CONTEXT-FREE GRAMMARS
Linguistics 680, Autumn 2008

(1) **Context-Free Grammars (CFGs)**

A CFG is an ordered quadruple $\langle T, N, D, P \rangle$ where

a. T is a finite set called the **terminals**;

b. N is a finite set called the **nonterminals**

c. D is a finite subset of $N \times T$ called the **lexical entries**;

d. P is a finite subset of $N \times N^{+}$ called the **phrase structure rules** (PSRs).

(2) **CFG Notation**

a. '$A \rightarrow t$ ' means $\langle A, t \rangle \in D$.

b. '$A \rightarrow A_0 \ldots A_{n-1}$' means $\langle A, A_0 \ldots A_{n-1} \rangle \in P$.

c. '$A \rightarrow \{s_0, \ldots s_{n-1}\}$' abbreviates $A \rightarrow s_i$ $(i < n)$.

(3) **A 'Toy' CFG for English**

T = {Fido, Felix, Mary, barked, bit, gave, believed, heard, the, cat, dog, yesterday}

N = {S, NP, VP, TV, DTV, SV, Det, N, Adv}

D consist of the following lexical entries:

NP → {Fido, Felix, Mary}
VP → barked
TV → bit
DTV → gave
SV → {believed, heard}
Det → the
N → {cat, dog}
Adv → yesterday

P consists of the following PSRs:

S → NP VP
VP → {TV NP, DTV NP NP, SV S, VP Adv}
NP → Det N

(4) **Context-Free Languages (CFLs)**

a. Given a CFG $\langle T, N, D, P\rangle$, we can define a function C from N to (T-)languages (we write C_A for $C(A)$) as described below.

b. The C_A are called the **syntactic categories** of the CFG (and so a nointerminal can be thought of as a name of a syntactic category.

c. A language is called **context-free** if it is a syntactic category of some CFG.

(5) **Historical Notes**

- Up until the mid 1980's an open research questions was whether NLs (considered as sets of word strings) were context-free languages (CFLs).
- Chomsky maintained they were not, and his invention of transformational grammar (TG) was motivated in large part by the perceived need to go beyond the expressive power of CFGs.
- Gazdar and Pullum (early 1980's) refuted all published arguments that NLs could not be CFLs.
- Together with Klein and Sag, they developed a context-free framework, generalized phrase structure grammar (GPSG), for syntactic theory.
- But in 1985, Shieber published a paper arguing that Swiss German cannot be a CFL.
- Shieber's argument is still generally accepted today.

(6) **Defining the Syntactic Categories of a CFG**

a. We will recursively define a function $h : \omega \to \wp(T^*)^N$.

b. Intuitively, for each nonterminal A, the $h(n)(A)$ are successively larger approximations of C_A.

c. Then C_A is defined to be $C_A =_{\text{def}} \bigcup_{n\in\omega} h(n)(A)$.

d. To define h, we use RT with the parameters X, x, F set as follows:

   i. $X = \wp(T^*)^N$

   ii. x is the function that maps each $A \in N$ to the set of length-one strings t for all lexical entries $A \rightarrow t$.

   iii. F is the function from X to X that maps a function $L : N \rightarrow \wp(T^*)$ to the function that maps each nonterminal A to the union of $L(A)$ with the set of all strings that can be obtained by applying a PSR $A \rightarrow A_0 \ldots A_{n-1}$ to strings $s_0, \ldots, s_{n-1}$, where, for each $i < n$, s_i belongs to $L(A_i)$. In other words:
   $F(L)(A) = F(L) \cup \bigcup\{L(A_0) \bullet \ldots \bullet L(A_{n-1}) \mid A \rightarrow A_0 \ldots A_{n-1}\}$.

   iv. Given these values of X, x, and F, the RT guarantees the existence of a unique function h from ω to functions from N to $\wp(T^*)$.