
PERSIAN SUPPLEMENT

to the

TIDES Standard for the Annotation of Temporal Expressions

July 2008

Karine Megerdooian

Contact: karine@mitre.org
<http://time2.mitre.org>

Approved for public release; distribution unlimited.

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.
© 2008 The MITRE Corporation. All rights reserved.

MITRE

Table of Contents

1	Introduction	1
2	Determining What Kinds of Expressions to Annotate	2
2.1	Markable Expressions.....	2
2.2	Non-Markables	3
2.2.1	Non-Markable Parts of Speech: Prepositions and Subordinating Conjunctions	3
2.2.2	Non-Markable Point and Duration Expressions	4
2.2.3	Non-Markable Frequency Expressions.....	6
2.2.4	Non-Markable Proper Names	6
3	Capturing the Meaning of Temporal Expressions.....	8
3.1	General Notes on Persian.....	8
3.1.1	Ambiguities in Persian.....	8
3.1.2	Syntax of Noun Phrases	8
3.1.3	Conversational Persian	10
3.2	Precise Temporal Expressions	10
3.2.1	Calendar Dates.....	10
3.2.1.1	Decades, Centuries, Millennia, and BCE.....	12
3.2.2	Times of Day	13
3.2.2.1	Time Zones	15
3.2.3	Units of Weeks	16
3.2.4	Duration.....	16
3.2.5	Choosing Between Point and Duration Format	17
3.3	Fuzzy Temporal Expressions.....	18
3.3.1	Past, Present and Future.....	18
3.3.2	Seasons	19
3.3.3	Weekends	19
3.3.4	Morning, Afternoon, and Night	19
3.4	Modified Temporal Expressions.....	21
3.4.1	Language Specific Differences.....	24
3.4.1.1	ON_OR_BEFORE and ON_OR_AFTER tokens	24
3.4.1.2	Structural Position.....	25
3.4.1.3	Separated modifiers	26
3.4.1.4	The preposition لَ	26
4	Determining the Extent of the Annotations.....	28
4.1	Lexical and Morphological Criteria.....	28
4.1.1	Lexical Criteria	28
4.1.2	Morphological criteria	29
4.1.2.1	Pronouns	30
4.1.2.2	Copula verb.....	31
4.1.2.3	Object and topic marker لَ	31
4.2	Syntactic Criteria	32
4.2.1	Appositives	35
4.2.2	Range Expressions.....	35
4.2.3	Conjoined Expressions	36
4.2.4	Embedded Expressions.....	37
4.2.4.1	When to Create One Tag.....	37
4.2.4.2	When to Create Multiple Tags, with Embedding.....	39
4.2.4.3	When to Create Multiple Tags, without Embedding.....	40

List of Tables

Table 2-1 Sample Lexical Triggers and Non-Triggers	2
Table 2-2 Sample Non-Triggers	4
Table 3-1 Temporal Expressions with the <i>Ezafe</i>	9
Table 3-2 Present, Past, Future Tokens	18
Table 3-3 Modifier Tokens	22

1 Introduction

This document is a supplement to the *TIDES Standard for the Annotation of Temporal Expressions*, and is designed to assist system developers and annotators working with Persian (Farsi) language data.

This supplement consists of three parts:¹

Chapter 2. Determining What Kind of Expressions to Annotate

This corresponds to Chapter 3 of the *TIDES Standard for the Annotation of Temporal Expressions*, providing Persian examples of markable and non-markable expressions.

Chapter 3. Capturing the Meaning of Temporal Expressions

This section corresponds to Chapter 4 of the *TIDES Standard for the Annotation of Temporal Expressions*, discussing language-specific issues. This section follows the order provided in the original document and presents specific examples either where Persian differs from English in expressing time or when it was felt that additional language-specific information could clarify the annotation guidelines. The cases that are similar to English, however, are typically not discussed and one should refer directly to the original document for how to normalize the temporal expressions using the attributes of the <TIME2> tag.

Chapter 4. Determining the Extent of Annotations

This section corresponds to Chapter 5 of the *TIDES Standard for the Annotation of Temporal Expressions*, providing Persian examples for establishing where the temporal expressions begin and end. Language-specific differences are also discussed.

¹ The document includes numbers in both Persian and Latin numerals. In order to be able to see the Persian numbers correctly, please set the Numeral feature to ‘Context’ on MS Word. On MS Office 2007, this is achieved by selecting the Office button, choose Word Options, select Advanced, and under Show document content, set Numeral to ‘Context’ (Numeral is by default set to ‘Arabic’ referring to ‘European-Arabic’).

2 Determining What Kinds of Expressions to Annotate

This annotation standard is focused on 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.

2.1 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 ماه for “month” or روزانه for “daily”.

Table 2-1 Sample Lexical Triggers and Non-Triggers

Part of Speech	Lexical Triggers	Non-Triggers
Noun	دقیقه، ساعت، ثانیه، بعد از ظهر، نصف شب، شب، روز، پریروز، شبانهروز، بامداد، ماه، سال، تابستان، فصل، ترم، دهه، قرن، سده، هزاره، دوره، موقع، زمان، آینده، گذشته	لحظه، دم، مناسبت، فرصت، موقعیت، وهله، برنامه، مورد، دفعه، ابد، عمر، سال نوری، تاریخ (when meaning “history”)
Proper name	دوشنبه، فوریه، آبان، نوروز، عیدفطر، شب یلدا، کریسمس	
Specialized time patterns	8:00، 30/12/2007، 1994 2005	
Adjective	گذشته، جاری، آینده، اخیر، سابق، کنونی، آتی، روزانه، پیش، بیشتر، بعد، دیروزی، امروزی	بعدی، قبلی، دیگر، معاصر، مدرن، هم‌دوره، همزمان، جدید، مکرر، مجدد، پیوسته، زود، متعاقب، ابدی، مدام، همیشگی
Adverb	فعلاً، اخیراً، سالانه، ماهانه، ششماهه، امروزه، هر ساعت، کوتاهمدت، دراز مدت	فورا، آنآ، همزمان، پیشاپیش، باز، بار دیگر، بالاخره، بلافاصله، دوباره، قبلاً، بعداً، احیاناً، معمولاً، مکررانه، همچنان، هنوز، بزودی، مادام العمر، سرانجام، عاقبت، بموقع، بجا، بوقت
Time noun/adverb	الآن، حالا، دیروز، امروز، اینک، زمانه، اکنون، هم اکنون	
Number	2 (مثال: 2 می‌رسیم)، پنجم (مثال: پنجم ژانویه)، شصت (مثال: دهه شصت)	

To be a trigger, the referent must also be able to be oriented on a timeline, or at least oriented with relation to a time (past, present, future). Table 2-1 contains a sampling of lexical triggers. The table also includes examples of closely related temporal concepts that are not considered triggers. Note that the same term can be a lexical trigger in some contexts and not in others since many words have ambiguous meanings. It is therefore vital for the annotator to consider the context of temporal expressions, rather than simply being triggered by the string. For instance, the term تاریخ is used temporally and is a trigger in تاریخ امروز 20 فوریه است “today’s **date** is February 20th”, but it is not a trigger in تاریخ اسلام و ایران “the **history** of Islam and Iran” where it has no temporal meaning. These non-markable terms are discussed further in Section 2.2.

Markable expressions also include pronouns (such as آن) that can co-refer with a markable time expression. See the section “Pronouns and Elided Elements” in the *TIDES Standard for the Annotation of Temporal Expressions*.

2.2 Non-Markables

In contrast to the triggers presented above, the non-triggers (i.e., the non-markables), although they can be temporal in their semantics, are *as a class* less amenable to being pinned down to a timeline. For 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 جوجه‌ها رو آخر پاییز می‌شمارند “don’t count your chickens until they hatch” (lit. “one counts the chickens at the end of fall”) or صد سال به این سالها “many happy returns” (lit. “one hundred years to these years”).

Likewise, note that adjective non-triggers are permitted within the extent of a markable expression, as in در قرون معاصر “in the modern era”, هفته قبلی “the previous week/the week before that” or تا چند روز دیگر “in a few more days”. They are not markable on their own, as in نویسندگان معاصر “the modern writers”. To give another example, the adjective بعدی listed as a non-trigger in Table 2-1 would be included in the markable expression روز بعدی “the next day” but it will not be marked in تصمیم بعدی “the next decision” where the *syntactic head* is تصمیم “decision” and therefore not a lexical trigger for temporal annotation. Hence, معاصر and بعدی are not intrinsically terms that indicate a temporal expression, but whether they are markable or not depends on the context in which they appear (e.g., if the noun that they modify is a lexical trigger). However, starting with the 2005 guidelines the extent may also include non-triggers that are syntactic heads of markable expressions when the semantic head is a trigger word; e.g., partitives like بیشتر هفته “much of the week”.

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

Table 2-1 above shows that some parts of speech contain both triggers and non-triggers. In contrast, subordinating conjunctions which introduce clauses are never triggers; that is, they

never appear as the syntactic head of an annotated expression. The following table shows some examples of this part of speech. In addition, prepositions that introduce noun phrases are not considered triggers and are excluded from the annotated expression. These are exemplified in Table 2-2 as well.

Table 2-2 Sample Non-Triggers

Part of Speech	Non-Triggers
Subordinating Conjunction	وقتی که، زمانی که، در حالیکه، همینکه، تاموقعی که، در ضمن اینکه، هر دفعه(که)، هر وقت که، زودتر از اینکه
Preposition	در، از، تا، طی، در خلال، حین، قبل از، بعد از، پس از، پیش از، در طول

Note that in Persian, conjunctions often contain *که*. In addition, in conversational Persian *که* can be used to mean “when” which is also a conjunction and should not be marked as in ...کوچک که بودم... “when I was young...”.

In Persian, there are two types of prepositions: those that are derived from nouns and participate in the *ezafe* construction² (e.g., *طی این دو ساعت* “during these two hours”), and a limited set of prepositions that do not take the *ezafe* (e.g., *تا سال گذشته* “until last year”). We treat both instances as a preposition and do not include them within the underlined annotated expressions. This can give rise to some ambiguities as most ‘nominal prepositions’ can also be used as simple nouns. Hence in the phrase *در این مدت* “in this period”, *مدت* is used as a temporal noun and is a lexical trigger. However, in the phrase *به مدت دو سال* “for two years” the term *مدت* is part of the compound preposition *به مدت* “for” and is not included in the temporal expression.

In certain instances, the preposition behaves more like a modifier of the temporal expression; in these cases, the preposition should be included within the annotated expression. See Section 3.4 “Modified Temporal Expressions” for examples.

2.2.2 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:

دولت پاکستان قبلاً گفته بود که انتخابات عمومی برگزار نخواهد شد.
The Pakistani government had *previously* said that the general elections will not take place.

معرفی برترین وبلاگهای فارسی همزمان با جشن تولد پرشین بلاگ بود.
The introduction of the best Persian weblogs was *simultaneous with* Persian Blog’s anniversary.

سرانجام وردپرس فارسی بعد از ماهها تلاش تیمی در دسترس عموم قرار گرفت.
After months of teamwork Persian Wordpress has *finally* been released to the public.

² The *ezafe* links elements of the noun phrase and preposition phrase. It is discussed in more detail in Section 3.1.

بالاخره هنر آشکار می کند که چه کسی بوده ایم و حالا به چه کسی تبدیل شده ایم.
Art will *eventually* reveal who we used to be and who we have now become.

این انفجار توجه جمعیت بزرگی از رهگذران را جلب کرد و انفجار بعدی موجب مجروح شدن تعدادی از آنها شد.
This explosion attracted a large crowd and the *subsequent* explosion caused injuries to a number of them.

کسانی که فرزندانشان کوچک هستند نمیدانند که چه آینده ای درپیش دارند.
Those who have young children do not know what future awaits them (lit. what future they have *ahead*).

These expressions are admittedly borderline in terms of markability. With the anchoring attributes ANCHOR_VAL and ANCHOR_DIR, it might well be feasible to capture some of the semantics of these terms. For example, consider the following sentence:

ریچارد هاس امروز در گفت و گویی با هفته نامه نیوزویک گفته است که پاکستان فصلی سخت در پیش دارد.
Richard Haass said *today* in an interview with the Newsweek weekly that *a difficult season* lies *ahead* for Pakistan.

This sentence may be tagged as shown below (where the reference date is the 22nd of October, 2007). For more information on Anchoring, see Section 4.2.4 of the *TIDES Standard for the Annotation of Temporal Expressions*.

ریچارد هاس </TIME2> امروز <TIME2 VAL="2007-10-22"> در گفت و گویی با هفته نامه نیوزویک گفته است
که پاکستان </TIME2> فصلی سخت <TIME2>
</TIME2> درپیش <TIME2 VAL="FUTURE_REF" ANCHOR_DIR="AFTER" ANCHOR_VAL="2007-10-22">
دارد.

