TUT: LINGUISTIC NOTES

TABLE OF CONTENTS

0. Preface

These notes describe and motivate the representation of the syntactic structure of complex linguistic expressions according to TUT annotation schema. When needed, some quotations are reported to support the choices we have made.

The notes are especially thought as a support for the activity of human annotators (which, in face of a piece of text, must determine the correct links among words), and to any user of the treebank, for understanding how dependencies can be used to face difficult linguistic problems.

At the beginning developed for Italian, which is the reference language of TUT, they have been extended with some observation on English in order to support the comprehension of the annotation scheme of non-Italian speakers (see also TUT English corpus in the web site).

1. Traces and co-indexing

1.1. Introduction

Traces are elements lexically empty, but they are useful to represent explicitly the arguments of predicates (i.e. complements of verbs, nouns, etc.). Traces are useful to express deletions (as in pro-drop or equi) or movement (as in raising). Some examples are:

1.1 *Amo Maria ([I] love Maria)*
1.2 *Luigi ama Maria e Marco Claudia (Luigi loves Maria and Marco Claudia)*
1.3 *Luigi vuole incontrare Maria (Luigi wants to meet Maria)*
1.4 *Luigi la vuole incontrare (Luigi her wants to meet: Luigi wants to meet her)*

The representation of traces is done inserting in the annotation extra lines modelled on the basis of the lines of lexically realized words having the following form:

> ***i word (syntinfo) govinfo***

where i is the line number (i.e. index of the position in the linear order of the sentence) of the word, **syntinfo** are the morpho-syntactic info of the word (see "The syntactic categories" at the web site), and **govinfo** is the link to the governor, which is expressed in the usual form (i.e. line number of the governor and link label).

The general form of a trace line instead is:

> ***i.n T [js] (syntinfo) govinfo***

where i is the index of the governor, **n** is the trace counter (an integer which is 10 for the first trace, 11 for a possible second trace on the same index, etc.), j is the index of the referent of the trace, **s** is the trace type (see below), **syntinfo** are a copy of the syntinfo of the referent (if any), and **govinfo** is the link to the governor, expressed in the usual form. In some cases, there is no referent and **syntinfo** takes a special form that will be described below.

The trace type can be (see the examples in the following subsections):
-   f (full): it means that the trace refers to the full subtree rooted in the referent;
-   w (word): it means that the trace refers just to the single referent word;
-   p (partial): it means that the trace refers to the part of the subtree rooted in the referent which does not include the trace itself.

In case the referent is absent, j and **s** are empty and **syntinfo** can assume three different forms:
-   generic: (GENERIC-T PRON PERS gender number person)
    For generic traces, "gender" "number" and "person" usually are ALLVAL (which is the value which unifies with anything). However, in some cases, agreement can force them to

assume specific values: this does not happen for subject of infinitives, but the values are recoverable from finite verbs.

Ex: Usando T un nuovo test, si è visto che ... (ALLVAL ALLVAL ALLVAL)
(By using T a new test, it has been noticed that)
In tutti i paesi vediamo T nuove chances di democrazia (ALLVAL PL 1)
(In all countries, T see[pl,1st] new chances of democracy)

- anaphoric: (ANAPH-T PRON PERS gender number person)
Anaphoric traces are used for intersentential references. They usually have well-defined values for "gender" "number" and "person". They are determined either by manual inspection of the context, or by means of agreement constraints.
Ex. E' un prodotto per VIP che ... (M SING 3)
(Is T a product for VIP that ...)

- deictic: (DEITT-T PRON PERS gender number person)
Deictic traces are reserved for first and second person traces. Again, this is determined on the basis of the person of the verb.
Ex. Dico T solo che ... (ALLVAL SING 1)
*(Say[sing,1st] T only that ...)*

In the figures, we do not report the whole trace line, but we insert only the **T** marker followed by square brackets and the co-reference index (if exist) within them. Instead, an index I appearing within square brackets to the right of a word W indicates that W plays the role of referent for the trace co-referenced by I.

1.2. Pro-drop (the missing subject)

Italian is a pro-drop language, so that the subject is often missing. In such a case, the actual subject can either appear in some previous sentence, or simply be some "generic", not specified, entity.

In the present version of the treebank, we do not account for intersentential links; this is also due to the fact that the initial nucleus of the treebank included isolated sentences instead of whole pieces of text. So, many traces concerning missing subjects have a 'generic' reference.

Example 1.5: "Era andato a Torino" ([] had gone[m,sing] to Torino)[1]
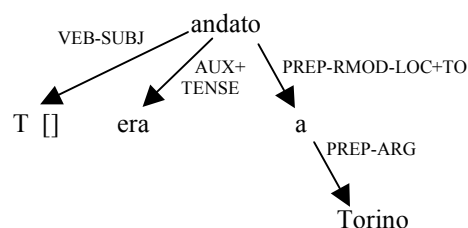
1.3. Modals

For modals, we assume that the modal verb governs the infinite verb via a *VERB+MODAL-INDCOMPL* link. Of course, since the subordinate verb is in the infinitive, it lacks the subject; so, a trace is inserted, whose referent is the subject of the modal.

---

[1] As it is usual in drawing trees, the links between nodes are represented as downward arrows. However, in the treebank, they are implemented as pointers from the dependent to the governor.

Example 1.6: "... il Papa vuole sottolineare il legame ..."
(... the Pope wants to stress the link ...”

This is an example of "full" (f) co-indexing
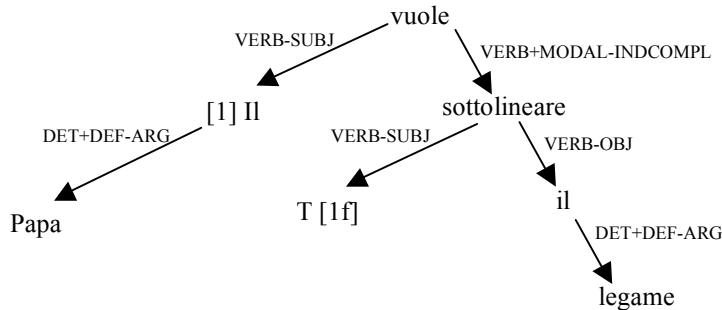
Slightly more complex is the case where the subordinate clause is passive:

Example 1.7: "Questo può essere
paragonato a ..."
(This may be compared
to ...”
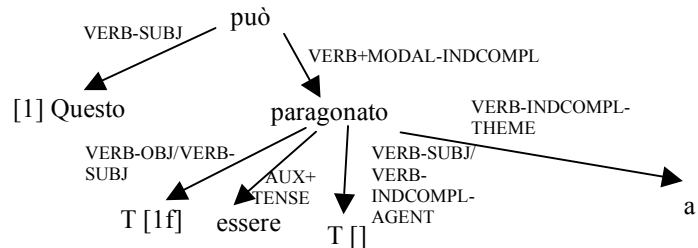
Remember that, although our approach is monostratal, we record in the label the results of some transformation, so that OBJ/SUBJ must be read as "deep OBJ, surface SUBJ". Of course, the second trace is not needed in case the agent complement appears explicitly.

1.4. Gapping

Gapping involves "word" traces, since single words, instead of whole constituents are cancelled.

Example 1.8: "l'esercito entro' a Belgrado e la polizia a Tirana"
(the army entered [in] Belgrado and the police [in] Tirana)
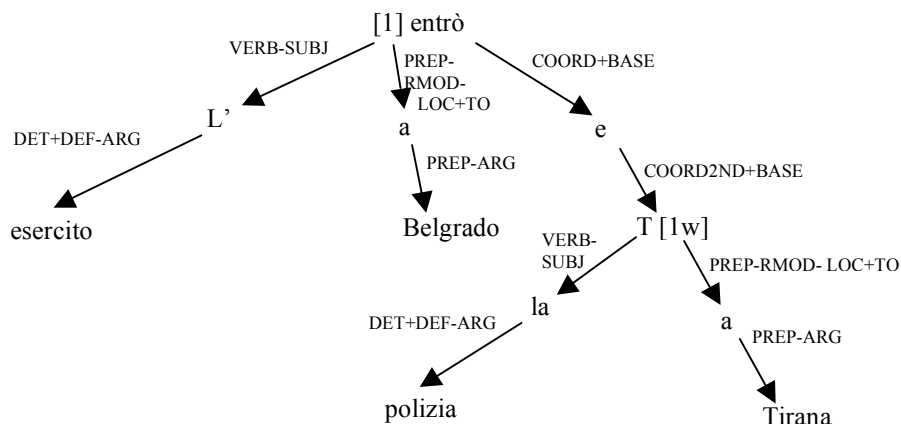
In some cases, more than one word is gapped:

Example 1.9: "l'esercito decise di entrare a Belgrado e la polizia a Tirana"
(the army decided to enter [in] Belgrado and the police [in]
Tirana)
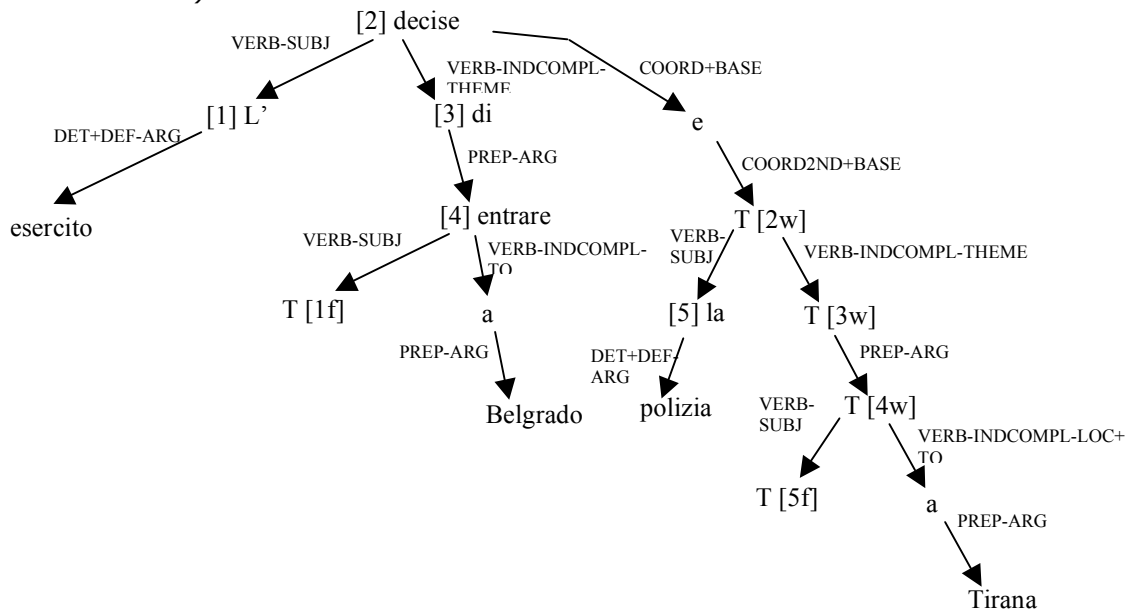
1.5. Co-reference and pronouns

Co-reference is not limited to traces. It concerns also anaphora. The next sentence includes a first example of comparative structures. Comparatives will be handled in more detail in section 12.1.

Example 1.10: "... un test piu' raffinato di quello solito ..."
(... a test more sophisticated than that usual ... – than the usual
one)
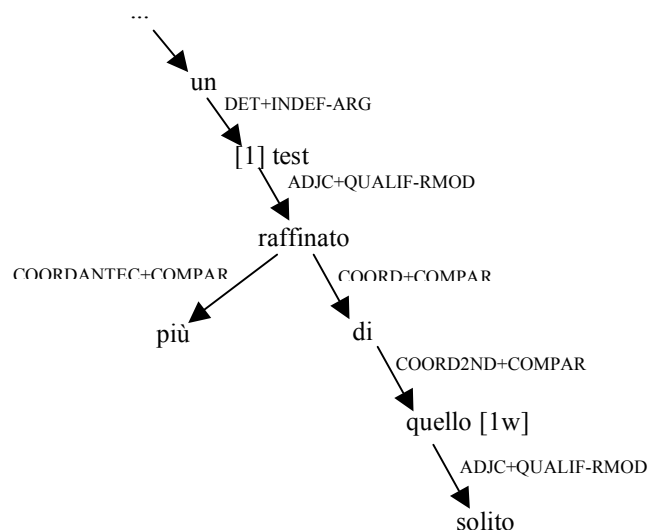
This is another example of "word" (w) co-indexing.

Example 1.11: "... con mezzi alternativi a quelli stradali ..." (with vehicles alternative to the road ones)

```
                         …
                          ↓
                         con
                          │ PREP-ARG
                          ↓
                      [1] mezzi
                           │ ADJC+QUALIF-RMOD
                           ↓
                       alternativi
                            │ ADJC-ARG
                            ↓
                            a
                             │ PREP-ARG
                             ↓
                         quelli [1w]
                              │ ADJC+QUALIF-RMOD
                              ↓
                           stradali
```

Here, we have word co-indexing, but in this case it would be equivalent to adopt a partial (p) co-indexing, since the only subtree governed by "mezzi" is the one including the co-indexed word.

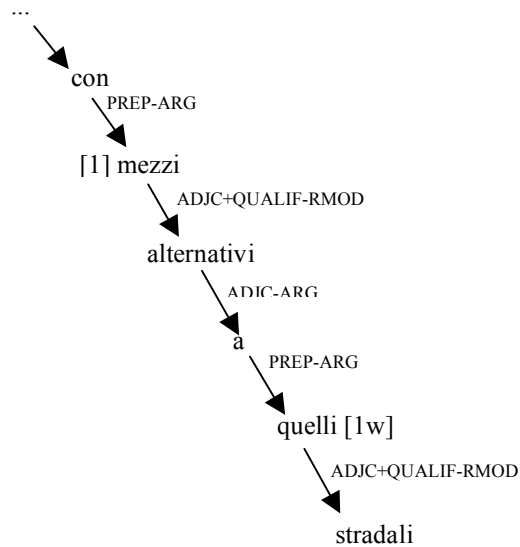Differently from the previous example, "word" (w) co-indexing is required here. In fact, "aerei" must not be included in the referent. With "p" co-indexing, we would get " ... alternativi a 'mezzi aerei' stradali ...", which is not the intended meaning.

