

Spanish Tree Bank:
Specifications
Version 5

Antonio Moreno Sandoval
Susana López Ruesga
Fernando Sánchez León

Laboratorio de Lingüística Informática
Universidad Autónoma de Madrid

30 April 1999

Índice General

1	Introduction	3
2	Basic encoding strategies	4
3	Annotation schema	5
4	Inventory of units	6
4.1	Parts of speech	6
4.2	Phrases and clauses	11
5	Problematic issues	16
5.1	Ambiguity and ambivalence	16
5.2	Null elements	18
5.2.1	Null subjects	19
5.2.2	Ellipsed material	20
5.3	Pronouns vs. noun modifiers	21
5.4	Complex constituents	21
5.4.1	Postclitics	21
5.4.2	Amalgams <i>del</i> and <i>al</i>	22
5.4.3	Periphrastic or semi-auxiliary verbs	23
5.4.4	Compound PPs	25
5.4.5	Discontinuous constituents	26
5.5	“Se” constructions	27
5.5.1	“Se” that corresponds to “le”	27
5.5.2	”Se” in a reflexive or reciprocal construction	27
5.5.3	Pronominal ”se”	28
5.5.4	”Se” as an ”intransitive-marker”	29
5.5.5	Impersonal ”se”	30
5.6	Impersonal/existential constructions	31
5.6.1	Unipersonal constructions	31
5.6.2	Existential constructions	31
5.6.3	Impersonal ”se”	32
5.7	Predicative complements	34
5.8	Criteria for the categorial status assignment	36

5.8.1	Infinitives	36
5.8.2	Gerunds	38
5.8.3	Participles	39
5.8.4	Subordinate clauses as Nouns and Adjectives	40
5.9	Grammaticalisation and Lexicalisation	41
5.10	Conditional, Comparative and Coordinated Structures	46
5.10.1	Conditionals	47
5.10.2	Comparisons	48
5.10.3	Coordinations	51
5.11	Special categories and features	57
5.11.1	Connectives	57
5.11.2	Idiomatic Expressions and Foreign Words	62
5.11.3	Dates and Hours	64
5.11.4	Titles	65
5.11.5	Maths operators, Scores and Measures	66
5.11.6	Punctuation	67
6	Bibliography and References	73
A	Periphrastic verbs	74
B	Control Verbs	81
B.1	Object Control	81
B.2	Subject Control	82
B.3	Subject and Object Control	82
C	Multi-word Units	83
C.1	Multi-word Adverbs	83
C.2	Multi-word Conjunctions	85
C.3	Multi-word Prepositions	85

Capítulo 1

Introduction

This report presents a preliminary version of the Specifications for the Spanish Tree Bank being developed at UAM with funding support by NYU. The aim is to provide a sort of guidelines for encoding the syntactic annotation of sentences. Two are the sources of this report:

- **Previous documents on syntactic annotation**, concretely Marcus et al.(1993), *Building a Large Annotated Corpus of English: The Penn Treebank*, Bies et al.(1995) *Bracketing Guidelines for Treebank II Style Penn Treebank Project*, and EAGLES' *Preliminary Recommendations for the Syntactic Annotation of Corpora* (Version of March 1996).
- **Experience** from a tentative annotation of the 1000 sentences, using Fernando Sánchez's morphosyntactic tagger as input.

The current corpus consists of 1000 sentences, with 14909 words, taken from:

- **El País Digital**, the newspaper on-line edition. Sentences 1-50, 101-150 and 201-1000. Each one is an isolated sentence, without context.
- **Organización Consumidores y Usuarios (OCU)** magazines. Sentences 51-100 and 151-200. The sentences are integrated in a context of several paragraphs.

The fifty-fifty division of the 200 first sentences, is intended for testing annotation within a pure syntactic environment and annotation with a discourse perspective. The other 700 sentences are again isolated sentences since from now on our intention is to enlarge the corpus.

Capítulo 2

Basic encoding strategies

The items are not ordered by precedence or importance.

- **Combination of Parts Of Speech (POS) tags and features:** the annotation format combines a typical POS label with feature values that specify the syntactic information for each (terminal and non-terminal) element.
- **Layers of annotation:** we annotate syntactic categories (i.e., parts-of-speech such as NOUNS, ADJECTIVES, ...), syntactic functions (e.g. SUBJ, OBJ1 (direct object), OBJ2 (indirect object)...), syntactic features (i.e. number, gender, tense, etc.) and some semantic features (HUMAN, TIME, etc.). The actual string is presented between ‘‘<...>’’, and its lexeme is between ‘ ‘... ’’.
- **Eliminate redundancy as possible:** to avoid any kind of duplication, the information is specified only once in the relevant layer. For instance, internal agreement features between a noun and its modifiers are not percolated up to the NP if they will not be used in further stages (for instance, the NP inside a PP). This strategy causes **underspecification**, since not all syntactic details are included. As EAGLES' recommendations put it: “It is normal for a parsing scheme to be selective in specifying the syntactic phenomena to be identified and labelled, and omission of some details is almost inevitable”.
- **When there is an uncertainty about how to assign a given feature, better do not do it.** This attitude produces ambivalence, but we prefer that, better than making arbitrary decisions.
- **Reflect the surface syntax:** we are very cautious about empty categories, and how to annotate them. Only two kinds of null elements are labelled: null subjects and ellipsed material.

Capítulo 3

Annotation schema

We follow a simple, vertical, indented format, close to the PROTEUS parse format:

```
(CATEGORY '<string>' 'lexeme' FEATURE1 FEATURE2 ... )
```

For the actual catalog of allowed categories and features, please consult Chapter 4, “Inventory of units”.

In order to label complex constituents, i.e. those with more than one element such as VP, verbs with clitics, semi-auxiliaries, etc., we use the balanced ‘(’’):

```
(CAT1 ...  
(CAT2 ...)  
( ... )  
(CATn ...))
```

In Spanish, the left-most element is the head of the complex constituent, except in complex verbs (i.e. analytical tenses such as “he comido”, or periphrastic constructions such as “tengo que comer”) where the lexical head is the right-most element.

Capítulo 4

Inventory of units

4.1 Parts of speech

% Adjective

(ADJ

MASC|FEM

SG|PL

COMPARATIVE|SUPERLATIVE

DISCONTINUOUS

INTERROGATIVE

IDIOM

REF

FOREIGN-WORD

COORDINATED)

% Adverb

(ADV

INTERROGATIVE|NORMAL|LOCATIVE|MODAL|NEG|TIME|REL

DEG|COMPARATIVE

IDIOM

FOREIGN-WORD

COORDINATED)

% Article

(ART

DEF | INDEF
MASC | FEM | NEUT
SG | PL)

% Auxiliary

(AUX

TENSED | UNTENSED
PRES | IMPERFECT | FUT | PAST | COND
INFINITE | GERUND | PART
MASC | FEM
IND | SUBJUNCTIVE | IMPERATIVE
PERFECT | PROGRESSIVE | INCHOATIVE | MODAL | PASSIVE | FUTURE | COPULATIVE
SG | PL
P1 | P2 | P3
POLITE
IDIOM
#CLITIC
DISCONTINUOUS
REF
COORDINATED)

% Conjunction

(C

COORDINATING | SUBORDINATING
NEG | FINITE | INFIN
COMPARATIVE | TIME
DISCONTINUOUS
IDIOM
REF)

% Demonstrative

(DEM

MASC | FEM | NEUT
SG | PL
COORDINATED)

% Maths-op

(MATHS-OP)

% Noun

(N

MASC | FEM | NEUT
SG | PL
ACRNM | PCENT | TRATM
PROPER
MEASURE
FOREIGN-WORD
IDIOM
TIME
LOCATIVE
COMPARATIVE
COORDINATED)

% Possesive

(POSS

MASC | FEM
SG | PL
P1 | P2 | P3
COORDINATED)

% Preposition

(PREP)

% Pronoun

(P

INTERROGATIVE | PERS | POSS | REL | QU | DEM | REFLEXIVE
P1 | P2 | P3
DEF | INDEF
NOM | NOM/OBL | ACC-DAT | ACC | DAT
MASC | FEM | NEUT | MASC/NEUT
LOCATIVE | TIME
SG | PL
NONHUMAN | HUMAN
DISCONTINUOUS
COORDINATED)

Feature implications:

REL > HUMAN

% Punctuation marks

(PUNCT "," COMMA)
(PUNCT "." PERIOD)
(PUNCT "-" DASH)
(PUNCT "(" PARENTHESIS)
(PUNCT ")" PARENTHESIS)
(PUNCT "{" BRACE)
(PUNCT "}" BRACE)
(PUNCT "[" BRACKET)
(PUNCT "]" BRACKET)
(PUNCT " " QUOTE)
(PUNCT ":" COLONS)
(PUNCT ";" SEMICOLONS)
(PUNCT "... " ELLIPSES)
(PUNCT "?" QUESTION)
(PUNCT ">" QUESTION)
(PUNCT "!" EXCLAMATION)
(PUNCT "<" EXCLAMATION)

NOTE: the feature QUOTE is empty because of problems with lists in
Lisp.

% Quantifier

(Q

MASC|FEM
SG|PL
NONHUM|DISTR|ROMAN
COORDINATED)

% Se-mark

(SE-MARK

PASSIVE|IMPERSONAL)

% Verb

(V

TENSED|UNTENSED
PRES|IMPERFECT|FUT|PAST|COND
INFINITE|GERUND|PART
MASC|FEM
IND|SUBJUNCTIVE|IMPERATIVE
PERFECT|PROGRESSIVE|INCHOATIVE|MODAL|PASSIVE|FUTURE|COPULATIVE
SG|PL
P1|P2|P3
POLITE
IDIOM
#CLITIC
DISCONTINUOUS
REF
COORDINATED)

Feature implications:

TENSED > {PRES|IMPERFECT|FUT|PAST}
{IND|SUBJUNCTIVE|IMPERATIVE}
{SG|PL} {P1|P2|P3}

UNTENSED > {INFINITE|GERUND|PART}

PART > {MASC|FEM}
{SG|PL}

4.2 Phrases and clauses

% Adjective Phrases

(ADJP

{ATTR}

{APPOS}

{COMPARATIVE|COMPARATIVE-1|COMPARATIVE-2}

{SUPERLATIVE}

{INTERROGATIVE}

{IDIOM}

{DISCONTINUOUS|DISCONTINUOUS-1|DISCONTINUOUS-2}

{MASC|FEM}

{SG|PL}

{COORDINATED}})

% Adverb Phrases

(ADVP

{INTERROGATIVE|NORMAL|LOCATIVE|MODAL|NEG|TIME|REL}

{DEG|COMPARATIVE}

{IDIOM}

{COORDINATED}})

% Clauses

(CL

```

{COMPLETIVE|INFINITIVE|RELATIVE|PARTICIPLE|GERUND}
{INTERROGATIVE|TIME|LOCATIVE|MODAL|EXPLICATIVE|CONCESIVE|NEG|CAUSE|FINAL|ADVERSATIVE}
{COMPARATIVE|COMPARATIVE-1|COMPARATIVE-2}
{CONDITIONAL|CONDITION|MAIN}
{COORDINATED|COORDINATED-1|COORDINATED-2}
{CORRELATIVE|CORRELATIVE-1|CORRELATIVE-2}
{DISCONTINUOUS|DISCONTINUOUS-1|DISCONTINUOUS-2}
{DISTRIBUTIVE|DISTRIBUTIVE-1|DISTRIBUTIVE-2}
{INDIRECT SPEECH|DIRECT SPEECH}
{IMPERSONAL|PASSIVE}
{ABS-PART}
{MASC|FEM}
{SG|PL}
{SUBJ|OBJ1|OBJ2|OBL|ATTR|APPOS}
{REF})

```

Feature implications:

```
\begin{verbatim}
```

```

ABS-PART > {MASC|FEM}
{SG|PL}
{REF}

```

```

(CON)
(DATE)
(... DISCONTINUOUS)

```

Feature implications:

```
DISCONTINUOUS > REF
```

```
% Hour
```

```
(HOUR)
```

% Noun Phrases

(NP

{SUBJ|OBJ1|OBJ2|OBL|ATTR|TIME|LOCATIVE|MEASURE|APPOS}
{REF|ID}
{COMPARATIVE|COMPARATIVE-1|COMPARATIVE-2}
{CORRELATIVE|CORRELATIVE-1|CORRELATIVE-2}
{MASC|FEM|NEUT}
{SG|PL}
{P1|P2|P3}
{IDIOM}
{POLITE}
{COORDINATED}})

Feature implications:

APPOS > REF

% Null element marks

* only in null subjects

? for ellipsed material different from null subjects.

% Predicative Complements

(PRED-COMPL

{REF}
{MASC|FEM}
{SG|PL}
{COORDINATED}})

% Prepositional Phrases

(PP

{"pval"}
{LOCATIVE|TIME|INTERROGATIVE}
{OBJ1|OBJ2|OBL}
{COMPARATIVE|COMPARATIVE-1|COMPARATIVE-2}
{CORRELATIVE|CORRELATIVE-1|CORRELATIVE-2}
{IDIOM}
{COMPOUND}
{COORDINATED}})

*pval= any single or multiword preposition.

% Quantifier Phrases

(QP

{COORDINATED}})

% Score

(SCORE)

% Sentences

(S

{IMPERSONAL}
{COORDINATED|CONDITIONAL|CORRELATIVE|DISCONTINUOUS|DISTRIBUTIVE}})

% Titles

(TITLE

{SUBJ|OBJ1|OBJ2}

{SG|PL}
{P3}
{COORDINATED}}

% Verb Phrases

(VP

{TENSED|UNTENSED}
{PRES|IMPERFECT|FUT|PAST|COND}
{INFINITE|GERUND|PART}
{MASC|FEM}
{DISCONTINUOUS|DISCONTINUOUS-1|DISCONTINUOUS-2}
{DISTRIBUTIVE|DISTRIBUTIVE-1|DISTRIBUTIVE-2}
{IND|SUBJUNCTIVE|IMPERATIVE}
{PERFECT|PROGRESSIVE|INCHOATIVE|MODAL|PASSIVE|FUTURE|COPULATIVE|IMPERSONAL}
{SG|PL}
{P1|P2|P3}
{POLITE}
{COORDINATED}}

Capítulo 5

Problematic issues

5.1 Ambiguity and ambivalence

We are here only concerned with syntactic ambiguity. Considering the sentence below, we can provide two valid parses:

”La quiebra de otro banco en Japón presagia una grave crisis financiera”.

