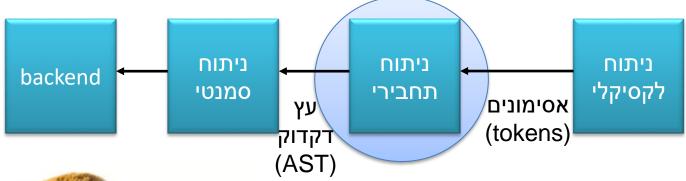
תזכורת מתרגולים אחרונים

• מבנה סכמתי של קומפיילר



- ניתוח תחבירי:
- LL(1) :Top Down
 - :Bottom up -
 - LR(0)
 - SLR •
 - LR(1) •



LR(1)

- SLR-יותר חכם מ lookahead נשתמש ב
- נפרק את מצבי (R(0) למצבים עדינים יותר, המכילים גם מידע על lookahead
 - עידון של ההקשר: מה מגיע אחרי <u>במסלול ספציפי</u> במקום בכל
 מקום אפשרי בשפה

$$A \rightarrow \alpha \bullet B\beta \Rightarrow A \rightarrow \alpha \bullet B\beta, t$$

LR(1) פריטי

- .closure-נעוץ רק בLR(1)-ו LR(0) ההבדל בין פריטי
 - $A \rightarrow \alpha \beta \in P$ כאשר ($A \rightarrow \alpha \bullet \beta$) הוא הוא (LR(0) פריט •
 - על קבוצת פריטים I מוגדר באופן אינדוקטיבי: על קבוצת (closure) אור \bullet
 - closure(I)=I :כסיס
- $(B \rightarrow \bullet \gamma) \in closure(I)$ גם B $\rightarrow \gamma \in P$ אז לכל (A $\rightarrow \alpha \bullet B\beta$) ∈ closure(I) צעד: אם
 - פונקצית המעברים של האוטומט:

$$\delta(I,X) = \bigcup \{closure(A \to \alpha X \bullet \beta) | (A \to \alpha \bullet X \beta) \in I \}$$

- $t \in T \cup \{\$\}$, $A \rightarrow \alpha \beta \in P$ כאשר ($A \rightarrow \alpha \bullet \beta, t$) הוא הוא (LR(1) •
- על קבוצת פריטים ${
 m I}$ מוגדר באופן אינדוקטיבי: (${
 m closure}$
 - closure(I)=I :בסיס
- X ∈ first(βt) אז לכל $B \rightarrow γ ∈ P$ אז לכל ($A \rightarrow α \bullet Bβ, t$) ∈ closure(I) צעד: אם
 - $(B \rightarrow \bullet \gamma, x) \in closure(I)$ גם
 - פונקצית המעברים של האוטומט:

$$\delta(I,X) = \bigcup \{closure(A \to \alpha X \bullet \beta, t) | (A \to \alpha \bullet X \beta, t) \in I \}$$

משמעות פריט (1)LR: זיהינו את מה שמשמאל לנקודה, אנו מצפים למצוא את מה שמימין לה, <u>ולאחר</u> מכן את האסימון המצורף לפריט.

דוגמא - קונפליקט R\R

$$0. S \rightarrow S$$

1. S → Aa

- 2. $S \rightarrow Ba$
- $3. A \rightarrow b$
- 4. B \rightarrow b

SLR	Follow(A)={a		
	Follow(B)={a		

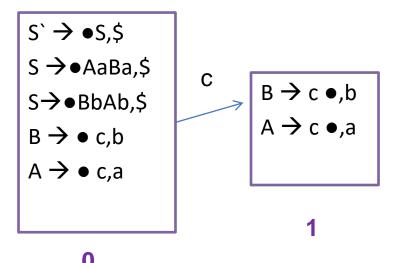
:אבל עבור הדקדוק הבא, עדיין יש בעיה

					` ,
$S^{\rightarrow} \bullet S^{\downarrow}$ $S \rightarrow \bullet Aa$			a	b	\$
S → •Aa	b []	0		s1	
$S \rightarrow \bullet Ba$ $A \rightarrow \bullet b$ $B \rightarrow \bullet b$		1	r3, r4		
$A \rightarrow \bullet b$	$B \rightarrow b \bullet$	•••			
$B \rightarrow \bullet b$					
0	1				

