ITALIAN CONTENT ANNOTATION BANK (I-CAB):

Temporal Expressions[1]

Bernardo Magnini, Matteo Negri*, Emanuele Pianta*,
Manuela Speranza*, Valentina Bartalesi Lenzi**, and Rachele Sprugnoli***
ITC-irst, Centro per la Ricerca Scientifica e Tecnologica
Povo 38050 (Trento) Italy
* {lavelli | magnini | negri | pianta | manspera}@itc.it

** CELCT, Trento 38100 Italy
{bartalesi | sprugnoli} @celct.it

February 2007

The content of the document is the result of the discussion which has taken place within the working group of ITC-irst composed by Luisa Bentivogli, Danilo Giampiccolo, Christian Girardi, Alberto Lavelli, Bernardo Magnini, Matteo Negri, Alessandra Novello, Emanuele Pianta, Lorenza Romano, Manuela Speranza, Rachele Sprugnoli, Alessandro Vallin and Roberto Zanoli.

TABLE OF CONTENTS

ABSTRACT

This document reports on Temporal Expression annotation for the Italian Content Annotation Bank (I-CAB) being developed at ITC-irst in conjunction with CELCT.
We describe the extensions to the English annotation guidelines that are required for Italian and provide a large number of examples and a detailed description of the benchmark.
The benchmark is intended as a reference work both for the annotation methodology and for the automatic recognition and normalization of temporal expressions in Italian.

1. INTRODUCTION TO THE TASK: TIMEX2

The TIMEX2 scheme has been developed by MITRE and SPAWAR[2] under the sponsorship of the DARPA[3] TIDES[4]. It covers both the detection and the normalization of time expressions in order to support the characterization of the ACE[5] program (Ferro et al. 2005)
A quick guide is reported below.

- **Markable expressions**: expressions that can reference to calendar dates, times of day, durations, sets of times, not-specific times and event-anchored expressions.
- **Tag element**: a special SGML (Standard Generalized Markup Language) tag is inserted around the text string: <TIMEX2> at the start of the expression and </TIMEX2> at the end.
- **Tag attributes**: five optional attributes can be contained in the TIMEX2 tag:

Table 1: TIMEX2 tag attributes

Attribute	Function	Example
VAL	Normalization of expressions based on the ISO 8601 standard	VAL="1964-10-16"
MOD	Normalization of temporal modifiers.	MOD="APPROX"
ANCHOR_VAL	Normalization of vague or partially anchored expressions in ISO format.	ANCHOR_VAL="1964-10-16"
ANCHOR_DIR	Captures the relative direction/orientation between VAL and ANCHOR_VAL.	ANCHOR_DIR="BEFORE"
SET	Identifies set-denoting expressions.	SET="YES"
COMMENT	Contains comments made by the annotator.	COMMENT="context garbled"

---

[2] Space and Naval Warfare Systems Command.
[3] Defense Advanced Research Projects Agency.
[4] Translingual Information Detection, Extraction, and Summarization.
[5] Automatic Content Extraction.

2. ITALIAN STANDARD FOR THE ANNOTATION OF TEMPORAL EXPRESSIONS

As it would be unwieldy to provide the full context for all examples, we assume, unless stated otherwise, that the reference date for all the examples in the remainder of this document is Friday, October 8, 2004, that is week 41. In some cases, though, the attributes are omitted.[6]

2.1 Determining what kinds of expressions to annotate

2.1.1 Markable expressions

Table 2.1: Sample lexical triggers and non-triggers

Part of Speech	Lexical Triggers	Non-Triggers
Noun	[il] futuro, [il] passato, [il] presente, alba, anno, autunno, decade, era, estate, fine settimana, giornata, giorno, inverno, mattina, mattinata, mese, mezzanotte, mezzogiorno, millennio, minuto, notte, periodo, pomeriggio, primavera, secolo, semestre, sera, settimana, stagione, vigilia, weekend	istante, episodio, occasione
Proper name	week days (lunedì, martedì etc. etc.), names of months (gennaio, febbraio, marzo etc. etc.)	Capodanno, Ferragosto, Natale, Pasqua, Vigilia di Natale
Specialized time patterns	8:00, 12/2/00, 1994	
Adjective	autunnale, estivo, invernale, primaverile, prossimo, recente, settimanale	frequente
Adverb	adesso, allora, annualmente, attualmente , finora, giornalmente, mensilmente, recentemente	ancora, dopo, frequentemente, già, immediatamente, mai, poi, presto, prima, sempre, tardi
Time noun/adverb	domani, dopodomani, ieri, oggi, ora	
Number	il primo (the first day of a month), Sessanta (referring to the decade of the Sixties)	

---

[6] We use the following notations:
- Examples are indented and are marked with a white bullet.
- The tags <TIMEX2>…</TIMEX2> are used to mark the end and the beginning of temporal expressions, except in tables.
- Other expressions are written in italics.

2.1.2 Non-markable expressions

- Sequencing and ordering expressions:
  - Si cerca, *nel frattempo*, di razionalizzare gli accessi.
- Manner adverbs, which say how soon or how quickly something is done:
  - L'uomo è stato soccorso *immediatamente*.
- Non-quantifiable durations:
  - Il servizio è *temporaneamente* bloccato.
- Frequency expressions with no time period given:
  - L'hanno incoronato per *due volte* campione del mondo.
- Proper names that designate something other than a temporal entity but happen to correspond to lexical triggers. for instance, temporal expressions in titles and names of organizations:
  - Al Palazzo delle Albere «*Il secolo* dell'Impero: principi, artisti e borghesi tra 1815 e 1915».

