Midterm 1: Mon Sep 29th 7-9pm.
Conflict: Tue 30th
DRES

$\{0^n 1^n \mid n \geq 0\}$ not regular!

$\{(^n)^n \mid n \geq 0\}$

Regular languages:

concatenation,

union,

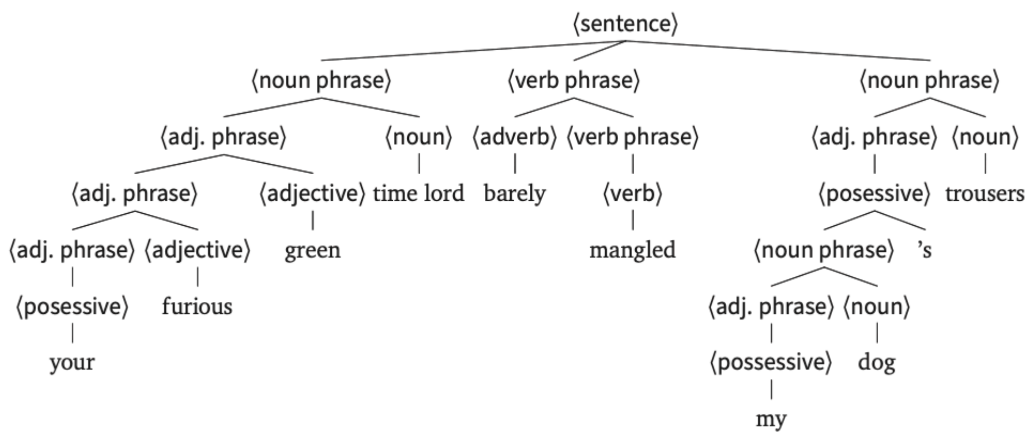
~~repetition~~

Recursion!

via context-free grammars

⟨sentence⟩ → ⟨noun phrase⟩⟨verb phrase⟩⟨noun phrase⟩
⟨noun phrase⟩ → ⟨adjective phrase⟩⟨noun⟩
⟨adj. phrase⟩ → ⟨article⟩ | ⟨possessive⟩ | ⟨adjective phrase⟩⟨adjective⟩
⟨verb phrase⟩ → ⟨verb⟩ | ⟨adverb⟩⟨verb phrase⟩

⟨noun⟩ → dog | trousers | daughter | nose | homework | time lord | pony | · · ·
⟨article⟩ → the | a | some | every | that | · · ·
⟨possessive⟩ → ⟨noun phrase⟩’s | my | your | his | her | · · ·
⟨adjective⟩ → friendly | furious | moist | green | severed | timey-wimey | little | · · ·
⟨verb⟩ → ate | found | wrote | killed | mangled | saved | invented | broke | · · ·
⟨adverb⟩ → squarely | incompetently | barely | sort of | awkwardly | totally | · · ·



A context-free grammar:

- a finite set Σ: the symbols or terminals
- a finite set Γ: disjoint from Σ: non-terminals or variables
- a finite set R of production rules of the form $A \to w$ with $A \in \Gamma$ and $w \in (\Sigma \cup \Gamma)^*$
- a starting non-terminal S

Ex: $\Sigma = \{0, 1\}$. $\Gamma = \{S, A, B, C\}$

$R:$ $S \to A$ $\quad S \to B$ $\quad A \to 0A$ $\quad A \to 0C$

$\quad B \to B1$ $\quad B \to C1$ $\quad C \to \varepsilon$ $\quad C \to 0C1$

$R:$ $S \to A \mid B$

$A \to 0A \mid 0C$

$B \to B1 \mid C1$

$C \to \varepsilon \mid 0C1$

For any $x, y, z \in (\Sigma \cup \Gamma)^*$ and a non-terminal A, applying $A \to y$ to the string $x\underline{A}z$, yields $x\underline{y}z$

Written as $xAz \rightsquigarrow xyz$.

Ex: $00\underline{C}1BA\underline{C}0 \rightsquigarrow 000\underline{0C11}BAC0$

$\rightsquigarrow 00C1BA\underline{0C1}0$

z derives from x, written $x \rightsquigarrow^* z$, if x becomes z after a finite sequence of applications.

Each $w \in (\Sigma \cup \Gamma)^*$ has a language
$$L(w) := \{x \in \Sigma^* \mid w \rightsquigarrow^* x\}$$

Given a context-free grammar $G = (\Sigma, \Gamma, R, S)$, the language generated from G is
$$L(G) = L(S)$$

A language is <u>context-free</u> if generated from some context-free grammar.