אינו פותר את הקונפליקט.SLRבדוגמא זו מעבר ל

$R\R$ פתרון קונפליקט – LR(1)

- $0. S \rightarrow S$
- 1. S→AaBb
- 2. S→BbAb
- 3. $B \rightarrow c$
- 4. $A \rightarrow c$



LR(1)

	а	b	С	\$
0			s1	
1	r4	r3		
•••				
•••				
•••				

בניית טבלת הניתוח

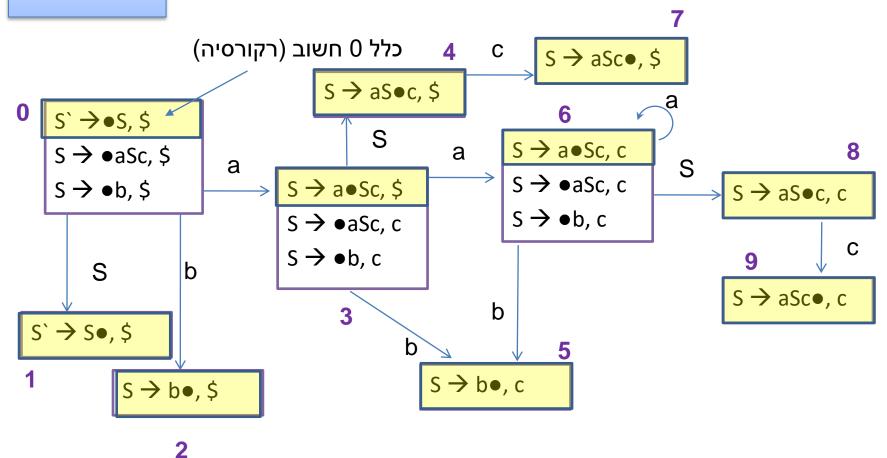
• וטבלת goto כמו goto כמו LR(0)/SLR

```
 \begin{aligned} \text{action[i , t]} &= & \begin{cases} & \text{SHIFT}_j & \delta(I_i \,,\, t) = I_j \\ & \text{REDUCE}_k & \text{rule k is A} \rightarrow \alpha, \quad (A \rightarrow \alpha \bullet, t) \in I_i \\ & \text{ACCEPT} & (S' \rightarrow S \bullet, \$) \in I_i \quad \text{and } t = \$ \\ & \text{ERROR} & \text{otherwise} \end{cases}
```

דוגמא נוספת: רקורסיה

 $2. S \rightarrow b$

1. $S \rightarrow aSc$



^{*}החלקים בצהוב הם הגרעין של המצב

 $0. S' \rightarrow S$

1. S \rightarrow aSc

 $2.S \rightarrow b$

דוגמא (1) LR - הרצה

LR(0)הרצה באופן דומה ל

actions				goto	
	а	b	С	\$	S
0	s3	s2			1
1				acc	
2				r2	
3	s6	s5			4
4			s7		
5			r2		
6	s6	s5			8
7				r1	
8			s9		
9			r1		

מחסנית	פעולה	קלט
(0,)	s3	aabcc\$
(0,)(3,a)	s6	abcc\$
(0,)(3,a)(6,a)	s5	bcc\$
(0,)(3,a)(6,a)(5,b)	r2	cc\$
(0,)(3,a)(6,a)(goto(6,S),S)		
(0,)(3,a)(6,a)(8,S)	s9	cc\$
(0,)(3,a)(6,a)(8,S)(9,c)	r1	c\$
(0,)(3,a)(goto(3,S),S)		
(0,)(3,a)(4,S)	s7	c\$
(0,)(3,a)(4,S)(7,c)	r1	\$
(0,)(goto(0,S),S)		
(0,)(1,S)	acc	\$

בעולם האמיתי Bottom-up

• הכלים הנפוצים לייצור קומפיילרים (compiler • compiler) הם Botoom-up:



