

# ITALIAN CONTENT ANNOTATION BANK (I-CAB): ORGANIZATION ENTITIES (V. 1.1)<sup>1</sup>

*Emanuele Pianta\**, *Manuela Speranza\**, *Bernardo Magnini\**,  
*Valentina Bartalesi Lenzi\*\**, and *Rachele Sprugnoli\*\**

\* *ITC-irst, Povo 38050 (Trento) Italy*  
{ [pianta](mailto:pianta@itc.it) | [manspera](mailto:manspera@itc.it) | [magnini](mailto:magnini@itc.it) } @itc.it

\*\* *CELCT, Trento 38100 Italy*  
{ [sprugnoli](mailto:sprugnoli@celct.it) | [bartalesi](mailto:bartalesi@celct.it) } @celct.it

*Maggio 2007*

The content of the document is the result of the discussion which has taken place within the working group of ITC-irst and CELCT composed by Valentina Bartalesi Lenzi, Christian Girardi, Bernardo Magnini, Matteo Negri, Emanuele Pianta, Lorenza Romano, Manuela Speranza, Rachele Sprugnoli.

---

<sup>1</sup> 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.



# TABLE OF CONTENTS

|  |           |
|--|-----------|
| <b>ABSTRACT .....</b>  | <b>1</b>  |
| <b>1. INTRODUCTION.....</b>  | <b>3</b>  |
| <b>2. THE ENTITY DETECTION TASK.....</b>                                 | <b>5</b>  |
| 2.1 SEMANTIC TYPES AND SUBTYPES OF ENTITIES .....                        | 6         |
| 2.2 REFERENCE CLASSES OF ENTITIES .....                                  | 7         |
| 2.3 SYNTACTIC CATEGORIES OF ENTITY MENTIONS: ACETYPES AND LDCTYPES ..... | 7         |
| 2.4 ATTRIBUTIVE USE OF ENTITY MENTIONS .....                             | 10        |
| 2.5 THE ANNOTATION TOOL.....   | 11        |
| 2.6 NOTATIONAL CONVENTIONS .....   | 11        |
| <b>3. GUIDELINES FOR THE ANNOTATION OF ORGANIZATION ENTITIES .....</b>   | <b>13</b> |
| 3.1 SEMANTIC SUBTYPES OF ORGANIZATION ENTITIES .....                     | 13        |
| <i>ORG-Government</i> .....  | 13        |
| <i>ORG-Commercial</i> .....  | 13        |
| <i>ORG-Educational</i> .....   | 13        |
| <i>ORG-Entertainment</i> .....   | 13        |
| <i>ORG-Non Governmental</i> .....  | 14        |
| <i>ORG-Media</i> .....   | 14        |
| <i>ORG-Religious</i> .....   | 14        |
| <i>ORG-Medical Science</i> .....   | 14        |
| <i>ORG-Sports</i> .....  | 15        |
| <i>ORG-Mixed</i> .....   | 15        |
| 3.2 REFERENCE CLASSES .....  | 15        |
| <i>Specific Referential (SPC)</i> .....                                  | 15        |
| <i>Generic Referential (GEN)</i> .....                                   | 15        |
| <i>Under-specified Referential (USP)</i> .....                           | 16        |
| <i>Negatively Quantified (NEG)</i> .....                                 | 16        |
| <i>Mapping between reference CLASSES and semantic SUBTYPES</i> .....     | 16        |
| 3.3 MENTION ACETYPES (ADAPTED FOR I-CAB) .....                           | 17        |
| <i>NAM (Names)</i> .....   | 17        |
| <i>NOM (Quantified Nominal Constructions)</i> .....                      | 17        |
| <i>PRO (Pronouns)</i> .....  | 17        |
| <i>MIX (Mixed MENTIONS)</i> :.....                                       | 17        |
| 3.4 MENTION LDCTYPES (ADAPTED FOR I-CAB) .....                           | 18        |
| <i>NAM (Names)</i> .....   | 18        |
| <i>NOM (Quantified Nominal Constructions)</i> .....                      | 18        |
| <i>HLS (Headless MENTIONS)</i> : .....                                   | 18        |
| <i>WHQ (WH-Words)</i> .....  | 19        |
| <i>ENCLIT (Enclitics)</i> .....  | 19        |
| <i>PROCLIT (Proclitics)</i> .....  | 20        |
| <i>PRO (Pronouns)</i> .....  | 20        |
| <i>BAR (Bare Nominal MENTIONS)</i> .....                                 | 20        |
| <i>PTV (Partitive Constructions)</i> .....                               | 21        |
| <i>APP (Appositional Constructions)</i> .....                            | 21        |

|   |           |
|---|-----------|
| <i>ARC (Complex Constructions taking a Relative Clause)</i> ..... | 23        |
| <i>CONJ (Conjoined constructions)</i> .....                       | 23        |
| 3.5 ATTRIBUTIVE USE .....   | 24        |
| <b><u>4. PRONOUNS AND ADJECTIVES.....</u></b>                     | <b>26</b> |
| 4.1 PERSONAL PRONOUNS.....  | 26        |
| 4.2 POSSESSIVE ADJECTIVES AND PRONOUNS .....                      | 26        |
| 4.3 INDEFINITE PRONOUNS .....                                     | 26        |
| 4.4 CLITICS .....   | 26        |
| 4.5 REFLEXIVE PRONOUNS.....                                       | 27        |
| 4.6 SYNTHESIS OF THE POSSIBLE INTERPRETATIONS OF <i>SI</i> .....  | 28        |
| <b><u>APPENDIX A: SPECIAL CASES.....</u></b>                      | <b>29</b> |
| TITLES OF BOOKS, CDs AND EXHIBITIONS .....                        | 29        |
| ARTICULATED PREPOSITIONS .....                                    | 29        |
| ORGANIZATION VERSUS OTHER SEMANTIC TYPES OF ENTITIES.....         | 29        |
| <i>a. Organization versus Person ENTITIES</i> .....               | 29        |
| <i>b. Organization versus Geo-Political ENTITIES</i> .....        | 30        |
| NICKNAME METONYMY .....   | 31        |
| HOW TO DEAL WITH DASHES, COLONS AND BRACKETS .....                | 31        |
| “CIRCA” AND “ALMENO” .....  | 32        |
| <b><u>APPENDIX B: INTER-ANNOTATOR AGREEMENT.....</u></b>          | <b>34</b> |
| <b><u>APPENDIX C: STATISTICAL DATA .....</u></b>                  | <b>36</b> |
| GENERAL STATISTICS .....  | 36        |
| DISTRIBUTION OF ORGANIZATION ENTITIES BY SEMANTIC SUBTYPE.....    | 36        |
| DISTRIBUTION OF ORGANIZATION ENTITIES BY REFERENCE CLASS .....    | 36        |
| STATISTICS ON VALUES OF THE ACETYPE ATTRIBUTE .....               | 36        |
| STATISTICS ON VALUES OF THE LDCTYPE ATTRIBUTE .....               | 37        |
| NUMBER OF ATTRIBUTIVE MENTIONS (ATR=“YES”) .....                  | 38        |
| <b><u>APPENDIX D: TEXT FILES .....</u></b>                        | <b>39</b> |
| TRAINING TEXT FILES DIVIDED BY DATE AND CATEGORY .....            | 39        |
| TEST TEXT FILES DIVIDED BY DATE AND CATEGORY .....                | XLIII     |
| <b><u>REFERENCES.....</u></b>                                     | <b>46</b> |
| <b><u>WEB SITES.....</u></b>                                      | <b>46</b> |

## **ABSTRACT**

This document reports on the annotation of Organization entities for the Italian Content Annotation Bank (I-CAB) being developed at ITC-irst in conjunction with CELCT.

