The Chomsky Hierarchy

Type 0: Unrestricted grammars

The largest family of grammars permits productions of the form:

$$\alpha \longrightarrow \beta$$

where α and β are arbitrary strings of grammar symbols with $\alpha \neq e$.

Example

$G = (V_{\mathbf{T}}, V_{\mathbf{N}}, P, S)$

$V_{\mathbf{T}} = \{\mathbf{a}\}$

$V_{\mathbf{N}} = \{A, B, C, D, E, S\}$

P:
1) S—> ACaB
2) Ca—>aaC
3) CB—>DB
4) CB—>E
5) aD—>Da
6) AD—>AC
7) aE—>Ea
8) AE—>e

Let us prove that this grammar generates $\mathbf{a}^4$

The Chomsky Hierarchy

$$S \xrightarrow{1} ACaB \xrightarrow{2} AaaCB \xrightarrow{3} AaaDB \xrightarrow{5} AaDaB \xrightarrow{5} ADaaB \xrightarrow{6}$$

$$ACaaB \xrightarrow{2} AaaCaB \longrightarrow$$

$$\xrightarrow{2} AaaaaCB = Aa^4CB \longrightarrow$$

$$\xrightarrow{4\ !!} Aa^4E \xrightarrow{7} Aa^3Ea$$

$$\xrightarrow{7} Aa^2Ea^2$$

$$\xrightarrow{7} AaEa^3$$

$$\xrightarrow{7} AEa^4 \xrightarrow{8} a^4$$

We can prove that this grammar generates the following language: $\{\mathbf{a}^{\boldsymbol{i}} : \boldsymbol{i}$ is a positive power of 2$\}$.

It was proved that type 0 grammars are equivalent to Turing machines and generate recursively enumerable languages.

The Chomsky Hierarchy

Type 1: Context-sensitive grammars

We place a restriction on productions

$$\alpha \longrightarrow \beta$$

as follows: $|\alpha| \leq |\beta|$ (length).

The term "context-sensitive" comes from a normal form for these grammars, where each production is as follows:

$$\alpha_1 A \alpha_2 \longrightarrow \alpha_1 \beta \alpha_2, \text{ with } \beta \neq e$$

These productions permit the replacement of nonterminal A by string β only "in the context" α_1, α_2.

The Chomsky Hierarchy

Almost any language one can think of is context-sensitive! Consider again the grammar of the previous example.

$$G = (V_T, V_N, P, S)$$

$V_T = \{a\}$ $V_N = \{A, B, C, D, E, S\}$

P: 1) S—> ACaB
2) Ca—>aaC
3) CB—>DB
4) CB—>E
5) aD—>Da
6) AD—>AC
7) aE—>Ea
8) AE—>e

There are two productions that violate the definition of context-sensitive grammar:

CB —> E
AE —> e

We can "correct" this grammar by creating new nonterminal symbols as "composite" strings like [CaB] which is a single symbol appearing in place of the string CaB.

The complete set of composite symbols we need is:

[ACaB], [Aa], [ACa], [ADa], [AEa],
[Ca],[Da],[Ea],[aCB], [CaB], [aDB], [aE], [DaB], and [aB]

The Chomsky Hierarchy

The new set of productions is as follows: (numbers correspond to old productions from Example 1) L(G) = $\{a^{2^i} : i \geq 1\}$

1) S—>ACaB	1) S—>[ACaB]
2) Ca—>aaC	2) [Ca]a—>aa[Ca] [Ca][aB]—>aa[CaB] [ACa]a—>[Aa]a[Ca] [ACa][aB]—>[Aa]a[CaB] [ACaB]—>[Aa][aCB] [CaB]—>a[aCB]
3) CB—>DB	3) [aCB]—>[aDB]
4) CB—>E	4) [aCB]—>aE]
5) aD—>Da	5) a[Da]—>[Da]a [aDB]—>[DaB] [Aa][Da]—>[ADa]a a[DaB]—>[Da][aB] [Aa][DaB]—>[ADa][aB]
6) AD—>AC	6) [ADa]—>[ACa]
7) aE—>Ea	7) a[Ea]—>[Ea]a [aE]—>[Ea] [Aa][Ea]—>[AEa]a
8) AE—>e	8) [AEa]—>a

