LOCAL ENTITY DETECTION AND RECOGNITION

ANNOTATION FOR EVALITA 2009

Bernardo Magnini*, Emanuele Pianta*,
Manuela Speranza*, Valentina Bartalesi Lenzi**, and Rachele Sprugnoli**
FBK-irst, Centro per la Ricerca Scientifica e Tecnologica
Povo 38100 (Trento) Italy
** { magnini | pianta | manspera}@fbk.eu*

*** CELCT, Center for the Evaluation of Language and Communication Technologies*
Povo 38100 (Trento) Italy
{bartalesi | sprugnoli} @celct.it

February 2009

TABLE OF CONTENTS

ABSTRACT

This document reports on entities and mentions annotation, for the Evalita 2009 evaluation campaign, based on the Italian Content Annotation Bank (I-CAB).

I-CAB[1], developed at CELCT in conjunction with FBK-irst, is intended as a reference work both for the annotation methodology and for the automatic detection and recognition of different types of entities (i.e. person, organizations, locations and geo-political entities) in Italian. The benchmark follows the ACE[2] standards but with some modifications to adapting them to the specific morphosyntactic features of Italian.

For the purpose of this evaluation campaign, we simplified the I-CAB annotation scheme in order to conform the Local Entity Detection and Recognition task in Evalita to the one developed in the ACE program evaluation.

1. THE ENTITY DETECTION TASK

Entity Detection is one of the tasks of the ACE (*Automatic Content Extraction*)[3] program. It "requires that selected types of entities mentioned in the source data be detected, their sense disambiguated, and that selected attributes of these entities be extracted and merged into a unified representation for each entity" (LDC 2005, p. 4).

The ACE guidelines crucially distinguishes between entities and entity mentions:

- "An entity is an object or set of objects in the world." (LDC 2005, p. 4)
- "A mention is a [textual] reference to an entity. Entities may be referenced in a text by their name, indicated by a common noun or noun phrase, or represented by a pronoun" (LDC 2005, p. 4).

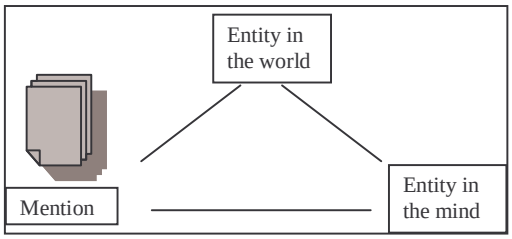
As shown above, the official definition of entity is 'object in the world'. However, in the ACE-LDC guidelines, this term is sometimes used in a way that is not compatible with this definition. Following the ACE-LDC guidelines, for example, the entity *10.000 persone* in *10.000 persone hanno partecipato alla sfilata* (= *10,000 people took part in the parade*), should be classified as underspecified as the exact number can not be quantified. In this case, it makes more sense to say that what is not quantified is actually the mental representation of it (entity in the mind). For the sake of clarity, when we refer to entities in general in this report, if not otherwise specified, we prefer to consider them as mental representations.

[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.

[2] Automatic Content Extraction Program.

[3] http://www.nist.gov/speech/tests/ace/index.htm or http://www.ldc.upenn.edu/Projects/ACE/

Figure 1: Semantic triangle of ENTITIES and Entity MENTIONS



According to the ACE-LDC guidelines, annotators should tag all mentions of each entity within a document; for each mention, they identify the maximal extent of the string that describes the entity and label the head of the mention. Annotators also group co-referring mentions (i.e. all mentions within a text which refer to the same entity).

Mentions can be nested; that is, a mention can contain mentions of other entities or even embedded mentions of the same entity.

Entities are classified from the semantic point of view, so we have different semantic types (e.g. persons, organizations, locations, etc.), subtypes (e.g. persons of subtype individual, group, etc., organizations of subtype sport, commercial, educational, etc.) and reference classes (e.g. specific referential, underspecified, etc.).

Mentions, on the other hand, are classified according to syntactic categories (e.g. proper names, common nouns, pronouns, etc.), and to metonymic versus literal style. In addition, for Geo-Political Entity mentions, the Role attribute can be annotated. Table 1 presents the complete hierarchy of annotation categories for entities and mentions.

Annotation categories may appear in the report in an abbreviated form (i.e. only the part in upper case).

Table 1: Annotation Categories

- ENTITY
  - Semantic TYPE
    - Semantic SUBTYPE
  - Reference CLASS
- MENTION
  - TYPE
  - METONYMY_MENTION
  - Role (only for Geo-Political Entity MENTIONS)

1.1 The annotation tool

For the annotation of I-CAB we have chosen Callisto, a freely distributed annotation tool developed at the MITRE Corporation. It supports linguistic annotation of textual sources for any Unicode-supported language and accepts files encoded as UTF-8, US-ASCII and several other character encodings. Callisto is written in Java, taking advantage of its portability and language support; it has been built with a modular design and utilizes standoff-annotation, allowing for unique tag-set definitions and domain dependent interfaces. Stand-off annotation support allows for many different annotation tasks to be represented. For the annotation of ENTITIES we have used the ACE Event task.

1.2 Notational conventions

All our examples are in italics. We have two notational conventions: a short form (which gives only information about the extent and head of MENTIONS) and an extended form.

Short notation:

The MENTION is enclosed in brackets and the head is underlined; when the head is the same as the extent, the underlining is omitted.

[*Un altro partecipante*] *ha testimoniato al processo*

[*Giovanni*] *è sposato da anni*

Extended notation:

MENTIONS are still enclosed in brackets and heads are underlined. For each MENTION, its TYPE is also provided.

[*Marco*] *è andato al cinema.*

TYPE=NAM

If an example contains more than one MENTION (of the same ENTITY or of different ones), we identify each entity and mention with a progressive index: E-1, E-2, etc., for ENTITIES; m-1, m-2, etc., for MENTIONS.

Two mentions referring to the same ENTITY:

[*Marco*]$_{E\text{-}1\ m\text{-}1}$ *è andato al cinema con la* [*propria*]$_{E\text{-}1\ m\text{-}2}$ *macchina.*

E-1 m-1 TYPE=NAM
m-2 TYPE=PRO

Two mentions referring to different ENTITIES:

[[*I figli di* [*Marco*]$_{E\text{-}1\ m\text{-}1}$]$_{E\text{-}2\ m\text{-}1}$ *sono andati al cinema.*

E-1 m-1 TYPE=NAM
E-2 m-1 TYPE=NOM

In some examples we add values of ENTITY attributes (e.g SUBTYPES and CLASSES); as far as metonymy is concerned, we only specify when METONYMY_MENTION=TRUE.

[*Il Siena*]$_{E\text{-}1\ m\text{-}1}$ *è* [*una squadra forte*]$_{E\text{-}1\ m\text{-}2}$

E-1 (ORG-Spo) (SPC) m-1 TYPE=NAM
METONYMY_MENTION=TRUE
m-2 TYPE=NOM

2. GUIDELINES FOR THE ANNOTATION OF ENTITY ATTRIBUTES

In this Section we describe annotation guidelines for Semantic TYPE, SUBTYPE and CLASS attributes.

2.1 Semantic TYPE

In the ACE project, seven semantic TYPES of ENTITIES were identified:

- Person: a single individual or a group of humans.
- Organization: corporations, agencies, and other groups of people defined by an established organizational structure.
- Geo-Political Entity: geographical regions defined by political and/or social groups (e.g. a nation, its region, its government, or its people).
- Location: geographical ENTITIES such as geographical areas and landmasses, bodies of water, and geological formations.
- Facility: buildings and other permanent man-made structures and real estate improvements.
- Vehicle: physical devices primarily designed to move an object from one location to another.
- Weapon: physical devices primarily used as instruments for physically harming or destroying other ENTITIES.

In I-CAB we have restricted our annotation to four of the semantic TYPES defined above:

- Person (PER)
- Organization (ORG)
- Geo-Political Entity (GPE)
- Location (LOC)

2.2 Semantic subtypes

For each semantic TYPE, various SUBTYPES are defined, that provide further semantic information.