I-CAB is a corpus of Italian news annotated with semantic information at different levels. The first level is represented by temporal expressions, the second level is represented by different types of entities (i.e. person, organizations, locations and geopolitical entities), and the third level is represented by relations between entities (e.g. the affiliation relation connecting a person to an organization).

As we intend I-CAB to become a benchmark for various automatic Information Extraction tasks, we followed a policy of reusing already available markup languages. In particular, we adopted the annotation schemes developed for the ACE Entity Detection and Time Expressions Recognition and Normalization tasks for English. We describe the extensions to the ACE guidelines needed to adapt them to the specific morpho-syntactic features of Italian and to include a wider range of entities (e.g. conjunctions), providing a large number of examples.



# 1. INTRODUCTION

This report presents the annotation of Organization entities in the Italian Content Annotation Bank (I-CAB), a corpus of semantically annotated documents for Italian containing annotations of temporal expressions (Lavelli et al. 2005), person entities, organization entities, location entities, geo-political entities and of a number of selected relations among such entities.

Following a policy of reusing already available markup languages, the annotation activity has been carried out adopting the formalisms developed within the American ACE program. However, due to the differences between English and Italian, part of the work has been dedicated to the revision and adaptation to Italian of the annotation guidelines.

The creation of I-CAB is part of the three-year project Ontotext funded by the Autonomous Province of Trento. Ontotext focuses on the study and development of innovative knowledge extraction techniques to produce new or less noisy information to be made available for the Semantic Web. Within the new research area of Ontology-Based Knowledge Extraction, Ontotext addresses three key research aspects: annotating documents with semantic and relational information, providing an adequate degree of interoperability of such relational information, and updating and extending the ontologies used for Semantic Web annotation. The concrete evaluation scenario in which algorithms will be tested with a number of large-scale experiments is the automatic acquisition of information about people from newspaper articles.

This document is structured as follows. Section 2 presents the Entity Detection task (valid for all types of entities) and introduces Callisto, the annotation tool we have chosen, as well as the notational conventions used throughout the document. Section 3 describes the annotation of Organization entities giving several examples and Section 4 provides details on how we deal with pronouns and adjectives referring to organization entities. Some special cases are described in Appendix A, while Appendix B presents data about the inter-annotator agreement. Appendix C shows statistics about the Organization entities annotated in I-CAB and, finally, Appendix D contains the complete list of the text files making up the corpus.





## 2. THE ENTITY DETECTION TASK

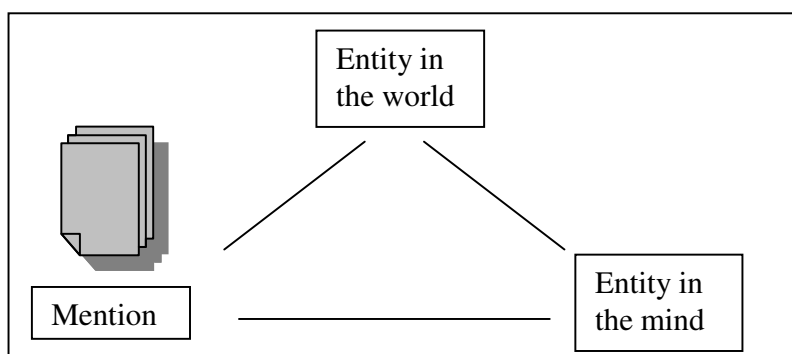
Entity Detection is one of the tasks of the ACE (*Automatic Content Extraction*)<sup>2</sup> program. It “requires that selected types of entities mentioned in the source data be detected, their sense disambiguated, and that selected attributes of these entities be extracted and merged into a unified 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.

**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.).

<sup>2</sup> <http://www.nist.gov/speech/tests/ace/index.htm> or <http://www ldc.upenn.edu/Projects/ACE/>

*Mentions*, on the other hand, are classified according to syntactic categories (e.g. proper names, common nouns, pronouns, etc.) and to attributive versus non attributive use. 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). These abbreviations coincide with the labels used in the annotated files, with the only exception of ACETYPE, which becomes simply TYPE, due to the design of the Callisto user interface (see Section 2.5).

**Table 1:** Annotation Categories

|   |
|---|
| <ul style="list-style-type: none"> <li>• ENTITY <ul style="list-style-type: none"> <li>○ Semantic TYPE <ul style="list-style-type: none"> <li>-Semantic SUBTYPE</li> </ul> </li> <li>○ Reference CLASS</li> </ul> </li> <li>• MENTION <ul style="list-style-type: none"> <li>○ ACETYPE</li> <li>○ LDCTYPE</li> <li>○ ATtRibutive use</li> </ul> </li> </ul> |
|---|

## 2.1 Semantic TYPES and SUBTYPES of ENTITIES

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, while adding the new semantic TYPE Mixed as follows:

- Person (PER)
- Organization (ORG)
- Geo-Political Entity (GPE)
- Location (LOC)
- Mixed (MIX): non-uniform ENTITY groups for which it is impossible to chose a single semantic TYPE (e.g., in *he and his company*, we have a group consisting of a PER and an ORG).

For each semantic TYPE, various SUBTYPES are defined, that provide further semantic information. For instance, for PER, we have the individual, group and indefinite semantic SUBTYPES.

## 2.2 Reference CLASSES of ENTITIES

Following the ACE-LDC guidelines, each ENTITY (entity in the mind) is assigned a CLASS that describes the kind of reference it makes to something in the world (entity in the world).

### Specific Referential (SPC)

An ENTITY is SPC when it refers to a particular, unique object (or set of objects).

*[Ciampi] è originario di [Livorno]*<sup>3</sup>

Notice that an entity can be SPC even if the author or reader is not aware of its name or anchor in the real world.

*Non ricordo come si chiama [il precedente Presidente]*

### Generic Referential (GEN)

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

*[La Camera] è eletta a suffragio universale e diretto*

### Under-specified Referential (USP)

We reserve the term underspecified for non-specific, non-generic reference. Underspecified references include quantified NP's in modal, future, conditional, hypothetical, negated, uncertain and question contexts: in all these cases the ENTITY/ENTITIES referenced cannot be verified, regardless of the amount of "effort".

*Non so [chi] verrà al concerto*

### Negatively Quantified (NEG)

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

*[Nessuno stato] accetterebbe questo accordo*

## 2.3 Syntactic categories of Entity MENTIONS: ACETYPES and LDCTYPES

The original ACE annotation scheme, called APF (ACE Program Format), includes four syntactic categories (ACETYPES). In a second phase, in 2004, the Linguistic Data Consortium (LDC) introduced a new annotation scheme called ALF (ACE LDC Format) which adds a number of new syntactic categories (LDCTYPES). We have decided to

---

<sup>3</sup> In our examples MENTIONS are enclosed in brackets and their head is underlined; when the head is the same as the extent, the underlining is omitted (see Section 2.6) .

follow both indications and thus to annotate each Entity MENTION with an ACETYPE and an LDCTYPE.

The original four ACETYPES are: NAM (proper name), NOM (quantified nominal constructions), PRE (pre-modifier), and PRO (pronoun). For the purpose of adapting the guidelines to Italian, we have transformed PRE into MOD (modifier). In addition, we have added the new ACETYPE MIX (see Section 3.3), so we end up with five ACETYPES (Table 3).

The LDCTYPES have been adopted with a few modifications:

(1) In order to adapt the guidelines to Italian, we have created two separate tags for pre-modifiers (PRE) and post-modifiers (POST), and two new tags, ENCLIT and PROCLIT, respectively for enclitics and proclitics that are attached at the end or the beginning of a word (see Section 4.4).