First analysis:

```
(S
  (NP SUBJ FEM SG P3
    (ART "<La>" "la" DEF FEM SG)
    (N "<quiebra>" "quiebra" FEM SG)
    (PP DE
      (PREP "<de>" "de")
      (NP
        (QP
          (Q "<otro>" "otro" MASC SG))
          (N "<banco>" "banco" MASC SG))
        (PP EN LOCATIVE
          (PREP "<en>" "en")
          (NP
            (N "<Japón>" "Japón" PROPER))))))
  (VP TENSED PRES IND SG P3
    (V "<presagia>" "presagiar" TENSED PRES IND SG P3)
    (NP OBJ1
      (ART "<una>" "una" INDEF FEM SG)
      (ADJP FEM SG
        (ADJ "<grave>" "grave" FEM SG))
      (N "<crisis>" "crisis" FEM SG)
      (ADJP FEM SG
```

(ADJ "<financiera>" "financiera" FEM SG))))))

Second analysis:

```
(S
  (NP SUBJ FEM SG P3
    (ART "<La>" "la" DEF FEM SG)
    (N "<quiebra>" "quiebra" FEM SG)
    (PP DE
      (PREP "<de>" "de")
      (NP
        (QP
          (Q "<otro>" "otro" MASC SG))
          (N "<banco>" "banco" MASC SG))))
    (PP EN LOCATIVE
      (PREP "<en>" "en")
      (NP
        (N "<Japón>" "Japón" PROPER)))
  (VP TENSED PRES IND SG P3
    (V "<presagia>" "presagiar" TENSED PRES IND SG P3)
    (NP OBJ1
      (ART "<una>" "una" FEM SG)
      (ADJP FEM SG
        (ADJ "<grave>" "grave" FEM SG))
      (N "<crisis>" "crisis" FEM SG)
      (ADJP FEM SG
        (ADJ "<financiera>" "financiera" FEM SG))))))
```

We do not want to provide both trees but only one. We may trust on the intuitions of the annotator and, we may only consider those ambiguities that can make any difference to the interpretation of the sentence. EAGLES recommends to use the format:

```
// x1 / x2 / ... / xn//
```

which is an approach we believe to be easier for a human reader to interpret.

Accordingly, our example above would be annotated this way:

```
(S
```

```

//(NP SUBJ FEM SG P3
  (ART "<La>" "la" DEF FEM SG)
  (N "<quiebra>" "quiebra" FEM SG)
  (PP DE
    (PREP "<de>" "de")
    (NP
      (QP
        (Q "<otro>" "otro" MASC SG))
        (N "<banco>" "otro" MASC SG))
      (PP EN LOCATIVE
        (PREP "<en>" "en")
        (NP
          (N "<Japón>" "Japón"PROPER)))))))/
(NP SUBJ FEM SG P3
  (ART "<La>" "la" DEF FEM SG)
  (N "<quiebra>" "quiebra" FEM SG)
  (PP DE
    (PREP "<de>" "de")
    (NP
      (QP
        (Q "<otro>" "otro" MASC SG))
        (N "<banco>" "banco" MASC SG))))
  (PP EN LOCATIVE
    (PREP "<en>" "en")
    (NP
      (N "<Japón>" "Japón"PROPER))))//
(VP TENSED PRES IND SG P3
  (V "<presagia>" "presagiar" TENSED PRES IND SG P3)
  (NP OBJ1
    (ART "<una>" "una" INDEF FEM SG)
    (ADJP FEM SG
      (ADJ "<grave>" "grave" FEM SG))
    (N "<crisis>" "crisis" FEM SG)
    (ADJP FEM SG
      (ADJ "<financiera>" "financiera" FEM SG))))))

```

5.2 Null elements

We will follow basically the Penn Treebank schema for empty elements, but we will only annotate **null subjects** (*) and **ellipsed material** (*?*) such as VP, PP, CL, etc. in coordinations, and also required objects. We do not mark traces (*T* in the Penn Treebank).

5.2.1 Null subjects

Null subjects can be found in two different constructions:

1. As a pro-drop language, Spanish usually omits implicit subjects, which nevertheless are recoverable from the verb agreement.
2. In the infinitive, participle and gerund clauses, which normally are referred to as *raising and control sentences*.

In both cases we will use the same format:

```
(NP * FUNCTION FEATURE1 FEATURE2 ...)
```

For null subjects in sentences with main verb tensed, the agreement features (i.e. number and person) are encoded:

”Regresamos ayer”.

```
(S
  (NP * SUBJ PL P1)
  (VP TENSED PRES IND PL P1
    (V "<Regresamos>" "regresar" TENSED PRES IND PL P1)
    (ADVP TIME
      (ADV "<ayer>" "ayer"TIME))))
```

For null subjects in untensed sentences, we need to specify a **reference index (REF)** in the empty NP and, an identity index (ID) in the coreferenced NP :

”Juan quiere leer un libro de Chejov”

```
(S
  (NP SUBJ ID-1 SG P3
    (N "<Juan>" "Juan" PROPER SG P3))
  (VP TENSED PRES IND SG P3
    (V "<quiere>" "querer" TENSED PRES IND SG P3)
    (CL INFINITIVE OBJ1
      (NP * SUBJ REF-1)
      (VP UNTENSED INFINITE
        (V "<leer>" "leer" UNTENSED INFINITE)
        (NP OBJ1
          (ART "<un>" "uno" INDEF MASC SG)
          (N "<libro>" "libro" MASC SG))
```

```

      (PP DE
        (PREP "<de>" "de")
          (NP
            (N "<Chejov>" "Chejov" PROPER)))))))))

```

The correferenced index is not always easy to assign. In those cases where there is not a clear candidate for the ID feature, then it is not specified:

”Alguien propuso cantar una canción”.

```

(S
  (NP SUBJ SG P3
    (P "<Alguien>" "alguien" INDEF SG P3))
  (VP TENSED PAST IND SG P3
    (V "<propuso>" "proponer" TENSED PAST IND SG P3)
    (CL INFINITIVE OBJ1
      (NP * SUBJ)
      (VP UNTENSED INFINITE
        (V "<cantar>" "cantar" UNTENSED INFINITE)
        (NP OBJ1
          (ART "<una>" "una" FEM SG)
          (N "<canción>" "canción" FEM SG)))))))))

```

We do not annotate NP movement in passives, as Penn Treebank does, nor mark any trace.

5.2.2 Ellipsed material

We use the Penn Treebank tag, ***?***, for marking any ellipsed element different from the empty subjects. However, we do not follow the Penn Treebank strategy of annotating ellipsis in comparatives, but only in coordinations. In addition, we include the REF feature in the null element, and the ID feature in the correferenced one, as we do with null subjects in untensed sentences. This is an example of verb ellipsis:

“Wall Street retrocedió el 1,44% y Tokio un 2,30%”.

```

(S COORDINATED
  (CL COORDINATED-1
    (NP SUBJ SG P3
      (N "<Wall Street>" "Wall_Street" PROPER))
    (VP TENSED PAST IND SG P3
      (V "<retrocedió>" "retroceder" TENSED PAST IND SG P3 ID-1)
      (NP OBJ1
        (ART "<el>" "el" DEF MASC SG)

```

```

      (N "<1,44%>" "catorce_coma_cuarenta_y_cuatro_por_ciento"
MASC SG PCENT))))
    (C "<y>" "y" COORDINATING)
    (CL COORDINATED-2
      (NP SUBJ SG P3
        (N "<Tokio>" "Tokio" PROPER))
      (VP TENSED PAST IND SG P3
        (V *?* REF-1)
        (NP OBJ1
          (ART "<un>" "un" INDEF MASC SG)
          (N "<2,30%>" "dos_coma_treinta_por_ciento" MASC SG PCENT))))))

```

5.3 Pronouns vs. noun modifiers

In Spanish many closed categories can be functionally either pronouns or noun modifiers (demonstrative, possessive, quantifier, etc.). E.g. “Todos queremos más” (*Todos* is a pronoun) vs. “Todos los hombres queremos más” (here it is a quantifier).

The way of distinguishing both is:

- **pronoun**: the category is P, and the relevant features are included (e.g. QU, DEM, POSS):

```
(P "<Todos>" "todo" QU MASC PL)
```

- **noun modifier**: just the proper category:

```
(Q "<Todos>" "todo" MASC PL)
```

5.4 Complex constituents

5.4.1 Postclitics

Clitics in Spanish can appear before (preclitics) or after (postclitics) the verb. Significantly, preclitics are always separate words (*se lo dio*), and postclitics are always joined to the verb (*darselo*). Annotating preclitics is not problematic, since they are pronouns, and therefore NPs. For postclitics we can split the input string into parts: “dar” “se” “lo”. But we want to show that both, verb and clitic(s), constitute a compound. We mark this using the following format:

```
(VP UNTENSED INFINITE
```

(V "<dárselo>" "dar" UNTENSED INFINITE #CLITIC DISCONTINUOUS
 ID-1
 (NP
 (P "se" PERS P3 SG DISCONTINUOUS REF-1))
 (NP OBJ1
 (P "lo" PERS P3 ACC MASC SG DISCONTINUOUS REF-1))))

This is an example from the corpus:

"colocarlo sobre el colchón"

(VP UNTENSED INFINITE
 (V "<colocarlo>" "colocar" UNTENSED INFINITE #CLITIC DISCONTINUOUS
 ID-1
 (NP OBJ1
 (P "lo" PERS P3 ACC MASC SG DISCONTINUOUS REF-1)))
 (PP SOBRE LOCATIVE
 (PREP "<sobre>" "sobre")
 (NP
 (ART "<el>" "el" DEF MASC SG)
 (N "<colchón>" "colchón" MASC SG))))

5.4.2 Amalgams *del* and *al*

There are only two amalgams (portmanteau words) in Spanish: *del* is a combination of preposition DE + article EL; *al* = preposition A + article EL. This is a problem for annotation since the PREP is the head of the PP and the article is part of the NP, which is definite. One possibility is to manually split the amalgam ("

"Los últimos partidos del equipo"

(NP SUBJ MASC PL P3
 (ART "<Los>" "el" DEF MASC PL)
 (ADJP MASC PL
 (ADJ "<últimos>" "último" MASC PL))
 (N "<partidos>" "partido" MASC PL)
 (PP DE
 (PREP "" "de")
 (NP
 (ART "el" DEF MASC SG)
 (N "<equipo>" "equipo" MASC SG))))

”La cadena emite noticias 24 horas al día”

```
(S
(NP SUBJ FEM SG P3
 (ART "<La>" "la" DEF FEM SG)
 (N "<cadena>" "cadena" FEM SG))
(VP TENSED PRES IND SG P3
 (V "<emite>" "emitir" TENSED PRES IND SG P3)
(NP OBJ1
 (N "<noticias>" "noticia" FEM PL))
(NP TIME
 (QP
 (Q "<24>" "veinticuatro"))
 (N "<horas>" "hora" FEM PL)
 (PP A
 (PREP "<al>" "a")
 (NP
 (ART "el" DEF MASC SG)
 (N "<día>" "día" MASC SG))))))
```

5.4.3 Periphrastic or semi-auxiliary verbs

The Spanish language has a number of so-called “periphrastic” or “semi-auxiliary verbs” which are a combination of TENSED VERB + PARTICLE (a PREP or a C) + UNTENSED VERB. The tensed verb is a kind of auxiliary verb with different meanings. For instance, “ir a” is distributionally equivalent to the future tense; “tener que” expresses modality, etc. The untensed verb is the actual main verb. There are two possible annotations:

1. as a regular AUX + MAIN VERB sentence (as passives, or compound tenses). In this case, AUX and particle build a compound constituent: (AUX ... (PREP or C ...)):

”Manuel tiene que ir al dentista el viernes”

```
(S
(NP SUBJ MASC SG P3
 (N "<Manuel>" "Manuel" PROPER))
(VP TENSED PRES IND MODAL SG P3
 (V "<tiene que ir>" "ir" TENSED PRES IND MODAL SG P3
 (AUX "<tiene que>" "tener_que" TENSED PRES IND SG P3)
 (V "<ir>" "ir" UNTENSED INFINITE))
 (PP A LOCATIVE
```

```

(PREP "<a1>" "a")
(NP
  (ART "el" DEF MASC SG)
  (N "<dentista>" "dentista" MASC SG)))
(NP TIME
  (ART "<el>" "el" DEF MASC SG)
  (N "<viernes>" "viernes" MASC SG))))

```

2. as a typical control sentence:

“El entrenador forzó a Juan a terminar el partido.”

```

(S
  (NP SUBJ SG P3
    (ART "<El>" "el" DEF MASC SG)
    (N "<entrenador>" "entrenador" MASC SG))
  (VP TENSED PAST IND SG P3
    (V "<forzó>" "forzar" TENSED PAST IND SG P3)
    (PP A OBJ1
      (PREP "<a>" "a")
      (NP REF-1
        (N "<Juan>" "Juan" PROPER))))
    (PP A OBJ2
      (PREP "<a>" "a")
      (CL INFINITIVE
        (NP * SUBJ REF-1)
        (VP UNTENSED INFINITE
          (V "<terminar>" "terminar" UNTENSED INFINITE)
          (NP OBJ1
            (ART "<el>" "el" DEF MASC SG)
            (N "<partido>" "partido" MASC SG))))))))
  (PUNCT "." PERIOD))

```

Therefore, we have to define the conditions for distinguishing between subject control and semi-auxiliar constructions, so, we include a list of most of the possible periphrastic verbs that can be found in Spanish (see Appendix A, where a non exhaustive list of verbal compounds may be found).

Periphrastic verbs can be classified according to the features they have in the following way:

1. Perfect
2. Progressive
3. Inchoative
4. Modal
5. Passive
6. Future

Within the feature PERFECT we may find periphrasis such as “haber de + participle”, “terminar de + infinitive”, etc.

As PROGRESSIVE we have, for example, “estar + gerund”, “venir a + infinitive”, “venir + gerund”, “volver a + infinitive”, etc.

The INCHOATIVE feature is present in verbs such as: “comenzar a + infinitive”, “empezar a + infinitive”, etc.

MODAL periphrasis are for example: “deber + infinitive”, “deber de + infinitive”, “haber de + infinitive”, “poder + infinitive”, “tener que + infinitive”, etc.

PASSIVE periphrastic verbs, are for example: “estar + participle”, etc.

The feature FUTURE is found in periphrasis such as: “ir a + infinitive”.