There are many ordering/sequencing terms that would need to be more fully explored before they could be added to the list of markables.³ Thus, although such terms are not “officially” markable, we encourage TIME2 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:

این سخن احمدی نژاد بلافاصله تیتیر یک خبرگزاری های غربی شد
Ahmadinejad's comment *immediately* made the major headlines in the West.

همین الان میخواستم پیام بیداتون کنم.
I was *just* about to come look for you. / I was going to come look for you *momentarily*.

Non-quantifiable durations are not markable:

رنال مادیرد موقتاً از رویارویی با دیگر قهرمانان اروپا گریخت.
Real Madrid has *temporarily* managed to avoid facing the other European champions.

این کلید پارامترها را بطور همیشگی و بدون قرار دادن آن در سطل اشغال حذف می کند
This key deletes the parameters *permanently* without placing them in the recycle bin.

³ Consider:

در پیش، بعدی، قبلی، پیشین، پسین، پس آیند، قبلا، بعدا، در ابتدا، پیش از این، بیشتر، سابقا، تاکنون، تا الان، تابحال، از این ببعده، پیشاپیش، بالاخره، سرانجام، اولش، اوایلش، اولیه، هنوز، تا حالا، دیگه هرگز، دیگه هیچوقت.

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

وزارت امور خارجه آمریکا اعلام کرد هنوز هیچ تاریخی برای مذاکره با ایران درباره اوضاع عراق تعیین نشده است.
The U.S. State Department announced that *no date* has yet been set for talks with Iran regarding the situation in Iraq.

“Time” when it means “Situation” or “Occasion” is not markable. In Persian, the meaning of “situation” or “occasion” is often indicated by the words *فرصت*, *دفعه*, *مرتبه* or *بار*. These words should not be marked:

به نظر من تابستان هم فرصت خوبی برای تفریح است و هم فرصت خوبی برای درس خواندن.

In my view, summers are both a good *time* for having fun and good *time* for studying.

بار اولش نیست؛ اما گفتم شاید این دفعه عوض شده باشه.

It's not his first *time*; but I thought maybe he has changed this *time*.

Although *لحظه* is generally translated as “second” or “instant”, it can also mean an “occasion” or “opportunity” as in the following sentence where it is not markable:⁴

هیچ لحظه عظیمی از کف نرفته است چراکه لحظه های بزرگ، هنوز و همیشه درپیش روی ما هستند.

We have not lost a giant *opportunity* as big *opportunities* are still and always ahead of us.

2.2.3 Non-Markable Frequency Expressions

Bare frequencies (frequency expressions with no time period given) are not markable:⁵

چرا اندازه گیریهای مکرر قند خون توسط فرد مبتلا به مرض قند حائز اهمیت است؟

Why are *frequent* measurements of blood sugar by the person who has diabetes so important?

منظورم اینه که کم کم به خاطر شروع ترم جدید دیگه نمیتونم زودزود آپ کنم.

What I mean is that gradually because of the start of the new term I won't be able to upload *frequently* anymore.

بارها گفته بودم که قهرمانی روش نویسنده نیست.

I had said *time and time again* that heroism is not the style of the writer.

منطقه ی غرب معمولاً به محافظه کاران رای می دهد.

The West *usually* votes for the conservatives.

همیشه راه دیگری هست.

There is *always* another way.

Other examples in this class include:

سه دفعه، عموماً، بطور کلی، بطور عادی، بطور معمول، معمولاً، بطور مداوم، بطور نادر، مکرراً، دائماً، هرگز، گاهی، گاهگاه، اغلب.

2.2.4 Non-Markable Proper Names

Proper names that designate something other than a temporal entity but happen to contain lexical triggers are not markable. The following list contains names of organizations, books, festivals,

⁴ Note that in this example هنوز (still) and همیشه (always) are also not markable.

⁵ In contrast, frequencies whose semantics include a temporal unit, such as روزانه or هر روز, are markable. See section 4.5 in the *TIDES Standard for the Annotation of Temporal Expressions*.

albums, and movies (shown underlined). Since these are not temporal entities, they are not markable. Some of these expressions may be ambiguous in Persian since proper names are not capitalized and punctuation such as quotation marks is not generally used; annotators should use the context of the sentence to determine the correct marking.

سازمان فلسطینی سیتامبر سیاه مسئولیت را بر عهده گرفت.

The Palestinian organization *Black September* claimed responsibility.

هزار و نهصد و هشتاد و چهار اثر جورج اورول رو کسی به فارسی نداره؟

Doesn't anyone have George Orwell's work *Nineteen Eighty Four* in Persian?

کیارستمی در جشنواره فیلم فجر حضور نداشت.⁶

Kiarostami did not participate in the *Fajr* Film Festival.

«روز سوم» جدیدترین فیلم سینمایی محمدحسین لطیفی در مقام کارگردان است.

Third Day is Mohamad Hossein Latifi's latest directorial film.

لینکین پارک قرار است در تاریخ 15 آوریل 2007 سومین آلبوم خود به نام دقایقی تا نیمه شب را وارد بازار کنند.

Linkin Park is supposed to issue its third album *Minutes to Midnight* on 15 April 2007.

ایمان بیاوریم به آغاز فصل سرد مجموعه شعری از فروغ فرخزاد و شامل ۷ شعر است.

Let us believe in *the beginning of the cold season* is a collection of poetry from Forugh Farrokhzad and includes 7 poems.

However, triggers that are functioning as temporal modifiers within **titles** (as opposed to proper names) are markable. Examples include titles of conferences and awards:

جایزه پولیتزر سال ۲۰۰۲

The 2002 Pulitzer Prize (lit: The Pulitzer prize of year 2002)

⁶ Note that *fajr* is a temporal expression according to the *TIME2 Persian Calendar Extensions* as it represents the time of the morning prayer between the break of dawn until sunrise.

3 Capturing the Meaning of Temporal Expressions

This section is a supplement to Section 4 in the *TIDES Standard for the Annotation of Temporal Expressions* which illustrates how the semantics of markable expressions are captured in the annotations. The focus of this section is to provide Persian language-specific examples that may raise issues for annotators.

3.1 General Notes on Persian

3.1.1 Ambiguities in Persian

As noted in Section 2.2, many expressions may be markable or not depending on context (e.g., does it refer to a particular time, or is it a non-temporally-anchored reference to an occasion?). It is therefore important to always consider the context of temporal expressions in a text.

For instance, the term همین الآن may mean *momentarily* as we have seen in Section 2.2.2, in which case it should not be annotated.

همین الآن میخوام بیام ببینتون کنم.

I was *just* about to come look for you. / I was going to come look for you *momentarily*

However, it can also mean *right now* or *nowadays* as shown below. In these cases, همین الآن is a temporal expression and should be tagged.

همین الآن چی گوش می‌کنید؟

What are you listening to *right now*?

همین الآن زندانی سیاسی در ایران در زیر انواع فشارها و شکنجه‌های جسمی و روحیست.

Nowadays, the political prisoner is under various types of psychological pressures and tortures in Iran.

3.1.2 Syntax of Noun Phrases

An important distinction in Persian, as compared to English, is the use of the اضافه (*ezafe*) in the formation of the noun phrase. The noun and its modifiers (i.e., adjectives) and possessors (i.e., possessive pronoun or possessor nouns) are linked to each other by using an affix known as the *ezafe*, which is pronounced as /e/ after consonants and /ye/ after vowels. In English, the adjectives simply precede the noun and sometimes ‘of’ is used to express possession, whereas in Persian the *ezafe* is used in both cases, as shown in the contrasting examples below:

گنجشک سیاه کوچک

[sparrow-Ez black-Ez small]

“small black sparrow”

گنجشک سیاه کوچک نیما

[sparrow-Ez black-Ez small-Ez Nima]

“Nima’s small black sparrow / the small black sparrow of Nima”

The ezafe construction plays an important role in the formation of temporal expressions as can be seen in the following parallel Persian and English examples. Each noun phrase in this table is a temporal expression and should be annotated as a unit.

Table 3-1 Temporal Expressions with the Ezafe

Persian Expression and Gloss	English Equivalent
عصر روز چهارشنبه evening-Ez day-Ez wednesday	Wednesday evening
روزهای آینده days-Ez future	the future days/the upcoming days
سال 1342 خورشیدی year-Ez 1342-Ez solar	1342 AP
فصل زمستان امسال season-Ez winter-Ez this.year	this year's winter = the winter of this year
سال های ابتدایی انقلاب years-Ez early-Ez revolution	the early years of the revolution
ساعت 8 شب دوشنبه هفته چهارم این ماه hour-Ez 8-Ez night-Ez monday-Ez week-Ez fourth-Ez this month	8 o'clock on monday night of the fourth week of this month

The difference in the noun phrase constructions in these two languages is crucial especially with respect to Section 4.2.4.1 “When to Create One Tag” (cf. Section 5.2.4.1 of the *TIDES Standard for the Annotation of Temporal Expressions*) as many instances that use a preposition in English are actually linked via the ezafe in Persian and therefore do not raise similar issues:

1. The preposition “of” in temporal expressions in English often corresponds to the ‘ezafe’ in Persian:

دوم سپتامبر
the second of December

تابستان 1964
the summer of 1964

صبح یازدهم سپتامبر
the morning of September 11th

2. Partitives formed in English using the “of” preposition are typically formed with the ‘ezafe’ in Persian:

تمام سال
all of the year

بقیه‌ی این هفته
the rest of this week

طرفهای آخر سال
near the end of the year

اغلب اوقات
most of the time

3. The preposition “in” in TOD expressions in English corresponds to the ‘ezafe’ in Persian:

یازده صبح
eleven in the morning

4. Prepositions used with “early” and “late” expressions in English correspond to the ‘ezafe’ in Persian:

اوایل آن سال
earlier in the year (cf. “earlier that year”)

دیر وقت شب
late at night

All of these noun phrase constructions should receive a single TIME2 tag when annotating the temporal expressions in Persian.

3.1.3 Conversational Persian

There is an instance of diglossia in Persian where both a literary dialect and a spoken, conversational variant coexist. Both may be used in writing depending on the author, topic, and document style or register. Hence the conversational or colloquial variant will be found more commonly in blogs, chats, dialogue, and personal correspondence. The examples provided here are extracted from both variants of the language. Furthermore, if an expression is markable in literary Persian, then its conversational variant and equivalent is also markable, and vice versa.

3.2 Precise Temporal Expressions

3.2.1 Calendar Dates

Persian language documents may use several alternative calendar systems, sometimes within the same text. The attribute ‘CAL’ should be used in the TIME2 tag to identify the calendar system being used. For detailed information, see the *TIME2 Persian Calendar Extensions* and *TIME2 Islamic Calendar Extensions*. The Persian calendar is a solar calendar and is used as the official calendar system in Iran and Afghanistan. In Persian, Anno Persicum or AP is indicated alternatively as ‘hejri-e khorshidi’ (هجری خورشیدی), ‘hejri-e shamsi’ (هجری شمسی), ‘khorshidi’ (خورشیدی) or ‘shamsi’ (شمسی). In Persian, the Gregorian calendar or the Christian year (CE) is

called ‘Miladi’ (میلادی). If the Islamic lunar calendar or AH is used in a document, it is referred to as ‘hejri-e ghamari’ (هجری قمری) or simply as ‘ghamari’ (قمری).⁷

The type of calendar system being referred to may be explicitly stated in the text as in the following examples (assume for all these examples that the reference date is Monday, October 22, 2007 which is equivalent to 30 Mehr 1386 AP):

روز دهم سپتامبر سال جاری میلادی
The tenth of September of the current Christian year
<TIME2 CAL=“ISO_EXT” VAL=“2007-09-10”>روز دهم سپتامبر سال جاری میلادی
قیام بهمن 1357 هجری خورشیدی
The rebellion of Bahman 1357 AP
<TIME2 CAL=“PERSIAN_FA” VAL=“1357-11”>قیام بهمن 1357 هجری خورشیدی
در سال 1372 شمسی
In 1372 AP
<TIME2 CAL=“PERSIAN_FA” VAL=“1372”>در سال 1372 شمسی
۱۴۲۹ هجری قمری اول محرم سال
The first of Moharram of 1429 AH
<TIME2 CAL=“ISLAMIC” VAL=“1429-01-01”>۱۴۲۹ هجری قمری اول محرم سال

In some cases, however, the annotator needs to determine which system is being used based on the bigger context. Hence to identify the correct calendar year for e.g., “the sixteenth of this month”, the annotator would need to first determine within the context of the document which date or calendar system is meant. In general, however, the month mentioned indicates what calendar system is being used: if the month is a member of the Iranian, Afghan or Pashto months then the calendar used is also the solar or AP year; if an Islamic Arabic month is used, then the calendar refers to the lunar or Islamic year; and finally, a European month marks the Gregorian calendar.