Example 2.<u>Context-sensitive grammar</u> :

$G = (V_T, V_N, P, S)$ $V_T = \{a,b,c\}$ $V_N = \{S,R,T\}$

P:
1) S—> aSRT
2) S—> aRT
3) TR—> RT
4) aR—> ab
5) bR—> bb
6) bT—> bc
7) cT—> cc

Derivation for $a^n b^n c^n$:

$$n-1 \left\{ \begin{array}{l} S \Rightarrow aSRT \underset{1}{\Rightarrow} \\ \Rightarrow a(aSRT)\ RT \underset{1}{\Rightarrow} \\ \ldots\ldots\ldots\ldots \end{array} \right.$$

$$\underset{1}{\Rightarrow} a^{n-1}(SRT)(RT)\ldots(RT) \underset{2}{\Rightarrow} a^n RT\ \underbrace{RT(RT)\ \ldots\ (RT)}_{n-2} \Rightarrow$$

$$1 \left\{ \Rightarrow a^n R^2 T^2 (RT)\ldots(RT) \underset{3}{\Rightarrow} \right.$$

$$2 \left\{ \begin{array}{l} \Rightarrow a^n R^2 TRT^2 (RT)\ldots(RT) \underset{3}{\Rightarrow} \\ \Rightarrow a^n R^3 T^3 (RT)\ldots(RT) \underset{3}{\Rightarrow} \end{array} \right.$$

. . . .

$$n-1 \left\{ \begin{array}{l} \Rightarrow a^n R^{n-1} T^{n-2} RT^2 \underset{3}{\Rightarrow} \\ \ldots\ldots\ldots\ldots \\ \Rightarrow a^n R^n T^n \underset{3}{\Rightarrow} \end{array} \right.$$

The Chomsky Hierarchy

$\Rightarrow a^{\mathbf{n}}bR^{\mathbf{n-1}}T^{\mathbf{n}} \Rightarrow a^{\mathbf{n}}b^{\mathbf{2}}R^{\mathbf{n-2}}T^{\mathbf{n}} \Rightarrow \ldots \Rightarrow a^{\mathbf{n}}b^{\mathbf{n}}T^{\mathbf{n}} \Rightarrow$

4 **5** **5** **5**

==========================

n-1

$\Rightarrow a^{\mathbf{n}}b^{\mathbf{n}}cT^{\mathbf{n-1}} \Rightarrow a^{\mathbf{n}}b^{\mathbf{n}}c^{\mathbf{2}}T^{\mathbf{n-2}} \Rightarrow \ldots \Rightarrow a^{\mathbf{n}}b^{\mathbf{n}}c^{\mathbf{n}}$

6 **7** **7** **7**

==========================

n-1

Run time

Total:
1. n-1 times
2. 1
3. n(n-1)/2 = 1+2+3+. . . + (n-1)
4. 1
5. n-1
6. 1
7. n-1

---

n(n+5)/2

The Chomsky Hierarchy

Type 2: Context-free grammars

Each production is as follows

$$A \longrightarrow B$$

where $A \in V_N$, $B \in (V_T \cup V_N)^+$

These productions permit the replacement of A independently of any context.

Example: $G = (V_T, V_N, P, S)$,

$V_T = \{a, b\}$ $V_N = \{S, A, B\}$

P:

1)	S —> aB	5)	A —> a
2)	S —> bA	6)	B —> bS
3)	A —> aS	7)	B —> aBB
4)	A —> bAA	8)	B —> b

What language does this grammar generate?

Answer: All the strings with equal numbers of symbols **a** and **b**.

$$S \xrightarrow{1} aB \xrightarrow{8} ab$$

$$S \xrightarrow{1} aB \xrightarrow{6} abs \xrightarrow{2} abbA \xrightarrow{5} abba$$

$$S \xrightarrow{2} bA \xrightarrow{5} ba$$

$$S \xrightarrow{2} bA \xrightarrow{4} bbAA \xrightarrow{5} bbaA \xrightarrow{5} bbaa$$

What about Pascal Grammar?

