

**ACE (Automatic Content Extraction)
Chinese Annotation Guidelines for
TIMEX2 (Summary)**

Version 1.2 20050610

1	Determining what kinds of expressions to annotate.....	3
1.1	Markable Expressions	3
1.2	Non-Markables	4
2	Capturing the meaning of temporal expressions.....	4
2.1	Precise Temporal Expressions.....	5
2.1.1	Calendar Dates.....	5
2.1.2	Times of Day.....	5
2.1.3	Units of Weeks (week-based format).....	6
2.1.4	Duration	6
2.2	Fuzzy Temporal Expressions	7
2.2.1	Indeterminate Precision	7
2.2.2	Past, present, future	7
2.2.3	Seasons.....	8
2.2.4	Fiscal Year.....	8
2.2.5	Yearly quarters and halves	8
2.2.6	Weekends.....	8
2.2.7	Morning, Afternoon, and Night	8
2.2.8	Unspecified components of calendar dates and times of day.....	8
2.2.9	Unspecified numbers in durations.....	8
2.2.10	Combined week-based and month-based expressions	9
2.3	Modified Temporal Expressions	9
2.4	Set-denoting Temporal Expressions	9
2.5	Non-Specific Temporal Expressions	9
2.6	Event-Anchored Temporal Expressions	10
2.7	Culturally-Determined Expressions	10
3	Determin the extent of Annotations.....	10
3.1	Noun modifier	11
3.2	Range Expressions and Conjoined Expressions.....	11
3.3	Embedded Expressions	11

1 Determining what kinds of expressions to annotate

This section introduces you to the types of expressions that require annotations. We are interested in temporal expressions. Such expressions can reference calendar dates, times of day, 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 meanings.

1.1 Markable Expressions

To be markable, the syntactic head of the expression must be an appropriate lexical trigger. Each lexical trigger is a word or numeric expressions whose meaning conveys a temporal unit or concept, such as “day” or “monthly”. Furthermore, to be a trigger, the referent must be able to be oriented on a timeline, or at least oriented with relation to a time (past, present, future). The following list contains a sampling of lexical triggers. The table also includes examples of closely related temporal concepts that are not considered triggers.

Table 1.1 Sample Lexical Triggers and Non-Triggers

Part of Speech	Lexical Trigger	Non-Trigger
Noun	时间, 3点钟, 5号, 分, 分钟, 小时, 凌晨, 早晨, 上午, 中午, 下午, 午夜, 白天, 晚上, 傍晚, 当晚, 天, 今天, 明天, 周末, 日, 月, 月份, 上/中/下旬, 夏天, 季节, 季度, 年, 财年, 周年, 十年, 世纪, 百年, 千年, 时代, 学期, 那天, 那日, 今日, 未来, 暑期	即时, 任期, 光年, 有生之年, 历史, 年会, 周刊, 久, 当代, 现代, 近代, 之前, 之后,
Proper Name	(unique identifier for temporally-defined events): 除夕, 圣诞前夜, 国庆, 新年, 五一, 清明, 夏至, 立秋, 冬至	
Specialized time pattern	8:00, 12/2/00, 1994, 60年代, 民国2年, 嘉庆元年	
Adjectives	前(as in 前总统), 整天, 短期, 长期	早, 下, 上
Adverbs		同时, 有时, 然后, 现有
Time noun/adverb	将来, 过去, 时间, 时刻, 现在, 近来, 近日,	

	当前，最近，目前，如今，当今，以前，以往，当时，	
Localizer	期间	

1.2 Non-Markables

The non-triggers (i.e., the non-markables), although they are temporal in meaning, are as a class less amenable to being pinned down to a timeline.

Note that adjective non-triggers are permitted within the extent of a markable expression, as in “早些年”，“下个月”，“上个星期”，etc. They are not markable on their own, as in “上次会议”，“下次讨论”，etc.

Here is a list of words/phrases that are non-markables:

在，从，自，到，至，于，

就，才，还，然后，其次，后来，接着，其后，经常，常常，偶尔，时常，多次，几次，一般，通常，反复，有时，从来

Although 同时 often refers to simultaneous events, it is not markable when it is used idiomatically, for example to highlight inconsistency in actions or facts, or when functioning as a conjunction. It is markable when used as an anaphor, referring to an explicit time mentioned elsewhere in the text, and used like 当时.