2.2.1 Person

Individual: when the ENTITY refers to a single person.

[*Quella <u>ragazza</u>*]$_{E\text{-}1\ m\text{-}1}$ *si chiama* [*Francesca*]$_{E\text{-}1\text{-}m\text{-}2}$
E-1 (PER-Indiv.) m-1 TYPE=NOM
m-2 TYPE=NAM

[*Ciampi*]$_{E\text{-}1\ m\text{-}1}$ *è nato nel 1920*
E-1 (PER-Indiv.) m-1 TYPE=NAM

Group: when the ENTITY refers to more than one person. This includes family names and ethnic and religious groups that do not have a formal organization unifying them.

[*Quei bambini*]$_{E\text{-}1\text{-}m\text{-}1}$ *sono disubbidienti*
E-1 (PER-Group)m-1 TYPE=NOM

More examples:

[*Gli avvocati*] *non lavorano gratis*
[*I Rossi*] *sono originari di Pisa*
[*La mia famiglia*] *abita in centro*
[*Gli arabi*] *parlano una lingua appartenente alla famiglia semitica*
[*I Cristiani*] *professano una religione monoteista*

Indefinite: when it is not possible to judge from the context whether the ENTITY refers to one or more than one person.

Non sappiamo ancora [*chi*]$_{E\text{-}1\text{-}m\text{-}1}$ *l'abbia rubato*
E-1 (PER-Indef.) m-1 TYPE=PRO

2.2.2 Organization

Government: government organizations are those that are of, relating to, or dealing with the structure or affairs of government, politics, or the state**.** Also military organizations that are connected to the government of a state are to be tagged with this SUBTYPE.

[*Il Senato*]$_{E\text{-}1\ m\text{-}1}$ *è* [*un organismo politico*]$_{E\text{-}1\ m\text{-}2}$
E-1 (ORG-Gov.) m-1 TYPE=NAM
m-2 TYPE=NOM

L' auto è stata posta sotto sequestro [*dalla Guardia di Finanza*]$_{E\text{-}1\ m\text{-}1}$
E-1 (ORG-Gov.) m-1 TYPE=NAM

Please notice that the entire government of a Geo-Political Entity is excluded from this SUBTYPE and should be tagged GPE.ORG (see Appendix A).

Commercial: commercial organizations are those that are focused primarily upon providing ideas, products, or services for profit.

Nel maggio 1963 [*la società*]$_{E\text{-}1\ m\text{-}1}$ *cambiò la denominazione sociale*
E-1 (ORG-Com.) m-1 TYPE=NOM

More examples:

[*Alitalia*] *ha un rialzo del 2,25%*
Sono Rita e Tamara che gestiscono [*il bar*]

Educational: institutions that are focused primarily upon the promulgation of learning/education.

[*L'ateneo*] $_{E\text{-}1\ m\text{-}1}$ *avrà un nuovo rettore*
E-1 (ORG-Edu.) m-1 TYPE=NOM

Entertainment: entertainment organizations are those whose primary activity is entertainment.

[*I Rem*] $_{E\text{-}1\ m\text{-}1}$ *presenteranno il* [*loro*] $_{E\text{-}1\text{-}m\text{-}2}$ *nuovo album*
E-1 (ORG-Ent.) m-1 TYPE=NAM

m-2 TYPE=PRO

Non Governmental: several types of organizations whose main role is advocacy, charity or politics (in a broad sense):

- (Para-)Military organizations

  Nel muro di silenzio [*delle nuove Brigate Rosse*] *si è finalmente aperto uno squarcio*

- Political parties

  Fini ([*AN*]*) si è incontrato ieri con Bertinotti (*[*Rifondazione Comunista*]*)*

- Professional Regulatory

  È stato appena eletto il nuovo presidente [*dell'Ordine degli Avvocati*]

- Charitable and no-profit organizations

  [*Le associazioni di volontariato*] *promuoveranno l'iniziativa*

  [*Gli « Amici della Neonatologia Trentina »*] *organizzano un incontro in mattinata*

- International Regulatory and Political Bodies

  Il cardinale sottolinea il ruolo che deve avere [*l'Onu*]

- Labor and industrial unions

  Arrivato l' ok [*dei sindacati*] *al piano industriale 2005 - 2008*

  È andato a buon fine l'accordo con [*Assoartigiani*]

Media: media organizations are those whose primary interest is the distribution of news or publications. They can be private or public.

Arrestato l' inviato di [*Al Jazira*] *E-1 m-1*
E-1 (ORG-Med.) m-1 TYPE=NAM

Religious: religious organizations are those that are primarily devoted to issues of religious worship.

[*L' Arcidiocesi di Trento*] *E-1 m-1 propone un nuovo convegno ecumenico*
E-1 (ORG-Rel.) m-1 TYPE=NAM

Era presente anche Igor Vyzhanov [*del Patriarcato Ortodosso di Mosca*] *E-1 m-1*
E-1 (ORG-Rel.) m-1 TYPE=NAM

Medical Science: medical science organizations are those whose primary activity is the application of medical care or the pursuit of scientific research. They can be private or public.

La rassegna è stata promossa [*dal CNR*] *E-1 m-1*
E-1 (ORG-Med Sci.) m-1 TYPE=NAM

Il giovane venne soccorso dall' elicottero [*del 118*] *E-1 m-1*
E-1 (ORG-Med Sci.) m-1 TYPE=NAM

Sports: sports organizations are those that are primarily concerned with participating in or governing organized sporting events. They can be professional, amateur, or scholastic.

Ho colto al volo l' opportunità di giocare in [*A1*] *E-1 m-1*
E-1 (ORG-Spo.) m-1 TYPE=NAM

More examples:
[*Juve*] - [*Roma*] *1 – 1*

[*Luna Rossa*] *è in testa alla classifica*
È stato designato il nuovo CT [*della squadra*]

2.2.3 Geo-Political Entities

Continent: taggable MENTIONS of the ENTITIES of any of the seven continents (i.e. North America, South America, Antarctica, Europe, Asia, Africa, and Australia).

[*L'Asia*]$_{E\text{-}1\ m\text{-}1}$ *è* [*il continente più esteso*]$_{E\text{-}1\ m\text{-}2}$
E-1 (GPE-Cont.) m-1 TYPE=NAM
m-2 TYPE=NOM

Nation: taggable MENTIONS of the ENTITIES of any nation.

Nell'incidente sono stati coinvolti due turisti appena arrivati [*dalla Germania*]$_{E\text{-}1\ m\text{-}1}$
E-1 (GPE-Nat.) m-1 TYPE=NAM

More examples:
La governabilità [*della nostra nazione*] *è in crisi*
Chi sarà il prossimo presidente [*degli USA*]?

State or Province: taggable MENTIONS of the ENTITIES of any state, province, or canton of any nation.

[*Gli Stati Uniti*] $_{E\text{-}1\ m\text{-}1}$ *sono composti da* [*50 stati*] $_{E\text{-}2\ m\text{-}1}$
E-1 (GPE-Nat.) m-1 TYPE=NAM
E-2 (GPE- State or Prov.) m-1 TYPE=NOM

Italian regions (e.g. Toscana/*Tuscany*) and provinces (e.g. Provincia di Firenze/*Province of Florence*) belong to this SUBTYPE.

Sono stati contestati i finanziamenti concessi [*dalla Provincia Autonoma*] $_{E\text{-}1\ m\text{-}1}$
E-1 (GPE-State or Prov.) m-1 TYPE=NAM

More examples:
Molti ettari del territorio [ligure] sono andati a fuoco nell'incendio di ieri
[*La Florida*] *ha un clima splendido*

County or District: Taggable MENTIONS of the ENTITIES of any county, district, prefecture, or analogous body of any state/province/canton.

[*Le contee*] $_{E\text{-}1\ m\text{-}1}$ *sono solitamente divise in* [*diversi distretti*] $_{E\text{-}2\ m\text{-}1}$
E-1 (GPE-County or Dist.) m-1 TYPE=NOM
E-2 (GPE-County or Dist.) m-1 TYPE=NOM

Italian municipalities (e.g. Comune di Firenze/*Municipality of Florence*) and the so called "comunità montane" (e.g. Comunità Montana Valle di Fiemme) are annotated as County-or-District.

