

ITALIAN CONTENT ANNOTATION BANK (I-CAB):
TEMPORAL EXPRESSIONS (V. 1.0)¹

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

Maggio 2005

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 Zanolì.

¹ This work has been supported by the ONTOTEXT (From Text to Knowledge for the Semantic Web) project, funded by the Autonomous Province of Trento under the FUP-2004 research program, and by the WebFaq (Flexible Access and Quality on the Web) project funded by the Autonomous Province of Trento under the FUP-2002 research program.

TABLE OF CONTENTS

ABSTRACT	2
1. INTRODUCTION TO THE TASK: TIMEX2.....	3
2. ITALIAN STANDARD FOR THE ANNOTATION OF TEMPORAL EXPRESSIONS.....	4
2.1 DETERMINING WHAT KINDS OF EXPRESSIONS TO ANNOTATE	4
2.1.1 <i>Markable expressions</i>	4
2.1.2 <i>Non-markable expressions</i>	5
2.2 DETERMINING THE EXTENT OF THE ANNOTATIONS	5
2.2.1 <i>Lexical Criteria</i>	5
2.2.2 <i>Syntactic Criteria</i>	6
2.2.3 <i>Multiwords</i>	6
2.2.4 <i>Appositives</i>	6
2.2.5 <i>Range Expressions</i>	6
2.2.6 <i>Conjoined Expressions</i>	6
2.2.7 <i>Embedded Expressions</i>	6
2.3 CAPTURING THE MEANING OF TEMPORAL EXPRESSIONS	7
2.3.1 <i>Precise Temporal Expressions</i>	7
2.3.2 <i>Fuzzy Temporal Expressions</i>	8
2.3.3 <i>Modified Temporal Expressions</i>	10
2.3.4 <i>Set-Denoting Temporal Expressions</i>	11
2.3.5 <i>Non-Specific Temporal Expressions</i>	12
2.3.6 <i>Culturally-Determined Expressions</i>	12
2.4 OPEN ISSUES	12
3. THE ANNOTATION TOOL: CALLISTO.....	14
4. CORPUS CHARACTERISTICS	15
5. ANNOTATORS.....	17
6. NUMBER AND TYPES OF TEMPORAL EXPRESSIONS	18
6.1 GENERAL STATISTICS ON THE CORPUS.....	18
6.2 STATISTICS ON TIMEX2 ATTRIBUTES USED IN THE CORPUS.....	18
6.2.1 <i>VAL attribute</i>	18
6.2.2 <i>SET attribute</i>	19
6.2.3 <i>NON_SPECIFIC attribute</i>	19
6.2.4 <i>MOD attribute</i>	20
6.2.5 <i>ANCHOR_DIR attribute</i>	21
7. APPENDIX A: THE ISO 8601 STANDARD	22
8. APPENDIX B: EXAMPLE OF AN ANNOTATED FILE IN SGML FORMAT.	23
9. APPENDIX C: TEXT FILES	24
9.1 TRAINING TEXT FILES DIVIDED BY DATE AND CATEGORY	24
9.2 TEST TEXT FILES DIVIDED BY DATE AND CATEGORY	29
10. APPENDIX D: INDEX OF THE TEMPORAL EXPRESSIONS OCCURRING IN THIS DOCUMENT.	32
REFERENCES	34

ABSTRACT

This document reports on Temporal Expression annotation for the Italian Content Annotation Bank (I-CAB) being developed at ITC-irst.

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.

This work is part of a larger effort aiming at the realization of an Entity-Relation annotated corpus for Italian. The long term goal is to produce a reference benchmark for Information Extraction tasks, including both entities (e.g. temporal expressions, persons, organizations and locations) and a number of relations among such entities (e.g. date of birth of a person).

1. INTRODUCTION TO THE TASK: TIMEX2

The TIMEX2 scheme has been developed under DARPA² TIDES³ sponsorship by MITRE and SPAWAR⁴. It covers both the detection and the normalization of time expressions in order to support the characterization of the ACE⁵ program.

The major reference for TIMEX2 is *2003 Standard for the Annotation of Temporal Expressions* by Ferro et al.

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"
NON_SPECIFIC	Identifies non-specific expressions.	NON_SPECIFIC="YES"
COMMENT	Contains comments made by the annotator.	COMMENT="context garbled"

² Defense Advanced Research Projects Agency.

³ Translingual Information Detection, Extraction, and Summarization.

⁴ Space and Naval Warfare Systems Command.

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, pomeriggio, primavera, secolo, semestre, sera, settimana, stagione, vigilia, weekend	istante, episodio, occasione, periodo
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	ex, frequente
Adverb	adesso, allora, annualmente, finora, giornalmente, mensilmente, recentemente	ancora, dopo, frequentemente, già, immediatamente, mai, poi, presto, prima, sempre, tardi
Time noun/adverb	domani, dopodomani, ieri, oggi, ora	

⁶ We use the following notations:

- Examples are indented and are marked with a white bullet.
- For temporal expressions inside sentence examples the tags <TIMEX2>...</TIMEX2> are used to mark the end and the beginning.
- Other expressions are written in italics.

Number	il primo (the first day of a month), Sessanta (referring to the decade of the Sixties)	
--------	--	--

2.1.2 Non-markable expressions

- *Periodo* (and its plural *periodi*) is not a lexical trigger even if the equivalent English word *period* is considered markable.
- *Tempo* is never tagged. The only exception is the expression *a suo tempo*:
 - `<TIMEX2 VAL="PAST_REF" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="BEFORE">a suo tempo</TIMEX2>`
- *Tempi* is not tagged, but there are two exceptions, e.g. when *tempi* is preceded by *ultimi* or followed by *recenti*:
 - `<TIMEX2 VAL="PAST_REF" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="BEFORE">negli ultimi tempi</TIMEX2>` è apparsa concreta l'eventualità...
 - Pertanto non può aver partecipato ai lavori di ristrutturazione in casa Dusevich in `<TIMEX2 VAL="PAST_REF" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="BEFORE">tempi recenti</TIMEX2>`.
- Expressions like *l'ordine del giorno* and *all'ordine del giorno* are not tagged although the word *giorno* is a lexical trigger.
- Adjectives like *autunnale*, *estivo*, *invernale*, *primaverile*, *prossimo*, *recente*, *settimanale* are tagged only if they occur with a noun that is lexical trigger:
 - Groff terrà un vero e proprio live con tutte le canzoni del suo tour *estivo*.
 - Una riflessione sulla Passione di Gesù ispirata dal *recente* film di Gibson.
 - L'arrivo di `<TIMEX2 VAL="XXXX-WI">una stagione invernale</TIMEX2>`.
 - La mostra infatti verrà smontata per essere proposta `<TIMEX2 VAL="2005">il prossimo anno</TIMEX2>` a Bolzano.
- Names of holiday and expressions that refer to alternative calendars (see paragraph 2.3.6 “Culturally-Determined Expressions”).
- 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 Lexical Criteria

As in English, a temporal expression cannot start with a simple preposition (*di*, *a*, *da*, *in*, *con*, *su*, *per*, *tra*, *fra*) but it can start with an article:

- per `<TIMEX2>la settimana prossima</TIMEX2>`
- prima di `<TIMEX2>giovedì</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>`

2.2.2 Syntactic Criteria

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>

2.2.3 Multiwords

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

- 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

2.2.4 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.

