# Part Of Speech (POS) Tagging Methods

1. **Basics of Grammar -Foundations of POS Tagging:** Understanding the basic grammar is a crucial part for POS tagging.
   1. ***Part Of Speech***— Part of Speech defines the function and role of word in a sentence. There are many POS but we will go through the most common one’s.

|  |  |  |  |
| --- | --- | --- | --- |
| POS | Definition | Examples | Context in POS Tagging |
| Noun(NN) | Words that name a person, place, thing, or idea. | **Proper Nouns:** John, Paris **Common Nouns:** dog, city **Abstract Nouns:** happiness, freedom | Often follows determiners (e.g., the cat) or adjectives (e.g., big house). |
| Pronoun(PRP) | Words that replaces nouns | he, she, it, they, who **Subcategories:** Personal Pronouns: I, you, he Reflexive: myself, yourself Relative: who, which, that | Replaces a noun to avoid repetitions. |
| Verb(VB) | Words that describe actions, states, or occurrences. | **Action Verbs:** run, jump **Linking Verbs:** is, seem **Auxiliary Verbs:** have, do, will | **Verb Forms:** Base Form (VB): run Past Tense (VBD): ran Present Participle (VBG): running Past Participle (VBN): run Third Person Singular (VBZ): runs |
| Adjective (JJ) | Words that modify or describe nouns | **Big, red, beautiful** | Often precedes nouns (e.g., big dog). Comparative (e.g., bigger) and superlative (e.g., biggest) forms are common. |
| Adverb (RB) | Words that modify verbs, adjectives, or other adverbs. | quickly, very, too | Often ends in -ly (e.g., quickly) but not always (e.g., fast). |
| Prposition(IN) | Words that show relationships between nouns (or pronouns) and other words in a sentence. | On, in , at, by | Usually followed by noun phrases (e.g., on the table). |
| Conjuction (CC) | Words that connect words, phrases, or clauses. | **Coordinating:** and, but, or **Subordinating:** because, although | Connects elements of similar or different syntactic nature. |
| Interjection(UH) | Words that express emotions or exclamations. | **Oh, wow, ouch** | Often standalone or followed by punctuation. |

* 1. Syntax and Sentence structure: POS tagging not only identifies parts of speech but also examines how they combine to form sentences.

A sentence typically consists of:

* Subject: The doer (noun or pronoun).
* Predicate: The action or state (verb + complements).
* Object: The receiver of the action (noun or pronoun).  
  Example: The cat (subject) chased (predicate) the mouse (object).

It also involves identifying phrase types:

* Noun Phrase (NP): The big dog.
* Verb Phrase (VP): is running fast.
* Prepositional Phrase (PP): on the table.
* Grammar rules, such as S → NP VP, define how phrases combine to form valid sentences. These rules are key to understanding sentence structure and syntax.
  1. POS ambiguity: Words can have multiple POS tags depending on context.

E.g. “I will book a ticket.” – here “book” acts as verb

“This is a great book.” – Here “book” acts as noun

We can resolve this ambiguity using surrounding words or syntactic structure.

* 1. Roles of POS tagging in grammar.
* Linguistic analysis: Helps in syntactic and semantic analysis.
* Applications: Grammar checking, text parsing, Machine Translation, etc.

## Tagging Methods:

1. Rule-Based Methods:

**Lexical Rules**:

* 1. Direct mappings for common words to their POS tags.
  2. Examples:
     1. "the" → DT (Determiner)
     2. "is" → VBZ (Verb, 3rd person singular present)
     3. "and" → CC (Conjunction)

**Suffix-Based Rules**:

* 1. Words are tagged based on suffix patterns, which indicate their likely POS.
  2. Examples:
     1. Words ending in -ly → RB (Adverb): "quickly" → RB
     2. Words ending in -ing → VBG (Gerund/Present Participle): "running" → VBG
     3. Words ending in -ed → VBD (Past Tense Verb): "walked" → VBD

**Proper Noun Handling**:

* 1. Capitalized words (not at the beginning of a sentence) are treated as NNP (Proper Nouns).
  2. Example: "John" → NNP.

**Fallback Rule**:

* 1. If no rule matches, a word is tagged as NN (Noun) by default.

**Context-Aware Adjustments**:

* 1. Modifies tags based on surrounding words' context.
  2. Examples:
     1. If a DT (Determiner) precedes a word, the word is likely an NN (Noun).
     2. If an JJ (Adjective) precedes a word, the word is likely an NN (Noun).

### **Usage in Grammar Checking**

1. **Tagging Output**:
   * The tagged words provide a basis for applying grammar rules (e.g., Subject-Verb Agreement).
2. **Error Identification**:
   * Rules like determiners followed by non-nouns or mismatched subjects and verbs are detected.
3. **Correction**:
   * The errors identified can be corrected using predefined patterns or context-aware logi
4. **Statistical Models**:

Implement HMM-based POS tagging using libraries like Python’s NLTK.

1. Experiment with MaxEnt models and CRFs.
2. **Neural Network Models**:
3. Build a BiLSTM-based POS tagger using TensorFlow or PyTorch.
4. Fine-tune transformer models like BERT for POS tagging.
5. **Experimentation**:
6. Use datasets like the Penn Treebank to train and evaluate models.
7. Compare traditional and modern approaches.
8. **Research Papers**:
9. Read foundational works (e.g., Ratnaparkhi’s MaxEnt model, Lafferty’s CRF paper).
10. Explore state-of-the-art models on arXiv.
11. **Projects**:
12. Build a POS tagging pipeline.
13. Integrate POS tagging into an NLP application (e.g., question answering).