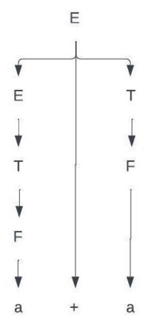
Problem 1: Parse Trees (10)

Let $G = (V, \Sigma, R, S)$ be a grammar with $V = \{E, T, F\}$, $\Sigma = \{+, \times, (,), a\}$ and these rules:

- $E \to E + T | T$
- $T \to T \times F | F$
- $F \to (E) | a$

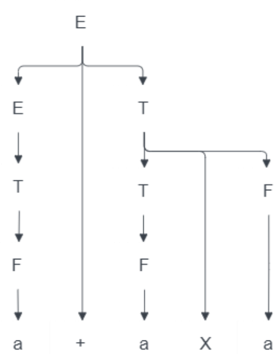
Give parse trees and derivations for each of the following strings:

A. a + a



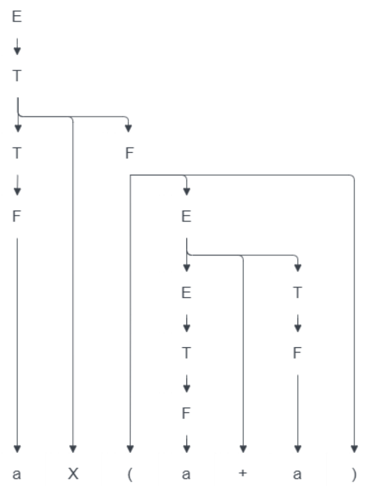
$S \to E \to E + T \to E + F \to E + a \to T + a \to F + a \to a + a$

B. a + a × a



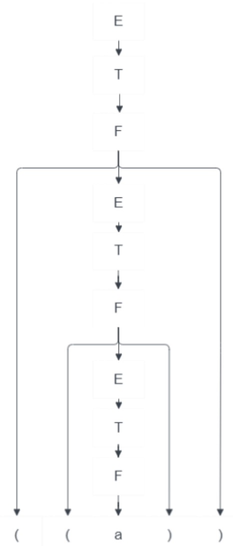
$S \to E \to E + T \to T + T \to F + T \to a + T \to a + T \times F \to a + T \times a \to a + F \times a \to a + a \times a$

C. a × (a + a)



$E \to T \to T \times F \to F \times F \to a \times F \to a \times (E) \to a \times (E) \to a \times (E + T) \to$
$a \times (E + F) \to a \times (E + a) \to a \times (T + a) \to a \times (F + a) \to a \times (a + a)$

D. ((a))



$E \to T \to F \to (E) \to (T) \to (F) \to ((E)) \to ((T)) \to ((F)) \to ((a))$

Problem 2: Context Free Grammar (20)

Give context-free grammars generating the following languages, all with $\Sigma = \{\, \boldsymbol{a}, \boldsymbol{b} \,\}$:

A. $\{\, s \mid s$ starts and ends with the same symbol $\}$

$V = \{A, B, C\}$,
$\Sigma = \{ \boldsymbol{a}, \boldsymbol{b} \}$
$S \in A$
$R = \{A \rightarrow aBa|bBb, \quad B \rightarrow CB|C, \quad C \rightarrow \boldsymbol{a}|\boldsymbol{b} \}$

B. { *s* | the length of *s* is odd }

$V = \{A, B, C\}$,
$\Sigma = \{ \boldsymbol{a}, \boldsymbol{b} \}$
$S \in A$
$R = \{A \rightarrow BCB|C, \quad B \rightarrow BCC|C, \quad C \rightarrow \boldsymbol{a}|\boldsymbol{b} \}$

C. { s | $s = s^R$ }

$V = \{A, B, C\}$,
$\Sigma = \{ \boldsymbol{a}, \boldsymbol{b} \}$
$S \in A$
$R = \{A \rightarrow aBa|bBb|C, \quad B \rightarrow A|C, \quad C \rightarrow \boldsymbol{a}|\boldsymbol{b} \}$

D. { $a^n b^n$ | $n \geq 0$ }

$V = \{A, B, C\}$,
$\Sigma = \{ \boldsymbol{a}, \boldsymbol{b} \}$
$S \in A$
$R = \{A \rightarrow aBb|C, \quad B \rightarrow A|C, \quad C \rightarrow \boldsymbol{ab} \}$