- 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:

- 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.

2.2.5 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:

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

2.2.6 Conjoined Expressions

Two temporal expressions joined with a conjunction are tagged separately:

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

The following example is syntactically different so it is treated in a different way:

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

2.2.7 Embedded Expressions

When to create one tag:

- a) Hierarchically-related expressions
 - <TIMEX2>novembre 1943</TIMEX2>
 - <TIMEX2>venerdì sera</TIMEX2>

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

When to create separate tags:

- a) Time-Anchored expressions
 - <TIMEX2>Un anno</TIMEX2> da <TIMEX2>ora</TIMEX2>
- b) All the other cases
 - Il concerto è <TIMEX2>venerdì</TIMEX2> <TIMEX2>alle otto di sera</TIMEX2>
 - Il concerto è <TIMEX2>martedì</TIMEX2> dopo <TIMEX2>le 12 PM</TIMEX2>
 - <TIMEX2>ieri</TIMEX2>, <TIMEX2>alle 11</TIMEX2>
 - <TIMEX2>ieri mattina</TIMEX2> <TIMEX2>alle 11</TIMEX2>

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**

- o 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.

- o 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.
- o Cerchiamo di divertirci <TIMEX2 VAL="P4Y" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="STARTING">nei prossimi 4 anni</TIMEX2>.
- o 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 context whether the span occurred before or after the reference time.

- o 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

- Fra i film in concorso è quello che personalmente <TIMEX2 VAL="PRESENT_REF" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="AS_OF">finora</TIMEX2> preferiamo.
- Uno sguardo <TIMEX2 VAL="FUTURE_REF" ANCHOR_VAL="2004-10-08" ANCHOR_DIR="AFTER">al futuro</TIMEX2>
- 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:

- <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

- quando mancano <TIMEX2 VAL="P4M" MOD="LESS_THAN">poco meno di 4 mesi</TIMEX2>
- era stato per <TIMEX2 VAL="P6M" MOD="MORE_THAN">oltre sei mesi</TIMEX2> in Iraq
- Per arrivare a Bari impiego <TIMEX2 VAL="PT2H" MOD="EQUAL_OR_LESS">al massimo due ore</TIMEX2>.
- 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>.
- <TIMEX2 VAL="2004-09" MOD="START">all'inizio di settembre</TIMEX2> ha dovuto giocoforza riparare qualche foratura imprevista.
- <TIMEX2 VAL="199" MOD="START">dai primi anni novanta</TIMEX2> la cooperativa ha iniziato ad accumulare debiti.
- L'impasse ha portato, <TIMEX2 VAL="2004-07" MOD="MID">alla metà dello scorso luglio</TIMEX2>, all'interruzione delle trattative.
- molto attivo <TIMEX2 VAL="2004-SU" MOD="END">in questa fine estate</TIMEX2>
- 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*:

- <TIMEX2 VAL="1994" MOD="APPROX">una decina di anni fa</TIMEX2>
- 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.

- Andranno in onda <TIMEX2 VAL="XXXX-W05" SET="YES">tutti i venerdì</TIMEX2>.
- <TIMEX2 VAL="P1M" SET="YES">Ogni mese</TIMEX2> viene trattata una voce.

- Un abbonamento che costa 52 euro <TIMEX2 VAL="XXXX" SET="YES">all'anno</TIMEX2>.
- 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:

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

2.3.5 Non-Specific Temporal Expressions

These expressions do not refer to a specific time, therefore we identify them by adding the NON_SPECIFIC="YES" attribute to the TIMEX2 tag. In all the other cases, the attribute is omitted from the annotation altogether.

- Passano <TIMEX2 VAL="PXY" NON_SPECIFIC="YES">gli anni</TIMEX2>, ma non il successo degli zaini Seven.
- Raffiche di vento a 180 chilometri <TIMEX2 VAL="PT1H" NON_SPECIFIC="YES">all'ora</TIMEX2>.
- Amicizia vuol dire passare <TIMEX2 VAL="XXXX-XX-XXTAF" NON_SPECIFIC="YES">pomeriggi</TIMEX2> e <TIMEX2 VAL="XXXX-XX-XXTEV" NON_SPECIFIC="YES">serate</TIMEX2> e <TIMEX2 VAL="XXXX-XX-XXTNI">nottate</TIMEX2> insieme a leggere.

2.3.6 Culturally-Determined Expressions

The interpretation of some temporal expressions requires cultural or domain-specific knowledge. In English, these expressions are markable but receive no VAL.:

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

In Italian we don't tag names of holidays like *Capodanno*, *Ferragosto*, *Natale*, *Pasqua*, *Vigilia di Natale*:

- Il Mercatino di Natale di Piazza Fiera.

Expressions that refer to calendars different from the Gregorian one are not markable. In English these expressions are tagged when they have a lexical trigger as their head (like *season* in *baseball season*), but they receive no value.

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

2.4 Open Issues

a) The reasons for not annotating the following expressions have not been discussed yet (this was inherited by an annotator who left IRST before the annotation of the Corpus was completed):

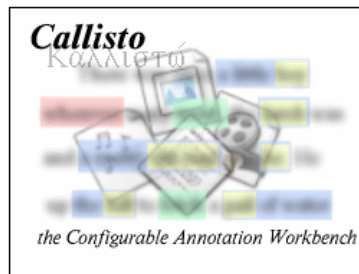
- adjectives like *autunnale*, *estivo*, *invernale*, *primaverile*, *prossimo*, *recente*, *settimanale* without a lexical trigger noun;
- *periodo* and *periodi*;

- culturally-determined expressions like *Natale*;
- b) We have decided not to tag *giorno* in the expressions *all'ordine del giorno* and *l'ordine del giorno*.
- c) We have decided to tag the expressions *ore e ore*, *giorno dopo giorno* and *24 ore su 24* in the following ways:
 - <TIMEX2 VAL="PTXH">ore e ore</TIMEX2> di bombardamenti.
 - <TIMEX2 VAL="XXXX-XX-XX" SET="YES">giorno dopo giorno</TIMEX2> si arricchisce di informazioni.
 - A disposizione <TIMEX2 VAL="XXXX-XX-XXTXX" SET="YES">24 ore su 24</TIMEX2>.
- d) 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.
 - <TIMEX2 VAL="PXW" ANCHOR_VAL="2004-w41" ANCHOR_DIR="AFTER">Nelle prossime settimane</TIMEX2> si farà un'analisi approfondita.

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.