Example 1.12: "... con mezzi aerei alternativi a quelli stradali ..." (with air vehicles alternative to the road ones)

```
              …
               ↓
              con
               │ PREP-ARG
               ↓
           [1] mezzi
   ADJC+QUALIF-RMOD ╱    ╲ ADJC+QUALIF-RMOD
               aerei   alternativi
                            │ ADJC-ARG
                            ↓
                            a
                             │ PREP-ARG
                             ↓
                         quelli [1w]
                              │ ADJC+QUALIF-RMOD
                              ↓
                           stradali
```
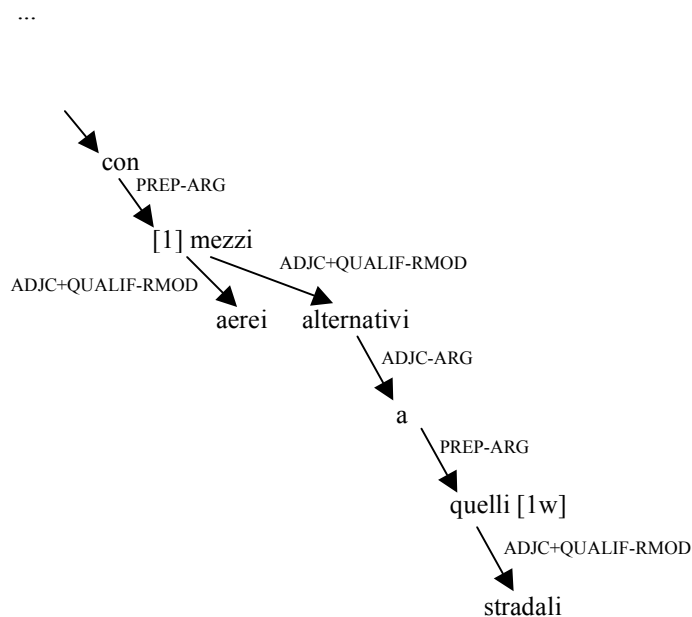
Example 1.13: "... con mezzi di soccorso alternativi a quelli stradali ..." (with rescue vehicles alternative to the road ones)
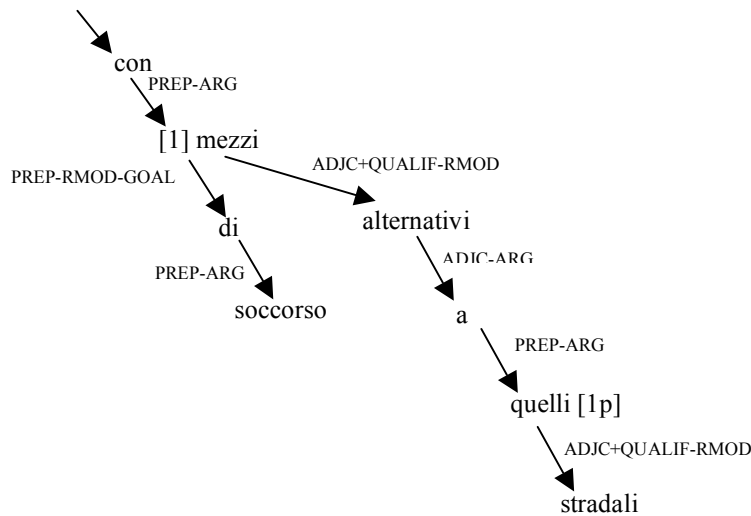
This is the opposite, with respect to Ex. 1.11; in fact the co-indexing should enable to solve "quelli stradali" as "mezzi di soccorso stradali" (i.e. by also copying the first subtree governed by "mezzi")

2. Punctuation and parentheticals

2.1. Introduction

Punctuation marks are not very relevant from a linguistic point of view, since its main task is just to provide information about the syntactic structure. When the markers are used to delimit a constituent, they will be attached to the governor of that constituent.
In the case of markers of end-of-sentence (periods, question marks, etc.), they are attached to the head of the sentence.

2.2. Parentheticals

In the case of parentheticals, the surrounding markers (open and closed parenthesis, quotes, commas or hyphens) are attached to the same entry to which the head of the parenthetical is attached.

Example 2.1: "Il professore, probabilmente, era in ritardo."
(The professor, probably, was late)
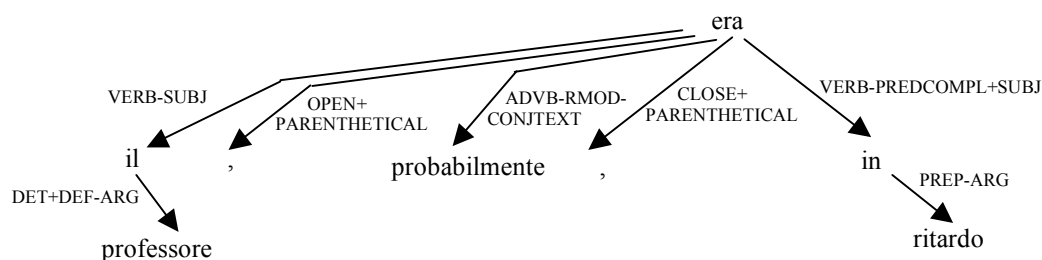
A similar structure is the one marked by quotes. But the OPEN and CLOSE relations (with +QUOTES) are attached to the head of the constituent included between quotes.

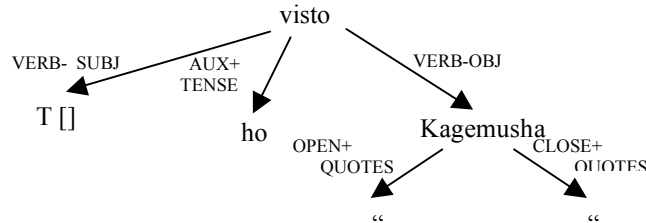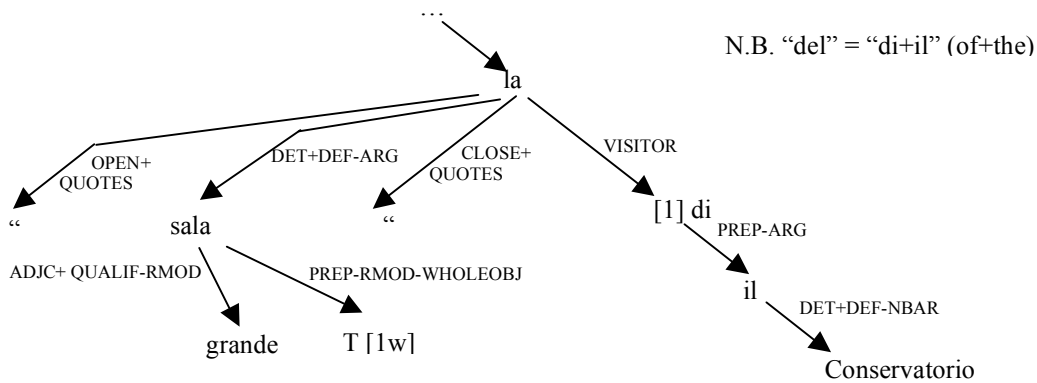*Example 2.2: Ho visto "Kagemusha"*
          ([I] have seen "Kagemusha")

In some cases, however, this approach forces the insertion of a trace; in fact, the quotes may 'separate' a dependent from its governor:

Example 2.3: "la sala grande" del Conservatorio
          ("the big room" of the Conservatory)

N.B. "del" = "di+il" (of+the)

In the previous example, there appears a VISITOR label, which, following Hudson, characterizes a "semantically empty" relation.

2.3. Separators and hyphens

In many cases, punctuation marks (usually commas) play the role of separating two parts of the sentences, perhaps to indicate the presence of some prosodic pattern (e.g. a pause). In these cases, because of the projectivity of the representation, the two parts have a common ancestor (possibly one of them) and there are no crossing arcs; the punctuation mark is attached as a dependent on that ancestor.

Example 2.4: "Dai tedeschi orientali, Krenz è visto come un generale."
          (By the eastern Germans, Krenz is seen as a general.)

9
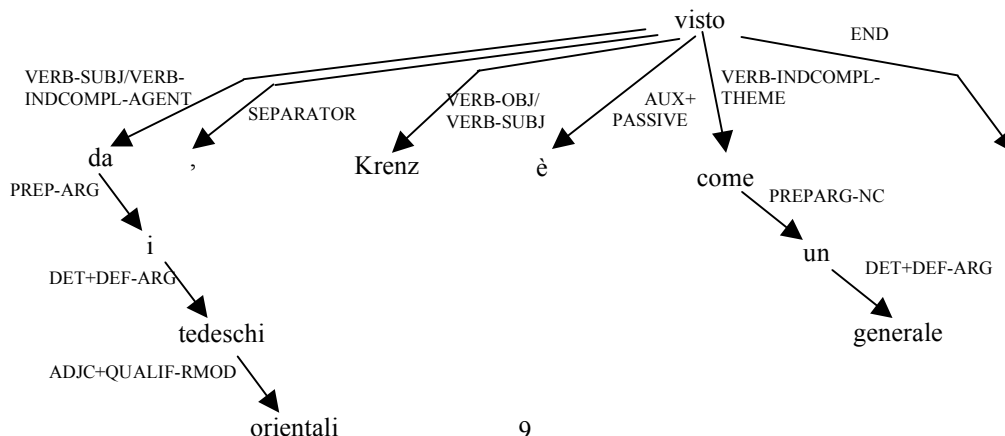
The hyphen that separes two nouns (e.g. "padre-padrone" father-master)), two adjectives (e.g. "italo-americano" (Italo-American)) or a couple of words belonging to different grammatical categories (e.g. "post-comunista" (post-communist)), is linked as a separator to the head word of the couple.

Example 2.5: "padre-padrone"
(father-master)

padre
SEPARATOR   NOUN-APPOSITION
-        padrone
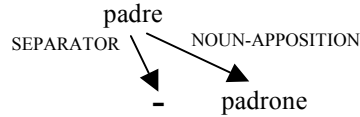
Example 2.6: "La ferrovia Tirana-Durazzo."
(The railway Tirana-Durazzo.)

La
DET+DEF-ARG

ferrovia
ADJC+QUALIF-RMOD

Tirana
SEPARATOR   NOUN-RMOD-DENOM
-        Durazzo
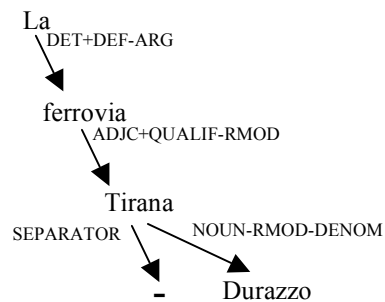
Example 2.7: "Parla di 250-300
milioni."
((It) talks about
250-300 millions.)

Parla
VERB-INDCOMPL-THEME

di
PREP-ARG

250
SEPARATOR   NUM-ARG-RANGE
-        300
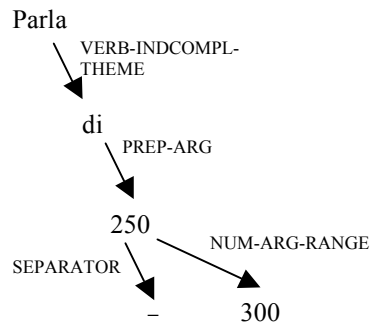
Example 2.8: "Un post-comunista."
(post-communist)

Un
DET+INDEF-ARG

post
SEPARATOR   PREP-ARG
-        comunista
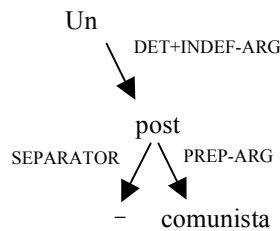
Example 2.9: "Un maxi-impegno."
(maxi-engagement)

Un
DET+INDEF-ARG

impegno
ADJC-RMOD   SEPARATOR
maxi   -
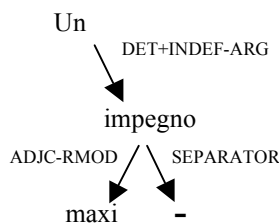
Example 2.10: "Una impresa italo-sloveno-ungherese."
        (an Italo-Sloveno-Hungarian)
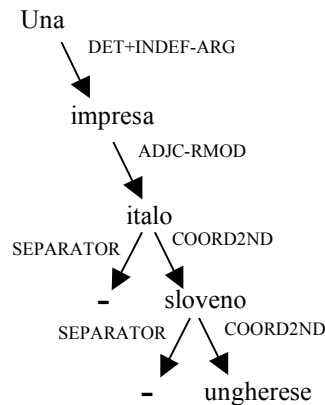
3. Locutions

3.1. Introduction

Locutions are multi-word expressions to which cannot be applied the standard compositionality principle to evaluate their meaning. So, the meaning must be assigned to the locution as a whole: since it is not productive, then it has to be specified explicitly in the dictionary.

LOCUTIONS (Serianni – "Grammatica italiana" - II ed. Torino, UTET, 1991 p.493-512)

From the functional point of view, locutive expressions (formed by more than one word) are equivalent to the expressions belonging to the same grammatical category and formed by a single word. Within the different categories, the set of locutive expressions may be expanded almost indefinitely and its boundaries are not always sharp:

a) Prepositional locutions are composed of two prepositions (usually, the first one is proper, and the second improper, ex. "dietro a" – behind of – "insieme con" – together with) or by a noun with one or more prepositions (ex. "in cima a" – on top of – "nell'intento di" – with the aim of – "per mezzo di" – by means of). As the improper prepositions, locutions enable one to make explicit the semantic content, which would remain implicit if simple prepositions were used ("la lettera l'ho spedita per tua sorella" → "la lettera l'ho spedita a causa di tua sorella" – the letter [it] I have sent for your sister → – the letter [it] I have sent because of your sister).

b) Conjunctive locutions are, for example, "con tutto ciò" (with all of this), "dal momento che" (because of), "visto che" (having seen that)

c) Verbal locutions usually include a noun without article or an adjective (es. "avere bisogno" – have need – andare a cavallo – go with horse: ride – "farsi bello" – to boast – "farsi vivo" – to turn up).

d) Adverbial locutions usually correspond to an adverb ("letteralmente" → "alla lettera" – literally → at the letter), but often the correspondence is illusory, or it is valid just in some cases ("sono letteralmente distrutto" → "*sono distrutto alla lettera" – I'm literally destroyed → I'm destroyed at the letter).

11

It is possible to identify some characteristic types:
- with a preposition ("a stento" - barely, "con sforzo" – with effort, "senza ira" – without rage, "di sicuro" - certainly).The nature of the adverbial phrase becomes more uncertain when there are attributes between the preposition and the noun: "con ogni sforzo" – with any effort, "con ogni possibile sforzo" – with any possible effort, "senza la minima ira" - without the least rage)
- with double preposition ("a faccia a faccia" – face to face, "a pezzo a pezzo" – piece by piece)
- with the prepositions "di ... in" ("di bene in meglio" – better and better, "di tanto in tanto" – from time to time
- doubling a noun ("passo passo" – very slowly, "via via" - gradually)
- doubling an adjective ("bel bello" - unhurriedly)
- doubling an adverb ("or ora" – just now)

In the treebank, we categorize the locutions in two classes: fixed and flexible.

From a semantic point of view, the distinction is somewhat fuzzy: fixed locutions have lost almost completely their compositional flavour, while flexible locutions can be seen as metaphors, where the components have lost (or perhaps shifted) their basic meaning.

From the syntactic point of view, the separation is more clear-cut: fixed locutions are treated as invariable sequences of tokens (and so are handled by the lexical analyser), while flexible locutions are not recognized by the lexical analyser, but by the syntactic processor (parser). Often, fixed locutions have an internal structure incompatible with the standard rules of the grammar, while this does not happen for flexible locutions. Moreover, morphological variability is allowed for flexible locutions, but not for fixed locutions: since "avere bisogno" – have need, i.e. need – is flexible, then the verb can be inflected Finally, fixed locutions cannot include any extra word (since "a pezzo a pezzo" – piece by piece - is a fixed locution, "a piccolo pezzo a piccolo pezzo" – small piece by small piece - cannot be recognized as an instance of that locution), while this does not happen for flexible locutions ("avere bisogno" vs. "avere tanto bisogno" – have much need).

3.2. Representation of fixed locutions

Since within fixed locutions the relations among words are not relevant from the semantic point of view, and since the link labels are viewed as a bridge towards semantics, it is clear that the link labels are, in this case, irrelevant. Similarly, it is not useful to represent explicitly the internal syntactic structure of fixed locutions (if any). So the elements of fixed locutions are linked each other via the arc CONTIN+LOCUT, and they are represented as a simple right-branching structure, where the first word of the locution is the head. All "syntinfo" components of the lines are the same, and the last item within the "syntinfo" list is LOCUTION.