E. { $x_1\#x_2\#\ldots\#x_n$ | $n \geq 0$, each $x_i \in \Sigma^*$ and for some i and j, $x_i = x_j^R$ }

$V = \{A, B, C\}$,
$\Sigma = \{ \boldsymbol{a}, \boldsymbol{b} \}$
$S \in A$
$R = \{$
$A \rightarrow BCD|BED, \quad D \rightarrow \#|Da|Db|\epsilon, \quad B \rightarrow \#|aB|bB|\epsilon$
$C \rightarrow aCa|bCb|a|b|\epsilon \quad E \rightarrow aBa|bBb|\#B$
$\}$

Problem 3: CFLs versus Regular Languages (15)

Let $G = (V, \Sigma, R, S)$ be a grammar with $V = \{S, T, U\}$, $\Sigma = \{0, \#\}$, and the set of rules:

- $S \rightarrow TT|U$
- $T \rightarrow 0T|T0|\#$
- $U \rightarrow 0U00|\#$

A. Describe *L*(*G*) in English.

L(G) is made up of all strings comprised of 0 intersected by at most two # and at least one #. The strings follow the rules where there can only either be one or two #, if there is one # then every time a zero is added on the left of # there will be two added on the right of # (via rule U) or if there are two # there will be any number of zeroes in between, after, or before the two # (via rule T).

B. Prove that *L*(*G*) is not regular.

Assume the language is regular, with $pumping\ length\ = p$. Consider a pumping string $s\ = (0^p\#0^p0^p)$. So, we have $s = xyz$ with x the prefix, y the cycle string and z the suffix, so that:

$PL1: xy^iz \in L\ for\ all\ i \in N\ or\ xy^*z \in L$

$PL2: |y| > 0\ PL3: |xy| \leq p$

By definition of $PL1$, $PL2$ and our pumping string s, all xyz divisions of s must be:

- $x = 0^a$
- $y = 0^b$, where b must be greater than 0
- $z = \#0^p0^p$
  - Since $|xy| \leq p$, the first 0^p must be made of x and y, with the rest made of z
- So, the resulting string xyz should be equal to:
$$0^a0^b\#0^p0^p = 0^{a+b}\#0^p0^p = 0^p\#0^p0^p$$
- by $PL1$, set $i = 2$ and consider $xy^iz = xyyz$
  - Since $xyyz = 0^{a+b}0^b\#0^p0^p$ which simplified is $0^{p+b}\#0^p0^p$
  - That means it violates our language, since there is only one # it must follow the $U \to 0U00$ rule, which $xyyz$ breaks
  - Then by definition of $L, xyyz \notin L, \to\leftarrow PL1$

So, for every possible construction of $\boldsymbol{y}$, $\boldsymbol{xyyz \notin L}$, $\to\leftarrow$ PL1. By contradiction, the language must be irregular.

Problem 4: Chomsky Normal Form, Easy (5)

Convert these rules to CNF.

- $A \rightarrow BAB|B|\epsilon$
- $B \rightarrow bb|\epsilon$

  Add start S and remove ϵ:
  $S_0 \rightarrow A|\epsilon$
  $A \rightarrow BAB|B|A|BB|BA|AB$
  $B \rightarrow bb$

  Remove singe rewrites:
  $S_0 \rightarrow BAB|bb|BB|BA|AB|\epsilon$
  $A \rightarrow BAB|bb|BB|BA|AB$
  $B \rightarrow bb$

  Remove mixed/multiple terminals:
  $S_0 \rightarrow BAB|B_0B_0|BB|BA|AB|\epsilon$
  $A \rightarrow BAB|B_0B_0|BB|BA|AB$
  $B \rightarrow B_0B_0$
  $B_0 \rightarrow b$

  Remove long rewrites to get CNF:
  $S_0 \rightarrow BA_1|B_0B_0|BB|BA|AB|\epsilon$
  $A \rightarrow BA_1|B_0B_0|BB|BA|AB$
  $B \rightarrow B_0B_0$
  $A_1 \rightarrow AB$
  $B_0 \rightarrow b$

Problem 5: Chomsky Normal Form, Less Easy (15)

Convert these rules to CNF.