هفت ثور 1358
7 Sawr 1358
<TIME2 CAL=“PERSIAN_AR” VAL=“1358-02-07”>هفت ثور 1358
پنجشنبه بیست و سوم آذر 1385
Thursday twenty third of Azar 1385
<TIME2 CAL=“PERSIAN_FA” VAL=“1385-09-23”>پنجشنبه بیست و سوم آذر 1385
8 اکتبر 1492
8 October 1492
<TIME2 CAL=“ISO_EXT” VAL=“1492-10-08”>8 اکتبر 1492
اولین روز ماه می

⁷ The common abbreviations used are م for میلادی, ه ش for شمسی, and ه ق for قمری.

The first day of the month of May

<TIME2 CAL="ISO_EXT" VAL="2007-05-01"> اولین روز ماه می

Some documents may list alternative calendar dates referring to the same day. In these cases, each instance should be tagged separately as shown:

چهارشنبه ۲۹ فروردین ۱۳۸۶ هجری شمسی مطابق با ۱۸ آوریل ۲۰۰۷ میلادی برابر با ۲۹ ربیع الاول ۱۴۲۸ هجری قمری
Wednesday 29 Farvardin 1386 AP corresponding to 18 April 2007 CE equivalent to 29 Rabi-ol-Awwal 1428 AH

<TIME2 CAL="PERSIAN_FA" VAL="1386-01-29"> چهارشنبه ۲۹ فروردین ۱۳۸۶ هجری شمسی

<TIME2> با ۱۸ آوریل ۲۰۰۷ میلادی <TIME2 CAL="ISO_EXT" VAL="2007-04-18"> برابر با

<TIME2> ۲۹ ربیع الاول ۱۴۲۸ هجری قمری <TIME2 CAL="ISLAMIC" VAL="1428-03-29">

If no calendar system has been mentioned in the context of the document, annotate the text according to the default (i.e., Gregorian) system.

زمین‌لرزه روز گذشته در پرو 337 کشته داد.

The previous day's earthquake in Peru left 337 dead.

زمین‌لرزه <TIME2> روز گذشته <TIME2 CAL="ISO_EXT" VAL="2007-10-21"> در پرو 337 کشته داد.

پریروز بمبی قوی در خیابان الرائد بغداد، نزدیک خانه ما، منفجر شد.

The day before yesterday a powerful bomb exploded in the Al-Raed street of Baghdad near our house.

<TIME2> پریروز <TIME2 CAL="ISO_EXT" VAL="2007-10-20"> بمبی قوی در خیابان الرائد بغداد، نزدیک خانه ما، منفجر شد.

3.2.1.1 Decades, Centuries, Millennia, and BCE

As with regular dates, the calendar system used to represent decades (دهه)، centuries (قرن، سده)⁸ and millennia (هزاره) are generally explicitly marked or can be determined by the context of the document.

حزب کمونیست ایران در دهه 60 هجری شمسی تشکیل شد.

The Iranian Communist Party was formed in the 60s AP.

حزب کمونیست ایران در <TIME2> دهه 60 هجری شمسی <TIME2 CAL="PERSIAN_FA" VAL="136"> تشکیل شد.

ناسیونالیسم در سده 20 م، بستر بسیاری از اندیشه‌های خشونت‌پرور بوده است.

Nationalism in the 20th century CE has been the feeding ground for many of the violent ideologies.

ناسیونالیسم در <TIME2> سده 20 م <TIME2 CAL="ISO_EXT" VAL="19">، بستر بسیاری از اندیشه‌های خشونت‌پرور بوده است.

بحران زیست محیطی در هزاره سوم!

An environmental crisis in the third millennium!

بحران زیست محیطی در <TIME2> هزاره سوم <TIME2 CAL="ISO_EXT" VAL="2">!

Similarly, the pre-calendar dates may be explicitly tagged for the calendar system used as in "second millennium BCE" or "the 8th century (BCE)", "سده هشتم (پیش از میلاد)"

⁸ is generally translated as 'centennial' but is equivalent to 'century' in expressing temporal expressions.

هنر قالیبافی ایران به 5000 سال پیش می‌رسد.
The art of carpet weaving in Iran goes back to 5000 years ago.

هنر قالیبافی ایران به <TIME2> 5000 سال پیش </TIME2> <TIME2 CAL="ISO_EXT" VAL="BC2993"> می‌رسد.

To simplify the annotations in the rest of this document, the CAL attribute will not be included in the following examples unless the Persian or Islamic calendar is explicitly used.

3.2.2 Times of Day

In formal contexts, the 24-hour time period may be used to express time in Persian. However, “a.m.” and “p.m.” are most often expressed using words for the part of day such as صبح, سحر, “a.m.” and “p.m.” are most often expressed using words for the part of day such as صبح, سحر, بعدازظهر “afternoon”, عصر “evening”, or شب “night” (for all examples, assume that the reference date is October 22, 2007).

چهارمین نشست ساعت پانزده پنجشنبه 22 آذر ماه ، در سالن آمفی تئاتر کانون برگزار شد.

The fourth session took place at 15:00 on Thursday, the 22 of Azar in the Amphitheatre Hall of the Center.

چهارمین نشست

<TIME2> ساعت پانزده پنجشنبه 22 آذر ماه </TIME2> <TIME2 CAL="PERSIAN_FA" VAL="1386-09-22T15:00"> ، در سالن آمفی تئاتر کانون برگزار شد.

ساعت 8/30 دقیقه صبح است و من دو استکان چای خورده ام و با چند نفری بحث بیکاری کرده ام.

It's 8:30 in the morning and I have had two cups of tea and have discussed unemployment with couple of people.

<TIME2> ساعت 8/30 دقیقه صبح </TIME2 VAL="2007-10-22T08:30"> خورده ام و با چند نفری است و من دو استکان چای بحث بیکاری کرده ام.

روز جمعه آب تهران از ساعت 5 تا 7 عصر در کلیه مناطق قطع خواهد بود.

On Friday the water of Tehran will be interrupted in all areas from 5 to 7 in the evening.

<TIME2> روز جمعه </TIME2 VAL="2007-10-26"> آب تهران از

<TIME2> ساعت 5 تا </TIME2 VAL="2007-10-26T17:00">

<TIME2> 7 عصر </TIME2 VAL="2007-10-26T19:00"> در کلیه مناطق قطع خواهد بود.

جمع آوری زباله های شهری در راس ساعت 9 شب است.⁹

The city's trash collection takes place at exactly 9 o'clock at night.

جمع آوری زباله های شهری در <TIME2> راس ساعت 9 شب </TIME2 VAL="T21:00"> است.

In Persian, the relative placement of the word “night” next to the name of a day is extremely important as it may refer to two distinct days. Hence, جمعه شب (lit. Friday night) refers to “Friday night” while شب جمعه (lit. night of Friday) means “Thursday night”.

⁹ In this sentence, راس (exactly) is the modifier of the temporal expression and is therefore included in the annotation. See Section 3.4 “Modified Temporal Expressions” for discussion.

The words *نیمه‌شب* and to a lesser extent *نصف‌شب* are used to express “midnight” in Persian, but they actually translate into “middle of the night” and are therefore ambiguous. The annotator needs to use the context to determine the time meant by the author. Hence, *midnight* can be expressed as follows:

عزاداری محرم، بعد از ساعت ۱۲ شب ممنوع شد.

The mourning for Muharram is prohibited after *midnight* (lit. after 12 at night).

عزاداری محرم، بعد از <TIME2> ساعت ۱۲ شب <TIME2 VAL=“2007-10-21T24:00”> ممنوع شد.

دوباره زن و شوهر کنار هم نشستند و ساعت تیک‌تاک می‌کرد؛ بالاخره نیمه شب شد.

Husband and wife were once again sitting next to each other and the clock was ticking; it finally turned *midnight* (lit. middle of the night)

دوباره زن و شوهر کنار هم نشستند و ساعت تیک‌تاک می‌کرد؛ بالاخره <TIME2> نیمه شب <TIME2 VAL=“2007-10-21T24:00”> شد.

Nevertheless, both *نیمه‌شب* and *نصف‌شب* are often used to mean some time in the night (usually between 1am and 4am) and it can therefore cover a wide range of time expressions. In most instances, a specific time is used along with these expressions as shown in the examples below:¹⁰

ساعت ۱:۳۰ بعد از نیمه شب شد.

It turned 1:30 in the night.

<TIME2> ساعت ۱:۳۰ بعد از نیمه شب <TIME2 VAL=“2007-10-22T01:30”> شد.

این مسیج ساعت سه نیمه شب به دست ما رسید.

We received this message at three in the night.

این مسیج <TIME2> ساعت سه نیمه شب <TIME2 VAL=“2007-10-22T03:00”> به دست ما رسید.

الآن اینجا ساعت سه و ربع نصف شب است و هنوز خوابم نبرده.

It is now three and a quarter in the night here and I am still not asleep.

<TIME2> الآن <TIME2 VAL=“2007-10-22T03:15”> اینجا

<TIME2> ساعت سه و ربع نصف شب <TIME2 VAL=“2007-10-22T03:15”> است و هنوز خوابم نبرده.

The words, *نیمروز* “midday” and *ظهر* “noon” can both be annotated as 12 o’clock:

بی‌دلیل به ساعتش نگاه می‌انداخت: درست نیمروز بود.

He glanced at his watch for no reason: it was *exactly midday*.

بی‌دلیل به ساعتش نگاه می‌انداخت: درست <TIME2> نیمروز <TIME2 VAL=“2007-10-22T12:00”> بود.

The term *شبانه‌روز* in Persian literally means “nightly day” and refers to the 24-hour period that forms a whole day; it refers to both daytime and nighttime. In the following example, *شبانه‌روز* (although referring to a 24-hour period as can be seen in the English translation) can be expressed as a specific day for the purposes of the TIME2 annotation.

هشت نفر در ناآرامیهای شبانه روز اخیر در پایتخت ارمنستان کشته شدند.

Eight people were killed in the unrest in the Armenian capital *the past 24 hours*.

هشت نفر در ناآرامیهای <TIME2> شبانه روز اخیر <TIME2 VAL=“2007-10-21”> کشته شدند

¹⁰ See Section 3.3.4 on annotating usages of *نیمه‌شب* and *نصف‌شب* that do not seem to refer to a specific time.

3.2.2.1 Time Zones

To express different time zones in Persian, the explicit location is mentioned in the text. Therefore, the relative time of each location with respect to Universal coordinated Time (UTC) should be determined and indicated in the annotation. Hence, a “Z” in the first example indicates that the time is given in UTC or Greenwich Meridian Time, and the “+01” in the second sentence indicates that the local time (in this case, Central European Time) is one hour ahead of UTC.

یک پس لرزه بزرگ با شدت 5.3 در ساعت سه و 6 دقیقه و 32 ثانیه بامداد به وقت گرینویچ در منطقه رخ داد.
A large aftershock with a magnitude of 5.3 occurred at 3 o'clock and 6 minutes and 32 seconds in the morning at Greenwich time.

یک پس لرزه بزرگ با شدت 5.3 در
<TIME2 VAL="2007-10-22T03:06:32Z"> ساعت سه و 6 دقیقه و 32 ثانیه بامداد به وقت گرینویچ در منطقه رخ داد.

شروع رسمی جلسه 7 شب به وقت اروپای مرکزی میباشد.
The official start of the meeting is at 7 pm Central European Time.

شروع رسمی جلسه <TIME2 VAL="2007-10-22T19:00+01"> 7 شب به وقت اروپای مرکزی میباشد.

ساعت تحویل سال 1387 خورشیدی ساعت 9 و 18 دقیقه روز پنجشنبه 1 فروردین 1387 به وقت ایران خواهد بود.
The time for the transition to the new year 1386 AP will be at 9:18 on Thursday, Farvardin 1, 1387 at the time of Iran.

<TIME2 VAL="PERSIAN_FA" VAL="1387"> ساعت تحویل <TIME2 VAL="PERSIAN_FA" VAL="1387"> سال 1387 خورشیدی
<TIME2 VAL="PERSIAN_FA" VAL="1387"> ساعت 9 و 18 دقیقه روز پنجشنبه 1 فروردین 1387 به وقت ایران
<TIME2 CAL="PERSIAN_FA" VAL="1387-01-01T09:18+3:30"> خواهد بود.

مراسم تشییع جنازه شادروان مهستی هنرمند فقید ایرانی روز جمعه 29 ژوئن ساعت 20 به وقت استکهلم (11 به وقت لس آنجلس) در گورستان وست وود مموریال پارک در لس آنجلس صورت می گیرد.

The funeral ceremony of the late Mahasti, the deceased Iranian artist, is taking place on Friday 29 June at 8 pm at the time of Stockholm (11 at the time of Los Angeles) at Westwood Memorial Park cemetery in Los Angeles.

مراسم تشییع جنازه شادروان مهستی هنرمند فقید ایرانی
<TIME2 VAL="2007-06-29T20:00+01"> روز جمعه 29 ژوئن ساعت 20 به وقت استکهلم
<TIME2 VAL="2007-06-29T11:00-08"> 11 به وقت لس آنجلس (<TIME2 VAL="2007-06-29T11:00-08">) در گورستان وست وود مموریال پارک در لس آنجلس صورت می گیرد.