Ex. "in tutto e per tutto" (in all and for all, i.e. completely)
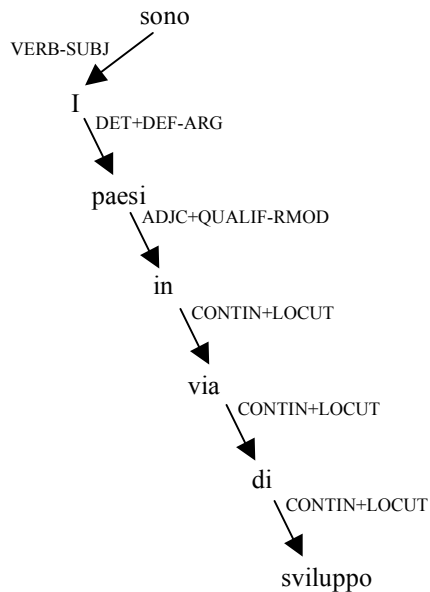1 in (IN_TUTTO_E_PER_TUTTO ADV MANNER LOCUTION) [...;...]
2 tutto (IN_TUTTO_E_PER_TUTTO ADV MANNER LOCUTION) [1;CONTIN+LOCUT]
3 e (IN_TUTTO_E_PER_TUTTO ADV MANNER LOCUTION) [2;CONTIN+LOCUT]
4 per (IN_TUTTO_E_PER_TUTTO ADV MANNER LOCUTION) [3;CONTIN+LOCUT]
5 tutto (IN_TUTTO_E_PER_TUTTO ADV MANNER LOCUTION) [4;CONTIN+LOCUT]

Example 3.1: I paesi in via di
sviluppo sono ...
(The countries in
course of development
are ..., i.e. the
developing countries are)

sono

VERB-SUBJ

I

DET+DEF-ARG

paesi

ADJC+QUALIF-RMOD

in

CONTIN+LOCUT

via

CONTIN+LOCUT

di

CONTIN+LOCUT

sviluppo

For locutions governing an item, the item is linked to the first word of the locution (e.g. in "in mano ai rapitori", "i rapitori" is a dependent of the word "in" as a PREP-ARG), as you can see also in the following example.

Example 3.2: Mario mangiava un po'
di minestra
(Mario was eating a
bit of soup)

mangiava

VERB-SUBJ   VERB-OBJ

Mario   un

CONTIN+LOCUT   DET+QUANTIF-ARG

pò   minestra

CONTIN+LOCUT

di
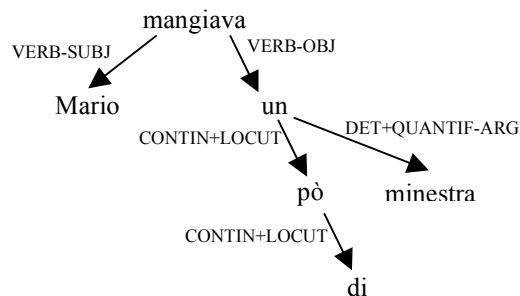
Fixed locutions are also used to represent multi-word names (e.g. "Stati Uniti" – United States -, "San Francisco"), but with CONTIN+DENOM in place of CONTIN+LOCUT.

Example 3.3: Mario abitava negli
Stati Uniti
(Mario was living in
the United States)

abitava

VERB-SUBJ   VERB-INDCOMPL-LOC+IN

Mario   in

PREP-ARG

gli

DET+DEF-ARG

Stati

CONTIN+DENOM

Uniti
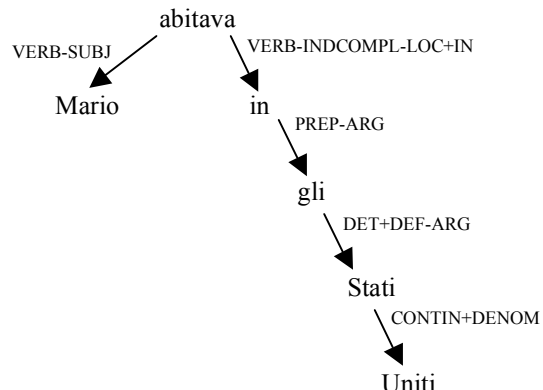
N.B. "negli" = "in+il" (in+the)

3.3. Representation of flexible locutions

Flexible locutions are represented in such a way that both the underlying syntactic structure, their original semantic connection, and their conventional character are made explicit. The links connecting the words are the same as if the locution were not a locution; the semantic flavour is given by the insertion in the link label of the standard label that is used for non-locutions; the fact that it actually is a locution is marked by suffixing the standard functional syntactic label with "+LOCUT".

Example 3.4: Essa venne al mondo in Italia (She came to the world in Italy, i.e. she was born in Italy)

N.B. "al" = "a+il" (to+the)
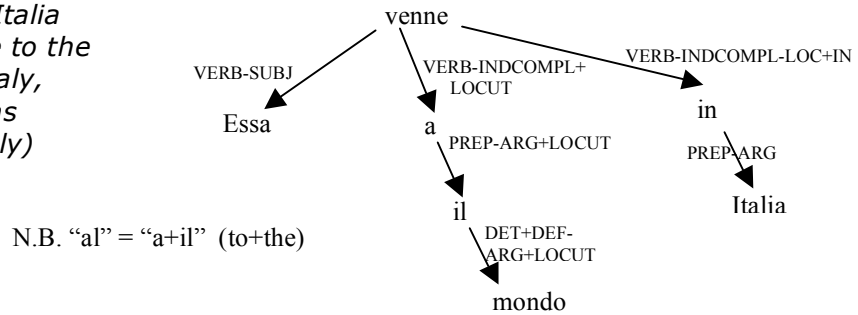
Example 3.5: Avevo bisogno di te (Had [sing, 1ˢᵗ] need of you, i.e. I needed you)
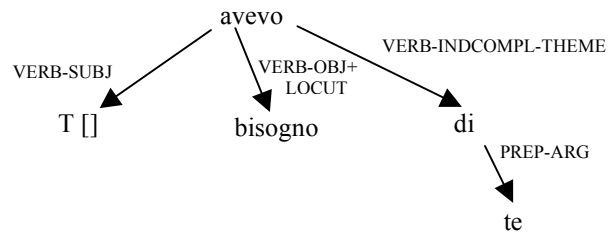
Example 3.6: Ero in linea con le sue idee (Was in line with her [or his] ideas, i.e. I agreed with her)
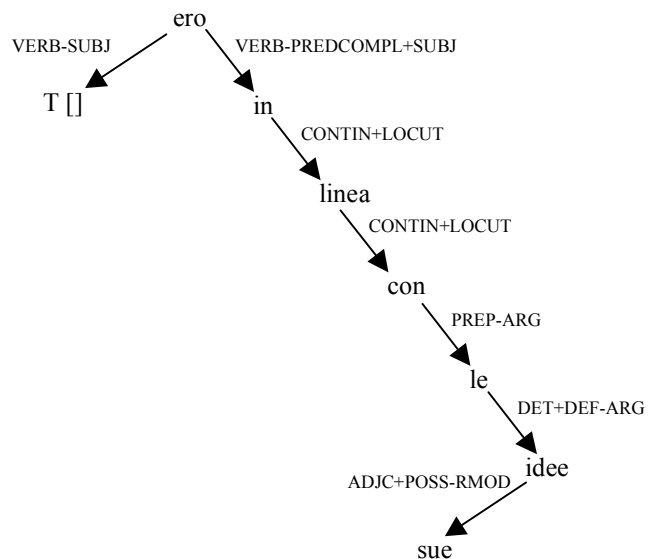
4. Numbers and proper nouns

The reason why numbers and names have been included in the same section is that often numbers play the roles of identifiers. In particular, this happens for one of the interesting use of numbers, i.e. for expressing dates. Before considering that, however, let's see the representation of standard numerals.

4.1. Numbers and counting

Example 4.1: "un esercito di quindicimila uomini"
(an army of fifteen-thousand men)

```
un
  \ DET+INDEF-ARG
   esercito
      \ PREP-RMOD-MEMBEROF
       di
          \ PREP-ARG
           quindicimila
              \ DET+QUANTIF-ARG
               uomini
```
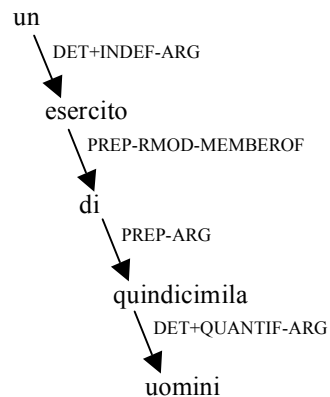
Example 4.2: "in 48 ore" (in 48 hours)

```
in
  \ PREP-ARG
   48
      \ DET+QUANTIF-ARG
       ore
```
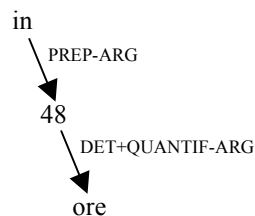
A numeral can be also used in order to indicate the position of an item in a list. In this case we use the relation NUM-RMOD-LISTPOS and we represent the structure as follows.

Example 4.3: "1. Introduction"

```
              Introduction
NUM-RMOD-LISTPOS  /   \ SEPARATOR
              1
```
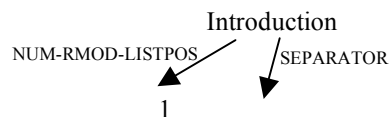
The constructs containing numerals indicating dates are not especially problematic from the syntactic point of view, but a format able to represent their syntactic and semantic similarities can be convenient.

4.2. Numbers, times and dates

All dates or temporal expressions are usually linked to their heads by a modifier (RMOD) grammatical relation including the semantic label TIME and an appropriate morphosyntactic label.

Arrivò ((He/she) came)	presto (early)	ADVB-RMOD-TIME
	ieri sera (yesterday night)	
	alle 8 di sera (at 8 of night, i.e. at 8 p.m.)	PREP-RMOD-TIME
	nelle prime ore del pomeriggio (in the early hours of the afternoon)	
	nel primo trimestre '96 (in the first term `96)	
	la mattina (the morning)	NOUN-RMOD-TIME
	lunedì mattina (Monday morning)	
	giovedì 26 ottobre (thursday 26 October, i.e. thursday October the 26th)	
	il 13 ottobre 1999 (the 13 October 1999, i.e. October the 13th 1999)	

The internal structure of a temporal modifier can include grammatical relations with various semantic labels, in order to indicate that the dependent is the part of the day (PARTOFDAY, e.g. sera (night), mattina (morning)), or the DAY (lunedì (monday), 26), or the MONTH (gennaio (january), maggio (may)) or the YEAR (1996, '96) of its head. These relations are morphosyntactically labeled according to the morphosyntactic features of the word involved (e.g. ieri (yesterday) is an adverb, sera (night) is a noun, '96 is a date, 1996 and 26 are numerals).

Example 4.4: "Arrivò ieri sera"
((He/she) came yesterday night)
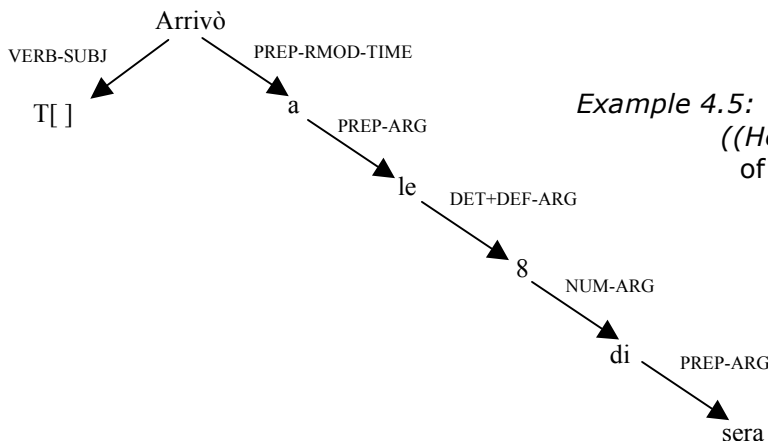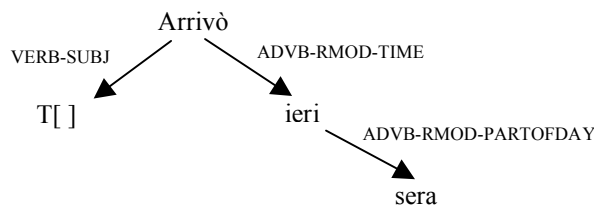
Example 4.5: "Arrivò alle 8 di sera"
((He/she) came at 8 of night, i.e. at 8 p.m.)

Example 4.6: "Arrivò nelle prime ore del pomeriggio*"*
   ((He/she) came in the early hours of the afternoon*)*

```
                        Arrivò
        VERB-SUBJ              PREP-RMOD-TIME
          T[ ]                    ne
                                     PREP-ARG
                                       le
                                          DET+DEF-ARG
                                            ore
                    ADJC-RMOD              PREP-RMOD-PARTOF
                       prime                  de
                                                 PREP-ARG
                                                   il
                                                      DET+DEF-ARG
                                                        pomeriggio
```
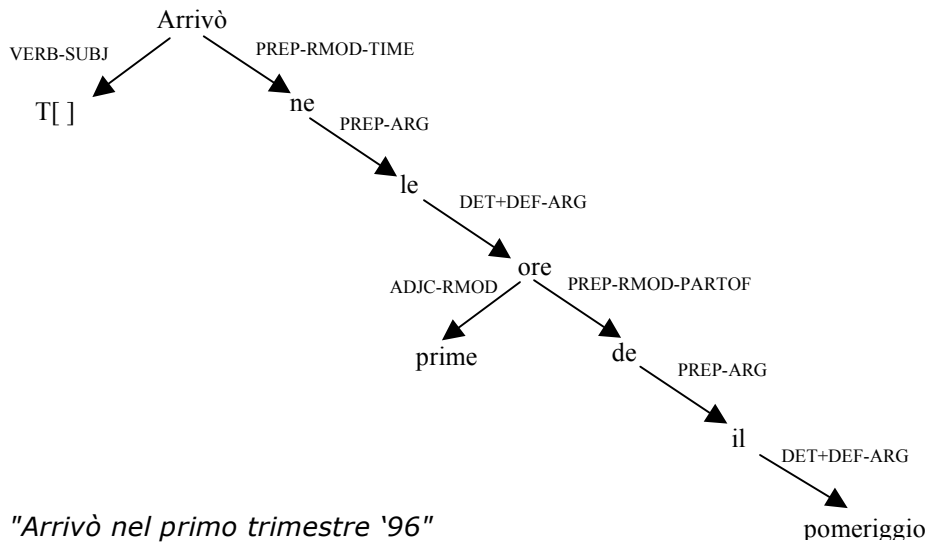
Example 4.7: "Arrivò nel primo trimestre '96"
   ((He/she) came in the first term '96)

```
              Arrivò
  VERB-SUBJ        PREP-RMOD-TIME
    T[ ]              ne
                         PREP-ARG
                           il
                              DET+DEF-ARG
                                trimestre
              ADJC-RMOD           DATE-RMOD-YEAR
                primo               '96
```

But:

```
trimestre
      NUM-RMOD-YEAR
         1996
```
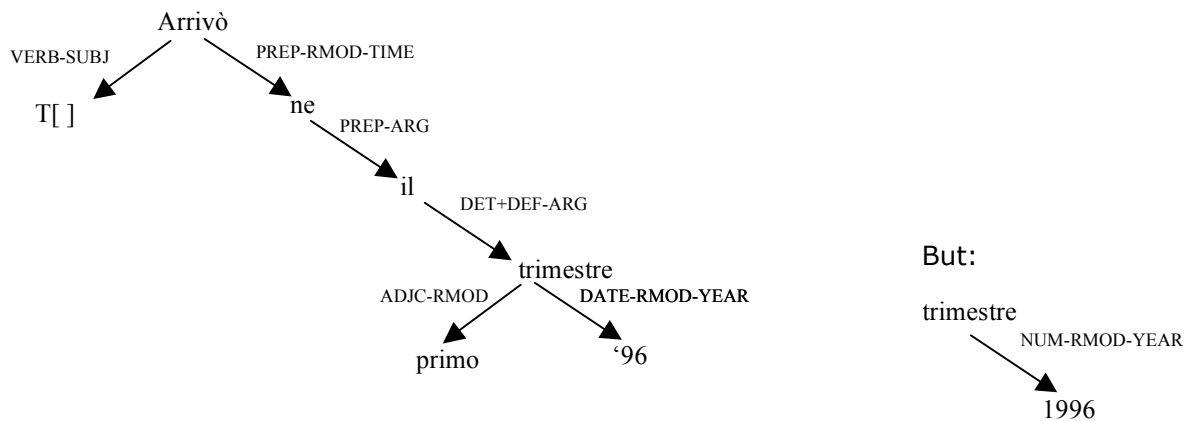
In this way is represented "gli anni '80" (the years '80, i.e. the '80 years)

