Computational Linguistics (CL3.101)

Summer 2021, IIIT Hyderabad

20 July, Tuesday (Lecture 24) – Syntax 5

Taught by Prof. Radhika Mamidi

Parsing

Chomsky's model of generative grammar (or phrase structure grammar) focuses on producing all grammatical sentences of a language. As it did not show the relations between sentences, transformational grammar was developed to account for this.

In computational linguistics, phrase structure grammar is studied in terms of context-free grammars or CFGs.

Probabilistic CFGs

In a PCFG, each rule $N^i \rightarrow \zeta^j$ is associated with a probability $P(N^i \rightarrow \zeta^k)$, such that the sum $\sum_j P(N^i \rightarrow \zeta^j)$ is 1.

The probability of a derivation is the product of the probabilities of all the rules that go into it. Thus a PCFG gives some idea of the plausibility of a sentence. However, this is not a very good idea, since it is not lexicalised – it is better for grammar induction.

Top-Down and Bottom-Up Parsing

Top-down parsing involves starting with the start variable S and developing the tree downwards till the terminals.
Bottom-up parsing involves taking the terminals and analysing them into constituents repeatedly until we reach S.

Ambiguities

Ambiguities in sentences can be structural, lexical, referential etc. Syntax, however, is only concerned with structural ambiguities. Parsing algorithms can generate all possible parse trees for a given sequence of terminals.