Recursive Descent part 2

Nullable-First-Follow

Recursive Descent

ניתוח רקורסיבי יורד

תבנית כללית לפונקצית Parse_X

```
Parse_X()
{
        t=next_token();
        switch(t) {
                case t1:  ……... break;
                case t2: ………... break;
                …
                case tk: ………... break;

                default: error();
        };
}
```

לכל אחד מהאסימונים ב- First(X)
בהתאם לכלל הגזירה שמתחיל
באותו אסימון

מכיוון שאין גורמים שמאליים
ישירים או עקיפים:
אין שני כללים שמתחילים
באותו אסימון

תבנית כללית לפונקציית Parse_X

```
Parse_X()
{
      t=next_token();
      switch(t) {
            case a:  ……... break;
            case b: ………... break;
            …
            …
            case u: case v: ….  : back_token(); break;
            default: error();
      };
}
```

לכל אחד מהאסימונים ב- First(X)
בהתאם לכלל הגזירה שמתחיל
באותו אסימון

אם *X* אפיס:
לכל אחד מהאסימונים
ב-Follow(X)

שגיאות תחביר

$$int\ x;$$
$$int\ y;$$
$$real\ c,$$
$$real\ r;$$

- דוגמא:

מהי בעצם השגיאה??

- במקום נקודה-פסיק מופיע פסיק.
- הפסיק בסדר והשגיאה בהמשך. הכוונה היתה ל- $real\ c, real r;$
- משהו אחר?

התאוששות משגיאות

- כאשר יש שגיאה תחבירית, לא ניתן לגזור את כל הקלט מהמשתנה ההתחלתי.
- Panic mode – הפסקה מוחלטת של תהליך הקומפילציה.
- ממשיכים בניתוח התחבירי, אבל:
  - מאיזו נקודה בקלט ממשיכים?
  - לפי איזו תחזית ממשיכים בניתוח?
- אסטרטגיות שונות.
  - באף אסטרטגיה לא **מובטחת** התאוששות מוצלחת. זו רק אסטרטגיה, שעלולה גם להיכשל.

רעיונות שונים לפתרון

- אם נתקלים באסימון שלא ציפינו לו, נתייחס אליו כאילו הוא כן האסימון שציפינו לו ונמשיך כרגיל מאותה נקודה (כמובן שנדווח על השגיאה).
זוהי אסטרטגיה טובה למשל במקרה שמצפים למילה שמורה ובמקומה מופיע ID , אבל מספיק "קרוב" למה שציפינו לו (למשל $wile$ במקום $whie$).
ואז ההנחה היא שזה לא ID אלא המילה השמורה החסרה.

- אם ראינו משהו שלא ציפינו לו, נפסיק את נסיון הניתוח מהמשתנה הנוכחי ונדלג על קטע שלם, עד שנמצא נקודה שנוכל להמשיך ממנה לפי התחזית הנוכחית.
נניח שהפרסר מנסה לגזור משהו מהמשתנה STATEMENT ונתקל באסימון לא-צפוי. יכול להיות שכדאי לדלג עד לאסימון **;** ומשם להמשיך כפי שהיינו ממשיכים מאותו מקום אילו לא היתה שגיאה.

מימוש אסטרטגיה:
דילוג עד ל- Follow(CurrentVar)

```
Error_Recoverry(X)
{       repeat{
                t = next_token()
        } until (t ∈ FOLLOW(X)  OR  t == TOK_EOF )
        back_token();
}
```

בהמשך:
למימוש
בפרוייקט

Nullable, First, Follow

- כפי שראינו, כאשר כותבים את הפונקציה $Parse\_X$ יש לדעת:
  - $Nullable(X)$: האם X אפיס
  - $First(X)$: איזה אסימונים יכולים להופיע בתחילת משהו שנגזר מ-X
  - $Follow(X)$: איזה אסימונים יכולים להופיע מייד אחרי משהו שנגזר מ- X