Example 4.8: "Arrivò la mattina"
   ((He/she) came the morning)

```
              Arrivò
  VERB-SUBJ        NOUN-RMOD-TIME
    T[ ]              la
                         DET+DEF-ARG
                           mattina
```
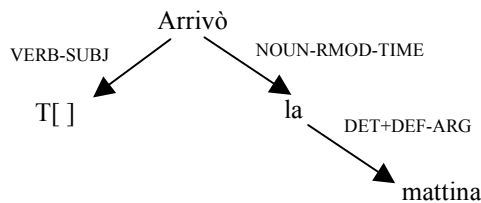
Example 4.9: "Arrivò lunedì mattina*"*
   ((He/she) came Monday morning*)*

```
              Arrivò
  VERB-SUBJ        NOUN-RMOD-TIME
    T[ ]              lunedì
                         ADVB-RMOD-PARTOFDAY
                           mattina
```

17
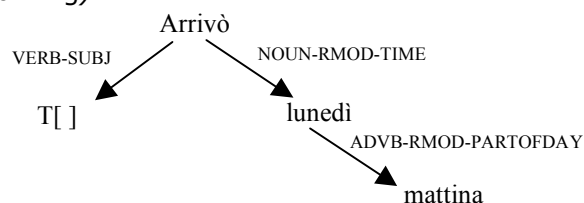
Example 4.10: "Arrivò giovedì 26 ottobre"
   ((He/she) came thursday 26 October, i.e. thursday October the 26th)

Arrivò
VERB-SUBJ — T[ ]
NOUN-RMOD-TIME — giovedì
NUM-RMOD-DAY — 26
NOUN-RMOD-MONTH — ottobre
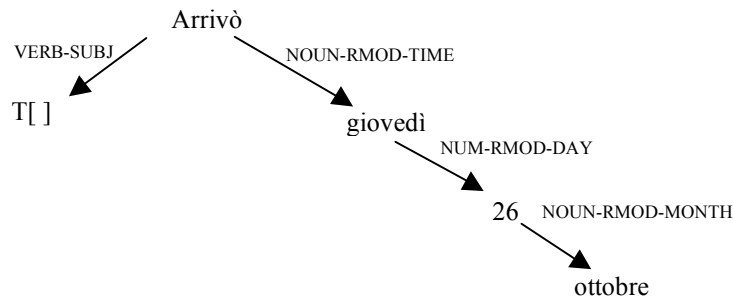
Example 4.11: "Arrivò il 13 ottobre 1999"
   ((He/she) came the 13 October 1999, i.e. October the 13th 1999)

Arrivò
VERB-SUBJ — T[ ]
NOUN-RMOD-TIME — il
DET+DEF-ARG — 13
NOUN-RMOD-MONTH — ottobre
NUM-RMOD-YEAR — 1999
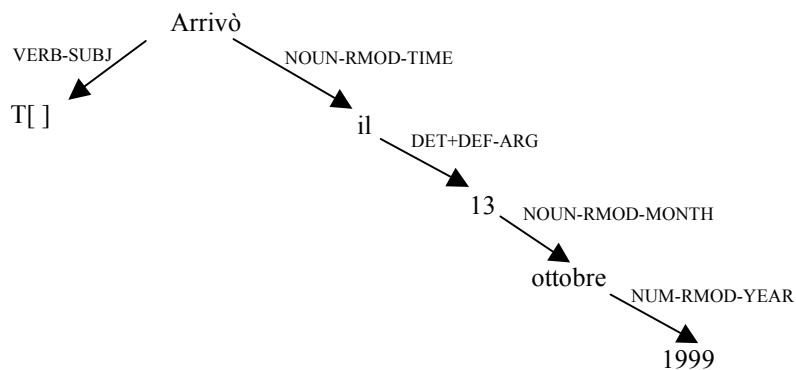
Example 4.12: "Se ne prevede la stampa per il 2003"
   (Impers of-it foresee the print for the 2003, i.e.
   its print is foreseen for 2003)

prevede
SUBJ-IMPERS — Se
VISITOR — [1] ne
VERB-OBJ — la
PREP-RMOD-TIME — per
DET+ DEF-ARG — stampa
NOUN-OBJ — T [1f]
PREP-ARG — il
DET+DEF-ARG — 2003
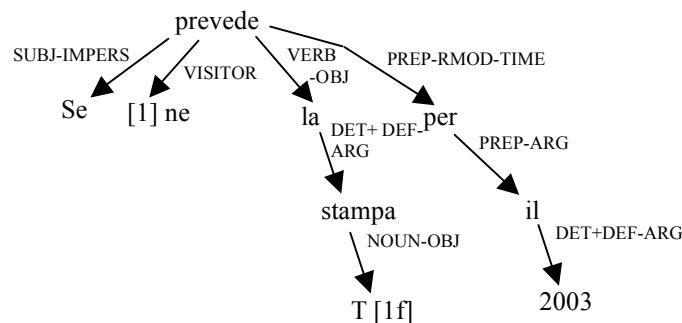
This last example also includes an interesting case of movement: the argument of "stampa" (print) has been moved before the verb into an enclitic position ("ne"). So, the clitic is attached to the verb as a VISITOR (it does not contribute to the selection of the arguments of "foresee"), and a trace has been inserted below "stampa" (the link is labelled as NOUN-OBJ, which marks the OBJ of NOUNS).

4.3. Numbers and percents

Percents are represented this way, Also if the depending preposition is not a true partitive (like in "una quota del 60 % sul totale").

Example 4.13: "L'88 % dei consensi" (the 88 % of agreements)

L'
DET+ DEF-ARG
88
NUM-ARG-PERCENT   NUM-ARG-PARTITIVE
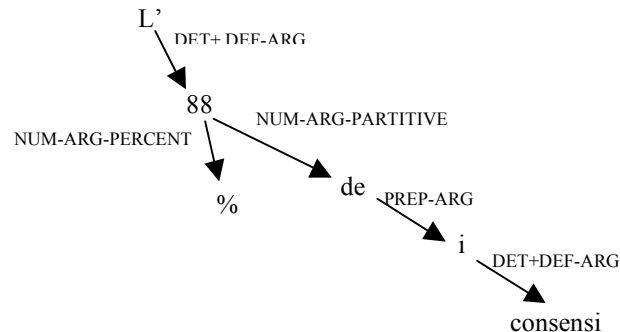%   de
PREP-ARG
i
DET+DEF-ARG
consensi

If the preposition specifies a temporal range, this is simply labelled as NUM-ARG-TIME+DISTRIB.

Example 4.14: "50 % al mese" (50 % each month)

50
NUM-ARG-PERCENT   NUM-ARG-TIME+DISTRIB
%   a
PREP-ARG
il
DET+DEF-ARG
mese
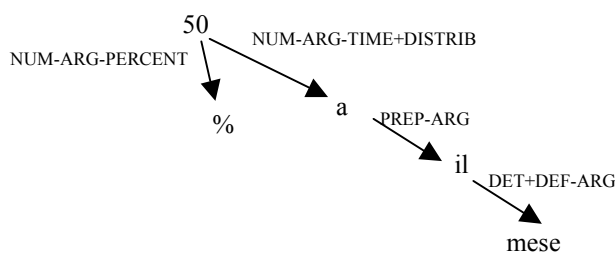
(The relation NUM-ARG-DISTRIB is instead annotated in "800 dollari al kilo" (800 dollars each kg), between "800" and "al".)

But if % is expressed in words ("per cento"), the word "cento" is linked as PREPARG-PERCENT to the numeric value

Example 4.15: "0,4 per cento" (0,4 per cent)

0,4
NUM-ARG-PERCENT
per
PREP-ARG-PERCENT
cento
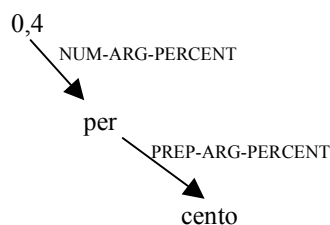
4.4. Distances

Distance is expressed by making the "da (from)" modifier depend on the measure unit, like in the following case:

8 60 (|60| NUM 60) [6;PREPARG-NC]
9 chilometri (CHILOMETRO NOUN COMMON M PL) [8;NBAR-QUANTIF]
10 dalla (DA PREP MONO) [**9;PREPMOD**]
10.1 dalla (IL ART DEF F SING) [10;PREPARG-NC]
11 capitale (CAPITALE NOUN COMMON F SING) [10.1;NBAR-DEF]

(See also the representation of ranges in par. 12.2)

4.5. Names

Names are rather simple structures. They can stand alone (in which case they can require or not the article), or appear together with the class of the named entity (ex. "la pizzeria Partenope" – the pizzeria Partenope). Finally, they can be single-word or multi-word. In all cases, the connectio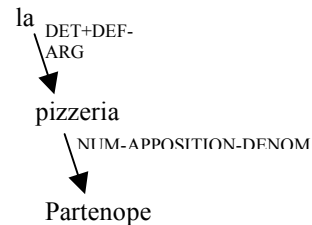ns are standard; the link between the class and the name is labelled as NUM-APPOSITION-DENOM, while for multi-word, the word after the first is linked to the previous on as CONTIN+DENOM (as anticipated in the section on locutions).

Example 4.16: "la pizzeria Partenope"
(the pizzeria Partenope)

la
DET+DEF-ARG

pizzeria
NUM-APPOSITION-DENOM

Partenope

Example 4.17: "la Magneti Marelli" (the Magneti Marelli)

la
DET+DEF-ARG

Magneti
CONTIN+DENOM

Marelli
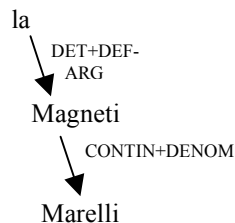
5. Determiners and predeterminers

In this section, beyond standard determiners and predeterminers, we also present the treatment of demonstrative adjectives ("questo" – this – "quello" – that ...), indefinite adjectives ("molti" – many – "alcuni" – some - ...). Also numerals are handled in the way described here, when they are used for counting (see ex. 4.1 and 4.2).

In general, determiners play the role of head of a nominal group, so that a head noun depends on them. According to the type of determiner, the link from the determiner to the noun is labelled DET+INDEF-ARG, DET+DEF-ARG, DET+QUANTIF-ARG. However, in some cases more than one element of this class may appear before the noun, as in "i molti amici" (the many friends), "queste tre ragazze" (these three girls), etc. In these cases, the article or the demonstrative are taken as the head of the whole construct, while the other element is attached as a dependent of the noun.

Example 5.1: Giorgio ha mangiato la torta
(Giorgio has eaten the cake)

```
                              mangiato
           VERB-SUBJ         /    |    \    VERB-OBJ
                         /  AUX+TENSE  \
                    Giorgio     ha        la
              DET+DEF-ARG                    \
                                             torta
```
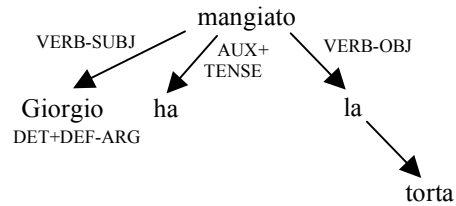
Example 5.2: Giorgio ha mangiato tre torte
(Giorgio has eaten three cakes)

```
                              mangiato
          VERB-SUBJ          /    |    \    VERB-OBJ
                         /  AUX+TENSE  \
                    Giorgio     ha       tre
                                           \  DET+QUANTIF-ARG
                                            torte
```
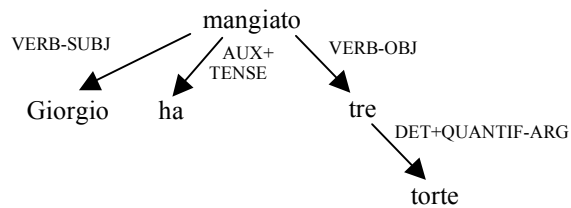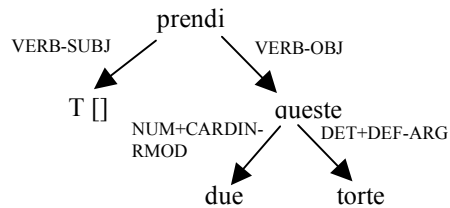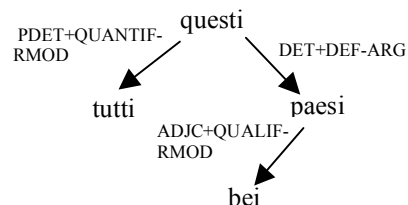
Example 5.3: Prendi queste due mele
(Take these two apples)

```
                          prendi
          VERB-SUBJ      /      \     VERB-OBJ
                     T []        queste
                        NUM+CARDIN-  /  \  DET+DEF-ARG
                        RMOD
                               due      torte
```

Predeterminers are handled as modifiers of the head determiner, linked to it via an arc labelled as PDET+QUANTIF-RMOD (see section 16.3. for a similar structure in English):

Example 8.5: tutti questi bei paesi
(all these nice countries)

```
                            questi
        PDET+QUANTIF-       /    \     DET+DEF-ARG
        RMOD
                        tutti      paesi
                            ADJC+QUALIF-  \
                            RMOD
                                          bei
```
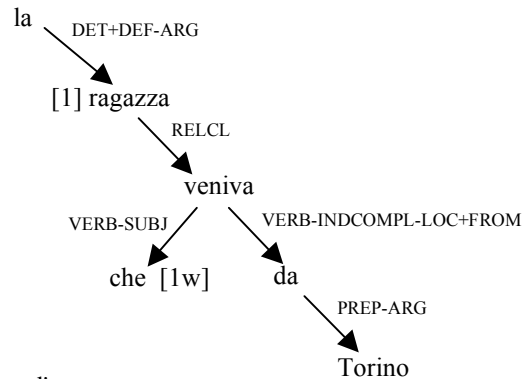
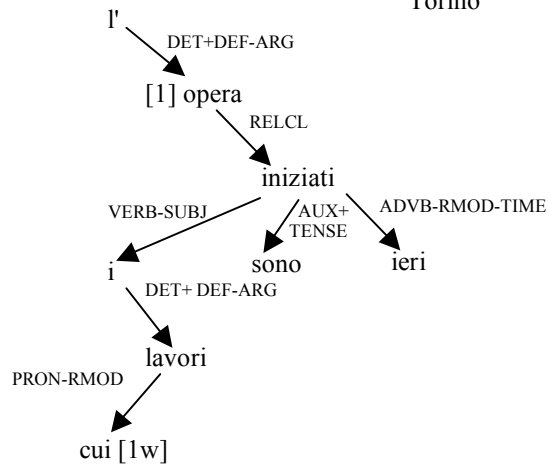## 6. Relative clauses

6.1. Standard relative clauses

There are different approaches to the representation of relative clauses; in some of them, the head of the relative clause is the verb, in others, the head is the relative pronoun. We have chosen the first alternative, so that the relative pronoun depends on the verb as a standard

argument. The link that connects the verb of the relative clause to its parent (the noun governing the clause) is labelled as RELCL.

