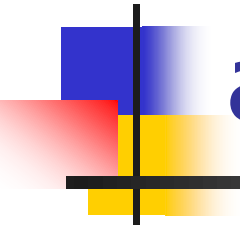


Chapter 11
A Hierarchy of Formal Languages and Automata

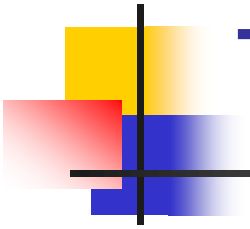
The Chomsky Hierarchy

Recursive and Recursively Enumerable Languages

Unrestricted Grammars

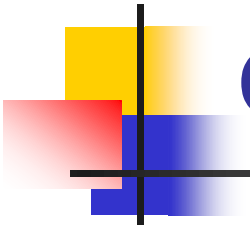
Context –Sensitive Grammars and Languages

The Chomsky Hierarchy



- Each of the four grammar classes has a simple yet powerful automaton
- A useful connection between grammars, automata, and languages
  - Type: 0, 1, 2, 3
  - Languages: r.e.L, CSL, CFL, RL
  - Grammar production restrictions: (X and Y)
    - $X \rightarrow Y$
  - Automata: TM, LBM, PDA, FA

To prove a language to be in certain language class



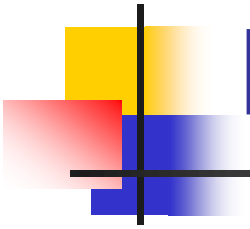
- You may show the language can be generated from associated grammar class
- For example,
  - G = 2 $\Rightarrow$ Automata =PDA
  - G = 0 $\Rightarrow$ Automata =TM

Chomsky Hierarchy of Grammars

The Chomsky Hierarchy of Grammars

Type	Name of Languages Generated	Production Restrictions $X \rightarrow Y$	Acceptor
0	Phrase-structure = recursively enumerable	X = any string with nonterminals Y = any string	TM
1	Context-sensitive	X = any string with nonterminals Y = any string as long as or longer than X	TM's with bounded (not infinite) TAPE, called linear-bounded automata LBA's‡
2	Context-free	X = one nonterminal Y = any string	PDA
3	Regular	X = one nonterminal $Y = t\,N$ or $Y = t$ t terminal N nonterminal	FA

‡The size of the tape is a linear function of the length of the input.

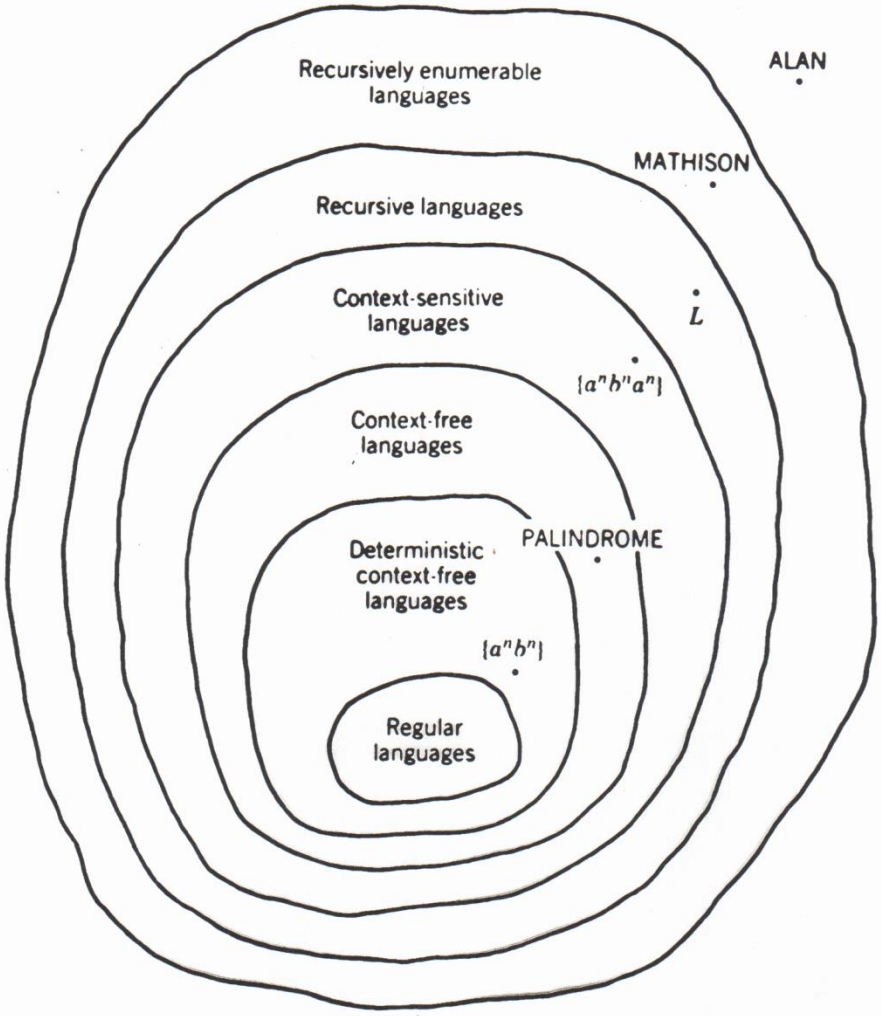


Note: exceptions

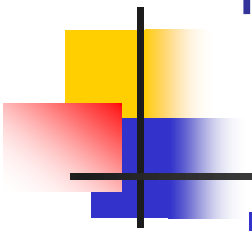
- Which type (0, 1, 2, 3) of grammar is the following grammar belong to?
  - S → aSb | λ

THE CHOMSKY HIERARCHY

These six classes of languages form a nested set as shown in the Venn diagram below.



