DERIVATIONS IN REWRITING SYSTEMS

ASSUME STRINGS OVER SOME ALPHABET Γ

FOR GRAMMARS, $\Gamma = V \cup \Sigma$

FOR SEMI-THUE SYSTEMS $\Gamma = \Sigma$

ASSUME RULES $\alpha_i \rightarrow \beta_i \quad 1 \le i \le n$

SIMPLE APPLICATION IS, IF $w = \delta \alpha_i \gamma$

THEN $w \Rightarrow w'$ IF $w' = \delta \beta_i \gamma$

THIS IS A ONE-STEP DERIVATION.

WE EXTEND TO $\stackrel{i}{\Rightarrow}, \stackrel{*}{\Rightarrow}, \stackrel{+}{\Rightarrow}$

$w \stackrel{0}{\Rightarrow} w'$ IFF $w = w'$

$w \stackrel{i+1}{\Rightarrow} w'$ IFF $\exists w'', w \stackrel{i}{\Rightarrow} w'' \Rightarrow w' \quad i \ge 0$

IFF $\exists w'', w \Rightarrow w'' \stackrel{i}{\Rightarrow} w' \quad i \ge 0$

$w \stackrel{*}{\Rightarrow} w'$ IFF $w \stackrel{i}{\Rightarrow} w'$ FOR SOME $i \ge 0$

$\stackrel{*}{\Rightarrow}$ IS THE REFLEXIVE, TRANSITIVE CLOSURE OF $\Rightarrow$

$w \stackrel{+}{\Rightarrow} w'$ IFF $w \stackrel{i}{\Rightarrow} w'$ FOR SOME $i > 0$

$\stackrel{+}{\Rightarrow}$ IS THE TRANSITIVE CLOSURE OF $\Rightarrow$

$\Rightarrow$ AND ITS VARIANTS DENOTE THE OPERATION (OR RELATION) OF DERIVATION

DERIVATIONS IN GRAMMARS

LET $G = (V, \Sigma, R, S)$ $S \in V$

LET $\Rightarrow$ BE THE DERIVATION OPERATION (RELATION) ASSOCIATE WITH R APPLIED TO STRINGS OVER $V \cup \Sigma$

NOTE: ANY STRING OVER $V \cup \Sigma$ IS CALLED A <u>SENTENTIAL FORM</u>

ANY STRING OVER Σ IS A <u>SENTENCE</u>

WE CARE MOSTLY ABOUT SENTENCES DERIVABLE FROM THE TRIVIAL SENTENTIAL FORM S

$$\mathcal{L}(G) = \{ w \mid w \in \Sigma^* \ \& \ S \stackrel{*}{\Rightarrow} w \}$$