Nullable / First / Follow

סימונים : $\overset{*}{\Rightarrow}$ $\Rightarrow$

- $A \Rightarrow \alpha$: ניתן לגזור מ- A את α בצעד גזירה אחד (זה קורה רק אם יש כלל גזירה $A \rightarrow \alpha$).
- $A \overset{*}{\Rightarrow} \alpha$: ניתן לגזור מ- A את α (באפס או יותר צעדים)
- דוגמא:

- $S \rightarrow aSb \mid \# \mid \varepsilon$

$S \overset{*}{\Rightarrow} \#$, $S \overset{*}{\Rightarrow} aaa\#bbb, S \overset{*}{\Rightarrow} aaaaaaSbbbbbb$, ……

משתנים אפיסים (Nullable)

- $\boldsymbol{Nullable(X) = True} \quad \Longleftrightarrow \quad \boldsymbol{X \overset{*}{\Rightarrow} \varepsilon}$
- אם קיים כלל גזירה $\varepsilon \to X$, אזי ברור ש- X אפיס.
- זהו??
- נתבונן בדקדוק הבא.
- האם רק C אפיס?

$$
\begin{aligned}
\boldsymbol{S} &\to \boldsymbol{ab} \mid \boldsymbol{BC} \mid \boldsymbol{D} \\
\boldsymbol{B} &\to \boldsymbol{bB} \mid \boldsymbol{C} \\
\boldsymbol{C} &\to \boldsymbol{cC} \mid \boldsymbol{\varepsilon} \\
\boldsymbol{D} &\to \boldsymbol{aD} \mid \boldsymbol{cc}
\end{aligned}
$$

משתנים אפיסים (Nullable)

- $\boldsymbol{Nullable(X) = True} \quad \Longleftrightarrow \quad \boldsymbol{X \overset{*}{\Rightarrow} \varepsilon}$

- אם קיים כלל גזירה $\varepsilon \rightarrow X$, אזי ברור ש- X אפיס.
- אם קיים כלל גזירה $X \rightarrow Y_1 Y_2 \cdots Y_k$ כאשר המשתנים $Y_1, Y_2, \cdots, Y_k$ הם אפיסים, אז גם המשתנה X אפיס.

אלגוריתם למציאת כל המשתנים האפיסים

1. איתחול: לכל משתנה X:
   - $Nullable(X) = False$
   - אם קיים כלל גזירה $X \to \varepsilon$, אז $Nullable(X) = True$
2. לכל כלל גזירה מהצורה $X \to Y_1 Y_2 \cdots Y_k$:
   - אם לכל $1 \le i \le k$ מתקיים $Nullable(Y_i) = True$,
     אז $Nullable(X) = True$
3. חזור על 2 כל עוד יש שינוי

דוגמא

$$S \rightarrow ab \mid BC \mid D$$
$$B \rightarrow bB \mid C$$
$$C \rightarrow cC \mid \varepsilon$$
$$D \rightarrow aD \mid cc$$

	איתחול	איטרציה 1	איטרציה 2	איטרציה 3
S	$-$	$-$	$+$	$+$
B	$-$	$+$	$+$	$+$
C	$+$	$+$	$+$	$+$
D	$-$	$-$	$-$	$-$

האם האלגוריתם הזה סופי??

- האם מספר האיטרציות סופי? כמה איטרציות לכל היותר יהיו??
- חוזרים על שלב 2 כל עוד היה שינוי כלשהו.
- בכל שינוי, מתווסף לפחות משתנה אחד אל "קבוצת האפיסים".
- מספר האיטרציות חסום ע"י מספר המשתנים, $|V|$. כלומר, סופי.

$First(X)$

$$\boldsymbol{First(X) = \left\{t \in T \;\mid X \overset{*}{\Rightarrow} t\alpha\right\}}$$

$First(X) =$ קבוצת האסימונים שיכולה להופיע בהתחלה של משהו שנגזר מ-X

- אם קיים כלל גזירה $X \to t\alpha$ כאשר t הוא אסימון:
  אז $t \in First(X)$.
- אם קיים כלל גזירה $X \to Y\alpha$ כאשר Y הוא משתנה:
  אז $First(Y) \subseteq First(X)$.
- זהו??