(2) In addition, we have created a new LDCTYPE, CONJ (see Section 3.4), to annotate conjunctions of ENTITIES (e.g. *uomini e donne anziani; madre e figlio*). This allows us to mark the co-reference with anaphoric MENTIONS which might follow in the text (e.g. *essi = they, queste persone = these people*, etc.). A CONJ (see Table 2) is a complex MENTION whose head corresponds to the entire MENTION. It also contains nested MENTIONS, in the sense that the different parts of a CONJ are also annotated (e.g. *uomini* and *donne* in the first example, and *madre* and *figlio* in the second, are independent MENTIONS)<sup>4</sup>.

**Table 2:** Some CONJ patterns

|                   |                              |
|-------------------|------------------------------|
| x e y             | <i>x and y</i>               |
| x, y e z          | <i>x, y and z</i>            |
| tra x e y         | <i>between x and y</i>       |
| con x e y         | <i>with x and y</i>          |
| x o y             | <i>either x or y</i>         |
| x, y o z          | <i>x, y or z</i>             |
| con x e con y     | <i>with x and with y</i>     |
| x ma anche y      | <i>x but also y</i>          |
| x, y ma anche z   | <i>x, y but also z</i>       |
| della x e della y | <i>of the x and of the y</i> |

The adapted lists of the ACETYPES and LDCTYPES used in the annotation of I-CAB are presented in Table 3. The mapping between the set of ACETYPES and LDCTYPES is presented in Tables 5 and 6 (our adaptations in italic).

<sup>4</sup> 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*.

**Table 3:** ACETYPES and LDCTYPES for the annotation of I-CAB

| ACETYPES               | LDCTYPES   |
|------------------------|--|
| NAM (proper name)      | NAM (proper name)                                    |
| NOM (nominal compound) | NOM (nominal compound)                               |
| PRO (pronoun)          | PRO (pronoun)  |
| MOD (modifier)         | PRE (pre-modifier)                                   |
| MIX (mixed)            | POST (post-modifier)                                 |
|                        | BAR (bare)   |
|                        | WHQ (question word)                                  |
|                        | HLS (headless)                                       |
|                        | PTV (partitive)                                      |
|                        | ENCLIT (enclitic)                                    |
|                        | PROCLIT (proclitic)                                  |
|                        | APP (appositional construction)                      |
|                        | ARC (appositional construction with relative clause) |
|                        | CONJ (conjunction)                                   |

**Table 4:** Mapping from ACETYPES to LDCTYPES

| ACETYPE     | LDCTYPE        |
|-------------|----------------|
| <b>NOM</b>  | NOM            |
|             | BAR            |
|             | APP            |
|             | ARC            |
|             | <i>CONJ</i>    |
|             | PTV            |
| <b>NAM</b>  | NAM            |
|             | APP            |
|             | ARC            |
|             | <i>CONJ</i>    |
| <b>PRO</b>  | PRO            |
|             | <i>PROCLIT</i> |
|             | <i>ENCLIT</i>  |
|             | WHQ            |
|             | HLS            |
|             | PTV            |
|             | APP            |
|             | ARC            |
|             | <i>CONJ</i>    |
|             | <b>MOD</b>     |
| <i>POST</i> |                |
| <b>MIX</b>  | <i>CONJ</i>    |

**Table 5:** Mapping from LDCTYPES to ACETYPES

| LDCTYPE        | ACETYPE    |
|----------------|------------|
| <b>NOM</b>     | NOM        |
| <b>BAR</b>     | NOM        |
| <b>NAM</b>     | NAM        |
| <b>APP</b>     | NOM        |
|                | NAM        |
|                | PRO        |
| <b>ARC</b>     | NOM        |
|                | NAM        |
|                | PRO        |
| <b>CONJ</b>    | NOM        |
|                | NAM        |
|                | PRO        |
|                | <i>MIX</i> |
| <b>PRO</b>     | PRO        |
| <b>ENCLIT</b>  | PRO        |
| <b>PROCLIT</b> | PRO        |
| <b>PRE</b>     | MOD        |
| <b>POST</b>    | MOD        |
| <b>WHQ</b>     | PRO        |
| <b>HLS</b>     | PRO        |
| <b>PTV</b>     | PRO        |
|                | NOM        |

## 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]*

Appositional constructions (APP, e.g. [*La cantante Madonna*]), appositional constructions with relatives (ARC, e.g. [*La cantante Madonna che è in tour*]), and conjunctions (CONJ, e.g. [*Madonna e Prince*]) 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.

## 2.4 Attributive use of Entity MENTIONS

MENTIONS can be used attributively and these uses have to be marked. As a general rule, a MENTION is marked as ATR in the following cases (see Section 3.5):

- when it is a predicate complement (e.g. *la capitale italiana* in *Roma è la capitale italiana*) or an objective complement (e.g. *Little John* in *Lo chiamano Little John*)
- when it occurs in the final part/s of an appositional construction (e.g. *azienda leader nel settore* in *Microsoft, azienda leader nel settore*, see Section 3.4) or in the central part of an ARC (e.g. *avvocato* in *Franco, avvocato, che vive in America*, see Section 3.4);
- when it is preceded by expressions such as *in qualità di*, *come* (= *as*), e.g. *insegnante* in *Luigi lavora come insegnante* (see APP in Section 3.4).

## 2.5 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 Temporal Expressions we have used the TIMEX2 task, whereas for the annotation of ENTITIES we have used the ACE2004 task.

For each MENTION, Callisto provides two separate slots for the ACETYPES and the LDCTYPES, so we found no problems in annotating each MENTION with both kinds of information. For the annotation of the attributive uses, Callisto provides a specific slot which can be filled with the value ATR or left empty.

## 2.6 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'altra azienda*] è fallita

[*Marco*] è sposato da anni

### Extended notation:

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

[*Marco*] è andato al cinema.

ACETYPE=NAM

LDCTYPE=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*]<sub>E-1 m-1</sub> è andato al cinema con la [*propria*]<sub>E-1 m-2</sub> macchina.

E-1 m-1 ACETYPE=NAM

LDCTYPE=NAM

m-2 ACETYPE=PRO

LDCTYPE=PRO

Two mentions referring to different ENTITIES:

[[*I figli* di [*Marco*]<sub>E-1 m-1</sub>]<sub>E-2 m-1</sub> sono andati al cinema.

E-1 m-1 ACETYPE=NAM

LDCTYPE=NAM

E-2 m-1 ACETYPE=NOM

LDCTYPE=NOM

In some examples we add values of ENTITY attributes (e.g SUBTYPES and CLASSES); as far as attributive use is concerned, we only specify when ATR=YES.

[*Marco*]<sub>E-1 m-1</sub> è [*un ingegnere*]<sub>E-1 m-2</sub>

E-1 (PER-Indiv.) (SPC) m-1 ACETYPE=NAM

LDCTYPE=NAM

m-2 ACETYPE=NOM

LDCTYPE=NOM

ATR=YES



### 3. GUIDELINES FOR THE ANNOTATION OF ORGANIZATION ENTITIES

In this Section we describe annotation guidelines for Organization ENTITIES.

#### 3.1 Semantic SUBTYPES of Organization ENTITIES

**ORG-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]<sub>E-1 m-1</sub> è [un organismo politico]<sub>E-1 m-2</sub>*  
E-1 (ORG-Gov.) m-1 ACETYPE=NAM  
LDCTYPE=NAM  
m-2 ACETYPE=NOM  
LDCTYPE=NOM  
ATR=YES

*L'auto è stata posta sotto sequestro [dalla Guardia di Finanza]<sub>E-1 m-1</sub>*  
E-1 (ORG-Gov.) m-1 ACETYPE=NAM  
LDCTYPE=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).