Nuova iniziativa [*del comune di Trento*] $_{E\text{-}1\ m\text{-}1}$ *per la raccolta differenziata*
E-1 (GPE-County or Dist.) m-1 TYPE=NAM

[La comunità montana Valle di Fiemme] E-1 m-1 ha organizzato un'escursione
E-1 (GPE-County or Dist.) m-1 TYPE=NAM

Population Center: Taggable MENTIONS of the ENTITIES of any GPE below the level of County-or-District.

[*Un intero vilaggio*] $_{E\text{-}1\ m\text{-}1}$ *sopravvive grazie al nostro potabilizzatore*

E-1 (GPE-Population Cent.) m-1 TYPE=NOM

Grande afflusso di turisti a [*Pisa*] $_{E\text{-}1\ m\text{-}1}$
E-1 (GPE- Population Cent.) m-1 TYPE=NAM

The lowest levels of Italian administrative division (e.g. localities and the so called "circoscrizioni") are annotated as Population-Center.

Depedri è il nuovo presidente [*della circoscrizione5*] $_{E\text{-}1\ m\text{-}1}$
E-1 (GPE- Population Cent.) m-1 TYPE=NAM

Cluster: named groupings of GPEs that can function as political ENTITIES.
[*L'Unione Europea*] $_{E\text{-}1\ m\text{-}1}$ *condanna la discriminazione*
E-1 (GPE-Cluster) m-1 TYPE=NAM

Other examples are: *Medio Oriente, Europa dell'Est, Sud-est Asiatico* and *America Latina*.

Special: a closed set of GPEs for which the conventional labels do not straightforwardly apply (e.g. *Autorità Palestinese, Riserve Indiane*).
Sotto analisi la situazione [*dell'attuale Palestina*] $_{E\text{-}1\ m\text{-}1}$
E-1 (GPE-Special) m-1 TYPE=NAM

Former GPEs, i.e. *Impero Austro-Ungarico* or *Repubblica Socialista Federale della Jugoslavia*, are also annotated as GPE-Special.

2.2.4 Location

Address: a location denoted as a point such as in a postal system or abstract coordinates.
L'ufficio postale di [*Piazza Vicenza*] $_{E\text{-}1\ m\text{-}1}$ *è chiuso per lavori*
E-1 (Address) m-1 TYPE=NAM

Il negozio si è spostato [*al n.25, via Banchi di Sotto*], $_{E\text{-}1\ m\text{-}1}$
E-1 (Address) m-1 TYPE=NAM

Boundary: a one-dimensional location such as a border between GPE's or other locations.
La guerriglia continua [*sul confine*] $_{E\text{-}1\ m\text{-}1}$
E-1 (Boundary) m-1 TYPE=NOM

Si sono incontrati in un villaggio di [*frontiera*] $_{E\text{-}1\ m\text{-}1}$
E-1 (Boundary) m-1 TYPE=NOM

Celestial: a location which is otherworldly or entire-world-inclusive.
L'eclisse di [*luna*] $_{E\text{-}1\ m\text{-}1}$ *è stata spettacolare*
E-1 (Celestial) m-1 TYPE=NAM

Rappresentanti da [*tutto il mondo*] $_{E\text{-}1\ m\text{-}1}$ *sono giunte oggi per la conferenza*
E-1 (Celestial) m-1 TYPE=NOM

Water-Body: bodies of water, natural or man-made.
Sono in secca [*il Po*] $_{E\text{-}1\ m\text{-}1}$ *e* [*altri fiumi minori*] $_{E\text{-}2\ m\text{-}1}$
E-1 (Water-Body) m-1 TYPE=NAM
E-2 (Water-Body) m-1 TYPE=NOM

More examples:

La piena [*del lago artificiale*] *fa paura a molti*
[*Il ghiacciaio*] *si sta sciogliendo rapidamente*

Land-Region-natural: geologically or ecosystemically designated, non-artificial locations (e.g. valli/*valleys*, monti/*mountains*, colline/*hills*, isole/*islands*, penisole/*peninsulas*, deserti/*deserts*, boschi/*woods*, campagna/*countryside* e spiagge/*beaches*).

Spero di fare un'escursione [*sulle Alpi*] $_{E\text{-}1\ m\text{-}1}$
E-1 (Land-Region-nat.) m-1 TYPE=NAM

[*Le valli*] $_{E\text{-}1\ m\text{-}1}$ *si stanno ripopolando*
E-1 (Land-Region-nat.) m-1 TYPE=NOM

Region-International: Taggable locations that cross national borders.

Arrivano sempre nuovi immigrati [*dall'Africa settentrionale*] $_{E\text{-}1\ m\text{-}1}$
E-1 (Region-Inter.) m-1 TYPE=NOM

Vorrei andare a lavorare [*all'estero*] $_{E\text{-}1\ m\text{-}1}$
E-1 (Region-Inter.) m-1 TYPE=NOM

Chissà cosa pensano di noi [*oltreoceano*] $_{E\text{-}1\ m\text{-}1}$*!*
E-1 (Region-Inter.) m-1 TYPE=NOM

Region-General: Taggable locations that do not cross national borders.

Il governo propone nuovi finanzianti per lo sviluppo [*dell'Italia meridionale*]$_{E\text{-}1\ m\text{-}1}$
E-1 (Region-Gen.) m-1 TYPE=NOM

Il negozio ha aperto in [*un'altra zona della città*] $_{E\text{-}1\ m\text{-}1}$
E-1 (Region-Gen.) m-1 TYPE=NOM

Maltempo previsto domani [*nel nord-est*] $_{E\text{-}1\ m\text{-}1}$
E-1 (Region-Gen.) m-1 TYPE=NOM

2.3 Reference CLASSES

Reference CLASSES describe the kind of reference each ENTITY makes to something in the world.

Specific Referential (SPC)

An ENTITY is SPC when it refers to a particular, unique object (or set of objects), whether or not the author or reader is aware of the name of the ENTITY or its anchor in the real world. For example, in the sentence *Ho visto Francesca passeggiare con un bambino*, both *Francesca* and *un bambino* are to be annotated as SPC (in the first case, the author and the reader are aware of the name of the ENTITY, in the second they are not).

Generic Referential (GEN)

An ENTITY is GEN when it does not refer to a particular, unique object (or set of objects) but a general type or class of objects. GEN ENTITIES are typically used in laws and rules.

[*La Camera*] *è composta da 630 membri*
[*Gli avvocati*] *non lavorano gratis*

Co-reference between GEN ENTITIES is generally admitted:

[*Un lago*]$_{E\text{-}1\ m\text{-}1}$ *è* [*una massa d'acqua dolce raccolta nelle cavità terrestri*]$_{E\text{-}1\ m\text{-}2}$
E-1 m-1 TYPE=NOM
m-2 TYPE=NOM

Under-specified Referential (USP)

It is a non-specific, non-generic reference. It includes:

a) quantified NP's in modal, future, conditional, hypothetical, negated, uncertain, question contexts (in all cases the ENTITY/ENTITIES referenced cannot be verified, regardless of the amount of "effort");
Non so [*quanti comuni*] *firmeranno l'accordo*
b) imprecise quantifications;
[*Tutti*] *sanno quando ci sarà il corteo.*
c) MENTIONS of a large number of ENTITIES where the actual members of the set are not identifiable and the number used is an estimate;
[*Oltre 10.000 scuole*] *hanno sottoscritto l'iniziativa*
d) impersonal and passive pronoun *si;*
[*Si*] *dice che cadrà molta neve* (impersonal, "*si*" means "people")
[*Si*] *vende carne* (passive)
e) NPs that the annotator cannot classify.