$First(X)$

נתבונן בדקדוק הבא:

$$
\begin{aligned}
&\boldsymbol{S \rightarrow ab \mid CB \mid Dc} \\
&\boldsymbol{B \rightarrow bB \mid D} \\
&\boldsymbol{C \rightarrow cC \mid \varepsilon} \\
&\boldsymbol{D \rightarrow dDCe \mid ec}
\end{aligned}
$$

- איזה אסימונים יכולים להופיע בתחילת משהו שנגזר מ-S ?
- a בגלל הכלל $S \rightarrow ab$
- כל האסימונים ב- $First(C)$ וב- $First(D)$ בגלל הכללים $S \rightarrow CB$ ו- $S \rightarrow Dc$.
- מה לגבי האסימון b למשל?

$$S \Longrightarrow CB \Longrightarrow B \Longrightarrow bB \Longrightarrow \dots\dots$$

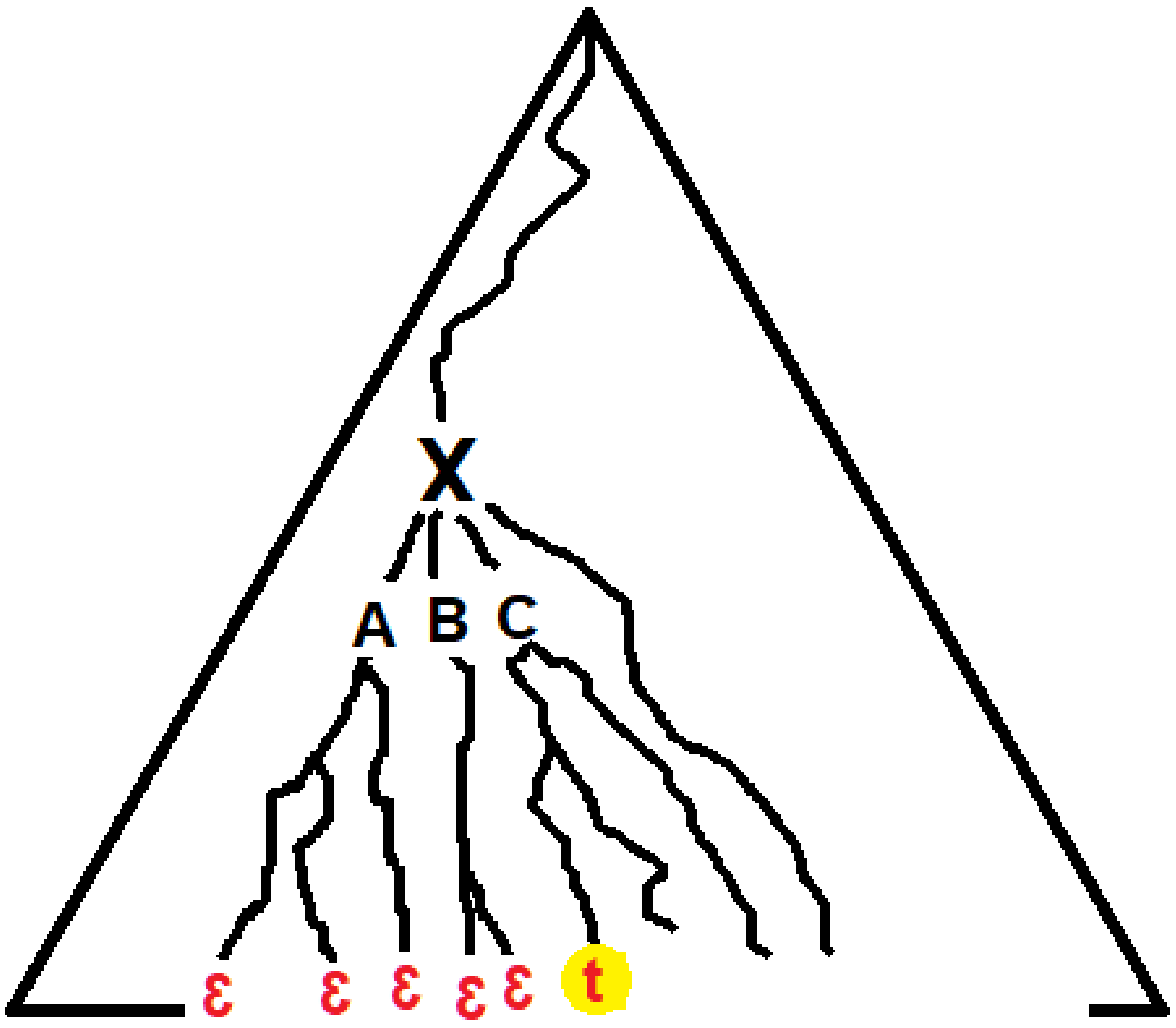
$First(X)$

1. אם קיים כלל גזירה $X \to t\alpha$ כאשר t הוא טרמינל:
   אז $t \in First(X)$.

2. אם קיים כלל גזירה $X \to Y\alpha$ כאשר Y הוא משתנה:
   אז $First(Y) \subseteq First(X)$.

3. אם קיים כלל גזירה מהצורה $X \to Y_1 Y_2 \cdots Y_k t\alpha$ כאשר $Y_1, Y_2, \cdots, Y_k$ אפיסים:
   אז $t \in First(X)$.

4. אם קיים כלל גזירה מהצורה $X \to Y_1 Y_2 \cdots Y_k Y\alpha$ כאשר $Y_1, Y_2, \cdots, Y_k$ אפיסים:
   אז $First(Y) \subseteq First(X)$.



X
A B C
ε ε ε ε ε t

אלגוריתם לחישוב קבוצות $First$

1. איתחול:
   - לכל משתנה X: $First(X) = \emptyset$