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

*Nel maggio 1963 [la società]<sub>E-1 m-1</sub> cambiò la denominazione sociale*  
E-1 (ORG-Com.) m-1 ACETYPE=NOM  
LDCTYPE=NOM

More examples:

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

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

*[L'ateneo]<sub>E-1 m-1</sub> avrà un nuovo rettore*  
E-1 (ORG-Edu.) m-1 ACETYPE=NOM  
LDCTYPE=NOM

**ORG-Entertainment:** entertainment organizations are those whose primary activity is entertainment.

*[I Rem]<sub>E-1 m-1</sub> presenteranno il [loro]<sub>E-1 m-2</sub> nuovo album*  
E-1 (ORG-Ent.) m-1 ACETYPE=NAM  
LDCTYPE=NAM  
m-2 ACETYPE=PRO  
LDCTYPE=PRO

**ORG-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]*

**ORG-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] <sub>E-1 m-1</sub>*

E-1 (ORG-Med.) m-1 ACETYPE=NAM

LDCTYPE=NAM

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

*[L' Arcidiocesi di Trento] <sub>E-1 m-1</sub> propone un nuovo convegno ecumenico*

E-1 (ORG-Rel.) m-1 ACETYPE=NAM

LDCTYPE=NAM

*Era presente anche Igor Vyzhanov [del Patriarcato Ortodosso di Mosca] <sub>E-1 m-1</sub>*

E-1 (ORG-Rel.) m-1 ACETYPE=NAM

LDCTYPE=NAM

**ORG-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] <sub>E-1 m-1</sub>*

E-1 (ORG-Med Sci.) m-1 ACETYPE=NAM

LDCTYPE=NAM

*Il giovane venne soccorso dall'elicottero [del 118] <sub>E-1 m-1</sub>*

E-1 (ORG-Med Sci.) m-1 ACETYPE=NAM

LDCTYPE=NAM

**ORG-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 [AI]<sub>E-1 m-1</sub>*  
 E-1 (ORG-Spo.) m-1 ACETYPE=NAM  
 LDCTYPE=NAM

More examples:

*[Juve] - [Roma] 1 - 1*  
*[Luna Rossa] è in testa alla classifica*  
*È stato designato il nuovo CT [della squadra]*

**ORG-Mixed:** non-uniform groups of organizations for which it is impossible to choose a single semantic SUBTYPE (e.g., in the following example, we have a group consisting of an ORG-NonGov and an ORG-Comm).

*[[I sindacati]<sub>E-1 m-1</sub> e [l'azienda]<sub>E-2 m-1</sub>]<sub>E-3 m-1</sub> hanno finalmente trovato un accordo*  
 E-1 (ORG-Non Gov.) m-1 ACETYPE=NOM  
 LDCTYPE=NOM  
 E-2 (ORG-Comm.) m-1 ACETYPE=NOM  
 LDCTYPE=NOM  
 E-3 (ORG-Mixed) m-1 ACETYPE=NOM  
 LDCTYPE=CONJ

## 3.2 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, *Bar Stazione* in the sentence *Francesca gestisce il bar Stazione da tre anni* is to be annotated as SPC; however, also *bar* in *Francesca ha aperto un nuovo bar* is 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*

Co-reference between GEN ENTITIES is generally admitted:

*[I gruppi integralisti]<sub>E-1 m-1</sub> sono numerosi e il [loro]<sub>E-1 m-2</sub> potere è enorme*  
 E-1 m-1 ACETYPE=NOM  
 LDCTYPE=NOM

E-1 m-2 ACETYPE=PRO  
LDCTYPE=PRO

### 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 [quante scuole] parteciperanno all'iniziativa*
- b) imprecise quantifications;  
*[Molti ospedali italiani] hanno problemi di budget*
- 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) NPs that the annotator cannot classify.

USP reference of type (a), (c), and (d) above all admit co-reference between each other. Co-reference does not occur, usually, between USP reference of type (b).

### 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.

*[Nessuna squadra trentina]<sub>E-1 m-1</sub> è retrocessa*

E-1 (NEG) m-1 ACETYPE=NOM  
LDCTYPE=NOM

NEG ENTITIES can not co-refer:

*[Nessuna]<sub>E-1 m-1</sub> è retrocessa, [nessuna]<sub>E-2 m-1</sub> è stata promossa*

E-1 (NEG) m-1 ACETYPE=PRO  
LDCTYPE=PRO

E-2 (NEG) m-1 ACETYPE=PRO  
LDCTYPE=PRO

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

*[Nessuna]<sub>E-1 m-1</sub> [sì]<sub>E-1 m-2</sub> è meritata la promozione*

E-1 (NEG) m-1 ACETYPE=PRO  
LDCTYPE=PRO

m-2 ACETYPE=PRO  
LDCTYPE=PRO

### Mapping between reference CLASSES and semantic SUBTYPES

Organization ENTITIES can be of any semantic SUBTYPE, independently of their reference CLASSES.

Here are some examples:

[*Le scuole*] chiudono a Giugno (GEN-Edu.)

*Non so quale sia [l'ospedale] migliore della zona* (USP-Med Sci.)

[*Nessun giornale*] ha riportato la notizia (NEG-Media)

### 3.3 MENTION ACETYPES (adapted for I-CAB)

**NAM (Names):** proper nouns and nicknames.

[*La Microsoft Corporation*] ha sede a Redmond, USA

[*Il Carroccio*] ha molti sostenitori nell'Italia Settentrionale

Mentions of foreign organizations have been annotated as proper nouns (type="NAM") if they were the literal translation of the original name, whereas they 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*.

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

[*Le associazioni no-profit*] sono sempre più numerose

[*La mia azienda*] sta assumendo nuove figure professionali

**PRO (Pronouns):** all pronouns and headless MENTIONS have ACETYPE PRO.

[*Molte*] sono [*le associazioni italiane nel mondo*]

[*La ditta*]<sub>E-1 m-1</sub> sta aumentando i [*suoi*]<sub>E-1 m-2</sub> profitti

E-1 m-1 ACETYPE=NOM

LDCTYPE=NOM

E-1 m-2 ACETYPE=PRO

LDCTYPE=PRO

**MOD (Modifiers):** as Italian, unlike English, does not admit nouns in modifier position (cfr. *police station* versus *stazione di polizia*), this ACETYPE does not apply for organization ENTITIES.

**MIX (Mixed MENTIONS):** it is used for conjunctions of ENTITIES when the two ENTITIES are not of the same ACETYPE. For example, when one is NAM and the other is NOM, as in the following example (E-3 m-1):

[[*La Trentino Trasporti*]<sub>E-1 m-1</sub> e [*altre ditte trentine*]<sub>E-2 m-1</sub>]<sub>E-3 m-1</sub> sono in crescita

E-1 m-1 TYPE=NAM

LDCTYPE=NAM

E-2 m-1 TYPE=NOM

LDCTYPE=NOM

E-3 m-1 TYPE=MIX  
LDCTYPE=CONJ

### 3.4 MENTION LDCTYPES (adapted for I-CAB)

**NAM (Names):** proper nouns and nicknames.

*[Il Gruppo Folkloristico Canazei] terrà un'esibizione in piazza*

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM

*[Il Toro]<sub>E-1 m-1</sub> è in testa alla classifica*

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM

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

*Il presidente [dell'associazione]<sub>E-1 m-1</sub> ha parlato al congresso*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

**HLS (Headless MENTIONS):** 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 (notice that they all have ACETYPE PRO).

*La situazione [delle aziende nel territorio] <sub>E-1 m-1</sub> è critica: [il 10 %] <sub>E-2 m-1</sub> è in fallimento*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

E-2 m-1 ACETYPE=PRO  
LDCTYPE=HLS

*Nuova classifica [sulle migliori università del mondo] <sub>E-1 m-1</sub>: [due] <sub>E-2 m-1</sub> sono europee e [tre] <sub>E-3 m-1</sub> sono statunitensi*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

E-2 m-1 ACETYPE=PRO  
LDCTYPE=HLS

E-3 m-1 ACETYPE=PRO  
LDCTYPE=HLS

*[Varie squadre] <sub>E-1 m-1</sub> si contendono il titolo ma solo [la più forte] <sub>E-2 m-1</sub> vincerà*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

E-2 m-1 ACETYPE=PRO  
LDCTYPE=HLS

Also indefinite pronouns having the same form as their corresponding adjectives have LDCTYPE HLS. See for instance *molti* (=many), *alcuni* (=some), *tutti* (=every), *altri* (=other), *quanti* (=how many), *entrambi* (=both), *gli stessi* (=the same), *tutt'e due* (=both).

