CSE 396 Lecture Thu Mar $\boxed{8}$ 7 Spring 201$\boxed{8}$ 9

The first two pages were read "robotically" from last year's notes.

Top Hat #4268

A context-free grammar (CFG) - an object is a 4-tuple $G = (V, \Sigma, R, S)$ where

- V is a finite alphabet of variables, aka nonterminals
- Σ, also called T, is a finite alphabet of terminals

We suppose $V \cap \Sigma = \emptyset$

- S, a member of V, is the starting variable aka. the start symbol
- R is a finite set of rules of the form

$$A \to X \text{ where } A \in V \text{ and } X \in (\Sigma \cup V)^*.$$

Really $R \subseteq V \times (V \cup \Sigma)^*$. Also called P for productions

The lecture branched off here - Generally was covered after giving examples first

Def^n: Given $X, Y \in (\Sigma \cup V)^*$, we write

$X \underset{G}{\Longrightarrow} Y$ "X derives Y in one step of G"

if X can be broken as $X = U \cdot A \cdot W$ and there is a rule $A \to Z$ in R s.t. $Y = U \cdot Z \cdot W$.

Also define $X \overset{0}{\underset{G}{\Rightarrow}} X$ for any $X \in (\Sigma \cup V)^*$

and for $K \geq 1$, $X \overset{K}{\underset{G}{\Rightarrow}} Z$ if there is a Y in $(\Sigma \cup V)^*$ s.t. $X \overset{K-1}{\underset{G}{\Rightarrow}} Y$ and $Y \underset{G}{\Rightarrow} Z$.

I.e., X derives Z in K steps of the grammar G.

Finally: $\boxed{X \overset{*}{\underset{G}{\Rightarrow}} Z \text{ if } X \overset{K}{\underset{G}{\Rightarrow}} Z \text{ for some } K \geq 0}$

Language has Terminal Strings Only

$\boxed{L(G) = \{ w \in \Sigma^* : S \overset{*}{\underset{G}{\Rightarrow}} w \}.}$

Then $L(G)$ is called a Context-Free Language (CFL).

However: A string $X \in (\Sigma \cup V^*)$ is called a **Sentential Form** of G if $S \overset{*}{\underset{G}{\Rightarrow}} X$

The lecture branched off here — Chomsky was covered after giving examples first

Origin Noam Chomsky, 1956-57 Book: **Syntactic Structures**:

"followed by a"

$$S \longrightarrow \underbrace{N}_{\text{Noun Phrase}}\ \underbrace{V}_{\text{Verb Phrase}}$$

A sentence can be a Noun Phrase Verb Phrase

Variables are often written as "tokens" <noun_phrase> as <verb_phrase> **BNF**

Examples: First, $L(G)$ can be nonregular:

① $L = \{a^n b^n : n \geq 0\}$.

$S \to \varepsilon$

$$S \to aSb \mid \varepsilon$$

Clear that $V = \{S\}$, $\Sigma = \{a, b\}$, R has these two rules.

Example Derivation: $S \underset{G}{\Rightarrow} aSb \Rightarrow \underbrace{aa}_{u}S\underbrace{bb}_{w} \Rightarrow \underbrace{aa}_{u}\underbrace{bb}_{w}$.

used rule $S \to \varepsilon$ here, $z = \varepsilon$

Thus $S \underset{G}{\Rightarrow}^{*} aabb$. "Clearly" $L = \{x \in \Sigma^* : S \Rightarrow^* x\} =_{def} L(G)$.

② $L = \{x \in \{a,b\}^* : x = x^R$, i.e., x is a palindrome$\}$ PAL

$$S \to aSa \mid bSb \mid a \mid b \mid \varepsilon$$

so that we can derive e.g. $S \Rightarrow aSa \Rightarrow abSba \Rightarrow abba$.

Then $L(G) = L$. To get EVENPAL $= \{x \cdot x^R : x \in \Sigma^*\}$ remove rules $S \to a \mid b$.

How about PAL's sister language DOUBLEWORD $= \{xx : x \in \Sigma^*\}$?

FACT: There is no CFG for this language! Will prove later with the CFL Pumping Lemma in §2.3.

③ $\Sigma = \{$'(', ')'$\}$ BAL $= \{x \in \Sigma^* : x$ is a balanced string of parentheses.$\}$

G

$$S \to (S) \mid SS \mid \varepsilon$$

"If x and y are balanced then so are (x) and xy."

$x = (()())()$

$((S))S$ would $\Rightarrow$ make x impossible

$$S \underset{G}{\Rightarrow} SS \underset{G}{\Rightarrow} (S)S \underset{G}{\Rightarrow} (SS)S \Rightarrow ((S)S)S$$

use $S \to \varepsilon$

$$\Rightarrow (()S)S \Rightarrow (()(S))S \Rightarrow (()())S \Rightarrow^2 (()())() = x.$$

④ Expressions $E \Rightarrow a \mid 0 \mid 1 \mid (E+E) \mid (E-E) \mid (E*E) \mid (E/E)$ -izabin

Start Symbol E. Sound, but not comprehensive because it mandates full parenth

(5) CFGs for Natural Human Languages.

Noam Chomsky used S to stand for "Sentence"

$S \to N V$

or more properly

$S \to \langle NP \rangle \langle VP \rangle$

"can be a"

N ~~for~~ or ⟨NP⟩ for a noun or noun phrase

V or ⟨VP⟩ for a verb or verb phrase.

A for an adjective.

A noun phrase can be a noun or a noun preceded by an adjective — one or more of those.

usually in french.

$\langle NP \rangle \to N \mid A \langle NP \rangle \mid \langle NP \rangle A$

Transcribing the rest of the lecture: Although French has different rules — such as usually putting adjectives after nouns rather than before as in English — the point is that it has rules. It has rules that are equally simple enough to be expressed via a CFG. So do all other human languages that have been observed. Chomsky's "Rationalist Thesis" is that in ways independent of childhood upbringing, we are "wired for language" — well in particular, wired for grammar. How strong is this wiring? Chomsky's famous example sentence

"Colorless green ideas sleep furiously"

Sounds cogent when we first hear it — even though it is nonsensical: "colorless" and "green" contradict each other, as basically do "sleep" and "furiously." The analogy raised by a class question is that a program with a division-by-zero line x = y/0; will still pass the compiler since it is grammatical, though it will bomb when run. However, even when the meaning of a sentence is completely clear midway through it, we still get uncomfortable if the sentence doesn't .... [I left the room to underscore this] .... finish!
Thus our brains' wiring and robot wiring may be less far apart than we think...

END