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

十月革命
早春二月
家春秋三部曲
半夜鸡叫
辛亥革命
五一路

However, triggers that are functioning as temporal modifiers within titles are markable. Examples include titles of conferences, “2001 广州秋季交易会”，and awards, “2004 年诺贝尔化学奖”.

2 Capturing the meaning of temporal expressions

This section provides details on how to systematically encode the meaning behind each temporal expression.

2.1 Precise Temporal Expressions

A precise temporal expressions is one in which you can confidently determine the calendar date, time of day, or duration that is intended. When annotating naturally occurring text and speech, you use the context of the document to make that determination.

2.1.1 Calendar Dates

[1995 年 10 月], 中国长江三峡工程开发...

徐老主持了[今天]的会议。

[2 月 28 日], 两家公司签订了合同。

这块雨花石是他[5 年]前在南京郊外捡到的。

[公元 1924 年], 黄埔军校正式开课。

据介绍, 大连市[去年]外贸出口在种种不利条件下仍保持了持续增长的势头。[全年]出口额达 25.24 亿美元, 比[前年]增长了两成半以上。

[去年], 三资企业出口比[前年同期]增长了 15.7%。

会议于[四月十八日]至[十九日]在北京举行。

[一九九二年]至[一九九六年]上海国内生产总值达...

[三年]前的今天]是[一个不平常的日子]。

在[刚刚过去的 1999 年]继我国政府恢复对香港行使主权之后, ...

从[九十年代初], 广东省经济发展了。

中国已确定了[未来五年]高技术研究重点, 并着手制定[下个世纪]的研究计划。

在[新千年]来临之际, ...

广西考古研究发现[5000 万年]前的鱼化石。

[公元前 2100 年左右], 美索布达米亚人已有了乘法表。

2.1.2 Times of Day

[周五晚 19 时]—[23 时]

[6日晚八点零五分] 海府路发生车祸。

[当晚 10 时半]还有一次大碰撞

停火于[23 日中午当地时间 12 时]生效

巴总理谢里夫在[28 日上午五时十五分]发表的电视讲话中说...

[2003 年 08 月 11 时 04 分]

新华社北京[五月三十日]电 据中国地震台网测定，[北京时间今天十九时三十七分]，在北京丰台附近发生三点一级地震。[今天十九时四十二分]，在原地附近有发生一次三点一级地震。

伊拉克官员说，多国部队的飞机[13 日当地时间凌晨 4 点]投下的两枚导弹...

2.1.3 Units of Weeks (week-based format)

法国外长朱佩[上周]也表示...

他定于[下周]出访美国

2.1.4 Duration

工期缩短了[18 个月]。

碰撞[15 分钟]内，碰撞地区开始转向...

为期[15 年]的八六三计划

代表团将在华盛顿向国会和白宫展开为期[一周]的游说活动。

三资企业的审批速度已由[原来的一个月左右]缩短到[七天]。

夏威夷将是他[9 天]美国之行的最后一站。

就职仪式上，布什发表了[12 分钟]的就职演说。

法国酷热[三周]死亡 14000 人，希拉克被指责。

[过去五年]是[外商投资增长最快的五年]。

会谈中，双方对朝中两国作为朝鲜停战协定签字一方，在[过去四十多年]间为在朝鲜半岛保障和平的努力表示满意。

随着[苏哈托时代]的结束，印尼政治严冬已过去，实行[多年]的党禁也土崩瓦解。苏哈托倒台[才十天]，新的政党已纷纷诞生。

[五天]以后拿报告。

张学良[连续三个晚上]观看京剧表演。

[今年][26岁]的澜伟身手不凡。

到[今天]已经历了[十一个春秋]。

[几个星期]以来，古巴和也门等国要求安理会举行正式会议。

[三年]来， ...

历时[半月以上]的内战

2.2 Fuzzy Temporal Expressions

2.2.1 Indeterminate Precision

奥萨马[三周]前放话袭美

[两年]后每个人的薪水将重新审查。

[多年]前，他创作了一首歌曲

[一年]来，欧盟国家和中国有了[比以]往更频繁的高层互访。

[接下来的几个月]会有更激烈的反抗

2.2.2 Past, present, future

Also many temporal expressions refer in general terms to the past, the present, or the future.