2. עבור על כללי הגזירה ובצע (בהתאם לצורה של כל כלל):
   1. $\boldsymbol{X \to t\alpha}$: הוסף את t ל- $First(X)$
   2. $\boldsymbol{X \to Y\alpha}$: הוסף את $First(Y)$ ל- $First(X)$
   3. $\boldsymbol{X \to Y_1Y_2 \cdots Y_kt\alpha}$ **כאשר** $\boldsymbol{Y_1, Y_2, \cdots, Y_k}$ **אפיסים** : הוסף את t ל- $First(X)$
   4. $\boldsymbol{X \to Y_1Y_2 \cdots Y_kY\alpha}$ **כאשר** $\boldsymbol{Y_1, Y_2, \cdots, Y_k}$ **אפיסים** : הוסף את $First(Y)$ ל- $First(X)$
3. חזור על 2 כל עוד היה שינוי

דוגמא

$$S \rightarrow \overset{1}{ab} \mid \overset{2}{CB} \mid \overset{3}{Dc}$$

$$B \rightarrow \overset{4}{bB} \mid \overset{5}{D}$$

$$C \rightarrow \overset{6}{cC} \mid \overset{7}{\varepsilon}$$

$$D \rightarrow \overset{8}{dDCe} \mid \overset{9}{ec}$$

ε אינו אסימון!
לכן אינו שייך
לאף קבוצת First

	Nullable	First
S	$-$	a , c, b, d, e
B	$-$	b, d, e
C	$+$	c
D	$-$	d , e

האם האלגוריתם הזה סופי??

- האם מספר האיטרציות סופי? כמה איטרציות לכל היותר יהיו??
- חוזרים על שלב 2 כל עוד היה שינוי כלשהו.
- בכל שינוי, מתווסף לפחות אסימון אחד אל אחת מקבוצות ה-$First$
- יש $|V|$ קבוצות $First$ וגודלה של כל אחת הוא לכל היותר $|T|$
- מספר האיטרציות חסום ע"י $|T| \times |V|$. כלומר, סופי.