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

3. THE ANNOTATION TOOL: CALLISTO



Callisto is an annotation tool for linguistic annotation of textual data, developed by the MITRE Corporation with the initial funding of the U.S. government.

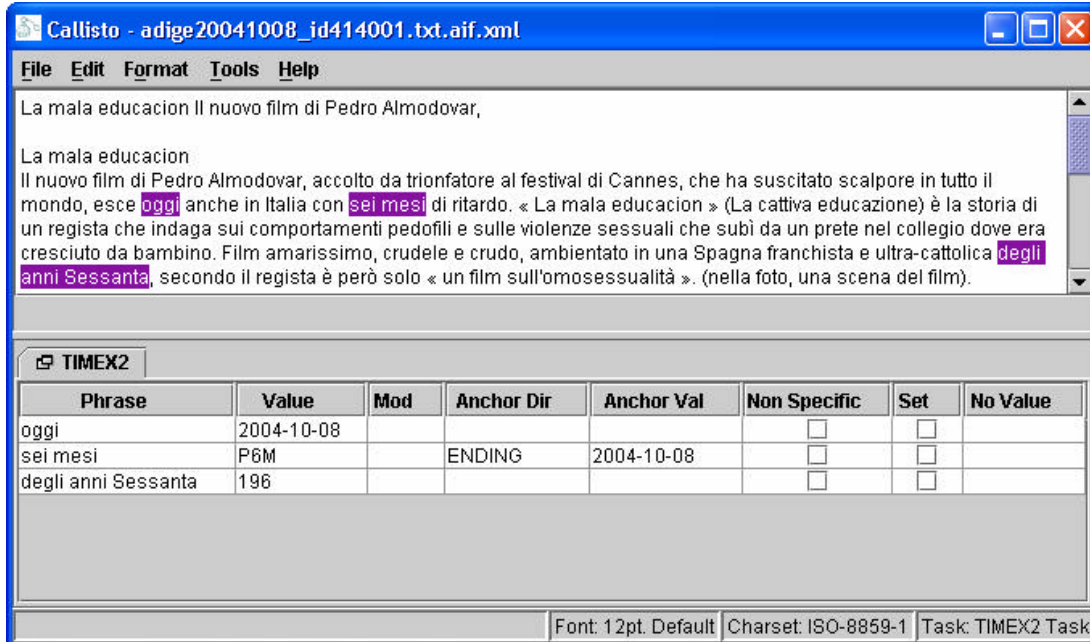
It is written in Java, so it is possible to annotate any Unicode-supported language, and it accepts files encoded as UTF-8 and US-ASCII.

Callisto has been built with a modular design that permits to create extensions for a specific domain. It utilizes standoff-annotation, allowing for unique tag-set definitions and domain dependent interfaces.

As illustrated in the figure below, Callisto's main window is divided into two parts:

- the *Main Text Pane* (on top) in which the text cannot be edited but only annotated;
- the *Editor Pane* (on bottom), customized for the annotation task at hand.

Figure 3.1 Callisto's main window



Currently Callisto recognizes and handles four distinct types of files: raw text files, sgml (including html, xml) files, Ace Pilot Format (APF) files, and Atlas Interchange Format (AIF) files. All data annotated with this version of Callisto are saved in AIF format.

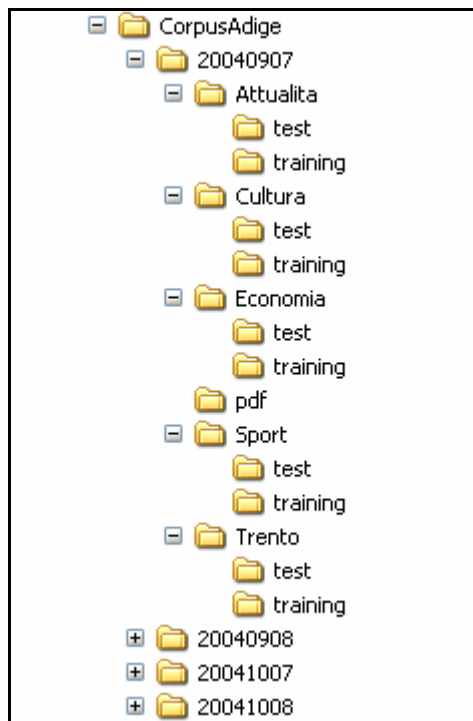
Callisto is available at <http://callisto.mitre.org>; a new version (V.2.0) was released in March 2005.

4. CORPUS CHARACTERISTICS

The **Adige Corpus** consists of 524 files derived from the local newspaper “L’Adige”. Articles belong to four days: September, 7th and 8th 2004 and October, 7th and 8th 2004. Five categories were chosen: Attualità (*News Stories*), Cultura (*Cultural News*), Economia (*Economy News*), Sport (*Sports News*), Trento (*Local News*).

Christian Girardi downloaded the articles in pdf and html formats from the website <http://www.ladige.it/>. He created a text document with UTF-8 encoding for each article and a directory for each day inside the main directory “**Adige**”. Each day contains six directories: one for each category plus one that includes pdf documents. Alessandra Novello created the distinction between Training and Test files under each category directory (see figure 4.1.1).

Figure 4.1.1 Structure of the Adige Corpus



Each article has its own header file (.hdr extension) containing some general information such as the number of words and the title:

Figure 4.1.2 Example of header file

```
WCOUNT=99
PAGINA=12
AUTORE=
SUBJECTS=
NOMI=
KEYWORD=
KIND=news
GENERE=
LUOGHI=
TABELLE=
PDF=20040907/pdf/AD07CUT1.pdf
DATA=07/09/2004
OWNER='Adige
TITOLO=Concerto di altissimo livello Dall'Ungheria a Merano
PERSONE=
SEZIONE=Cultura
ORGANIZZAZIONI=
```

5. ANNOTATORS

Danilo Giampiccolo, Alessandra Novello and Alessandro Vallin started the discussion on the annotation of temporal expression for Italian in October 2004.

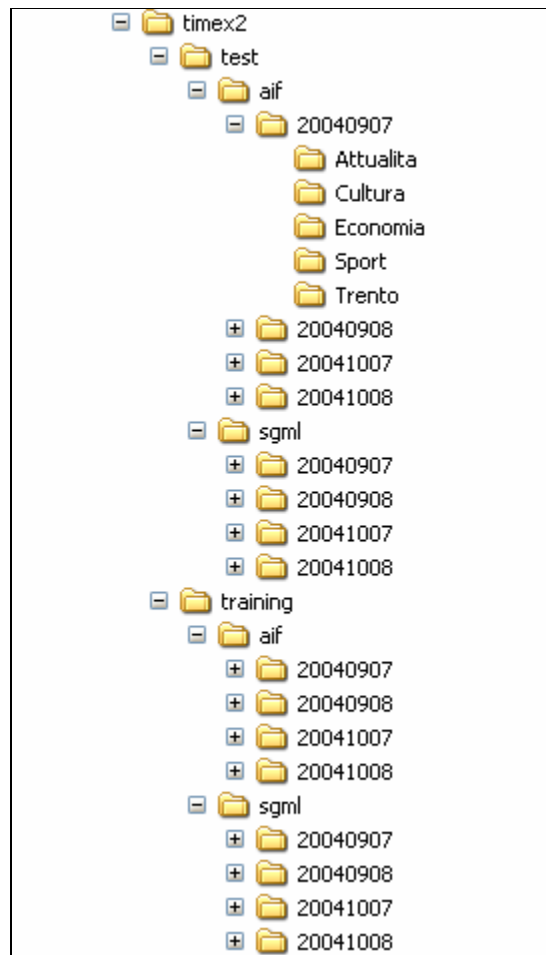
Then, Alessandra Novello annotated most of the documents included in the training section of the Corpus.