Context may need to be used to determine the local time for annotation purposes as in the following example referring to voting booths in Oman on October 27th, 2007:

حوزه‌های رای‌گیری ساعت هفت به وقت محلی بر روی رای‌دهندگان گشوده شد.
The voting booths were opened to the voters at seven local time.

حوزه‌های رای‌گیری <TIME2 VAL="2007-10-27T07:00+04"> ساعت هفت به وقت محلی بر روی رای‌دهندگان گشوده شد.

Note that Iran generally operates Daylight Savings Time (DST) between 1 Farvardin (March 21) and 1 Mehr (September 23) when the time is 4.5 hours ahead of Greenwich Mean Time (i.e., GMT+4:30), and the rest of the year it's 3.5 hours ahead of Greenwich time (i.e., GMT+3:30). However, similar to the annotation of time zones in Section 4.2.2.1 of the *TIDES Standard for*

the Annotation of Temporal Expressions, if DST is not overtly indicated in the expression, it does not need to be included in the tag.

3.2.3 Units of Weeks

If a Persian or Islamic calendar is used in the document and the text refers to a week-based temporal expression, follow the guidelines described in the *TIME2 Persian Calendar Extensions* and *TIME2 Islamic Calendar Extensions*. As an example consider the temporal expression below:

این مقاله در دوشنبه هفته چهارم اردیبهشت 1386 نوشته شده است.

This article was written on the *Monday of the fourth week of Ordibehesht of 1386*.

این مقاله در `<TIME2>` دوشنبه هفته چهارم اردیبهشت 1386 `<TIME2 CAL="PERSIAN_FA" VAL="1386-W08-3">` نوشته شده است.

As Ordibehesht is the second month of the year in the Persian calendar system, week 4 of Ordibehesht would coincide with week 8.¹¹ Monday is day 3 since day 1 of each week is Saturday.

3.2.4 Duration

An expression of duration indicates a period of time, indicating **how long** something lasts. Durations that refer to specific periods of time can be oriented or anchored with respect to other points or periods of time. The annotation of duration events in Persian is similar to the guidelines described for English, but three specific points can be emphasized:

The term شبانه‌روز in Persian refers to the 24-hour period that forms a whole day; it refers to both daytime and nighttime. When referring to a duration, شبانه‌روز can be tagged as the TIME2 ‘Day’ token, as exemplified below where we translate شبانه‌روز as “a 24-hour period”.

هفت و نیم ساعت بخوابد. يك متخصص بيماريهاي مغز و اعصاب گفت بطور میانگین هر فرد طی شبانه روز باید
A specialist of brain and nerve diseases said that on average each person should sleep *seven and half hours* during a 24-hour period.

يك متخصص بيماريهاي مغز و اعصاب گفت بطور میانگین هر فرد طی
`<TIME2 VAL="P1D" SET="YES">` شبانه روز باید
`<TIME2 VAL="PT7.5H" SET="YES">` هفت و نیم ساعت بخوابد

WITHIN, STARTING and ENDING. As in English, an expression of age is markable when it is an adjective phrase. In Persian the expression would be formed with ساله, which is itself formed on the noun سال “year” as in the examples below.

شاعر 89 ساله پنجشنبه شب درگذشت.

¹¹ Note that week 01 of the year is the first week of Farvardin that contains a Tuesday according to the *TIME2 Persian Calendar Extensions* and given the fact that the year 1386 AP began on a Wednesday (i.e., Farvardin 1 fell on a Wednesday), the first week begins on Saturday, 4th of Farvardin and the fourth week of Farvardin begins on Saturday the 25th.

The 89-year-old poet passed away on Thursday night.

شاعر <TIME2 VAL="P89Y" ANCHOR_DIR="ENDING" ANCHOR_VAL="2007"> 89 ساله </TIME2>
<TIME2 VAL="2007-10-18TNI"> پنجشنبه شب </TIME2> درگذشت.

این نمایشگاه بازگوکننده تاریخ پنجاه ساله فضاوردی جهان بود.
This exposition was representing the *fifty year* history of space travel.

این نمایشگاه بازگوکننده تاریخ
<TIME2 VAL="P50Y" ANCHOR_DIR="ENDING" ANCHOR_VAL="2007"> پنجاه ساله </TIME2>
فضانوردی جهان بود.

Choosing between STARTING and ENDING. The tense used in the sentence can help determine whether an ongoing duration is involved. For instance, in the example below the use of the present tense است سه روز [lit. "three day is"] instead of the past tense بود سه روز [lit. "three day was"] suggests that the strike is still continuing and the annotator should count back three days to enter the ANCHOR_VAL.

زندانیان سه روز است که به اعتصاب غذا دست زده‌اند.
It has been *three days* that the prisoners have been on a hunger strike.

زندانیان
<TIME2 VAL="P3D" ANCHOR_DIR="STARTING" ANCHOR_VAL="2007-10-20"> سه روز </TIME2>
است که به اعتصاب غذا دست داده‌اند.

If the sentence indicates both a start and an end date, the STARTING value is preferred:

روز 5 سپتامبر درست بعد از آن حادثه زندانیان برای سه روز متوالی به اعتصاب غذا دست زدند.
On the 5th of September, right after that event, the prisoners went on a hunger strike for *three consecutive days*.

<TIME2 VAL="2007-09-05"> روز 5 سپتامبر درست بعد از آن حادثه زندانیان برای
<TIME2 VAL="P3D" ANCHOR_DIR="STARTING" ANCHOR_VAL="2007-09-05"> سه روز متوالی
به اعتصاب غذا دست زدند.

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. The following two examples in Persian show the same word بعد used in two different contexts:

Point in Time:

سه روز بعد همدیگر را دیدیم.
Three days later we saw each other.

<TIME2 VAL="2007-10-19"> سه روز بعد همدیگر را دیدیم.

Duration:

بعد از سه روز همدیگر را دیدیم.
After *three days* we saw each other.

بعد از <TIME2 VAL="P3D"> سه روز همدیگر را دیدیم.

In the first example above, بعد means "later" and modifies the temporal expression; in the second example, it is part of the compound preposition بعد از "after" and is not included in the annotation

of the temporal expression. Thus, it is often the context, and not the expression itself, which dictates the format of the annotation.

3.3 Fuzzy Temporal Expressions

3.3.1 Past, Present and Future

Many temporal expressions refer in general terms to the past, the present, or the future. These can be expressed with the tokens listed in the table below, which also includes sample markable and non-markable expressions.

Table 3-2 Present, Past, Future Tokens

Token	Sample Markable Expressions	Sample Non-Markable Expressions
PRESENT_REF	امروز، امروزه (with the “nowadays” meaning) الآن، همین الآن، حالا (the “nowadays” meaning) اینک، اکنون، کنونی، امروزی در این زمان/زمانه، این روزها فعلاً (“for the time being”)	فوراً، آنآ
FUTURE_REF	آینده، روزهای آینده، چند سال/ماه/روز دیگر فرداها، فردا (meaning “in the future”)	بعداً، بالاخره، پس از آن، بزودی، درپیش
PAST_REF	گذشته، اخیر، سابق، دیروز، این اواخر، ماهها/سالها پیش خیلی وقت پیش، چند وقت پیش، کمی پیش، زمانی، به موقعی (“at one time”, “once”)	قبلاً، قبل از این، سابقاً

Several examples are shown below, but in general the annotation guidelines for these tokens follow the guidelines described for English in Section 4.3.2 of the *TIDES Standard for the Annotation of Temporal Expressions*.

مهاجرین دیروز، از مهاجرین امروز حمایت کنید!
Yesterday’s immigrants, support today’s immigrants!

از مهاجرین <TIME2 VAL=“PAST_REF” ANCHOR_DIR=“BEFORE” ANCHOR_VAL=“2007-10-22”>دیروز</TIME2>
مهاجرین
<TIME2 VAL=“PRESENT_REF” ANCHOR_DIR=“AS_OF” ANCHOR_VAL=“2007-10-22”>امروز</TIME2>
حمایت کنید!

The sentence below, repeated from Section 3.1.1, exemplifies the use of همین الآن to mean “nowadays” or “these days”:

همین الآن زندانی سیاسی در ایران در زیرانواع فشارها و شکنجه های جسمی و روحیست.
Nowadays, the political prisoner is under various types of psychological pressures and tortures in Iran.

<TIME2 VAL=“PRESENT_REF” ANCHOR_DIR=“AS_OF” ANCHOR_VAL=“2007-10-22”>همین الآن</TIME2>
زندانی سیاسی در ایران در زیرانواع فشارها و شکنجه های جسمی و روحیست.

The following sentence discusses an interview with Mr. Modarressi carried out on 14 December 2006, and the author claims that Mr. Modarressi has countered his own position just a few days earlier in a written article. Thus 14 December 2006 can be used as the Anchor value in this example:

همین آقای مدرس‌ی در مقاله دیگری در همین رابطه چند روز قبلاً ۱۸۰ درجه عکس این مصاحبه 14 دسامبرش نوشته است.

Mr. Modarressi himself, in a previous article he has written *a few days earlier* on this topic, has taken a 180 degree stance against his *14 December* interview.

همین آقای مدرس‌ی در مقاله دیگری در همین رابطه

<TIME2 VAL="PAST_REF" ANCHOR_DIR="BEFORE" ANCHOR_VAL="2006-12-14">چند روز قبلاً</TIME2> 14 دسامبرش<TIME2 VAL="2006-12-14"> نوشته است.

پروین به توصیه مسئولان باشگاه از بازیکنان پرسپولیس خواست چند روز دیگر صبر کنند و هنوز اعتصاب نکنند. Parvin, following the recommendation of the club staff asked the players of Persepolis to wait *a few more days* and not to go on strike yet.

پروین به توصیه مسئولان باشگاه از بازیکنان پرسپولیس خواست

<TIME2 VAL="FUTURE_REF" ANCHOR_DIR="AFTER" ANCHOR_VAL="2007-10-22">چندروز دیگر</TIME2> صبر کنند و هنوز اعتصاب نکنند.

3.3.2 Seasons

Note that for the Persian calendar, the seasons align perfectly with the year so that no season falls in between two years (as does Winter in the Gregorian calendar).

3.3.3 Weekends

For a description of how to annotate weekends in a Persian calendar system, see the *TIME2 Persian Calendar Extensions*.

3.3.4 Morning, Afternoon, and Night

Periods of the day are often used vaguely in Persian. For instance, *نصف‌شب* and *نیمه‌شب* can be used to mean “midnight” or they can refer to a time at night typically between 1 and 5 a.m. (also see examples in Section 3.2.2). Note that in the sentences in this section, only the English translations of *نیمه‌شب* and *نصف‌شب* are shown in italic, and other temporal expressions are ignored. In the two examples below, *نیمه‌شب* and *نصف‌شب* refer to “midnight”:

یه موقع به ساعت نگاه کردم دیدم از نصف شب گذشته

I suddenly looked at the watch and saw that it was past *midnight*.

قرار بود در نیمه شب آن روز رمز عملیات آژاکس با عنوان «حالا دقیقاً نیمه شب است» به صورت رمز از رادیو بی‌بی‌سی اعلام شود.

The code for Operation Ajax was supposed to be announced as a code from Radio BBC on *midnight* of that day under the name of “the time is exactly *midnight*”.

However, in general نیمه‌شب and نصف‌شب mean any time between midnight and sunrise, so they are best translated into English as “a.m.”:

آی اس پی من نرخش اینه: ساعت 6 تا 1 نیمه‌شب 100 تومن؛ ساعت 1 نیمه‌شب تا 8 صبح 90 تومن.

The rate for my ISP is this: 6 to 1 am 100 tumans; 1 am to 8 am 90 tumans.

[Lit.: The rate for my ISP is this: 6 to 1 middle of the night 100 tumans; 1 middle of the night to 8 of morning 90 tumans]

And in some rare cases, نیمه‌شب and نصف‌شب may even be interpreted as starting after dusk when the dark sets as in the examples below¹². Note that in this case the tokens would span a new day boundary.

نیمه شب شد و وقت نماز شب فرا رسید.

It turned the *middle of the night* and the time of the evening prayer arrived.

نماز شب از نصف شب به بعد آغاز می‌شود و تا طلوع فجر ادامه پیدا می‌کند

The evening prayer starts from *middle of the night* on and continues until the rising of Fajr (aurora).

These terms can therefore be tricky where, depending on the context, they can be used as indicating midnight exactly and at other times, they refer to a range in the night. The tagged examples below show the annotation for each instance.

(1) نیمه‌شب and نصف‌شب meaning midnight

یه موقع به ساعت نگاه کردم دیدم از نصف شب گذشته.

I suddenly looked at the watch and saw that it was past *midnight*.

یه موقع به ساعت نگاه کردم دیدم از <TIME2 VAL="T24:00"> نصف شب </TIME2> گذشته.