CAN LIMIT TO $\stackrel{+}{\Rightarrow}$ SINCE S IS NOT A SENTENCE

DFA TO RIGHTLINEAR

$M = (Q, \Sigma, \delta, q_0, F)$

$G = (Q, \Sigma, R, q_0)$

R: $q \to ap$ WHENEVER $\delta(q, a) = p$

$q \to \lambda$ WHENEVER $q \in F$

PROVE: $\delta^*(q_0, w) \in F \Leftrightarrow q_0 \overset{*}{\Rightarrow} w$

LEMMA: SHOW $\delta^*(q, w) = p \Leftrightarrow q \overset{*}{\Rightarrow} wp$

USE INDUCTION ON $|w|$

BASIS: $|w| = 0$ $\delta^*(q, \lambda) = q$

$q \overset{0}{\Rightarrow} \lambda q$, i.e., $q = \lambda q$

IH: $|w| = k$. ASSUME IF $|w| = k$ THEN

$\delta^*(q, w) = p \Leftrightarrow q \overset{*}{\Rightarrow} wp$

IS: $|w| = k+1$. HENCE $w = xa$, $|x| = k$

BY IH $\delta^*(q, x) = p \Leftrightarrow q \overset{*}{\Rightarrow} x \cdot p$

LET $\delta^*(q, xa) = \delta(\delta^*(q, x), a) = \delta(p, a) = t$

BY CONSTRUCTION $\delta(p, a) = t \Leftrightarrow p \to atER$

THUS, $q \overset{*}{\Rightarrow} xp \Rightarrow xat$

AND SO $q \overset{*}{\Rightarrow} wt$

FINAL STEP DFA TO RIGHT LINEAR

LEMMA SHOWS $\delta^*(q, w) = p \Leftrightarrow q \overset{*}{\Rightarrow} wp$

CLEARLY THEN $\delta^*(q_0, w) = p \Leftrightarrow q_0 \overset{*}{\Rightarrow} wp$

BUT THEN $w \in \mathcal{L}(Q)$ IFF $\delta^*(q_0, w) \in F$

AND SO $w \in \mathcal{L}(Q)$ IFF $q_0 \overset{*}{\Rightarrow} wp$, $p \in F$
IFF $q_0 \overset{*}{\Rightarrow} w$ $(p \to \lambda)$

BUT THEN
$w \in \mathcal{L}(Q) \Leftrightarrow w \in \mathcal{L}(G)$

CONTEXT-FREE GRAMMARS

$G = (V, \Sigma, R, S)$

V: FINITE SET OF VARIABLES (NON-TERMINALS)
Σ: FINITE SET OF TERMINALS
S: START SYMBOL, $S \in V$
R: RULES, EACH OF FORM

$$A \to \alpha \qquad A \in V \quad \alpha \in (V \cup \Sigma)^*$$

EXAMPLE NON-REGULAR CFLS

$S \to aSb \mid \lambda \qquad L = \{a^n b^n \mid n \geq 0\}$

$S \to \lambda \mid a \mid b \mid aSa \mid bSb$

$L = \{$PALINDROMES OVER $\{a, b\}\}$

$S \to aSbS \mid bSaS \mid \lambda$

$L = \{$STRINGS OVER $\{a, b\}$ WITH EQUAL NUMBER OF a'S + b'S $\}$

USING $\overset{k}{\Rightarrow}$ NOTATION

THIS NOTATION IS USEFUL FOR INDUCTIVE PROOFS

CONSIDER $G = (\{S\}, \{a,b\}, R, S)$ WHERE

$R: S \to \lambda \mid a \mid b \mid aSa \mid bSb$

CLAIM: $\mathcal{L}(G) = \{ w \mid w \in \{a,b\}^* \text{ \& } w \text{ IS A PALINDROME} \}$

LEMMA: $S \overset{*}{\Rightarrow} \beta$, WHERE β CONTAINS A NON-TERMINAL IFF $\beta = xSx^R$, WHERE $x \in \{a,b\}^*$. WE ATTACK BY CONSIDERING $S \overset{k}{\Rightarrow} \beta$, SHOWING $\beta = xSx^R$ FOR ANY AND ALL $x \in \{a,b\}^*$, $|x| = k$

BASE: $k=0$, $S \overset{0}{\Rightarrow} S$ BY DEFN. OF $\overset{0}{\Rightarrow}$. BUT $S = \lambda S \lambda^R$ AND THIS IS ONLY STRING OF FORM xSx^R, $|x|=0$

IH: $k=n$. ASSUME $S \overset{n}{\Rightarrow} \beta$ IFF $\beta = xSx^R$ $\forall x, x \in \{a,b\}^*, |x| = n$

IS: $k=n+1$. SHOW $S \overset{n+1}{\Rightarrow} \beta$ IFF $\beta = xSx^R$, $x \in \{a,b\}^*$, $|x| = n+1$

BY DEFN. OF $\overset{LH}{\Rightarrow}$, $S \overset{n+1}{\Rightarrow} \beta$ IFF $S \Rightarrow \alpha \overset{n}{\Rightarrow} \beta$ BY RULES IN R AND OUR CONSTRAINT THAT WE RETAIN A VARIABLE (S) IN THE DERIVATION,

