

**SYNTACTIC ANALYSIS**

|  |  |
| --- | --- |
| **lexical analysis** | **syntactic analysis** |
| **Data Cleaning** and **Feature Extraction** | * Find the **roles played by words** in a sentence, * **Interpret the relationship between words**, * **Interpret the grammatical structure** of sentences. |

Syntactic analysis:

|  |  |  |
| --- | --- | --- |
| **Right word** | **Right form(past/present/future/continous)** | **Right place** |

Syntactic analysis is defined as analysis that tells us the logical meaning of certainly given sentences or parts of those sentences. We also need to consider rules of grammar in order to define the logical meaning as well as the correctness of the sentences.

It is also known as **syntax analysis or parsing**. The word ‘parsing’ is originated from the Latin word ‘pars’ which means ‘part’. The syntactic analysis deals with the **syntax of Natural Language**. In syntactic analysis, **grammar rules** have been used.

Consider the following sentence:

Sentence: **School go a boy**

The above sentence does not logically convey its meaning, and its grammatical structure is not correct. So, **Syntactic analysis tells us whether a particular sentence conveys its logical meaning or not** and **whether its grammatical structure is correct or not.**

**Part Of Speech Tagging**

Part-of-speech (POS) tagging is a popular Natural Language Processing process which refers to **categorizing words in a text (corpus)** in correspondence with **a particular part of speech**, depending on the definition of the word and its context.

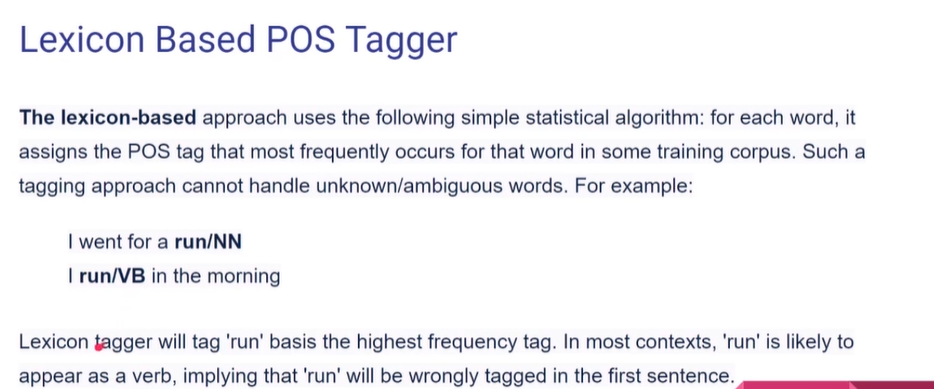


we can see each word has its own lexical term written underneath, however, having to constantly write out these full terms when we perform text analysis can very quickly become cumbersome — especially as the size of the corpus grows. Thence, we use a short representation referred to as “tags” to represent the categories.

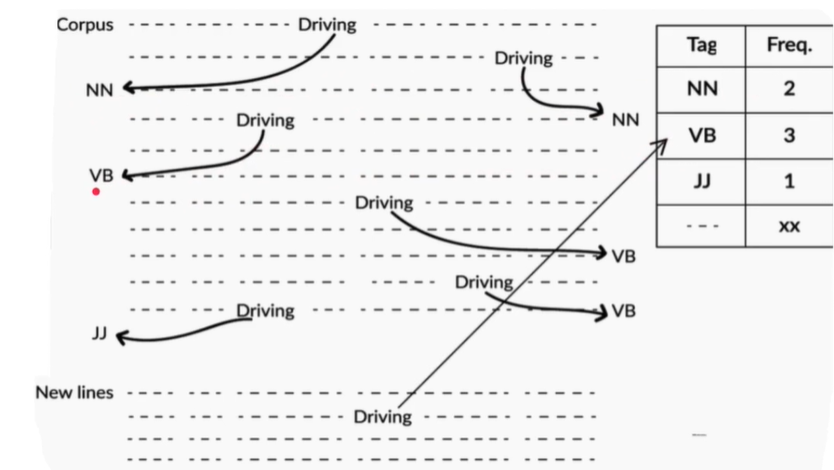


**Techniques for POS tagging:**

|  |  |
| --- | --- |
| **lexicon based POS tagging** |  |
| **Rule-based POS tagging** | Built on lexicon based approach and on this pre defined rules are applied  The rule-based POS tagging models apply a set of handwritten rules and use contextual information to assign POS tags to words. These rules are often known as context frame rules. One such rule might be: “If an ambiguous/unknown word ends with the suffix ‘ing’ and is preceded by a Verb, label it as a Verb”. |
| **Transformation Based Tagging** | use a pre-defined set of handcrafted rules as well as automatically induced rules that are generated during training. |
| **Deep learning models** | Various Deep learning models have been used for POS tagging such as Meta-BiLSTM which have shown an impressive accuracy of around 97 percent. |
| **Stochastic (Probabilistic) tagging** | finds out the most frequently used tag for a specific word in the annotated training data and uses this information to tag that word in the unannotated text. But sometimes this approach comes up with sequences of tags for sentences that are not acceptable according to the grammar rules of a language. One such approach is to calculate the probabilities of various tag sequences that are possible for a sentence and assign the POS tags from the sequence with the highest probability. Hidden Markov Models (HMMs) are probabilistic approaches to assign a POS Tag. |







Since Driving is tagged maximum as Verb, in the new line it will not be tagged as noun even if it is a noun