On the other hand, indefinite pronouns having different form than their corresponding adjective have LDCTYPE PRO. See for instance *nessuno* (=none) as pronoun and *nessun* (=any) as adjective.

*Presentato il bilancio annuale [dell'azienda]<sub>E-1 m-1</sub>: [nessuno]<sub>E-2 m-1</sub> è così in espansione*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

E-2 m-1 ACETYPE=PRO  
LDCTYPE=PRO

**WHQ (WH-Words):** interrogative and relative pronouns

*[La ditta [che]<sub>E-1 m-2</sub> ha sede a Trento]<sub>E-1 m-1</sub>*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

m-2 ACETYPE=PRO  
LDCTYPE=WHQ

*[Quelle [che]<sub>E-1 m-2</sub> vedi nel grafico]<sub>E-1 m-1</sub> sono [le cooperative della zona]<sub>E-1 m-3</sub>*

E-1 m-1 ACETYPE=PRO  
LDCTYPE=PRO

m-2 ACETYPE=PRO  
LDCTYPE=WHQ

m-3 ACETYPE=NOM  
LDCTYPE=NOM  
ATR=YES

The annotation of the correlative pronoun “chi” (=those who) includes the pronoun and also the subordinate clause:

*Con grossa delusione di [chi pensava di farcela]<sub>E-1 m-1</sub>*

E-1 m-1 ACETYPE=PRO  
LDCTYPE=WHQ

The annotation of the interrogative pronoun “chi” includes only the pronoun:

*[Chi]<sub>E-1 m-1</sub> ha vinto il titolo di miglior gruppo emergente?*

E-1 m-1 ACETYPE=PRO  
LDCTYPE=WHQ

**ENCLIT (Enclitics):** enclitics that are attached at the end of a verb ([incontrarli]/to meet them) or of the adverb *ecco* ([Eccoli]/!Here they are!). We take all the word as extent and the prefix as head.

*Adoro i [REM]<sub>E-1 m-1</sub>! Potrei [ascoltarli]<sub>E-1 m-2</sub> per ore*

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM

m-2 ACETYPE=PRO  
LDCTYPE=PROCLIT

**PROCLIT (Proclitics):** proclitics that precede the main verb and are attached to another word. We take all the word as extent and only the prefix as head.

*[La compagnia]<sub>E-1 m-1</sub> voleva esibirsi ma i manifestanti [glielo]<sub>E-1 m-2</sub> hanno impedito*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

m-2 ACETYPE=PRO  
LDCTYPE=PROCLIT

Notice that in Italian proclitics can also occur as free morphemes. In this case they are tagged as ACETYPE PRO and LDCTYPE PRO.

*Lui doveva salvare [la ditta]<sub>E-1 m-1</sub> e invece è stato quello che [la]<sub>E-1 m-2</sub> ha affossata*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

m-2 ACETYPE=PRO  
LDCTYPE=PRO

**PRO (Pronouns):** all remaining pronouns

Among others:

a) personal pronouns:

*[Loro]<sub>E-1 m-1</sub> non sanno come evitare il fallimento*

E-1 m-1 ACETYPE=PRO  
LDCTYPE=PRO

b) possessive pronouns:

*[L' Associazione Pranic Healing]<sub>E-1 m-1</sub> presenta la [propria]<sub>E-1 m-2</sub> attività*

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM

m-2 ACETYPE=PRO  
LDCTYPE=PRO

c) indefinite pronouns whose form differ from the corresponding adjective, such as pronoun *nessuno* (=none) – adj. *nessun/nessuna* (=no). Other examples are *ciascuno* (=every), *quest'ultimo* (=the latter), *quelli/e* (=those), *questi/e* (=these):

*[Qualcuna]<sub>E-1 m-1</sub> sarà quotata presto in borsa*

E-1 m-1 ACETYPE=PRO  
LDCTYPE=PRO

**BAR (Bare Nominal MENTIONS):** unquantified nominal constructions without pre-modifiers and articles. They can be singular or plural, and are very common in appositional constructions.

*[Sindacati]<sub>E-1 m-1</sub> in marcia a Roma*



E-1 m-1 ACETYPE=NOM  
LDCTYPE=BAR

*Il concorso è aperto a [band di tutti i generi]<sub>E-1 m-1</sub>*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=BAR

**PTV (Partitive Constructions):** they have two elements, the part and the whole. The first element (the part) quantifies over the second element (the whole).

*[Alcune [delle squadre]<sub>E-2 m-1</sub>]<sub>E-1 m-1</sub> sono già condannate alla retrocessione*

E-1 (SPC) m-1 ACETYPE=PRO  
LDCTYPE=PTV

E-2 (SPC) m-1 ACETYPE=NOM  
LDCTYPE=NOM

NB: In Italian, *parte* (= *part*) and *maggioranza* (= *majority*) are nouns so they are tagged as NOM PTV, whereas, their English correspondent are tagged as PRO PTV in the ACE-LDC corpus.

*[Parte [degli Atenei italiani]<sub>E-2 m-1</sub>]<sub>E-1 m-1</sub> è contraria alla nuova legge*

E-1 (SPC) m-1 ACETYPE=NOM  
LDCTYPE=PTV

E-2 (SPC) m-1 ACETYPE=NOM  
LDCTYPE=NOM

*[La maggioranza [delle Ong]<sub>E-2 m-1</sub>]<sub>E-1 m-1</sub> era favorevole*

E-1 (SPC) m-1 ACETYPE=NOM  
LDCTYPE=PTV

E-2 (SPC) m-1 ACETYPE=NOM  
LDCTYPE=NOM

Group nouns (*gruppo*, *famiglia*, *equipe*) can occur 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 gruppo di [aziende]<sub>E-1 m-2</sub>]<sub>E-1 m-1</sub> della valle ha inaugurato un nuovo business*

E-1 (SPC) m-1 ACETYPE=NOM  
LDCTYPE=NOM

m-2 ACETYPE=NOM  
LDCTYPE=BAR

**APP (Appositional Constructions):** they consist of a pivot nominal phrase and one or more appositions.

We consider as appositional constructions any combination of various types of contiguous co-referring noun phrases<sup>5</sup>:

- Noun – noun: [la ditta, leader nel settore], lavora tantissimo

<sup>5</sup> Contiguous noun phrases can be separated by a comma, but not by colons, parenthesis, dashes or similar punctuation marks (see Appendix A, "How to deal with dashes, colons and brackets").

- Noun – proper name: [l'azienda Trentofrutta] è in fase di ristrutturazione
- Noun – pronoun: [il contingente, quello arrivato ora in Iraq], è a Nassiriya
- Proper name – noun: [La Juve, squadra forte], non giocherà nelle Coppe
- Proper name – pronoun: [I Rem, i più famosi tra i giovani], sono in tour in Italia
- Pronoun – noun: [loro, le cooperative], fanno fronte comune
- Pronoun – proper name: [uno dei musei più importanti, il Mart], è in Trentino
- Pronoun – pronoun: [loro, quelle che giocano in A], hanno sponsor importanti

In most of the cases, the second part of the appositional construction (and similarly also the third, the fourth, etc.) is ATR (see Section 3.5).

The ACETYPE of any APP is the same as the ACETYPE of the pivot of the APP (generally, the first part).

