Part of Speech Tags (POSTAG)

Contents

What Are POSTAGS?

What do all those symbols (VB, JJ, etc) that appear in the output of CoreNLP (the CoNLL table) under the label **POSTAG** mean? They are Part of Speech tags and, these tags mean whatever they meant in your original training data. You are free to invent your own tags in your training data, as long as you are consistent in their usage. Training data generally takes a lot of work to create, so a pre-existing corpus is typically used. These usually use the **Penn Treebank** or **Brown Corpus** tags.

The most common part of speech (POS) tag schemes are those developed for the Penn Treebank and Brown Corpus. Penn Treebank is probably the most common, but both corpora are available with NLTK.

Penn Treebank POS Tags

POS Tag	Description	Example
CC	coordinating conjunction	and
CD	cardinal number	1, third
DT	Determiner	the
EX	existential there	*there* is
FW	foreign word	d'hoevre
IN	preposition/subordinating conjunction	in, of, like
JJ	Adjective	big
JJR	adjective, comparative	bigger
JJS	adjective, superlative	biggest
LS	list marker	1)

MD	Modal	could, will
NN	noun, singular or mass	door

NNS	noun plural	doors
NNP	proper noun, singular	John
NNPS	proper noun, plural	Vikings
PDT	Predeterminer	*both* the boys
POS	possessive ending	friend*'s*
PRP	personal pronoun	I, he, it
PRP$	possessive pronoun	my, his
RB	Adverb	however, usually, naturally, here, good
RBR	adverb, comparative	better
RBS	adverb, superlative	best
RP	Particle	give *up*
TO	To	*to* go, *to* him
UH	Interjection	uhhuhhuhh
VB	verb, base form	take
VBD	verb, past tense	took
VBG	verb, gerund/present participle	taking
VBN	verb, past participle	taken
VBP	verb, sing. present, non-3d	take
VBZ	verb, 3rd person sing. Present	takes
WDT	wh-determiner	which

WP	wh-pronoun	who, what
WP$	possessive wh-pronoun	whose
WRB	wh-abverb	where, when

The official annotation guidelines including full descriptions can be found [here](#) (GZip-compressed Postscript file). This includes confusing parts of speech, capitalization, and other conventions.

Brown Corpus POS Tags

The Brown Corpus POS tags are very similar, and there is the potential for some confusion. However, there are differences. For example, the Penn Treebank has three types of adjective (JJ, JJR, JJS) but the Brown Corpus divides JJS into JJS and JJT.

The Brown Corpus also has rules for combining tags. For example, the colloquial "wanna" means "want to" and is tagged "VB+TO" ("want/VB to/TO"). Similarly, a suffix asterisk indicates a negative, so that "aren't" becomes "BER*".

The [Brown Corpus manual](#) is available [here](#).