Following this strategy we may find some inconsistencies of features since we may have verbs with two features which are, for example, IMPERFECT and PERFECT one comes from the auxiliary verb and the other from the main verb. Both the auxiliary and the main verb forms a compound verb, as we have said, so, the decision we have taken is to leave the feature IMPERFECT on the auxiliary verb, and that the feature PERFECT, from the main verb, prevails over the features included in the compound verb.

NOTE: Regarding subject control constructions, we also include a list (see Appendix B).

5.4.4 Compound PPs

Some time and locative adjuncts are prepositional phrases expressing an interval by means of two correlative prepositions. We have decided to treat those cases as a compound PP because of their semantics. The pval value is compounded by the two prepositions (e.g. DE-A, see example below), and the feature COMPOUND is added in these cases.

”de aqui a diciembre”

(PP DE-A TIME COMPOUND)

```

(P P DE
  (PREP "<de>" "de")
  (NP TIME
    (N "<aqui>" "aquí" TIME)))
(P P A
  (PREP "<a>" "a")
  (NP TIME
    (N "<diciembre>" "diciembre" TIME)))

```

”desde Madrid hasta Barcelona”

```

(P P DESDE-HASTA LOCATIVE COMPOUND
  (P P DESDE
    (PREP "<desde>" "desde")
    (NP LOCATIVE
      (N "<Madrid>" "Madrid" PROPER)))
  (P P HASTA
    (PREP "<hasta>" "hasta")
    (NP LOCATIVE
      (N "<Barcelona>" "Barcelona" PROPER))))

```

5.4.5 Discontinuous constituents

In the preceding sections we have dealt with some compound or multi-word units. It is frequent to find some of these units broken by external elements. For example, “tener en cuenta” is a lexicalised multi-word verb that can appear broken, as in “tener siempre en cuenta”. We want to stress that this can happen with any category, not only with verbs.

Our treatment is based on the inclusion of an special feature: DISCONTINUOUS, and the reference and identification marks:

```

(... )
(CL
  (C "<pero>" "pero" COORDINATING)
  (VP UNTENSED GERUND
    (V "<teniendo>" "tener_en_cuenta" UNTENSED GERUND DISCONTINUOUS
REF-1)
    (ADVP TIME
      (ADV "<siempre>" "siempre")))
  (V "<en cuenta>" DISCONTINUOUS ID-1)
  (CL COMPLETIVE OBJ1
    (C "<que>" "que" SUBORDINATING FINITE)
    (...))))

```

5.5 “Se” constructions

One of the major problematic issues in Spanish is the one that involves the word “se”. Even in the way of naming this “se” there are many disagreements (some linguists call it pronoun, others clitic, etc.). Here we are going to treat it as a pronoun in the majority of cases, except when it functions as an intransitive marker or as an impersonal marker.

We have established five different annotations involving “se-constructions” depending on the kind of “se” that appears. These five annotations are the following ones:

5.5.1 “Se” that corresponds to “le”

In this case, “se” is a pronoun that substitutes “le” due to phonetic reasons. “Se” in this construction is treated as a postclitic (see point 5.4.) but it precedes the verb. Example:

”Se lo dio”

```
(S
  (NP * SUBJ SG P3)
  (VP TENSED PAST IND SG P3
    (V "<se lo dio>" "dar" TENSED PAST IND SG P3
      #CLITIC
      (NP OBJ1
        (P "<se>" "le" PERS ACUS SG P3))
      (NP OBJ2
        (P "<lo>" "lo" PERS DAT SG P3))
      (V "<dio>" "dar" TENSED PAST IND SG P3))))
```

5.5.2 “Se” in a reflexive or reciprocal construction

This “se” is also treated as a pronoun, and as such it changes according to person and number (“me” SG P1, “te” SG P2, “se” SG P3, “nos” PL P1, “os” PL P2, “se” PL P3). When this pronoun appears the subject and the patient of the action are the same. The difference between reflexive and reciprocal is a semantic one, so we are going to consider both equally. Example:

”Yo me lavo”

```
(S
  (NP SUBJ SG P1
    (P "<yo>" "yo" PERS NOM SG P1))
  (VP TENSED PRES IND SG P1
```

```

(NP
  (P "<me>" "me" REFLEXIVE SG P1 DISCONTINUOUS REF-1))
  (V "<lavo>" "lavarse" TENSED PRES IND SG P1 #CLITIC DISCONTINUOUS
ID-1)))

```

5.5.3 Pronominal "se"

This type of "se" appears in constructions where "se" is attached to the verb, that is, "se" is part of the verb, both appear together in the lexicon, for example "dormirse", "ponerse", etc. This type of "se" is going to be annotated as a compound word; it can precede or follow the verb, but we will treat it as a part of the verb either way. When the "se" precedes the verb, we will treat it as a pronoun with its own features plus the discontinuous and reference marks, to maintain its autonomy as a pronoun that forms part of the verb. If it follows the verb we will leave it in that way, adding the feature #CLITIC. Examples:

"A partir de ahora se abre el interrogante respecto a la recuperación del piloto"

```

(S
  (VP TENSED PRES IND SG P3
    (PP A-PARTIR-DE TIME
      (PREP "<A partir de>" "a_partir_de")
      (NP
        (N "<ahora>" "ahora" TIME))))
    (NP
      (P "<se>" "se" PERS P3 DISCONTINUOUS REF-1))
      (V "<abre>" "abrirse" TENSED PRES IND SG P3 #CLITIC DISCONTINUOUS
ID-1))
    (NP SUBJ MASC SG P3
      (ART "<el>" "el" DEF MASC SG)
      (N "<interrogante>" "interrogante" MASC SG)
      (PP RESPECTO-A
        (PREP "<respecto a>" "respecto_a")
        (NP
          (ART "<la>" "el" DEF FEM SG)
          (N "<recuperación>" "recuperación" FEM SG)
          (PP DE
            (PREP "<del>" "de")
            (NP
              (ART "el" DEF MASC SG)
              (N "<piloto>" "piloto" MASC SG))))))
      (PUNCT "." PERIOD))

```

”Tratan de hacerse con el control de la calle”

```
(S
(NP * SUBJ ID-1 PL P3)
(VP TENSED PRES IND PL P3
(V "<tratan>" "tratar" TENSED PRES IND PL P3)
(PP DE OBL
(PREP "<de>" "de")
(CL INFINITIVE
(NP * SUBJ REF-1)
(VP UNTENSED INFINITE
(V "<hacerse>" "hacerse" UNTENSED INFINITE #CLITIC)
(PP CON OBL
(PREP "<con>" "con")
(NP
(ART "<el>" "el" DEF MASC SG)
(N "<control>" "control" MASC SG)
(PP DE
(PREP "<de>" "de")
(NP
(ART "<la>" "el" DEF FEM SG)
(N "<calle>" "calle" FEM SG))))))))))
(PUNCT "." PERIOD))
```

5.5.4 ”Se” as an ”intransitive-marker”

Here, what we have are three different constructions (that have been called ”passive se”, ”inacusative se” and ”middle se”) which have in common the loss of an argument. They are distinguishable by semantic means, but, as we are only concerned with syntactic criteria, they are to be annotated all the same. In this annotation, ”se” is not treated as a pronoun but as a mark used to make passive constructions. Examples:

- Inacusative se.

”El libro se rompió”

```
(S
(NP SUBJ MASC SG P3
(ART "<El>" "el" DEF MASC SG)
(N "<libro>" "libro" MASC SG))
(VP TENSED PAST IND PASSIVE SG P3
(SE-MARK "<se>" "se" PASSIVE)
(V "<rompió>" "romper" TENSED PAST IND SG P3)))
```

- Passive se.

”La obra se representó”

```
(S
  (NP SUBJ FEM SG P3
    (ART "<La>" "la" DEF FEM SG)
    (N "<obra>" "obra" FEM SG))
  (VP TENSED PAST IND PASSIVE SG P3
    (SE-MARK "<se>" "se" PASSIVE)
    (V "<representó>" "representar" TENSED PAST IND SG P3)))
```

- Middle se.

”Las casas prefabricadas se construyen fácilmente”

```
(S
  (NP SUBJ FEM PL P3
    (ART "<Las>" "la" DEF FEM PL)
    (N "<casas>" "casa" FEM PL)
    (ADJP FEM PL
      (ADJ "<prefabricadas>" "prefabricada" FEM PL)))
  (VP TENSED PRES IND PASSIVE PL P3
    (SE-MARK "<se>" "se" PASSIVE)
    (V "<construyen>" "construir" TENSED PRES IND PL P3)
    (ADVP NORMAL
      (ADV "<fácilmente>" "fácilmente" NORMAL))))
```

5.5.5 Impersonal ”se”

These constructions have the characteristic that there is no syntactic subject. ”Se” is a pronoun that marks (or rather creates) the impersonality. Example:

”Se vende piso”

```
(S IMPERSONAL
  (VP TENSED PRES IND SG P3
    (SE-MARK "<se>" "se" IMPERSONAL))
  (V "<vende>" "vender" TENSED PRES IND SG P3)
  (NP OBJ1
    (N "<piso>" "piso" MASC SG))))
```

SEE point 5.10.3. Impersonal constructions with “se”.

5.6 Impersonal/existential constructions

There are only some special verbs in Spanish able to produce impersonal sentences, that is, sentences that lack a subject in syntactic terms. A characteristic common to all of them is that they do not inflect: they can only take 3rd person singular. We have divided these constructions into three different types depending on the verb that appears on them, or according to the use or not of the particle "se".

5.6.1 Unipersonal constructions

Here we include atmospheric verbs, such as "llover", "nevar", "relampaguear", etc. In languages such as English or French, there is an expletive subject ("It rains", "Il pleut"). This expletive absorbs the syntactic function of the subject of the sentence, though it does not have a semantic content, but in Spanish this expletive does not exist. For this reason, and also because Spanish speakers do not feel the presence of a subject in sentences such as "Llovió durante toda la noche", we have established the following annotation:

"Llovió durante toda la noche"

```
(S IMPERSONAL
 (VP TENSED PAST IND SG IMPERSONAL
  (V "<llovió>" "llover" TENSED PAST IND SG IMPERSONAL)
  (PP DURANTE TIME
   (PREP "<durante>" "durante")
   (NP
    (ADJP FEM SG
     (ADJ "<toda>" "todo" FEM SG))
    (ART "<la>" "la" DEF FEM SG)
    (N "<noche>" "noche" FEM SG))))))
```

5.6.2 Existential constructions

These constructions involve existential verbs such as "hay" (=There is/are), "hace" or "parece". These verbs also lack a subject, and they appear in the 3rd person singular. The annotation we will follow here is the next one:

"Hay cinco libros sobre la mesa"

```
(S IMPERSONAL
 (VP TENSED PRES IND SG IMPERSONAL
  (V "<hay>" "hay" TENSED PRES IND SG)
  (NP
```

```

(QP
  (Q "<cinco>" "cinco"))
(N "<libros>" "libro" MASC PL))
(PP SOBRE LOCATIVE
  (PREP "<sobre>" "sobre")
  (NP
    (ART "<la>" "la" DEF FEM SG)
    (N "<mesa>" "mesa" FEM SG))))))

```

5.6.3 Impersonal "se"

As we said above, "se" is a mark that makes a sentence impersonal when it appears. Most part of verbs can enter in such constructions; the "impersonal" value comes from the fact that when appearing with "se" they inflect in the 3rd person singular. There is no syntactic subject and they have not or they do not imply reference to any definite subject as they would do if the particle "se" were eliminated. A preliminary classification of the verbs that cooccur with "se" is:

1. Intransitive verbs (for example "se vive bien en Barcelona")
2. Transitive verb + Inanimate Direct Object (for example "se vende pan")
3. Transitive verb + Animate Direct Object (here, if the Direct Object is definite and bears the preposition "a" the verb cannot be used in plural; for example: "se ha auxiliado a los heridos")
4. Transitive verb + CL (for example: "se cree que el tiempo mejorará" or "se recomienda fumar con moderación")
5. Transitive verbs in absolute constructions (for example "se lee poco en esta ciudad")

This "se" is going to be treated as a mark of impersonality. Its anotation will be as follows:

"Se vende pan"

```

(S IMPERSONAL
  (VP TENSED PRES IND SG IMPERSONAL
    (SE-MARK "<se>" "se" IMPERSONAL)
    (V "<vende>" "vender" TENSED PRES IND SG P3)
    (NP OBJ1
      (N "<pan>" "pan" MASC SG))))))

```

"Se habla español"

(S IMPERSONAL
(VP TENSED PRES IND SG IMPERSONAL
(SE-MARK "<se>" "se" IMPERSONAL)
(V "<habla>" "hablar" TENSED PRES IND SG P3)
(NP OBJ1
(N "<español>" "español" MASC SG))))

"Se ha auxiliado a los heridos"

(S IMPERSONAL
(VP TENSED PRES IND SG IMPERSONAL
(SE-MARK "<se>" "se" IMPERSONAL)
(AUX "<ha>" "haber" TENSED PRES IND SG P3)
(V "<auxiliado>" "auxiliar" UNTENSED PART PAST MASC SG)
(PP A OBJ1
(PREP "<a>" "a")
(NP
(ART "<los>" "el" DEF MASC PL)
(N "<heridos>" "herido" MASC PL))))

"Se cree que el tiempo mejorará"

(S IMPERSONAL
(VP TENSED PRES IND SG IMPERSONAL
(SE-MARK "<se>" "se" IMPERSONAL)
(V "<creo>" "creer" TENSED PRES IND SG P3)
(CL COMPLETIVE OBJ1
(C "<que>" "que" COMPLETIVE)
(NP SUBJ MASC SG P3
(ART "<el>" "el" DEF MASC SG)
(N "<tiempo>" "tiempo" MASC SG))
(VP TENSED FUT IND SG P3
(V "<mejorará>" "mejorar" TENSED FUT IND SG P3))))

"Se lee poco en esta ciudad"

(S IMPERSONAL
(VP TENSED PRES IND SG IMPERSONAL
(SE-MARK "<se>" "se" IMPERSONAL)
(V "<lee>" "leer" TENSED PRES IND SG P3)
(ADVP
(ADV "<poco>" "poco"))
(PP EN LOCATIVE

```

(PREP "<en>" "en")
(NP
  (DEM "<esta>" "este" FEM SG)
  (N "<ciudad>" "ciudad" FEM SG))))))

```

”Se vive bien en Barcelona”

```

(S IMPERSONAL
  (VP TENSED PRES IND SG P3
    (SE-MARK "<se>" "se" IMPERSONAL)
    (V "<vive>" "vivir" TENSED PRES IND SG P3)
    (ADVP
      (ADV "<bien>" "bien"))
    (PP EN LOCATIVE
      (PREP "<en>" "en")
      (NP LOCATIVE
        (N "<Barcelona>" "Barcelona" PROPER))))))

```

Again, this annotation schema may create other inconsistency, which is that we have verbs with two features that are incompatible: IMPERSONAL and P3. What we have decided here is to erasure the feature P3 (SEE the annotated sentences on the corpus).