Example 6.1: "la ragazza che veniva da Torino"
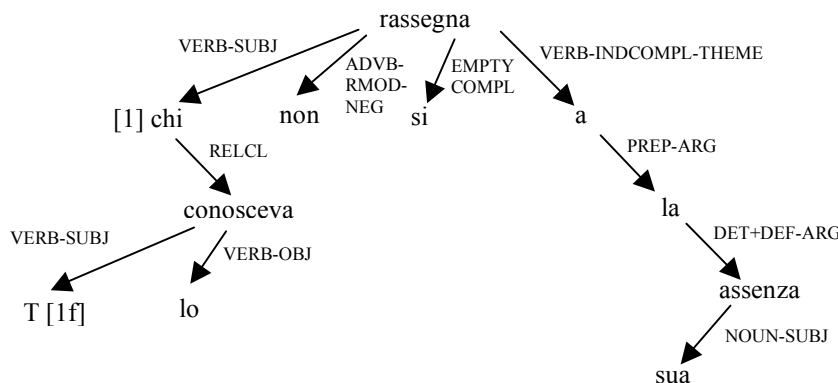*(the girl who came from Torino)*

la → DET+DEF-ARG → [1] ragazza
[1] ragazza → RELCL → veniva
veniva → VERB-SUBJ → che [1w]
veniva → VERB-INDCOMPL-LOC+FROM → da
da → PREP-ARG → Torino

Example 6.2: "l'opera i cui lavori sono iniziati ieri"
(the enterprise whose works have begun yesterday)

l' → DET+DEF-ARG → [1] opera
[1] opera → RELCL → iniziati
iniziati → VERB-SUBJ → i
iniziati → AUX+TENSE → sono
iniziati → ADVB-RMOD-TIME → ieri
i → DET+ DEF-ARG → lavori
lavori → PRON-RMOD → cui [1w]

6.2. Independent relative pronouns

Some relative pronouns lack an antecedent, i.e. they play the combined role of referent and reference to themselves. A simple example is "Chi mi ama mi segua" (Who loves me, follow me). In this case, "chi" is both the subject of "ama" (love) and the subject of "segua" (follow). This use of "chi", in fact, could be paraphrased as "colui che" (the one who). In order to represent these structures, what must be done is to attach the pronoun explicitly to one of its governor, and implicitly (via a trace) to the other. For various reasons, we opted for the attachment of the pronoun to the main clause and for the insertion of the trace in the relative pronoun position.

*Example 6.3: "Chi lo conosceva non si rassegna alla sua assenza"*
(Who knew him, does not resign himself to his absence)

rassegna → VERB-SUBJ → [1] chi
rassegna → ADVB-RMOD-NEG → non
rassegna → EMPTY COMPL → si
rassegna → VERB-INDCOMPL-THEME → a
[1] chi → RELCL → conosceva
conosceva → VERB-SUBJ → T [1f]
conosceva → VERB-OBJ → lo
a → PREP-ARG → la
la → DET+DEF-ARG → assenza
assenza → NOUN-SUBJ → sua

## 6.3. Relative clauses introduced by "dove" (where) and "quando" (when)

Among the various syntactic roles played by some locative and temporal words, there is also the one of relative pronoun. In their simplest use, they behave exactly as the usual relative pronouns ("che" – that, who, which).

Example 6.4: "il luogo dove l'ho incontrato"
        (the place where [I] him have met,
   *i.e. the place where I met him)*
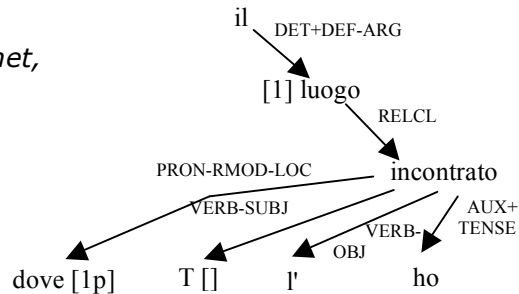
In some cases, however, the pronoun has a double function: it acts both as an argument of a local governor and as an argument of the relative clause. This happens, for instance, in "l'ho visto da dove mi trovavo" (I saw him from where I was). This is an instance of "independent relative pronoun", examined in the previous section, so it is represented in the same way:

Example 6.4:
"l'ho visto da dove mi trovavo"
([I] him have seen from where
myself was, i.e. I saw him
*from where I was)*
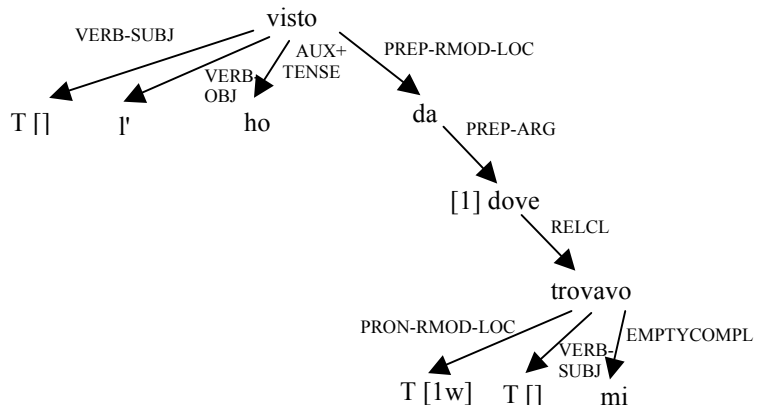
Example 6.5: "dalla finestra ho visto da dove arrivava"
        (from the window, [I] have seen from where [s/he] arrived, i.e.
        I've seen from the window where s/he was arriving from)
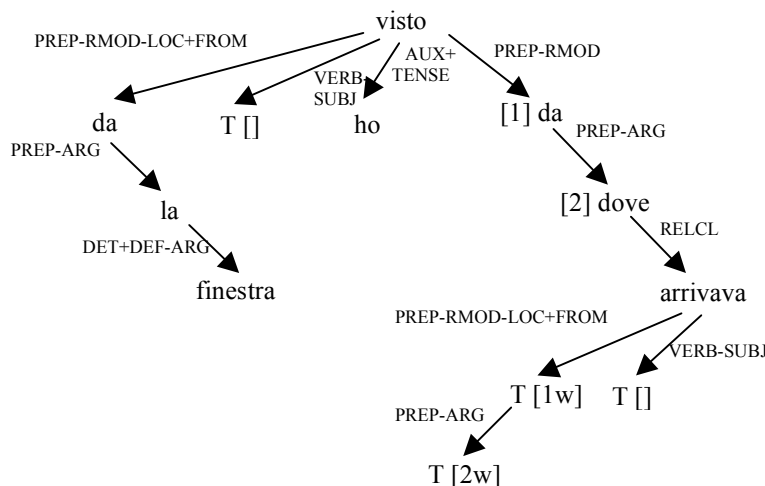
6.4. Reduced relative clauses

A special case of relative clause is the 'reduced relative clause', where there is no relative pronoun, and the verb appears in the participle, without auxiliaries. As in standard cases, we represent these constructs by attaching the verb of the relative clause as a dependent of the governing noun; moreover, we introduce a trace co indexed with the governing noun, and dependent on the verb. The label of the link connecting the trace is SUBJ (in case of present participles, and of past participles of intransitive verbs), otherwise it is VERB-OBJ/VERB-SUBJ.

Example 6.5: "le sinfonie di Beethoven registrate da Toscanini"
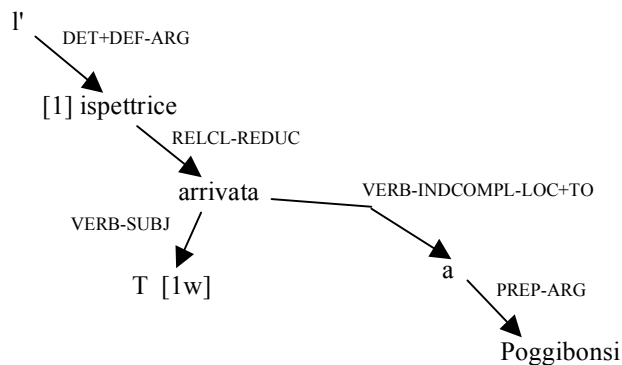 *(the symphonies by Beethoven recorded by Toscanini)*
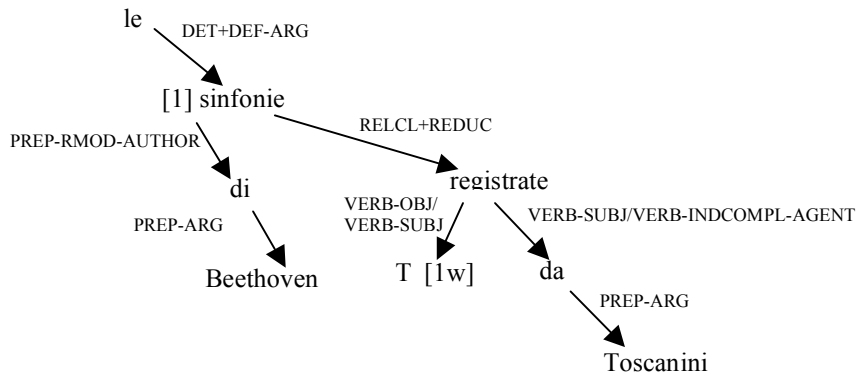
Of course, the agent complement may be missing; in such a case, a second trace is introduced in its place.

Example 6.6: "l'ispettrice inviata
  a Poggibonsi"
  (the inspector sent
  to Poggibonsi)
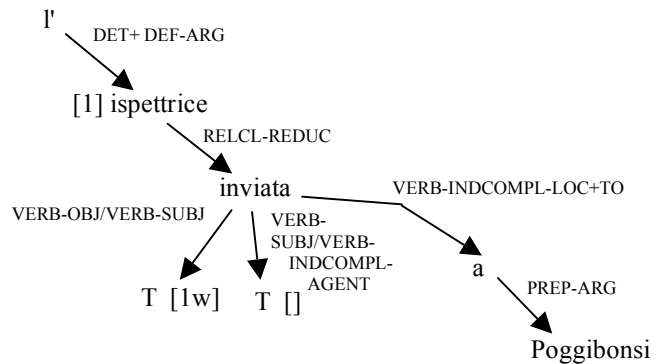
Example 6.7: "l'ispettrice arrivata a Poggibonsi"
  (the inspector arrived to Poggibonsi)

PARTICIPLES AND ADJECTIVES (Renzi, "Grande grammatica italiana di consultazione", vol.II p.330-332)

Some adjectives, as "ammalato" (ill), "appassionato" (passionate), "atteso" (expected), preoccupato (worried), sconosciuto (unknown) are morpho-syntactically similar to participles, and sometimes it may be difficult to understand how must they be classified.

Granted that in our formalism a verbal participle depending on a noun (or a pronoun) has to be annotated as a reduced relative clause, while an adjective depending on a non is linked to it via an ADJC-RMOD arc, we can envision the following criteria to keep apart the two cases:

- the adjectives can be in different degrees, i.e. it is possibile to form the superlative via the suffix "-issimo", or via the adverbs "molto" and "assai" (both meaning 'very') ex. "attesissimo" (very expected), "amatissimo" (very loved), "molto preoccupato" (very worried), *"uccisissimo" (very killed), *"mangiatissimo" (very eaten), *"molto parlato" (very spoken).

- adjectives cannot include a clitic ex "la notizia comunicatagli" (the news communicated to him) *"la notizia sconosciutagli" (the news unknown to him)

- adjectives (as predicates) can be associated only with the verb "essere" (to be), while participles accept the auxiliary "venire" (to come, used for passivization) ex. "Gianni è ascoltato da tutti" vs. "Gianni viene ascoltato da tutti" (Gianni is listened by everybody) "Gianni è preoccupato" vs. *"Gianni viene preoccupato" (Gianni is worried).

Psychological adjectives, as "affascinato" (fascinated), "attratto" (attracted), "commosso" (moved), "disgustato" (disgusted), "impaurito" (frightened), "interessato" (interested), "preoccupato" (worried), "spaventato" (frightened) are a subset of these adjectives. They take a complement expressing the cause of the psychological state (the complement may be governed by various prepositions (as "da", "per", "a", "di") and may be realized by the clitic pronoun "ne". Ex. "Gianni è affascinato dalla matematica" (Gianni is fascinated by mathematics) "Gianni ne è affascinato" (Gianni 'by it' is fascinated)

7. Predicative Complements

The subject-predicate relationship, which usually holds between a subject and a verb phrase (ex. "[Giovanni]-SUBJ [mangia la minestra]-PRED" - Giovanni eats the soup), can hold also between a subject and another type of phrase (ex. "Credevo [Giovanni]-SUBJ [felice]-PRED": I believed Giovanni glad). This type of predication is always expressed via a verb, and can be an argument of the verb or an additional dependent. So, among the various predicative complements, (Renzi vol. II) distinguishes two basic types:

- additional predicative complement (predicative modifier in our terminology), which may be added to many verbs (except copulas);
- argumental predicative complement (predicative complement in our terminology), which must be governed by a copula or by a verb having copulative function.

Moreover, predicative complements can refer either to the subject or to the object of the governing verb. Let's see some examples:

additional predicative complement:
    'Mangia le patate crude' (OBJ): s/he eats the potatoes raw
   'Maria nuota nuda' (SUBJ); Maria swims naked
argumental predicative complement:
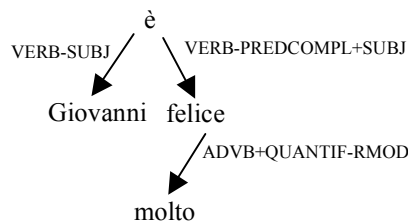   'Credevo Giovanni felice' (OBJ): I believed Giovanni glad
    'Lei sembrava contenta' (SUBJ): she seemed content.

Predicative complements do not present particular problems for the annotation formalisms. They are kept apart argumentality (argumental: PREDCOMPL, accessory: RMODPRED), and to the referent (+SUBJ or +OBJ).
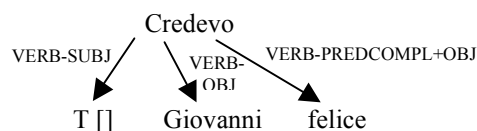
Example 7.1: Maria corre affannata (Mary runs breathless)

```
              corre
    VERB-SUBJ /     \ RMODPRED+SUBJ
             Maria   affannata
```
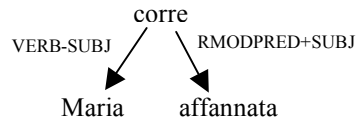
Example 7.2: Giovanni e' molto felice (Giovanni is very glad)

```
              è
    VERB-SUBJ /   \ VERB-PREDCOMPL+SUBJ
         Giovanni  felice
                      \ ADVB+QUANTIF-RMOD
                       molto
```

Example 7.3: Credevo Giovanni felice ([I] believed Giovanni glad)

```
              Credevo
   VERB-SUBJ /  VERB-  \ VERB-PREDCOMPL+OBJ
            /   OBJ     \
          T []  Giovanni  felice
```

8. Auxiliaries and modals

In the treebank, auxiliaries are dependent on the main verb, while modals govern the verb. The idea is that auxiliaries play just the role of feature specifiers (tense, passivization, progressive), so that they do not have an autonomous semantic content, while modals contribute to the meaning evaluation fully.

Example 8.1: Giovanni ha visto Lucia (Giovanni has seen Lucia)

```
              visto
   VERB-SUBJ / AUX+ \ VERB-OBJ
            / TENSE  \
       Giovanni  ha   Lucia
```
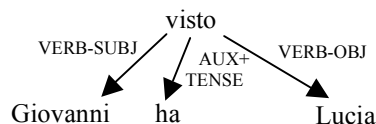
Example 8.2: Giovanni è stato visto da Lucia (Giovanni has been seen by Lucia)

```
                 visto
  VERB-OBJ/VERB-SUBJ / AUX+ \ VERB-SUBJ/VERB-INDCOMPL-AGENT
                    / PASSIVE \
              Giovanni  stato   da
                        AUX+      \ PREPA-RG
                        TENSE      \
                          è        Lucia
```
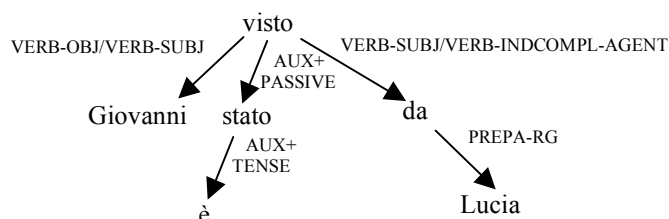
Example 8.3: Giovanni deve vedere Lucia (Giovanni must see Lucia)
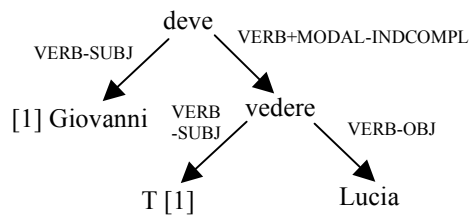

The arc linking the modal with the depending verb is VERB+MODAL-INDCOMPL (this is one of the two complements of modals, the other being the subject). Note the trace which makes the subject of the governed verb coincide with the subject of the governing modal. Some troubles arise in presence of clitics, which can be moved before the modal. In these cases, we assume that the clitic is a VISITOR of the modal, while its proper place (marked by a trace) is below the governed verb.

