Penn Treebank POS Tag Set

The Penn treebank POS tag set has 36 POS tags plus 12 others for punctuations and special symbols. These are listed below. For more details, refer to paper by Marcus, Marcinkiewicz and Santorini that appeared in Computational Linguistics, June 1993 issue 19(2), pages 313-330. (http://acl.ldc.upenn.edu/J/J93/J93-2004.pdf)

Many examples below were taken from http://www.comp.leeds.ac.uk/amalgam/tagsets/upenn.html

1. **CC** Coordinating Conjunction
   - and, or
   - both as in “both x and y”.

2. **CD** Cardinal Number
   - 371, 1
   - one, two as in “one dollar”

3. **DT** Determiner
   - all an another any both each either every many much neither no
   - some such that the them these this those

4. **Ex** Existential There

5. **FW** Foreign Word
   - ich jeux habeas jour salutaris oui corporis

6. **IN** Preposition/subordinating Conjunction
   - among upon in into below atop until over under towards to
   - whether despite if

7. **JJ** Adjective
   - third ill-mannered regrettable calamitous clean nice

8. **JJR** Adjective, Comparative
   - cleaner nicer

9. **JJS** Adjective, Superlative
   - cleanest nicest

10. **LS** List Item Marker
•

11. **MD** *Modal*
   • can could may might must need ought shall
   • cannot can’t shouldn’t

12. **NN** *Noun, singular or mass*
   • machine computer air wind

13. **NNS** *Noun plural*
   • machines computers

14. **NNP** *Proper Noun, Singular*
   • Philadelphia Delaware Eagles

15. **NNPS** *Proper Noun, plural*
   • Americas

16. **PDT** *Predeterminer*
   • all both half

17. **POS** *Possessive ending*
   • ’s

18. **PRP** *Personal pronoun*
   • him himself we

19. **PPS** *Possesive pronoun*
   • her our ours

20. **RB** *Adverb*
    • quickly swiftly
    • always – “always bugging me”

21. **RBR** *Adverb, Comparative*
    • further greater more

22. **RBS** *Adverb, Superlative*
    • farther best hardest most
23. **RP** *Particle*
   - across up

24. **SYM** *Symbol, mathematical or scientific*
   - =, &

25. **TO** *to*

26. **UH** *Interjection*
   - goodbye, shucks, heck, oops

27. **VB** *Verb, base form*
   - hit, assign run

28. **VBD** *Verb, past tense*
   - hit assigned ran

29. **VBG** *Verb, gerund/present participle*
   - hitting

30. **VBN** *Verb, past participle*
   - assigned

31. **VBP** *Verb, non-3rd person singular, present*
   - displeases

32. **VBZ** *Verb, 3rd person singular, present*
   - displeases

33. **WDT** *wh-determiner*
   - that which whichever what

34. **WP** *wh-pronoun*
   - that which whom

35. **WPS** *Possessive wh-pronoun*
   - whose

36. **WRB** *Wh-adverb*
   - how however wherein why
Here are a few examples of common words that can have different POS tags.

- **one**: Consider the different contexts in which it can appear. For each of these contexts, “one” will have to be assigned a different POS tag. “one dollar”, “one of the books”, “the one which I liked”.

- **both**: “both chairs”, “both the chairs”, “both hockey and cricket” “both of the chairs” “saw them both”

- **best**: “tried best”, “the best hitter”, “the best”

- **up**: “up the stairs”, “call him up”