5.7 Predicative complements

In Spanish there is an special kind of complements called predicative complements (“complementos predicativos”). These complements say something about the subject, for example in “Juan fue elegido presidente”, the NP “presidente” predicates something about the subject “Juan”.

There is the possibility of having what has been called “secondary predication”, that is, there are two predications as in “Juan trabajó cansado” where two things are predicated about the subject:

1. that Juan worked
2. that when he worked he was tired.

Or in “Juan fue elegido presidente”, where we have also two predications:

1. that Juan was ellected
2. that what he was ellected was to be president

Secondary predication can have its own subject, for example in “Juan encontró el coche roto”, where “el coche roto” is the predicative complement

of "Juan", and "roto" is a secondary predication that has its own subject "coche".

A final remark we want to pose here, before including the annotation schema we may follow to deal with these complements, is that of the criteria which may help us to distinguish between an attribute and a predicative complement. Most of these criteria is semantic (for example aspect) but we may rely only on the syntactic ones. These syntactic criteria are those of pronominal substitution and ellipsis: when we have an attribute, it cannot be elided but it can be pronominalized, that is, it can be substituted by the pronoun "lo". For example:

```
Juan es [el presidente]
*Juan es ...
Juan lo es
```

However, predicative complements can actually be omitted but they cannot be substituted by a pronoun, for example:

```
Encontraron el coche [roto]
Encontraron el coche
*Encontraron el coche lo
```

The annotation we may follow here is the following one:

```
(PRED-COMPL
  (ADJ/NOUN/ ... )
 (...))
```

Examples:

"Eligieron a Tomás presidente"

```
(S
  (NP * SUBJ PL P3)
  (VP TENSED PAST IND PL P3
    (V "<Eligieron>" "elegir" TENSED PAST IND PL P3)
    (PP A OBJ1
      (PREP "<A>" "a")
      (NP ID-1
        (N "<Tomás>" "Tomás" PROPER))))
  (PRED-COMPL REF-1 MASC SG
    (N "<presidente>" "presidente" MASC SG))))
```

"Juan encontró el coche roto"

```
(S
```

```

(NP SUBJ SG P3
 (N "<Juan>" "Juan" PROPER))
(VP TENSED PAST IND P3 SG
 (V "<encontró>" "encontrar" TENSED PAST IND P3 SG)
 (NP OBJ1 ID-1
 (ART "<el>" "el" DEF MASC SG)
 (N "<coche>" "coche" MASC SG))
 (PRED-COMPL REF-1 MASC SG
 (ADJ "<roto>" "roto" MASC SG))))

```

As it can be seen from these examples, we are going to state the subject to which the predicative complement refers to, by adding the marks ID-1 and REF-1.

5.8 Criteria for the categorial status assignment

When annotating we have faced the problem of having several categories that do not fulfill their "canonical" functions. That is to say, we have found examples of infinitives that, despite being verbs, function as nouns. As there are quite some of these examples we will deal with them one by one, clarifying on each case their characteristics and the annotation they will have.

5.8.1 Infinitives

In Spanish, infinitives have been sometimes considered a mixed category composed of both, a nominal and a verbal part. Nowadays the trend is to consider them verbs with the peculiarity that they do not inflect for time, mood, person or number. Here we have divided them into:

Infinitives as Nouns

The infinitive is a verbal masculine noun; it is the noun of the verb. Some infinitives have been lexicalized permanently as nouns, for example "pesar", "haber", "deber", etc.; these infinitives admit plural "pesares", "haberes", "deberes" and they agree for gender and number with adjectives, for example, "alegre despartar", "deberes penosos", etc. Also, all infinitives can take definite and indefinite articles, demonstratives, possessives, etc.

Infinitives can function on the sentences in the same way nouns do: subject, attribute, complement of another noun, complement of an adjective and complement of a verb.

If we have an infinitive that presents these characteristics, we are going to consider it a noun, forgetting its status of verb.

Its analysis will be as follows:

”Se hizo un recuento de haberes”

(S IMPERSONAL
(VP TENSED PAST IND SG IMPERSONAL
(SE-MARK "<Se>" "se" IMPERSONAL)
(V "<hizo>" "hacer" TENSED PAST IND P3 SG)
(NP
(ART "<un>" "un" INDEF MASC SG)
(N "<recuento>" "recuento" MASC SG)
(PP DE
(PREP "<de>" "de")
(NP
(N "<haberes>" "haberes" MASC PL))))))

Infinitives as Verbs

The infinitive maintains its verbal quality although it cannot express mood, nor time, nor person by itself because it lacks inflection. Having verbal features, its characteristics are:

1. It can take the passive form.
2. The simple infinitive expresses imperfect action and complex perfect action (NOTE: imperfect and perfect in the sense of not finished and finished).
3. It can take clitics.
4. It can appear in constructions with adverbs.
5. It can have explicit subject, ”Querer Juan el puesto y conseguirlo fue todo uno”

Infinitives can be direct or circumstantial complements of a verb. In these cases they have a subordinate use, and they will be treated as constructions of control.

1. Direct Object. With perception and wish verbs, the infinitive and its subject form a clause that functions as the direct object of the main verb. If the subject of the infinitive is a proper noun, it will take the preposition ^a to enter the construction, for example, ”Mandaron volver a Juan”. If the subject of the infinitive is pronominal, we will use the corresponding clitic, for example ”Te vi correr”.
2. Oblique Object. For example, ”Se contentaría con recibir una carta cada mes”.

3. With some prepositions the infinitives form fixed phrases with special meanings that are equivalent to circumstantial subordinate clauses. For example:

- (a) Al + infinitive: TIME clause
- (b) A/de + infinitive: CONDITIONAL clause
- (c) Con/para/por + infinitive: MODAL clause

In the cases where we find infinitives with these characteristics, they will be treated as verbs, head of clauses. Its annotation will be as follows:

”Oigo sonar las campanas”

```
(S
(NP * SUBJ SG P1)
(VP TENSED PRES IND SG P1
(V "<oigo>" "oir" TENSED PRES IND SG P1)
(CL INFINITIVE OBJ1
(VP UNTENSED INFINITE
(V "<sonar>" "sonar" UNTENSED INFINITE)
(NP SUBJ FEM PL
(ART "<las>" "la" DEF FEM PL)
(N "<campanas>" "campanas" FEM PL))))))
```

5.8.2 Gerunds

The most general meaning and uses of gerunds are:

1. Simple gerund expresses a durative and imperfect action.
2. Complex gerund expresses a perfect action, previous to that of the main verb.
3. The only preposition that can go with them is "en".
4. They admit clitics.

Gerunds can function as adverbs. In this way its major function is that of modifying the verb as a modal adverb. Also, it has been said that gerunds can function as adjectives, for example in "decreto nombrando gobernador". Here we are going to ignore this traditional classification, and we will treat all gerunds as verbs that function as heads of gerund clauses. Their annotation then will be:

”Encontré a tu padre escribiendo”

```

(S
  (NP * SUBJ SG P1)
  (PP A OBJ1
    (PREP "<A>" "a")
    (NP
      (POSS "<tu>" "tuyo" MASC SG P2)
      (N "<padre>" "padre" ID-1 MASC SG)))
  (CL GERUND MODAL
    (NP * SUBJ REF-1)
    (VP UNTENSED GERUND
      (V "<escribiendo>" "escribir" UNTENSED GERUND))))

```

5.8.3 Participles

Participles, as it happened with infinitives and gerunds, are untensed verbs, but, in this case, they agree with the noun they go with on gender and number.

The so-called passive participle, when functioning as a predicative complement, agrees with the subject. It appears with intransitive, transitive and pronominal verbs. (NOTE that when the participle of the transitive verbs is used as a predicate with the auxiliary verb "ser" it forms the passive voice).

The passive participle can also function as a predicative complement of a direct object. It appears with transitive and pronominal verbs. When it occurs with a transitive verb it has a passive meaning, but when it occurs with intransitive or pronominal verbs its meaning is no more passive.

When the participle functions as an attribute to a noun, what we have is a clause that can be of two types, depending on the insertion or not of the noun they refer to on the clause. In this way, we have the so-called "participio conjunto" where the noun is not included in the clause, and the "participio absoluto" where the noun is actually included.

Our annotation regarding participles will differentiate between those participles that work as predicative complements or "participio conjunto" that may be treated as adjectives, and those participles functioning as "participio absoluto" that may be treated as verbs forming a clause. The participles working as adjectives will be annotated:

"Hojas del árbol caídas"

```

(NP
  (N "<hojas>" "hoja" FEM PL)
  (PP DEL
    (PREP "<del>" "de")
    (NP

```

```
(ART "el" DEF MASC SG)
(N "<árbol>" "árbol" MASC SG)
(ADJP FEM PL
  (ADJ "<caídas>" "caído" FEM PL))))))
```

and those participles considered verbs will be annotated as clauses:

”Muerto el perro, se acabó la rabia”

```
(S IMPERSONAL
  (CL ABS-PART TIME
    (VP UNTENSED PART MASC SG
      (V "<Muerto>" "morir" UNTENSED PART MASC SG)
      (NP SUBJ MASC P3 SG
        (ART "<el>" "el" DEF MASC SG)
        (N "<perro>" "perro" MASC SG))))))
  (PUNCT ", " COMMA)
  (VP TENSED PAST IND SG IMPERSONAL
    (SE-MARK "<se>" "se" IMPERSONAL)
    (V "<acabó>" "acabar" TENSED PAST IND P3 SG)
    (NP OBJ1
      (ART "<la>" "la" DEF FEM SG)
      (N "<rabia>" "rabia" FEM SG))))))
```

5.8.4 Subordinate clauses as Nouns and Adjectives

Adjectival subordinate clauses are also called relative clauses because they are introduced in the main clause by the relation between the pronouns and their antecedents. Relative pronouns on the subordinate clauses have all the syntactic functions a pronoun can have, and they can also cooccur with prepositions if necessary. If the relative ”que” appears with the definite article and with a preposition, we can have two possibilities:

1. They are relatives with the antecedent included on their own meaning (relatives without expressed antecedent). For example, ”Los que nada saben, todo lo creen saber” which can be rephrased as ”Aquellos que nada saben...”
2. The article ”el” has its syntactic value and the antecedent of the relative is expressed in another part of the sentence. For example, ”La casa en la que pasé los días de mi infancia subsiste todavía”.

Here we are going to treat them as a compound, that is to say, we will annotate them as $\bar{j}el\ que_i$; and their categorial status will depend on the function they have: if they function as an adjective (i.e. clause with expressed

antecedent) they will be treated as adjectives, and if they function as a noun (i.e. clause without expressed antecedent) they will be considered so. An example of the second case::

”La casa en la que pasé los días de mi infancia subsiste todavía”.

```
(S
(NP SUBJ SG P3
(ART "<La>" "la" DEF FEM SG)
(N "<casa>" "casa" FEM SG)
(PP EN LOCATIVE
(PREP "<en>" "en")
(CL RELATIVE
(NP
(P "<la que>" "la_que" REL DEF FEM SG))
(NP * SUBJ SG P1)
(VP TENSED PAST IND SG P1
(V "<pasé>" "pasar" TENSED PAST IND SG P1)
(NP OBJ1
(ART "<los>" "el" DEF MASC PL)
(N "<días>" "día" MASC PL)
(PP DE
(PREP "<de>" "de")
(NP
(POSS "<mí>" "mío" FEM P1 SG)
(N "<infancia>" "infancia" FEM SG))))))))))
(VP TENSED PRES IND P3 SG
(V "<subsiste>" "subsistir" TENSED PRES IND P3 SG)
(ADVP TIME
(ADV "<todavía>" "todavía" TIME))))
```

5.9 Grammaticalisation and Lexicalisation

Grammaticalisation is a process by which two or more words which cooccur together are felt to be one with the same syntactic behaviour, despite the fact that, at the beginning of the process, both words could have belonged to different categories. This is a very slow process.

An example of grammaticalisation can be the use of an adjective as an adverb, as in ”Lo hizo muy rápido” or ”Sucedió muy rápido”. Another example of grammaticalisation, this time less obvious, will be the use of ^{al} as a temporal conjunction, as in ^{al} salir el sol” where ^{al} is not the amalgam of the preposition A plus the definite article EL, but rather, a temporal conjunction (^{al} salir el sol” = ”cuando el sol salió”). Example:

“El fabricante de esta muñeca, Mattel, revolucionó el mercado al crear una muñeca que semejaba una mujer joven ... y sexy.”

