









## 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* construction<sup>2</sup> (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.

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









## 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”



























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) *and* نیمه‌شب 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* “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 *بعد از دو ساعت* “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 *بیش از دو ساعت* “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

**Table 3-3 Modifier Tokens**

	Token	Sample Expressions
<b>Points</b>	BEFORE	بیش از (بیش از یکصد سال قبل)
	AFTER	کمتر از (کمتر از یک دهه پیش)
<b>Durations</b>	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) بالا تر از (بالا تر از 18 سال)
	EQUAL_OR_LESS_THAN	حداکثر (حداکثر 15 روز کاری)
	EQUAL_OR_MORE_THAN	حداقل (حداقل پنج دقیقه) اقلاً (اقلاً دو ماه)
<b>Points and Durations</b>	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.<sup>13</sup>

<sup>13</sup> See section 4.4.1 of the *TIDES Standard* for anchoring approximate offsets.











## 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 preposition<sup>16</sup>) 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”

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





درچنین روزی در سال 1285 خورشیدی پروین اعتصامی شاعر برجسته ایرانی به دنیا 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"> چنین روزی در  
<TIME2> سال 1285 خورشیدی <TIME2 CAL="PERSIAN\_FA" VAL="1285"> پروین اعتصامی شاعر برجسته  
ایرانی به دنیا آمد.

شاعران جوان تجربه های فردی خود را با نشانه های قرنی 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> قرنی 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 این تابستونی که شبیه اون زمستونه هست رو نمی خوام.  
I don't want *this summer* that is similar to *that winter*.

من <TIME2> این 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 مان “our” is included within the annotation of the temporal expression مان هفته گذشته “our last week” as in its English counterpart. It is, however, not included in the annotation of مان جلسه هفته گذشته “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 <TIME2 VAL="2007-W38"> مان

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 قرن بیست و یکمی “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>
دهه‌های اخیر	<i>recent decades</i>	پنج ساله	<i>five years old</i>
نصف شب سه‌شنبه‌ها	<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”> آیت 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 قتل های زنجیره ای، از تاریک ترین نقاط تاریخ معاصر است.  
*The Seventies, the years of the chain murders, was one of the darkest points of modern history.*

<TIME2> </TIME2> دههی هفتاد، <TIME2 CAL=“PERSIAN\_FA” VAL=“137”> سالهای قتل های زنجیره ای </TIME2 CAL=“PERSIAN\_FA” VAL=“137”> از تاریک ترین 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 سیاه ترین سال هاست.  
*30 Khordad '60, is the day of the beginning of one of the blackest years.*

دهه شصت میلادی، دهه آرمان های 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.<sup>18</sup>  
*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 می شود.  
*The art exhibit is being held in Tabriz from the twenty first until the twenty eighth of Mordad .*

<sup>18</sup> 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> 4 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 > مدت شش ماه یا بیشتر < TIME2 VAL="P6M" MOD="EQUAL\_OR\_MORE"> تجربه کرده اید، ممکن 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 18<sup>th</sup> of Azar*  
 نوامبر ۱۹۴۳ *November 1943*  
 دوازدهم ژانویه سال ۱۹۵۵ میلادی *the 12<sup>th</sup> 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*.