No two of these categories are really the same



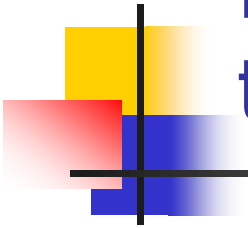
- $\{a^n b^n\}$ is deterministic context-free but not regular
- Palindrome is context free but not deterministic context-free
- $\{a^n b^n C^n\}$ is context-sensitive but not context-free

Chomsky Hierarchy of Grammars

(Make sure that you are able to fill it up!)

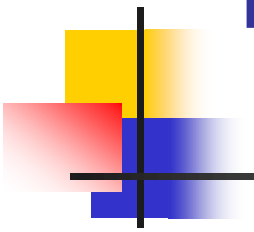
Type	Language	Production Restrictions	Acceptor	Representative

Every TM T over the alphabet Σ divides the set of strings Σ^* into three classes

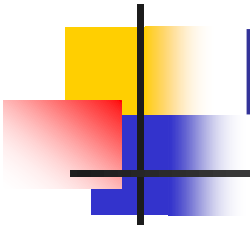


1. ACCEPT (T) is the set of all strings leading to a final state. This is also called the language accepted by T.
2. REJECT (T) is the set of all strings that halt in a non-final state.
3. LOOP (T) is the set of all other strings that loop forever while running on T.

Definition 11.1

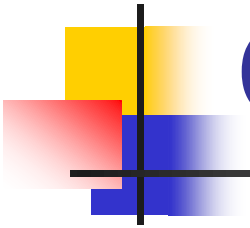


- A language L is said to be recursively enumerable if there exists a TM that accepts it



Definition 11.2

- A language L on Σ is said to be recursive if there exists a TM T that accepts L and that halts on every w in Σ^+
- In other words, a language is recursive if and only if there exists a membership algorithm
  - Note, in this case Loop(T) = ϕ

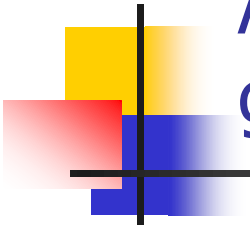


Question:

- Is recursive language a subset of recursively enumerable language or the other way around?

Definition 11.3
Unrestricted Grammar

- A grammar G = (V, T, S, P) is called unrestricted if all the productions are of the form
  - $u \rightarrow v$
  - Where u is in $(V \cup T)^+$ and v is in $(V \cup T)^*$
  - u should contain at least one non-terminal

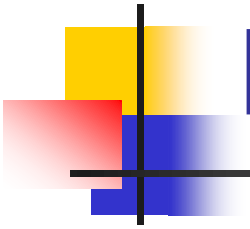


Theorem 11.6
Any language generated by an unrestricted grammar is recursively enumerable

- Proof (highlight)
  - We can simulate the derivations of unrestricted grammar on a TM $\Rightarrow$ it is recursively enumerable

Definition 11.4

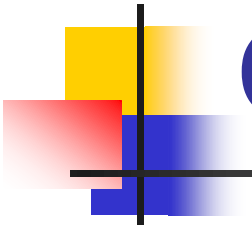
- A grammar G = (V, T, S, P) is said to be context-sensitive if all productions are of the form
  - $x \rightarrow y$
  - Where $x, y \in (V \cup T)^+$
  - and $|x| \leq |y|$



Definition 11.5

- A language L is said to be context-sensitive if there exists a context sensitive grammar G, such that L = L(G).

Context Sensitive Grammar Construction



- None of the examples is easy
- The next fifty years will see natural language processing as a dominant issue

Example 11.2

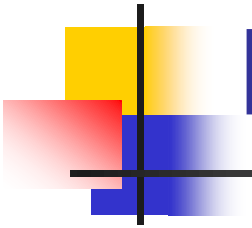
Show that L = $\{a^n b^n c^n : n \geq 1\}$ is CSL

- Construct a CSG for L
  - S $\rightarrow$ abc |aAbc
  - Ab $\rightarrow$ bA
  - Ac $\rightarrow$ Bbcc
  - bB $\rightarrow$ Bb
  - aB $\rightarrow$ aa |aaA
- Note: "A" is playing a role of messenger

In-class exercises

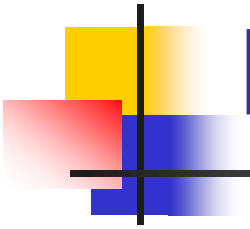
(need to be turned in together with HW)

- Page 294 #1 (b)
  - Find context-sensitive grammars for the following language:
  - $L = \{a^n b^n a^{2n} : n \geq 1\}$
- Turing Machine as Computer:
  - (a) 3-interger adder;
  - (b) adder for any number of integers



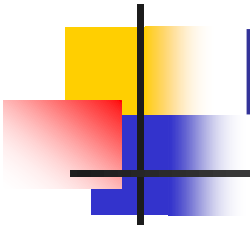
Definitions Review

- NPDA, DPDA
- Turing Machine
- Chomsky Hierarchy
- CSL, CSG



Review Exercises -1

- Give a formal definition for the following terms: TM, CSG, Chomsky Hierarchy
- Construct a PDA in two steps (algorithm and TG) for the following languages:
  - $L=\{a^n b^m : n \leq m \leq 2n\}$
  - $L=\{a^n b^{3n} : n \geq 0\}$
  - $L=\{wcw^R : w \in \{a, b\}^*$



Review Exercises -2

- Construct a TM in TG that will accept the following languages on {a, b},
  - L = {w: |w| is multiple of 2}
  - $L=\{a^n b^{3n} : n \geq 0\}$
- Construct a TM in TG that is an adder for any number of integers