

1] Syntax analysis is the process of checking the grammatical correctness of a sentence and identifying the relationship between the words in an input sentence.

2] Hence, this becomes a challenging task to write the grammar rules for any free order languages and Marathi is a free order language.

3] Hence as part of grammar is a description of how words are grouped and connected to each other in a sentence. Syntax usually entails the transformation of a linear sequence of tokens into a hierarchical syntax tree.

4] Parsing, syntax analysis, or syntactic analysis is the process of analysing a string of symbols, either in natural language, computer languages or data structures, conforming to the rules of a formal grammar. The term parsing comes from Latin pars, meaning part.

5] What is Part-of-speech (POS) tagging?

It is a process of converting a sentence to forms – list of words, list of tuples (where each tuple is having a form (word, tag)). The tag in case of is a part-of-speech tag, and signifies whether the word is a noun, adjective, verb, and so on

- ▶ An input sentence is processed by a parser according to the productions of a grammar and builds one or more constituent structures that satisfy the grammar.
- ▶ A parser is a procedural interpretation of the grammar. It permits a grammar to be evaluated against a collection of test sentences, helping learners to find mistakes in their grammatical analysis.
- ▶ The CYK parser model proposed it reads CFG grammar and converts into CNF format if required.

Packages

1] Use of punkt:

This tokenizer divides a text into a list of sentences by using an unsupervised algorithm to build a model for abbreviation words, collocations, and words that start sentences. It must be trained on a large collection of plaintext in the target language before it can be used.