$\alpha = aSa$ OR $\alpha = bSb$

BY IH, $\beta = xSx^R$ FOR ANY AND ALL $x \in \{a,b\}, |x| = n$

COMBINING, WE GET $S \overset{n+1}{\Rightarrow} axSx^Ra$ OR $S \overset{n+1}{\Rightarrow} bxSx^Rb$

ax & bx GIVE US ALL AND ONLY THOSE STRINGS OF LENGTH $n+1$ IN $\{a,b\}^*$. THIS PROVES IS AND HENCE THE ORIGINAL HYPOTHESIS

THEOREM: AS ALL STRINGS IN $\{w \mid w \in \{a,b\}^*, w \text{ A PALINDROME}\}$ ARE OF FORM xx^R, xax^R OR xbx^R GETS APPLYING $S \to \lambda$ OR $S \to a$ OR $S \to b$ GETS ALL AND ONLY THESE FORMS AND THE APPLICATION OF SUCH RULE IS THE ONLY MEANS OF DERIVING A SENTENCE IN $\mathcal{L}(G)$

A PRACTICAL GRAMMAR (SORT OF)

$G = (\{E\}, \{a, +, -, *, /, (,)\}, R, E)$

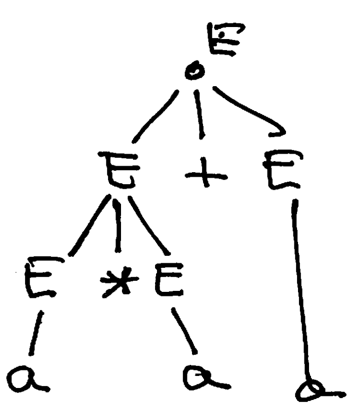
$R: E \rightarrow E+E \mid E-E \mid E \times E \mid E/E \mid (E) \mid a$

TERMINAL FOR VARIABLE OR CONSTANT

A DERIVATION

$E \Rightarrow E+E \Rightarrow E*E+E$
$\Rightarrow a*E+E \Rightarrow a*a+E$
$\Rightarrow a*a+a$

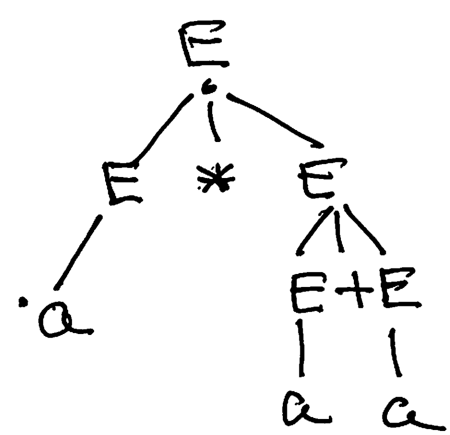
TREE VERSION



FRONTIER IS

$a*a+a$

OOPS: CAN GET BY



SAME FRONTIER

$E \Rightarrow E*E \Rightarrow a*E \Rightarrow a*E+E$
$\Rightarrow a*a+E \Rightarrow a*a+a$

AMBIGUITY

GRAMMAR IS <u>AMBIGUOUS</u> IF THERE IS A STRING, W, IN LANGUAGE SUCH THAT

W CAN BE DERIVED FROM S BY TWO DISTINCT LEFTMOST DERIV.
(ALWAYS REWRITE LEFTMOST NON-TERMINAL BEFORE OTHERS)
$S \overset{*}{\underset{LM}{\Longrightarrow}} w$ & $S \overset{*}{\underset{LM}{\Longrightarrow}} w$ WHERE INTERMEDIATES DIFFER

SAME IDEA

W CAN BE DERIVED FROM S BY TWO DISTINCT RIGHTMOST DERIVATIONS
$S \overset{*}{\underset{RM}{\Longrightarrow}} w$ & $S \overset{*}{\underset{RM}{\Longrightarrow}} w$ WHERE PATHS DIFFER