USP reference of type (a), (c), and (e) above all admit co-reference between each other. Co-reference does not occur, usually, between USP reference of type (b) and (d), with only one exception: co-reference is admitted when the imprecise quantification or the impersonal pronoun "si" co-refer with some pronominal element in the same clause. For example, in the sentences: ***ci si*** *può capire anche senza la guerra* / ***people*** *can understand* ***each other*** *without making war* and ***tutti si*** *potrebbero capire anche senza la guerra* / ***everybody*** *could understand* ***each other*** *without making war*.

[*Tutti*]$_{E\text{-}1\ m\text{-}1}$ [*si*]$_{E\text{-}1\ m\text{-}2}$ *potrebbero capire anche senza la guerra*
E-1 (USP) m-1 TYPE=PRO
m-2 TYPE=PRO

Negatively Quantified (NEG)

An ENTITY belongs to the referential CLASS NEG when it has been quantified so as to refer to the empty set of the type of object mentioned.

[*Nessun paese*] $_{E\text{-}1\ m\text{-}1}$ *è stato sanzionato*
E-1 (NEG) m-1 TYPE=NOM

NEG ENTITIES can not co-refer:

[*Nessuno*] $_{E\text{-}1\ m\text{-}1}$ *ha chiamato,* [*nessuno*] $_{E\text{-}2\ m\text{-}1}$ *ha risposto*
E-1 (NEG) m-1 TYPE=PRO
E-2 (NEG) m-1 TYPE=PRO

The only exception is when the NEG ENTITY co-refers with some pronominal element in the same clause.

[*Nessuno*] $_{E\text{-}1\ m\text{-}1}$ [*si*] $_{E\text{-}1\ m\text{-}2}$ *è ferito nell'incidente*
E-1 (NEG) m-1 TYPE=PRO
m-2 TYPE=PRO

2.4 Mapping between reference CLASSES and semantic SUBTYPES

Organization, Geo-Political and Location ENTITIES can be of any semantic SUBTYPE, independently of their reference CLASSES.

Le elezioni in [*Italia*] *saranno a Giugno* (SPC-Nation)
[*Le scuole*] *chiudono a Giugno* (GEN-Edu.)
Non so quale sia [*l'ospedale*] *migliore della zona* (USP-Med Sci.)
[*Nessuna montagna*] *è più alta!* (NEG- Land-Region-nat.)

For what concern the Person ENTITIES, SPC ENTITIES can be of any semantic SUBTYPE. NEG ENTITIES have semantic SUBTYPE indefinite, as an empty set has no number, whereas all GEN ENTITIES have SUBTYPE group (see Table 2). As for USP ENTITIES, we distinguish between imprecise quantifications, estimates, impersonal and reflexive pronoun 'si', and quantified NP's in modal, future, conditional, hypothetical, negated, uncertain, question contexts.

Table 2: Mapping between reference CLASSES and semantic SUBTYPES for Person ENTITIES

Reference CLASS		**Semantic SUBTYPE**
SPC		any
GEN		PER-Group
NEG		PER-Indefinite
USP	a) modal, future, etc., context	any
	b) imprecise quantifications	PER-Group
	c)estimates	PER-Group
	d) impersonal and reflexive 'si'	PER-Indefinite

Here are some examples of CLASS-SUBTYPE combinations related to the pronoun *chi*:

[*Chi partecipa all'incontro*] *ha diritto di votare* (GEN-Group)
[*Chi ha votato contro*] *lo ha fatto per le ragioni più disparate* (SPC-Group)
Non riesco a immaginare [*chi*] *possa aver votato contro!* (USP-Indefinite, because one or more people may have voted against)

3. GUIDELINES FOR THE ANNOTATION OF ENTITY MENTIONS

In this Section we describe annotation guidelines for Entity MENTIONS.

3.1 MENTION extent and MENTION head rules

For each MENTION, we record its full *extent*. The *extent* of a MENTION consists of the entire nominal phrase including all modifiers, prepositional phrases and relative clauses (e.g. *Ho incontrato* [*degli uomini* [*che*] *amano gli scacchi*]). In case of ambiguous structures, the extent annotated should be the maximal extent. In case of a discontinuous constituent, the extent goes to the end of the constituent, even if that means including tokens that are not part of the constituent (e.g *Ho incontrato* [*degli uomini, ieri al bar,* [*che*] *amano gli scacchi*]).

In addition, for each simple MENTION, the syntactic *head* is marked (e.g: [*Un altro partecipante importante*]). In most cases, the syntactic head of nominal phrases consists of a single word.

We can have heads composed of more than one word in two cases:

(i) proper names: the whole proper name is considered to be the head of the nominal phrase (i.e. both first and family name);

[*Carlo Azeglio Ciampi*] *è toscano*
[*Il saggio Carlo Azeglio Ciampi*] *è toscano*

(ii) expressions whose meaning has a certain degree of non-compositionality: the whole expression is considered to be the head. When in doubt, annotators refer to a reference dictionary (De Mauro 2000) and annotate the whole expression as head if it is recorded as idiomatic expression.

Giovanni è proprio [*un uccello del malaugurio*]

3.1.1 Complex Construction

Appositional constructions (e.g. [*La cantante Madonna*]), appositional constructions with relatives (e.g. [*La cantante Madonna che è in tour*]), and conjunctions (e.g. [*Madonna e Prince*], see Table 3 for other types of conjunction constructions), are complex constructions where the extent rules for simple MENTIONS are hard to apply. Each complex construction has special extent rules and simple MENTIONS within the extent of complex ones are further annotated. According to the ACE-LDC guidelines it is not necessary to annotate heads of complex constructions. However, the annotation tool we have chosen, i.e. Callisto, requires that every MENTION has a head so we have decided to annotate the whole extent as head.

These complex constructions contain nested MENTIONS, in the sense that the different parts of them are also annotated (e.g. *cantante* and *Madonna* in the first

example, *cantante*, *Madonna* and *che* in the second, and *Madonna* and *Prince* in the last one are independent MENTIONS)[4].

Table 3 Some conjunction constructions

x e y	*x and y*
x, y e z	*x, y and z*
tra x e y	*between x and y*
con x e y	*with x and y*
x o y	*either x or y*
x, y o z	*x, y or z*
con x e con y	*with x and with y*
x ma anche y	*x but also y*
x, y ma anche z	*x, y but also y*
della *x* e della *y*	*of the x and of the y*

3.2 Syntactic categories of Entity MENTIONS: TYPES

In Evalita 2009 three syntactic TYPES are evaluated: NAM (proper name), NOM (quantified nominal constructions), and PRO (pronoun).

NAM (Names): proper nouns and nicknames.

- Person ENTITIES

  [*Napolitano*] *è di origine campana*
  [*Pinturicchio*] *sta giocando bene*

- Organization ENTITIES

  [*La Microsoft Corporation*] *ha sede a Redmond, USA*
  [*Il Carroccio*] *ha molti sostenitori nell'Italia Settentrionale*

The head of NAM ORGs with composite names (e.g. “Vigili del Fuoco”, “protezione civile”) is identified by the whole name. In the case where they are mentioned using abbreviated forms (i.e. a part of the name is omitted), they are also annotated as NAM.

[*I Vigili del Fuoco*] *dipendono dal Ministero dell’Interno*
E-1 m-1 TYPE=NAM

[*I vigili*] *sono stati subito allertati dopo la fuga di gas*
E-1 m-1 TYPE=NAM

Mentions of foreign organizations have been annotated as proper nouns (TYPE=“NAM”) if they were the literal translation of the original name, whereas they

[4] Please notice that our annotation differs from the ACE-LDC guidelines where conjunctions of entities are not marked explicitly. In the case of *old men and women = uomini e donne anziani* (constructions of conjoined heads that share the same modifiers), they would annotate two simple MENTIONS with the same extent (*old men and women*), the first having *men* as head, the second having *women* as head. In the case of *mother and child* (constructions of conjoined heads without common modifiers), they would only annotate two distinct MENTIONS, *mother* and *child*.

have been annotated as nominal constructions (TYPE="NOM") if they were considered a cultural transposition of the concept expressed by the original word. Following this rule, *Dipartimento di Stato Americano* is annotated as NAM since it is the direct translation of *U.S. Department of State*. On the contrary *Polizia francese* is NOM because the official name of the French police is *Gendarmerie*.

