names of festivals, holidays and other occasions of religious observance, remembrance of famous massacres, etc. as “holidays.” Some of these expressions, like “Shrove Tuesday” and “Thanksgiving Day,” contain trigger words. Others, like “Thanksgiving,” “Christmas,” and “Diwali,” do not. A 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. We have seen several examples of holidays already throughout this document. Some are single non-specific expressions:

Bars were packed with people intent on taking part in the ritual of the *New Year’s Day* hangover.
Bars were packed with people intent on taking part in the ritual of the <TIMEX2 NON_SPECIFIC=”YES”>New Year’s Day</TIMEX2> hangover.

Some 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” GRANULARITY="G1D" NON_SPECIFIC=”YES”>Super Bowl Sunday</TIMEX2>.

Holidays can also be definite and referential, but should not receive a VAL:

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

3.9 Coreferring Temporal Expressions

In this section we simply show how expressions can corefer in various ways. In most cases, the coreference relation is not one of complete identity of reference.

I’m a creature of the *1960s*, *the days of free love*.[32]

[31] Although “Christmas” is markable, it doesn’t provide an anchoring date value because it does not have a VAL itself.

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

The contractor submitted a proposal on *Tuesday*. *The day after that*, the contract was awarded. *That night*, they had a big party.[33]
The contractor submitted a proposal on <TIMEX2 VAL="1999-07-13">Tuesday</TIMEX2>. <TIMEX2 VAL="1999-07-14">The day after <TIMEX2 VAL="1999-07-13">that</TIMEX2> </TIMEX2>, the contract was awarded. <TIMEX2 VAL="1999-07-14-TNI">That night</TIMEX2>, they had a big party.

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

At *11:59 p.m.*, Mayor Rudolph W. Giuliani sat on a stage at 45th Street and Broadway and pushed a button. *Sixty seconds later*, the ball atop 1 Times Square completed its slow descent into retirement.[34]
At <TIMEX2 VAL="1998-12-31-T23:59">11:59 p.m.</TIMEX2>, Mayor Rudolph W. Giuliani sat on a stage at 45th Street and Broadway and pushed a button. <TIMEX2 VAL="1998-12-31-T24:00:00">Sixty seconds later</TIMEX2>, the ball atop 1 Times Square completed its slow descent into retirement.[35]

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

---

[32] This is a case of apposition (syntactically-determined coreference).
[33] "that [day]" is coreferential with "Tuesday". The demonstrative "That" in "That night" is referring to "The day after that", i.e., to Wednesday.
[34] Assume for this example that the most recent New Year's Eve is implied.
[35] The granularity of the first expression is in minutes, while the granularity of the second expression is in seconds.

4 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.

4.1 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., the expression cannot start with a preposition) or a clause of any type (for example, 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

4.2 Syntactic Criteria

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

that cold day
the next day

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

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

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

Note that the inclusion of pre- and post modifiers applies only to the modifiers of the time expression. The time expression may be included within a larger phrase whose head is not part of the time expression; this head and the non-temporal modifiers of this head are not included within the scope of the TIMEX2 tag. For example, in the following, the string "accountant who holds much promise" is not included in the TIMEX2 tag, and neither is the non-temporal premodifier of that string, namely "a 22-year-old."

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

In this next example, neither the determiner "the" before "8:00" nor the noun phrase "this 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="1999-07-15-T08:00">8:00</TIMEX2> meeting <TIMEX2 VAL="1999-07-15-TMO">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?

4.2.1 Appositives

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

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

4.2.2 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="1999-07-15-T15">3</TIMEX2>-<TIMEX2 VAL="1999-07-15-T18">6 pm today</TIMEX2>.

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

4.2.3 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">now</TIMEX2> and <TIMEX2 VAL="1999-07-19-TMO">Monday morning</TIMEX2>.

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

Britain is staying outside the currency union for *at least the next year* or *two*.
Britain is staying outside the currency union for <TIMEX2 VAL="P1Y" MOD="EQUAL_OR_MORE">at least the next year</TIMEX2> or <TIMEX2 VAL="P2Y" MOD="EQUAL_OR_MORE">two</TIMEX2>.[36]

4.2.4 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.

4.2.4.1 When to Create One Tag

An expression is treated as an indivisible syntactic unit if there is no intervening element between temporal terms in the following situations:

(a) Two terms express values for a single unit of time. E.g.,

twelve o'clock midnight (where both twelve o'clock and midnight express values for a ***TOD*** unit.)

(b) Two terms express values for units that are hierarchically related. E.g.,

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

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

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

twelve o'clock midnight
Friday evening
8:00 p.m. Friday
Tuesday the 18th
November 1943
*Fall 199*8

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

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

Prepositions introduce syntactically embedded phrases and therefore typically introduce an embedded TIMEX2 tag. However, there are three notable exceptions in which you should ignore the preposition and create a single tag:

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

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

2. Prepositions like "to, of 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

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

   eleven in the morning

4.2.4.2 When to Create Multiple Tags, with Embedding

There are two 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>.[37]

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.

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

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

4.2.4.3 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 appositives, 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.

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

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

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

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

Note that we do not analyze any of the above expressions as having postmodifiers. The expression “some Thursdays in 1998” does not express a temporal sequence (cf. “two weeks from today,” an example from the previous section). Thus “in 1998” is considered syntactically independent of “some Thursdays.” In other 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

Glossary

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

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*."

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*." Such expressions are markable, but they are not given a time line value.

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 value.

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.

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 phrase as a whole 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.).

Index