In April 2004 Rachele Sprugnoli substituted Alessandra Novello and carried on the work annotating the last part of the training section (20041007 *Sports News* – 20041008 *Local News* and *Sports News*) and all the test section.

After the annotation, each text file was saved in the *aif* format and was exported in the *sgml* format. Aif and sgml files are divided into separate directories, as it is showed in the figure below.

The annotation of temporal expressions ended in May 2005.

Figure 6 Structure of the annotated corpus.



6. NUMBER AND TYPES OF TEMPORAL EXPRESSIONS

6.1 General statistics on the Corpus

TRAINING + TEST

Number of files = 524

Number of words = 182553

Average number of words per file = 348

Number of TIMEX2 tags = 3927

Number of words within TIMEX2 tags = 7708 [4.22% of the total]

Average length of temporal expressions = 2.0 words

Number of characters within TIMEX2 tags = 39251

Average number of characters per temporal expression = 10.0

TRAINING

Number of files = 333

Number of words = 113198

Average number of words per file = 339

Number of TIMEX2 tags = 2482

Number of words within TIMEX2 tags = 4790 [4.23% of the total]

Average length of temporal expressions = 1.9

Number of characters within TIMEX2 tags = 24163

Average number of characters per temporal expression = 9.7

TEST

Number of files = 191

Number of words = 69355

Average number of words per file = 363

Number of TIMEX2 tags = 1445

Number of words within TIMEX2 tags = 2918 [4.21% of the total]

Average length of temporal expressions = 2.0

Number of characters within TIMEX2 tags = 15088

Average number of characters per temporal expression = 10.4

6.2 Statistics on TIMEX2 attributes used in the Corpus

6.2.1 VAL attribute

- Periods: VAL="P..."

	# OCCURRENCES	# WORDS	AVERAGE LENGTH
Training	501/2482 (20.18%)	1017/113198(0.90%)	2.0 words
Test	384/1445 (26.57%)	853/69355 (1.23%)	2.2 words

- Points

	# OCCURRENCES	# WORDS	AVERAGE LENGTH
Training	1981/2482 (79.81%)	3773/113198 (3.33%)	1.9 words
Test	1061/1445 (73.42%)	2065/69355 (2.98%)	1.9 words

- VAL="PRESENT_REF"

	# OCCURRENCES	# WORDS	AVERAGE LENGTH
Training	165/2482 (6.65%)	203/113198 (0.18%)	1.2 words
Test	54/1445 (3.74%)	74/69355 (0.11%)	1.4 words

- VAL="FUTURE_REF"

	# OCCURRENCES	# WORDS	AVERAGE LENGTH
Training	47/2482 (1.90%)	97/113198 (0.08%)	2.1 words
Test	20/1445 (1.38%)	42/69355 (0.06%)	2.1 words

- VAL="PAST_REF"

	# OCCURRENCES	# WORDS	AVERAGE LENGTH
Training	50/2482 (2.01%)	97/113198 (0.08%)	1.9 words
Test	28/1445 (1.94%)	57/69355 (0.06%)	2.0 words

- VAL="XXXX-XX-XX"

	# OCCURRENCES	# WORDS	AVERAGE LENGTH
Training	24/2482 (0.97%)	59/113198 (0.05%)	2.5 words
Test	69/1445 (4.77%)	112/69355 (0.16%)	1.6 words

- VAL="" (empty value)

	# OCCURRENCES	# WORDS	AVERAGE LENGTH
Training	5/2482 (0.20%)	13/113198 (0.01%)	2.6 words
Test	4/1445 (0.28%)	9/69355 (0.01%)	2.3 words

6.2.2 SET attribute

- SET="YES"

	# OCCURRENCES	# WORDS	AVERAGE LENGTH
Training	115/2482 (4.63 %)	175/113198 (0.15%)	1.5 words
Test	41/1445 (2.84%)	73/69355 (0.10%)	1.8 words

6.2.3 NON_SPECIFIC attribute

- NON_SPECIFIC="YES"

	# OCCURRENCES	# WORDS	AVERAGE LENGTH
Training	38/2482 (1.53 %)	126/113198 (0.11%)	1.6 words
Test	35/1445 (2.42%)	66/69355 (0.09%)	1.9 words

6.2.4 MOD attribute

VALUE FOR MOD	CORPUS SECTION	# OCCURRENCES	# WORDS	AVER. LENGTH
“AFTER”	Training	3/2482 (0.12 %)	11/113198 (0.01%)	3.7 words
	Test	-	-	-
“APPROX”	Training	23/2482 (0.93%)	79/113198 (0.07%)	3.4 words
	Test	28/1445 (1.94%)	87/69355 (0.12%)	3.1 words
“BEFORE”	Training	2/2482 (0.08%)	7/113198 (0.07%)	3.5 words
	Test	1/1445 (0.07%)	8/69355 (0.01%)	8.0 words
“END”	Training	31/2482 (1.25%)	98/113198 (0.09%)	3.2 words
	Test	13/1445 (0.90%)	37/69355 (0.05%)	2.8 words
“MID”	Training	8/2482 (0.32%)	22/113198 (0.02%)	2.8 words
	Test	3/1445 (0.21%)	9/69355 (0.01%)	3.0 words
“START”	Training	16/2482 (0.64%)	60/113198 (0.05%)	3.8 words
	Test	6/1445 (0.41%)	19/69355 (0.03%)	3.2 words
“ON_OR_AFTER”	Training	-	-	-
	Test	-	-	-
“ON_OR_BEFORE”	Training	-	-	-
	Test	-	-	-
“EQUAL_OR_LESS”	Training	5/2482 (0.20%)	20/113198 (0.02%)	4.0 words
	Test	1/1445 (0.07%)	3/69355 (0.004%)	3.0 words
“EQUAL_OR_MORE”	Training	2/2482 (0.08%)	5/113198 (0.004%)	2.5 words
	Test	5/1445 (0.35%)	16/69355 (0.02%)	3.2 words
“LESS_THAN”	Training	1/2482 (0.04%)	5/113198 (0.004%)	5.0 words
	Test	2/1445 (0.14%)	4/69355 (0.005%)	2.0 words
“MORE_THAN”	Training	10/2482 (0.40%)	32/113198 (0.03%)	3.2 words
	Test	14/1445 (0.97%)	36/69355 (0.05%)	2.6 words
TOTAL	Training	101/2482 (4.07%)	339/113198 (0.30%)	3.3 words
	Test	73/1445 (5.05%)	219/69355 (0.31%)	3.0 words

