THE SYNTACTIC CATEGORIES
(Parts Of Speech)

Table of contents

---

0. Preface

This document describes the syntactic categories, syntactic subcategories, and syntactic features appearing in the treebank. The syntactic categories conform to the standard originated from the ILEX (Italian LEXicon) project, carried out in cooperation with IRST-ITC, The University of Venezia, and the University of Piemonte Orientale. Subcategories and features are non-standard.

We also include here a few lines on locutions; more details on them may be found in the document on syntactic structures (Linguistic Notes).

1. List of defined categories

1. ADJ (adjectives)
2. ADV (adverbs)
3. ART (articles)
4. CONJ (conjunctions)
5. DATE (dates)
6. INTERJ (interjections)
7. MARKER (markers)
8. NOUN (nouns)
9. NUM (numbers)
10. PHRAS (phrasal)
11. PREDET (predeterminers)
12. PREP (prepositions)
13. PRON (pronouns)
14. PUNCT (punctuation)
15. SPECIAL (special symbols)
16. VERB (verbs)

2. Comments and examples

This paragraph provides the user with general information about the elements included in the various categories. More detailed information is given in section 3. The two sections overlap partially, but they also complement each other. In the present one, we show some examples of usage, while in 3, we give examples of the involved words, but out of context (in the reported examples, the English translations are 'literal': they reflect the Italian form and not the correct English expression).

1. ADJ: It includes various types of adjectives: standard (qualificative) (bello -nice-, buono -good), interrogative ('quali' fiori vuoi comprare -'which' flowers do you want to buy-), deictic ('questi' fiori sono belli -'these' flowers are nice-), exclamative ('che' bei fiori! -'what' nice flowers!-).

2. ADV: It includes standard adverbs (spesso -often-, bene -well-) and question adverbs (quando -when-, perchè -why-). Probably, this is the most complex category, because the types (subcategories) partially include semantic information.

3. ART: No special comment on articles (un -a-, il -the-).

4. CONJ: Conjunctions, both coordinating (e -and-, o -or-) and subordinating ('mentre' mangiava, leggeva il giornale - 'while' she was eating, she was reading the newspaper-; lo ha baciato 'perchè' lo amava - she kissed him 'because' she loved him-)

5. DATE: The dates, but just when they have been recognized on the basis of their structure (i.e. by the tokenizer). For instance, '10/5/98' will be recognized as a single element (a single output line), and it will get the category DATE. On the contrary, '10 maggio 1998' -10 May 1998- will be taken as three separate elements (a NUM, a NOUN, and another NUM).

6. INTERJ: interjections as 'oh', 'ah'.

7. MARKER: This category has been created in order to handle typographic and formatting markers. Currently, it is used just for some markers which appear in the text of the used corpus, which are of the form <P Prose>, <N Smith John>. In principle, this POS can be associated with any kind of extra-text markers (ex. LaTex or HTML commands). However, this facility is currently very limited and there is no interface enabling a user to define easily a set of markers).

8. NOUN: common and proper nouns

9. NUM: numbers, both in numeric form (123.451) and in character form (centotrentasette -onehundredthirtyseven-).

10. PHRAS: phrasals, i.e. words playing the role of entire sentences (as 'sì' -yes- and 'no').

11. PREDET: predeterminers (i.e. 'tutto' -all-, 'ambedue' -both-)

12. PREP: both the normal prepositions ('di' -of-, 'a' -to-, 'da' -from-, ...) and the so-called 'polysyllabic' prepositions ('durante' -during-, 'sopra' -above-, 'davanti' -before-, ...)

13. PRON: beyond the personal pronouns ('io' -I-, 'tu' -you-, ...), also clitics (mangiando'lo' -eating+it-), the relative pronouns (la ragazza 'che' hai visto -the girl 'whom' you saw-, la casa 'dove' sono nato -the house 'where' I was born-, ...), the interrogative pronouns ('chi' hai incontrato -'who' did you meet-), the indefinite ones ('molti' credono in lui -'many' believe in him-) and the exclamative ones ('che' hai fatto! -'what' have you done!-)

14. PUNCT: various punctuation marks, as periods, commas, parentheses, hyphens, and so on.

15. SPECIAL: special symbols, i.e. all characters which are not standard punctuation marks (ex. $, #, &, % ...).

16. VERB: main verbs, but also auxiliars and modals. It must be noted that in all cases where the corresponding lemma is not explicitly present in the dictionary with another category, the past and present participles will be tagged as VERB. For instance, if 'interesting' appears as an ADJ in the dictionary, it is up to the tagger to choose between the ADJ and VERB (gerund) reading. Otherwise, it will appear in the input as a VERB.

3. The syntactic types (subcategories)

1. ADJ (adjectives)
   - DEITT (deictic: altro, fa, prossimo, scorso, ...)
   - DEMONS (demonstrative: questo, quello)
   - EXCLAM (exclamative: che)
   - INDEF (indefinite: nessun, alcuni, molti, qualsiasi, ...)
   - INTERR (interrogative: che, quale, quanto)
   - ORDIN (ordinal: primo, ventesimo, ultimo, ...)
   - POSS (possessive: altrui, mio, nostri, ...)
   - QUALIF (qualificative: bello, grande, italiano, ...)