[近日]，来自全国各地的四十多位科学家...

[过去]中国的对外开放以商品贸易为主

对[过去]的事件是否回顾一下历史就足够了呢？

中国[目前]最早创办的私人学校

[当前]一些国营企业有困难

考虑到[当前]形势，我们有必要...

[以前]他访问过中国

香港人民对香港的[未来]充满希望

[近期]目标是...

俄罗斯[不久]前参加了北约的和平伙伴计划

[现阶段]的工作主要集中在...

两岸同胞希望[早日]实现和平统一。
[现在]的空姐不如[从前]漂亮

2.2.3 Seasons

[今年]夏天]天气闷热。
据悉，温州市政府已做出部署，[今冬][明春]将在全市...
中央党校[[今年]秋季]新入校党员学员 825 人。

2.2.4 Fiscal Year

[八月二十三日]美国总统签署了[九五财年]海外活动拨款法
该公司在[本财年]难以达到预期盈利目的。

2.2.5 Yearly quarters and halves

第一季度
第二季度
第三季度
第四季度
上一季度
上半年
下半年
一个半年

2.2.6 Weekends

美国国务卿[这个周末]离开巴格达。
讲座在[下两个周末]举行。
很多人喜欢在[长周末]出游。

2.2.7 Morning, Afternoon, and Night

[今天早上][8 点钟]的会
总裁[今天早晨]离开纽约前往北京。
会谈[昨晚]结束

2.2.8 Unspecified components of calendar dates and times of day

[星期天]的报纸已经丢了。

他说莱温斯基在[[1995 年九月], [十月]或[十一月]的某个星期六的下午]在白
宫。

2.2.9 Unspecified numbers in durations

西班牙已经接受联合国的要求，准备在[近几天]内派出一架运输机...

[这几天]诞生的其他新当计有...

什叶派和逊尼派穆斯林之间的恩怨已经延续了[几个世纪]。

2.2.10 Combined week-based and month-based expressions

[[1998 年秋天], 一个星期五的下午], 他们两人出发去西藏。

2.3 Modified Temporal Expressions

Temporal expressions can be quantified or modified in some way. For example, “1999” is an unmodified expression, but “1999 年末” is modified. In general, quantifier modifiers such as 大约, 不超过, 将近 and post modifiers such as XX 前期, XX 中期, XX 末, XX 晚期 should be in the extent of the tagged temporal expressions.

会议定于[今年年底]召开。
[九十天以上]的合作
[两年多]来的两国经贸合作
从[现在]起用[三年多的时间]...
成立[不足一年]的网络公司纷纷倒闭
[十九世纪中期]
[本世纪中叶]

2.4 Set-denoting Temporal Expressions

Some temporal expressions indicate sets of times, such as 每天, 每个十月, 每年, 年年, 天天.

他们[每个星期二]的晚上看“百万富翁”。
这家人家[月月]都去扫墓。
该高速列车[每小时]运行 100 公里。
医生建议[每天]至少运动[一小时]。
[几乎每周]都有中国官员访问美国

2.5 Non-Specific Temporal Expressions

Some temporal expressions do not reference a specific time. These fall into two distinct categories:

- Generic: Lexical triggers used generically, as in “我喜欢[夏天]”, “这里[8 月份]最热”. Such expressions specify a whole class of temporal entities rather than referring to a specific time.
- Indefinite: Lexical triggers that are used referentially but are indefinite, as in “选举在[某个星期二]举行”.
- Idioms containing temporal expressions: Chinese is particularly rich in idiomatic expressions, some of which contain temporal expressions, as in “[春]风荡漾”, “小孩的脾气, [春天]的天”, “对同志像[春天]般温暖, 对敌人像[冬天]般冷酷”, “在[辞旧迎新]之际”

2.6 Event-Anchored Temporal Expressions

我记得[小平同志去世那天]...