[[Trenitalia]<sub>E-1 m-1</sub> [azienda in crisi]<sub>E-1 m-2</sub>]<sub>E-1 m-3</sub> cerca nuovi fondi

|     |     |             |   |
|-----|-----|-------------|---|
| E-1 | m-1 | ACETYPE=NAM | ← |
|     |     | LDCTYPE=NAM |   |
|     | m-2 | ACETYPE=NOM |   |
|     |     | LDCTYPE=BAR |   |
|     |     | ATR=YES     |   |
|     | m-3 | ACETYPE=NAM |   |
|     |     | LDCTYPE=APP |   |

NB: the extent of the head is the whole MENTION

Constructions in which expressions such as *in qualità di* and *come* (= *as*) are used to connect two nouns, are *not* annotated as APP. Still, MENTIONS introduced by these expressions are ATR.

[Un ponte per...]<sub>E-1 m-1</sub> in qualità di [portavoce]<sub>E-1 m-2</sub> ha espresso parere positivo

|     |     |             |
|-----|-----|-------------|
| E-1 | m-1 | ACETYPE=NAM |
|     |     | LDCTYPE=NAM |
|     | m-2 | ACETYPE=NOM |
|     |     | LDCTYPE=BAR |
|     |     | ATR=YES     |

[La Margherita]<sub>E-1 m-1</sub> si affermò come [partito di maggioranza]<sub>E-1 m-2</sub>

|     |     |             |
|-----|-----|-------------|
| E-1 | m-1 | ACETYPE=NAM |
|     |     | LDCTYPE=NAM |
|     | m-2 | ACETYPE=NOM |
|     |     | LDCTYPE=BAR |
|     |     | ATR=YES     |

On the other hand, if the expression *in qualità di*, *come* is implicit, we annotate an appositional construction:

Nuova operazione milionaria con [[General Electric]<sub>E-1 m-1</sub> [finanziatore d'eccezione]<sub>E-1 m-2</sub>]<sub>E-1 m-3</sub>

|     |     |             |
|-----|-----|-------------|
| E-1 | m-1 | ACETYPE=NAM |
|     |     | LDCTYPE=NAM |
|     | m-2 | ACETYPE=NOM |
|     |     | LDCTYPE=BAR |

ATR=YES  
m-3 ACETYPE=NAM  
LDCTYPE=APP

**ARC (Complex Constructions taking a Relative Clause):** an appositional construction with an adjacent relative clause (WHQ). The relative clause must refer to the initial, referential (SPC) MENTION of the ENTITY rather than the latter, attributive (ATR) MENTION(S) of the ENTITY. Each sub-part of an ARC-construction is tagged.

General rules:

1. the second part of an ARC is attributive
2. the ACETYPE of an ARC is the same as the ACETYPE of the pivot
3. the pivot, its apposition/s and the WHQ MENTION co-refer

[[Gli Outline]<sub>E-1 m-1</sub>, [il gruppo musicale]<sub>E-1 m-2</sub>, [che]<sub>E-1 m-3</sub> *si è appena esibito*]<sub>E-1 m-4</sub>  
*sta avendo un enorme successo*

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
m-2 ACETYPE=NOM  
LDCTYPE=NOM  
ATR=YES  
m-3 ACETYPE=PRO  
LDCTYPE=WHQ  
m-4 ACETYPE=NOM  
LDCTYPE=ARC

**CONJ (Conjoined constructions):** all conjunctions of ENTITIES. The head coincides with the MENTION extent.

[[Margherita]<sub>E-1 m-1</sub> e [Ds]<sub>E-2 m-1</sub>]<sub>E-3 m-1</sub> *propongono nuovi emendamenti*

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
E-2 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
E-3 m-1 ACETYPE=NAM  
LDCTYPE=CONJ

*Sequestrati i registri di* [[5 associazioni]<sub>E-1 m-1</sub> e [10 cooperative]<sub>E-2 m-1</sub>]<sub>E-3 m-1</sub>

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM  
E-2 m-1 ACETYPE=NOM  
LDCTYPE=NOM  
E-3 m-1 ACETYPE=NOM  
LDCTYPE=CONJ

*Nuovo corteo di* [[CISL]<sub>E-1 m-1</sub> e [UIL]<sub>E-2 m-1</sub> [che]<sub>E-3 m-2</sub> *scendono in piazza*]<sub>E-3 m-1</sub>

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
E-2 m-1 ACETYPE=NAM

LDCTYPE=NAM  
 E-3 m-1 ACETYPE=NAM  
 LDCTYPE=CONJ  
 m-2 ACETYPE=PRO  
 LDCTYPE=WHQ

### 3.5 Attributive Use

Example of MENTIONS of Organization ENTITIES used attributively:

- predicate complements:

[Izvestia]<sub>E-1 m-1</sub> è [un quotidiano autorevole nel panorama russo]<sub>E-1 m-2</sub>

E-1 m-1 ACETYPE=NAM  
 LDCTYPE=NAM  
 m-2 ACETYPE=NOM  
 LDCTYPE=NOM  
 ATR=YES

- object complements (Ital. complemento predicativo dell'oggetto): a complement that is used to predicate a description of a direct object.

[La]<sub>E-1 m-1</sub> chiameranno [Fondazione Kessler]<sub>E-1 m-2</sub>

E-1 m-1 ACETYPE=PRO  
 LDCTYPE=PRO  
 m-2 ACETYPE=NAM  
 LDCTYPE=NAM  
 ATR=YES

- in appositional constructions:

[[Il Santa Chiara]<sub>E-1 m-1</sub>, [l'ospedale cittadino]<sub>E-1 m-2</sub>]<sub>E-1 m-3</sub>, apre un nuovo reparto

E-1 m-1 ACETYPE=NAM  
 LDCTYPE=NAM  
 m-2 ACETYPE=NOM  
 LDCTYPE=NOM  
 ATR=YES  
 m-3 ACETYPE=NAM  
 LDCTYPE=APP (the extent of the head is the whole MENTION)

- in ARC constructions:

[[La Roma]<sub>E-1 m-1</sub>, [padrona di casa]<sub>E-1 m-2</sub>, [che]<sub>E-1 m-3</sub> ha nettamente perso]<sub>E-1 m-4</sub> ha contestato l'arbitraggio

E-1 m-1 ACETYPE=NAM  
 LDCTYPE=NAM  
 m-2 ACETYPE=NOM  
 LDCTYPE=BAR  
 ATR=YES  
 m-3 ACETYPE=PRO  
 LDCTYPE=WHQ

m-4 ACETYPE=NAM  
LDCTYPE=ARC

- preceded by expressions such as in *qualità di, come (= as)*:

[La CGIL]<sub>E-1 m-1</sub> opera come [organizzazione sindacale]<sub>E-1 m-2</sub> dal 1906

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
m-2 ACETYPE=NOM  
LDCTYPE=BAR  
ATR=YES

## 4. PRONOUNS AND ADJECTIVES

### 4.1 Personal pronouns

They are all ACETYPE = PRO and LDCTYPE = PRO.

*Le date del tour dei Rem sono disponibili sul [loro] sito web*

### 4.2 Possessive adjectives and pronouns

Unlike in English, in Italian possessive adjectives and possessive pronouns have the same form.

They are both annotated with ACETYPE and LDCTYPE PRO.

[*La Serie A*]<sub>E-1 m-1</sub> *inizia oggi il* [*suo*]<sub>E-1 m-2</sub> *campionato*  
E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
m-2 ACETYPE=PRO  
LDCTYPE=PRO

NB: “a sua volta” is considered an idiomatic expression so the possessive is not annotated.