- Geo-Political ENTITIES

In [*Italia*] *è iniziato il periodo elettorale*
[*Arezzo*] *si trova in* [*Toscana*]

The head of NAM GPEs with composite names (e.g. "Comune di Trento", "provincia di Bolzano") is identified by the whole name. In the case where they are mentioned using abbreviated forms (i.e. a part of the name is omitted), they are also annotated as NAM.

Nuovi finanziamenti da parte [*della Provincia di Trento*] $_{E\text{-}1\ m.1}$
E-1 m-1 TYPE=NAM

I finanziamenti [*della Provincia*]$_{E\text{-}1\ m\text{-}1}$ *soddisfano le organizzazioni in* [*Trentino*]$_{E\text{-}1\ m\text{-}2}$
E-1 m-1 TYPE=NAM
m-2 TYPE=NAM

- Location ENTITIES

Sono molti i turisti sul [*Lago di Garda*]
[*L'Himalaya*], *chiamata anche* [*Tetto del mondo*], *è lunga circa 2.400Km.*

Modifier MENTIONS that derive from the transformation of proper names are annotated with TYPE=NAM. Modifiers are those MENTIONS which occur in a modifying position, either before or after the word they modify. It is immaterial whether or not the word being modified is a taggable ENTITY.

Le ferrovie [*francesi*] $_{E\text{-}1\ m\text{-}1}$ *sono molto efficienti*
E-1 m-1 TYPE=NAM
Lo sport nazionale è il calcio

In the previous example "nazionale" is not taggable because it derives from the tranformation of "nazione", that it is not a proper name.

Similarly, "provinciale", "regionale" and "comunale" are taggable if the ENTITY from which they derive by transformation is of TYPE=NAM.

[*La Regione Toscana*] $_{E\text{-}1\ m\text{-}1}$ *ha stanziato nuovi finanziamenti. I contributi* [*regionali*] $_{E\text{-}1\ m\text{-}2}$ *verranno divisi tra vari enti.*
E-1 m-1 TYPE=NAM
m-2 TYPE=NAM

In Italia ci sono pochi parcheggi comunali gratuiti

NOM (Quantified Nominal Constructions): nouns quantified with determiners, quantifiers, or possessives.

- Person ENTITIES

[Il presidente dell'azienda] non è nel suo ufficio
[*Il mio vicino*] *è partito ieri*

- Organization ENTITIES

[*Le associazioni no-profit*] *sono sempre più numerose*
[*La mia azienda*] *sta assumendo nuove figure professionali*

- Geo-Political ENTITIES

[*Il Paese*] *ha bisogno di riforme*
[*La mia città*] *ha pochi parcheggi gratuiti*

- Location ENTITIES

[*Le periferie di Parigi*] *sono di nuovo sotto controllo*
Ci siamo trasferiti [*sulla collina*]

PRO (Pronouns): all pronouns and headless MENTIONS.

Headless MENTIONS are constructions in which the nominal head is not explicitly expressed. Following the ACE convention, we assign as head the rightmost modifier, i.e. the one which falls directly before the spot where the head would be. This is the case of superlative adjectives (when the noun they modify is elided), percentages, and numerals used as pronouns.

[*Il 30 %*]$_{E\text{-}1\ m\text{-}1}$ *è biondo*
E-1 m-1 TYPE=PRO

[*Due*] $_{E\text{-}1\ m\text{-}1}$ *sono europei e* [*tre*] $_{E\text{-}2\ m\text{-}1}$ *sono asiatici*
E-1 m-1 TYPE=PRO
E-2 m-1 TYPE=PRO

[*Il più forte*]$_{E\text{-}1\ m\text{-}1}$ *vincerà*
E-1 m-1 TYPE=PRO

Also enclitics that are attached at the end of a verb ([incontrarlo]/to meet him) or of the adverb 'ecco' ([Eccolo]!/Here he is!), and proclitics that precede the main verb and are attached to another word are annotated with TYPE="PRO". In this case we take all the word as extent and only the affix as head.

[*Aladino*]$_{E\text{-}1\ m\text{-}1}$ *non è il vero nome,* [*glielo*]$_{E\text{-}1\ m\text{-}2}$ *hanno appiccicato*
E-1 m-1 TYPE=NAM
m-2 TYPE=PRO

In partitive constructions, the first element (i.e. the part) that quantifies over the second element (i.e. the whole) is annotated with TYPE="PRO". An exception occurs with the Italian nouns *parte* (= *part*) and *maggioranza (= majority*) that are tagged as NOM whereas, their English correspondent are tagged as PRO in the ACE-LDC corpus.

[*alcuni* [*dei soci*]$_{E\text{-}2\ m\text{-}1}$]$_{E\text{-}1\ m\text{-}1}$ *sono in riunione*
E-1 (SPC) m-1 TYPE=PRO
E-2 (SPC) m-1 TYPE=NOM

[*parte* [*dei consiglieri*]$_{E\text{-}2\ m\text{-}1}$]$_{E\text{-}1\ m\text{-}1}$ *si è astenuta*
E-1 (SPC) m-1 TYPE=NOM
E-2 (SPC) m-2 TYPE=NOM

[*la* <u>*maggioranza*</u> [*dei* <u>*deputati*</u>] $_{E\text{-}2\ m\text{-}1}$] $_{E\text{-}1\ m\text{-}1}$ *era favorevole*

E-1 (SPC) m-1 TYPE=NOM
E-2 (SPC) m-2 TYPE=NOM

Group nouns (*gruppo, famiglia, equipe*) can occurr in pseudo-partitive expressions (e.g. *un gruppo di* meaning *un gruppo formato da*). In this case we don't have a partitive construction and the group expression (pseudo-part) co-refers with the pseudo-whole expression.

[*un* <u>*gruppo*</u> *di* [<u>*bambini*</u>]$_{E\text{-}1\ m\text{-}2}$] $_{E\text{-}1\ m\text{-}1}$ *gioca a calcio*

E-1 (SPC) m-1 TYPE=NOM
m-2 TYPE=NOM

- Person ENTITIES

[*Loro*] *non conoscono l'inglese*
[*Molti*] *non lo sanno*
[*Qualcuno*] *verrà*
[[*Suo*] $_{E\text{-}2\ m\text{-}1}$ <u>*figlio*</u>] $_{E\text{-}1\ m\text{-}1}$ *è nato nel 2000*
E-1 m-1 TYPE=NOM
E-1 m-2 TYPE=PRO

- Organization ENTITIES

[*Molte*] *sono* [*le* <u>*associazioni*</u> *italiane nel mondo*]
[*La* <u>*ditta*</u>] $_{E\text{-}1\ m\text{-}1}$ *sta aumentando i* [*suoi*] $_{E\text{-}1\ m\text{-}2}$ *profitti*
E-1 m-1 TYPE=NOM
E-1 m-2 TYPE=PRO

- Geo-Political ENTITIES

[*Molti*] *sono* [*i* <u>*comuni*</u> *interessati al progetto*]
[*La* <u>*città*</u>] $_{E\text{-}1\ m\text{-}1}$ *vuole espandere i* [*suoi*] $_{E\text{-}1\ m\text{-}2}$ *confini*
E-1 m-1 TYPE=NOM
m-2 TYPE=PRO

- Location ENTITIES

In questo periodo sono [*molte*] [*le* <u>*spiagge*</u> *affollate*]
[*Il* <u>*fiume*</u>] $_{E\text{-}1\ m\text{-}1}$ *sta per rompere i* [*suoi*] $_{E\text{-}1\ m\text{-}2}$ *argini*
E-1 m-1 TYPE=NOM
m-2 TYPE=PRO

Particular attention must be given to the annotation of the pronoun 'si', that can have several interpretations. Table 4 shows whether it is annotated or not in I-CAB and Table 5 reports some examples.