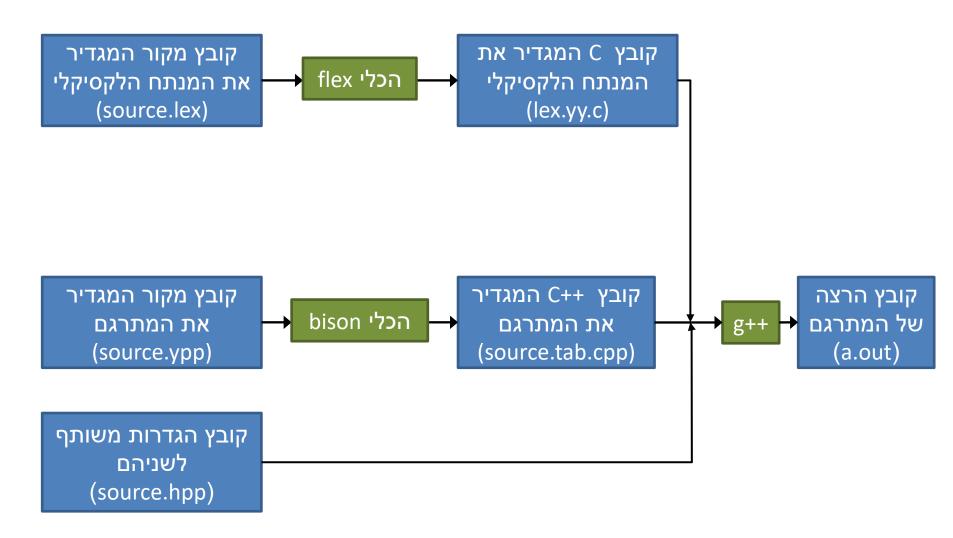




bison

- bison הוא כלי אוטומטי לייצור מהדרים (RTFM). – קלט:
 - קובץ מקור בשפת bison המגדיר את סכימת התרגום.
 - פלט:
 - ניתוח תחבירי ב: שיטת Bottom-up
 - + הרחבה ל: פתרון קונפליקטים.
 - מצפה לניתוח לקסיקלי:
 - (yylex) Lex יכול להיות מסופק ע"י כלי
 - יכול להיות מוגדר בדרכים אחרות

בעזרת bison סכימת השימוש ב



bison המנתח הנוצר על ידי

- מכיל את הפונקציות הבאות:
- . פונקציית הניתוח הלקסיקלי: yylex()
 - .Lex נוצרת ע"י כלי
- . פונקציית הניתוח התחבירי הראשית: yyparse()
 - .bison נוצרת אוטומטית ע"י •
 - מחזירה 0 אם הניתוח הסתיים בהצלחה או 1 אחרת.
- פונקציה ש yyparse קוראת לה בכל פעם: yyerror() שמתגלית שגיאה.
 - המשתמש יכול לדרוס אותה ע"מ לטפל בעצמו בשגיאות.
 - .yyparse מטרתה העיקרית היא קריאה לmain() –
- ניתן לנצלה לשם איתחולים ופעולת סיום מיוחדת של המנתח.

מבנה קובץ הקלט של bison

```
%{
                                       חלק 1:
         C user declarations
%}
declarations
                                       חלק 2:
rules
<u>%%</u>
                                       חלק 3:
C user routines
```

- קובץ מופרד ל3 חלקים: הצהרות, חוקים, וקוד ++C.
- שתי השורות המפרידות בין החלקים חייבות להופיע (מודגשות בקו תחתון).

מבנה קובץ הקלט של מבנה

```
%{
                                      חלק 1:
        user declarations
%}
declarations
%%
                                      חלק 2:
rules
<u>%%</u>
                                      חלק 3:
user routines
```

- על השגרות המופיעות בחלק 3 יש להכריז בחלק 1 בין }% ל- {% .
 - כל הרשום בחלק 3 ובחלק 1 בין }% ל- {% (כתוב בשפת ++3)
 מועתק כמו שהוא לקובץ המקור שנוצר ע"י bison.