(S

(NP SUBJ ID-1 MASC SG P3
(ART "<El>" "el" DEF MASC SG)
(N "<fabricante>" "fabricante" MASC SG)
(PP DE
(PREP "<de>" "de")
(NP
(DEM "<esta>" "este" FEM SG)
(N "<muñeca>" "muñeca" FEM SG)
(PUNCT "," COMMA)
(NP APPOS REF-1
(N "<Mattel>" "Mattel" PROPER))))
(PUNCT "," COMMA))
(VP TENSED PAST IND SG P3
(V "<revolucionó>" "revolucionar" TENSED PAST IND SG P3)
(NP OBJ1
(ART "<el>" "el" DEF MASC SG)
(N "<mercado>" "mercado" MASC SG)
(CL TIME
(C "<a1>" "a1")
(CL INFINITIVE
(NP * SUBJ REF-1)
(VP UNTENSED INFINITE
(V "<crear>" "crear" UNTENSED INFINITE)
(NP OBJ1
(ART "<una>" "un" INDEF FEM SG)
(N "<muñeca>" "muñeca" FEM SG)
(CL RELATIVE
(NP SUBJ SG P3
(P "<que>" "que" REL P3))
(VP TENSED IMPERFECT IND SG P3
(V "<semejaba>" "semejar" TENSED IMPERFECT IND SG

P3)

(NP OBJ1
(ART "<una>" "un" INDEF FEM SG)
(N "<mujer>" "mujer" FEM SG)
(ADJP FEM SG
(ADJ COORDINATED
(ADJ "<joven>" "joven" FEM SG)
(PUNCT "... " ELLIPSIS)
(C "<y>" "y" COORDINATING)

(ADJ "<sexy>" "sexy" FEM SG))))))))))
(PUNCT "." PERIOD))

We have decided to include here the relative pronoun "como", which can be found functioning as a preposition, a conjunction or an adverb, despite its canonical categorial status of pronoun. Examples:

"Muchas de estas grabaciones sirven además como complemento didáctico en el reconocimiento de especies"

(S
(NP SUBJ FEM PL P3
(P "<Muchas>" "mucho" QU FEM PL)
(PP DE
(PREP "<de>" "de")
(NP
(DEM "<estas>" "este" FEM PL)
(N "<grabaciones>" "grabación" FEM PL))))
(VP TENSED PRES IND PL P3
(V "<sirven>" "servir" TENSED PRES IND PL P3)
(ADVP NORMAL
(ADV "<además>" "además" NORMAL))
(PP COMO OBL
(PREP "<como>" "como")
(NP
(N "<complemento>" "complemento" MASC SG)
(ADJP MASC SG
(ADJ "<didáctico>" "didáctico" MASC SG))))
(PP EN
(PREP "<en>" "en")
(NP
(ART "<el>" "el" DEF MASC SG)
(N "<reconocimiento>" "reconocimiento" MASC SG)
(PP DE
(PREP "<de>" "de")
(NP
(N "<especies>" "especie" FEM PL))))))
(PUNCT "." PERIOD))

"Los médicos se muestran reticentes a abrir plazos, tanto para la vida normal como para su actividad profesional."

```

(S
(NP SUBJ ID-2 MASC PL P3
(ART "<Los>" "el" DEF MASC PL)
(N "<médicos>" "médico" MASC PL))
(VP TENSED PRES IND PL P3
(NP
(P "<se>" "se" PERS P3 DISCONTINUOUS REF-1))
(V "<muestran>" "mostrarse" TENSED PRES IND PL P3 #CLITIC
DISCONTINUOUS ID-1)
(NP ATTR MASC PL
(N "<reticentes>" "reticente" MASC PL)
(PP A
(PREP "<a>" "a")
(CL INFINITIVE
(NP * SUBJ REF-2)
(VP UNTENSED INFINITE
(V "<abrir>" "abrir" UNTENSED INFINITE)
(NP OBJ1
(N "<plazos>" "plazo" MASC PL)
(CL COMPARATIVE
(CL COMPARATIVE-1
(PUNCT "," COMMA)
(C "<tanto>" "tanto")
(PP PARA
(PREP "<para>" "para")
(NP
(ART "<la>" "el" DEF FEM SG)
(N "<vida>" "vida" FEM SG)
(ADJP FEM SG
(ADJ "<normal>" "normal" FEM SG))))))
(CL COMPARATIVE-2
(C "<como>" "como")
(PP PARA
(PREP "<para>" "para")
(NP
(POSS "<su>" "suyo" P3 SG)
(N "<actividad>" "actividad" FEM SG)
(ADJP FEM SG
(ADJ "<profesional>" "profesional" FEM SG))))))))))
(PUNCT "." PERIOD))

```

"Conde profirió ayer un grito al ser preguntado por la composición del accionariado que acompañaba a Banesto en la operación de la Esquina del

Bernabéu, calificada como presunta estafa"

(S
(NP SUBJ MASC SG P3
(N "<Conde>" "Conde" PROPER))
(VP TENSED PAST IND SG P3
(V "<profirió>" "proferir" TENSED PAST IND SG P3)
(ADVP TIME
(ADV "<ayer>" "ayer" TIME))
(NP OBJ1
(ART "<un>" "un" INDEF MASC SG)
(N "<grito>" "grito" MASC SG)))
(CL TIME
(C "<al>" "al")
(CL INFINITIVE
(VP UNTENSED INFINITIVE PASSIVE
(V "<ser preguntado>" "preguntar" UNTENSED PASSIVE
(AUX "<ser>" "ser" UNTENSED INFINITE)
(V "<preguntado>" "preguntar" UNTENSED PART MASC SG))
(PP POR
(PREP "<por>" "por")
(NP
(ART "<la>" "el" DEF FEM SG)
(N "<composición>" "composición" FEM SG)
(PP DE
(PREP "" "de")
(NP ID-1
(ART "el" DEF MASC SG)
(N "<accionariado>" "accionariado" MASC SG)
(CL RELATIVE
(NP SUBJ REF-1 SG P3
(P "<que>" "que" REL P3))
(VP TENSED IMPERFECT IND SG P3
(V "<acompañaba>" "acompañar" TENSED IMPERFECT
IND SG P3)
(PP A OBL
(PREP "<a>" "a")
(NP
(N "<Banesto>" "Banesto" PROPER)))
(PP EN
(PREP "<en>" "en")
(NP ID-2
(ART "<la>" "el" DEF FEM SG)

```

(N "<operación>" "operación" FEM SG)
(PP DE
  (PREP "<de>" "de")
  (NP
    (N "<La Esquina del Bernabéu>" "la_Esquina_del_Bernabéu"
      PROPER)
    (PUNCT "," COMMA)
    (PRED-COMPL REF-2 FEM SG
      (V "<calificada>" "calificar" UNTENSED
        PART FEM SG)
      (NP OBJ1
        (ADJP FEM SG
          (ADVP MODAL
            (ADV "<como>" "como" MODAL))
            (ADJ "<presunta>" "presunto" FEM SG))
          (N "<estafa>" "estafa" FEM SG)))))))))))))
(PUNCT "." PERIOD))

```

NOTE: Other examples of grammaticalisation can be found in some of the connectives that appear in Section 5.15.1., and also, in some of the multiword units that appear in Appendix C.

Lexicalisation, on the other hand, is a process by which two or more words that at the beginning were felt to have separate meanings, end up having a unique meaning. We may find several examples such as "coches de alquiler" "reloj de pulsera" "con dos palmos de narices", etc. Whenever we find one of these lexicalized items, we may treat them as compound words in the following way:

```

(NP
  (N «coches de alquiler» "coche_de_alquiler" MASC SG))

(NP
  (N «reloj de pulsera» "reloj_de_pulsera" MASC SG))

```

5.10 Conditional, Comparative and Coordinated Structures

All these structures have in common the characteristic that they are one constituent built up from two smaller parts. For example, if we have a comparative sentence, it has, at the same time two daughters: one where the comparison is established, and one where the compared element appears. These constructions will receive a similar treatment.

5.10.1 Conditionals

The structure where a condition is established is called CONDITIONAL CLAUSE and it is composed of: CONDITION and MAIN clauses as in the following example:

“Los jugadores del Fórum amenazan con no jugar el domingo si no les pagan”.

```
(S
  (CL CONDITIONAL
    (CL MAIN
      (NP SUBJ MASC PL P3
        (ART "<Los>" "el" DEF MASC PL)
        (N "<jugadores>" "jugador" MASC PL)
        (PP DE
          (PREP "<del>" "de")
          (NP
            (ART "el" DEF MASC SG)
            (N "<Fórum>" "Fórum" PROPER))))))
      (VP TENSED PRES IND PL P3
        (V "<amenazan>" "amenazar" TENSED PRES IND PL P3)
        (PP CON OBL
          (PREP "<con>" "con")
          (CL INFINITIVE
            (VP UNTENSED INFINITE
              (ADVP NEG
                (ADV "<no>" "no" NEG))
              (V "<jugar>" "jugar" UNTENSED INFINITE)
              (NP TIME
                (ART "<el>" "el" DEF MASC SG)
                (N "<domingo>" "domingo" MASC SG))))))))
      (CL CONDITION
        (C "<si>" "si" SUBORDINATING)
        (NP * SUBJ PL P3)
        (VP TENSED PRES IND PL P3
          (ADVP NEG
            (ADV "<no>" "no" NEG))
          (NP OBJ2
            (P "<les>" "él" PERS P3 DAT PL))
          (V "<pagan>" "pagar" TENSED PRES IND PL P3))))
      (PUNCT "." PERIOD))
```

5.10.2 Comparisons

Comparative clauses are composed of two other clauses named COMPARATIVE-1 and COMPARATIVE-2, as in:

“La sorpresa es que venden más copias que cuando salieron en vinilo”.

```
(S
  (NP SUBJ FEM SG P3
    (ART "<la>" "el" DEF FEM SG)
    (N "<sorpresa>" "sorpresa" FEM SG))
  (VP TENSED PRES IND SG P3
    (V "<es>" "ser" TENSED PRES IND SG P3)
    (CL COMPLETIVE
      (C "<que>" "que" SUBORDINATING FINITE)
      (NP * SUBJ PL P3)
      (VP TENSED PRES IND PL P3
        (V "<venden>" "vender" TENSED PRES IND PL P3)
        (CL COMPARATIVE OBJ1
          (CL COMPARATIVE-1
            (C "<más>" "más" COMPARATIVE)
            (N "<copias>" "copia" FEM PL))
          (CL COMPARATIVE-2
            (C "<que>" "que" SUBORDINATING FINITE)
            (CL TIME
              (C "<cuando>" "cuando" TIME)
              (NP * SUBJ PL P3)
              (VP TENSED PAST IND PL P3
                (V "<salieron>" "salir" TENSED PAST IND PL P3)
                (PP EN
                  (PREP "<en>" "en")
                  (NP
                    (N "<vinilo>" "vinilo" MASC SG))))))))))
  (PUNCT "." PERIOD))
```

NOTE: When a comparison is established in lower levels, that is, when phrases and not clauses are compared, the treatment will be the same but within the scope of the correspondent category. For example:

“Desde luego hay que ser un gran aficionado a estos vinos para pagar por ellos una media de 1.270 ptas (2 veces más que los rioja, por ejemplo)”.

```
(S IMPERSONAL
  (CON
```

(CON "<Desde luego>" "desde_luego")
(VP TENSED PRES IND MODAL SG IMPERSONAL
(V "<hay que ser>" "ser" TENSED PRES IND MODAL SG IMPERSONAL
(AUX "<hay que>" "haber_que" TENSED PRES IND SG IMPERSONAL)
(V "<ser>" "ser" UNTENSED INFINITE))
(NP ATTR MASC SG
(ART "<un>" "un" INDEF MASC SG)
(ADJP MASC SG
(ADJ "<gran>" "grande" MASC SG))
(N "<aficionado>" "aficionado" MASC SG)
(PP A OBL
(PREP "<a>" "a")
(NP
(DEM "<estos>" "este" MASC PL)
(N "<vinos>" "vino" MASC PL))))
(PP PARA
(PREP "<para>" "para")
(CL INFINITIVE
(NP * SUBJ)
(VP UNTENSED INFINITE
(V "<pagar>" "pagar" UNTENSED INFINITE)
(PP POR OBL
(PREP "<por>" "por")
(NP
(P "<ellos>" "él" PERS P3 MASC PL))))
(NP OBJ2 ID-1
(ART "<una>" "un" INDEF FEM SG)
(N "<media>" "media" FEM SG)
(PP DE
(PREP "<de>" "de")
(NP
(QP
(Q "<1.270>" "mil_doscientas_setenta"))
(N "<ptas>" "peseta" MEASURE))))
(NP APPOS REF-1
(PUNCT "(" PARENTHESIS)
(NP COMPARATIVE
(NP COMPARATIVE-1
(QP
(Q "<2>" "dos"))
(N "<veces>" "vez" FEM PL))
(C "<más que>" "más_que" COMPARATIVE)
(NP COMPARATIVE-2
(ART "<los>" "el" DEF MASC PL)

```

      (N "<rioja>" "rioja" MASC SG)))
(PUNCT "," COMMA)
(PP POR
  (PREP "<por>" "por")
  (NP
    (N "<ejemplo>" "ejemplo" MASC SG)))
(PUNCT ")" PARENTHESIS))))))
(PUNCT "." PERIOD))

```

But, if the two terms of the comparison cannot be established because of ellision, gapping or whatever, the COMPARATIVE feature will appear alone, as in:

“Quienes no quieran gastarse más de mil pesetas en un tinto crianza con esta DO tienen pocas opciones: sólo 9 marcas”.

```

(S
  (CL RELATIVE SUBJ ID-1 PL P3
    (NP SUBJ PL P3
      (P "<Quienes>" "quien" REL HUMAN P3 PL))
    (VP TENSED PRES SUBJUNCTIVE PL P3
      (ADVP NEG
        (ADV "<no>" "no" NEG))
      (V "<quieran>" "querer" TENSED PRES SUBJUNCTIVE PL P3)
      (CL INFINITIVE OBJ1
        (NP * SUBJ REF-1)
        (VP UNTENSED INFINITE
          (V "<gastarse>" "gastar" UNTENSED INFINITE #CLITIC)
          (NP COMPARATIVE
            (C "<más de>" "más_de")
            (QP
              (Q "<mil>" "mil"))
              (N "<pesetas>" "peseta" MEASURE)
            (PP EN
              (PREP "<en>" "en")
              (NP
                (ART "<un>" "un" INDEF MASC SG)
                (N "<tinto>" "tinto" MASC SG)
              (NP
                (N "<crianza>" "crianza" FEM SG))
              (PP CON
                (PREP "<con>" "con")
                (NP
                  (DEM "<esta>" "este" FEM SG)

```

```

(N "<DO>" "do" MASC SG)))))))))
(VP TENSED PRES IND PL P3
(V "<tienen>" "tener" TENSED PRES IND PL P3)
(NP OBJ1 ID-2
(QP
(Q "<pocas>" "poco))
(N "<opciones>" "opción" FEM PL)
(NP APPOS REF-2
(PUNCT ":" COLONS)
(NP
(QP
(ADVP NORMAL
(ADV "<sólo>" "sólo" NORMAL))
(Q "<9>" "nueve))
(N "<marcas>" "marca" FEM PL))))))
(PUNCT "." PERIOD))

```

5.10.3 Coordinations

Coordinated clauses can also be thought of as one superconstituent that consists of two smaller ones. In these cases, we are going to name the "superconstituent" category under COORDINATED and the two "subconstituents" will be left unnamed. For example:

"Fischler pronosticó ayer que las negociaciones para la reforma de la OMC del aceite de oliva serán largas y que las actuales condiciones podrían revisarse en dos años."

```

(S
(NP SUBJ SG P3
(N "<Fischler>" "Fischler" PROPER))
(VP TENSED PAST IND SG P3
(V "<pronosticó>" "pronosticar" TENSED PAST IND SG P3)
(ADVP TIME
(ADV "<ayer>" "ayer" TIME))
(CL COMPLETIVE OBJ1 COORDINATED
(CL COMPLETIVE
(C "<que>" "que" SUBORDINATING FINITE)
(NP SUBJ FEM PL P3
(ART "<las>" "el" DEF FEM PL)
(N "<negociaciones>" "negociación" FEM PL)
(PP PARA
(PREP "<para>" "para")
(NP