6.2.5 ANCHOR_DIR attribute

VALUE FOR ANCHOR_DIR	CORPUS SECTION	# OCCURRENCES	# WORDS	AVER. LENGTH
“AS_OF” ⁷	Training	165/2482 (6.65%)	203/113198 (0.18%)	1.2 words
	Test	54/1445 (3.74%)	74/69355 (0.11%)	1.4 words
“AFTER” without VAL=“FUTURE_REF”	Training	21/2482 (0.85%)	53/113198 (0.05%)	2.5 words
	Test	8/1445 (0.55%)	22/69355 (0.03%)	2.8 words
“AFTER” with VAL=“FUTURE_REF”	Training	47/2482 (1.90%)	97/113198 (0.08%)	2.1 words
	Test	20/1445 (1.38%)	42/69355 (0.06%)	2.1 words
“BEFORE” without VAL=“PAST_REF”	Training	23/2482 (0.93%)	58/113198 (0.05%)	2.5 words
	Test	28/1445 (1.94%)	63/69355 (0.09%)	2.3 words
“BEFORE” with VAL=“PAST_REF”	Training	50/2482 (2.01%)	97/113198 (0.08%)	1.9 words
	Test	28/1445 (1.94%)	57/69355 (0.06%)	2.0 words
“ENDING”	Training	169/2482 (6.81%)	353/113198 (0.31%)	2.1 words
	Test	105/1445 (7.27%)	212/69355 (0.30%)	2.0 words
“STARTING”	Training	36/2482 (1.45%)	82/113198 (0.07%)	2.3 words
	Test	28/1445 (1.94%)	95/69355 (0.14%)	3.4 words
“WITHIN”	Training	41/2482 (1.65%)	89/113198 (0.08%)	2.2 words
	Test	10/1445 (0.69%)	25/69355 (0.04%)	2.5 words
TOTAL	Training	552/2482 (22.24%)	1032/113198 (0.91%)	1.9 words
	Test	281/1445 (19.45%)	590/69355 (0.85%)	2.1 words

⁷ The value “AS_OF” is present only in combination with VAL=“PRESENT_REF”.

7. APPENDIX A: 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 53⁸ (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 (for the week format)

⁸ If it is a leap year.

8. APPENDIX B: EXAMPLE OF AN ANNOTATED FILE IN SGML FORMAT.

Michelin, manifestazione in Provincia. Poi l'incontro con l'azienda Lavoratori in piazza contro le crisi

```
TRENTO - <TIMEX2 val="2004-10-07" mod="" set=""
non_specific="" anchor_val="" anchor_dir=""
comment="">Ieri</TIMEX2> la Stl, <TIMEX2 val="2004-10-08"
mod="" set="" non_specific="" anchor_val="" anchor_dir=""
comment="">oggi</TIMEX2> la Michelin e tutte le altre
industrie a rischio chiusura o delocalizzazione. A partire
<TIMEX2 val="2004-10-08T09:30" mod="" set=""
non_specific="" anchor_val="" anchor_dir=""
comment="">dalle 9.30</TIMEX2>, i lavoratori del comparto,
in sciopero, manifesteranno davanti alla sede del consiglio
provinciale in piazza Dante. Un presidio deciso
unitariamente da Cgil, Cisl e Uil, per portare il tema
delle crisi industriali all'attenzione dei consiglieri,
impegnati nei lavori d'aula. Una delegazione sindacale
verrà ricevuta dal presidente della giunta provinciale,
Lorenzo Dellai, che ha convocato per <TIMEX2 val="2004-10-
13" mod="" set="" non_specific="" anchor_val=""
anchor_dir="" comment="">mercoledì prossimo</TIMEX2> un
vertice allargato su questo problema, al quale
parteciperanno tutte le parti sociali.
I lavoratori coinvolti dallo sciopero di <TIMEX2 val="2004-
10-08" mod="" set="" non_specific="" anchor_val=""
anchor_dir="" comment="">oggi</TIMEX2> dovrebbero essere un
migliaio, parte dei quali è attesa in piazza Dante. Per
l'occasione, i sindacati hanno preparato dei manifesti-
necrologi per ogni azienda che ha chiuso o delocalizzato la
produzione.
<TIMEX2 val="2004-10-08TAF" mod="" set="" non_specific=""
anchor_val="" anchor_dir="" comment="">Nel
pomeriggio</TIMEX2>, poi, l'appuntamento forse più
importante: l'incontro tra i sindacati di categoria e i
vertici di Michelin Italiana, per un primo confronto sulla
gestione dell'annunciata chiusura di Spini di Gardolo.
```

9. APPENDIX C: TEXT FILES

9.1 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 – SPORT 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_id405593.txt
11. adige20040907_id405594.txt
12. adige20040907_id405595.txt
13. adige20040907_id405596.txt
14. adige20040907_id405597.txt
15. adige20040907_id405598.txt
16. adige20040907_id405599.txt
17. adige20040907_id405600.txt
18. 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_id405617.txt
15. adige20040907_id405618.txt
16. adige20040907_id405619.txt
17. adige20040907_id405620.txt
18. adige20040907_id405621.txt
19. adige20040907_id405622.txt
20. adige20040907_id405623.txt

21. adige20040907_id405624.txt
22. adige20040907_id405625.txt
23. adige20040907_id405626.txt
24. adige20040907_id405627.txt
25. adige20040907_id405630.txt
26. adige20040907_id405631.txt
27. adige20040907_id405632.txt
28. adige20040907_id405633.txt
29. adige20040907_id405634.txt
30. adige20040907_id405635.txt

20040908

ATTUALITÀ - NEWS STORIES

1. adige20040908_id405656.txt
2. adige20040908_id405657.txt
3. adige20040908_id405658.txt
4. adige20040908_id405659.txt
5. adige20040908_id405660.txt
6. adige20040908_id405661.txt
7. adige20040908_id405662.txt
8. adige20040908_id405663.txt
9. adige20040908_id405664.txt
10. adige20040908_id405666.txt
11. adige20040908_id405667.txt
12. adige20040908_id405669.txt
13. adige20040908_id405670.txt
14. adige20040908_id405671.txt
15. adige20040908_id405672.txt
16. adige20040908_id405673.txt

CULTURA - CULTURAL NEWS

1. adige20040908_id405684.txt
2. adige20040908_id405685.txt
3. adige20040908_id405686.txt
4. adige20040908_id405687.txt
5. adige20040908_id405688.txt
6. adige20040908_id405689.txt
7. adige20040908_id405690.txt
8. adige20040908_id405691.txt
9. adige20040908_id405692.txt
10. adige20040908_id405693.txt

