

Simply speaking, parsing in NLP is the process of determining the syntactic structure of a text by analyzing its constituent words based on an underlying grammar (of the language).

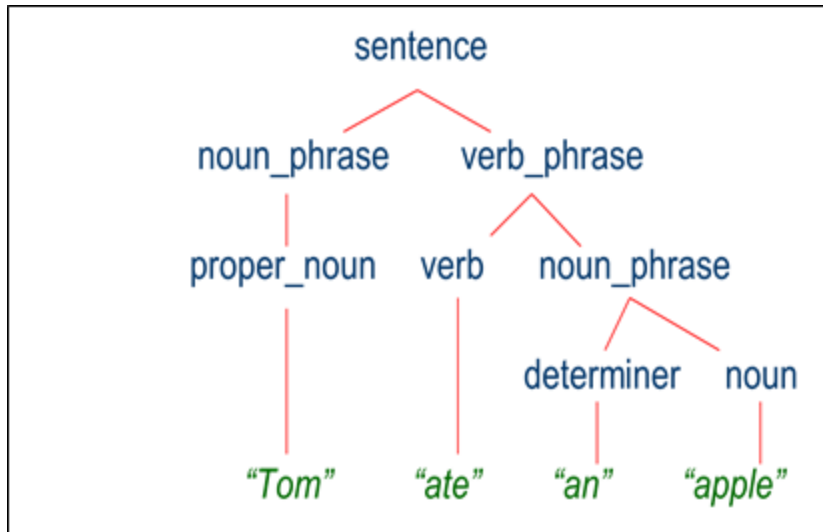
See this example grammar below, where each line indicates a rule of the grammar to be applied to an example sentence “Tom ate an apple”.

Example Grammar:

```
sentence -> noun_phrase, verb_phrase  
noun_phrase -> proper_noun  
noun_phrase -> determiner, noun  
verb_phrase -> verb, noun_phrase  
proper_noun -> [Tom]  
noun -> [apple]  
verb -> [ate]  
determiner -> [an]
```

Then, the outcome of the parsing process would be a parse tree like the following, where **sentence** is the root, intermediate nodes such as **noun_phrase**, **verb_phrase** etc. have children - hence they are called **non-terminals** and finally, the leaves of the tree ‘**Tom**’, ‘**ate**’, ‘**an**’, ‘**apple**’ are called **terminals**.

Parse Tree:



Existing parsing approaches are basically statistical, probabilistic, and machine learning-based. Some notable tools to use for parsing are: Stanford parser (The Stanford Natural Language Processing Group), OpenNLP (Apache OpenNLP Developer Documentation) etc.

A parse tree

- John ate the apple.

1. $S \rightarrow NP VP$

2. $VP \rightarrow V NP$

3. $NP \rightarrow NAME$

4. $NP \rightarrow ART N$

5. $NAME \rightarrow John$

6. $V \rightarrow ate$

7. $ART \rightarrow the$

8. $N \rightarrow apple$