Example 8.3: Giovanni la deve vedere (Giovanni her must see)
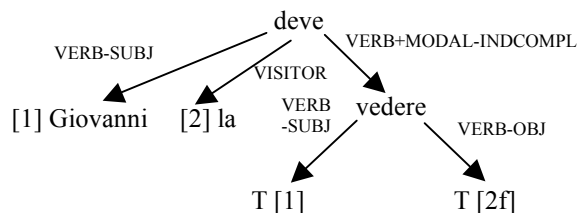
9. Reflexives

In principle, reflexives do not present any special difficulty. However, they are characterized by the clitic pronoun "si", which can play at least three different roles. So, there is a problem of ambiguity, and this section aims at explaining how to solve it.

The three different roles of "si" are the following:
- True reflexives:
    - VERB-OBJ: "Luca si lava" (Luca washes himself), "Marta si vede nello specchio" (Marta sees herself in the mirror)
    - VERB-INDOBJ: "Enzo si è mandato un messaggio" (Enzo sent a message to himself)
    - Reciprocals: "Carlo e Marta si sono baciati ieri" (Carlo and Marta kissed each other yesterday)
- Locutionary reflexives: "Enrico non si sente bene" (Enrico does not feel [himself] good)
- Impersonal subject: "Qui si mangia bene" (Here, [people] eat well)

Of course, ambiguity is often hard to solve: "Non si sente bene" could mean either 'S/he doesn't hear her/himself well', 'People does not hear well', or 'S/he does not feel good', according to the options that the subject has been 'drop', or that it has been 'impersonalized', or that the locutionary interpretation of "si sente" is selected. Usually, verb meaning may help (mangiare - to eat - can hardly be interpreted as a reflexive, but consider "si mangia le unghie" - he eats his nails [to himself]), but the only strong preference concerns locutions: the locutionary interpretation is acceptable (and preferred) just in case it appears in the dictionary. Of course, the choice has to be made by the lexicographer, whose intuitions may be supported by the possibility of paraphrasing "si xxx" with "xxx se stesso" ("se stesso" approximately corresponding to "he himself"); the latter form expresses more explicitly the reflexivity and cannot be applied to locutions.

In the treebank, the reflexives are not marked in any way (apart from the syntactic subcategory of the clitic, which is REFL-IMPERS), since the arc labels are the standard ones (VERB-OBJ, VERB-INDOBJ). Locutionary reflexives get the label EMPTYCOMPL, since the occurrence of "si" is semantically empty. Impersonal subjects are linked to the verb via VERB-SUBJ/VERB-SUBJ-IMPERS.
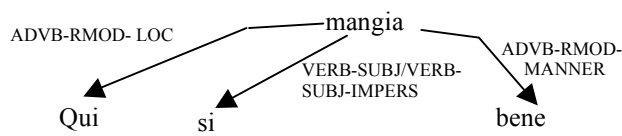
Example 9.1: Luca si lava (Luca washes himself)

```
                      lava
         VERB-SUBJ   ╱    ╲  VERB-OBJ
                    ╱      ╲
              [1] Luca    si [1]
```
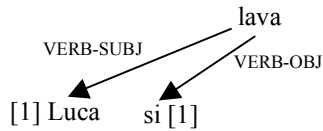
Example 9.2: Enrico non si sente bene (Enrico does not feel good)

```
                           sente
      VERB-SUBJ    ADVB-RMOD-        ADVB-RMOD-
                     NEG               MANNER
                          EMPTYCOMPL
       [1] Enrico       non   si         bene
```
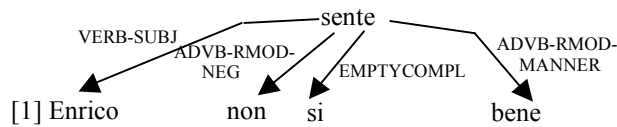
Example 9.3: Qui si mangia bene (Here, you can eat well)

```
   ADVB-RMOD- LOC            mangia
                  VERB-SUBJ/VERB-    ADVB-RMOD-
                   SUBJ-IMPERS         MANNER
            Qui      si                bene
```

Similar considerations apply to the clitics "mi" (1st person, singular), "ti" (2nd, singular), "ci" (1st, plural), "vi" (2nd, plural), apart that, of course, the impersonal reading is not available. The only extra observation concerns the contemporary presence of an impersonal and another reflexive. In these cases, even in case both of them should be "si", the first of them is changed to "ci":

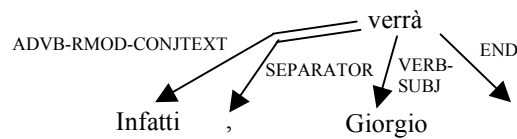*Example 9.4: Qui non ci si lava (Here, people do not wash themselves)*

```
                                   lava
    ADVB-RMOD-LOC        ADVB-              VERB-SUBJ/VERB-SUBJ-IMPERS
                         RMOD-      VERB-
                         NEG        OBJ
       Qui      non      ci        si
```
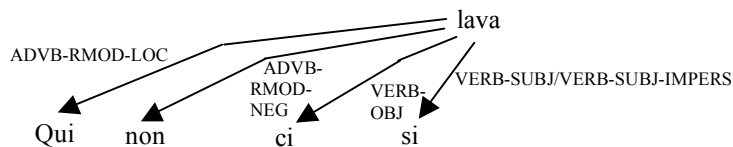
10. Adverbs

Adverbs (as their name suggests) are verbal dependents. Apart form standard adverbs (as "spesso" - often - "naturalmente" - obviously - etc.), there are also adverbs whose role is similar to the one of conjunctions, since they act as a kind of connection to the context. This is the case of adverbs as "pero'" (but, however), "infatti" (in fact), "anche" (also), etc. We decide to represent them as other adverbs, i.e. to insert them as dependents of the verb, and as adjuncts (so that the label is ADVB-RMOD). However, their morppho-syntactic role of connectors is marked by the first part of the label, for which we adopted the name ADVERS. For instance, in the next example, ADVERS makes explicit the adversative function of "però".

Example 10.1: "Giorgio pero' vorrebbe venire"
*(Giorgio, however, would like to come)*

vorrebbe

VERB-SUBJ   ADVB-RMOD-ADVERS   VERB+MODAL-INDCOMPL

[1] Giorgio   però   venire

VERB-SUBJ

T [1]
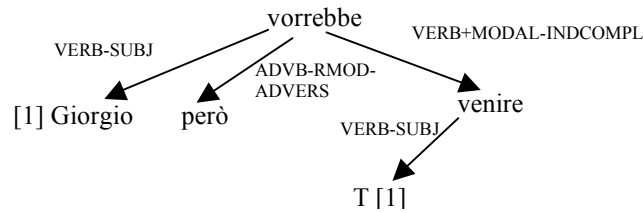
Example 10.2: "Infatti, Giorgio verrà."
(In fact, Giorgio will come.)

verrà

ADVB-RMOD-CONJTEXT   SEPARATOR   VERB-SUBJ   END
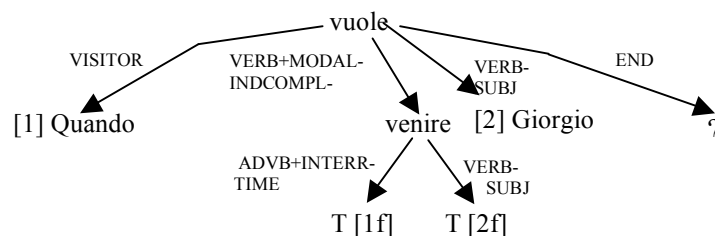
Infatti   ,   Giorgio

Interrogatives are treated exactly in the same way. Notice that, except in case of extraction from a subordinate, no trace is involved, since we exploit the rather free word order of Italian.

Example 10.3: "Quando verra' Giorgio?"
(When come$_{+Fut}$ Giorgio?, i.e. when will Giorgio come?)

verrà

ADVB+INTERR-RMOD-TIME   VERB-SUBJ   END
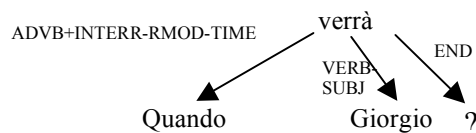
Quando   Giorgio   ?
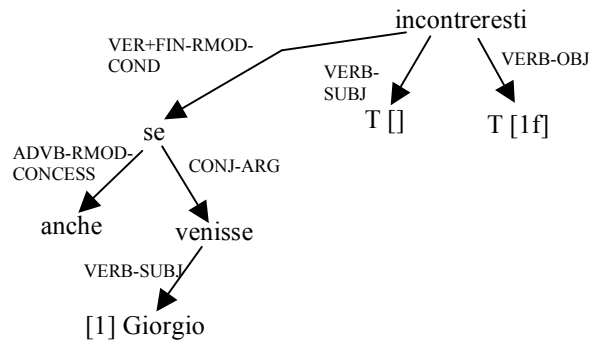
But:
Example 10.4: "Quando vuole venire Giorgio?"
(When wants to come Giorgio?, i.e. when wants Giorgio come?)

vuole

VISITOR   VERB+MODAL-INDCOMPL-   VERB-SUBJ   END

[1] Quando   venire   [2] Giorgio   ?

ADVB+INTERR-TIME   VERB-SUBJ
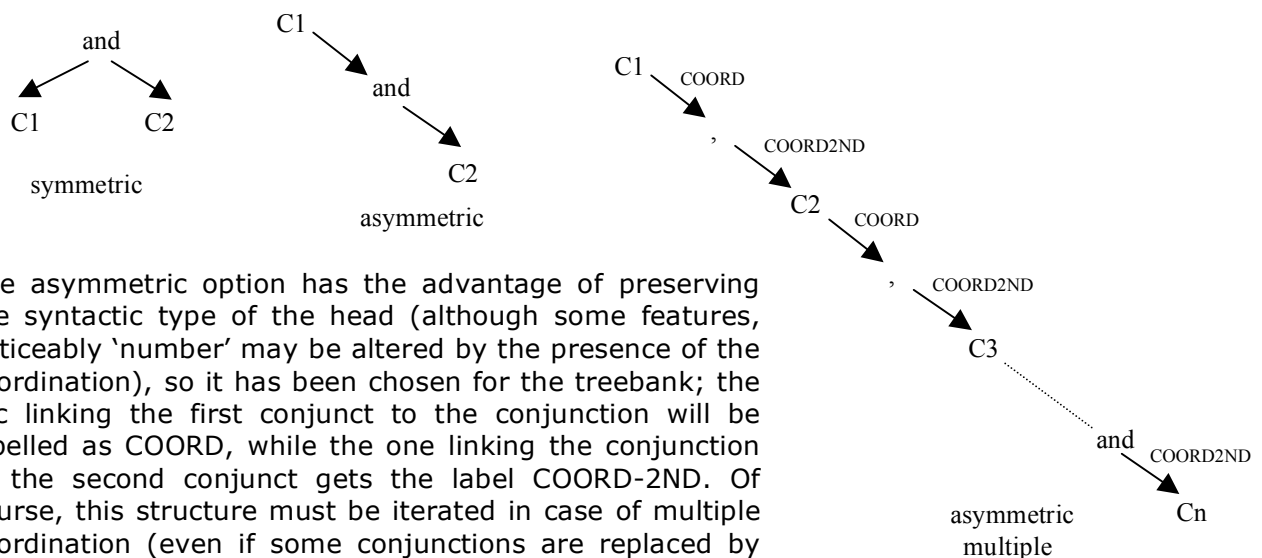
T [1f]   T [2f]

In some cases, however, the adverb is not linked to the main verb, but to a following conjunction:

Esempio 10.5: "Anche se Giorgio venisse, non lo incontreresti"
(even if Giorgio came, not him meet$_{2nd,sing,condiz}$, i.e.
even if Giorgio comes, you would not meet him)

incontreresti

VER+FIN-RMOD-COND   VERB-SUBJ   VERB-OBJ

se   T []   T [1f]

ADVB-RMOD-CONCESS   CONJ-ARG

anche   venisse

VERB-SUBJ

[1] Giorgio

11. Coordination

Coordination is one of the most complex aspects of natural language. In the analysis of coordinate structures there are two alternatives: symmetric representation (with the conjunction as head) or asymmetric representation (where the first conjunct is taken as head):

and

C1   C2

symmetric

C1

and

C2

asymmetric

C1   COORD

,   COORD2ND

C2   COORD

,   COORD2ND

C3

and   COORD2ND

Cn

asymmetric multiple

The asymmetric option has the advantage of preserving the syntactic type of the head (although some features, noticeably 'number' may be altered by the presence of the coordination), so it has been chosen for the treebank; the arc linking the first conjunct to the conjunction will be labelled as COORD, while the one linking the conjunction to the second conjunct gets the label COORD-2ND. Of course, this structure must be iterated in case of multiple coordination (even if some conjunctions are replaced by commas; in particular, the arc labels are alternations of COORD and COORD2ND.

Example 11.1: "Lucia ha incontrato Giorgio e Giovanni"
(Lucia has met Giorgio and Giovanni)

incontrato

VERB-SUBJ   AUX+TENSE   VERB-OBJ

Lucia   ha   Giorgio

COORD

e

COORD2ND
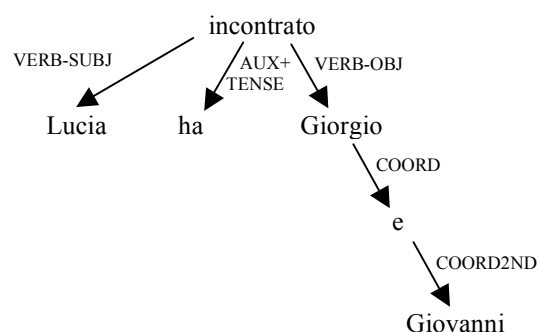
Giovanni

Example 11.2: "Lucia vuole salutare Giorgio e baciare Giovanni"
(Lucia wants to greet Giorgio and to kiss Giovanni)
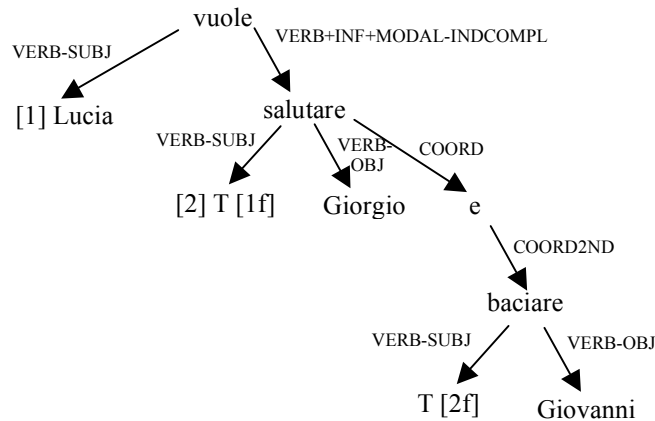
12. Balanced structures

In this section, we present the representation adopted for linguistic structures characterized by the presence of pairs of elements correlated or coordinated. The difference with respect to standard coordination is that both conjuncts are introduced by a particle, and not only the second one: "cosí ... come ..." (as ... as ...), "tanto ... quanto ..." (so much ... as ...), "piú ... che ..." (more ... than ...), "tra ... e ..." (between ... and ...), "da ... a ..." (from ... to ...).

12.1. Comparatives

In comparative structure, a comparison is expressed; the comparison may concern a property of two individuals or two properties of one or two individuals. It is possible that the two compared elements are expressed via different syntactic structures; but the most common case is the one where and adjectival phrase is involved. In case of adjectives, the comparison may be expressed via a morphological change (ex. "maggiore" - greater), and this is one of the three forms that an adjective may assume: positive, comparative, and superlative.