ECONOMIA - ECONOMY NEWS

1. adige20040908_id405704.txt
2. adige20040908_id405705.txt
3. adige20040908_id405706.txt

4. adige20040908_id405707.txt
5. adige20040908_id405708.txt
6. adige20040908_id405710.txt
7. adige20040908_id405711.txt
8. adige20040908_id405712.txt
9. adige20040908_id405713.txt
10. adige20040908_id405714.txt

SPORT - SPORTS NEWS

1. adige20040908_id405848.txt
2. adige20040908_id405849.txt
3. adige20040908_id405850.txt
4. adige20040908_id405851.txt
5. adige20040908_id405852.txt
6. adige20040908_id405853.txt
7. adige20040908_id405854.txt
8. adige20040908_id405855.txt
9. adige20040908_id405856.txt
10. adige20040908_id405857.txt
11. adige20040908_id405858.txt
12. adige20040908_id405859.txt
13. adige20040908_id405860.txt
14. adige20040908_id405861.txt
15. adige20040908_id405862.txt
16. adige20040908_id405863.txt
17. adige20040908_id405864.txt
18. adige20040908_id405865.txt
19. adige20040908_id405866.txt
20. adige20040908_id405867.txt
21. adige20040908_id405868.txt
22. adige20040908_id405869.txt
23. adige20040908_id405871.txt
24. adige20040908_id405874.txt
25. adige20040908_id405875.txt
26. adige20040908_id405877.txt
27. adige20040908_id405878.txt
28. adige20040908_id405880.txt

TRENTO - LOCAL NEWS

1. adige20040908_id405881.txt
2. adige20040908_id405882.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. adige20040908_id405890.txt
10. adige20040908_id405894.txt
11. adige20040908_id405896.txt
12. adige20040908_id405897.txt
13. adige20040908_id405898.txt
14. adige20040908_id405900.txt
15. adige20040908_id405901.txt
16. adige20040908_id405902.txt
17. adige20040908_id405903.txt
18. adige20040908_id405904.txt
19. adige20040908_id405906.txt
20. adige20040908_id405907.txt
21. adige20040908_id405908.txt
22. adige20040908_id405910.txt
23. adige20040908_id405914.txt
24. adige20040908_id405915.txt
25. adige20040908_id405916.txt
26. adige20040908_id405917.txt
27. adige20040908_id405918.txt

20041007

ATTUALITÀ - NEWS STORIES

1. adige20041007_id413699.txt
2. adige20041007_id413700.txt
3. adige20041007_id413701.txt
4. adige20041007_id413702.txt
5. adige20041007_id413703.txt
6. adige20041007_id413704.txt
7. adige20041007_id413705.txt
8. adige20041007_id413706.txt
9. adige20041007_id413708.txt
10. adige20041007_id413709.txt
11. adige20041007_id413710.txt
12. adige20041007_id413711.txt

CULTURA - CULTURAL NEWS

1. adige20041007_id413719.txt
2. adige20041007_id413720.txt
3. adige20041007_id413721.txt
4. adige20041007_id413722.txt
5. adige20041007_id413723.txt
6. adige20041007_id413724.txt
7. adige20041007_id413725.txt
8. adige20041007_id413726.txt
9. adige20041007_id413727.txt

10. adige20041007_id413728.txt

ECONOMIA - ECONOMY NEWS

1. adige20041007_id413743.txt
2. adige20041007_id413744.txt
3. adige20041007_id413745.txt
4. adige20041007_id413746.txt
5. adige20041007_id413747.txt
6. adige20041007_id413748.txt
7. adige20041007_id413750.txt
8. adige20041007_id413751.txt

SPORT - SPORTS NEWS

1. adige20041007_id413887.txt
2. adige20041007_id413888.txt
3. adige20041007_id413890.txt
4. adige20041007_id413891.txt
5. adige20041007_id413892.txt
6. adige20041007_id413893.txt
7. adige20041007_id413894.txt
8. adige20041007_id413895.txt
9. adige20041007_id413897.txt
10. adige20041007_id413898.txt
11. adige20041007_id413899.txt
12. adige20041007_id413900.txt
13. adige20041007_id413901.txt
14. adige20041007_id413902.txt
15. adige20041007_id413903.txt
16. adige20041007_id413904.txt
17. adige20041007_id413905.txt
18. adige20041007_id413906.txt

TRENTO - LOCAL NEWS

1. adige20041007_id413916.txt
2. adige20041007_id413917.txt
3. adige20041007_id413918.txt
4. adige20041007_id413919.txt
5. adige20041007_id413920.txt
6. adige20041007_id413921.txt
7. adige20041007_id413922.txt
8. adige20041007_id413923.txt
9. adige20041007_id413924.txt
10. adige20041007_id413925.txt
11. adige20041007_id413926.txt
12. adige20041007_id413927.txt
13. adige20041007_id413928.txt

14. adige20041007_id413929.txt
15. adige20041007_id413930.txt
16. adige20041007_id413931.txt
17. adige20041007_id413932.txt
18. adige20041007_id413933.txt
19. adige20041007_id413934.txt
20. adige20041007_id413935.txt
21. adige20041007_id413936.txt
22. adige20041007_id413937.txt
23. adige20041007_id413938.txt
24. adige20041007_id413939.txt
25. adige20041007_id413940.txt
26. adige20041007_id413941.txt
27. adige20041007_id413942.txt
28. adige20041007_id413943.txt
29. adige20041007_id413944.txt
30. adige20041007_id413945.txt

20041008

ATTUALITÀ - NEWS STORIES

1. adige20041008_id413973.txt
2. adige20041008_id413974.txt
3. adige20041008_id413975.txt
4. adige20041008_id413976.txt
5. adige20041008_id413977.txt
6. adige20041008_id413978.txt
7. adige20041008_id413979.txt
8. adige20041008_id413980.txt
9. adige20041008_id413981.txt
10. adige20041008_id413982.txt
11. adige20041008_id413984.txt
12. adige20041008_id413985.txt
13. adige20041008_id413986.txt
14. adige20041008_id413987.txt

CULTURA - CULTURAL NEWS

1. adige20041008_id413995.txt
2. adige20041008_id413996.txt
3. adige20041008_id413997.txt
4. adige20041008_id413998.txt
5. adige20041008_id413999.txt
6. adige20041008_id414000.txt
7. adige20041008_id414001.txt
8. adige20041008_id414002.txt
9. adige20041008_id414003.txt
10. adige20041008_id414004.txt

11. adige20041008_id414005.txt
12. adige20041008_id414007.txt

ECONOMIA - ECONOMY NEWS

1. adige20041008_id414017.txt
2. adige20041008_id414018.txt
3. adige20041008_id414019.txt
4. adige20041008_id414020.txt
5. adige20041008_id414021.txt
6. adige20041008_id414022.txt
7. adige20041008_id414023.txt
8. adige20041008_id414024.txt
9. adige20041008_id414025.txt

