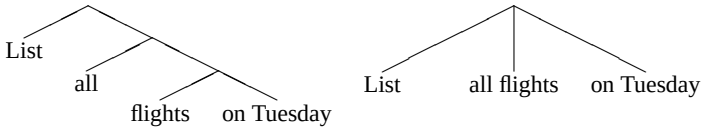
Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007
3/28/07: **Lecture 26 aid — Context-free grammars**

Topics: Context-free grammars (CFGs): a model of families of syntactic structures.

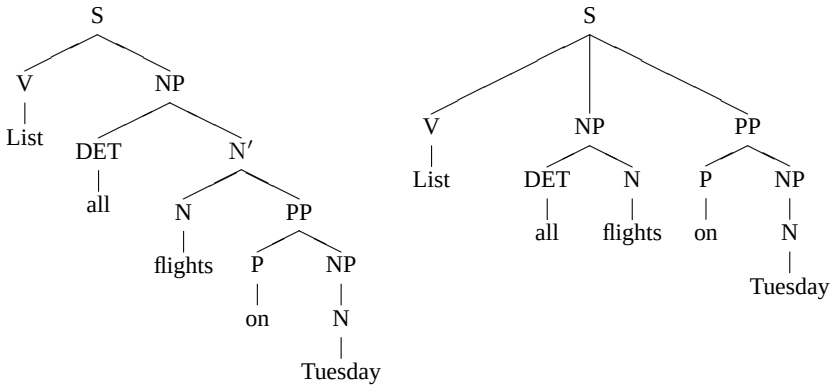
Announcements: The undergraduate coordinator for Information Science has asked that the following announcement be passed on by instructors of INFO courses:

> Information Science spring 2007 open house, 5-6pm, 315 Upson Hall, dinner will be served. All are welcome - it's a chance for prospective students to meet current IS/ISST majors and pick up info materials, and also for everyone to just chat.

I. Tree representation of possible syntactic analyses, reduced form Here, arguments of or modifiers to an item are shown in a sibling relationship with the item in question.



II. An alternative representation with constituent labels Other analyses are possible.



NP=noun phrase; VP=verb phrase; PP=prepositional phrase. N′ (pronounced "N-bar") is a "bare noun phrase". The labels just above the leaves represent parts of speech (POS); DET=determiner.

III. Context-free grammar (CFGs): definition Four components must be specified:

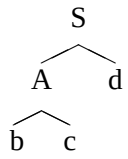
- the *terminals*: a finite set of at least one symbol;
- the *non-terminals*: a finite set of at least one symbol, where no symbol can be both a terminal and a non-terminal;

(OVER)

- a single designated *start non-terminal* (a.k.a. *start symbol*; and
- the *rewrite rules*: a finite set of at least one rule describing how a single non-terminal can be decomposed into a sequence of terminals and/or nonterminals (possibly intermixed[1]).

IV. Related concepts

We can think of CFGs as generating *parse trees*. In a parse tree, the interior nodes are labelled by non-terminals, the root is labelled with the start non-terminal, the leaves are labelled by terminals, and the children of an internal node represent, in order, the result of rewriting the non-terminal labelling the node according to one of the rewrite rules in the grammar. That is, if we have the following parse tree:



then the CFG generating this parse tree must contain the rewrite rules $A \rightarrow bc$ and $S \rightarrow Ad$.

A CFG is said to generate a given *sentence* (or *string*) — a sequence of terminal symbols — if it generates a parse tree in which the labels of the leaves, in order, form the sentence.

The *language* generated by a given CFG is the set of sentences the CFG generates.

A set of sentences is a *context-free language (CFL)* if there exists a CFG that generates all and only the sentences in that set.

V. A linguistically-motivated CFG

- 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

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

$$\underbrace{\text{a} \ldots \text{a}}_{n}\underbrace{\text{b} \ldots \text{b}}_{n}\underbrace{\text{c} \ldots \text{c}}_{m}\underbrace{\text{d} \ldots \text{d}}_{m}$$

[1] And possible *empty*, although in this class we will try to avoid dealing with empty strings.