قرار بود در نیمه شب آن روز رمز عملیات آژاکس با عنوان «حالا دقیقاً نیمه شب است» به صورت رمز از رادیو بی‌بی‌سی اعلام شود.

The code for Operation Ajax was supposed to be announced as a code from Radio BBC on *midnight of that day* under the name of “*now the time is exactly midnight*”.

قرار بود در <TIME2> نیمه شب آن روز <TIME2 VAL="1953-08-18T24:00"> رمز عملیات آژاکس با عنوان

<TIME2> «حالا <TIME2 VAL="1953-08-18T24:00">

<TIME2> دقیقاً نیمه شب <TIME2 VAL="1953-08-18T24:00"> است» به صورت رمز از رادیو بی‌بی‌سی اعلام شود.

(2) نیمه‌شب and نصف‌شب corresponding to the English “a.m.”

آی اس پی من نرخش اینه: ساعت 6 تا 1 نیمه‌شب 100 تومن؛ ساعت 1 نیمه‌شب تا 8 صبح 90 تومن.

The rate for my ISP is this: 6 to 1 am 100 tumans; 1 am to 8 am 90 tumans.

¹² As the time for the evening prayer actually begins after sunset and lasts until sunrise, “middle of the night” here can refer to a time before midnight, although the general native speaker judgment is to treat these words as referring to a time after midnight.

آی اس پی من نرخش اینه:

<TIME2> ساعت 6<TIME2 VAL="T18:00"> تا </TIME2> 1 نیمه‌شب<TIME2 VAL="T01:00"> 100 تومن؛
<TIME2> ساعت 1 نیمه‌شب<TIME2 VAL="T01:00"> تا </TIME2> 8 صبح<TIME2 VAL="T08:00"> 90 تومن.

(3) نیمه‌شب and نصف‌شب meaning some time in the middle of the night (between midnight and sunrise). Although these terms do not directly correspond to the concept of “night” in English as they generally involve a time after midnight only, we will tag them as TNI (see Section 4.3.7 of the *TIDES Standard for the Annotation of Temporal Expressions* for the annotation of periods of the day).

نیمه شب دیروز در کوچه های مه گرفته به دنبال او می گشتیم.

Yesterday night in the foggy streets, we were searching for him. [Lit. In the *middle of the night of yesterday* ...]

<TIME2> نیمه شب دیروز<TIME2 VAL="2007-10-21TNI"> در کوچه های مه گرفته به دنبال او می گشتیم.

In the following example, the author uses the same term, نصف‌شب, to refer to both *midnight* and *middle of the night* (referring to a time between midnight and sunrise), clearly showing that the annotator needs to take context into account when tagging these items.

شب بر همه خوش، البته نصف شب - الان ساعت 4:22 صبح دیگه از نصف شب هم گذشته.

Good night to all, well *middle of the night* actually – It's now 4:22 in the morning and it's way past *midnight*.

<TIME2> شب<TIME2 VAL="TNI"> بر همه خوش، البته </TIME2> نصف‌شب<TIME2 VAL="TNI">

-<TIME2 VAL="2007-10-22T24:00"> الان </TIME2>

<TIME2> ساعت 4:22 صبح<TIME2 VAL="2007-10-22T24:00"> دیگه از

</TIME2> نصف‌شب<TIME2 VAL="2007-10-21T24:00"> هم گذشته.

3.4 Modified Temporal Expressions

If a temporal expression is quantified or modified in some way, the modifying element should be expressed in the annotation. For example, سال 1386 “1386” is an unmodified expression, but 1386 سال “late 1386” is modified. In general, we want the annotation to capture the basic semantics of quantifier modifiers (e.g., حدود، تقریباً، بیش از) and lexicalized aspect markers (e.g., آغاز، اوایل). Table 3-3 represents some of the modifier tokens in Persian.

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 Section 4 of this document on Determining the Extent of the Annotations). For example, the expression “after two hours” is not considered a modified expression for our purposes because the compound preposition بعد از “after” is not included within the extent of the TIME2 tag. However, the expression “more than two hours” is a modified temporal expression and the adjective + preposition unit بیش از “more than” is included within the extent of the tag. In this case, the MOD attribute is used to capture the semantics of the modifier within the scope of the TIME2 expression. The distinct annotations for these two examples can be seen below:

بعد از دو ساعت

after two hours

<TIME2> بعد از دو ساعت<TIME2 VAL="PT2H">

بیش از دو ساعت

more than two hours

<TIME2 VAL="PT2H" MOD="MORE_THAN">بیش از دو ساعت</TIME2>

Table 3-3 Modifier Tokens

	Token	Sample Expressions
Points	BEFORE	بیش از (بیش از یکصد سال قبل)
	AFTER	کمتر از (کمتر از یک دهه پیش)
Durations	LESS_THAN	کمتر از (کمتر از 14 ساعت) قریب (قریب 28 سال) نزدیک به (نزدیک به هفده سال)
	MORE_THAN	بیشتر از (بیشتر از هشت ماه) بالاخر از (بالاخر از 18 سال)
	EQUAL_OR_LESS_THAN	حداکثر (حداکثر 15 روز کاری)
	EQUAL_OR_MORE_THAN	حداقل (حداقل پنج دقیقه) اقلاً (اقلاً دو ماه)
Points and Durations	START	آغاز (آغاز سده بیست و یکم) شروع (شروع سال 1387) ابتدا (ابتدای قرن بیستم) اوایل (اوایل زمستان همان سال) سپیده دم (سپیده دم سال نو میلادی) زود (صبح زود)
	MID	اواسط (اواسط این ماه)
	END	پایان (پایان قرن هشتم ق.م.) آخر (آخر تابستان امسال) اواخر (اواخر آبان 84) دیر وقت (دیر وقت شب)
	APPROX	حدود (حدود 12 سال پیش) تقریباً (تقریباً ساعت 9 صبح) کم و بیش (کم و بیش 14 سال)

The following are several sentences exemplifying the usage of the various modifier elements with temporal expressions.

کمی بیش از صد سال پیش؛ انگلستان، آلمان و ایتالیا یک ناوگان مشترک را به ساحل ونزوئلا ارسال کردند.

A bit more than a hundred years ago, Britain, Germany and Italy sent a joint fleet to the coast of Venezuela.

<TIME2> کمی بیش از صد سال پیش

؛<TIME2 VAL="1907" MOD="BEFORE" ANCHOR_DIR="AFTER" ANCHOR_VAL="1906">

انگلستان، آلمان و ایتالیا یک ناوگان مشترک را به ساحل ونزوئلا ارسال کردند.¹³

¹³ See section 4.4.1 of the *TIDES Standard* for anchoring approximate offsets.

دیده بان حقوق بشر همچنین از حکومت ایران می خواهد مجازات مرگ برای همه افرادی که در هنگام ارتکاب جرم کمتر از 18 سال داشتند را به سرعت و بطور کامل لغو کند.

Human Rights Watch also wants the Iranian government to quickly and completely nullify the death penalty for all individuals who at the time the crime was committed were *less than eighteen years old*.

دیده بان حقوق بشر همچنین از حکومت ایران می خواهد مجازات مرگ برای همه افرادی که در هنگام ارتکاب جرم کمتر از 18 سال <TIME2 VAL="P18Y" MOD="LESS_THAN"> داشتند را به سرعت و بطور کامل لغو کند.

هیئت مدیره مکلف است اقلأ دو ماه قبل از تشکیل مجمع عمومی سالانه ترازنامه و حساب سود و زیان شرکت و گزارش عملکرد خود را به حسابرس ارائه دهد.

The board of directors is required to produce the company's balance-sheet and profit and loss account and its revenue report to the auditor *at least two months prior to the formation of the annual public assembly*.

هیئت مدیره مکلف است <TIME2> اقلأ دو ماه قبل از تشکیل مجمع عمومی

<TIME2 SET="YES" VAL="XXXX"> سالانه

<TIME2 VAL="P2M" MOD="EQUAL_OR_MORE">

ترازنامه و حساب سود و زیان شرکت و گزارش عملکرد خود را به حسابرس ارائه دهد.

ایران امروز در آغاز قرن بیست و یکم قرار دارد.

Iran today finds itself in the *beginning of the twenty-first century*.

ایران <TIME2> امروز <TIME2 VAL="PRESENT_REF" ANCHOR_DIR="AS_OF" ANCHOR_VAL="2007-10-22">

در <TIME2> آغاز قرن بیست و یکم <TIME2 VAL="20" MOD="START"> قرار دارد.

صبح زود قبل از خوردن هر چیزی، سطح گلیکوژن در ماهیچه ها و کبد پایین است.

Early in the morning, before eating anything, the level of glycogen in the muscles and liver is low.

<TIME2> صبح زود <TIME2 VAL="TMO" MOD="START"> قبل از خوردن هر چیزی، سطح گلیکوژن در ماهیچه ها و کبد پایین است.

اما کاربران نمی توانند تا اواسط ماه مارس این سرویس را دانلود کنند.

But users cannot download this service until *the middle of March*.

اما کاربران نمی توانند تا <TIME2> اواسط ماه مارس <TIME2 VAL="2008-03" MOD="MID"> این سرویس را دانلود کنند.

دولت ایران تقریباً دو سال پیش آنها را دستگیر کرد.

The Iranian government arrested them *approximately two years ago*.

دولت ایران <TIME2> تقریباً دو سال پیش <TIME2 VAL="2005" MOD="APPROX"> آنها را دستگیر کرد.

In addition, elements meaning *only* and *exactly* are considered modifiers and should be included in the extent of the annotation. However, they do not need to be represented with a MOD attribute (there are other temporal expressions in the following examples but only the expressions under discussion are highlighted):

مثل خیلی از گروه های دیگر فقط سه روز مانده به اجرا، مجوز گرفتند.

Like so many other groups they received their authorization *only three days before the execution of the plan*.

امروز (یکشنبه، ۱۳۸۴/۴/۱۹) دقیقاً 30 روز از اعتصاب غذای من میگذرد.

Today (Sunday, 4/19/1384) is *exactly 30 days since my hunger strike*.

دقیقا روز چهارشنبه سوری امتحانات من تموم شد.

On *ChaharShanbe Suri exactly* I finished my exams.

عصر، درست سر ساعت شش، جلو در شرقی پارک لاله بودم.

In the evening, *right at six o'clock*, I was in front of the Eastern gate of the Laleh Park.

Note that a number of the sample expressions in Table 3-3 can also be used as adverbs. These should not be included in the annotation, even if they appear next to the temporal expression, as shown in the examples below with **اقلا** “at least” and **دیروقت** “late” (shown in bold) as they are not modifying a temporal expression. Thus, the first sentence does not mean “I was happy at least today but also possibly in more future days” but rather that “I have generally not been very happy but at least, I was happy today”. Hence, **اقلا** is not a modifier of “today” but is a sentential adverb.

اقلا امروز خوشحال بودم...

At least today I was happy...

دیشب تا دیروقت پرندگان هیچکاک سرگرم کرده بود.

Last night Hichcock's “The Birds” had amused me until **late**.

حاجی، به علت مشغله زیاد، اغلب شبها مجبور بود دیروقت به خانه بیاید.

Haji, because he had too much to do, had to come home **late** on most nights.

3.4.1 Language Specific Differences

There are some issues specific to Persian that should be considered, mostly having to do with the distinct structures of Persian and English:

3.4.1.1 ON_OR_BEFORE and ON_OR_AFTER tokens

These tokens do not exist in Persian as there are no expressions such as “no less than” or “no more than” in this language. To express a similar concept, the token “less than” or “more than” has to be used along with a negation on the verb:

اعتبار نسخه های آمپول یا محلول های مخدر بیش از دو روز نخواهد بود.

The validity of the prescriptions for syringes or drug solutions will not be for *more than two days*.

[i.e., The prescriptions for syringes or drug solutions will be valid for *no more than two days*]

In these cases, there is no way of capturing the negation meaning in the annotation as it is not overtly stated in the temporal expression itself. Only the temporal expression should be annotated and tagged according to its appropriate MOD value.¹⁴

<TIME2 VAL=“P2D” MOD=“MORE_THAN”> بیش از دو روز </TIME2> اعتبار نسخه های آمپول یا محلول های مخدر نخواهد بود.

¹⁴ Note that constructions such as in the following example are not equivalent to the English sense of “no more than”. As the translation of نه shows, the negation here is the adverb meaning “not”; it is not part of the temporal expression.

فکر کنم حسین پناهی هم همین بوده که گفته زندگی رو دوست دارم اما نه بیشتر از 20 سال

I think it was for this very reason that Hossein Panahi has said I love life but not *more than 20 years*

3.4.1.2 Structural Position

Some modifier tokens can appear either before or after the temporal element. For example, “late” can be following the noun as in the first two sentences below where it is linked to the head noun by the ‘ezafe’. It can also occur preceding the temporal noun as in the last example below; here, the two elements are not linked with the ‘ezafe’ yet “late” still modifies the temporal expression.

