MAIN IDEAS OF THE REQUIRED READING :

The paper mainly focuses on a particular system of determining the syntactic structure of text by analyzing its constituent words based on an underlying grammar. This grammatical analysis is represented by means of a tree structure and the linguistically annotated text corpus is referred to as a ‘**tree-bank**’. The system that automatically analyses a sentence with reference to its syntactic structure is called a ‘**parser**’ and the tree-like representations it may produce are ‘tree-bank style trees’. For a parser to be able to parse sentences or text, it needs a training corpus or tree-bank so that it can learn from the data. The ‘**Penn Wall Street Journal Tree-bank’** is the first large-scale treebank published and the parser mentioned in the paper uses this data for both training and testing. The paper talks about an increase in metrics like precision and recall and a reduction in the error rate of parses obtained by the mentioned parser when compared to those obtained as a result of other existing parsers. Next, it explains the differences in performance by tracing the particular decisions made in the construction of the parsing system.

Design of the parsing system :
The parser follows a ‘probabilistic generative model’ which means that it assigns probabilities to all possible parses for a sentence and then chooses the parse for which the probability is the highest. To understand how these probabilities are assigned, certain terms that need to be explained are :
Lexical Head - the word that determines the syntactic category of a phrase or sentence (for eg. in the phrase ‘boiling hot water’, ‘water’ is the head). ‘**Context free grammar**’ CFG **-** is a set of rules (called production rules) that provides a simple and mathematically precise mechanism by which phrases in natural language are built from smaller blocks. Probabilistic context free grammar (PCFG) is a CFG with probabilities assigned to each rule.
A parse is like a bag of context-free grammar rules specifying how each parse constituent is expanded. Given the ‘heads’ for each constituent, the probability of a parse of a sentence is determined by first determining the probability of its head, then the probability of the form of the constituent given the head, and finally recursively finding the probabilities of sub-constituents.
So, we now know that the probabilities of constituents or an expansion of constituents is **conditioned** on the ‘head’ among other things. The parser mentioned in the paper distinguishes itself from other parsers (in terms of performance) owing to the manner in which these probabilities are conditioned as well as **smoothed** (assign some non-zero probability to events that were plausible in reality but were not found in the training data used to estimate probabilities).

The parser follows a ‘**maximum-entropy**’ approach to conditioning and smoothing. Maximum Entropy is a guiding principle in assigning probabilities to events. (A probability distribution with maximum entropy incorporates the least possible information.) When computing probabilities of constituents, the parser chooses from a defined set of features. The parsing system benefits from this maximum entropy inspired approach because of its flexibility. It allows for probability computation to be factored into a sequence of features and hence changing the features can easily change the probability model. Finally, the maximum entropy model also does not require features to be independent of each other. So, if we do not have enough examples of conditioning events in the training corpus to ensure that the empirically obtained conditional probability is accurate, the max-entropy model can include features for all involved conditioning events.