- $A \rightarrow ABA|B|a|ab$
- $B \rightarrow BCB|C|b|bc|\epsilon$
- $C \rightarrow CD|DC|c$
- $D \rightarrow D|\epsilon$

  Add start S and remove ϵ:
  $S_0 \rightarrow A|\epsilon$
  $A \rightarrow ABA|B|a|ab|AA$
  $B \rightarrow b|bc|B$

  Remove single rewrites:
  $S_0 \rightarrow ABA|c|b|bc|a|ab|AA|\epsilon$
  $A \rightarrow ABA|c|b|bc|a|ab|AA$

$B \rightarrow b|bc$

Remove mixed/multiple terminals:
$S_0 \rightarrow ABA|c|b|B_0C_0|a|A_0B_0|AA|\epsilon$
$A \rightarrow ABA|c|b|B_0C_0|a|A_0B_0|AA$
$B \rightarrow b|B_0C_0$
$A_0 \rightarrow a$
$B_0 \rightarrow b$
$C_0 \rightarrow c$

Remove long rewrites to get CNF:
$S_0 \rightarrow AB_1|c|b|B_0C_0|a|A_0B_0|AA|\epsilon$
$A \rightarrow AB_1|c|b|B_0C_0|a|A_0B_0|AA$
$B \rightarrow b|B_0C_0$
$A_0 \rightarrow a$
$B_0 \rightarrow b$
$C_0 \rightarrow c$
$B_1 \rightarrow BA$

Problem 6: Non-Context-Free Languages (20)

Show that these languages are not context-free:

A. (10) The language of palindromes over {0, 1} containing equal numbers of 0's and 1's.

Assume the language is regular, with $pumping\ length\ = p$. Consider a pumping string $s\ = 1^p0^p1^p$ So, we have $s\ =\ uvxyz$, so that:

$PL1: uv^ixy^iz \in L\ for\ i \in \mathbb{N}$
$PL2: |vy| >\ 0,\qquad PL3: |vxy|\ \leq p$

By definition of $PL1$, $PL2, PL3$ and our pumping string s, all $uvxyz$ divisions of s must be:

- Since $|vxy| \leq p$, by PL3 and PL2
  - Either v, x and y represent one of the digits
  - Either v and x represent one digit and y represent another
  - Or v represents one digit and x and y represents another
- by $PL1$, set $i = 0$ and consider $uv^ixy^iz = uxz$
  - By definition of palindromes, the leftmost 1 has to be equal to the rightmost 1 and 0 must be an even length

- In every case $uvvxyyz$ would mean that at least one number would not be able to follow the definition of palindromes, thus $uvvxyyz$ violates our language
- Then by definition of $L, xyyz \notin L, \rightarrow\leftarrow PL1$

So, for every possible construction of **v** and $\boldsymbol{y}$, $\boldsymbol{uxz} \notin \boldsymbol{L}$, →← PL1. By contradiction, the language must be irregular

B. (10) The language of strings over {1, 2, 3, 4} with equal numbers of 1's and 2's **and** equal numbers of 3's and 4's.

Assume the language is regular, with $pumping\ length = p$. Consider a pumping string $s = 1^p 2^p 3^{2p} 4^{2p}$. So we have $s = uvxyz$, so that:

$PL1: uv^i xy^i z \in L\ for\ i \in \mathbb{N}$
$PL2: |vy| > 0, \qquad PL3: |vxy| \leq p$

By definition of $PL1$, $PL2, PL3$ and our pumping string s, all $uvxyz$ divisions of s must be:

- Since $|vxy| \leq p$, by PL3 and PL2
  - Either v, x, and y represent one symbol
  - Either v represents one symbol and x and y represent another
  - Or v, x represents one symbol and y represents another
- by $PL1$, set $i = 2$ and consider $uv^i xy^i z = uvvxyyz$
  - In every case, $uvvxyyz$ would cause at least one symbol to become unequal with its counterpart (1 and 2 or 3 and 4)
  - Then by definition of $L, xyyz \notin L, \rightarrow\leftarrow PL1$

So, for every possible construction of **v** and $\boldsymbol{y}$, $\boldsymbol{uvvxyyz} \notin \boldsymbol{L}$, →← PL1. By contradiction, the language must be irregular

---

Problem 7: CFG/PDA Generation (15)

Let Σ={0,1} and *L* be the language of all strings with at least one **1** in their second half. Give both a CFG that generates *L* and a PDA that recognizes it.

CFG
$V = \{A, B, C\}$
$\Sigma = \{0,1\}$
$S \in A$
$R = \{A \rightarrow CAC | CB1, \qquad B \rightarrow CBC | \epsilon, \qquad C \rightarrow 0|1|\epsilon\}$

PDA

$Q = \{A, B, C\}$

$\Sigma = \{0,1\}$

$\delta: (q_{l-1}, \epsilon, \epsilon) = \{(r, u_1)\}$

$\Gamma = \{0,1,\$\}$

$q_0 \in Q$

$F \subseteq \{q_2\}$