در استان تهران از دیروقت دوشنبه شب یعنی فردا شب هوا حالت برفی ملایم و بارانی دارد.

In the Tehran province from *late Monday night* which is tomorrow night the weather will be rainy and there will be light snow.

صبحار که کلمه‌ای مندرآوردی است معادل برانچ انگلیسی است یعنی غذایی که دیروقت صبح و به جای صبحانه و نهار میل میکنید

“Sobhar” which is a new coinage is equivalent to the English “brunch” meaning a meal that you eat *late morning* and instead of breakfast and lunch.

دیشب دیروقت مطلبی رو نوشتم و چون دیروقت بود آن را پست نکردم، بایستی باز خوانی میشد.

Late last night I wrote a piece and because it was late I didn't send it, it had to be reread

Similarly, compound tokens that consist of the combination of an adjective and a preposition such as *کمتر از* “less than” or *بیشتر از* “more than” can sometimes appear in different orders around the temporal expression. For instance, the most common way to say “more than ten years” would be *سال ده بیشتر از* [Lit. more than ten year]. However, one may also encounter the construction *از ده سال بیشتر* [Lit. from ten year more], where the “more” and “than/from” parts have been separated but still appear near the temporal expression. These two constructions should be annotated differently: in the first sentence below, the modifier token is included in the extent of the annotation; in the second example, the preposition “from” is excluded from the annotation but the modifying adjective “more” is included. However, both constructions should include a MOD attribute defined as MORE_THAN.

بیش از بیست درصد از نیروی کار این کشور برای بیشتر از ده سال گذشته، بیکار بودند.

More than twenty percent of this country's workforce was unemployed for *more than the past ten years*.

بیش از بیست درصد از نیروی کار این کشور برای <TIME2> بیشتر از ده سال گذشته

<TIME2 VAL="P10Y" MOD="MORE_THAN" ANCHOR_DIR="ENDING" ANCHOR_VALUE="2007"> بیکار بودند.

مقصودم این است که وقتی دادگاه رفتی و از ده سال بیشتر بهت دادند، من در خدمت حاضریم.

What I mean is that when you go to court and they give you *more than ten years* [Lit. from ten year more], I am at your service.

مقصودم این است که وقتی دادگاه رفتی و از <TIME2> ده سال بیشتر <TIME2 VAL="P10Y" MOD="MORE_THAN"> بهت دادند، من در خدمت حاضریم.

3.4.1.3 Separated modifiers

In some instances, the modifying token may not be adjacent to the temporal expression as in the following examples where the modifier is in bold and the relevant temporal expression is underlined (italic in the English translation):

یک سال خورشیدی تقریباً معادل 365 روز، 5 ساعت و 49 دقیقه است.

A solar year is **approximately** equal to 365 days, 5 hours and 49 minutes.

این تقویم دیواری دقیقاً در روز اول سپتامبر 2008 یعنی 11 شهریور سال 1387 منتشر خواهد شد.

This wall calendar will be printed **exactly** on the first day of September 2008 which is 11 Shahrivar of 1387.

جان فلیکس انتونی سینا در 23 آوریل 1977 دیده به جهان گشود تقریباً همیشه سه اردیبهشت روز تولدش.

John Felix Anthony Cena was born on 23 April 1977, it's **almost** on three Ordibehesht his birthday.

سه بار کمر بند سنگین وزن که اقلاً از اول سال دوهزار و پنج تا آخره دو هزار هفت دستش بود

Three time heavyweight championship- that he has held **at least** from the beginning of two thousand five until towards the end of two thousand eight.

In these cases, as in the similar instances in English, the modifier token should not be included in the extent of the annotation and the MOD value will not be represented in the tag.

3.4.1.4 The preposition تا

The preposition تا can sometimes modify a duration expression when it is used in the meaning of “up to” in English. For instance, in the following sentence, تا is used to refer to an amount that is equal to or less than the temporal amount specified in the expression (i.e., in three years or less, the number of internet users will be 35 million people). In these cases, although prepositions are generally not included in the temporal expression tags, the preposition تا is treated as a modifier token. It is included within the extent of the annotation and should be represented with the MOD value “EQUAL_OR_LESS”.

هم اکنون تعداد استفاده کنندگان از اینترنت در ایران هفت و نیم میلیون نفر است ولی انتظار می رود تا سه سال دیگر این تعداد به سی و پنج میلیون برسد.

Currently, the number of Internet users in Iran is seven and half million people but it is expected that up to three years from now this number will read thirty five million.

The following are additional examples for the modifier usage of the preposition تا in Persian with annotations. Note that these always occur with duration expressions.

شخصی که از قبل دارای گواهینامه و یا تجربه رانندگی نباشد تا یکسال حق دادن امتحان رانندگی را ندارد.

An individual who has not previously held a driver's license or does not have driving experience does not have the right to take the driving test for up to a year.

شخصی که از قبل دارای گواهینامه و یا تجربه <TIME2> تا یکسال <TIME2 VAL=“P1Y” MOD=“EQUAL_OR_LESS”>، رانندگی نباشد حق دادن امتحان رانندگی را ندارد.

او می دانست که نان آورش تا شش ماه دیگر در زندان خواهد ماند.

He knew that his breadwinner will be in jail for up to six more months.

او می دانست که نان آورش <TIME2> تا شش ماه دیگر <ANCHOR_VALUE=“2007-10-22”>، حق دادن امتحان رانندگی را ندارد. <TIME2 VAL=“P6M” MOD=“EQUAL_OR_LESS” ANCHOR_DIR=“STARTING”>

این برنامه پربیننده چنان دست و پاگیر دست اندرکاران آن شد که تا سه سال امکان فعالیت مجدد تصویری را از آنان گرفت.
This highly watched show was so constraining for those involved with it that for *up to three years* it did not allow them to participate in another picture.

این برنامه پربیننده چنان دست و پاگیر دست اندرکاران آن شد که <TIME2> تا سه سال
<TIME2 VAL="P3Y" MOD="EQUAL_OR_LESS" ANCHOR_VALUE="2007">
امکان فعالیت مجدد تصویری را از آنان گرفت.

In the following example, however, *تا* is not used as a modifier since the temporal expression is interpreted as until the year 2005 (but not including it).

سازمان ملل: جمعیت جهان تا سال 2050 به 9 میلیارد نفر خواهد رسید.
The United Nations: The population of the world will reach 9 Billion by *the year 2050*.

سازمان ملل: جمعیت جهان تا <TIME2> سال 2050 <TIME2 VALUE="2050"> به 9 میلیارد نفر خواهد رسید.

The sentences below show further usages of the non-modifying usage of the preposition *تا* usually means “until” or “up until” in English. In these cases, it is excluded from the extent of the annotation.¹⁵

شایعه شده که کنسرت‌های پاپ ایرانی در ترکیه لغو شده اما مطمئن باشین تا آخر شهریور برگزار می‌شن.
There is a rumor that Iranian pop concerts in Turkey have been cancelled but be sure that they will be held by *the end of Shahrivar*.

نوشیدن مشروبات الکلی از ابتدای انقلاب ایران در سال ۱۳۵۷ ممنوع شده و این ممنوعیت تا به امروز به قوت خود باقیست.
Drinking of alcoholic beverages has been banned since the beginning of the Iranian revolution in 1357 AP and this prohibition remains still strong until *today*.

تا پیرروز هم بازیگر مرد فیلم را پیدا نکرده بودیم تا اینکه مردی فرانسوی پیدا شد و قرار است بهار امسال در ایتالیا فیلمبرداری را شروع کنیم.
Even up until *the day before yesterday* we had not found the male actor for the movie until a French man was found and we are supposed to start filming in Italy in the spring of this year.

این فیلم تا حداکثر 30 سپتامبر در قفسه خواهد بود.
This movie will be sitting on the shelf until *at most the 30th of September*.

Note that another usage of the preposition *تا* is to present a range as in the sentences below, where it should also be excluded from the annotation (see Section 4.2.2 for more on Range Expressions).

موسسه بین المللی مطالعات استراتژیک در بریتانیا طی گزارشی اعلام کرد که ایران تا زمان بدست آوردن مواد لازم برای تولید بمب اتمی دو تا سه سال فاصله دارد.

The International Institute for Strategic Studies in the UK has announced in a report that Iran is *two to three* years away from obtaining the necessary materials for the creation of the Atomic bomb.

دانشجویان دانشگاه امیرکبیر بعد از ماه‌ها شکنجه و آزار و اذیت در زندان به چهار تا شش سال زندان محکوم شده‌اند.
The students of Amir Kabir University, after months of torture and persecution in jail, have been sentenced to *four to six years* in prison.

¹⁵ There are other temporal expressions in these sentences; however, only the ones introduced by *تا* are underlined.

4 Determining the Extent of the Annotations

This section contains rules for determining where each temporal expression begins and ends. We refer to this as the *extent* of the expression.

4.1 Lexical and Morphological Criteria

4.1.1 Lexical Criteria

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

- noun (including proper noun): e.g., امروز “today”, جمعه “Friday”
- noun phrase (NP): e.g., اون روزا “those days”, روز جمعه “Friday” (lit. “day of Friday”), این دو سال اخیر “the last two years”, یک ساعت تمام “a whole hour”, دو هفته پیش “two weeks ago”
- adjective: e.g., کنونی “current”
- adverb: e.g., اخیرا “recently”
- adjective/adverb phrase: e.g., تقریبا نیم ساعت “nearly half hour”, پنجاه و دو ساله “fifty-two year old”

The full extent cannot be a preposition phrase (i.e., the expression cannot start with a preposition¹⁶) or a clause of any type (i.e., the expression cannot start with a subordinating conjunction). Thus, from the following phrases we get the annotated expressions listed below:

Context	Extent of TIME2 Expression
قبل از پنجشنبه “before Thursday”	پنجشنبه “Thursday”
از صبح زود “from (the) early morning”	صبح زود “(the) early morning”
سرانجام بعد از اتمام اعتصاب کارگران در روز پنجشنبه “finally after the end of the workers’ strike on Thursday”	روز پنجشنبه “Thursday”
طی دو سال گذشته “over the last two years”	دو سال گذشته “the last two years”

¹⁶ But see Section 3.4 on “Modified Temporal Expressions” for a discussion of prepositions that act as modifiers.

Note that certain prepositions in Persian are in fact compound elements, generally consisting of a noun and a preposition. Examples include “before”, “after”, “in/during”. These compound elements should be treated as prepositions and should not be included in the annotated expression. However, as discussed in the last chapter, certain compound elements consisting of an adjective and a preposition that can be used to modify a temporal expression, such as “less than” and “more than”, should be included within the extent of the annotation. Similarly, modifier expressions like “about”, “close to” and تا (when used to mean “up to”) function more as modifiers of temporal expressions, and although they are prepositions, they should be included in the extent of the tag.

4.1.2 Morphological criteria

A number of elements such as pronouns, indefinite markers, and object markers, may be written as separate entities in Persian orthography or they can appear as affixes joined to the temporal expression. This section discusses which ones of these morphemes should be included within the extent of the annotation and which ones should be excluded.

The following are examples illustrating various elements such as pronouns, indefinite articles, relativizers, and definite articles, which should be included within the extent of the annotation if they are modifying a temporal expression. These elements are included in the annotation whether they appear as affixes attached to the previous element (e.g., the pronoun in the first example below) or as separate entities (as in the second example with ما “our”).

Pronouns

امروز روز تولدم بود.
Today was my birthday.

<TIME2> امروز <TIME2 VAL=“2007-10-22”> روز تولدم <TIME2> بود.

آینده‌ی ما اعتیاد و بی‌کاریست و یا در بهترین شکل مهاجرت از ایران!
Our future is addiction and unemployment or in the best case scenario emigration from Iran!

<TIME2> آینده‌ی ما
<TIME2 VAL=“FUTURE_REF” ANCHOR_DIR=“AFTER” ANCHOR_VAL=“2007-10-22”>
اعتیاد و بی‌کاریست و یا در بهترین شکل مهاجرت از ایران!

Indefinite article or relativizer

The indefinite article has the forms ای / بی / ی and usually means “a/an” as in کتابی “a book”, or “some” if it is used with a plural noun as in کتابهایی “some books”. The relativizer has the same form as the indefinite article but is used on a noun or adjective when the latter is followed by که “that” which introduces a relative clause as in ... کتابی که من خریدم “the book that I bought”.¹⁷

¹⁷ The form بی / ی is also used to form adjectives as in ما، انسانهای قرن بیست و یکمی “us, twenty-first century humans”. In here, the ی affix is not an indefinite article meaning “a twenty first century” but rather turns the whole thing into

در چنين روزي در سال 1285 خورشیدی پروين اعتصامي شاعر برجسته ايراني به دنيا آمد.
On such a day in 1285 AP Parvin Etessami the famous Iranian poet was born.