2.2 Determining the Extent of the Annotations

2.2.1 Syntactic Criteria

As in English, the full extent cannot be a prepositional phrase (i.e. the expression cannot start with a preposition) but it can start with an article:
  - in <TIMEX2>maggio</TIMEX2>
  - per <TIMEX2>la settimana prossima</TIMEX2>
  - durante <TIMEX2>il 2003</TIMEX2>

An exception, due to Italian morphology, is made with contracted prepositions: we want to incorporate the article but it is not possible to separate it from the preposition so we include everything in the tag:
  - <TIMEX2>dalla settimana prossima</TIMEX2>

Moreover, a temporal expression cannot start with a subordinating conjunction:
  - durante <TIMEX2>il 2003</TIMEX2>
  - nel corso <TIMEX2>della giornata</TIMEX2>

The full extent of the tagged expressions includes all premodifiers and postmodifiers of the time expression:
  - <TIMEX2>quella fredda giornata</TIMEX2>
  - <TIMEX2>più di un anno</TIMEX2>
  - <TIMEX2>non meno di un anno</TIMEX2>
  - <TIMEX2>il futuro dei nostri popoli</TIMEX2>
  - <TIMEX2>ieri l'altro</TIMEX2>

Exceptions:
  - un massimo di <TIMEX2>12 anni</TIMEX2>
  - appena <TIMEX2>3 mesi</TIMEX2>

Note that words like "circa,", "intorno", "almeno", "fa" and "tutto/tutta" function more as modifiers than prepositions, and thus should be included in the extent of the tag:
  - <TIMEX2>circa tre ore</TIMEX2>
  - <TIMEX2>intorno alle 11:00</TIMEX2>

- o <TIMEX2>almeno una settimana</TIMEX2>
- o <TIMEX2>tutto il mese</TIMEX2>
- o <TIMEX2>tutta la giornata</TIMEX2>
- o <TIMEX2>tre giorni fa </TIMEX2>

The word "dopo" can be either an adverb, a preposition or an adjective: it is included in the extent only in this last case.

- o dopo <TIMEX2>3 ore</TIMEX2>
- o <TIMEX2>3 ore</TIMEX2> dopo il concerto
- o <TIMEX2>3 ore dopo</TIMEX2> se ne andò

2.2.2 Multiwords

The expressions *per ora*, *fin d'ora* and *dopo domani* are considered as single units, and so the preposition is included:

- o la corsa <TIMEX2 VAL="PRESENT_REF" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="AS_OF" COMMENT="">per ora</TIMEX2> inarrestabile dei prezzi petroliferi è ormai al centro del dibattito.

Also *di recente* has to be tagged as a single unit because it is an adverbial phrase.

2.2.3 Appositives

In English, an appositive to a temporal expression is considered as a post- modifier and is thus included within the extent. If the appositive phrase itself contains a temporal expression, we have a nested tag.

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

In Italian we do not admit nested tags so we choose to annotate two separate expressions:

- o Sono una creatura <TIMEX2 VAL="196">degli anni '60 </TIMEX2>, <TIMEX2 VAL="196">gli anni dell'amore libero</TIMEX2>.

In this way we annotate two expressions as in English with the only difference being the extension of one of them.

2.2.4 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:

- o È stato approvato il bilancio <TIMEX2 VAL="2004">2004</TIMEX2>-<TIMEX2 VAL="2006">2006</TIMEX2>.

2.2.5 Conjoined Expressions

Two temporal expressions joined with a conjunction are tagged separately:

- o Le selezioni cominceranno <TIMEX2 VAL="2005">nel 2005</TIMEX2> o <TIMEX2 VAL="2006">nel 2006</TIMEX2>.

Please notice that conjoined expressions in times of day are treated in a different way because they are syntactically different:

- o Alle ore <TIMEX2 VAL="T12:20">dodici e venti</TIMEX2>.

2.2.6 Embedded Expressions

When to create one tag:

a)   Hierarchically-related expressions
   o   <TIMEX2>novembre 1943</TIMEX2>
   o   <TIMEX2>venerdì sera</TIMEX2>
   o   <TIMEX2>venerdì ore 20:00</TIMEX2>
   o   <TIMEX2>venerdì, ore 20:00</TIMEX2>
   o   <TIMEX2>venerdì (ore 20:00) </TIMEX2>
b)   Prepositions *di* and *del*
   o   <TIMEX2>ottobre del 1963</TIMEX2>
   o   <TIMEX2>estate del 1964</TIMEX2>
   o   <TIMEX2>l'estate di quest'anno</TIMEX2>
   o   <TIMEX2>fine del giorno</TIMEX2>
   o   <TIMEX2>alle 8:00 PM di venerdì</TIMEX2>
   o   <TIMEX2>la mattina del 15 Dicembre</TIMEX2>
   o   <TIMEX2>alle undici di mattina</TIMEX2>
c)   Prepositions and conjunctions in Times of Day expressions
   o   <timex2>dodici e venti</timex2>
   o   <timex2>mezzogiorno e mezza</timex2>
   o   <TIMEX2>dieci   minuti   alle   tre</TIMEX2>   or   <TIMEX2>tre   meno dieci</TIMEX2>
   o   <TIMEX2>cinque alle otto</TIMEX2>

When to create separate tags:
d)   Time-Anchored expressions
   o   <TIMEX2>Un anno</TIMEX2> da <TIMEX2>ora</TIMEX2>
e)   All the other cases
   o   Il concerto è <TIMEX2>venerdì</TIMEX2> <TIMEX2>alle otto di sera</TIMEX2>
   o   Il concerto è <TIMEX2>martedì</TIMEX2> dopo <TIMEX2>le 12 PM</TIMEX2>
   o   <TIMEX2>ieri</TIMEX2>, <TIMEX2>alle 11</TIMEX2>
   o   <TIMEX2>ieri mattina</TIMEX2> <TIMEX2>alle 11</TIMEX2>

2.2.7 Special extensions

- *ore e ore*
   o   <TIMEX2>ore e ore</TIMEX2> di bombardamenti.
- *giorno dopo giorno*
   o   <TIMEX2>giorno dopo giorno</TIMEX2> si arricchisce di informazioni.
- *24 ore su 24*
   o   A disposizione <TIMEX2>24 ore su 24</timex2>
- *da un giorno all'altro*
   o   Invece <TIMEX2>da un giorno all'altro</TIMEX2> al numero verde i richiedenti si sono sentiti dire che il bus non c'era più.
- *ordine del giorno*
   o   Molta routine all'ordine <TIMEX2>del giorno</TIMEX2> del consiglio federale.