```

```

(ART "<la>" "el" DEF FEM SG)
(N "<reforma>" "reforma" FEM SG)
(PP DE
  (PREP "<de>" "de")
  (NP
    (ART "<la>" "el" DEF FEM SG)
    (N "<OMC>" "OMC" PROPER)
    (PP DE
      (PREP "<del>" "de")
      (NP
        (ART "el" DEF MASC SG)
        (N "<aceite>" "aceite" MASC SG)
        (PP DE
          (PREP "<de>" "de")
          (NP
            (N "<oliva>" "oliva" FEM SG))))))))))
(VP TENSED FUT IND PL P3
  (V "<serán>" "ser" TENSED FUT IND PL P3)
  (ADJP ATTR FEM PL
    (ADJ "<largas>" "largo" FEM PL)))
(C "<y>" "y" COORDINATING)
(CL COMPLETIVE
  (C "<que>" "que" SUBORDINATING FINITE)
  (NP SUBJ FEM PL P3
    (ART "<las>" "el" DEF FEM PL)
    (ADJP FEM PL
      (ADJ "<actuales>" "actual" FEM PL))
      (N "<condiciones>" "condición" FEM PL))
  (VP TENSED COND IND MODAL PL P3
    (V "<podrían revisarse>" "revisarse" TENSED COND IND MODAL
      PL P3
        (AUX "<podrían>" "poder" TENSED COND IND PL P3)
        (V "<revisarse>" "revisar" UNTENSED INFINITE #CLITIC))
    (PP EN TIME
      (PREP "<en>" "en")
      (NP
        (QP
          (Q "<dos>" "dos"))
          (N "<años>" "año" MASC PL))))))
(PUNCT "." PERIOD))

```

Coordination within phrases is also possible, so the category that consists of two coordinated phrases or nucleus will be also added the feature coordinated as in:

“Las palabras de Chirac suscitaron reacciones de “agradecimiento” entre los líderes de la derecha y una actitud de mayor reserva en los partidos de izquierda”.

```
(S
  (NP SUBJ P3 PL
    (ART "<Las>" "el" DEF FEM PL)
    (N "<palabras>" "palabra" FEM PL)
    (PP DE
      (PREP "<de>" "de")
      (NP
        (N "<Chirac>" "Chirac" PROPER))))))
  (VP TENSED PAST IND PL P3
    (V "<suscitaron>" "suscitar" TENSED PAST IND PL P3)
    (NP OBJ1 COORDINATED
      (NP OBJ1
        (N "<reacciones>" "reacción" FEM PL)
        (PP DE
          (PREP "<de>" "de")
          (NP
            (PUNCT ">" QUOTE)
            (N "<agradecimiento>" "agradecimiento" MASC SG)
            (PUNCT "<" QUOTE)))
          (PP ENTRE LOCATIVE
            (PREP "<entre>" "entre")
            (NP
              (ART "<los>" "el" DEF MASC PL)
              (N "<líderes>" "líder" MASC PL)
              (PP DE
                (PREP "<de>" "de")
                (NP
                  (ART "<la>" "el" DEF FEM SG)
                  (N "<derecha>" "derecha" FEM SG))))))))))
    (C "<y>" "y" COORDINATING)
    (NP OBJ1
      (ART "<una>" "un" INDEF FEM SG)
      (N "<actitud>" "actitud" FEM SG)
      (PP DE
        (PREP "<de>" "de")
        (NP
          (ADJP COMPARATIVE FEM SG
            (ADJ "<mayor>" "grande" COMPARATIVE FEM SG))
          (N "<reserva>" "reserva" FEM SG))
        (PP EN LOCATIVE
```

```

(PREP "<en>" "en")
(NP
  (ART "<los>" "el" DEF MASC PL)
  (N "<partidos>" "partido" MASC PL)
  (PP DE
    (PREP "<de>" "de")
    (NP
      (N "<izquierda>" "izquierda" FEM SG)))))))))
(PUNCT "." PERIOD))

```

Example:

“Al parecer, Sagrario Larralde y Rosa Muñoz, de la orden Hermanas de la Caridad de Santa Ana, fueron secuestradas con la intención de que curasen y prestaran ayuda a combatientes hutus heridos.”

```

(S
  (CON
    (CON "<Al parecer>" "al_parecer")
    (PUNCT "," COMMA))
  (NP SUBJ ID-1 PL P3 COORDINATED
    (NP SUBJ SG P3
      (N "<Sagrario Larralde>" "Sagrario_Larralde" PROPER))
    (C "<y>" "y" COORDINATING)
    (NP SUBJ SG P3
      (N "<Rosa Muñoz>" "Rosa_Muñoz" PROPER)))
    (PUNCT "," COMMA)
    (PP DE APPOS REF-1
      (PREP "<de>" "de")
      (NP
        (NP FEM SG
          (ART "<la>" "el" DEF FEM SG)
          (N "<orden>" "orden" FEM SG))
        (N "<Hermanas de la Caridad de Santa Ana>" "hermana_de_la_caridad_de_Santa_Ana"
          FEM PL)))
      (PUNCT "," COMMA))
    (VP TENSED PAST IND PASSIVE PL P3
      (V "<fueron secuestradas>" "secuestrar" TENSED PAST IND PASSIVE
        PL P3
          (AUX "<fueron>" "ser" TENSED PAST IND PL P3)
          (V "<secuestradas>" "secuestrar" UNTENSED PAST PART FEM PL))
      (PP CON OBL
        (PREP "<con>" "con")
        (NP
          (ART "<la>" "el" DEF FEM SG)

```

```

(N "<intención>" "intención" FEM SG)
(PP DE
  (PREP "<de>" "de")
  (CL
    (C "<que>" "que" SUBORDINATING FINITE)
    (NP * P3 PL)
    (VP TENSED IMPERFECT SUBJUNCTIVE PL P3
      (V COORDINATED
        (V "<curasen>" "curar" TENSED IMPERFECT SUBJUNCTIVE
          PL P3)
          (C "<y>" "y" COORDINATING)
          (V "<prestaran>" "prestar" TENSED IMPERFECT SUBJUNCTIVE
            PL P3))
        (NP OBJ1
          (N "<ayuda>" "ayuda" FEM SG))
          (PP A OBJ2
            (PREP "<a>" "a")
            (NP
              (N "<combatientes>" "combatiente" MASC PL)
              (ADJP MASC PL
                (ADJ "<hutus>" "hutu" MASC PL))
              (ADJP MASC PL
                (ADJ "<heridos>" "herido" MASC PL))))))
          (PUNCT "." PERIOD))

```

NOTE: following this strategy, we may find coordinated verb phrases, for example, that may have different tenses. What we have decided here is to include the shared features by both verbs which are not contradictory. This also happens with coordinated NNPP, when for example one of the coordinated NNPP has a FEM feature and the other a MASC feature. In these cases what we do is not to include the gender feature in the coordinated NP. For example:

```

(NP SUBJ PL P3 COORDINATED
  (NP SUBJ FEM SG P3
    (N "<María>" "María" PROPER))
  (C "<y>" "y" COORDINATING)
  (NP SUBJ MASC SG P3
    (N "<Pedro>" "Pedro" PROPER)))

```

Notice also that the noun phrase coordinated has the feature PL since we have two lower noun phrases that are SG and we sum up their features.

We may also have cases where two coordinated categories do not share all the features. In these cases, what we have decided is to eliminate from the

higher phrase (the one with the feature COORDINATED) those features that are not the same in both conjoints. Finally, there is a last remark about coordination: when what we have is a coordinated sentence, that is, a sentences made of two or mor clauses with different subjects and different predicates, we have decided to attact the feature COORDINATED to the S node, and the features COORDINATED-1, COORDINATED-2, etc., to each of the clauses that form the given sentence. For example:

```
(S COORDINATED
  (CL COORDINATED-1
    (NP SUBJ MASC PL P3
      (ART "<Los>" "e1" DEF MASC PL)
      (N "<azulgrana>" "azulgrana" MASC SG))
    (VP TENSED FUT IND PL P3
      (ADVP NORMAL
        (ADV "<sólo>" "sólo" NORMAL))
      (V "<celebrarán>" "celebrar" TENSED FUT IND PL P3)
      (NP OBJ1
        (ART "<un>" "un" INDEF MASC SG)
        (N "<entrenamiento>" "entrenamiento" MASC SG)
        (PP A
          (PREP "<a1>" "a")
          (NP
            (ART "e1" DEF MASC SG)
            (N "<completo>" "completo" MASC SG))))))
    (PUNCT ", " COMMA))
  (C "<y>" "y" COORDINATING)
  (CL COORDINATED-1
    (NP SUBJ MASC PL P3
      (ART "<los>" "e1" DEF MASC PL)
      (N "<blancos>" "blanco" MASC PL))
    (VP TENSED PAST IND SG P3
      (NP
        (P "<se>" "se" PERS P3 DISCONTINUOUS REF-1))
      (V "<prepararon>" "prepararse" TENSED PAST IND SG P3 #CLITIC
DISCONTINUOUS ID-1)
      (ADVP TIME
        (ADV "<ayer>" "ayer" TIME))
      (PP SIN
        (PREP "<sin>" "sin")
        (NP
          (QP
            (Q "<cinco>" "cinco"))
```

```
(N "<titulares>" "titular" MASC PL))))  
(PUNCT "." PERIOD))
```

5.11 Special categories and features

5.11.1 Connectives

As we said in the introduction, we have tried a fifty-fifty division in the 200 first sentences intended for testing annotation within a pure syntactic environment and annotation with a discourse perspective. Within the annotation with discourse perspective, we have to deal with those elements that join sentences, giving the texts coherence and cohesion. We are going to name this elements connectives in a general sense, and we will not make further distinctions as to the type they belong to, their pragmatic content, etc., since we are not interested in analysing them deeply, but only in their influence upon syntax.

A connective can be defined as the unit that joins an utterance or sentence to a previous element, both implicitly or accesible from the context. The meaning of the connective, provides us with a number of intructions that guide the inferences to be reached from the two related members. In this way, more particular contexts, that will not be evident, can be easily obtained. Also, some inferences may be reinforced or eliminated.

Therefore, two main factors can be distinguished within connectives:

- a purely grammatical one: the instructions their meaning provide;
- a pragmatic one: the inferential processes that unchain on their relation with the context.

Connectives are linguistic units that, due to their meaning, join two members, but that do not reflect a syntactic relation. Not all of them are conjunctions, nor do they have the syntactic function of link.

So, connectives are linguistic means that allow the cohesion of textual units. Cohesion is not an end in the use of connectives, but a consequence. Speakers use them for the hearer to reach the pertinent inferences of the sentences. The text is not a grammatical unit as the syllable, the sentence or the noun phrase are, but the intention of the speakers. So then, coherence is not a criterion for good formation equal to syntactic grammaticallity. The text and the sentence belong to different environments.

There are several elements that can be considered connectives, for example, coordinating and subordinating conjunctions, adverbs and adverbials, prepositional phrases, verb phrases, etc. Here we will not make any consideration as to which category does a connective belong to, since we will treat equally a VP or a PP if they are functioning as connvectives. Their annotation will follow this pattern:

(CON
 (CON "< >" " "))

Also, as we said above, we will not distinguish any type within the scope of connectives, so they may appear all together under the same head CON. Some of the connectors we can find in Spanish are the following ones:

a decir verdad	de aquí
a este respecto	de cualquiera de las formas
a fin de cuentas	de forma/manera/modo/suerte que
a lo mejor	de hecho
a mi modo de ver	de igual manera
a mi parecer	de igual modo
a modo de resumen	de la misma manera
a propósito	de otra parte
a saber	de otro lado
a su vez	de otro modo
a todo esto	de todos modos/formas/maneras
aclaro	de un lado
además	de veras
ahora	de verdad
ahora bien	del mismo modo
al contrario	desde luego
al fin	desde mi punto de vista
al fin y al cabo	después
al menos	después de todo
al parecer	dicho de forma más simple
ante todo	dicho de otra manera
antes al contrario	dicho de otro modo
antes bien	dicho en dos palabras
antes por el contrario	dicho en otras palabras
aparte (de)	dicho en otros términos
asimismo (o así mismo)	dicho entre paréntesis
así	dicho sea de paso
así pues	digamos
así que	digo
así y todo	digámoslo así
aun	dos (tres, cuatro...)
aun así	efectivamente
aun con todo	empero
aunque	en breves palabras
bien pensado	en cambio
claro (está)	en caso contrario
como es natural	en conclusión
como hemos visto	en concreto
con otras palabras	en conjunto
con toda seguridad	en consecuencia
con todo (y con eso)	en contraste
conque	en cualquier caso
contrariamente	en definitiva
de ahí	en dos palabras

en efecto	ni
en el fondo	no obstante
en ese caso	o
en fin	o sea
en mi opinión	o séase
en otras palabras	ojalá
en otros términos	opuestamente
en pocas palabras	para empezar
en primer lugar	para terminar
en principio	pero
en realidad	por Dios
en resumen	por así decir
en segundo (tercer) lugar	por cierto
en sentido amplio	por consiguiente
en suma	por contra
en síntesis	por descontado
en tal caso	por desgracia
en todo caso	por el contrario
en un primer momento	por ende
en una palabra	por esa razón
en verdad	por eso (mismo)
en último lugar	por fin
entonces	por fortuna
entre paréntesis	por lo demás
es decir	por lo general
es que	por lo menos
esto es	por lo pronto
gracias a Dios	por lo que parece
grosso modo	por (lo) tanto
hasta	por lo visto
hombre	por mejor decir
igualmente	por otra parte
incluso	por otro lado
luego	por suerte
mas	por supuesto
mejor dicho	por un lado
menos mal (que)	por una parte
miento	por último
mientras tanto	porque
muy al contrario	precisamente
muy por el contrario	precisamente por eso
más aún	primer X
más bien	primeramente
más fácil	primero

pues
pues bien
puesto que
quiero decir
quizá(s)
sea como sea
segundo (tercer) X
segundo (tercero)
según (tu, ella...)
según parece
según se ve
será que
si bien
sin duda (alguna)
sin embargo
sin lugar a dudas
sino
sobre todo (porque)
tal vez
también
tampoco
todavía más
uno
vale decir
vamos
y
ya que

NOTE: The information about connectors and the list that appear in this section have been taken from: Portolés, J. (1993), "La distinción entre los conectores y otros marcadores del discurso en español", VERBA 20, pp. 141-170.