> COMPARATIVE AND SUPERLATIVE ADJECTIVES (Serianni – "Grammatica italiana" - II ed., 1991 p.209-219)
>
> Three different degrees are encoded for the qualificative adjective: positive, comparative, superlative.
>
> 1. For the comparative degree, there are three different forms of comparatives, i.e. equality, majority, minority (the last two indicate inequality):
> o Equality: the adjective may be preceded by correlative terms, as "tanto" (as much), "altrettanto" (as much ... as), "così" (so): "Io sono così indignato come te" (I am so irritated as you [are]; "E' un'auto tanto veloce quanto silenziosa" (it is a car as fast as noiseless . The second term is introduced by "quanto" (as) or by "come" (as): "Lui è capace quanto te" (He is skilled as you [are]), "ho un'auto veloce come la tua" ([I] have a car as fast as yours).
> o Inequality: the adjective is introduced by "più" (more: majority) or "meno" (less: minority); the second term is introduced by "di" (than) or "che" (than): "Mario è più gentile con me che con te" (Mario is more kind

with me than with you), "Mario è più furbo che intelligente" (Mario is more cunning than intelligent), "L'ha detto più per scherzare che per offenderti" (He told it more for joking than for offending you), "più che mai" (more than ever).

2. For the superlative degree, the strength of a quality is expressed at its maximum degree, either in relative or in absolute terms:

o Relative: the adjective is modified by "più" (more) or "meno" (less), preceded by the definite article; it plays here a role similar to the one of a demonstrative: "Mario è il più bravo" (Mario is the most clever). The second term is introduced when it is needed) by "di" (of), "tra", "fra" (between, among): "il più simpatico di/tra voi" (the most likeable of/among you); when missing, the second term is understood to refer to all individuals for which a comparison may be made: "il più simpatico" = "il più simpatico di tutti" (the most likeable = the most likeable among everybody)

o Absolute: it expresses the maximum degree, without involving comparisons (although it may be claimed that there is an implicit comparison with an average value). The adjective is morphologically modified by adding the suffix "-issimo": "Mario è bellissimo e altissimo" (Mario is very nice and very tall). Actually, the adverb "molto" (very) can also be used, but the strength is somewhat reduced: "Mario è molto alto" seems to express a weaker statement than "Mario è altissimo".

In the treebank, a comparative based on adjectives is represented by taking the adjective as the head of the construct, and making both the antecedent (adverb or conjunction) of the comparative conjunction and the second term as dependent on the adjective.

Example 12.1: "Maria é piú bella di
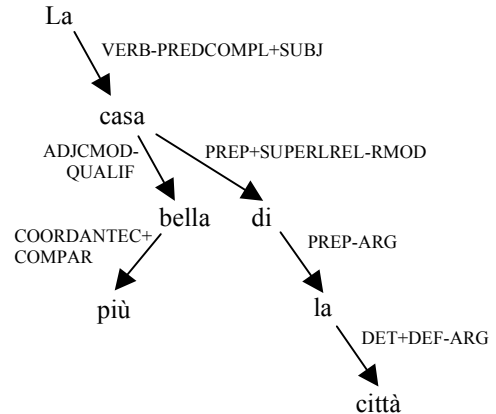*Michela"*
(Maria is more nice
than Michela)
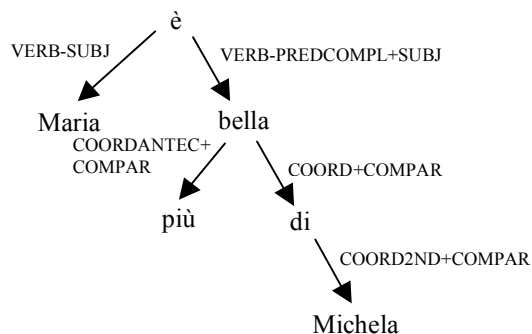
Example 12.2: "Maria é piú studiosa
che intelligente"
(Maria is more studious
than intelligent)
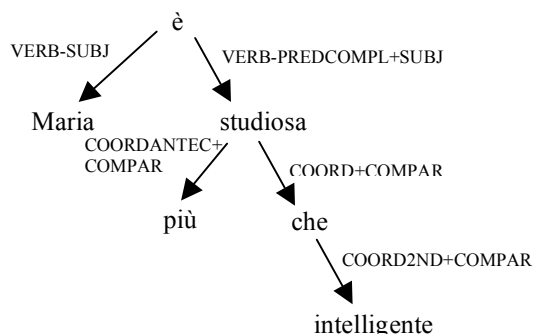
Relative superlative is represented by attaching the adverbial modifier ("più" – more -, "meno" – less -) to the adjective; the second term (if it is present) is attached to the noun which the adjective refers to.
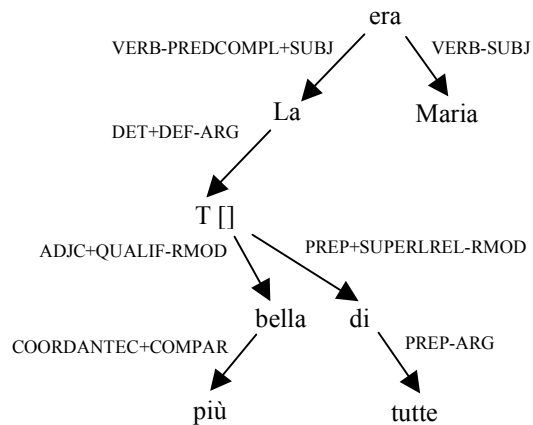
Example 12.3: "La casa piú bella della cittá"
(The house most fine of the city,
i.e. the finest house of the city)

La
VERB-PREDCOMPL+SUBJ
casa
ADJCMOD-QUALIF
PREP+SUPERLREL-RMOD
bella   di
COORDANTEC+COMPAR   PREP-ARG
più   la
DET+DEF-ARG
città

In case the prepositional modifier is absent (*la casa più bella*) the label SUPERLREL is on the relation linking the adjective with the noun (ADJC+QUALIF+SUPERLREL-RMOD) and the relation linking "più" is ADVB+COMPAR-RMOD (since the adverb is not an antecedent of something).

In case the noun is absent, we introduce a (non-coindexed) trace standing for the noun implicitly referred to (in the following example 'girl', or 'woman').
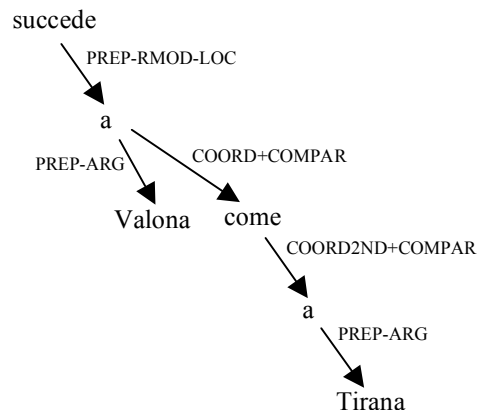
Example 12.4: "La piú bella di tutte era Maria"
(the nicest of all was Maria)

era
VERB-PREDCOMPL+SUBJ   VERB-SUBJ
La   Maria
DET+DEF-ARG
T []
ADJC+QUALIF-RMOD   PREP+SUPERLREL-RMOD
bella   di
COORDANTEC+COMPAR   PREP-ARG
più   tutte

Beyond adjectives, the comparison may involve also nouns; in such a case, the comparison implicitly refers to the amount of something (in case of mass nouns) or to the count of some individuals (in case of countables).

In the example that follows, the comparison concerns a prepositional phrase. Note that the first term ("a Valona") could be introduced by "così" (so), and that the comparison could be expressed as "tanto ... quanto ..." (so ... as ..).

Esempio 12.6: "... succede a Valona
come a Tirana"
(... [it] happens at Valona
as at Tirana)

succede
PREP-RMOD-LOC
a
PREP-ARG   COORD+COMPAR
Valona   come
COORD2ND+COMPAR
a
PREP-ARG
Tirana

> COMPARATIVES (Renzi, "Grande grammatical italiana di consultazione"-vol.II, pp.832-853)
>
> There are adjectives which are intrinsically comparative, as "migliore" (better) and "peggiore" (worse), which correspond, from a semantic point of view, to "più" (more) and "meno" (less) followed by an adjective ("migliore" → "più buono" - better → more good; "peggiore" → "meno buono" - worse → less good); they may be followed by a comparative structure introduced by "di" (than), or "a", in the case of "inferiore" (inferior) and "superiore" (superior); they may also be used as superlatives ("il migliore" - the best). In place of "migliore" and "peggiore" it is also possible to use the corresponding adverbial forms "meglio" and "peggio".
>
> When the first comparison term is a VP, it may be associated with the adverbials "più" (more) or "meno" (less), possibly in the forms "di più" (more) and "di meno" (less): "Gianni lavora più di prima, Giorgio lavora di meno" (Gianni works more than before, Giorgio works less). Among these 4 adverbial forms, only "più" cannot appear alone: *"Gianni lavora più".
>
> Only "più" and "meno" can appear as modifiers of a NP: "Gianni ha letto più libri di te" (Gianni read more books than you [did]), *"Gianni ha letto di più libri di te".

Example 12.7:"rende di piú la canapa indiana del frumento"
         (returns more the cannabis than+the wheat, i.e.
         The cannabis returns more than the wheat)
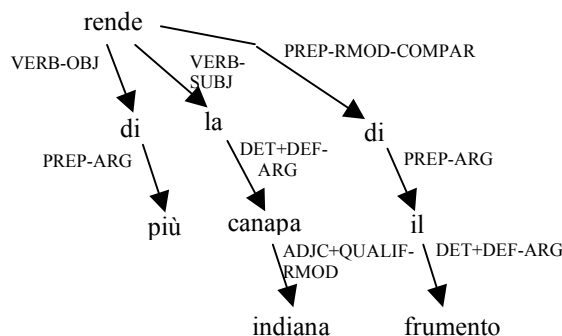

```
12.2. The construct "da ... a ..." (from ... to ...)
```

This construct expresses a range or a path. In the treebank, it has been represented by letting the first preposition to act as the head of the whole construct; it has as dependents the 'initial' element (i.e. the one following "da" (from), and the second preposition (linked to the first via PREPARG-RANGE).

Example 12.8: "Da Roma a Milano ci vogliono sei ore"
         (From Roma to Milano it takes six hours)
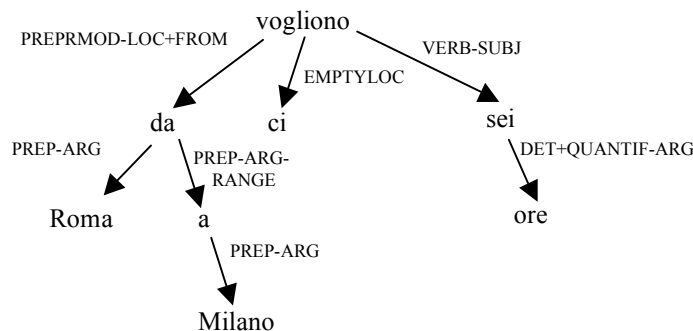
The construct "tra term1 e term2" (between term1 and term2) is not represented as a balanced structure because often the relation holds between two terms of the same class and the second term is merged with the first: "la casa si trova tra due alberi" (the house lies between two trees) instead of "la casa si trova tra un albero e un albero" (the house lies between a tree and a tree).

THE PREPOSITION "TRA" and "FRA" (Serianni "Grammatica italiana" – II ed., 1991 p.351)

The preposition "tra" (between, among; "fra" is an exact synonymous) expresses the intermediate position between two or more reference points. If both the terms of the relation (spatial, temporal, etc.) are made explicit, the preposition is used once, and the two terms are connected via "e" (and): "il ponte tra Venezia e Mestre" (the bridge between Venezia and Mestre), "verrò tra le otto e le dieci" (I will come between eight o'clock and ten o'clock).
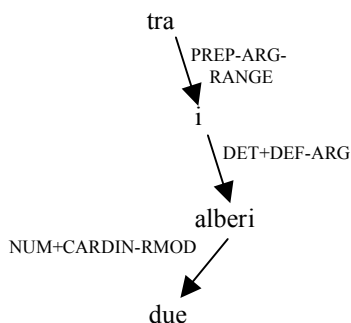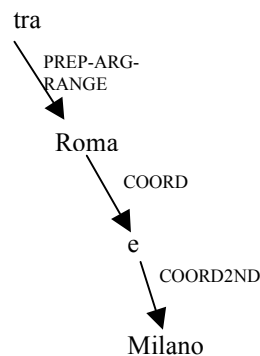
However, if the relation holds between elements of the same class, the second term is merged with the first, to avoid redundancy: "la casa si trova tra due alberi" (the house lies between two trees) instead of "la casa si trova tra un albero e un albero" (the house lies between a tree and a tree).

An analogous simplification occurs when the first term is determined on a deictic basis: "verrò fra tre giorni" (I will come in three days), "trovi il bar tra due chilometri" ([you] find the bar after two chilometers). Here, the range appears in place of the limiting points, and the interpretation refers to the last point, instead of to the interval itself.

"Tra" is used to express various complements:
- locative: state ("si trova tra due ponti" - it lies between two bridges); distance ("fra due chilometri" - after two km)
- temporal: time interval (transition between past and present)
- partitive: "il clima è tra i peggiori" - the weather is amongst the worse ones; "alcuni di noi" - some of us
- company and reciprocals: "una serata tra amici" (an evening among friends); "accordo tra banca e cliente" (agreement between bank and customer)

Example 12.9: "... tra Roma e Milano ..."
(between Roma and Milano)

tra
PREP-ARG-RANGE
Roma
COORD
e
COORD2ND
Milano

tra
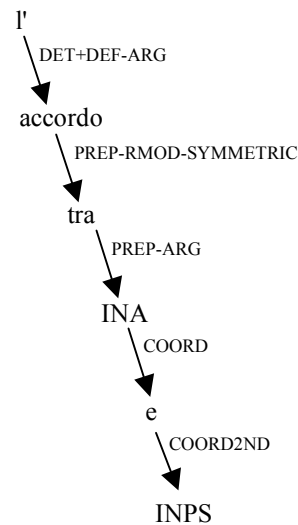PREP-ARG-RANGE
i
DET+DEF-ARG
alberi
NUM+CARDIN-RMOD
due

Example 12.10: "... tra i due alberi ..."
(between the two trees)

> "TRA" and NOMINALIZATION (from Renzi "Grande grammatica italiana di consultazione", vol.I, p.351)
>
> "Tra" is one of the prepositions that introduce NP's which are complements of nouns obtained from verbs via nominalization; it is used when the verb expresses a reciprocal relation between two complements: "Giovanni e Mario lottarono" (Giovanni and Mario fought) or "Giovanni lottò contro Mario" (Giovanni fought against Mario) → "la lotta tra Giovanni e Mario" (the fight between Giovanni and Mario)

In this case, the representation has been chosen according to the subjectness of the arguments (which is kept after the nominalization).
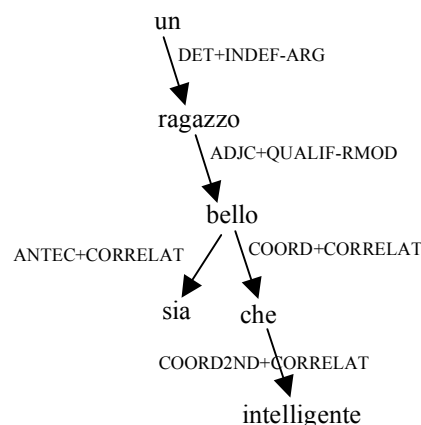
Example 12.11: "l'accordo tra INA e INPS"
*(the agreement between INA and INPS)*

```
l'
 | DET+DEF-ARG
 ↓
accordo
  \ PREP-RMOD-SYMMETRIC
   ↓
  tra
    \ PREP-ARG
     ↓
    INA
      \ COORD
       ↓
       e
        \ COORD2ND
         ↓
        INPS
```

12.3. Correlative coordination: 'sia … sia …' (both … and), 'nè … nè …' (neither … nor)

These construct are similar to the previous ones, but the first conjunction has just a syntactic 'balancing' role, without a relevant semantic import. So, differently from the representation above, we assume that the head of the construct is the item following the first conjunction (as in examples 12.2 and 12.5). Notice that in "sia ... sia ...", the second occurrence of the the conjunction "sia" is often replaced by "che".