2.3 Capturing the Meaning of Temporal Expressions

2.3.1 Precise Temporal Expressions
We have six types of precise temporal expressions:

a) **Calendar dates**
   - Una conferenza dei capigruppo di Montecitorio, convocata sempre per <TIMEX2 VAL="2004-10-08">oggi</TIMEX2>.
   - Il sanguinoso attentato è la risposta alla carneficina sciita di <TIMEX2 VAL="2004-10-01">sette giorni fa</TIMEX2>.

As in English, we consider the first position of the value attribute as the millennium component, the second as the century component, the third as the decade component, and the fourth as the year component. Thus we can have a value with just one number (*3* could represent *the next millennium*), two numbers (*20* would represent *the 21st century*), or three (*196* would represent *the 1960s*).

   - Ma <TIMEX2 VAL="199">negli anni '90</TIMEX2> era valida anche per l'italiano.

b) **Times of Day**
   - Seguiranno testimonianze e, <TIMEX2 VAL="2004-10-08T21:00">alle 21</TIMEX2>, proposte teatrali di Cooperative, Gruppi e Centri da tutta Italia.

c) **Week-Based format**
   - Le ho detto che notificheremo il piano <TIMEX2 VAL="2004-W42">la prossima settimana</TIMEX2>.

d) **Durations**

An expression of duration indicates a period of time that can be oriented or anchored with respect to other points or periods of time. The ANCHOR attributes, indeed, are used to capture a wide range of anchored durations.

   - Esordio <TIMEX2 VAL="2004-10-08">oggi</TIMEX2>, con <TIMEX2 VAL="P1D" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="WITHIN" comment="">un'intera giornata</TIMEX2> da vivere sul territorio.
   - Cerchiamo di divertirci <TIMEX2 VAL="P4Y" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="STARTING">nei prossimi 4 anni</TIMEX2>.
   - Aleen Bailey ha <TIMEX2 VAL="P24Y" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="ENDING">24 anni</TIMEX2>.

STARTING and ENDING are used when we have a precise anchor; BEFORE and AFTER values are used when the end or begin points of a duration cannot be anchored precisely, but it is clear from the context whether the span occurred before o after the reference time.

   - Dicevano che avrebbe vissuto altri <TIMEX2 VAL="P10Y" ANCHOR_VAL="2004-09-08" ANCHOR_DIR="AFTER">dieci anni</TIMEX2>.

Table 2.3.1: Anchor attributes

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

e) **Unspecified Components of Calendar Dates and Times of Day**
   We use the placeholder character, *X*, for each unfilled position in the value of a component
   - o Orario: <TIMEX2 VAL="XXXX-W02">martedì</TIMEX2>-<TIMEX2 VAL="XXXX-W05">venerdì</TIMEX2>.
f) **Unspecified Plural Numbers in Durations**
   The X character is also used in the representation of durations as an indicator of an unspecified plural number.
   - o Dopo <TIMEX2 VAL="PTXH">ore</TIMEX2> di lavoro, stanco o arrabbiato dall'ennesima lite con la moglie.

2.3.2 Fuzzy Temporal Expressions

1) Past, present and future
   - o Fra i film in concorso è quello che <TIMEX2 VAL="PRESENT_REF" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="AS_OF">finora</TIMEX2> preferiamo.
   - o Uno sguardo <TIMEX2 VAL="FUTURE_REF" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="AFTER">al futuro</TIMEX2>.
   - o La situazione è radicalmente diversa rispetto <TIMEX2 VAL="PAST_REF" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="BEFORE">al passato</TIMEX2>.

Table 2.3.2.1: Present, past, future tokens

Token	ANCHOR_DIR	Sample Markable Expressions	Sample Non-Markable Expressions
PRESENT_REF	AS_OF	ora, oggi, adesso [all with the "nowadays" meaning] (in) questo momento al momento (in) questi giorni tuttora per ora il/nel presente	immediatamente istantaneamente
FUTURE_REF	AFTER	al/il/nel futuro (in) futuro tra/fra qualche giorno/mese/anno domani [unless further resolvable to a particular date]	più tardi a breve
PAST_REF	BEFORE	recentemente (in) passato al/il/nel passato qualche giorno/mese/anno fa ieri [unless further resolvable to a particular date]	prima precedentemente

2) Seasons

The following example illustrates a point expression that refers to a season:

- o <TIMEX2 VAL="2004-SU">Questa estate</TIMEX2>, con il caldo infernale di Baghdad, in città non c'era acqua e luce.

In duration expressions the season tokens is used in the same manner as the ISO "PnW" format for periods of weeks and it can be combined with anchoring attributes.

Table 2.3.2.2: Season tokens

Token	Sample Expressions
SP	primavera
SU	estate
FA	autunno
WI	inverno

3) Weekends

- o La Coppa del Mondo di salto con gli sci <TIMEX2 VAL="2004-W40-WE">lo scorso weekend</TIMEX2> ha celebrato la terza prova in Polonia.

Table 2.3.2.3: Weekend token

Token	Sample Expressions
WE	weekend fine settimana

4) Part-of-day tokens

- o L'incontro con i sindacati, svoltosi <TIMEX2 VAL="2004-10-07TAF">ieri pomeriggio</TIMEX2>.