5.11.2 Idiomatic Expressions and Foreign Words

In Spanish there are certain expressions that are considered to be a "block", for example, expressions such as "nada más y nada menos que". These are felt by native speakers as a compound comparative adverb.

In these cases, we have opted for analysing these expressions within the correspondant category they may fall in (that is, ADVERB, CONJUNCTION, etc.), and we have decided to add the feature IDIOM.

Example:

"Sagliocco mantenía, según ha descubierto ahora el fisco italiano, nada más y nada menos que 344 apartamentos alquilados "en negro", es decir, sin contrato y sin pasar, por lo tanto, por el filtro de la Hacienda Nacional".

```
(S
  (NP SUBJ SG P3
    (N "<Sagliocco>" "Sagliocco" PROPER))
  (VP TENSED IMPERFECT IND SG P3
    (V "<mantenía>" "mantener" TENSED IMPERFECT IND SG P3)
    (CL INDIRECT SPEECH
      (PUNCT "," COMMA)
      (PP SEGÚN
        (PREP "<según>" "según")
        (VP TENSED PRES IND PERFECT SG P3
          (V "<ha descubierto>" "descubrir" TENSED PRES IND PERFECT
            SG P3
              (AUX "<ha>" "haber" TENSED PRES IND SG P3)
              (V "<descubierto>" "descubrir" UNTENSED PART MASC SG))
            (ADVP TIME
              (ADV "<ahora>" "ahora" TIME))))
          (NP SUBJ MASC SG P3
            (ART "<el>" "el" DEF MASC SG)
            (N "<fisco>" "fisco" MASC SG)
            (ADJP MASC SG
              (ADJ "<italiano>" "italiano" MASC SG))))
            (PUNCT "," COMMA))
          (NP OBJ1 ID-1
            (QP
              (ADVP IDIOM
```

(ADV "<nada más y nada menos que>" "nada_más_y_nada_menos_que"
 IDIOM))
 (Q "<344>" "trescientos_cuarenta_y_cuatro"))
 (N "<apartamentos>" "apartamento" MASC PL)
 (PRED-COMPL REF-1 MASC PL
 (V "<alquilados>" "alquilar" UNTENSED PART MASC PL)
 (PP EN
 (PUNCT ">" QUOTE)
 (PREP "<en>" "en")
 (NP
 (N "<negro>" "negro" MASC SG))
 (PUNCT "<<" QUOTE))
 (CON
 (PUNCT "," COMMA)
 (CON "<es decir>" "es_decir")
 (PUNCT "," COMMA))
 (PP SIN COORDINATED
 (PP SIN
 (PREP "<sin>" "sin")
 (NP
 (N "<contrato>" "contrato" MASC SG)))
 (C "<y>" "y" COORDINATING)
 (PP SIN
 (PREP "<sin>" "sin")
 (CL INFINITIVE
 (VP UNTENSED INFINITE
 (V "<pasar>" "pasar" UNTENSED INFINITE)
 (CON
 (PUNCT "," COMMA)
 (CON "<por lo tanto>" "por_lo_tanto")
 (PUNCT "," COMMA))
 (PP POR
 (PREP "<por>" "por")
 (NP
 (ART "<el>" "el" DEF MASC SG)
 (N "<filtro>" "filtro" MASC SG)
 (PP DE
 (PREP "<de>" "de")
 (NP
 (ART "<la>" "el" DEF FEM SG)
 (N "<Hacienda>" "hacienda" FEM SG)
 (ADJP FEM SG
 (ADJ "<nacional>" "nacional" FEM SG)))))))))))))
 (PUNCT "." PERIOD))

Foreign words are going to be led to the annotator's intuitions, that is, if he/she thinks the word in question is a noun it will be labelled as such, and the feature FOREIGN-WORD will be added to it. The possible categories we believe may be found are: NOUN, CON, ADJ, and ADV. These words will not receive a syntactic function and, also, they will be treated as compound items if they have two or more components.

Example:

“Mea culpa de Clinton por el genocidio de Ruanda”.

```
(S
  (NP
    (PUNCT " " QUOTE)
    (N "<Mea culpa>" "mea_culpa" FOREIGN-WORD)
    (PUNCT " " QUOTE)
    (PP DE
      (PREP "<de>" "de")
      (NP
        (N "<Clinton>" "Clinton" PROPER))))
  (PP POR
    (PREP "<por>" "por")
    (NP
      (ART "<el>" "el" DEF MASC SG)
      (N "<genocidio>" "genocidio" MASC SG)
      (PP DE
        (PREP "<de>" "de")
        (NP
          (N "<Ruanda>" "Ruanda" PROPER))))))
  (PUNCT "." PERIOD))
```

5.11.3 Dates and Hours

There are several ways of expressing dates and hours in Spanish. Some of these ways are:

```
"El 5 de enero de 1997"
"Hoy es 17 de mayo"
"El 14 del 3 del 74"
"A 9 de Octubre"
"Con fecha de 10 de marzo del presente año"
‘‘Las tres y cuarto’’
‘‘Las 9 y media de la mañana’’
‘‘Las 16,30 h.’’
‘‘Las cinco menos diez de la tarde’’
```

‘‘Las doce en punto de la noche’’

In order to avoid the possible problems that may arise from the analysis in depth of these structures, we have decided to create an special category for them, namely DATE. So, an expression such as "El 5 de mayo de 1997" instead of being annotated in this way:

```
(NP
  (ART "<el>" "el" DEF MASC SG)
  (N "<5>" "cinco")
  (PP DE
    (PREP "<de>" "de")
    (NP
      (N "<mayo>" "mayo" MASC SG)))
  (PP DE
    (PREP "<de>" "de")
    (NP
      (N "<1997>" "mil_novecientos_noventa_y_siete"))))
```

will be annotated:

```
(DATE
  (DATE "<El 5 de mayo de 1997>" "el_cinco_de_mayo_de_mil_novecientos_noventa_y_siete"))
```

In the same way, the annotation of the "hours" may create the same problems as the dates, so we will annotate them under an special category named HOUR. So now, an expression as "Las nueve y media de la mañana" will be annotated as follows:

```
(HOUR
  (HOUR "<Las nueve y media de la mañana>" "las_nueve_y_media_de_la_mañana"))
```

5.11.4 Titles

In the same way as it happened with dates and hours, the "developed" analysis of titles will lead to problematic situations, such as "En *Si te dicen que caí* el autor muestra..." The full analysis of this title into a conditional clause may give us an ungrammatical result: How can it be possible to have a conditional clause introduced by the preposition EN? In order to avoid this, we have created a category for titles so that the previous example will be annotated as follows:

```
(S
  (PP EN LOCATIVE
    (PREP "<En>" "en")
    (TITLE
```

```

        (TITLE "<Si te dicen que caí>" "si_te_dicen_que_caí"))
(NP SUBJ MASC SG P3
  (ART "<el>" "el" DEF MASC SG)
  (N "<autor>" "autor" MASC SG))
(VP TENSED PRES IND SG P3
  (V "<muestra>" "mostrar" TENSED PRES IND SG P3))
(...))

```

5.11.5 Maths operators, Scores and Measures

There are certain expressions that involve *mathematical formulae*. Some of these formulae are expressed by means of mathematical symbols such as +, -, x, ·, etc., but sometimes we may find them in a written form, for example “más”, “menos”, “por”, “entre”, etc. For those formulae, both in a symbol or in a written form, we have created a special category:

```
(MATHS-OP      )
```

which stands for mathematical operator. We will use it in expressions such as: “Uno de cada 10 españoles fuma habitualmente”. In the case of numbers, we transform the number into a word for the lemma features. The annotation of these examples will be as follows:

```

Uno de cada 10 españoles fuma habitualmente”
(S
  (NP SUBJ SG P3
    (QP
      (Q "<uno de cada 10>" "uno_de_cada_diez"
        (Q "uno" MASC SG)
        (MATHS-OP "de_cada")
        (Q "diez" MASC PL)))
      (N "<españoles>" "español" MASC SG))
    (VP TENSED PRES IND SG P3
      (V "<fuma>" "fumar" TENSED PRES IND SG P3)
      (ADVP NORMAL
        (ADV "<habitualmente>" "habitualmente" NORMAL))))

```

In the frame of this annotational schema, we are going to treat *scores* as an special category named SCORE. So, sentences such as “El Real Madrid ganó 2-0 al Barça” will be annotated as follows:

```
(S
```

```

(NP SUBJ SG P3
  (ART "<El>" "el" DEF MASC SG)
  (N "<Real Madrid>" "Real_Madrid" PROPER))
(VP TENSED PAST IND SG P3
  (V "<ganó>" "ganar" TENSED PAST IND SG P3)
  (SCORE
    (SCORE "<2-0>" "dos_a_cero")))
  (PP A OBL
    (PREP "<a1>" "a")
    (NP
      (ART "el" DEF MASC SG)
      (N "<Barça>" "Barça" PROPER))))
(PUNCT "." PERIOD))

```

Measure expressions are going to be annotated in the same way as any other phrase, but they will be added the feature MEASURE. For example:

"La mesa mide 29 cm. de alto"

```

(S
  (NP SUBJ FEM SG P3
    (ART "<La>" "el" DEF FEM SG)
    (N "<mesa>" "mesa" FEM SG))
  (VP TENSED PRES IND SG P3
    (V "<mide>" "medir" TENSED PRES IND SG P3)
    (NP
      (QP
        (Q "<29>" "veintinueve"))
        (N "<cm.>" "centímetro" MEASURE)
        (PP DE
          (PREP "<de>" "de")
          (NP
            (N "<alto>" "alto" MASC SG))))))
  (PUNCT "." PERIOD))

```

Note that the lexeme string of "cm." does not appear in its abbreviated form but in the "full" one.

5.11.6 Punctuation

According to EAGLES, "in the syntactic annotation of written texts, there is a tradition of treating punctuation marks as 'words' for the purpose of parsing". We are going to treat them as marks of major syntactic boundaries, and, as such, they will be bracketed in the following format:

(PUNCT "....." FEATURE)

Medial or mid-sentence punctuation, that is, brackets, parentheses, commas, colons, etc., are going to be treated as sisters of the constituent they surround. For example:

"Juan, mi tío, trabajó en esa tienda".

```
(S
  (NP SUBJ ID-1 SG P3
    (N "<Juan>" PROPER)
    (PUNCT "," COMMA))
  (NP APPOS REF-1 MASC SG
    (POSS "<mi>" "mio" MASC SG P1)
    (N "<tío>" "tío" MASC SG)
    (PUNCT "," COMMA))
  (VP TENSED PAST IND SG P3
    (V "<trabajó>" "trabajar" TENSED PAS IND SG P3)
    (PP EN LOCATIVE
      (PREP "<en>" "en")
      (NP
        (DEM "<esa>" "esa" FEM SG)
        (N "<tienda>" "tienda" FEM SG))))
  (PUNCT "." PERIOD))
```

Final punctuation as a rule is a child of the highest level of structure. For example:

"Diego estuvo en Barcelona."

```
(S
  (NP SUBJ SG P3
    (N "<Diego>" "Diego" PROPER))
  (VP TENSED PAST IND SG P3
    (V "<estuvo>" "estar" TENSED PAST IND SG P3)
    (PP EN LOCATIVE
      (PREP "<en>" "en")
      (NP LOCATIVE
        (N "<Barcelona>" "Barcelona" PROPER))))
  (PUNCT "." PERIOD))
```

Initial punctuation may be also treated as a child of the highest level of structure, but in this case, it will appear opening the constituency structure. For example:

”¿Quieres otra cosa?”

```
(S
(PUNCT "¿" QUESTION)
(NP * SUBJ SG P2)
(VP TENSED PRES IND SG P2
(V "<Quieres>" "querer" TENSED PRES IND SG P2)
(NP OBJ1
(QP
(Q "<otra>" "otro"))
(N "<cosa>" "cosa" FEM SG)))
(PUNCT "?" QUESTION)
(PUNCT "." PERIOD))
```

Colons and semicolons have to receive necessarily an special treatment, since they convey in many cases syntactic, semantic and/or pragmatic meanings or functions. In our corpus, there is a quite fair representation of the use of both punctuation marks in Spanish, but not all their uses are included. For this reason, and also, because a bigger and deeper study of Spanish punctuation system should be done in order to cover all the casuistic, the treatment of colons and semicolons is going to be that of any other punctuation marks, that is, they will be the sister or daughter of their closer constituent, depending on their position inside the sentence.

Examples:

”No haga caso: se pueden obtener resultados igualmente buenos con colchones y bases de distintas marcas.”

```
(S
(NP * SUBJ SG P2)
(VP TENSED PRES IMPERATIVE SG P2 POLITE
(ADVP NEG
(ADV "<No>" "no" NEG))
(V "<haga>" "hacer" TENSED PRES IMPERATIVE SG POLITE)
(NP OBJ1
(N "<caso>" "caso" MASC SG)))
(PUNCT ":" COLONS)
(VP TENSED PRES IND MODAL PL P3
(V "<se pueden obtener>" "obtener" TENSED PRES IND MODAL PASSIVE
PL P3
(SE-MARK "<se>" "se" PASSIVE)
(AUX "<pueden>" "poder" TENSED PRES IND PL P3)
(V "<obtener>" "obtener" UNTENSED INFINITE))
```

(NP OBJ1
 (N "<resultados>" "resultado" MASC PL)
 (ADVP NORMAL
 (ADV "<igualmente>" "igualmente" NORMAL))
 (NP
 (N "<buenos>" "bueno" MASC PL)))
 (PP CON
 (PREP "<con>" "con")
 (NP
 (N COORDINATED
 (N "<colchones>" "colchón" MASC PL)
 (C "<y>" "y" COORDINATING)
 (N "<bases>" "base" FEM PL))
 (PP DE
 (PREP "<de>" "de")
 (NP
 (ADJP FEM PL
 (ADJ "<distintas>" "distinto" FEM PL))
 (N "<marcas>" "marca" FEM PL))))))
 (PUNCT "." PERIOD))

”Todas las decisiones que afectan a la comunidad de vecinos se toman por acuerdo de la junta de propietarios: estas juntas se convocan al menos una vez al año, notificándose a cada propietario 6 días de antelación.”