در <TIME2 CAL="PERSIAN_FA" VAL="1285-XX-XX"> چنين روزی در
<TIME2> سال 1285 خورشیدی <TIME2 CAL="PERSIAN_FA" VAL="1285"> پروين اعتصامي شاعر برجسته
ايراني به دنيا آمد.

شاعران جوان تجربه های فردی خود را با نشانه های قرن آشوبی در ساحت شعر این دهه اعلام کرده اند.
The young poets have proclaimed their personal experiences with signs of a tumultuous century upon this decade's poetry scene.

شاعران جوان تجربه های فردی خود را با نشانه های <TIME2> قرن آشوبی <TIME2 VAL="XX"> در ساحت شعر
<TIME2> این دهه <TIME2 VAL="200"> اعلام کرده اند.

Definite article

The definite article is used only in the colloquial language as in کتابه “the book” or اون کتابه “that book”:

من اين تابستونی که شبیه اون زمستونه هست رو نمی خوام.
I don't want *this summer* that is similar to *that winter*.

من <TIME2> اين تابستونی <TIME2 VAL="2007-SU"> که شبیه
<TIME2> اون زمستونه <TIME2 VAL="XXXX-WI"> هست رو نمی خوام.

4.1.2.1 Pronouns

As already mentioned, certain elements such as possessive pronouns, if they are part of the temporal expression, should be tagged within the annotation. For instance, the pronoun “our” is included within the annotation of the temporal expression هفته گذشته مان “our last week” as in its English counterpart. It is, however, not included in the annotation of جلسه هفته گذشته مان “our last week’s meeting = our meeting of last week” because in this case, the pronoun is actually modifying the syntactic head of the full noun phrase, namely جلسه “meeting” (i.e., “our meeting” not “our week”). The two distinct annotations are illustrated below:

هفته گذشته مان

our last week = this last week of ours

<TIME2> هفته گذشته مان <TIME2 VAL="2007-W38">

جلسه هفته گذشته مان

our last week's meeting = our meeting of last week

جلسه <TIME2> هفته گذشته مان <TIME2 VAL="2007-W38">

Note that at times the annotator may need to exclude an attached affix from the extent of the annotation. In certain cases, the meaning obtained may not be as easy to distinguish since the possessor pronoun is always placed at the end of the noun phrase in Persian. For instance, the

an adjective. As expected, it needs to be included within the annotation of the expression “twenty-first century”.

sentence [Lit. poets of this century of our] is ambiguous based on whether it is referring to “our century” or “our poets”. The annotator should use the context and the meaning of the sentence in each case to determine the correct annotation. The two possible annotations are shown below:

شاعران این قرنمان
The poets of *this century of ours* / The poets of *our century* [refers to “our century”]
< TIME2 VAL=“2007-W38”> این قرنمان </TIME2>

شاعران این قرنمان
Our poets of *this century* [refers to “our poets”]
شاعران < TIME2 VAL=“2007-W38”> این قرن </TIME2> مان

4.1.2.2 Copula verb

The copula verb is the verb بودن ‘to be’ in the present tense. The copula verb is never included within the annotation of the expression as shown in the following examples where the temporal expression is underlined and the copula is shown in bold:

امشب شش سال و چند ساعت است که او رفته است.
Tonight it's six years and a few hours that he is gone.

گفت راستی نمی دونی اونا مال کدوم دهه اند؟ گفتم دقیق نمیدونم شاید هفتاد یا هشتاد!
She said hey don't you know which *decade* they **are** from? I said not really maybe *seventies* or *eighties*.

The copula verb may appear attached to the preceding temporal element, but it should still be excluded from the extent of the annotation. Hence, in the first example below the suffix “is” and in the second sentence the suffix “am” are not included in the annotation.

کمی بیش از دو ساله که با همسرم آزاده در شهر وینیپگ، کانادا، زندگی می کنیم.
It's a bit more than two years that my spouse Azadeh and I are living in the city of Winnipeg, Canada.

شرم دارم که انسان این قرنم !!
I am ashamed that I am a human from *this century*.

In the following example, the final adjective has both an indefinite article and a copula verb (در دناک + ی + ه); the indefinite marker is included but the copula verb is excluded from the annotation as in English:

زمستان سرد در دناکیه.
It is a cold painful winter.

4.1.2.3 Object and topic marker را

In Persian, specific objects are marked with را as in کتاب را خریدم “I bought the book”. The same affix can appear on *topicalized* elements (the topic of the discussion) as in من پاریس خریدم این کتاب را, which in English would translate to “As for this book, I bought it in Paris”. The affix را should never be included within the extent of the annotation, whether it appears attached to or detached from the temporal expression. Note that را can also be written as رو or و in colloquial writing, in which case it is often joined to the preceding word.

این کودکان مجبورند به دنبال يك لقمه نان، تمام روز را در مترو، اتوبوس و گوشه خیابان آدامس بفروشند.
These kids are forced to sell gum *all day long* in the metro, bus and the street corner, in search of a piece of bread.

صحنه ی ریجکت شدگان دم در سفارت اون دو هفته رو برام کرده بود جهنم...
The scene of those rejected at the door of the embassy made *those two weeks* hell for me...

امیدوارم که حالتان خوب باشه و سیزده امسالو به خوبی گذرونده باشید.
I hope that you are feeling well and that *this year's Sizdah Bedar* was spent well.

شرکت هما پروازهای امروز را لغو کرده است.
The Homa company has cancelled *today's* flights.

4.2 Syntactic Criteria

The full extent of the tagged expressions includes all **premodifiers** of the time expression, among which are determiners (هر، آنها، این)، numerals (صدها، هزاران)، quantifiers (یک، چهارده، صدها، هزاران)، (طولانی‌ترین، بهترین)، (خیلی، چند، بعضی) and superlative adjectives (طولانی‌ترین، بهترین)، as well as all **postmodifiers** in the noun phrase which in Persian include possessive noun phrases, pronouns, and all nominal/adjectival postmodifiers. For example, all of the following represent the full extent of markable temporal expressions:

روز بعد	<i>the next day</i>	فقط یک سال پیش	<i>just a year ago</i>
صبح زود	<i>early in the morning</i>	تنها یه ساعت	<i>only one hour (long)</i>
این روز های سرد زمستانی	<i>these cold winter days</i>	بیش از یک ماه	<i>more than a month</i>
اوایل تابستون همون سال	<i>early summer of that year</i>	دقیقا دو سال پیش	<i>exactly two years ago</i>
بهار آینده	<i>next spring</i>	چند سال پیشتر	<i>several years earlier</i>
دهه های اخیر	<i>recent decades</i>	پنج ساله	<i>five years old</i>
نصف شب سه شنبه ها	<i>midnight on Tuesdays</i>	سه هفته بعدش	<i>three weeks later</i>
آینده خود را	<i>its own future</i>	ساعتها پیش	<i>hours earlier</i>
آینده کشورمون	<i>our country's future</i>	دیشب دیروقت	<i>late last night</i>
سال های ابتدایی انقلاب	<i>the early years of the revolution</i>	کمی بیشتر از ده سال بعد	<i>a bit more than ten years later</i>
غمگین ترین روز جهان	<i>the saddest day in the world</i>	خیلی روزها	<i>many days</i>
مقدس ترین روز مذهبی یهودی	<i>the holiest Jewish religious day</i>	یک ساعت و ۱۰ دقیقه زودتر از موعد مقرر	<i>an hour and 10 minutes before schedule</i>

Note that what is referred to as adjective phrases in the English examples in *the TIDES Standard for the Annotation of Temporal Expressions* are often noun phrases in Persian as their structure is slightly different as in “a year ago” (یک سال پیش).

The full extent also includes all **postmodifiers** of the time expression, among which are adverbs, preposition phrases, and dependent clauses. For example, all of the following represent markable temporal expressions (in these examples, the heads are underlined and the postmodifiers are shown in italic):

یک روز بعد از زلزله‌ی تهران	one <u>day</u> <i>after the Tehran earthquake</i>
آینده‌ی جامعه‌ی ما	the <u>future</u> <i>of our society</i>
نزدیک به سه دهه حکومت ملایان	close to three <u>decades</u> <i>of the regime of the Mullahs</i>
و افسین ماههای پیش از نابودی میهن قشنگم	the last <u>months</u> <i>prior to the destruction of my beautiful country</i>
همین هفت سال پیش، که میرعلایی را کشتند	just seven <u>years</u> <i>ago when Mir-Alayi was murdered</i>
یک روز تاریخی برای ایرانیان و فارسی‌زبانان	a historic <u>day</u> <i>for Iranians and Persian speakers</i>

In Persian, however the postmodifier may sometimes be separated from the temporal head since structural elements may move around more freely. As a general rule, the postmodifier should be included within the extent of the annotation of the temporal expression if it appears adjacent to the syntactic head, but not if it is separated by intervening elements. For instance, the sentence “It has been 20 years since the death of Gholamhossein Saedi” can be translated into Persian as in the first sentence below, where the temporal noun سال “year” and the postmodifier are adjacent and can be included within the same annotation:

بیست و دو سال از مرگ غلامحسین ساعدی گذشت.

It has been *twenty years since the death of Gholamhossein Saedi.*

<TIME2> بیست و دو سال از مرگ غلامحسین ساعدی

<TIME2 VAL=“P20Y” ANCHOR_DIR=“ENDING” ANCHOR_VAL=“2007-10-22”> گذشت.

But the following structure where the verb intervenes between the temporal element and the postmodifier is quite common, especially in colloquial style. In this case, the postmodifier should be ignored in the annotation.

بیست و دو سال گذشت از مرگ غلامحسین ساعدی.

Twenty years have passed since the death of Gholamhossein Saedi.

<TIME2> بیست و دو سال

از مرگ غلامحسین ساعدی گذشت.

Non-adjacent temporal nouns and postmodifiers are especially common when the postmodifier is a relative clause as in the following examples where the intervening verb is shown in bold:

همین هفت سال پیش بود که میرعلایی را کشتند.

It was *just seven years ago* that Mir Allayi was murdered.

خانم یکی از آشناهامون از اولین روزی صحبت می کرد که مجبور بود تنها دخترش رو به یکی از این دبستان های اسلامی برسونه.

The wife of one of our relatives was talking about *the first day* that she had to take her daughter to one of these Islamic schools all by herself.

Note also that the inclusion of adjacent pre- and post modifiers in a temporal expression applies only to the modifiers of the time expression (i.e., the head of the phrase has to be itself a markable 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 TIME2 tag. For example, in the following, the head noun پیام “message” and the postmodifier آیت الله خامنه ای “of Ayatollah Khamenei” are not included in the TIME2 tag, and neither is the non-temporal premodifier of that string, namely the determiner این “this.”

این پیام اخیر آیت الله خامنه ای به ملت بزرگ ایران است.

This is Ayatollah Khamenei’s *recent* message to the great people of Iran.

این پیام

آیت </TIME2> اخیر <TIME2 VAL=“PAST_REF” ANCHOR_DIR=“BEFORE” ANCHOR_VAL=“2007-10-22”>
الله خامنه ای به ملت بزرگ ایران است.

In this next example, neither the noun دانشجویان “University students” after 2 ساعته “two-hour long” nor the preceding noun جلسه ای “meeting (of)” is included within the scope of either TIME2 tag in Persian.

من امروز صبح به جلسه ای 2 ساعته دانشجویان نتونستم برسم.

I was not able to make it to the *two-hour long* meeting of the University students *this morning*.

من </TIME2> امروز صبح <TIME2 VAL=“2007-10-22TMO”> به جلسه ای
</TIME2> 2 ساعته <TIME2 VAL=“PT2H”> دانشجویان نتونستم برسم.

Similarly, the strings پیکنیک “picnic” and در بلک استرپ “in Black Strap” are not included within the scope of the TIME2 tag below:

عکسی از پیکنیک هفته پیش در بلک استرپ.

A photo from *last week’s* picnic in Black Strap.

عکسی از پیکنیک </TIME2> هفته پیش <TIME2 VAL=“2007-W46”> در بلک استرپ.

Finally, in this last example, as the syntactic head is سن “age” and is not a temporal expression, the head noun and its modifier حداقل “at least” are not included in the extent of the annotation.

برای افتتاح حساب های سپرده سرمایه گذاری حداقل سن 18 سال تمام الزامی می باشد.

To establish investment accounts a minimum age of *18 years completed* is required.

برای افتتاح حساب های سپرده سرمایه گذاری حداقل سن

</TIME2> 18 سال تمام <TIME2 VAL=“P18Y”> الزامی می باشد.

4.2.1 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. In the following example, “the years of the chain murders” is an appositive that includes a temporal expression and it receives an embedded annotation.

دهه‌ی هفتاد، سالهای قتل‌های زنجیره‌ای، از تاریک‌ترین نقاط تاریخ معاصر است.
The Seventies, the years of the chain murders, was one of the darkest points of modern history.