Table 2.3.2.4: Part-of-day tokens

Token	Position in VAL	Sample Expressions
MO	point: hour duration: PnMO	mattina, mattinata, alba
MI	point: hour duration: PnMI	mezzogiorno, mezzodì
AF	point: hour duration: PnAF	pomeriggio
EV	point: hour duration: PnEV	sera, serata
NI	point: hour duration: PnNI	notte, nottata
DT	point: hour duration: PnDT	mattina+pomeriggio (basically, daytime or working hours)

2.3.3 Modified Temporal Expressions

Table 2.3.3: Modifier tokens

	MOD value	Sample Expressions
POINTS	BEFORE	più di
	AFTER	meno di
	ON_OR_BEFORE	non meno di
	ON_OR_AFTER	non più di
DURATIONS	LESS_THAN	meno di
	MORE_THAN	più di oltre
	EQUAL_OR_LESS	non più di al massimo
	EQUAL_OR_MORE	almeno
POINTS AND DURATIONS	START	all'inizio di/del nei primi
	MID	metà
	END	alla fine di/del nel tardo
	APPROX	intorno circa

- o quando mancano <TIMEX2 VAL="P4M" MOD="LESS_THAN">poco meno di 4 mesi</TIMEX2>.
- o era stato per <TIMEX2 VAL="P6M" MOD="MORE_THAN">oltre sei mesi</TIMEX2> in Iraq.
- o Per arrivare a Bari impiego <TIMEX2 VAL="PT2H" MOD="EQUAL_OR_LESS">al massimo due ore</TIMEX2>.
- o la popolazione sotto <TIMEX2 VAL="P14Y" MOD="EQUAL_OR_LESS" ANCHOR_VAL="2004-12-08" ANCHOR_DIR="ENDING">i 14 anni</TIMEX2> e sopra <TIMEX2 VAL="P70Y" MOD="EQUAL_OR_MORE" ANCHOR_VAL="2004-12-08" ANCHOR_DIR="ENDING">i 70</TIMEX2>.
- o <TIMEX2 VAL="2004-09" MOD="START">all'inizio di settembre</TIMEX2> ha dovuto giocoforza riparare qualche foratura imprevista.
- o <TIMEX2 VAL="199" MOD="START">dai primi anni novanta</TIMEX2> la cooperativa ha iniziato ad accumulare debiti.
- o L'impasse ha portato, <TIMEX2 VAL="2004-07" MOD="MID">alla metà dello scorso luglio</TIMEX2>, all'interruzione delle trattative.
- o molto attivo <TIMEX2 VAL="2004-SU" MOD="END">in questa fine estate</TIMEX2>.
- o Questa è la decisione adottata <TIMEX2 VAL="2003-10-08" MOD="APPROX">circa un anno fa</TIMEX2>.

APPROX is also used with expressions like *una decina*, *un paio*:
- o Mi sono laureata <TIMEX2 VAL="1994" MOD="APPROX">una decina di anni fa</TIMEX2>.

o Passare nel giro di <TIMEX2 VAL="P2M" MOD="APPROX">un paio di mesi</TIMEX2> dal campo alla panchina.

2.3.4 Set-Denoting Temporal Expressions

Set-denoting expressions tell how often something happened. The attribute SET is placed to YES in all these expressions. If the expression does not denote a set, the attribute is omitted from the annotation altogether.

- o Andranno in onda <TIMEX2 VAL="XXXX-WXX-5" SET="YES">tutti i venerdì</TIMEX2>.
- o <TIMEX2 VAL="P1M" SET="YES">Ogni mese</TIMEX2> viene trattata una voce.
- o Un abbonamento che costa 52 euro <TIMEX2 VAL="XXXX" SET="YES">all'anno</TIMEX2>.
- o Nel 2° turno <TIMEX2 VAL="XXXX-WXX-X" SET="YES">infrasettimanale</TIMEX2> di A1.

Frequency expressions formed with the word *volta*, or its plural *volte*, preceded by a numeral adjective, are all included in the extent of the tag but their semantics are ignored. Only the granularity of the head is captured:

- o Variano <TIMEX2 VAL="XXXX-WXX" SET="YES">due volte a settimana</TIMEX2>.

2.3.5 Culturally-Determined Expressions

The interpretation of some temporal expressions requires cultural or domain-specific knowledge. As in English, these expressions are markable but receive no VAL.
This is the case of holiday names like *Capodanno, Ferragosto*, *Natale*, *Pasqua*, and *Vigilia di Natale*:

- o L'azienda aprirà entro <TIMEX2>Natale</TIMEX2>.

Expressions that refer to calendars different from the Gregorian one are tagged when they have a lexical trigger as their head (like *season* in *baseball season*), but they receive no value.

- o Con l'approssimarsi dell'apertura <TIMEX2>dell'anno scolastico</TIMEX2> il problema sta diventando sempre più scottante.
- o Sorprendono invece parecchio le due sconfitte di Pieve di Bono e Pinzolo Campiglio, entrambe retrocesse dalla Promozione <TIMEX2>nell'ultima stagione calcistica</TIMEX2>.

2.4 Specific Issues

a) We have interpreted the event's property expressed by the verb as not having influence on temporal expressions such as *nei mesi scorsi*, *nei prossimi giorni*. As a consequence, all the plural non quantified temporal expressions containing the adjectives *prossimo*, *scorso*, *futuro*, *passato*, etc., have been interpreted as periods, independently of the verb contained in the sentence.

- o Il presidente ha intenzione di fare <TIMEX2 VAL="PXW" ANCHOR_VAL="2004-W41" ANCHOR_DIR="AFTER">nelle prossime settimane</TIMEX2> un'analisi approfondita della situazione.

This choice, which may appear questionable in some cases (see the following example), was mainly motivated by the necessity of guaranteeing an acceptable level of inter annotator agreement.

- o Il disegno di legge presentato <TIMEX2 VAL="PXD" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="BEFORE">nei giorni scorsi</TIMEX2> da Roberto Bombarda.

b) If the word *periodo* (or its plural *periodi*) refers to a temporal expression already definite in the text, it must be tagged with the same annotation used previously:

- o Era <TIMEX2 VAL="2002-08">l'agosto 2002</TIMEX2>. In <TIMEX2 VAL="2002-08">quel periodo</TIMEX2> faceva molto caldo.

c) *Mensile*, *annuale*, *settimanale* etc.etc. are points or durations according to the context but, as a general rule, we can say that if both a point and a duration interpretation is possible, the point representation is preferred over the duration representation.

