GPS 4: Mon Sep 21st
HW 4: Tue Sep 22nd

last GPS + HW before MT1

MT1: Mon Sep 28 7pm-9pm
CMT1: Tue Sep 29

Regular: union + concatenation + repetition
(Kleene closure)

Not regular: $\{0^n 1^n \mid n \geq 0\}$

: balanced parantheses

(()())()

: regular expressions

Need to add unbounded recursion.

Context-free grammars.

<sentence> → <noun phrase><verb phrase><noun phrase>

<noun phrase> → <adjective phrase><noun>

<noun> → dog | trousers | daughter...

<adjective phrase> → <article> | <possessive> | <adjective phrase><adjective>

⟨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 | · · ·

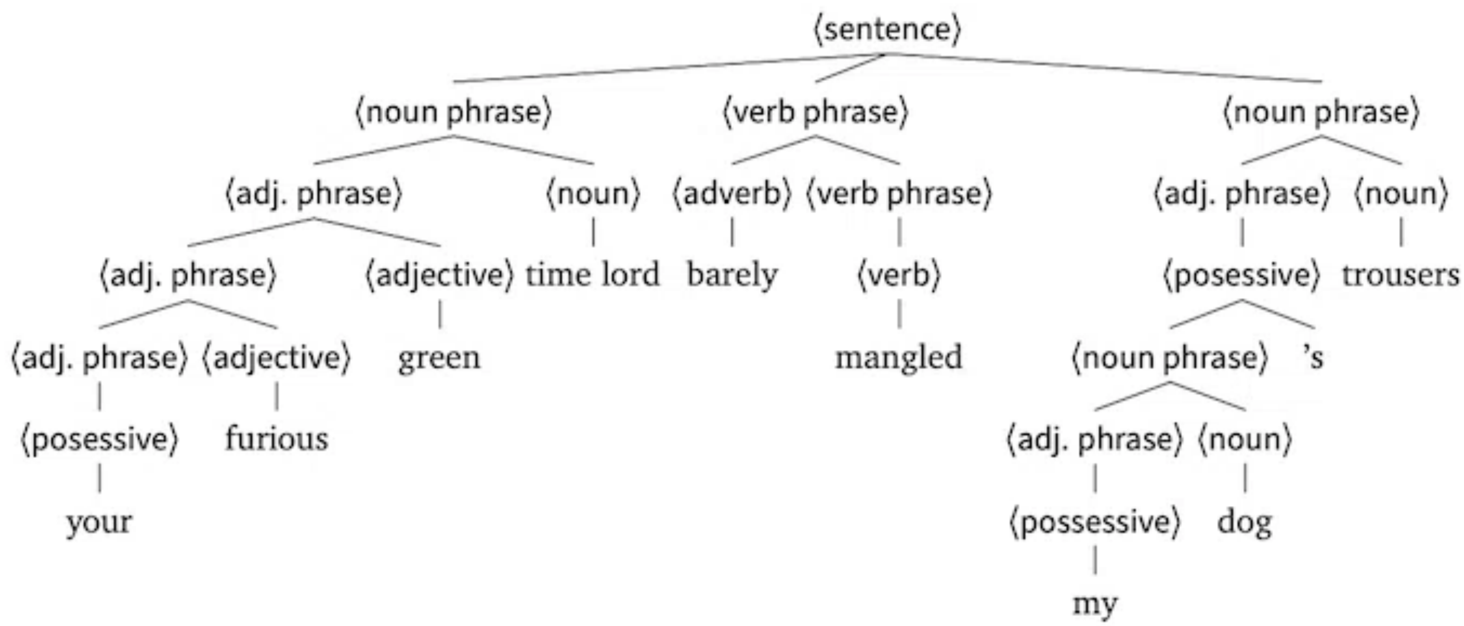
⟨verb⟩ → ate | found | wrote | killed | mangled | saved | invented | broke | · · ·

⟨adverb⟩ → squarely | incompetently | barely | sort of | awkwardly | totally | · · ·

$\langle \text{sentence} \rangle \rightsquigarrow \langle \text{n.p.} \rangle \langle \text{v.p.} \rangle \langle \text{n.p.} \rangle$

$\rightsquigarrow \underline{\langle \text{adj.p} \rangle \langle \text{noun} \rangle} \langle \text{v.p.} \rangle \langle \text{n.p.} \rangle$

$\rightsquigarrow^{*} \text{your furious} \langle \text{noun} \rangle \langle \text{v.p.} \rangle \langle \text{n.p.} \rangle$



A context-free grammar (CFG) consists of ~~five~~ four components.

- A finite set Σ called terminals (or symbols)
- A finite set T disjoint from Σ of non-terminals
- A finite set R of production rules of the form $A \to w$ where $A \in T$ & $w \in (\Sigma \cup T)^*$
- A starting non-terminal $S \in T$

Multiple rules with one LHS writen as $A \to w_1 \mid w_2 \mid w_3$

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

$$
\begin{aligned}
S &\to A \mid B \\
A &\to 0A \mid 0C \\
B &\to B1 \mid C1 \\
C &\to \varepsilon \mid 0C1
\end{aligned}
$$

We <u>apply</u> a production rule $A \to w_0$ to a string w_1 by replacing any A in w_1 with w_0.

i.e. For any $x, y, z \in (\Sigma \cup \Gamma)^*$ & a non-terminal $A \in T$,

we apply production $A \to y$ to the string xAz to get string xyz.