Table 4: Annotation of the different types of 'si'

Truly reflexive pronoun	YES
Reciprocal pronoun	YES
Impersonal pronoun	YES
Passive *si*	YES
Benefactive pronoun	YES
Pseudo-reflexive	NO

Table 5: Examples of annotations of 'si'

Intanto [*si*] *chiedono soldi*	PER (USP) TYPE=PRO	passive
[*Si*] *può ipotizzare un miglioramento*	PER (USP) TYPE=PRO	impersonal
È importante che [*si*] *prendano delle decisioni*	PER (USP) TYPE=PRO	passive
Giovanni [*si*] *è visto un bel film dopo cena*	PER (SPC) TYPE=PRO	benefactive
[*Si*] *è vista arrivare un mazzo di fiori*	PER (USP) TYPE=PRO	benefactive

3.3 METONYMY_MENTION

Mention style is either literal or metonymic. This is currently encoded in the APF as an attribute called METONYMY_MENTION, which is either true (for metonymic style of reference) or false (for literal style of reference). A metonymy occurs when the name of one entity is used to refer to another entity. We annotate the co-reference between the mention and the entity to which the mention refers in the context.

The METONYMY_MENTION recognition is optional in Evalita 2009.

a. City name for Sports Team

Names of GPEs used to refer to sport teams are annotated as ORG; the MENTION is co-referenced with the sport organization and it is marked as a metonymy:

[*La Russia*] $_{E\text{-}1\ m\text{-}1}$ *ha conquistato la medaglia d'oro.* [*La squadra*] $_{E\text{-}1\ m\text{-}2}$ *ha meritato la vittoria.*

E-1 (ORG-Spo.) m-1 TYPE=NAM
METONYMY_MENTION=TRUE
m-2 TYPE=NOM

b. Capital City or Government Seat names standing for Country's Government

In the case where the capital city is used to refer to the nation's government, its SUBTYPE should be Nation (i.e. GPE-Nation); in addition, it is marked as a metonymy:

[*Parigi*] *firmerà presto il patto*
E-1 (GPE-Nat.) m-1 TYPE=NAM
METONYMY_MENTION=TRUE

Government seats used to refer to the nation's government are to be tagged according to the ENTITY to which they refer.

[*Il Cremlino*] *ha fatto sapere che non firmerà il nuovo accordo sul gas*
E-1 (GPE-Nat.) m-1 TYPE=NAM
METONYMY_MENTION=TRUE

3.4 Geo-Political MENTION Roles

For each MENTION of a GPE ENTITY in the text, the role (PER, ORG, LOC, or GPE) invoked by its context can be tagged.

- GPE.ORG: *La Francia ha firmato l'accordo con la Germania*
- GPE.PER: *I francesi attendono con ansia le prossime elezioni*
- GPE.LOC: *Il G8 si è riunito ieri in Francia*
- GPE.GPE: *La Francia produce dell'ottimo vino*

The GPE Role recognition is optional in Evalita 2009.

GPE.ORG

We annotate with role ORG GPEs that are responsible for decisions to take military actions, for political communication events such as announcements, agreements, statements, denials, etc., as in the following examples:

[*La Cina*] $_{E\text{-}1\ m\text{-}1}$ *ha annunciato l'invio di nuove truppe al sud*
E-1 m-1 TYPE=NAM
ROLE=GPE.ORG

The role for governments should always be GPE.ORG:

Sarà fondamentale l'esito della prossima seduta [*del governo*] $_{E\text{-}1\ m\text{-}1}$
E-1 m-1 TYPE=NOM
ROLE=GPE.ORG

GPEs modifying government organizations reflect a relationship between the organizations and the governmental aspect of the GPE, so they are assigned a GPE.ORG role.

[*La Corte dei Conti* [*della Repubblica Italiana*] $_{E\text{-}2\ m\text{-}1}$] $_{E\text{-}1\ m\text{-}1}$ *ha aperto un'inchiesta*
E-1 (ORG) m-1 TYPE=NAM
E-2 (GPE) m-1 TYPE=NAM
ROLE=GPE.ORG

GPE.PER

Mentions referring to the entire population of a GPE, or to most of the population of a GPE, are treated as GPE.PER. On the other hand, if a MENTION refers just to a group of people, it has to be annotated as PER.

[*I Cubani*] $_{E\text{-}1\ m\text{-}1}$ *hanno aspettato per anni questo momento*
E-1 (GPE) m-1 TYPE=NOM
ROLE=GPE.PER

[*Gli Americani repubblicani*] $_{E\text{-}1\ m\text{-}1}$ *devono scegliere il nuovo candidato presidenziale*
E-1 (PER) m-1 TYPE=NOM

GPE.LOC

GPE.LOC is used when a MENTION of a GPE ENTITY primarily references the territory or geographic position of the GPE.

Ci troviamo a [*Pisa*] $_{E\text{-}1\ m\text{-}1}$ *alle 17:00*

E-1 m-1 TYPE=NAM
ROLE=GPE.LOC

GPEs with role LOC are generally introduced by location prepositions, such as “a”, “da”, “per”, “attraverso”, etc.

Arriverò a [*Roma*] *domani*
A [*Bologna*] *ci sarà una fiera*
Vengo da [*Milano*]
Il treno non passa per [*Vicenza*]

Geo-Political ENTITIES that indicate routes (e.g. air, train, bus and car routes), are annotated as GPE.LOC.

Sull’autostrada [*Salerno*] $_{E\text{-}1\ m\text{-}1}$ – [*Reggio Calabria*] $_{E\text{-}2\ m\text{-}1}$ *ci sono molti cantieri*
E-1 m-1 TYPE=NAM
ROLE=GPE.LOC
E-2 m-1 TYPE=NAM
ROLE=GPE.LOC

In the case where we have a GPE contained in another GPE (e.g. a city contained in a Region) in a nested form as in the following: [GPE1, [GPE2]], the role of the contained GPE depends on the context, whereas the role of the other is always LOC.

[*Orvieto,* [*Umbria*]$_{E\text{-}2\ m\text{-}1}$] $_{E\text{-}1\ m\text{-}1}$*, ha annunciato la costruzione di una nuova funivia*
E-1 m-1 TYPE=NAM
ROLE=GPE.ORG
E-2 m-1 TYPE=NAM
ROLE=GPE.LOC

GPE.GPE

GPE.GPE is used when a MENTION refers to an indistinct amalgam of more than one of the aspects of a GPE or when none of the roles stands out in the context.

In particular, GPE.GPE should always be used when military activities (e.g. invasions, bombings, etc). are considered to be acts carried out by and directed at entire nations and therefore are associated with GPEs. Both the aggressors and the victims in these cases are marked GPE.GPE.

Nel 1979 [*l’Unione Sovietica*] $_{E\text{-}1\ m\text{-}1}$ *invase* [*l’Afghanistan*] $_{E\text{-}2\ m\text{-}1}$

E-1 m-1 TYPE=NAM
ROLE=GPE.GPE

E-2 m-1 TYPE=NAM
ROLE=GPE.GPE

APPENDIX A: Special Cases

Titles of books, CDs and exhibitions

Names of people appearing in book/CD titles, in names of organizations or events, etc., are not to be annotated.

Articulated prepositions

According to the ACE-LDC guidelines, definite and indefinite articles are considered as part of the textual realization of an ENTITY, while prepositions are not. This is problematic for Italian articulated prepositions, where a definite article and a preposition are merged. We have decided that this type of prepositions should be included in the extent of the MENTION, so as to consistently include all the articles.

Annotation of difficult semantic TYPES of ENTITIES

a. Person versus Geo-Political ENTITIES

A MENTION that refers to the entire population of a Geo-Political ENTITY is annotated as GPE, rather than PER.