- o Il bilancio <TIMEX2 SET="YES" VAL="XXXX">annuale</TIMEX2> è stato positivo.

d) *autunnale*, *estivo*, *primaverile*, *invernale*, *recente*, *ex* are always to be tagged, even if they do not occur with a lexical trigger:

- o Entra in vigore l'orario <TIMEX2 VAL="XXXX-WI">invernale</TIMEX2> per il servizio extraurbano.
- o Andreatta dichiara di voler comprare l'<TIMEX2 ANCHOR_DIR="BEFORE" ANCHOR_VAL="2004-10-08" VAL="PAST_REF">ex</TIMEX2> azienda .
- o La risonanza magnetica - ha aggiunto il medico - ha dimostrato la <TIMEX2 ANCHOR_DIR="BEFORE" ANCHOR_VAL="2004-10-07" VAL="PAST_REF">recente</TIMEX2> lussazione subita dal giocatore.

On the contrary, in the same position, *prossimo*, *scorso*, *ultimo*, *successivo* and *precedente* are not to be tagged:

- o Una differenza sostanziale rispetto allo studio *precedente*, affidato a professori della Cattolica di Milano.
- o Sono stati destinati tutti i 3,9 milioni di euro dell'utile che la holding ha realizzato nel corso dell'esercizio <TIMEX2 VAL="2003">2003</TIMEX2>. Un livello ben lontano dai 24,2 milioni di euro incassati <TIMEX2 VAL="2002">l'anno precedente</TIMEX2>.

e) *serata* and its plural *serate* have to be tagged also if they have the meaning of "evening performance/party":

- o Non capisco perchè un cantante viene a Trento e chiede 30.000 euro per <TIMEX2 VAL="TEV">una serata</TIMEX2>.

f) The expressions *da <anni>* and *per <anni>* are always annotated as periods (VAL="PX..."). If they have a general meaning, the only attribute to use is VAL. If those expressions refer to something happened for a long time but already finished, we use: VAL="PXY" ANCHOR_VAL="specific_year" ANCHOR_DIR="ENDING". Otherwise, if they refer to something which end cannot be anchored precisely, we annotate as follows: VAL="PXY" ANCHOR_VAL="specific_year" ANCHOR_DIR="BEFORE".

- o Aspetta inutilmente notizie del marito scomparso da <TIMEX2 VAL="PXY">anni </TIMEX2>.
- o Nato a Milano ma da <TIMEX2 ANCHOR_DIR="BEFORE" ANCHOR_VAL="2004" VAL="PXY">anni</TIMEX2> a Trento senza fissa dimora.

- o Villa Igea darà alle vittime della guerra che insanguinava da <TIMEX2 ANCHOR_DIR="ENDING" ANCHOR_VAL="2004" VAL="PXY">anni</TIMEX2> il Corno d'Africa la possibilità di studiare.
g) The trigger *tempo* (and its plural *tempi*) is one of the more difficult words in terms of markability. In many cases it is not referring to a time interval or it is part of an adverbial or adjectival phrase:
    - o Dipendenti con contratto a *tempo indeterminato* che per continuare il proprio lavoro dovrebbero andare in aspettativa.
    - o Tosi si dice favorevole al principio per cui chi trae beneficio dal turismo deve pagare, ma *allo stesso tempo* è contrario se diventa una tassa per tutti.
    - o *Il primo tempo* della partita della Juventus è stato veramente noioso.
    - o Sono arrivata in sala quando era già iniziato *il secondo tempo* del film.

APPENDIX A: Inter-annotator Agreement

Inter-annotator agreement has been evaluated on the dual annotation of a corpus of ten randomly chosen news stories, for a total of about 5,204 words.

The most commonly used measure to characterize inter-annotator agreement is the kappa statistic (Cohen 1960), which measures pair wise agreement among a set of coders making category judgments taking into consideration agreement obtained by chance. In the case of Temporal Expressions (TEs), however, annotators can theoretically choose to tag any sequence of adjacent tokens in a sentence. This makes it necessary to consider every possible sequence as a candidate, which would give exceedingly low results in terms of kappa, as annotation would become a binary categorization problem with an extremely skewed distribution (only a minimum number of the candidate sequences, in fact, are TEs).

For this reason we have used the kappa statistic to simply measure the agreement in determining whether each token is or is not part of any TE, and we have obtained k=0.958. However, this measure does not take into account the extent of the annotated TEs, so we have also compared the two annotated versions using the Dice coefficient. The Dice coefficient is computed as in [1], where C is the number of common annotations, while A and B are respectively the number of annotations provided by the first and the second annotator.

$$[1] \; Dice=2C/(A+B)^7$$

The Dice coefficient is 0.955 for TE detection and 0.931 for TE bracketing. Agreement in normalization has been measured on the TEs uniformly bracketed. Table 5 reports, for each attribute, the cases where the two annotators agreed in assigning or not assigning a value for the attribute and, for the attributes which admit a restricted number of values, it also reports the kappa statistic[8].

Table 5: Agreement in attribute value assignment

	agreement	kappa statistic
VAL	92.2% (142/154)	-
ANCH._VAL	92.2% (142/154)	-
ANCH._DIR	90.3% (139/154)	0.749
MOD	99.3% (153/154)	0.886
SET	98.7% (152/154)	0.744

---

[7] Notice that the Dice coefficient has the same value of the F_1 measure computed considering any of the two annotators as the reference.

[8] As observed in (Di Eugenio, Glass 2004) the kappa statistic could be affected by bias and prevalence problems. By also calculating kappa according to the (Siegel, Castellan 1988) definition we verified there are no bias problems (values are equal), but the natural skewing of the distribution of categories does affect kappa (e.g. for the SET attribute).

