Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007
3/30/07: **Lecture 27 aid — CFGs and parsing**

Topics: More on CFGs, and algorithms for parsing with respect to a given CFG.

I. Parse trees (see the previous lecture aid for formal definitions) Assuming a fixed particular CFG G, a parse tree has for a sentence (sequence of terminals) has root labeled with the start symbol, branches given by the rewrite rules, and the sentence as its leaves.

II. Example CFG

- Terminals (vocabulary): mom, dad
- Non-terminals (category/constituent labels): S
- Start non-terminal (type for a full sentence): S
- Rewrite rules: S →mom dad, S →mom S dad

III. A linguistically-motivated CFG Self-check: confirm that rules 1, 3, 4, 6, 7, 8, 9, 10, 11, and 12 can be combined to generate the left-hand parse tree in item (I) on the previous lecture aid (the one where "on Tuesday" modifies "flights").

- Terminals: list, all, flights, on, Tuesday
- Non-terminals: S, NP, N′, PP, V, DET, N, P
- Start non-terminal: S
- Rewrite rules:

(1)	S	→	V NP	(7)	N	→	flights
(2)	S	→	V NP PP	(8)	N′	→	N PP
(3)	V	→	list	(9)	PP	→	P NP
(4)	NP	→	DET N′	(10)	P	→	on
(5)	NP	→	DET N	(11)	NP	→	N
(6)	DET	→	all	(12)	N	→	Tuesday

IV. Example CFL All and only sentences of the following form, where $\ell \geq 1, m \geq 0$:

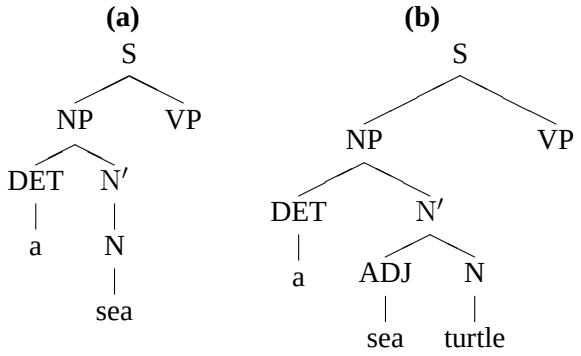
$$\underbrace{\mathrm{a}\ldots\mathrm{a}}_{\ell}\underbrace{\mathrm{b}\ldots\mathrm{b}}_{\ell}\underbrace{\mathrm{c}\ldots\mathrm{c}}_{m}\underbrace{\mathrm{d}\ldots\mathrm{d}}_{m}$$

V. Possible CFG for the above? Could we have the following rewrite rules, where the terminals are a,b,c,d; the non-terminals are S and E; and the start symbol is S?

E →a E b E →a b
E →c E d E →c d
S →E E
S →E

(OVER)

VI. Example Earley-style partial parse trees The sentence being parsed is "a sea turtle swam to shore", and we assume some reasonable CFG for English is being used.



The point is that the leftmost leaves correspond to the first words of the sentence being parsed: Earley's algorithm tries to "guess" at how the parse tree will grow, but "anchors" its guesses against the words of the sentence in a left-to-right manner.

VII. Parse states Assume a fixed sentence $w_1 w_2 \ldots w_n$ and CFG with start non-terminal S. The general form of a parse state is

$$(X \rightarrow \alpha \bullet \beta, i, j),$$

where

- α and β are some "stuff" (a sequence of zero or more terminals or nonterminals) such that $X \rightarrow \alpha\beta$ is a rewrite rule in the CFG,
- i and j range between 0 and n, and are *usually* interpreted as indicated endpoints of some subsequence of the sentence, and
- if $1 \leq i \leq j$, then we have inferred through the parsing process that α (i.e., the "stuff" before the "dot") can be rewritten into the sentence subsequence $w_i w_{i+1} \cdots w_j$.