(S
 (CL
 (NP SUBJ ID-1 PL P3
 (QP
 (Q "<Todas>" "todo"))
 (ART "<las>" "el" DEF FEM PL)
 (N "<decisiones>" "decisión" FEM PL)
 (CL RELATIVE
 (NP SUBJ REF-1 PL P3
 (P "<que>" "que" REL P3))
 (VP TENSED PRES IND PL P3
 (V "<afectan>" "afectar" TENSED PRES IND PL P3)
 (PP A OBL
 (PREP "<a>" "a")
 (NP
 (ART "<la>" "el" DEF FEM SG)
 (N "<comunidad>" "comunidad" FEM SG)
 (PP DE
 (PREP "<de>" "de"))

(NP
 (N "<vecinos>" "vecino" MASC PL)))))))))
 (VP TENSED PRES IND PL P3 PASSIVE
 (SE-MARK "<se>" "se" PASSIVE)
 (V "<toman>" "tomar" TENSED PRES IND PL P3)
 (PP POR
 (PREP "<por>" "por")
 (NP
 (N "<acuerdo>" "acuerdo" MASC SG)
 (PP DE
 (PREP "<de>" "de")
 (NP
 (ART "<la>" "el" DEF FEM SG)
 (N "<junta>" "junta" FEM SG)
 (PP DE
 (PREP "<de>" "de")
 (NP
 (N "<propietarios>" "propietario" MASC PL)))))))))
 (PUNCT ":" COLONS)
 (CL
 (NP SUBJ PL P3
 (DEM "<estas>" "este" FEM PL)
 (N "<juntas>" "junta" FEM PL))
 (VP TENSED PRES IND PL P3 PASSIVE
 (SE-MARK "<se>" "se" PASSIVE)
 (V "<convocan>" "convocar" TENSED PRES IND PL P3)
 (ADVP NORMAL
 (ADV "<al menos>" "al_menos" NORMAL))
 (NP TIME
 (ART "<una>" "un" INDEF FEM SG)
 (N "<vez>" "vez" FEM SG)
 (PP A
 (PREP "<al>" "a")
 (NP
 (ART "el" DEF MASC SG)
 (N "<año>" "año" MASC SG)))
 (PUNCT "," COMMA))
 (CL GERUND
 (VP UNTENSED GERUND
 (V "<notificándose>" "notificar" UNTENSED GERUND #CLITIC)
 (PP A
 (PREP "<a>" "a")
 (NP
 (QP

(Q "<cada>" "cada" DISTR))
(N "<propietario>" "propietario" MASC SG))
(PP CON
(PREP "<con>" "con")
(NP
(QP
(Q "<6>" "seis"))
(N "<días>" "día" MASC PL)
(PP DE
(PREP "<de>" "de")
(NP
(N "<antelación>" "antelación" FEM SG)))))))))
(PUNCT "." PERIOD))

Capítulo 6

Bibliography and References

- Alcina Franch, J., (1989) *Gramática española*, Ariel, Barcelona, 7th edition.
- Bies et al. (1995) *Bracketing Guidelines for Treebank II Style Penn Treebank Project*.
- Bosque, I., (1989) *Las Categorías Gramaticales*, Síntesis, Madrid.
- EAGLES' *Preliminary Recommendations for the Syntactic Annotation of Corpora* (Version of March 1996)
- Gili Gaya, S., (1985) *Curso superior de Sintaxis española*, Bibliograf, Barcelona, 15th edition.
- Marcos Marín, F., (1984) *Curso de Gramática española*, Cincel, Madrid, 2nd edition.
- Marcus et al. (1993) *Building a Large Annotated Corpus of English: The Penn Treebank*.
- Moreno Cabrera, J.C., (1991), *Curso universitario de Lingüística General. Tomo I: Teoría de la gramática y sintaxis general*, Síntesis, Madrid.
- Portolés, J., (1993) "La distinción entre los conectores y otros marcadores del discurso en español", in VERBA 20, pp. 141-170.
- Seco, R., (1985), *Manual de gramática española*, Aguilar, Madrid, 10th edition.
- Seco, M., (1985), *Gramática esencial del español*, Aguilar, Madrid, 9th edition.

Apéndice A

Periphrastic verbs

1. COMENZAR A + INFINITIVE: INCHOATIVE

Ana comenzó a estudiar chino en enero

```
(S
  (NP SUBJ)
  (...)
  (VP TENSED PAS IND INCHOATIVE P3 SG
    (V "<comenzó a estudiar>" "estudiar" TENSED PAS IND INCHOATIVE
      P3 SG
        (AUX "<comenzó a>" "comenzar_a" TENSED PAST IND P3 SG)
        (V "<estudiar>" "estudiar" UNTENSED INFINITE))
    (NP OBJ1
      (...))))
```

2. DEBER + INFINITIVE: MODAL

Juan debe tener unos 40 años

```
(S
  (NP SUBJ)
  (...)
  (VP TENSED PRES IND MODAL SG P3
    (V "<debe tener>" "tener" TENSED PRES IND MODAL SG P3
      (AUX "<debe>" "deber" TENSED PRES IND SG P3)
      (V "<tener>" "tener" UNTENSED INFINITE))
```

(NP OBJ1
(...)))

3. DEBER DE + INFINITIVE: MODAL

Debes de tener más cuidado

(S
(SUBJ * SG P3)
(VP TENSED PRES IND MODAL SG P2
(V "<debes de tener>" "tener" TENSED PRES IND MODAL SG P2
(AUX "<debes de>" "deber_de" TENSED PRES IND SG P2)
(V "<tener>" "tener" UNTENSED INFINITE))
(NP OBJ1
(...)))

4. EMPEZAR A + INFINITIVE: INCHOATIVE

Pedro empezó a estudiar a las 9 de la mañana.

(S
(NP SUBJ MASC SG P3
(...))
(VP TENSED PAST IND INCHOATIVE SG P3
(V "<empezó a estudiar>" "estudiar" TENSED PAST IND INCHOATIVE
SG P3
(AUX "<empezó a>" "empezar_a" TENSED PAST IND SG P3)
(V "estudiar" UNTENSED INFINITE))
(...)))

5. ESTAR + GERUND: PROGRESSIVE

Está terminando con mi paciencia.

(S
(NP SUBJ * SG P2)
(VP TENSED PRES IND PROGRESSIVE SG P3

(V "<está terminando" "terminar" TENSED PRES IND PROGRESSIVE
SG P3

(AUX "<está>" "estar" TENSED PRES IND SG P3)

(V "terminando" UNTENSED GERUND))

(...)))

6. ESTAR + PARTICIPLE: PASSIVE

La casa está construida sobre una colina.

(S

(NP SUBJ

(...))

(VP TENSED PRES IND PASSIVE SG P3

(V "<está construida>" "construir" TENSED PRES IND PASSIVE
SG P3

(AUX "<está>" "estar" TENSED PRES IND SG P3)

(V "construida" UNTENSED PAS PART FEM SG))

(PP LOCATIVE

(...)))

7. HABER DE + INFINITIVE: MODAL

Has de considerar todas las posibilidades.

(S

(NP SUBJ * SG P2)

(VP TENSED PRES IND MODAL SG P2

(V "<has de considerar>" "considerar" TENSED PRES IND MODAL
SG P2

(AUX "<has de>" "haber_de" TENSED PRES IND SG P2)

(V "<considerar>" "considerar" UNTENSED INFINITE))

(NP OBJ1

(...)))

8. HABER + PARTICIPLE: PERFECT

Pedro ha estudiado mucho hoy.

(S
 (NP SUBJ
 (...))
 (VP TENSED PRES IND PERFECT SG P3
 (V "<ha estudiado>" "estudiar" TENSED PRES IND PERFECT SG
 P3
 (AUX "<ha>" "haber" TENSED PRES IND SG P3)
 (V "<estudiado>" "estudiar" UNTENSED PAST PART))
 (...)))

9. IR A + INFINITIVE: FUTURE

Nino va a dormir con sus abuelos hoy.

(S
 (NP SUBJ
 (...))
 (VP TENSED PRES IND FUTURE SG P3
 (V "<va a dormir>" "dormir" TENSED PRES IND FUTURE SG P3
 (AUX "<va a>" "ir_a" TENSED PRES IND SG P3)
 (V "<estudiar>" "estudiar" UNTENSED INFINITE))
 (PP
 (...)))

10. IR + GERUND: PROGRESSIVE

El grupo fue hilando....

(S
 (NP SUBJ MASC SG P3
 (ART "<El>" "el" DEF MASC SG)
 (N "<grupo>" "grupo" MASC SG))
 (VP TENSED PAST IND PROGRESSIVE SG P3
 (V "<fue hilando>" "hilar" TENSED PAST IND PROGRESSIVE SG
 P3
 (AUX "<fue>" "ser" TENSED PAST IND SG P3)
 (V "<hilando>" "hilar" UNTENSED GERUND))
 (...)))

11. PODER + INFINITIVE: MODAL

Puedes tener problemas si sigues así.

```
(S
  (NP SUBJ * SG P2)
  (VP TENSED PRES IND MODAL SG P2
    (V "<Puedes tener>" "tener" TENSED PRES IND MODAL SG P2
      (AUX "<puedes>" "poder" TENSED PRES IND SG P2)
      (V "<tener>" "tener" UNTENSED INFINITE))
    (NP OBJ1
      (...))))
```

12. TENER QUE + INFINITIVE: MODAL

Tienes que controlar tus nervios.

```
(S
  (NP SUBJ * SG P2)
  (VP TENSED PRES IND MODAL SG P2
    (V "<tienes que controlar>" "controlar" TENSED PRES IND
      MODAL SG P2
      (AUX "<tienes que>" "tener_que" TENSED PRES IND SG P2)
      (V "<controlar>" "controlar" UNTENSED INFINITE))))
```

13. TERMINAR DE + INFINITIVE: PERFECT

Termina de hacer los deberes.

```
(S
  (NP SUBJ * SG P2)
  (VP TENSED PRES IND PERFECT SG P2
    (V "<termina de hacer>" "hacer" TENSED PRES IND PERFECT
      SG P2
      (AUX "<termina de>" "terminar_de" TENSED PRES IND SG P2)
      (V "<hacer>" "hacer" UNTENSED INFINITE))
    (NP OBJ1
      (...))))
```

14. VENIR A + INFINITIVE: PROGRESSIVE.

Juan vino a comprar el pan.

```
(S
  (NP SUBJ
    (...))
  (VP
    (V "<vino a comprar>" "comprar" TENSED PRES IND PROGRESSIVE
      SG P3
      (AUX "<vino a>" "venir_a" TENSED PRES IND SG P3)
      (V "<comprar>" "comprar" UNTENSED INFINITE))
    (NP OBJ1
      (...))))
```

15. VENIR + GERUND: PROGRESSIVE

Esto viene sucediendo desde hace mucho tiempo.

```
(S
  (NP SUBJ
    (...))
  (VP TENSED PRES IND PROGRESSIVE SG P3
    (V "<viene sucediendo>" "suceder" TENSED PRES IND PROGRESSIVE
      SG P3
      (AUX "<viene>" "venir" PRES IND SG P3)
      (V "<sucedendo>" "suceder" UNTENSED PART))
    (PP TIME
      (...))))
```

16. VOLVER A + INFINITIVE: PROGRESSIVE

Vuelve a llamarla dentro de una hora.

```
(S
  (NP SUBJ * SG P2)
  (VP TENSED PRES IND PROGRESSIVE SG P2
```

(V "<Vuelve a llamarla>" "llamar" TENSED PRES IND PROGRESSIVE
SG P2
(AUX "<vuelve_a>" "volver_a" TENSED PRES IND SG P2)
(V "<llamarla>" "llamar" UNTENSED INFINITE
(NP OBJ1
(P "la" PERS ACC FEM SG P3))))
(PP TIME
(...))))

Apéndice B

Control Verbs

B.1 Object Control

aconsejar
acusar
advertir
alentar
animar
apelar a
autorizar
capacitar
causar
convencer
dejar
desafiar
exhortar
facultar
forzar
hacer
impedir
imponer
incitar
inducir
influir
inspirar
instar
instruir
mandar
motivar
obligar
oir

ordenar
pedir
permitir
persuadir
presionar
prohibir
prometer
señalar
ver

B.2 Subject Control

amenazar
comprometerse
ofrecer
prometer
lamentar

B.3 Subject and Object Control

anhelar
ansiar
aspirar
decidir
desear
elegir
encargarse de
esforzarse
esperar
estar impaciente por
intentar
jurar
odiar
pretender
proponer
querer
rehusar
tener ganas de
tener la intención de
tratar de

Apéndice C

Multi-word Units

We want to clarify here that not all the units that appear here have a unique meaning since many of them can be found having a compositional meaning (i.e. sometimes they have a meaning taking all the components together, and sometimes they have another meaning each component its own one). Also it is possible to find them undertaking functions which are different from the ones they are supposed to have according to their category.

C.1 Multi-word Adverbs

a bordo
a buenas (horas/noches)
a continuación
a diario
a falta de
a menudo
a plazo fijo
a posteriori
a propósito
a prueba
a rienda suelta
a sangre caliente/fría/y fuego
a simple vista
a través
a última hora
a veces
ahora mismo
al margen
al menos
al mismo tiempo

ante todo
antes de Cristo
cara a cara
como si nada
con frecuencia
con locura
de cerca
de inmediato
de la noche a la mañana
de mal en peor
de memoria
de por sí
de segunda mano
de vez en cuando
dentro de poco
desde ya
en clave de
en cualquier caso
en definitiva
en especial
en su mayoría
en última instancia
entre semana
gota a gota
hoy en día
in situ
mal que bien
mientras tanto
nunca en la vida
paso a paso
poco a poco
por ahora
por ejemplo
por escrito
por horas
por separado
por tanto
sin duda
sin querer
sobre todo

C.2 Multi-word Conjunctions

a condición de (que)
a excepción de que
a medida que
al tiempo que
asi es que
caso de que
con el objeto de
con tal que
de ahí que
debido a que
en cuanto que
en lugar de que
hasta que
por consiguiente
por más que
puesto que
siempre que
tan pronto como
una vez que
ya que

C.3 Multi-word Prepositions

a cambio de
a comienzos de
a diferencia de
a excepción de
a finales de
a lo largo de
a partir de
a propósito de
acerca de
al margen de
con motivo de
conforme a
debajo de
en detrimento de
en función de
en vez del
la mayor parte/mayoría de

pese a
por lo que respecta a
referente a