APPENDIX B: Training Data

Number of files = 335
Number of words = 113634
Average number of words per file = 339
Number of TIMEX2 tags = 2931
Number of words within TIMEX2 tags = 5555 [4.89% of the total]
Average length of temporal expressions = 1.9
Number of characters within TIMEX2 tags = 28442
Average number of characters per temporal expression = 9.8

Statistics on TIMEX2 attributes used in the Corpus

VAL attribute

# VAL="PRESENT_REF"	**242**/2931 (**8.26%**)
# VAL= "FUTURE_REF"	**52**/2931 (**1.77%**)
# VAL= "PAST_REF"	**131**/2931 (**4.47%**)
# No VAL	**143**/2931 (**4.88%**)

MOD attribute

VALUE FOR MOD	# OCCURRENCES
"AFTER"	**2**/2931 (**0.07 %**)
"APPROX"	**31**/2931 (**1.06%**)
"BEFORE"	**-**
"END"	**31**/2931 (**1.06%**)
"MID"	**8**/2931 (**0.27%**)
"START"	**16**/2931 (**0.54%**)
"ON_OR_AFTER"	-
"ON_OR_BEFORE"	-
"EQUAL_OR_LESS"	**10**/2931 (**0.34%**)
"EQUAL_OR_MORE"	**8**/2931 (**0.27%**)
"LESS_THAN"	**1**/2931 (**0.03%**)
"MORE_THAN"	**12**/2931 (**0.41%**)
TOTAL	**119**/2931 (**4.06%**)

ANCHOR_DIR attribute

VALUE FOR *ANCHOR_DIR*	# OCCURRENCES
"AS_OF"[9]	**243**/2931 (**8.29%**)
"AFTER"	**72**/2931 (**2.46%**)
"BEFORE"	**164**/2931 (**5.59%**)
"ENDING"	**160**/2931 (**5.46%**)
"STARTING"	**43**/2931 (**1.47%**)
"WITHIN"	**24**/2931 (**0.82%**)
TOTAL	**706/**2931 (**24.00%**)

---

[9] The value "AS_OF" is present only in combination with VAL="PRESENT_REF".

APPENDIX C: The ISO 8601 standard

ISO 8601:2004 (*Data elements and interchange formats – Information interchange – Representation of dates and times*) is the standardized format for all dates, times, and periods of time requiring to be represented.

• Calendar dates:

$$YYYY-MM-DD$$

YYYY is the year in the Gregorian calendar, MM is the month of the year between 01 (January) and 12 (December), and DD is the day of the month between 01 and 31.

• Week dates:

$$YYYY-Www-D$$

YYYY is the year in the Gregorian calendar, ww is the week of the year between 01 (the first week) and 52 or 53[10] (the last week), and D is the day in the week between 1 (Monday) and 7 (Sunday).

• Time of day using the 24-hour timekeeping system:

$$hh:mm:ss$$

hh are the hours passed since midnight, mm are the minutes since the start of the hour, and ss are the seconds since the start of the minute.

• A specified time of a specified day:

$$YYYY-MM-DDThh:mm:ss$$

the capital letter T separates the date and time components.

• Durations:

$$PnYnMnDTnHnMnS$$
$$PnW \text{ (for the week format)}$$

---

[10] If it is a leap year.

APPENDIX D: Training text files divided by date and category

20040907
ATTUALITÀ - NEWS STORIES
1.   adige20040907_id405381.txt
2.   adige20040907_id405382.txt
3.   adige20040907_id405383.txt
4.   adige20040907_id405384.txt
5.   adige20040907_id405385.txt
6.   adige20040907_id405386.txt
7.   adige20040907_id405388.txt
8.   adige20040907_id405390.txt
9.   adige20040907_id405392.txt
10.   adige20040907_id405394.txt
11.   adige20040907_id405395.txt
12.   adige20040907_id405396.txt
13.   adige20040907_id405397.txt
14.   adige20040907_id405398.txt
15.   adige20040907_id405399.txt

CULTURA - CULTURAL NEWS
1.   adige20040907_id405408.txt
2.   adige20040907_id405409.txt
3.   adige20040907_id405410.txt
4.   adige20040907_id405411.txt
5.   adige20040907_id405412.txt
6.   adige20040907_id405414.txt
7.   adige20040907_id405417.txt
8.   adige20040907_id405418.txt
9.   adige20040907_id405419.txt
10.   adige20040907_id405420.txt
11.   adige20040907_id405424.txt
12.   adige20040907_id405425.txt

ECONOMIA - ECONOMY NEWS
1.   adige20040907_id405436.txt
2.   adige20040907_id405437.txt
3.   adige20040907_id405438.txt
4.   adige20040907_id405442.txt
5.   adige20040907_id405444.txt
6.   adige20040907_id405446.txt
7.   adige20040907_id405447.txt
8.   adige20040907_id405448.txt

SPORT - SPORTS NEWS
1.   adige20040907_id405581.txt
2.   adige20040907_id405582.txt
3.   adige20040907_id405583.txt
4.   adige20040907_id405585.txt
5.   adige20040907_id405586.txt
6.   adige20040907_id405588.txt
7.   adige20040907_id405589.txt
8.   adige20040907_id405590.txt
9.   adige20040907_id405591.txt
10.   adige20040907_id405592.txt
11.   adige20040907_id405593.txt
12.   adige20040907_id405594.txt
13.   adige20040907_id405595.txt
14.   adige20040907_id405596.txt
15.   adige20040907_id405597.txt
16.   adige20040907_id405598.txt
17.   adige20040907_id405599.txt
18.   adige20040907_id405600.txt
19.   adige20040907_id405601.txt

TRENTO - LOCAL NEWS
1.   adige20040907_id405602.txt
2.   adige20040907_id405603.txt
3.   adige20040907_id405604.txt
4.   adige20040907_id405605.txt
5.   adige20040907_id405606.txt
6.   adige20040907_id405607.txt
7.   adige20040907_id405608.txt
8.   adige20040907_id405609.txt
9.   adige20040907_id405610.txt
10.   adige20040907_id405611.txt
11.   adige20040907_id405612.txt
12.   adige20040907_id405613.txt
13.   adige20040907_id405615.txt
14.   adige20040907_id405616.txt
15.   adige20040907_id405617.txt
16.   adige20040907_id405618.txt
17.   adige20040907_id405619.txt
18.   adige20040907_id405620.txt
19.   adige20040907_id405621.txt
20.   adige20040907_id405622.txt
21.   adige20040907_id405623.txt