דוגמא

```
%{
               #include <iostream>
               #include "calc.h"
               using namespace std;
               #define YYSTYPE Node*
               int yylex();
               void yyerror(const char*);
%}
%token NUM
%left PLUS MINUS
%left MULT
%right LPAR RPAR
%%
               prog line
prog:
               | /*epsilon*/
line: exp '\n' {cout << $1->prettyPrint() << endl; }
               NUM { $$ = $1; }
exp:
                exp PLUS exp \{\$\$ = \text{new Add}(\$1,\$3);\}
                 \exp MINUS \exp \{\$\$ = new Sub(\$1,\$3);\}
                 exp MULT exp \{\$\$ = \text{new Mul}(\$1,\$3);\}
                LPAR exp RPAR \{\$\$ = \$2;\}
%%
void yyerror(const char*) { cout << "syntax error" << endl; }</pre>
int main() {
               return yyparse();
```

declarations

```
%{
                                         חלק 1:
         C user declarations
%}
declarations
<u>%%</u>
                                         :2 חלק
rules
<u>%%</u>
                                         :3 חלק
C user routines
```

- מיועד להצהרה על האסימונים של הדקדוק. •
- בחלק זה גם תוגדר עדיפות לבחירת אסימון. (על כך בהמשך.)

דוגמא

```
%{
               #include <iostream>
               #include "calc.h"
               using namespace std;
               #define YYSTYPE Node*
               int yylex();
               void yyerror(const char*);
%}
%token NUM
%left PLUS MINUS
%left MULT
%right LPAR RPAR
%%
               prog line
prog:
               | /*epsilon*/
line: exp '\n' {cout << $1->prettyPrint() << endl; }
               NUM { $$ = $1; }
exp:
                exp PLUS exp \{\$\$ = \text{new Add}(\$1,\$3);\}
                 \exp MINUS \exp \{\$\$ = new Sub(\$1,\$3);\}
                 exp MULT exp \{\$\$ = \text{new Mul}(\$1,\$3);\}
                LPAR exp RPAR \{\$\$ = \$2;\}
%%
void yyerror(const char*) { cout << "syntax error" << endl; }</pre>
int main() {
               return yyparse();
```

bison-ל flex קישור בין תאום אסימונים

.yylex-ט"י קריאה ל flex מפעיל את המנתח הלקסיקלי של bison מפעיל את המנתח הלקסיקלי של flex צריך לתאם בין האסימונים ש flex מייצר לבין האסימונים ש-פורא.

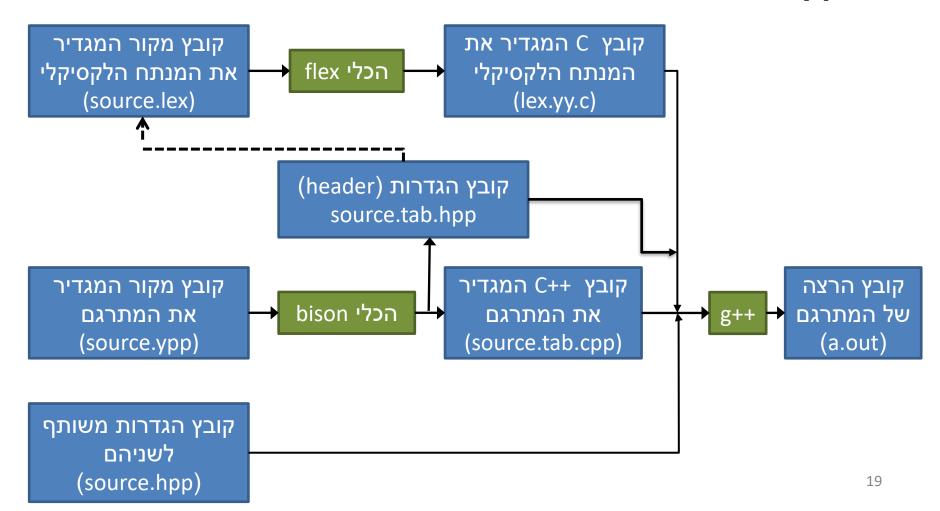
Bison: %token NUM

<u>Lex</u>: [0-9]+ {yylval=new Num(yytext) return **NUM**; }

- כל האסימונים שהוגדרו בקובץ ה bison יופיעו ב header כל האסימונים שהוגדרו בקובץ ה source.tab.hpp יופיעו ב
 - את הקובץ (include) צריך לצרוך (Lex את הקובץ source.tab.hpp על מנת ליצור את התיאום.