### 4.3 Indefinite pronouns

An indefinite pronoun is a pronoun referring to an identifiable but not specified person or thing. Examples of English indefinite pronouns are *any, none, some*, etc. When they refer to people, they are annotated with ACETYPE = PRO but, for the LDCTYPE, we have two options (see Section 2.4):

1. LDCTYPE = PRO. Pronouns whose form differs from the corresponding adjective: *alcuno, chiunque, ciascuno, nessuno, ognuno, qualcun altro, qualcuno, quelli/e, quest'ultimo, questi/e*
2. LDCTYPE = HLS. Indefinite pronouns having the same form as their corresponding adjectives: *alcuni, altri, entrambi, gli stessi, molti, quanti, tutt'e due, tutti, tutti gli altri*.

### 4.4 Clitics

Clitics referring to organizations are always annotated with ACETYPE PRO. As far as the LDCTYPE is concerned, annotation varies depending on whether they are attached to another word (verb, pronoun or adverb *ecco*) or else separate:

- if they are separate, they have LDCTYPE = PRO
- if they are attached they have
  - LDCTYPE = ENCLIT, when they follow the verb ([*incontrarli*]/to meet them) or the adverb *ecco* ([*Eccolo*]/!Here it is!)
  - LDCTYPE = PROCLIT, when they precede the verb ([*glielo*] hanno inviato/They sent it to them)

*Sono andati a* [*ascoltarli*]<sub>E-1 m-1</sub> *due mesi fa in concerto*  
E-1 (SPC) m-1 ACETYPE=PRO

LDCTYPE=ENCLIT

*La compagnia avrebbe voluto esibirsi, ma [glielo]<sub>E-1 m-1</sub> hanno impedito*

E-1 (SPC) m-1 ACETYPE=PRO  
LDCTYPE=PROCLIT

*[Li]<sub>E-1 m-1</sub> ho ascoltati ieri a teatro.*

E-1 (SPC) m-1 ACETYPE=PRO  
LDCTYPE=PRO

Please notice that with clitics attached to another word, the extent of the MENTION includes both the word and the pronoun, and has the pronoun as head.

When used with the past participle (without auxiliary verb), the infinitive, the gerund and the imperative, clitics are always attached to the verb.

## 4.5 Reflexive pronouns.

Reflexive pronouns can be of four different types:

1. **truly reflexive**, “riflessivi propri”: the object of the action is the same as the subject

*[La società]<sub>E-1 m-1</sub> [si]<sub>E-1 m-2</sub> dice soddisfatta dei guadagni*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM  
m-2 ACETYPE=PRO  
LDCTYPE=PRO

2. **reciprocal**, “riflessivi reciproci”: express mutual action or relationship among the referents of a plural subject, i.e. *each other* in English

*[Quelle due società]<sub>E-1 m-1</sub> [si]<sub>E-1 m-2</sub> contendono le azioni*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM  
m-2 ACETYPE=PRO  
LDCTYPE=PRO

3. **benefactive**, “riflessivi benefattivi”: the focus refers to the person or thing an action is being done for

*[La ditta]<sub>E-1 m-1</sub> [si]<sub>E-1 m-2</sub> è assicurata l'appalto*

E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM  
m-2 ACETYPE=PRO  
LDCTYPE=PRO

4. **pseudo-reflexive**, “riflessivi impropri”: occur with intransitive pronominal verbs  
*Si sta già occupando del caso* (“occuparsi” is an intrans. pronominal verb)

We annotate all reflexive pronouns with the only exception of the pseudo-reflexive pronouns.

## 4.6 Synthesis of the possible interpretations of *si*

For each possible interpretation of the pronoun *si*, Table 8 shows whether it is annotated or not in I-CAB.

**Table 8:** Annotation of the different types of *si*

|                         |     |
|-------------------------|-----|
| Truly reflexive pronoun | YES |
| Reciprocal pronoun      | YES |
| Benefactive pronoun     | YES |
| Pseudo-reflexive        | NO  |



## APPENDIX A: Special Cases

### Titles of books, CDs and exhibitions

Names of organizations appearing in book/CD titles, in names of events, etc., are not annotated, consistently with the criteria adopted for the annotation of temporal expressions<sup>6</sup>.

### 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.

### Organization versus other semantic TYPES of ENTITIES

There are some particular cases in which groups of people defined by an established organizational structure are annotated as Person ENTITIES or Geo-Political ENTITIES and not as Organization ENTITIES.

#### a. Organization versus Person ENTITIES

##### Political Parties

Political parties are usually ENTITIES of TYPE ORG and SUBTYPE Non Governmental. When the text mentions the people belonging to the party (e.g. *diessini*, *leghisti*) instead of the name of the party itself (e.g. *DS*, *Lega Nord*), the entity is classified either as a person or an organization, depending on the context:

- if the text refers to the party as a whole, or to its directives, we have an organization  
*[I leghisti] hanno votato contro l'emendamento*
- 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] hanno abbandonato l'aula*

In this example, the text refers to the people of the green party who were present in the chamber and left it.

##### 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.

ORG: *La maglietta [della Juventus] è in vendita nei negozi specializzati*

PER: *[La Juventus] ha fatto due gol ieri*

---

<sup>6</sup> See Lavelli et al., 2005.

### Authorities

An authority can be a person or an organization having official power so it is annotated as PER or as ORG depending on the context.

PER: *All'incontro sono intervenute [le autorità]*

ORG: *Vogliono aprire il dialogo con [le autorità civili della nazione]*

### Other specific cases

See Table 10 and 11 for other specific cases.

**Table 10:** Specific 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 11:** Specific cases of Organization

| ORGANIZATION  |
|---|
| cattolici   |
| CDA / organi di società                                   |
| Cori, orchestre, bande, accademie                         |
| dirigenza   |
| Guardia di Finanza, Carabinieri, il 118, Vigili del Fuoco |
| forze politiche   |
| giuria  |
| gruppi musicali (REM), band                               |
| esecutivo   |
| guardia nazionale irachena                                |
| militari fascisti   |
| presidenza europea, presidenza olandese                   |
| Procura   |
| resistenza islamica = HAMAS                               |
| sindacati, parti sociali                                  |

## b. Organization versus Geo-Political ENTITIES

According to the ACE-LDC guidelines, “Geo-Political Entities are composite ENTITIES comprised of a population, a government, a physical location, and a nation (or province, state, county, city, etc.). All MENTIONS of these four aspects of GPEs will be marked as GPE and co-referenced.” (LDC 2005, p. 13)

*[La Francia] è [una grande nazione]*

*Il G8 si riunirà a [Mosca]*

*[L'Unione Europea] ha approvato il progetto*

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*

## Nickname Metonymy

This kind of 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 and indicate that this is an example of Nickname Metonymy by selecting the corresponding check box, follow the ACE-LDC guidelines (LDC 2005, p. 33).

The most common example of Nickname Metonymy is when the name of a Geo-Political Entity is used to refer to a sport team:

[*La Russia*]<sub>E-1 m-1</sub> *ha conquistato la medaglia d'oro*  
E-1 (ORG-Spo.) m-1 ACETYPE=NAM  
LDCTYPE=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 an unique appositional construction. The following examples illustrate some of the guidelines we followed.