22.	adige20040907_id405624.txt	4.	adige20040908_id405707.txt
23.	adige20040907_id405625.txt	5.	adige20040908_id405708.txt
24.	adige20040907_id405626.txt	6.	adige20040908_id405710.txt
25.	adige20040907_id405627.txt	7.	adige20040908_id405711.txt
26.	adige20040907_id405630.txt	8.	adige20040908_id405712.txt
27.	adige20040907_id405631.txt	9.	adige20040908_id405713.txt
28.	adige20040907_id405632.txt	10.	adige20040908_id405714.txt
29.	adige20040907_id405633.txt		
30.	adige20040907_id405634.txt		
31.	adige20040907_id405635.txt		

SPORT - SPORTS NEWS

20040908

ATTUALITÀ - NEWS STORIES

1.	adige20040908_id405656.txt	1.	adige20040908_id405848.txt
2.	adige20040908_id405657.txt	2.	adige20040908_id405849.txt
3.	adige20040908_id405658.txt	3.	adige20040908_id405850.txt
4.	adige20040908_id405659.txt	4.	adige20040908_id405851.txt
5.	adige20040908_id405660.txt	5.	adige20040908_id405852.txt
6.	adige20040908_id405661.txt	6.	adige20040908_id405853.txt
7.	adige20040908_id405662.txt	7.	adige20040908_id405854.txt
8.	adige20040908_id405663.txt	8.	adige20040908_id405855.txt
9.	adige20040908_id405664.txt	9.	adige20040908_id405856.txt
10.	adige20040908_id405666.txt	10.	adige20040908_id405857.txt
11.	adige20040908_id405667.txt	11.	adige20040908_id405858.txt
12.	adige20040908_id405669.txt	12.	adige20040908_id405859.txt
13.	adige20040908_id405670.txt	13.	adige20040908_id405860.txt
14.	adige20040908_id405671.txt	14.	adige20040908_id405861.txt
15.	adige20040908_id405672.txt	15.	adige20040908_id405862.txt
16.	adige20040908_id405673.txt	16.	adige20040908_id405863.txt
		17.	adige20040908_id405864.txt
		18.	adige20040908_id405865.txt
		19.	adige20040908_id405866.txt

CULTURA - CULTURAL NEWS

1.	adige20040908_id405684.txt	20.	adige20040908_id405867.txt
2.	adige20040908_id405685.txt	21.	adige20040908_id405868.txt
3.	adige20040908_id405686.txt	22.	adige20040908_id405871.txt
4.	adige20040908_id405687.txt	23.	adige20040908_id405874.txt
5.	adige20040908_id405688.txt	24.	adige20040908_id405875.txt
6.	adige20040908_id405689.txt	25.	adige20040908_id405877.txt
7.	adige20040908_id405690.txt	26.	adige20040908_id405878.txt
8.	adige20040908_id405691.txt	27.	adige20040908_id405880.txt
9.	adige20040908_id405692.txt		
10.	adige20040908_id405693.txt		

TRENTO - LOCAL NEWS

ECONOMIA - ECONOMY NEWS

1.	adige20040908_id405704.txt	1.	adige20040908_id405881.txt
2.	adige20040908_id405705.txt	2.	adige20040908_id405882.txt
3.	adige20040908_id405706.txt	3.	adige20040908_id405883.txt
		4.	adige20040908_id405884.txt
		5.	adige20040908_id405885.txt
		6.	adige20040908_id405887.txt
		7.	adige20040908_id405888.txt

8.	adige20040908_id405889.txt		9.	adige20041007_id413727.txt
9.	adige20040908_id405890.txt		10.	adige20041007_id413728.txt
10.	adige20040908_id405894.txt			
11.	adige20040908_id405896.txt		ECONOMIA - ECONOMY NEWS	
12.	adige20040908_id405897.txt		1.	adige20041007_id413743.txt
13.	adige20040908_id405898.txt		2.	adige20041007_id413744.txt
14.	adige20040908_id405900.txt		3.	adige20041007_id413745.txt
15.	adige20040908_id405901.txt		4.	adige20041007_id413746.txt
16.	adige20040908_id405902.txt		5.	adige20041007_id413747.txt
17.	adige20040908_id405903.txt		6.	adige20041007_id413748.txt
18.	adige20040908_id405904.txt		7.	adige20041007_id413750.txt
19.	adige20040908_id405906.txt		8.	adige20041007_id413751.txt
20.	adige20040908_id405907.txt			
21.	adige20040908_id405908.txt		SPORT - SPORTS NEWS	
22.	adige20040908_id405910.txt		1.	adige20041007_id413887.txt
23.	adige20040908_id405911.txt		2.	adige20041007_id413888.txt
24.	adige20040908_id405914.txt		3.	adige20041007_id413890.txt
25.	adige20040908_id405915.txt		4.	adige20041007_id413891.txt
26.	adige20040908_id405916.txt		5.	adige20041007_id413892.txt
27.	adige20040908_id405917.txt		6.	adige20041007_id413893.txt
28.	adige20040908_id405918.txt		7.	adige20041007_id413894.txt
			8.	adige20041007_id413895.txt
			9.	adige20041007_id413897.txt
20041007			10.	adige20041007_id413898.txt
ATTUALITÀ - NEWS STORIES			11.	adige20041007_id413899.txt
1.	adige20041007_id413699.txt		12.	adige20041007_id413900.txt
2.	adige20041007_id413700.txt		13.	adige20041007_id413901.txt
3.	adige20041007_id413701.txt		14.	adige20041007_id413902.txt
4.	adige20041007_id413702.txt		15.	adige20041007_id413903.txt
5.	adige20041007_id413703.txt		16.	adige20041007_id413904.txt
6.	adige20041007_id413704.txt		17.	adige20041007_id413905.txt
7.	adige20041007_id413705.txt		18.	adige20041007_id413906.txt
8.	adige20041007_id413706.txt			
9.	adige20041007_id413708.txt		TRENTO - LOCAL NEWS	
10.	adige20041007_id413709.txt		1.	adige20041007_id413916.txt
11.	adige20041007_id413710.txt		2.	adige20041007_id413917.txt
12.	adige20041007_id413711.txt		3.	adige20041007_id413918.txt
			4.	adige20041007_id413919.txt
			5.	adige20041007_id413920.txt
CULTURA - CULTURAL NEWS			6.	adige20041007_id413921.txt
1.	adige20041007_id413719.txt		7.	adige20041007_id413922.txt
2.	adige20041007_id413720.txt		8.	adige20041007_id413923.txt
3.	adige20041007_id413721.txt		9.	adige20041007_id413924.txt
4.	adige20041007_id413722.txt		10.	adige20041007_id413925.txt
5.	adige20041007_id413723.txt		11.	adige20041007_id413926.txt
6.	adige20041007_id413724.txt		12.	adige20041007_id413927.txt
7.	adige20041007_id413725.txt			
8.	adige20041007_id413726.txt			