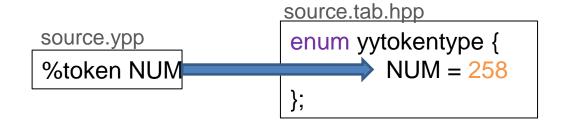
bison-ל flex קישור בין

תיאום בין האסימונים ש-flex מייצר לבין האסימונים ש-bison קורא. ע"י source.tab.hpp



bison-ל flex קישור בין קריאת אסימונים

- :bison •
- מניח את קיומה של yylex.
- המחזירה בכל פעם קבוע המתאים לאחד האסימונים שהוגדרו ע"י token...
- :Lex •
- כל מחרוזת שהוגדרה ב bison ע"י bison מתורגמת לקבוע ב source.tab.hpp
- מנתח לקסיקלי המחזיר את המזהה הקבוע X, יחזיר את אותו הקבוע
 שמוגדר בקובץ ההגדרות, ולכן יפורש ע"י המנתח התחבירי כאסימון



לישור בין flex קישור בין תכונות סמנטיות לאסימונים

```
Bison: %token NUM
```

Lex:
$$[0-9]+ \{yy|val=new Num(yytext);$$

return NUM; }

- yylval משתנה גלובלי **מטיפוס YYSTYPE** משתנה source.tab.hpp
- מכיל את התכונות הסמנטיות של האסימון האחרון שנקרא.

טיפוסי החזרה ב-Bison

• תכונות כל משתני הדקדוק הן מאותו הטיפוס – YYSTYPE.

- זהו גם הטיפוס של המשתנה הגלובלי yylval.
 - ברירת המחדל לטיפוס YYSTYPE היא -
- ניתן לדרוס את YYSTYPE כדי להגדיר תכונות בעלות טיפוסים שונים, (בקובץ source.hpp) ע"י למשל:

חוקים – כללי הגזירה

```
%{
                                   חלק 1:
        C user declarations
%}
declarations
<u>%%</u>
                                  :2 חלק
rules
<u>%%</u>
                                  חלק 3:
C user routines
```

כללי הגזירה

המשתנה באגף שמאל של החוק הראשון הוא המשתנה התחילי של הדקדוק.

זיהוי דקדוק ב-Bison

```
%%
                                    מנתח זה ירוץ על קלט
                                 ויסיים אם הקלט בשפה או
     prog line
prog:
       // /*epsilon*/
                                 אם לא yyerror-יקרא ל
line: exp '\n';
                                כלומר: מזהה דקדוק בלבד
       NUM
exp:
        exp PLUS exp
        exp MINUS exp
        exp MULT exp
        LPAR exp RPAR
%%
```

בניית AST

ל בתור /*action*/ נכתוב • קוד שנרצה להריץ בסוף הכלל

- מועתק (כמעט) כמו שהוא לקובץ הפלט
 - כדי לבנות את ה-AST:
- "להחזיר"(=לשים במחסנית) ערך מכל כלל את תת העץ שהוא מייצג
 - בהתבסס על תתי העצים של הילדים (או חלקם)
 - יש ב-Bison גישה לאיברי הכלל במחסנית •

פעולות עבור כללי הגזירה

- י גישה לאובייקטים (מסוג YYSTYPE) של משתנים וטרמינלים במחסנית:
 - המשתנה המופיע באגף שמאל של החוק \$\$ -
 - הסימן ה-n באגף ימין של החוק (n>0) n