2. ADV (adverbs)
   - ADFIRM (adfirmative: certo)
   - ADVERS (adversative: anzi, pero')
   - COMPAR (comparative: piu', meglio, peggio, cosi')
   - DOUBT (doubt: forse)
   - INTERR (interrogative: come, dove, perche', ...)
   - LIMIT (limit: solo, soltanto)
   - LOC (locative: sopra, intorno, lassu', sottoterra, ...)
   - MANNER (manner: cosi', volentieri, ...; this type includes also all the adverbs derived from adjectives by means of the -mente suffix (which roughly corresponds to -ly in English, ex. forte --> fortemente -strong --> strongly-)
   - NEG (negation: non, senza, neanche, nemmeno, ...)
   - QUANT (quantification: meno, circa, assai, troppo, ...)
   - REASON (motivation: infatti, quindi)
   - STRENG (strengthening: perfino, persino, anche)
   - SUPERL (superlative: benissimo)
   - TIME (time: poi, prima, ormai, spesso, ...)

3. ART (articles)
   - DEF (definite: il, la, gli, ...)
   - INDEF (indefinite: un, una, un', uno, degli, ...)

4. CONJ (conjunctions)
   - COORD (coordinative: e, o, ma, eppure, inoltre, ...)
   - SUBORD (subordinative: che, nonostante, poiche', quando, ...)
   - COMPAR (comparative: a, che, di, come)

5. DATE (dates)
   no type

6. INTERJ (interjections)
   no type

7. MARKER (markers)
   no type

8. NOUN (nouns)

- COMMON
- PROPER

9. NUM (numbers)

   No type

10. PHRAS (phrasals)

    No type

11. PREDET (predeterminers)

    No type

12. PREP (prepositions)
    - MONO (monosyllabic: di, a, da, in, ...)
    - POLI (polysyllabic: attorno, accanto, prima, sopra, ...)

13. PRON (pronouns)
    - DEMONS (demonstrative: cio', medesimo, questo, coloro, ...)
    - EXCLAM (exclamative: che, chi)
    - INDEF (indefinite: chiunque, nessuno, qualcosa, ...)
    - INTERR (interrogative: chi, che, quale, quanto)
    - LOC (locative: ne, ci, vi)
    - PERS (personal: io, tu, noi, lei)
    - POSS (possessive: mio, tuo, nostro, proprio, ...)
    - REFL-IMPERS (reflexive-impersonal: ci, vi, si, se)
    - RELAT (relative: che, quale, cui, come, dove, ...)

14. PUNCT (punctuation)

    No type

15. SPECIAL (special symbols)

    No type

16. VERB (verbs)
    - MAIN (all standard verbs, but also copulas)
    - AUX (auxiliaries: essere, avere, venire, stare)
    - MOD (modals: dovere, potere, volere)

4. Features

1. ADJ
   - Gender (M, F)
   - Number (SING, PL)

2. ADV

   No features

3. ART
   - Gender (M, F)
   - Number (SING, PL)

4. CONJ
   - Semtype (caus [poiche'], manner+time [come], tempo [dopo], loc [dove], conc [nonostante], reason [per], caus+reason [perche'], advers [pero', ma], caus [poiche', siccome], time [quando], cond [se], fin [percio', sicche'], neutral [che]).

5. DATE

   No features

6. INTERJ

No features

7. MARKER

   No features

8. NOUN:

   - Gender (M, F)
   - Number (SING, PL)

   There are two more features which can appear with nouns in case the noun derives from a verb. They specify what is the verb from which the noun derives and if that verb is transitive or not (their name is v-deriv, and v-trans).

   N.B. These features are very important for the module which assigns automatically the grammatical relations. For instance, with 'la caduta di Marco' -the fall of Marco-, since 'caduta' -fall- derives from 'cadere' -to fall-, which is intransitive, to the arc connecting the noun 'caduta' to the preposition 'di' is assigned the relation NOUN-SUBJ (nominal subject). On the contrary, with transitive verbs, the label assigned is NOUN-OBJ ('la distruzione della città' –the destruction of the city-); of course, in this case, it is just
   a preference, since counterexamples do exist.

   N.B.2 The derivation verb can assume the value 'dummy', in case one just wants to force the label assignment (NOUN-SUBJ or NOUN-OBJ) described in the previous note.

9. NUM:

   - Value, i.e. the numeric value (ex. trentatre' --> 33)

10. PHRAS:

    No features

11. PREDET:

    - Gender (M, F)
    - Number (SING, PL)

12. PREP:

    No features

13. PRON:

    - Gender (M, F)
    - Number (SING, PL)
    - Person (1, 2, 3)
    - Case (LSUBJ, LOBJ, LIOBJ, and various combinations of them concatenated with the separator '+'; ex. LOBJ+LIOBJ (this expresses ambiguity); LIOBJ stands for 'indirect object')

14. PUNCT

    No features

15. SPECIAL

    No features

16. VERB:

    - Mood (IND, INFINITE, CONG, PARTICIPLE, CONDIZ, GERUND, IMPER)
    - Tense (PRES, PAST, IMPERF, REMPAST, FUT)
    - Transitivity (TRANS, INTRANS, REFL)
    - Person (1, 2, 3)
    - Number (SING, PL)
    - Gender (M, F)

    Note that among the listed features, only mood, tense, and transitivity ar always present. For instance, for infinites and gerunds all other features are absent, and the gender appears only with past participles.

5. Locutions

Currently, the tagger recognizes a limited number of locutions (about 100). The term 'locution' is intended to mean a lemma composed of more than one word, as for instance, "più o meno" -more or less-, "per esempio" -for instance. In the output, locutions are identical to all other entries, except for the presence of the marker LOCUTIION at the end of the syntactic information described above.