(Say $xAz \rightsquigarrow xyz$)

Ex: $00C1BAC0$ rule $C \to 0C1$

apply rule to get $000C11BAC0$

or $00C1BA0C10$

z derives from x ($x \rightsquigarrow^* z$) if x

transforms into z via a sequence of applications.

$$00C1BAC0 \rightsquigarrow^* 000C11BA0C10$$

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

Given grammar $G = (\Sigma, \Gamma, R, S)$, the language $L(G)$ generated by G is $L(S)$.

Ex: $S \to \varepsilon \mid 0S1$ generates $\{0^n 1^n \mid n \geq 0\}$

L is context-free if generated by

a CFG.

Facts:

Thm: Every regular language is context-free.

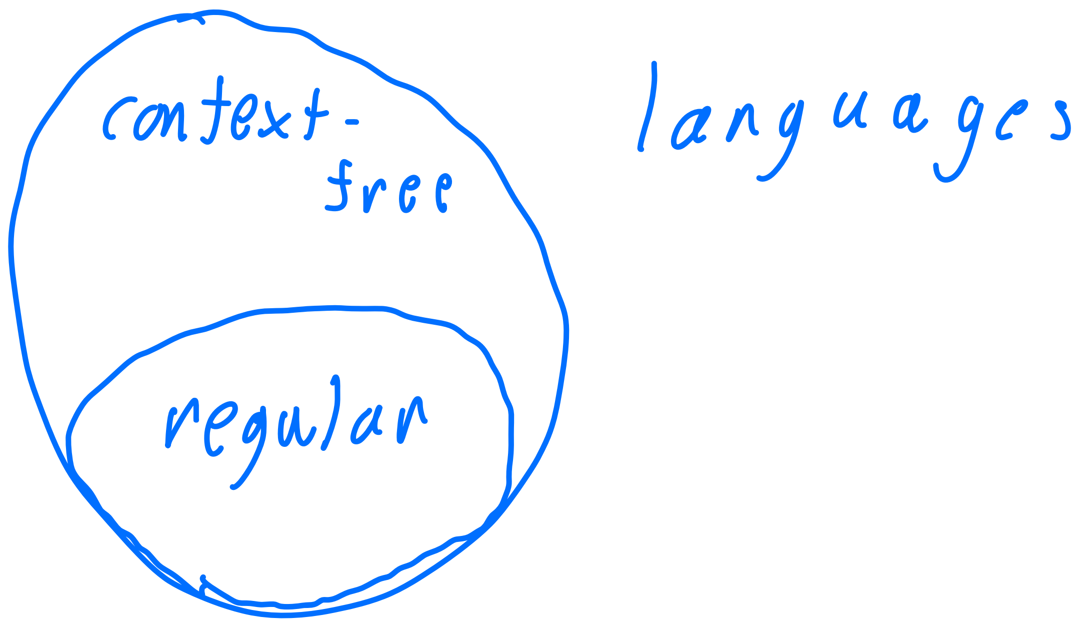
Thm: There are context-free languages that are not regular.

Thm: Not all languages are context-free.

Ex: $\{0^n 1^n 0^n \mid n \geq 0\}$

Ex: $\{ww \mid w \in \{0,1\}^*\}$

But! $\{w \circ w^R \mid w \in \{0,1\}^*\}$ is context-free.



Context-free languages are closed under union, concatenation, & Kleene closure.

Not! generally closed under intersection or complement.

Ex: $S \to A \mid B$

$A \to 0A \mid 0C$

$B \to B1 \mid C1$

$C \to \epsilon \mid 0C1$

$\{0^a 1^b \mid a \neq b\}$

$\{0^a 1^b \mid a > b\}$

$\{0^a 1^b \mid a < b\}$

$L(C) = \{0^n 1^n \mid n \geq 0\}$

Homework: give language of each non-terminal

Ex: balanced parantheses

$S \to \epsilon \mid SS \mid \text{"("}S\text{")"}$ ← is fine for this course!

ambiguous grammar: there are strings with multiple derivation

$(()()())$

$$S \rightsquigarrow (S) \rightsquigarrow (SS) \rightsquigarrow ((S)S) \rightsquigarrow (()S) \rightsquigarrow^{*} (()()())$$

$$S \rightsquigarrow (S) \rightsquigarrow (SS) \rightsquigarrow (S(S)) \rightsquigarrow (S()) \rightsquigarrow^{*} (()()())$$

L is <u>inherently ambiguous</u> if <u>all</u> grammars for L are ambiguous.

$S \to \epsilon \mid (S)S$ so not inherently ambiguous!

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

Want $L(A) = \{0^i 1^i \mid i \geq 0\}$

$A \to \epsilon \mid 0A1$

Want $L(C) = \{1^k 2^k \mid k \geq 0\}$

$C \to \epsilon \mid 1C2$

Want $L(B) = \{1^j \mid j \geq 1\}$

$B \to 1 \mid 1B$

$S \to ABC$

assumed to be starting non-terminal