$A \rightarrow B C NUM D E$

\$\$	\$1	\$2	\$3	\$4	\$5	הסימן
Α	В	С	NUM	D	E	מתייחס ל-

ביצוע פעולות

```
%%
                                                                     Reducing exp → exp PLUS exp:
           prog line
prog:
             /*epsilon*/
line : exp '\n' {cout << $1->prettyPrint() << endl; }</pre>
                                                                                     exp
                                                                                     PLUS
             NUM { $$ = $1; }
exp:
              exp PLUS exp {$$ = new Add($1,$3);}
exp MINUS exp {$$ = new Sub($1,$3);}
exp MULT exp {$$ = new Mul($1,$3);}
LPAR exp RPAR {$$ = $2;}
                                                                                     exp
```

%%

ביצוע פעולות

```
%%
                                                                         Reducing \exp \rightarrow \exp PLUS \exp :
             prog line
prog:
              /*epsilon*/
line : exp '\n' {cout << $1->prettyPrint() << endl; }</pre>
              NUM { $$ = $1; }
exp:
               exp PLUS exp {$$ = new Add($1,$3);}
exp MINUS exp {$$ = new Sub($1,$3);}
exp MULT exp {$$ = new Mul($1,$3);}
LPAR exp RPAR {$$ = $2;}
                                                                                          exp
                                                                                                         Add
%%
```

9\$=\$2 ו-2\$=\$2 ו-2\$=\$?

טיפול בדקדוקים רב משמעיים

ישנם אלמנטים בשפות תכנות המתוארים ע"י דקדוק דו • משמעי

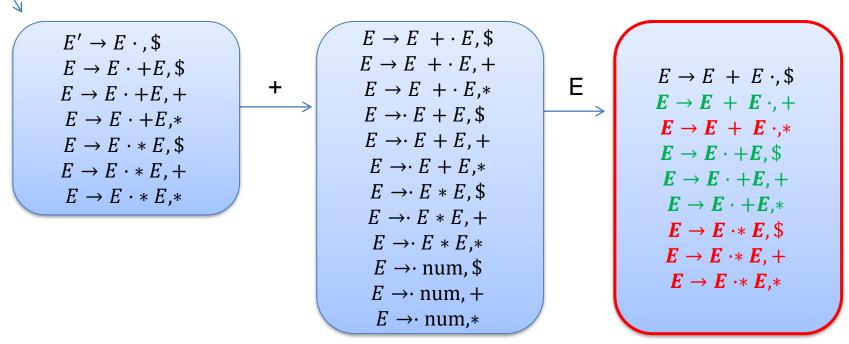
$$E \rightarrow E + E$$
 משמעי.

$$E
ightarrow E \stackrel{*}{-} E = 0$$
 לדוגמא, ביטויים אריתמטיים עם חיבור וכפל: $E
ightarrow num$

- דקדוק זה הוא דו משמעי על "1+2*3". לאחר קריאת "1+2" ישנן 2 התנהגויות:
- ר בצע shift, המשך לקרוא את הקלט ובצע shift. בצע הפלישי.
 - 1+(2*3) מתאים לפירוש
 - .2 בצע reduce ע"פ הכלל הראשון.
 - (1+2)*3 מתאים לפירוש

טיפול בדקדוקים רב משמעיים

ציור חלקי של האוטומט עבור הדקדוק: –



- במצב המסומן ישנו קונפליקט עבור הטרמינל * המסומן באדום.
 - בנוסף, ישנו קונפליקט גם עבור הטרמינל + המסומן בירוק.
 - מתאים לשני הפירושים של המחרוזת "1+2+3":
- . (עבור חיסור כן)+1+(2+3) או (2+3)+1. (עבור חיבור זה לא משנה, אבל עבור חיסור כן).

קדימות אופרטורים

- פתרון לקונפליקט בין כפל לחיבור: (2*3)+1 לעומת 3*(2+1): עדיפות לאסימונים
- . '+' שלאסימון '*' יש עדיפות גדולה מאשר לאסימון bison) ניתן לציין ל
 - . מבוצע ע"י סדר ההכרזה על אסימונים בקובץ ההגדרות.
 - האסימון בעל העדיפות הגבוהה ביותר, יוכרז אחרון.