[*Quella onlus*]<sub>E-1 m-1</sub> (*con sede in Italia*) *è operativa in Somalia*  
E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

[*Quella onlus*]<sub>E-1 m-1</sub> (*con sede in Italia*), *molto attiva, è operativa in Somalia*  
E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM

[*Quella onlus* (*con sede in Italia*) [*che*]<sub>E-1 m-2</sub> *è molto attiva*]<sub>E-1 m-1</sub> *opera in Somalia*  
E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM  
m-2 ACETYPE=PRO  
LDCTYPE=WHQ

[*Quella onlus*]<sub>E-1 m-1</sub> ([*che*]<sub>E-1 m-2</sub> *è molto attiva*) *opera in Somalia*  
E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM  
m-2 ACETYPE=PRO  
LDCTYPE=WHQ

[*Quella onlus*]<sub>E-1 m-1</sub> - [*che*]<sub>E-1 m-2</sub> *è molto attiva - è operativa Somalia*  
E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM  
m-2 ACETYPE=PRO  
LDCTYPE=WHQ

[*Due aziende*]<sub>E-1 m-1</sub> - [*ditte nel settore immobiliare*]<sub>E-1 m-2</sub> - *concludono oggi un accordo*  
E-1 m-1 ACETYPE=NOM  
LDCTYPE=NOM  
m-2 ACETYPE=NOM

LDCTYPE=BAR  
ATR=YES

[*La Brigata Friuli*] <sub>E-1 m-1</sub> (*[reggimento italiano]* <sub>E-1 m-2</sub>) *tornerà presto in Italia*

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
m-2 ACETYPE=NOM  
LDCTYPE=BAR  
ATR=YES

[*La Brigata Friuli*] <sub>E-1 m-1</sub> (*300 uomini*) (*[reggimento valoroso]* <sub>E-1 m-2</sub>) *tornerà presto in Italia*

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
m-2 ACETYPE=NOM  
LDCTYPE=BAR  
ATR=YES

[[*Inter*] <sub>E-1 m-1</sub> (*30 punti*), [*Palermo*] <sub>E-2 m-1</sub> (*27*) e [*Roma*] <sub>E-3 m-1</sub> (*26*)] <sub>E-4 m-1</sub>

E-1 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
E-2 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
E-3 m-1 ACETYPE=NAM  
LDCTYPE=NAM  
E-4 m-1 ACETYPE=NAM  
LDCTYPE=CONJ

### “Circa” and “almeno”

These two adverbs are included in the MENTION extent.

*Erano presenti alla fiera del tessile [circa 10 aziende]  
[Almeno 2 laboratori medici] hanno aderito alle sciopero*



## **APPENDIX B: Inter-annotator agreement**

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

Results are as follows:

- the Dice coefficient for organization entity detection is 0.857;
- limited to the entities detected by both annotators, the Dice coefficient for mention detection is 0.845;
- 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.



## APPENDIX C: Statistical Data

### General statistics

|   | Training | Test   | Total   |
|---|----------|--------|---------|
| Number of files                             | 335      | 190    | 525     |
| Number of words                             | 113,634  | 68,930 | 182,564 |
| Average number of words per file            | 339      | 362    | 347     |
| Number of Organization ENTITIES             | 2,236    | 1,048  | 3,284   |
| Number of MENTIONS of Organization ENTITIES | 4,336    | 1,965  | 6,301   |

### Distribution of Organization ENTITIES by semantic SUBTYPE

|                  | Training     | Test         | Total        |
|------------------|--------------|--------------|--------------|
| Government       | 331 (14.8%)  | 171 (16.3%)  | 502 (15.3%)  |
| Commercial       | 479 (21.4%)  | 202 (19.3%)  | 681 (20.7%)  |
| Educational      | 159 (7.1%)   | 94 (9.0%)    | 253 (7.7%)   |
| Entertainment    | 106 (4.7%)   | 18 (1.7%)    | 124 (3.8%)   |
| Non Governmental | 405 (18.1%)  | 122 (11.6%)  | 527 (16.0%)  |
| Media            | 47 (2.1%)    | 21 (2.0%)    | 68 (2.1%)    |
| Religious        | 30 (1.4%)    | 18 (1.7%)    | 48 (1.5%)    |
| Medical Science  | 54 (2.4%)    | 24 (2.3%)    | 78 (2.4%)    |
| Sports           | 591 (26.4%)  | 360 (34.4%)  | 951 (28.9%)  |
| Mixed            | 34 (1.5%)    | 18 (1.7%)    | 52 (1.6%)    |
| <b>TOTAL</b>     | <b>2,236</b> | <b>1,048</b> | <b>3,284</b> |

### Distribution of Organization ENTITIES by reference CLASS

|              | Training      | Test          | Total         |
|--------------|---------------|---------------|---------------|
| SPC          | 2,096 (93.7%) | 1,009 (96.3%) | 3,105 (94.5%) |
| USP          | 40 (1.9%)     | 11 (1.0%)     | 51 (1.6%)     |
| GEN          | 97 (4.3%)     | 28 (2.7%)     | 125 (3.8%)    |
| NEG          | 3 (0.1%)      | 0 (0.0%)      | 3 (0.1%)      |
| <b>TOTAL</b> | <b>2,236</b>  | <b>1,048</b>  | <b>3,284</b>  |

### Statistics on values of the ACETYPE attribute

- NAM

|          | # OCCURRENCES         |
|----------|-----------------------|
| Training | 2,702/4,336 (62.32 %) |
| Test     | 1,223/1,965 (62.24 %) |



- NOM

|          | # OCCURRENCES         |
|----------|-----------------------|
| Training | 1,252/4,336 (28.87 %) |
| Test     | 541/1,965 (27.53 %)   |

- PRO

|          | # OCCURRENCES      |
|----------|--------------------|
| Training | 359/4,336 (8.28 %) |
| Test     | 189/1,965 (9.62 %) |

- MIX

|          | # OCCURRENCES     |
|----------|-------------------|
| Training | 23/4,336 (0.53 %) |
| Test     | 12/1,965 (0.61 %) |

## Statistics on values of the LDCTYPE attribute

- APP

|          | # OCCURRENCES      |
|----------|--------------------|
| Training | 104/4,336 (2.40 %) |
| Test     | 43/1,965 (2.19 %)  |

- CONJ

|          | # OCCURRENCES      |
|----------|--------------------|
| Training | 201/4,336 (4.63 %) |
| Test     | 104/1,965 (5.29 %) |

- PROCLIT + ENCLIT

|          | # OCCURRENCES          |
|----------|------------------------|
| Training | 9 (0+9)/4,336 (0.21 %) |
| Test     | 1 (1+0)/1,965 (0.05 %) |

- 43+NOM

|          | # OCCURRENCES       |
|----------|---------------------|
| Training | 956/4,336 (22.05 %) |
| Test     | 416/1,965 (21.17 %) |

- PRO

|          | # OCCURRENCES      |
|----------|--------------------|
| Training | 162/4,336 (3.74 %) |
| Test     | 84/1,965 (4.27 %)  |

- BAR

|          | # OCCURRENCES      |
|----------|--------------------|
| Training | 178/4,336 (4.10 %) |

|              |                       |
|--------------|-----------------------|
| Test         | 66/1,965 (3.36 %)     |
| <b>• ARC</b> |                       |
|              | # OCCURRENCES         |
| Training     | 21/4,336 (0.48 %)     |
| Test         | 6/1,965 (0.31%)       |
| <b>• PTV</b> |                       |
|              | # OCCURRENCES         |
| Training     | 6/4,336 (0.14 %)      |
| Test         | 6/1,965 (0.31 %)      |
| <b>• NAM</b> |                       |
|              | # OCCURRENCES         |
| Training     | 2,520/4,336 (58.12 %) |
| Test         | 1,143/1,965 (58.17 %) |
| <b>• WHQ</b> |                       |
|              | # OCCURRENCES         |
| Training     | 167/4,336 (3.85 %)    |
| Test         | 80/1,965 (4.07%)      |
| <b>• HLS</b> |                       |
|              | # OCCURRENCES         |
| Training     | 12/4,336 (0.28%)      |
| Test         | 16/1,965 (0.81%)      |

**Number of attributive MENTIONS (ATR="yes")**

|          |                   |
|----------|-------------------|
|          | # OCCURRENCES     |
| Training | 236/4,336 (5.44%) |
| Test     | 79/1,965 (4.02%)  |

## APPENDIX D: 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

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

### 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

### 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

### 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

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://www.demauroparavia.it/>
- (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](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/>