CLiPS
COMPUTATIONAL LINGUISTICS & PSYCHOLINGUISTICS RESEARCH CENTER

Penn Treebank II tag set

Pattern (/pages/pattern) and MBSP (/pages/MBSP) assign meaningful tags to words and groups of words in a sentence. Each tag is a short code (such as "**DT**" for "determiner").

The tag set is based on the Penn Treebank Tagging Guidelines [pdf (ftp://ftp.cis.upenn.edu/pub/treebank/doc/tagguide.ps.gz)].

Part-of-speech tags

Part-of-speech tags are assigned to a single word according to its role in the sentence. Traditional grammar classifies words based on eight parts of speech: the verb (**VB**), the noun (**NN**), the pronoun (**PR**+**DT**), the adjective (**JJ**), the adverb (**RB**), the preposition (**IN**), the conjunction (**CC**), and the interjection (**UH**).

Tag	Description	Example
CC	conjunction, coordinating	*and, or, but*
CD	cardinal number	*five, three, 13%*
DT	determiner	*the, a, these*
EX	existential there	*there were six boys*
FW	foreign word	*mais*
IN	conjunction, subordinating or preposition	*of, on, before, unless*
JJ	adjective	*nice, easy*
JJR	adjective, comparative	*nicer, easier*
JJS	adjective, superlative	*nicest, easiest*
LS	list item marker	
MD	verb, modal auxillary	*may, should*
NN	noun, singular or mass	*tiger, chair, laughter*
NNS	noun, plural	*tigers, chairs, insects*
NNP	noun, proper singular	*Germany, God, Alice*
NNPS	noun, proper plural	*we met two Christmases ago*
PDT	predeterminer	*both his children*
POS	possessive ending	*'s*
PRP	pronoun, personal	*me, you, it*
PRP$	pronoun, possessive	*my, your, our*
RB	adverb	*extremely, loudly, hard*
RBR	adverb, comparative	*better*
RBS	adverb, superlative	*best*
RP	adverb, particle	*about, off, up*
SYM	symbol	*%*
TO	infinitival to	*what to do?*
UH	interjection	*oh, oops, gosh*
VB	verb, base form	*think*
VBZ	verb, 3rd person singular present	*she thinks*
VBP	verb, non-3rd person singular present	*I think*
VBD	verb, past tense	*they thought*
VBN	verb, past participle	*a sunken ship*
VBG	verb, gerund or present participle	*thinking is fun*
WDT	*wh*-determiner	*which, whatever, whichever*
WP	*wh*-pronoun, personal	*what, who, whom*
WP$	*wh*-pronoun, possessive	*whose, whosever*
WRB	*wh*-adverb	*where, when*
.	punctuation mark, sentence closer	*.;?**
,	punctuation mark, comma	*,*
:	punctuation mark, colon	*:*
(	contextual separator, left paren	*(*
)	contextual separator, right paren	*)*

Chunk tags

Chunk tags are assigned to groups of words that belong together (i.e. phrases). The most common phrases are the noun phrase (**NP**, for example *the black cat*) and the verb phrase (**VP**, for example *is purring*).

Tag	Description	Words	Example	%
NP	noun phrase	**DT**+**RB**+**JJ**+**NN** + **PR**	*the strange bird*	51
PP	prepositional phrase	**TO**+**IN**	*in between*	19
VP	verb phrase	**RB**+**MD**+**VB**	*was looking*	9
ADVP	adverb phrase	**RB**	*also*	6
ADJP	adjective phrase	**CC**+**RB**+**JJ**	*warm and cosy*	3
SBAR	subordinating conjunction	**IN**	*whether or not*	3
PRT	particle	**RP**	*up the stairs*	1
INTJ	interjection	**UH**	*hello*	0

The IOB prefix marks whether a word is inside or outside of a chunk.

Tag	Description
I-	inside the chunk
B-	inside the chunk, preceding word is part of a different chunk
O	not part of a chunk

A prepositional noun phrase (**PNP**) is a group of chunks starting with a preposition (**PP**) followed by noun phrases (**NP**), for example: *under the table*.

Tag	Description	Chunks	Example
PNP	prepositional noun phrase	**PP**+**NP**	*as of today*

Relation tags

Relations tags describe the relation between different chunks, and clarify the role of a chunk in that relation. The most common roles in a sentence are **SBJ** (subject noun phrase) and **OBJ** (object noun phrase). They link **NP** to **VP** chunks. The subject of a sentence is the person, thing, place or idea that is *doing* or *being* something. The object of a sentence is the person/thing affected by the action.

Tag	Description	Chunks	Example	%
-SBJ	sentence subject	**NP**	*the cat sat on the mat*	35
-OBJ	sentence object	**NP**+**SBAR**	*the cat grabs the fish*	27
-PRD	predicate	**PP**+**NP**+**ADJP**	*the cat feels warm and fuzzy*	7
-TMP	temporal	**PP**+**NP**+**ADVP**	*arrive at noon*	7
-CLR	closely related	**PP**+**NP**+**ADVP**	*work as a researcher*	6
-LOC	location	**PP**	*live in Belgium*	4
-DIR	direction	**PP**	*walk towards the door*	3
-EXT	extent	**PP**+**NP**	*drop 10 %*	1
-PRP	purpose	**PP**+**SBAR**	*die as a result of*	1

Anchor tags

Anchor tags describe how prepositional noun phrases (**PNP**) are attached to other chunks in the sentence. For example, in the sentence, *I eat pizza with a fork*, the anchor of *with a fork* is *eat* because it answers the question: "In what way do I eat?"

Tag	Description	Example
A1	anchor chunks that corresponds to **P1**	*eat with a fork*
P1	**PNP** that corresponds to **A1**	*eat with a fork*

Occurence estimate

The given percentages for chunk and relations tags are based on tenfold cross validation on sections 10 to 19 of the WSJ Corpus of the Penn Treebank II by Sabine Buchholz, from which we derived a rough indication. The estimate means that if a 100 chunk tags are found, about 50 would be **NP** tags and 35 would have a **SBJ** relation tag. About 30 of the chunks would be tagged as **NP-SBJ**, and 15 as **NP-OBJ**.

Reference: Buchholz, S. (2002). *Memory-Based Grammatical Relation Finding*. ILK, Tilburg University.