13.	adige20041007_id413928.txt	10.	adige20041008_id414004.txt
14.	adige20041007_id413929.txt	11.	adige20041008_id414005.txt
15.	adige20041007_id413930.txt	12.	adige20041008_id414007.txt
16.	adige20041007_id413931.txt		
17.	adige20041007_id413932.txt	**ECONOMIA - ECONOMY NEWS**	
18.	adige20041007_id413933.txt	1.	adige20041008_id414017.txt
19.	adige20041007_id413934.txt	2.	adige20041008_id414018.txt
20.	adige20041007_id413935.txt	3.	adige20041008_id414019.txt
21.	adige20041007_id413936.txt	4.	adige20041008_id414020.txt
22.	adige20041007_id413937.txt	5.	adige20041008_id414021.txt
23.	adige20041007_id413938.txt	6.	adige20041008_id414022.txt
24.	adige20041007_id413939.txt	7.	adige20041008_id414023.txt
25.	adige20041007_id413940.txt	8.	adige20041008_id414024.txt
26.	adige20041007_id413941.txt	9.	adige20041008_id414025.txt
27.	adige20041007_id413942.txt		
28.	adige20041007_id413943.txt	**SPORT - SPORTS NEWS**	
29.	adige20041007_id413944.txt	1.	adige20041008_id414146.txt
30.	adige20041007_id413945.txt	2.	adige20041008_id414147.txt
		3.	adige20041008_id414148.txt
20041008		4.	adige20041008_id414149.txt
ATTUALITÀ - NEWS STORIES		5.	adige20041008_id414150.txt
1.	adige20041008_id413973.txt	6.	adige20041008_id414151.txt
2.	adige20041008_id413974.txt	7.	adige20041008_id414152.txt
3.	adige20041008_id413975.txt	8.	adige20041008_id414153.txt
4.	adige20041008_id413976.txt	9.	adige20041008_id414154.txt
5.	adige20041008_id413977.txt	10.	adige20041008_id414155.txt
6.	adige20041008_id413978.txt	11.	adige20041008_id414156.txt
7.	adige20041008_id413979.txt	12.	adige20041008_id414157.txt
8.	adige20041008_id413980.txt	13.	adige20041008_id414158.txt
9.	adige20041008_id413981.txt	14.	adige20041008_id414159.txt
10.	adige20041008_id413982.txt	15.	adige20041008_id414160.txt
11.	adige20041008_id413984.txt	16.	adige20041008_id414163.txt
12.	adige20041008_id413985.txt		
13.	adige20041008_id413986.txt	**TRENTO - LOCAL NEWS**	
14.	adige20041008_id413987.txt	1.	adige20041008_id414176.txt
		2.	adige20041008_id414177.txt
CULTURA - CULTURAL NEWS		3.	adige20041008_id414178.txt
1.	adige20041008_id413995.txt	4.	adige20041008_id414179.txt
2.	adige20041008_id413996.txt	5.	adige20041008_id414180.txt
3.	adige20041008_id413997.txt	6.	adige20041008_id414181.txt
4.	adige20041008_id413998.txt	7.	adige20041008_id414182.txt
5.	adige20041008_id413999.txt	8.	adige20041008_id414183.txt
6.	adige20041008_id414000.txt	9.	adige20041008_id414184.txt
7.	adige20041008_id414001.txt	10.	adige20041008_id414185.txt
8.	adige20041008_id414002.txt	11.	adige20041008_id414186.txt
9.	adige20041008_id414003.txt	12.	adige20041008_id414187.txt

13.	adige20041008_id414188.txt	23.	adige20041008_id414198.txt
14.	adige20041008_id414189.txt	24.	adige20041008_id414199.txt
15.	adige20041008_id414190.txt	25.	adige20041008_id414206.txt
16.	adige20041008_id414191.txt	26.	adige20041008_id414207.txt
17.	adige20041008_id414192.txt	27.	adige20041008_id414208.txt
18.	adige20041008_id414193.txt	28.	adige20041008_id414209.txt
19.	adige20041008_id414194.txt	29.	adige20041008_id414210.txt
20.	adige20041008_id414195.txt	30.	adige20041008_id414211.txt
21.	adige20041008_id414196.txt		
22.	adige20041008_id414197.txt	**TOTAL: 335**	

APPENDIX E: Index of the Temporal Expressions occurring in this document

REFERENCES

- ACE: http://www.nist.gov/speech/tests/ace/index.htm and http://www.ldc.upenn.edu/Projects/ACE/
- DARPA: http://www.darpa.mil/
- Ferro, L., Gerber, L., Mani, I., Sundheim, B. and Wilson G., *TIDES 2005 Standard for the Annotation of Temporal Expressions* April 2005, Updated September 2005.
- http://timex2.mitre.org/annotation_guidelines/2005_timex2_standard_v1.1.pdf
- MITRE: http://www.mitre.org/
- TIDES: http://www.darpa.mil/ipto/programs/tides/index.htm
- TIMEX2 Quick Guide: http://timex2.mitre.org/annotation_guidelines/timex2-quick-guide-2003.pdf