GPE: [*Gli Italiani*] *amano la pasta*
PER: [*Gli italiani* [*che*] *vivono in America*] *sono tantissimi*

NB: *Arabs* (when the word does not refer to the inhabitants of Saudi Arabia) are PER because they do not belong to a unified political structure.

b. Person versus Organization ENTITIES

Some ENTITIES can be annotated as PER or as ORG according to the context

Political Parties

In general, political parties are Entities of type ORG. However, when the text mentions the people belonging to the party instead of the party itself, we have two cases, depending on the context:

- if the text refers to the party as a whole, or to its directives, we have an organization
  [*I Verdi*] $_{E\text{-}1\ m\text{-}1}$ *hanno votato contro l'emendamento*
  E-1 (ORG-NoGov.) m-1 TYPE=NAM
- if the text refers to a specific subset of the people belonging to the party (especially if they behave somehow differently from the others) we have a person ENTITY
  [*I Verdi*] $_{E\text{-}1\ m\text{-}1}$ *che hanno abbandonato l'aula sono stati espulsi* [*dal partito*] $_{E\text{-}2\ m\text{-}1}$
  E-1 (PER-Group) m-1 TYPE=NOM
  E-2 (ORG-NoGov.) m-1 TYPE=NOM
  In this example, the text refers to the people of the green party who were present in the chamber and left it.

Police and similar organizations

If the text refers to police organizations as a whole, they are annotated as ORG with TYPE="NAM"; if instead the text refers to the single persons in a special activity, they are annotated as PER with subtype Group and TYPE="NOM". Please notice the MENTION head extent.

Cambio al vertice [*nei vigili del fuoco*] $_{E\text{-}1\ m\text{-}1}$
E-1 (ORG-NoGov.) m-1 TYPE=NAM

[*I vigili del fuoco*] $_{E\text{-}1\ m\text{-}1}$ *hanno fatto irruzione nell'abitazione*
E-2 (PER-Group) m-1 TYPE=NOM

Sport teams

Similarly, sport teams can be either organizations or person ENTITIES depending on the context. If the text refers to the team in general, its management or its administration, we have an organization; if instead it refers to the players, we have a person ENTITY.

La maglietta [*della Juventus*] $_{E\text{-}1\ m\text{-}1}$ *è in vendita nei negozi specializzati*
E-1 (ORG-Spo.) m-1 TYPE=NAM

[*La Juventus*] *ha giocato bene ieri*
E-2 (PER-Group) m-1 TYPE=NOM

Other specific cases

See Tables 6 and 7 for other specific cases.

Table 6: Special cases of Person

PERSON
(centro) destra, (centro) sinistra
cast
centristi
comunità
cristiani, ebrei, musulmani, buddisti, islamici, induisti
critica (cinematografica)
opposizione, minoranza, maggioranza
resistenza irachena
vertici (of a company)

Table 7: Special cases of Organization

ORGANIZATION
cattolici
CDA / organi di società
cori, orchestre, bande, accademie
dirigenza
Guardia di Finanza, Carabinieri, 118, Vigili del Fuoco
forze politiche
giuria
gruppi musicali (REM)

esecutivo
guardia nazionale irachena
militari fascisti
presidenza europea, presidenza olandese
Procura
resistenza islamica (Hamas)
sindacati, parti sociali

c. Person versus no ENTITY annotation

Names of people are not annotated if they refer to something which is not a person. For instance, in *la legge Bossi-Fini* or *il governo Prodi*, the names do not identify persons anymore (i.e. the relation with the person has been lost), but they are used as the names of the law or the government itself.

On the other hand, when a preposition is used, names of people are annotated as PER:

La legge di [[*Bossi*] *e* [*Fini*]]
Il governo di [*Prodi*]

In some cases it is difficult to decide whether an ENTITY should be annotated as PER or whether it should not be annotated at all. In these cases, annotators refer to the definition provided in the dictionary (De Mauro 2000): if the *genus* in the definition refers to people, we annotate it as PER, otherwise, it is not annotated at all.

Following the definitions contained in the dictionary, it has been decided to annotate *corteo*, *soccorsi*, *presenze*, *casta*, and *categoria* as PER and not to annotate *volanti della polizia*, *assemblea dei soci* and *tavolo delle trattative*. See the following definitions:

YES Corteo: "gruppo di persone che sfilano nel corso di una manifestazione pubblica: *un c. di manifestanti dimostrava in piazza, un c. di studenti, di lavoratori, un c. militare, di protesta*" (De Mauro 2000)

NO Assemblea: "riunione spec. numerosa per discutere questioni importanti di interesse comune: *indire, convocare, organizzare, sciogliere un'a.; l'a. di fabbrica*" (De Mauro 2000)

d. Organization versus Geo-Political ENTITIES

A MENTION that refers to the entire governing body of a GPE is annotated as GPE, rather than ORG. It is important to differentiate between a part of the government (the executive branch, the courts) and the entire governing body.

GPE: [*Il governo italiano*] *ha emanato una nuova legge*
ORG: [*Il Senato* [*che*] *si è riunito ieri*] *ha votato la fiducia*

e. Location versus no ENTITY annotation

Every physical object implies a location because the space that each physical object occupies is the "location" of that object. Viewed from a certain angle, for example, in sentences like *il cappotto è sotto il tavolo, ho un'idea nuova in testa*, "il tavolo" and "testa" become locations. However, none of these are taggable location expressions because they do not fall within any of the subclasses defined in section 3.1.

Compass points are not to be tagged when they serve as adjectives or refer to directions, as in *le truppe marciarono verso nord*. Compass points should only be tagged when they refer to sections of a region, as in *nel nord-est*.

Syntactic TYPE of complex constructions

The TYPE of any appositional construction is the same as the TYPE of the first MENTION of the appositional construction.

[[*Il Brasile*]$_{E\text{-}1\ m\text{-}1}$ [*nazione in via di sviluppo*]$_{E\text{-}1\ m\text{-}2}$]$_{E\text{-}1\ m\text{-}3}$ *cerca nuovi mercati*

E-1 m-1 TYPE=NAM
m-2 TYPE=NOM
m-3 TYPE=NAM

The TYPE of any conjunction construction is the same as the TYPE of the first MENTION of the conjunction construction.

[[*Marco*]$_{E\text{-}1\ m\text{-}1}$ *e* [*il figlio*]$_{E\text{-}2\ m\text{-}1}$]$_{E\text{-}3m\text{-}1}$ *sono bellissimi*

E-1 m-1 TYPE=NAM
E-2 m-1 TYPE=NOM
E-3 m-1 TYPE=NAM

Semantic TYPE and SUBTYPE of conjunction constructions

The semantic TYPE and SUBTYPE of non-uniform ENTITY groups for which it is impossible to chose a single semantic TYPE, are the same as the semantic TYPE and SUBTYPE of the first ENTITY of the conjunction construction.

Soddisfatti [[Cimoli]$_{E\text{-}1\ m\text{-}1}$ *e [sindacati]* $_{E\text{-}2\ m\text{-}1}$*]* $_{E\text{-}3\ m\text{-}1}$

E-1 (PER-Ind)	m-1	TYPE=NAM
E-2 (ORG-NonGov)	m-2	TYPE=NOM
E-3 (PER-Ind)	m-3	TYPE=NAM

How to deal with dashes, colons and brackets

Dashes, colons and brackets may be relevant for deciding the extent of MENTIONS. This is particular important to decide when a sequence of nominal phrases constitutes two different MENTIONS or a unique appositional construction. The following examples illustrate some of the guidelines we followed.