SPORT - SPORTS NEWS

1. adige20041008_id414146.txt
2. adige20041008_id414147.txt
3. adige20041008_id414148.txt
4. adige20041008_id414149.txt
5. adige20041008_id414150.txt
6. adige20041008_id414151.txt
7. adige20041008_id414152.txt
8. adige20041008_id414153.txt
9. adige20041008_id414154.txt
10. adige20041008_id414155.txt
11. adige20041008_id414156.txt
12. adige20041008_id414157.txt
13. adige20041008_id414158.txt
14. adige20041008_id414159.txt
15. adige20041008_id414160.txt
16. adige20041008_id414163.txt

TRENTO - LOCAL NEWS

1. adige20041008_id414176.txt
2. adige20041008_id414177.txt
3. adige20041008_id414178.txt
4. adige20041008_id414179.txt
5. adige20041008_id414180.txt
6. adige20041008_id414181.txt
7. adige20041008_id414182.txt
8. adige20041008_id414183.txt
9. adige20041008_id414184.txt
10. adige20041008_id414185.txt
11. adige20041008_id414186.txt
12. adige20041008_id414187.txt
13. adige20041008_id414188.txt

14. adige20041008_id414189.txt
15. adige20041008_id414190.txt
16. adige20041008_id414191.txt
17. adige20041008_id414192.txt
18. adige20041008_id414193.txt
19. adige20041008_id414194.txt
20. adige20041008_id414195.txt
21. adige20041008_id414196.txt
22. adige20041008_id414197.txt
23. adige20041008_id414198.txt

24. adige20041008_id414199.txt
25. adige20041008_id414206.txt
26. adige20041008_id414207.txt
27. adige20041008_id414208.txt
28. adige20041008_id414209.txt
29. adige20041008_id414210.txt
30. adige20041008_id414211.txt

Total: 333 files

9.2 Test text files divided by date and category

20040907

ATTUALITÀ - NEWS STORIES

1. adige20040907_id405400.txt
2. adige20040907_id405401.txt
3. adige20040907_id405402.txt
4. adige20040907_id405403.txt
5. adige20040907_id405404.txt
6. adige20040907_id405405.txt
7. adige20040907_id405406.txt
8. adige20040907_id405407.txt

CULTURA - CULTURAL NEWS

1. adige20040907_id405426.txt
2. adige20040907_id405427.txt
3. adige20040907_id405428.txt
4. adige20040907_id405429.txt
5. adige20040907_id405430.txt
6. adige20040907_id405432.txt
7. adige20040907_id405433.txt
8. adige20040907_id405434.txt

ECONOMIA - ECONOMY NEWS

1. adige20040907_id405450.txt
2. adige20040907_id405451.txt
3. adige20040907_id405452.txt
4. adige20040907_id405453.txt
5. adige20040907_id405455.txt

SPORT - SPORTS NEWS

1. adige20040907_id405571.txt
2. adige20040907_id405572.txt
3. adige20040907_id405573.txt
4. adige20040907_id405574.txt

5. adige20040907_id405575.txt
6. adige20040907_id405576.txt
7. adige20040907_id405577.txt
8. adige20040907_id405578.txt
9. adige20040907_id405579.txt
10. adige20040907_id405580.txt

TRENTO - LOCAL NEWS

1. adige20040907_id405636.txt
2. adige20040907_id405637.txt
3. adige20040907_id405638.txt
4. adige20040907_id405639.txt
5. adige20040907_id405640.txt
6. adige20040907_id405641.txt
7. adige20040907_id405642.txt
8. adige20040907_id405643.txt
9. adige20040907_id405644.txt
10. adige20040907_id405645.txt
11. adige20040907_id405646.txt
12. adige20040907_id405647.txt
13. adige20040907_id405648.txt
14. adige20040907_id405649.txt
15. adige20040907_id405650.txt

20040908

ATTUALITÀ - NEWS STORIES

1. adige20040908_id405674.txt
2. adige20040908_id405675.txt
3. adige20040908_id405676.txt
4. adige20040908_id405677.txt
5. adige20040908_id405678.txt
6. adige20040908_id405679.txt
7. adige20040908_id405680.txt

8. adige20040908_id405681.txt
9. adige20040908_id405683.txt

CULTURA - CULTURAL NEWS

1. adige20040908_id405694.txt
2. adige20040908_id405695.txt
3. adige20040908_id405697.txt
4. adige20040908_id405698.txt
5. adige20040908_id405699.txt
6. adige20040908_id405700.txt
7. adige20040908_id405701.txt
8. adige20040908_id405702.txt

ECONOMIA - ECONOMY NEWS

1. adige20040908_id405715.txt
2. adige20040908_id405716.txt
3. adige20040908_id405717.txt
4. adige20040908_id405719.txt
5. adige20040908_id405722.txt

SPORT - SPORTS NEWS

1. adige20040908_id405833.txt
2. adige20040908_id405834.txt
3. adige20040908_id405836.txt
4. adige20040908_id405837.txt
5. adige20040908_id405838.txt
6. adige20040908_id405839.txt
7. adige20040908_id405840.txt
8. adige20040908_id405841.txt
9. adige20040908_id405842.txt
10. adige20040908_id405843.txt
11. adige20040908_id405844.txt
12. adige20040908_id405845.txt
13. adige20040908_id405846.txt
14. adige20040908_id405847.txt

TRENTO - LOCAL NEWS

1. adige20040908_id405919.txt
2. adige20040908_id405920.txt
3. adige20040908_id405921.txt
4. adige20040908_id405923.txt
5. adige20040908_id405925.txt
6. adige20040908_id405926.txt
7. adige20040908_id405927.txt
8. adige20040908_id405928.txt
9. adige20040908_id405929.txt

10. adige20040908_id405930.txt
11. adige20040908_id405931.txt
12. adige20040908_id405932.txt
13. adige20040908_id405933.txt
14. adige20040908_id405934.txt
15. adige20040908_id405936.txt
16. adige20040908_id405937.txt

20041007

ATTUALITÀ - NEWS STORIES

1. adige20041007_id413712.txt
2. adige20041007_id413713.txt
3. adige20041007_id413714.txt
4. adige20041007_id413715.txt
5. adige20041007_id413717.txt
6. adige20041007_id413718.txt

CULTURA - CULTURAL NEWS

1. adige20041007_id413735.txt
2. adige20041007_id413736.txt
3. adige20041007_id413737.txt
4. adige20041007_id413738.txt
5. adige20041007_id413740.txt
6. adige20041007_id413741.txt

ECONOMIA - ECONOMY NEWS

1. adige20041007_id413752.txt
2. adige20041007_id413753.txt
3. adige20041007_id413754.txt
4. adige20041007_id413755.txt

SPORT - SPORTS NEWS

1. adige20041007_id413907.txt
2. adige20041007_id413908.txt
3. adige20041007_id413909.txt
4. adige20041007_id413910.txt
5. adige20041007_id413911.txt
6. adige20041007_id413912.txt
7. adige20041007_id413913.txt
8. adige20041007_id413914.txt
9. adige20041007_id413915.txt