Example 12.12: " ... un ragazzo sia bello che intelligente ..."
(... a boy both nice and intelligent)

```
un
 \ DET+INDEF-ARG
  ↓
ragazzo
   \ ADJC+QUALIF-RMOD
    ↓
   bello
ANTEC+CORRELAT /   \ COORD+CORRELAT
     ↓           ↓
    sia         che
              COORD2ND+CORRELAT
                    ↓
               intelligente
```

Example 12.13: "Giorgio non mangia ne' carne ne' pesce"
(Giorgio [not] eats neither meat nor fish, e.g
*Giorgio does not eat either meat or fish)*
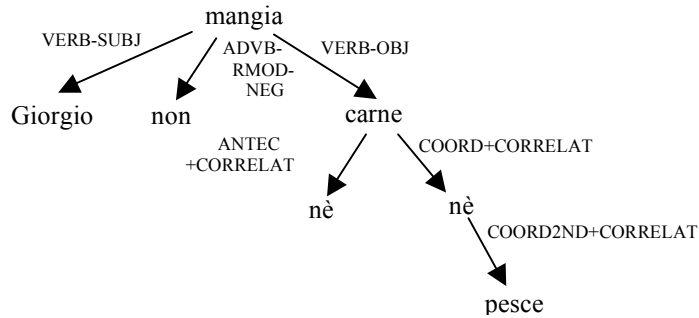
13. Clefts

A cleft construction is a complex sentence construction consisting of two clauses: a) a main clause containing a copula whose non-subject complement is the focus phrase, and b) a subordinate clause one of whose arguments is coindexed with that focus phrase. Together the two clauses (a and b) express a simple proposition, which can normally also be expressed in the form of a single clause ("[Fu con queste parole che] [lo studente guidò la manifestazione]" ([(It) was with these words that] [the student leaded the protest]) = "Lo studente guidò la manifestazione con queste parole" (The student leaded the protest with these words)). Moreover in implicit cleft construction the subordinate is an infinitive clause introduced by a preposition; in explicit the subordinate is a finite clause introduced by a conjunction.

Cleft constructions are represented considering the focus phrase as the predicative complement of the copula (of the main clause), and the other clause as the subject of the main clause.
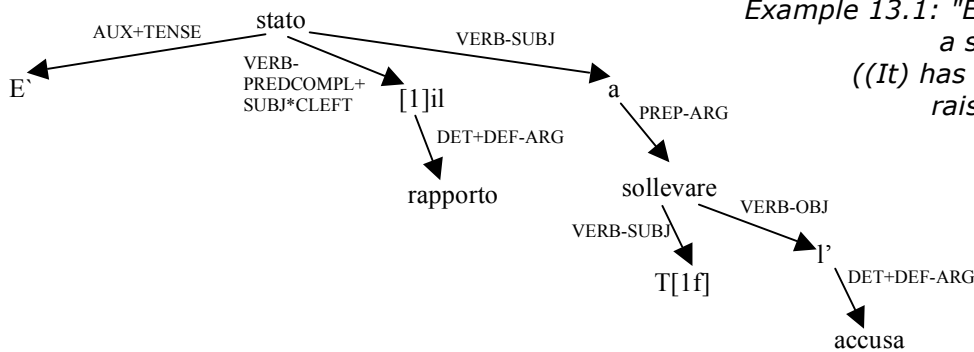
Example 13.1: "E` stato il rapporto
a sollevare l'accusa."
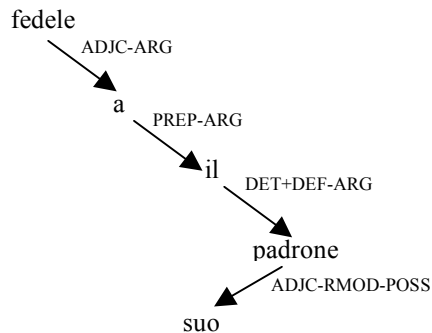*((It) has been the report to*
raise the accusation.)

Example 13.2: "Fu con queste parole che
lo studente guidò la manifestazione."
((It) was with these words that the
student leaded the protest.)
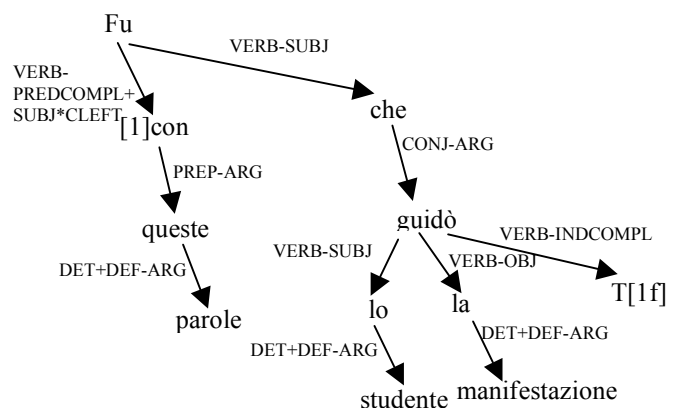
14. Adjectives

Some adjective is usually followed by a prepositional phrase ("fedele a …" (faithful to …), "tale che …" (such as …), …). In these cases the prepositional phrase is represented as an argument of the adjective.

Example 14.1: "Fedele al suo padrone."
(Faithful to its owner.)

```
fedele
     \ ADJC-ARG
      a
        \ PREP-ARG
         il
           \ DET+DEF-ARG
            padrone
                   \ ADJC-RMOD-POSS
                    suo
```

15. Causative constructions

The causative or factitive construction is characterized in Italian by a particular syntactic behavior.

CAUSATIVE CONSTRUCTION (Renzi, Grande grammatica italiana di consultazione, vol. 2)

A causative construction (Caus + Inf) involves a special Verb (Caus) and an infinite clause (Inf) depending on it.
Several arguments can be presented in order to show that these Verbs form a sort of compound with the depending infinitive verb
- First, only the whole compound (Caus + Inf) can be put in the negative form, but not Inf alone. For instance, the negation of "Piero dorme" (Piero sleeps) is "Piero non dorme" (Piero not sleeps), but the negation of "Paolo fa dormire Piero" (Paolo get (to) sleep Piero) is "Paolo non fa dormire Piero" (Paolo not get (to) sleep Piero), and not *"Paolo fa non dormire Piero" (Paolo get (to) not sleep Piero).
- Second, usually no lexical item can be inserted between Caus and Inf. But there are exceptions: "lo faccio subito uscire" (I get it immediately get out).
- Third, Inf does not form any constituent with its complements in Caus + Inf. An indirect evidence of that can be found observing the behavior of Caus + Inf in cleft clauses, where only constituents can be extracted. It's impossible to extract, e.g., Inf with its object: "Gli faccio mangiare la minestra" --> "'E mangiare la minestra che gli faccio" ((I) him get (to) eat the soup --> (It) is (to) eat the soup that (I) him get). But it is, instead, possible to extract to whole Caus + Inf with its object, because they form a constituent: "Preferisco fargli mangiare la minestra" --> "'E fargli mangiare la minestra che preferisco" ((I) prefer get him (to) eat the soup --> (It) is (to) get him (to) eat the soup that (I) prefer).
- Fourth, the transformation of the grammatical relations subcategorized by Inf is the most typical feature of Caus + Inf (Renzi,88). The change happens according to the following principles:
1) when Inf is intransitive, the subject of Inf becomes the object of (Caus + Inf)
2) when Inf is transitive or ditransitive, the subject of Inf becomes the agent complement or the indirect object of (Caus + Inf)

> 3) when Inf is ditransitive, the subject of Inf becomes the agent complement of (Caus + Inf). Examples: with Inf intransitive the subject of Inf becomes the object of (Caus + Inf): "Piero dorme" --> "Paolo fa dormire Piero" Piero sleeps --> Paolo get (to) sleep Piero) with Inf transitive the subject of Inf becomes the indirect object or agent complement of (Caus + Inf): "Piero mangia la torta" --> "Paolo fa mangiare a/da Piero la torta" (Piero eats the cake --> Paolo get (to) eat to/by Piero the cake) whit Inf ditransitive the subject of Inf becomes the agent complement of (Caus + Inf): "Piero da a Maria il libro" --> "Paolo fa dare a Maria il libro da Piero"(Piero gives to Maria the book --> Paolo get (to) give the book to Maria by Piero)
>
> In conclusion, the (Caus + Inf) subcategorizes the subject of Caus and all the grammatical relations subcategorized by Inf (except the subject of Inf). The subject of Inf is transformed in another grammatical relation according to the subcategorization features of Inf and is annotated using a transformation.

In order to take into account all the above mentioned features of (Caus + Inf), we can propose the following representation.

Example 15.1: "Giorgio fece dare a Maria un libro da Silvia."
(Giorgio get to give to Mary a book by Silvia, i.e.
Giorgio get Silvia to give Mary a book.)
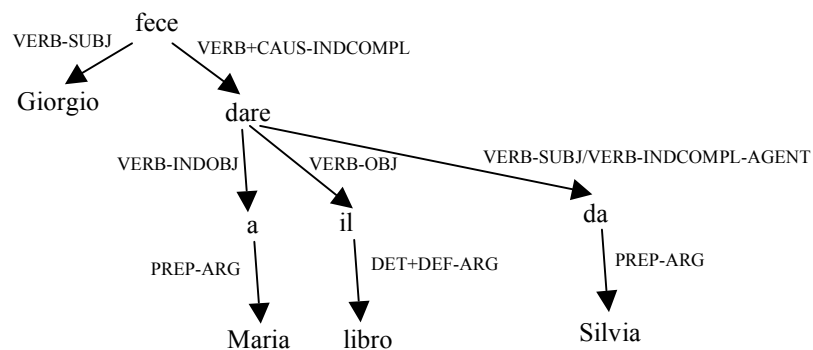
In this representation the relation between the causative Verb and the infinitive Verb is VERB+CAUS-INDCOMPL. The other relations are: the SUBJ of the causative Verb, the transformed SUBJ of the infinite Verb (VERB-SUBJ/VERB-INDCOMPL-AGENT or VERB-SUBJ/VERB-OBJ) and the other relations subcategorized by the infinite Verb. (A limit of this representation is that it does not account for the semantic relation between the subject of the causative and the subject of the infinitive verb.)

Example 15.2: "Alle parole egli fece seguire
i fatti."
(To the words he get
to follow the facts.)
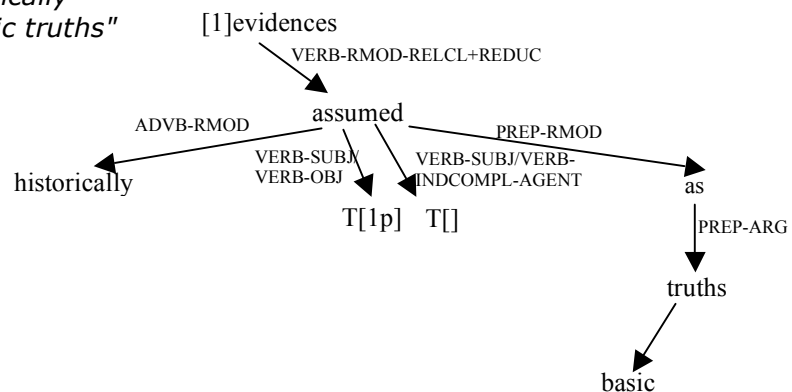

39
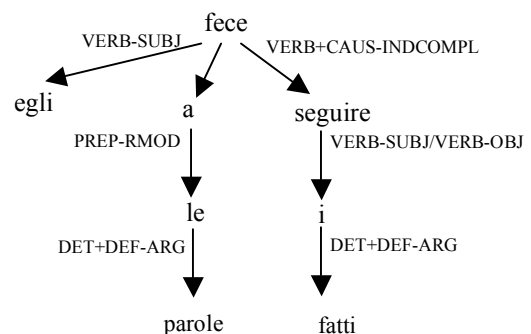
16. Applying TUT on English

A small set (200) of English sentences has been included in the TUT treebank just as a support for non-Italian speakers to the comprehension of the annotation scheme. We report below therefore some observations regarding peculiar features (and extensions) of the TUT annotation scheme for representing English structures and phenomena.

16.1. Participles

The participles used in adjectival position (i.e. "the required openness") have no traces, and are linked to the noun via ADJC+QUALIF-RMOD.
On the contrary, participles in reduced relative clause position (i.e. "evidences, historically assumed as basic truths") have a full representation as relative clauses (in the example above, two traces, one for the passivized subject (evidences), and one for the unknown agent complement, and a link upward labelled as VERB-RMOD+RELCL+REDUC).

Example 15.1: "Evidences historically
assumed as basic truths"

[1]evidences

VERB-RMOD-RELCL+REDUC

ADVB-RMOD   assumed   PREP-RMOD

historically   VERB-SUBJ/   VERB-SUBJ/VERB-   as
                VERB-OBJ   INDCOMPL-AGENT

                T[1p]   T[]   PREP-ARG

                                         truths

                                    basic

16.2. Present participle, gerund and noun

Present-participle versus gerund versus noun forms, are tagged according to their role, but the nominal interpretation is assigned just in case it appears explicitly in the lexicon. So, for "thinking":
 - intercultural thinking is the base of ... ---> NOUN
  - by thinking to the problem, you can solve it ---> GERUND
  - the thinking girl was in the garden ---> PARTICIPLE

Of course, the participle interpretation is the one used for progressives.

16.3. Such

"Such" in a pre-article position is tagged as a predeterminer and is linked to the article via the PDET-RMOD relation.

Example 15.1: "Such a process"

a

PDET-RMOD   DET+INDEF-ARG

such   process
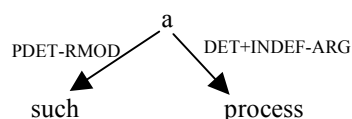
16.4. Hyphenated adjectives and nouns

Hyphenated adjectival pairs (e.g. "scientific-technological rationality") are represented by linking the hyphen to the first element of the pair (here "scientific") via a SEPARATOR link and the second element (here "technological") to the hyphen via COORD2ND+BASE N.B. If the hyphenated pair appears in the dictionary, it is considered as a single entry (e.g. "socio-economic").
On the contrary, hyphenated noun pairs are analyzed by assigning the first noun the role of NOUN-RMOD of the second noun (e.g. "world-culture").

16.5. Ordinal suffixes

The "ordinal" suffixes (as -nd in 2nd, -th in 18th) are tagged as adjectives of type "ORDINSUFF" and linked to the number by using the relation CONTIN+ORDSUFF. So "the 18th century":

    17 the (THE ART DEF ALLVAL ALLVAL) [16;PREP-ARG]
    18 18 (|18| NUM 18) [17;DET+DEF-ARG]
    19 th (TH ADJ ORDINSUFF ALLVAL ALLVAL) [18;CONTIN+ORDSUFF]
    20 century (CENTURY NOUN COMMON N SING) [18;DET+QUANTIF-ARG]

16.6. 'S genitive

Genitives ar represented by tagging "'s" as ART (with type GENITIVE and lemma XGENITIVE), and taking it as the head of the construction. So: "a radio receiver's ability":

    9 a (A ART INDEF ALLVAL SING) [12;DET+GENITIVE-RESTR]
    10 radio (RADIO NOUN COMMON N SING) [11;NOUN-RMOD]
    11 receiver (RECEIVER NOUN COMMON N SING) [9;DET+INDEF-ARG]
    12 's (XGENITIVE ART GENITIVE) [20;VERB-SUBJ]
    13 ability (ABILITY NOUN COMMON N SING) [12;DET+GENITIVE-ARG]

16.7. Compound verbs

Compound verbs (as "stand up") are represented in the standard way, but the adverbial-prepositional component is linked to it via a PARTICLE link.