W HAS TWO DISTINCT PARSE TREES TOPOLOGICALLY DIFFERENT WITH SAME FRONTIER

A LANGUAGE, L, IS <u>INHERENTLY AMBIGUOUS</u> IF ALL GRAMMARS FOR L ARE AMBIGUOUS

ARITHMETIC LANGUAGE IS NOT AMBIGUOUS

ALTERNATIVE GRAMMAR RULES

R: $E \rightarrow E+T \mid E-T \mid T$ LOWEST, LEFT TO RIGHT

$T \rightarrow T*F \mid T/F \mid F$ HIGHER, LEFT TO RIGHT

$F \rightarrow (E) \mid a$

$a*a+a$ CAN ONLY BE GOTTEN BY

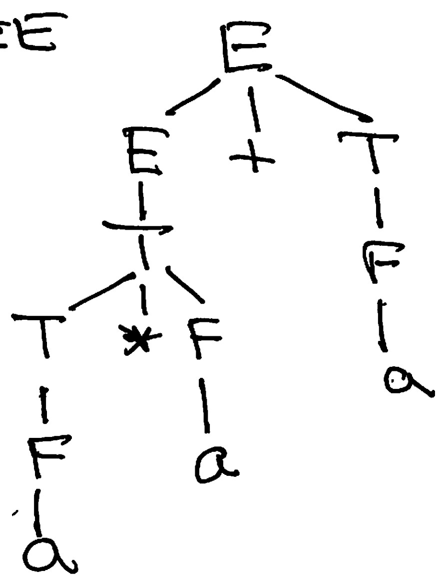
LEFTMOST

$$E \Rightarrow E+T \Rightarrow T+T \Rightarrow T*F+T$$
$$\Rightarrow F*F+T \Rightarrow a*F+T \Rightarrow a*a+T$$
$$\Rightarrow a*a+F \Rightarrow a*a+a$$

RIGHTMOST

$$E \Rightarrow E+T \Rightarrow E+F \Rightarrow E+a \Rightarrow T+a$$
$$\Rightarrow T*F+a \Rightarrow T*a+a$$
$$\Rightarrow F*a+a \Rightarrow a*a+a$$

TREE



AN INHERENTLY AMB. LANGUAGE

$L = \{a^i b^j c^k \mid i = j \text{ OR } j = k\}$

$S \to A\langle BC\rangle \mid \langle AB\rangle C$

$A \to aA \mid \lambda$

$C \to cC \mid \lambda$

$\langle BC\rangle \to b\langle BC\rangle c \mid \lambda$

$\langle AB\rangle \to a\langle AB\rangle b \mid \lambda$

CAN GET $i = j = k$ ON TWO PATHS AND THERE IS NO WAY TO AVOID THIS

SOME EASY CFL CLOSURES

$G_1 = (V_1, \Sigma, R_1, S_1)$ $G_2 = (V_2, \Sigma, R_2, S_2)$ $V_1 \cap V_2 = \emptyset$

UNION $G = (V_1 \cup V_2 \cup \{S\}, \Sigma, R, S)$

$R = R_1 \cup R_2 \cup \{S \to rhs(S_1) \mid rhs(S_2)\}$
OR $\{S \to S_1 \mid S_2\}$