<u>עדיפות לכללי גזירה.</u>

- ברירת המחדל היא עדיפות האסימון האחרון המופיע בכלל.
 - ביתן לציין עדיפות באופן מפורש ע"י שימוש ב prec t"

```
כאשר t הוא האסימון שאת עדיפותו רוצים לתת לכלל (ניתן להוסיף אסימון חדש למטרה זו).
```

```
%token '+'
%token '*'
%%
E: E '+' E
{...}
    | E '*' E
{...};
```

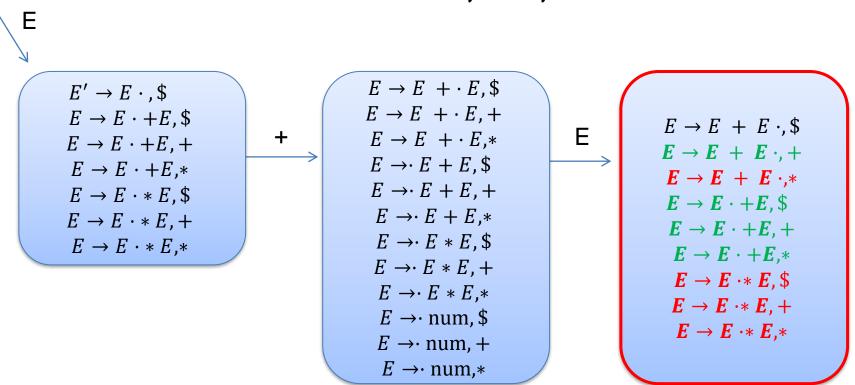
קדימות אופרטורים

- : (1+2)*3 לעומת 1+(2*3)•
 - עבור shift :shift/reduce מוצא קונפליקט bison אשר ר רמינל t אל מול t טרמינל t טרמינל t טרמינל אל מול
 - $A \rightarrow \alpha$ אם לטרמינל t אם אם shift יבוצע shift יבוצע •
 - אם לכלל עדיפות גבוהה מאשר לטרמינל. reduce יבוצע
 - אחרת הקונפליקט לא נפתר. •
- shift בדוגמא, במצב הבעייתי, קונפליקט זה ייפתר לטובת ה בדוגמא, במצב הבעייתי. E→E+E מאחר ולאסימון * עדיפות גדולה מאשר לכלל
 - התקבלה הפרשנות הראשונה.

```
%token '+'
%token '*'
%%
E: E '+' E {...}
| E '*' E {...};
```

למעשה בדוגמא זו צריך היה להשתמש ב אחסחassoc במקום ב token במקום ב קדימות. נדבר על כך בהמשך.

טיפול בדקדוקים רב משמעיים



.+ נראה כעת כיצד מתגברים על הקונפליקט עבור הטרמינל

אסוציאטיביות

: + פתרון לקונפליקט עבור הטרמינל

- לא ניתן להכריע בעזרת עדיפות מאחר ולאסימון '+' אותה העדיפות
 E→E+E כמו לכלל
 - . ניתן להגדיר את האסימון '+' כבעל אסוציאטיביות שמאלית
 - %token במקום עם left יוכרז עם •

```
%left '+'
%token '*'
%%
E: E '+' E {...}
| E '*' E {...};
```

אסוציאטיביות

: + פתרון לקונפליקט עבור הטרמינל

- יכריע bison אינו נפתר בעזרת קדימויות, S/R כאשר קונפליקט באופן הבא:
 - .reduce אם לאסימון אסוציאטיביות שמאלית יועדף ה
 - .shift אם לאסימון אסוציאטיביות ימנית יועדף ה
 - במקרה שלנו יועדף הreduce ונקבל את הפרשנות השנייה.

```
%left '+'
%token '*'
%%
E: E '+' E {...}
| E '*' E {...};
```

אסוציאטיביות

- הערה: עדיפות מוגדרת רק עבור אסימונים שהוכרזו עם token אסוציאטיביות, לא עבור אסימונים שהוגדרו ע"י
- כדי להגדיר אסימון עם עדיפות שאין %nonassoc ניתן להשתמש ב לו אסוציאטיביות.
 - שאלה: לאילו אופרטורים מתמטיים נרצה אסוציאטיביותימנית?
 - a^b^c :- למשל אופרטור החזקה
 - b ? a : b :- למשל תנאי טרינארי