The Chomsky Hierarchy

Derivation Trees for Context-free Grammars.

Definition: A tree is a derivation tree for the grammar

$$\mathbf{G = (V_T, V_N, P, S)} \quad \text{if:}$$

1) Every vertex has a label, which is a symbol from $V_T \cup V_N$
2) The label of the root is S.
3) If a vertex is interior and has label A then A must be in V_N.
4) If n has label A and vertices $n_1, n_2, \ldots, n_k$ are the sons of vertex n, in order from the left, with labels $A_1, A_2, \ldots, A_k$, respectively then;

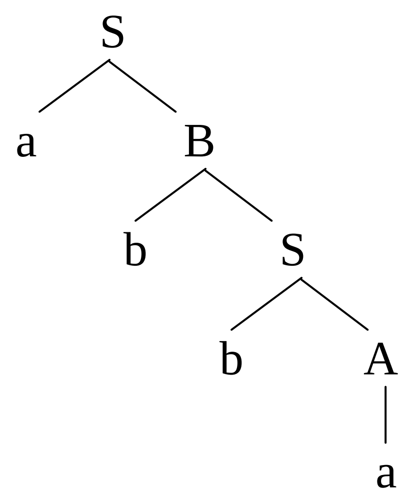
   $$A \longrightarrow A_1 A_2 \ldots A_k$$

   must be a production in P.

Example: Consider derivation S => abba from the previous example.

$$S \xrightarrow{1} aB \xrightarrow{6} abs \xrightarrow{2} abbA \xrightarrow{5} abba$$

The derivation tree is as follows:



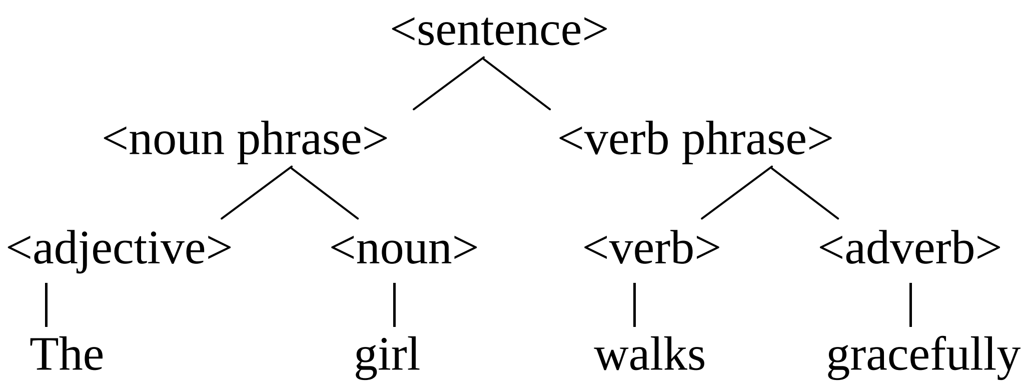
The Chomsky Hierarchy

What can we say about the grammar generating the phrase?

“**The girl walks gracefully**.”

It is context-free!

The derivation tree:



Type 3: Regular (automata) grammars

All productions are as follows:

A —> aB

or

A —> b,

where A, B $\in V_{\mathbf{N}}$, a, b $\in V_{\mathbf{T}}$.

This grammar is called **right-**linear. If all productions are of the form A —> Ba, A —>b, we call it **left-**linear. A right- or left-linear grammar is called a **regular** grammar.

The Chomsky Hierarchy

Example: P: S—> aA
A —> aA
A —> b

generates the language

$L(G) = \{a^{n}b \mid n=1, 2, \ldots\}$

$S \Rightarrow aA \Rightarrow aaA \Rightarrow aaaA \Rightarrow aaab.$

Example:

The language $\{0^{m}1^{n} \mid m, n = 1, 2, \ldots\}$ is regular but

$\{0^{n}1^{n} \mid n = 1, 2, \ldots\}$ is not!

Example: What about the grammar generating the sentence:

"**The girl walks gracefully**."

What about finite languages?