CONCATENATION

$R = R_1 \cup R_2 \cup \{S \to S_1 S_2\}$

$*$ $R = R_1 \cup \{S \to S_1 S \mid \lambda\}$

WE WILL SEE CLOSURE UNDER

INTERSECTION WITH REGULAR

SUBSTITUTION/HOMOMORPHISM

BUT LACK OF CLOSURE UNDER

INTERSECTION WITH CFL

COMPLEMENT

AN INTERESTING CFL

WE WILL PROVE THAT $\{ww \mid w \in \{a,b\}^*\}$
IS <u>NOT</u> A CFL

THE COMPLEMENT OF ABOVE HAS TWO PARTS

(a) ODD LENGTH STRINGS OVER $\{a,b\}$
CLEARLY REGULAR

$S \to aT \mid bT$
$T \to \lambda \mid aS \mid bS$

(b) $\{xy \mid x,y \in \{a,b\}^+ \;\&\; |x|=|y| \;\&\; x \neq y\}$

LOOKING AT (b) WE NEED ONE
TRANSCRITION ERROR FROM X TO Y

IT IS A $\exists$ RATHER THAN A $\forall$
AS IS WW

VIEWING THE STRINGS IN $\{xy \mid x, y \in \{a,b\}^+, |x|=|y|, x \neq y\}$

VIEW 1

$x_1 a x_2 y_1 b y_2$ OR

$x_1 b x_2 y_1 a y_2$

$|x_1| = |y_1| \quad |x_2| = |y_2|$

VIEW 2

$x_1 a y_1 \; x_2 b y_2$ OR

$\underbrace{x_1 b y_1}_{\text{MID}} \; \underbrace{x_2 a y_2}_{\text{MID}}$

$S \to AB \mid BA$

$A \to CAC \mid a$

$B \to CBC \mid b$

$C \to a \mid b$

BOTTOM UP VS TOP DOWN PARSING

BOTTOM UP USES INPUT TO DRIVE PROCESS

IT IS DRIVEN BY SHIFT/REDUCE

SHIFT IS PUSH CHARACTER ON A STACK

REDUCE IS REPLACE "HANDLE" OF TOP OF STACK WITH VARIABLE A WHERE $A \to$ "HANDLE".

THIS IS A REDUCE.

BOTTOM UP HATES RIGHT RECURSION
BUT LOVES LEFT RECURSION

EXAMPLE $E \to E+T \; .. \; T \to F \; ... \; F \to a$

$a + a$

SHIFT a
REDUCE a TO F
REDUCE F TO T
REDUCE T TO E
SHIFT +
SHIFT a
REDUCE a TO F
REDUCE F TO T
REDUCE E+T TO E

TOP DOWN VS BOTTOM UP PARSING

TOP DOWN IS PREDICTIVE AND COMMONLY

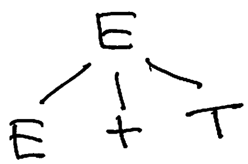
IMPLEMENTED USING RECURSIVE DESCENT

IF HAVE $E \rightarrow E+T \mid E-T \mid E$

WHICH RHS DO WE USE?

LET'S SAY INPUT IS $a+a$, THE PLUS

COULD HELP US PREDICT



THE PROBLEM IS WHEN SEE E AGAIN

WE WOULD DETERMINISTICALLY MAKE

SAME PREDICTION AND GET INFINITE

DESCENT ON E

TOP DOWN HATES LEFT RECURSION

BUT LOVES RIGHT RECURSION

[illegible] ...INED [illegible]AGES ON

[illegible] DIRECTED PARSING

BACK TO ARITHMETIC EXPRESSIONS

$$E \to E+T \mid E-T \mid T$$
$$T \to T*F \mid T/F \mid F$$
$$F \to (E) \mid a$$

IS LEFT RECURSIVE, THIS IS DISASTROUS FOR TOP DOWN

CONSIDER ANY NON-TERMINAL A

WHERE $A \to A\alpha_1 \mid A\alpha_2 \mid \cdots \mid A\alpha_k \mid \underbrace{\beta_1 \mid \cdots \mid \beta_j}_{\text{NOT LEFT REC. IN } A}$

CAN SEE WE GET

$$A \overset{*}{\Rightarrow} (\beta_1 + \beta_2 + \cdots \beta_j)(\alpha_1 + \alpha_2 + \cdots \alpha_k)^*$$

CAN REDO AS

$$A \to \beta_1 A' \mid \cdots \mid \beta_j A'$$
$$A' \to \alpha_1 A' \mid \cdots \mid \alpha_k A' \mid \lambda$$

$$\left.\begin{array}{l} E \to TE' \\ E' \to +TE' \mid -TE' \mid \lambda \end{array}\right\}$$

$$T \to FT'$$
$$T' \to *FT' \mid /FT' \mid \lambda$$
$$F \to (E) \mid a$$