Description of the Catalan and Spanish tag sets from CoNLL-2009 shared task
(Version 1.0, January 5, 2009)
----

1. Parts of speech and associated features (tags in POS and FEAT columns)

PoS tags

a	adjective
c	conjunction
d	determiner
f	punctuation mark
i	interjection
n	noun
p	pronoun
r	adverb
s	preposition
v	verb
w	date (behaves as a noun)
z	number (behaves either a determiner, a pronoun, or a noun)

PoS features:

Feature	Possible values[1]	Applies to
gen (gender)	m (masculine) f (feminine) c (common)	**a, d, n, p, s**
num (number)	s (singular) p (plural) c (common)	**a, d, n, p, s, v**
postype	negative	**R**
	demonstrative *possessive interrogative exclamative indefinite relative numeral	**d, p**
	article	**D**
	common proper	**N**
	main auxiliary semiauxiliary	**V**
	personal	**P**
	coordinating subordinating	**c**

---

[1] For attributes with possible values yes/no, "no" is the default value. If the attribute is not explicit, the default value is implied.

	preposition	**s**
	currency percentage	**z**
person	1 (first person) 2 (second person) 3 (third person)	**d, p, v**
mood	indicative subjunctive imperative infinitive gerund pastparticiple	**v**
tense	present imperfect future past conditional	**v**
possessornum	s (singular) p (plural) c (common)	**d, p** (the ones with * in the value for attribute "postype" in this chart)
case^2	nominative accusative dative oblique	**p**
punct	comma period *quotation colon semicolon hyphen slash revslash apostrophe mathsign *exclamationmark *questionmark *bracket *sqbracket *cubracket etc	**f**
contracted	yes no	**s**
punctenclose	open close	**f** (the ones with * in the value for attribute "punct" in this chart)

---

2. Syntactic dependencies (tags in DEPREL column)

a	adjective
c	conjunction (from word level)
d	determiner
f	punctuation mark
n	noun
p	pronoun
r	adverb
s	preposition (from word level)
v	verb (auxiliary or semi-auxiliary)
z	number
conj	subordinating conjunction (from syntactic node)
coord	coordinating conjunction (from syntactic node)
gerundi	gerund
grup.a	adjective group
grup.adv	adverb group
grup.nom	noun group
infinitiu	infinitive
inc	inserted element
morfema.pronominal	pronominal morpheme
morfema.verbal	verbal morpheme
neg	negation
participi	participle
prep	preposition (from syntactic node)
relatiu	relative pronoun (from syntactic)
S	clause
sa	adjective phrase (depends of a verb)
s.a	adjective phrase (does not depend of a verb)
sadv	adverb phrase
sentence	sentence (ROOT)
sn	noun phrase
sp	prepositional phrase
spec	specifier
ao	orational adjunct
atr	attribute
cag	agent complement
cc	adjunct
cd	direct object
ci	indirect object
cpred	predicative complement
creg	prepositional object
et	textual element
impers	impersonality mark
mod	non-argumental verb modifier
pass	passive mark
suj	subject

3. Semantic dependencies (tags in APRED columns)

APRED tags are split in two by a hyphen. The first part of the tag is compulsory, whereas the second part may be void.

The first part of the tag refers to arguments:

arg0 applies to the argumental complement closest to the verb, semantically (typically, the subject)

arg4 applies to the argumental complement farthest from the verb, semantically (typically, location-destination complements)

arg1, **arg2** and **arg3** are a gradation between those two edges: the higher the number, the further the complement, semantically (typically, arg1 would correlate to direct object, arg2 to indirect object and arg3 to location-origin complements)

argL applies to lexicalized complements

argM applies to non-argumental complements

The second part of the tag refers to thematic roles:

adv	Non-specific complement
agt	Agent
atr	Attribute
ben	Beneficiary
cau	Cause
des	Destination
efi	Final state
ein	Initial state
exp	Experiencer
ext	Extension
fin	Finality
ins	Instrument
loc	Location
mnr	Manner
ori	Origin
pat	Patient
src	Source
tem	Theme
tmp	Time

4. Predicates and Semantic classes (PRED column)

In PRED column one can find the predicate lemma and its semantic class (Lexical Semantic Structure, LSS) separated by a dot (e.g., *posicionar.a2*)

Verb semantic classes (LSS)

a1: Transitive-causative
a2: Transitive-agentive
a31: Ditransitive-agentive-locative
a32: Ditransitive-agentive-beneficiary

b1: Unaccusative-motion
b2: Unaccusative-state

c1: Existence-state
c2: Attributive-state
c3: Scalar-state
c4: Beneficiary-state

d1: Agentive-inergative
d2: Experiencer-inergative
d3: Source-inergative

"**0**": This tag has been assigned to all the adjectives/past-participles (a) with postype = qualificative (q), and posfunction = participle (p): 'aq---p' (e.g., aq0msp, aq0fsp, etc.). At this stage of the corpus development no specific LSS have been assigned to these adjective/past-participles, so we have assigned to all of them the dummy class "0". The form of the predicate is not the infinitive, but that of adjective/past-participle in its maculine singular variant: 'celebrat.0', 'situat.0', etc. There are 1,406 cases in Spanish and 3,171 cases for Catalan.

"_": In some coordinate structures where the verb has been elided we have taken as predicate the first word in the coordinated clause (it can have any PoS) and assigned the void semantic class "_". There are 256 cases in Spanish corpus and 216 cases in Catalan corpus.

General information about Verb Classes:

a: accomplishments: [+dynamic], [+telic]
b: achievements: [-dynamic], [+telic]
c: states: [-dynamic], [-telic]
d: activities: [+dynamic], [-telic]