SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 97
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Artificial Intelligence and Data Science

Course Name – 19AD601 – Natural Language Processing

III Year / VI Semester

Unit 3 – SYNTACTIC ANALYSIS

Topic 3- Normal form for Grammar

Normal form for grammars

- We say that two grammars are strongly equivalent if they generate the same set of strings and if they assign the same phrase structure to each sentence (allowing merely for renaming of the non-terminal symbols).
- Two grammars are weakly equivalent if they generate the same set of strings but do not assign the same phrase structure to each sentence.
- It is sometimes useful to have a normal form for grammars, in which each of the productions takes a particular form.
- For example, a context-free grammar is in Chomsky normal form (CNF) if it is €-free and if in addition each production is either of the form A□B C or A□a.
- That is, the right-hand side of each rule either has two non-terminal symbols or one terminal symbol.

Normal form for grammars

- We say that two grammars are strongly equivalent if they generate the same set of strings and if they assign the same phrase structure to each sentence (allowing merely for renaming of the non-terminal symbols).
- Two grammars are weakly equivalent if they generate the same set of strings but do not assign the same phrase structure to each sentence.
- It is sometimes useful to have a normal form for grammars, in which each of the productions takes a particular form.
- For example, a context-free grammar is in Chomsky normal form (CNF) if it is €-free and if in addition each production is either of the form A□B C or A□a.
- That is, the right-hand side of each rule either has two non-terminal symbols or one terminal symbol.

Normal form for grammars

Chomsky normal form grammars are binary branching, that is they have binary trees (down to the prelexical nodes).

Any context-free grammar can be converted into a weakly equivalent Chomsky normal form grammar. For example, a rule of the form A □B C D can be converted into the following two CNF rules
A □ B X
X □ C D

Sometimes using binary branching can actually produce smaller grammars. For example, the sentences that might be characterized as

VP □ VBD NP PP*

Normal form for grammars

are represented in the Penn Treebank by this series of rules:
VP → VBD NP PP
VP → VBD NP PP PP
VP → VBD NP PP PP PP
VP → VBD NP PP PP PP PP

but could also be generated by the following two-rule grammar:
VP → VBD NP PP
VP → VP PP

The generation of a symbol A with a potentially infinite sequence of symbols B with rule of the form A → A B is known as Chomsky-adjunction.

THANK YOU