[*Quel monte*] $_{E\text{-}1\ m\text{-}1}$ *(posizionato più a nord) ha un'altezza considerevole*

E-1 m-1 TYPE=NOM

[*Quel monte*] $_{E\text{-}1\ m\text{-}1}$ *(posizionato più a nord), molto alto, è meta di molti turisti*

E-1 m-1 TYPE=NOM

[*Quel monte (posizionato più a nord)* [*che*] $_{E\text{-}1\ m\text{-}2}$ *è molto alto*] $_{E\text{-}1\ m\text{-}1}$ *è meta di molti turisti*

E-1 m-1 TYPE=NOM
m-2 TYPE=PRO

[*Quel monte*] $_{E\text{-}1\ m\text{-}1}$ *(*[*che*] $_{E\text{-}1\ m\text{-}2}$ *è molto alto) è meta di molti turisti*

E-1 m-1 TYPE=NOM

m-2 TYPE=PRO

[*Quel <u>monte</u>*] $_{E\text{-}1\ m\text{-}1}$ - [*che*] $_{E\text{-}1\ m\text{-}2}$ *è molto alto - è meta di molti turisti*
E-1 m-1 TYPE=NOM
m-2 TYPE=PRO

[*Il <u>centro</u>*] $_{E\text{-}1\ m\text{-}1}$ - [*<u>zona</u> storica*] $_{E\text{-}1\ m\text{-}2}$ *- è ricco di attrattive*
E-1 m-1 TYPE=NOM
m-2 TYPE=NOM

[*Il <u>centro</u>*] $_{E\text{-}1\ m\text{-}1}$ *(*[*<u>zona</u> storica*] $_{E\text{-}1\ m\text{-}2}$*) è ricco di attrattive*
E-1 m-1 TYPE=NOM
m-2 TYPE=NOM
[*Il <u>Bondone</u>*] $_{E\text{-}1\ m\text{-}1}$ *(2180 metri s.l.m)* [*<u>rilievo</u> montuoso alle porte di Trento*] $_{E\text{-}1\ m\text{-}2}$ *è meta di molti turisti*
E-1 m-1 TYPE=NAM
m-2 TYPE=NOM

[*Everest*] $_{E\text{-}1\ m\text{-}1}$ *(8844 metri s.l.m),* [*Aconcuaga*] $_{E\text{-}2\ m\text{-}1}$ *(6959 s.l.m.) e* [*McKinley*] $_{E\text{-}3\ m\text{-}1}$ *(6194 s.l.m)*
E-1 m-1 TYPE=NAM
E-2 m-1 TYPE=NAM
E-3 m-1 TYPE=NAM

“Circa” and “almeno”

These two adverbs are included in the MENTION extent.

[*Circa 10 <u>persone</u>*] *stavano aspettando in ufficio*
[*Almeno 5 <u>laghi</u>*] *risultano sicuramente inquinati*

APPENDIX B: Inter-annotator agreement

We have adopted the matching criteria of the ACE 2005 distributed scorer:

- an entity is detected by both annotators if they detect at least a mention of that entity;
- a mention is detected by both annotators if the mutual fractional head overlap is at least 30%;
- the maximum extent difference allowed for mentions to be declared an extent match is 4 characters.

Therefore, if one annotates [Savani e Vujevic sempre meglio] / *Savani and Vujevic always better* as a mention while the other restricts the extent to *Savani* e *Vujevic*, we have agreement in mention detection, but no extent match.

The kappa statistic does not account for nested annotations. As this phenomenon is extremely frequent in the case of PEs, we have chosen to calculate the Dice coefficient instead and limit the use of the kappa statistic to the assignment of attributes.

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)[5]

Results of the inter-annotator agreement for each TYPE of ENTITY are reported in the following sections.

Person ENTITIES

Inter-annotator agreement has been evaluated on the dual annotation of a subset of ten randomly chosen news stories for a total of 4,657 words.

- the Dice coefficient for person ENTITIES detection is 0.906;
- limited to the ENTITIES detected by both annotators, the Dice coefficient for MENTION detection is 0.951;
- limited to the ENTITIES detected by both annotators, the kappa statistic is 0.937 for SUBTYPE assignment (i.e. Group, Individual or Indefinite) and 0.734 for class assignment (this relatively low value is due to the high prevalence of the SPC class and to some mismatches in the USP and GEN classes);
- limited to the MENTIONS detected by both annotators we have a 3.7% of extent mismatch.

Organization ENTITIES

Inter-annotator agreement has been evaluated on the dual annotation of a corpus of ten randomly chosen news stories for a total of 3,405 words.

Results are as follows:

- the Dice coefficient for organization ENTITIES detection is 0.857;
- limited to the ENTITIES detected by both annotators, the Dice coefficient for MENTION detection is 0.845;

[5] Notice that the Dice coefficient has the same value of the F1 measure computed considering any of the two annotators as the reference.

- limited to the ENTITIES detected by both annotators, the kappa statistic is 0.970 for subtype assignment and 1 for class assignment;
- limited to the MENTIONS detected by both annotators we have a 3.7% of extent mismatch.

Geo-Political ENTITIES

Inter-annotator agreement has been evaluated on the dual annotation of a corpus of ten randomly chosen news stories for a total of 4,741 words.

In particular, we calculated the Dice coefficient for the detection of ENTITIES and ENTITY MENTIONS and we used the kappa statistic to the assignment of attributes.

Results are as follows:

- the Dice coefficient for Geo-Political ENTITIES detection is 1;
- limited to the ENTITIES detected by both annotators, the Dice coefficient for MENTION detection is 0.980;
- limited to the ENTITIES detected by both annotators, the kappa statistic is 1 for subtype assignment and 1 for class assignment;
- limited to the ENTITIES detected by both annotators, the kappa statistic is 0.965 for Role assignment
- limited to the MENTIONS detected by both annotators we have a 0% of extent mismatch (total agreement).

Location ENTITIES

Inter-annotator agreement has been evaluated on the dual annotation of a corpus of ten randomly chosen news stories for a total of 4,868 words.

- the Dice coefficient for Location ENTITIES detection is 0.957;
- limited to the ENTITIES detected by both annotators, the Dice coefficient for MENTION detection is 0.938;
- limited to the ENTITIES detected by both annotators, the kappa statistic is 1 for SUBTYPE assignment;
- limited to the ENTITIES detected by both annotators, the kappa statistic is 0 for class assignment;
- limited to the MENTIONS detected by both annotators we have a 0% of extent mismatch (total agreement).

Note that, for what concerns class assignment, the inter-annotator agreement was acceptable (only 1 ENTITY has been differently annotated) but the kappa statistic is 0 because of the prevalence problem. In fact, the skewed distribution of the SPC, GEN, USP and NEG categories (the SPC class has a marked prevalence) highly increases the probability that the agreement is obtained by chance.

APPENDIX C: Text Files

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
23. adige20040907_id405625.txt
24. adige20040907_id405626.txt
25. adige20040907_id405627.txt
26. adige20040907_id405630.txt
27. adige20040907_id405631.txt
28. adige20040907_id405632.txt
29. adige20040907_id405633.txt
30. adige20040907_id405634.txt
31. 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_id405871.txt
23. adige20040908_id405874.txt
24. adige20040908_id405875.txt
25. adige20040908_id405877.txt
26. adige20040908_id405878.txt
27. 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_id405911.txt
24. adige20040908_id405914.txt
25. adige20040908_id405915.txt
26. adige20040908_id405916.txt
27. adige20040908_id405917.txt
28. 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: 335

Test text files divided by date and category

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_id405936.txt
15. 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:190

REFERENCES

- (De Mauro 2000) De Mauro, *Il dizionario della lingua italiana per il terzo millennio*, Torino, Paravia, 2000.
  On-line: http://old.demauroparavia.it/
- (Di Eugenio, Glass 2004) Di Eugenio, B., Glass, M. (2004). *The kappa statistic: A second look*. Computational Linguistics, 30(1):95--101.
- (Lavelli et al. 2005) Lavelli, Magnini, Negri, Pianta, Speranza, Sprugnoli, *Italian Content Annotation Bank (I-CAB): Temporal Expressions (V. 1.0.). Technical Report* T-0505-12, ITC-irst, Trento, 2005.
  On-line: http://tcc.itc.it/projects/ontotext/Publications/i-cab-v1.pdf
- (LDC 2005) Linguistic Data Consortium, *Automatic Content Extraction English Annotation Guidelines for Entities*, version 5.6.1 2005.05.23.
  On-line: http://projects.ldc.upenn.edu/ace/docs/English-Entities-Guidelines_v5.6.1.pdf
- (LDC 2004) Linguistic Data Consortium, *Mapping from LDC's Annotation Format to ACE Program Format,* version 1.0, 2004-01-30.
  On-line: http://projects.ldc.upenn.edu/ace/docs/Mapping-ALF-to-APF-v1-0.pdf

WEB SITES

- ACE: http://www.nist.gov/speech/tests/ace/index.htm
  http://www.ldc.upenn.edu/Projects/ACE/
- Callisto: http://callisto.mitre.org
- MITRE: http://www.mitre.org/