TRENTO - LOCAL NEWS

1. adige20041007_id413946.txt
2. adige20041007_id413947.txt
3. adige20041007_id413948.txt

4. adige20041007_id413949.txt
5. adige20041007_id413950.txt
6. adige20041007_id413951.txt
7. adige20041007_id413952.txt
8. adige20041007_id413953.txt
9. adige20041007_id413954.txt
10. adige20041007_id413955.txt
11. adige20041007_id413956.txt
12. adige20041007_id413958.txt
13. adige20041007_id413959.txt
14. adige20041007_id413960.txt
15. adige20041007_id413961.txt
16. adige20041007_id413962.txt
17. adige20041007_id413963.txt
18. adige20041007_id413964.txt
19. adige20041007_id413965.txt

20041008

ATTUALITÀ - NEWS STORIES

1. adige20041008_id413988.txt
2. adige20041008_id413989.txt
3. adige20041008_id413990.txt
4. adige20041008_id413991.txt
5. adige20041008_id413992.txt
6. adige20041008_id413993.txt
7. adige20041008_id413994.txt

CULTURA - CULTURAL NEWS

1. adige20041008_id414009.txt
2. adige20041008_id414010.txt
3. adige20041008_id414011.txt
4. adige20041008_id414012.txt
5. adige20041008_id414014.txt
6. adige20041008_id414015.txt

ECONOMIA - ECONOMY NEWS

1. adige20041008_id414026.txt
2. adige20041008_id414027.txt
3. adige20041008_id414029.txt

4. adige20041008_id414030.txt
5. adige20041008_id414031.txt

SPORT - SPORTS NEWS

1. adige20041008_id414164.txt
2. adige20041008_id414165.txt
3. adige20041008_id414166.txt
4. adige20041008_id414167.txt
5. adige20041008_id414168.txt
6. adige20041008_id414169.txt
7. adige20041008_id414171.txt
8. adige20041008_id414173.txt
9. adige20041008_id414174.txt
10. adige20041008_id414175.txt

TRENTO - LOCAL NEWS

1. adige20041008_id414213.txt
2. adige20041008_id414214.txt
3. adige20041008_id414215.txt
4. adige20041008_id414216.txt
5. adige20041008_id414217.txt
6. adige20041008_id414218.txt
7. adige20041008_id414219.txt
8. adige20041008_id414220.txt
9. adige20041008_id414221.txt
10. adige20041008_id414222.txt
11. adige20041008_id414223.txt
12. adige20041008_id414224.txt
13. adige20041008_id414225.txt
14. adige20041008_id414226.txt
15. adige20041008_id414227.txt
16. adige20041008_id414228.txt
17. adige20041008_id414229.txt
18. adige20041008_id414230.txt
19. adige20041008_id414231.txt
20. adige20041008_id414232.txt
21. adige20041008_id414233.txt

Total: 191 files

10. APPENDIX D: INDEX OF THE TEMPORAL EXPRESSIONS OCCURRING IN THIS DOCUMENT.

24 anni.....	<i>See</i> anni	futuro.....	4, 9, 13
24 ore su 24.....	13	gennaio.....	4
a breve.....	9	già.....	4
a suo tempo.....	<i>See</i> tempo	giornalmente.....	4
adesso.....	4, 9	giornata.....	4, 8
al massimo.....	11	giorni scorsi (nei).....	13
alba.....	4, 10	giorno.....	4, 5, 13
all'anno.....	12	giorno dopo giorno.....	13
all'inizio di/del.....	11	ieri.....	4, 9
all'ora.....	12	ieri pomeriggio.....	10
alla fine di/del.....	11	il primo di/del.....	5
alle 21.....	7	immediatamente.....	4, 9
allora.....	4	infrasettimanale.....	12
almeno.....	11	intorno.....	11
ancora.....	4	invernale.....	4, 5, 12
anni.....	8, 12	inverno.....	4, 10
anni '90.....	7	istantaneamente.....	9
anno.....	4	istante.....	4
anno scolastico.....	12	lunedì.....	4
annualmente.....	4	mai.....	4
autunnale.....	4, 5, 12	martedì.....	4, 8
autunno.....	4, 10	marzo.....	4
Capodanno.....	4, 12	mattina.....	4, 10
circa.....	11	mattinata.....	4, 10
decade.....	4	meno di.....	10
domani.....	4, 9	mensilmente.....	4
dopo.....	4	mese.....	4
dopo domani.....	6	metà.....	11
dopodomani.....	4	mezzanotte.....	4
due volte a settimana.....	12	mezzodì.....	10
episodio.....	4	mezzogiorno.....	4, 10
era.....	4	millennio.....	4
estate.....	4, 9, 10	minuto.....	4
estivo.....	4, 5, 12	momento (al).....	9
ex.....	4	Natale.....	4, 12, 13
febbraio.....	4	nei primi.....	11
Ferragosto.....	4, 12	nei prossimi 4 anni.....	8
fin d'ora.....	6	nel tardo.....	11
fine settimana.....	4, 10	non meno di.....	10
finora.....	4, 9	non più di.....	10
frequente.....	4	nottata.....	10
frequentemente.....	4	nottate.....	12

notte.....	4, 10	questo momento (in)	9
occasione.....	4	recente	4, 5, 12
oggi	4, 7, 8, 9	recentemente	4, 9
Ogni mese	11	scorso	13
oltre	10	scorso weekend (lo)	<i>See weekend</i>
ora	4, 9	secolo	4, 5
ordine del giorno	5	semestre.....	4
ore	8	sempre	4
ore e ore.....	13	sera	4, 10
Pasqua	4, 12	serata	10
passato.....	4, 9, 13	serate	12
per ora	6, 9	Sessanta.....	5
periodi	5, 12	sette giorni fa.....	7
periodo	4, 5, 12	settimana	4
più di	10	settimanale	4, 5, 12
più tardi	9	stagione	4, 12
poi	4	stagione invernale	<i>See stagione</i>
pomeriggi	12	tardi	4
pomeriggio	4, 10	tempi	5
precedentemente	9	tempi recenti.....	5
presente	4, 9	tempo.....	5
presto.....	4	tutti i venerdì	11
prima	4, 9	tuttora	9
primavera	4, 9	ultimi tempi.....	5
primaverile	4, 5, 12	un paio.....	11
prossima settimana (la)	8	una decina	11
prossime settimane (nelle)	13	venerdì.....	8
prossimo.....	4, 5, 12, 13	vigilia	4
prossimo anno (il)	5	Vigilia di Natale.....	4, 12
questi giorni (in).....	9	weekend	4, 10

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., Wilson G. (April 2004) “*TIDES 2003 Standard for the Annotation of Temporal Expressions*”: http://timex2.mitre.org/annotation_guidelines/2003_timex2_standard_v1_3.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