טיפול בדקדוקים רב משמעיים

- :shift/reduce סיכום קונפליקט •
- :t אם לכלל הגזירה עדיפות r ולטרמינל הקלט עדיפות
 - .reduce יבוצע r>t
 - .shift יבוצע r<t •
- אם r=t הפעולה תיקבע ע"פ האסוציאטיביות של הכלל (זהה לאסוציאטיביות הטרמינל):

reduce אסוציאטיביות שמאלית – אסוציאטיביות שמאלית –

. shift יבוצע – אסוציאטיביות ימנית –

אסוציאטיביות אינה מוגדרת שגיאה. –

- :reduce/reduce קונפליקט
- התנהגות ברירת המחדל של bison היא לבחור בכלל הגזירהשהופיע קודם בדקדוק.
 - . מומלץ לשנות את הדקדוק כך שיתמודד עם קונפליקט זה.

עבודה עם bison

- bison –d source.ypp *. tab.h, *.tab.cpp מייצר קבצי
- flex source.lex lex.yy.c מייצר קובץ
- g++ source.tab.cpp lex.yy.c

• bison –v source.ypp source.output מייצר קובץ עם מצבי האוטומט וטבלת המעברים. ניתן להשתמש באפשרות זאת על מנת לראות את הקונפליקטים, במידה וקיימים.

ישנו חומר עזר נוסף באתר: bison07_tips_bison.pdf Lex and Bison Examples and Guides(Course Material)

שאלות ממבחנים

שאלה 1

נתון G=(V, T, P, S) השייך ל LR(1). עבור כל אחד מהדקדוקים הבאים, רשמו האם הוא **בהכרח** ב LR(1):

$$t \in T$$
 , $X \notin V$ כאשר $G_1 = (V \cup \{X\}, T, P \cup \{X \rightarrow t\}, S)$. א.

$$G_2 = (V,T,P \cup \{S \rightarrow \epsilon\},S)$$
 .

שאלה 1

$$t \in T,X \notin V$$
 כאשר

$$G_1 = (V \cup \{X\}, T,P \cup \{X \rightarrow t\}, S)$$
 . A

- :LR(1) בהכרח ב•
- המשתנה X הוא משתנה חדש שאינו נגזר באגף ימין של אף כלל.
 - X→t בזמן ביצוע closure במצבי האוטומט, הכלל לעולם לא ייכנס.
 - האוטומט יהיה זהה לאוטומט של G, שהוא חסר קונפליקטים.

! משתנה לא ישיג X

שאלה 1

$$G_2 = (V, T, P \cup \{S \rightarrow \varepsilon\}, S)$$
 .

- .LR(1) אינו בהכרח ב •
- הכלל החדש יכול להפוך את הדקדוק לדו משמעי.לדוגמא:

$$G = \{S \to A, A \to \varepsilon\}$$

$$G_2 = \{S \to A, A \to \varepsilon, S \to \varepsilon\}$$

דקדוק דו משמעי: יש שתי גזירות ימניות ביותר בדקדוק לאותה מילה.

שאלה - אסוציאטיביות

```
응 {
   #include <stdio.h>
   #include <iostream>
   using namespace std;
   int yylex();
   void yyerror(const char*);
} %
 left 'a'
 right 'b'
응응
   : S'\n'
   : S 'a' S {cout << '1';}
      S 'b' S {cout << '2';}
      'c' {cout << '3';}
           {cout << '4';}
     /*epsilon*/ ;
```

נתונה תוכנית ה bison הבאה:

```
int yylex() {
    return getchar();
}

void yyerror(const char*) {
    cout << '5';
}

int main() {
    return yyparse();
}</pre>
```

כתבו מה יהיה פלט התוכנית עבור כל אחד מהקלטים הבאים (כל שורה מסתיימת בסימן ירידת שורה 'n'י):

1. cbdbc

응응

2. cadbc