[新年]五天]后，他退还了一些礼品。

很多俄罗斯人非常怀念[苏联是世界超级大国的岁月]。

[拉宾执政的年代]

[阿姆斯特朗登上月球 30 年]后，美国对太空的热情逐渐消退了。

他对[他登上领奖台的时刻]记忆犹新。

他[[五十岁生日]那一天]收到了一份特别的礼物

[他生日]是[这个月的 15 号]。

2.7 Culturally-Determined Expressions

There are other calendars in use throughout the world, such as the Chinese calendar or the Hindu calendar. Then there are those temporal units whose meanings are only understood with relation to a particular field of endeavor, such as academia (学期, 学年), sports (赛季), television (中央电视台的黄金时段), and political terms in office (任期). There are also expressions that are open to historical and individual interpretation such as “信息时代” and holidays, festivals and occasions of religious observance, remembrance of famous event. All those temporal units are markable.

全国人民欢庆[马年]的到来。

[千禧龙年]即将结束，大家在纷纷扰扰的情绪中送走了[不平静的庚辰年]。

[农历八月十五]是[全家团聚的日子]。

美国洛杉矶地区侨界[8 号]在华埠举行[中华民国 89 年双十国庆]大游行和升旗仪式。

杨基队在[常规赛季]遥遥领先。

爱国者队在[四个赛季]中三次获得超级碗杯。

[罗斯福的第一个任期]...

政府取缔了[选举期间]的一切游行示威活动。

他在[最后一个学年]不得不放弃钢琴，全身心投入高考复习。

大家身处[信息时代]...

这在[冷战时期]是不可想象。

他是[墨西哥黄金年代]的著名画家。

这件作品反映了[盛唐时期]的生活。

[911]可能是历史的转折点。

3 Determine the extent of Annotations

This section provides guidance on determining the textual boundaries and constituency of each annotated expressions. We need some rules for determining where each temporal expression begins and ends, i.e. the extent of the expression.

3.1 Noun modifier

The extent of a TIMEX2 expression is the lexical trigger and all its modifiers. In the case of a TIMEX2 that functions as a noun modifier, it may just be the lexical trigger itself, excluding even the attributive or genitive modifier marked by 的.

A trickier point is understanding and covering long extents where a TIMEX2 expression is the head of some long dependent phrase or clause. In Chinese, these dependencies are to the left of the lexical trigger.

[[四年]后的那个秋天]

[他离开我们的那一天]

[这]是[充满希望的一年]。

在[你生日]的[这一天], ...

华人以及其他非回教徒[最近]在[印尼暴乱期间]成为被攻击的对象, 包括[5月中]雅加达发生的[两天][狂暴行为期间]。

3.2 Range Expressions and Conjoined Expressions

If the text of a temporal range expression has explicit begin and end points, then it is considered a range expression, and the points are tagged separately.

他说, 印度于[11日]至[13日]进行了5次地下核试验。

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

系统将于[今天]和[明天]进行维护

萨达姆[从[现在]起半年]后或[一年]后还会故伎重演。

3.3 Embedded Expressions

One remaining question about extent is whether a text segment that contains more than one hierarchically-related or offset-related temporal element should be assigned one tag or two tags.

An expression is treated as an individual syntactic unit if

- There is no intervening element between temporal terms.
- The two terms express values for units that are hierarchically related.

[星期五晚上]

[1943年11月]

[18号星期二]

[98年夏天]

[1984年1月3日下午两点]

[三点差10分]

[早上8点]

[今年年初]

[那年秋天]

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

- 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.

[[三年]前的今天]
[[高三那年]的冬季]

- 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 later term is given one tag, and the pre-modifier is given another.

[[2004年六月]期间]
[他[九月二号]的生日]

In the following example, three TIMEX2 expressions occur in close proximity, but they are treated as three separate expressions. There is no embedding – the meaning of each one is resolved without reference to the others.

大会讨论了[一九九五年]至[一九九九年][五年]工作规划纲要。

3.3 localizers

Localizers in Chinese function both as preposition and as modifiers. We should be very careful when we decide whether or not to include the localizers in the extent of the timex2 mentions. Two localizers we need special attention are 前 and 后.

When 前 and 后 appear before a timex2 mention, it should be included in the extent.

When then appear after a timex2 mention, it should not be in the extent. Here are some examples:

[[三年]前的今天]
[[四年]后的那个秋天]
[[新年]五天]后，他退还了一些礼品。
萨达姆[从[现在]起半年]后或[一年]后还会故伎重演。
奥萨马[三周]前放话袭美
俄罗斯[不久]前参加了北约的和平伙伴计划
这块雨花石是他[5年]前在南京郊外捡到的。