Ex: $S \to A \mid B$

$A \to 0A \mid 0C$

$B \to B1 \mid C1$

$C \to \varepsilon \mid 0C1$

Which language:

$L(C)$: $C \rightsquigarrow \varepsilon$

$C \rightsquigarrow 0C1 \rightsquigarrow 01$

$L(C) = \{\varepsilon, 01, 0011, 000111, \ldots\}$

$= \{0^n 1^n \mid n \geq 0\}$ ✓

$B \rightsquigarrow^* \{C1^k \mid k \geq 1\} \Rightarrow L(B) = \{0^m 1^m 1^k \mid m \geq 0, k \geq 1\}$

$= \{0^m 1^n \mid m < n\}$

$L(A) = \{0^m 1^n \mid m > n\}$

$L(S) = \{0^m 1^n \mid m \neq n\}$ ← generated by grammer

context-free is a <u>strict</u> superset of the regular languages.

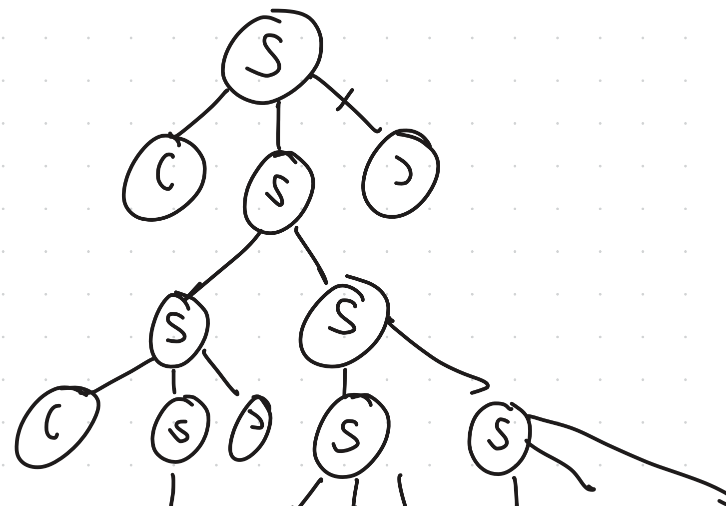
$\{0^n 1^n 0^n \mid n \geq 0\}$ is <u>not</u> context-free.

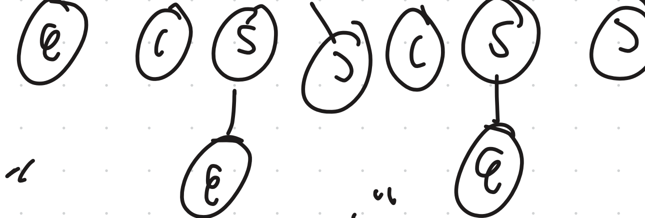
Balanced Parentheses:

$$S \to \varepsilon \mid (S) \mid SS$$

Issue: <u>ambiguous</u>: a string can have multiple derivations.

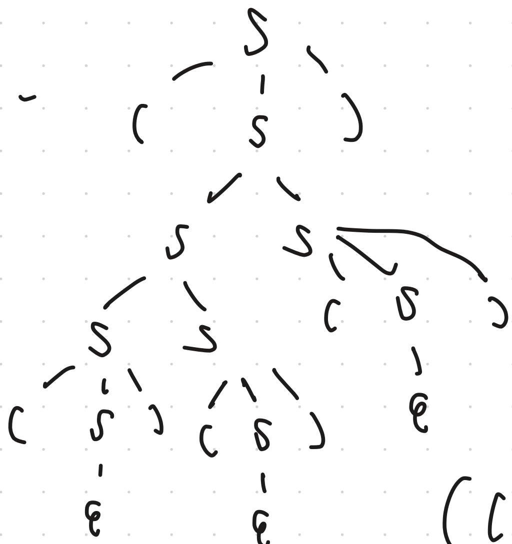
Seen with a <u>parse tree</u>:





Derived "(()()())"

- or -



(()()())

A language is <u>inherently ambiguous</u> if all generating grammars are ambiguous.

But! $S \to \epsilon \mid (S)S$ is <u>not</u> ambiguous.

Strings over $\{0,1,2\}$ of the form

$\{0^i 1^j 2^k \mid j > i+k\}$

$S \to ABC$

$A \to \epsilon \mid 0A1$ $\qquad \{0^i 1^i \mid i \geq 0\}$

$B \to 1 \mid B1$ $\qquad \{1^j \mid j \geq 1\}$

$C \to 1C2 \mid \epsilon$ $\qquad \{1^k 2^k \mid k \geq 0\}$

C D

$$\underbrace{(1+01)^*(1+10)^*}_{A} + \underbrace{1^*0}_{B}$$

$$S \to A \mid B$$
$$A \to CD$$
$$B \to 1B \mid 0$$
$$C \to 1C \mid 01C \mid \varepsilon$$
$$D \to 1D \mid 10D \mid \varepsilon$$