<TIME2 CAL=“PERSIAN_FA” VAL=“137”> دهه‌ی هفتاد، </TIME2> </TIME2> سالهای قتل‌های زنجیره‌ای
<TIME2 CAL=“PERSIAN_FA” VAL=“137”> از تاریک‌ترین نقاط تاریخ معاصر است .

The verb-final word order in Persian may make it difficult to distinguish an appositive from a predicate element. An appositive is generally described as a noun phrase set beside another noun phrase to define, modify, or identify it. Appositives are therefore usually optional and can be removed from the sentence without affecting the sentence structure. In the two examples below, however, the underlined phrase is a predicate nominal and not an appositive. A predicate nominal describes the subject noun phrase, is linked to it by the verb ‘to be’, and is not optional. Hence, the subject noun phrase and the predicate nominal should receive separate TIME tags. Note that in English the predicate nominal appears following the verb.

سی خرداد شصت، روز آغاز سیاه‌ترین سال هاست.
30 Khordad '60, is the day of the beginning of one of the blackest years.

دهه شصت میلادی، دهه آرمان‌های بزرگ بود.
The '60s, was the decade of big ideals.

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.

تعداد کمی گزارشات تاریخی راجع به جنگ کابل در 1992 الی 1996 وجود دارد.
There exist few historical reports about the war in Kabul from 1992 through 1996.

تعداد کمی گزارشات تاریخی راجع به جنگ کابل در <TIME2 VAL=“1992”>1992 </TIME2> الی <TIME2 VAL=“1996”>1996 </TIME2> وجود دارد.

ویزای هند را میتوان در 10-12 روز کاری دریافت کرد.¹⁸
One can obtain a visa for India in 10-12 workdays.

ویزای هند را میتوان در <TIME2 VAL=“P12D”>12 </TIME2> - <TIME2> 10 روز کاری <TIME2 VAL=“P10D”> دریافت کرد.

نمایشگاه هنری از بیست و یکم تا بیست و هشتم مرداد در تبریز برگزار می‌شود.
The art exhibit is being held in Tabriz from the twenty first until the twenty eighth of Mordad .

¹⁸ Note that there are no specific tags at this point for “workday” in TIME2.

نمایشگاه هنری از <TIME2> بیست و یکم <TIME2 CAL="PERSIAN_FA" VAL="1386-05-21"> تا <TIME2> بیست و هشتم مرداد <TIME2 CAL="PERSIAN_FA" VAL="1386-05-28"> در تبریز برگزار می‌شود.

یکشنبه ها : کارگاه آموزشی خصوصی ، ویژه خانمهای مسافر ۲ تا ۴ بعد از ظهر

Sundays: Private educational workshop, especially for traveling women 2 to 4 in the afternoon.

<TIME2> یکشنبه ها <TIME2 CAL="PERSIAN_FA" VAL="1386-WXX-2"> : کارگاه آموزشی خصوصی ،

ویژه خانمهای مسافر <TIME2> 2 <TIME2 CAL="PERSIAN_FA" VAL="1386-WXX-2T14"> تا

<TIME2> ۴ بعد از ظهر <TIME2 CAL="PERSIAN_FA" VAL="1386-WXX-2T16">

مایکروسافت اعلام کرد که این شرکت ظرف سه تا پنج سال آینده تبدیل به یکی از بازیگران اصلی بازار تبلیغات آنلاین خواهد شد.

Microsoft has announced that within *the next three to four years* this company will become one of the main players in the field of online advertising.

مایکروسافت اعلام کرد که این شرکت ظرف

<TIME2> سه <TIME2 VAL="P3Y" ANCHOR_VAL="2007" ANCHOR_DIR="AFTER">

<TIME2> پنج سال آینده <TIME2 VAL="P5Y" ANCHOR_VAL="2007" ANCHOR_DIR="AFTER">

تبدیل به یکی از بازیگران اصلی بازار تبلیغات آنلاین خواهد شد.

4.2.3 Conjoined Expressions

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

گروه بابک ریاحی پور روزهای ۲۹ و ۳۰ آذر و اول دی ماه در سالن اختصاصی کاخ نیاوران به اجرای کنسرت می‌پردازد.

Babak Riahipour's group will be performing in concert on *the days of 29 and 30 Azar and the first of Dey* in the private hall of the Niavaran Palace.

گروه بابک ریاحی پور <TIME2> روزهای ۲۹ <TIME2 CAL="PERSIAN_FA" VAL="1386-09-29"> و

<TIME2> ۳۰ آذر <TIME2 CAL="PERSIAN_FA" VAL="1386-09-30"> و

<TIME2> اول دی ماه <TIME2 CAL="PERSIAN_FA" VAL="1386-10-01"> در سالن اختصاصی کاخ نیاوران به

اجرای کنسرت می‌پردازد.

بین امروز و اواسط نوامبر شما نیاز خواهید داشت که بخود تکیه بیشتری داشته باشید.

Between *today and the middle of November* you may need to lean a bit more on yourself.

بین <TIME2> امروز <TIME2 VAL="2007-10-22"> و

<TIME2> اواسط نوامبر <TIME2 VAL="2007-11" MOD="MID"> شما نیاز خواهید داشت که بخود تکیه بیشتری

داشته باشید.

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

البته در این شیوه حکومت منافع آمریکایی ها بر منافع کوتاه و دراز مدت افغان ها اولویت دارد.

Of course in this style of government the interests of Americans will take precedence over *the short and long term* interests of Afghans.

البته در این شیوه حکومت منافع آمریکایی ها بر منافع <TIME2>

کوتاه <TIME2 VAL="FUTURE_REF" ANCHOR_VAL="2007-10-22" ANCHOR_DIR="AFTER">

و <TIME2> دراز مدت

< TIME2 VAL="FUTURE_REF" ANCHOR_VAL="2007-10-22" ANCHOR_DIR="AFTER">
افغان ها اولويت دارد.

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

اگر اضطراب و نگرانی شدید را به طور مداوم برای مدت شش ماه یا بیشتر تجربه کرده اید، ممکن است اختلال اضطراب منتشر داشته باشید.

If you have experienced long-lasting agitation and strong anxiety for a period of six months or more, you may have Generalized Anxiety Disorder.

اگر اضطراب و نگرانی شدید را به طور مداوم برای
< TIME2 VAL="P6M" MOD="EQUAL_OR_MORE"> مدت شش ماه یا بیشتر
است اختلال اضطراب منتشر داشته باشید.

In Persian, a conjunction may be represented by juxtaposing two numbers in order to indicate vagueness. For instance, "two three days" would mean "two or three days". These should be tagged as two distinct conjoined expressions.

دادسرای انقلاب: نازی عظیمای باید یکی دو سال در ایران بماند.

The public prosecutor's office of the Revolution: Nazi Azima must stay in Iran for about one - two years.

دادسرای انقلاب: نازی عظیمای باید < TIME2 > یکی < TIME2 VAL="P1Y">
< TIME2 > دو سال < TIME2 VAL="P2Y"> در ایران بماند.

این برنامه تا سه چهار هفته پیش که جلوییش را گرفتند پخش می شد.

This program was being broadcast until three - four weeks ago when they blocked it.

این برنامه تا < TIME2 VAL="2007-10-1"> سه < TIME2 >
< TIME2 > چهار هفته پیش < TIME2 VAL="2007-09-24"> که جلوییش را گرفتند پخش می شد.

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 (a) there is no intervening element between temporal terms and (b) the two terms express values for units that are hierarchically related. For example:

جمعه شب

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

نوامبر 1943

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 TIME2 tag:

جمعه شب *Friday evening*
 ساعت ۸ شب جمعه *8:00 p.m. Friday*
 روز یکشنبه ۱۸ آذرماه *Tuesday the 18th of Azar*
 نوامبر ۱۹۴۳ *November 1943*
 دوازدهم ژانویه سال ۱۹۵۵ میلادی *the 12th of January of 1955 in the Gregorian calendar*

Note that the familiar “day month year” format is considered a single TIME2 expression. A comma between *TOD* and *Calendar Date* expressions is ignored, and the whole is considered a single TIME2 expression in spite of the punctuation. Thus, the following are single TIME2 expressions (notice that the English versions of the date contain a comma, which is likewise ignored):

11:11:15، ۱۳۸۶ آبان ۲۲ سه‌شنبه، *Tuesday, 22 Aban 1386, 11:11:15*
 روز دوشنبه (۵ نوامبر) *Monday (5 November)*

Note however that if the expression includes two different calendars, they need to receive two distinct tags. In the following example, each underlined set of elements receives a separate annotation:

چهاردهم دسامبر ۲۰۰۶ (23 آذر ماه) *Fourteenth of December 2006 (23 Azar)*

In Persian, the elements within a noun phrase are joined to each other using the *ezafe* suffix. English, on the other hand, often uses prepositions to link the noun phrase constituents. Hence, the constructions described in section 5.2.4.1 of the *TIDES Standard for the Annotation of Temporal Expressions* are expressed without prepositions and with the *ezafe* construction in Persian and therefore the same issues do not arise (also see Section 3.1 in this document). Hence, prepositions in Persian introduce syntactically embedded phrases and therefore typically require a separate TIME2 tag. There are, however, several instances in which you should ignore the preposition in Persian and create a single tag:

1. The preposition “to” in time-of-day (*TOD*) expressions.

ده دقیقه به سه *ten minutes to three*

2. The preposition “from” in certain partitive constructions that are not formed with the ‘*ezafe*’.

هر ساعتی از شبانه روز *any hour in the day*

3. The preposition “after” in TOD expressions in the night.

ساعت ۲:۳۰ بعد از نیمه‌شب *2:30 am [lit. 2:30 after midnight]*

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

For instance, to say “(eight glasses of water) *per day*”, any of the following can be used:

در روز
 در هر روز
 در شبانه روز
 در طول روز
 در کل روز

In certain time expressions, other elements besides a preposition may intervene such as a verbal participle. These elements are part of the time expression and should be included in the annotation:

پانزده دقیقه مانده به دو *fifteen minutes left to two*

4.2.4.2 When to Create Multiple Tags, with Embedding

There are two situations in which a TIME2 tag is embedded in the extent of another in Persian:

- 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.

دو هفته بعد از عمل روز سه شنبه می‌توانید به خانه برگردید.
 You can return home *two weeks after Tuesday's operation*.

<TIME2 VAL="2007-10-12">روز سه شنبه</TIME2> عمل<TIME2 VAL="2007-10-26"> دو هفته بعد از عمل می‌توانید به خانه برگردید.

چهار سال پیش در همین روز زمین‌لرزه‌ای در شهر بام روی داد.
Four years ago today an earthquake took place in the city of Bam.

<TIME2 VAL="2007-10-22">چهار سال پیش در</TIME2> همین روز<TIME2 VAL="2003-10-22"> زمین‌لرزه‌ای در شهر بام روی داد.

- 2. 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.

تولدش در روز 20 اردیبهشت
His 20 Ordibehesht birthday

<TIME2 CAL="PERSIAN_FA" VAL="1386-02-20">تولدش در</TIME2> روز 20 اردیبهشت<TIME2 CAL="PERSIAN_FA" VAL="1386-02-20">

فصل عید

the holiday season

<TIME2> فصل <TIME2> عید <TIME2>

ضرب الاجل ماه دسامبر

The December deadline

<TIME2 VAL="2007-12"> <TIME2 VAL="2007-12"> ماه دسامبر </TIME2> ضرب الاجل </TIME2>

یک روز کاری 8 ساعته

An 8-hour work day

<TIME2 VAL="P1DT"> <TIME2 VAL="PT8H"> 8 ساعته </TIME2> یک روز کاری </TIME2>

Possessive Constructions. As possessions are usually formed with the ‘ezafe’ affix in Persian, the distinctions that arise in English cannot be found in this language. Hence, English can make a distinction between *this year’s summer* and *the summer of this year* and may annotate the two constructions differently. However, in Persian, both noun phrases will be represented simply as تابستان امسال [summer-Ez this year] and will receive a single tag.

Also note that in Persian, as most time expressions include the specific time of day, the whole expression should receive a single extent. In these expressions, both the time and the time of day part are describing values for the same **TOD** unit and they are linked with the *ezafe* affix:

ساعت دوازده شب *twelve o’clock at night*

ساعت دوازده ظهر *twelve o’clock noon*

صبح ساعت پنج *five o’clock in the morning*

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. In Persian, if the elements of a temporal expression are joined to each other with the *ezafe*, then in general the whole phrase should be treated as one unit even if the English translation may qualify for two distinct tags:

کنسرتمان ساعت هشت روز چهارشنبه خواهد بود.

Our concert will be at *eight o’clock Wednesday*

However, if a preposition or other parts-of-speech such as a verb intervene between the two temporal expressions, they should receive distinct tags:

کنسرتمان ساعت هشت در روز چهارشنبه خواهد بود.

Our concert will be at *eight o’clock on Wednesday*.

کنسرتمان روز چهارشنبه است سر ساعت هشت.

Our concert is on *Wednesday exactly at eight*.