אלגוריתם לחישוב קבוצות $First$

1. איתחול:
   - לכל משתנה X: $First(X) = \emptyset$
2. עבור על כללי הגזירה ובצע (בהתאם לצורה של כל כלל):
   1. $X \to t\alpha$: הוסף את t ל- $First(X)$
   2. $X \to Y\alpha$: הוסף את $First(Y)$ ל- $First(X)$
   3. **$X \to Y_1Y_2\cdots Y_kt\alpha$ כאשר $Y_1,Y_2,\cdots,Y_k$ אפיסים** : הוסף את t ל- $First(X)$
   4. **$X \to Y_1Y_2\cdots Y_kY\alpha$ כאשר $Y_1,Y_2,\cdots,Y_k$ אפיסים** : הוסף את $First(Y)$ ל- $First(X)$
3. חזור על 2 כל עוד היה שינוי

מקרים פרטיים K=0

$X \rightarrow Y_1 Y_2 \cdots Y_k Y \alpha$ כאשר $Y_1, Y_2, \cdots, Y_k$ אפיסים

- משמעותו של כלל 4:

אם יש כלל גזירה מהצורה: רצף של משתנים אפיסים ואחריהם משתנה $\boldsymbol{X \rightarrow Y}$
אז הוסף את $First(Y)$ אל $First(X)$.

דוגמא

- $X \rightarrow ABCDY\alpha$ וכן A,B,C,D אפיסים.
- לפי כלל 4, יש להוסיף ל- $First(X)$ את $First(Y)$.

אבל גם ABC היא תחילית אפיסה (ואז D "משחק" בתפקיד Y, כלומר הוא משתנה שנמצא מייד אחרי רצף של משתנים אפיסים) וכן הלאה...

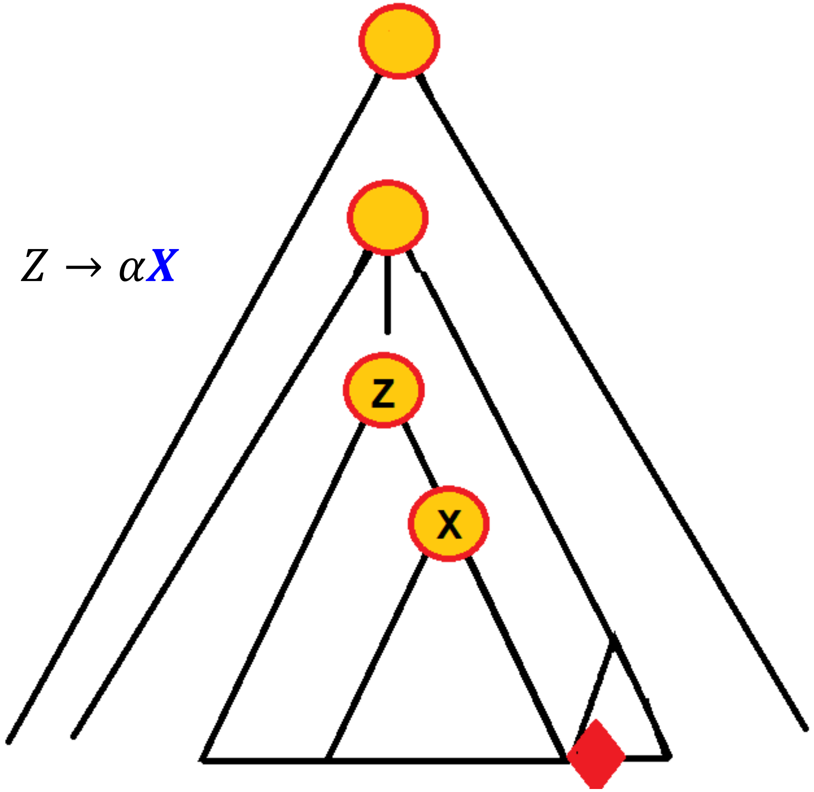
לכן יש להוסיף ל- $First(X)$ גם את:

$First(A)$, $First(B)$, $First(C)$, $First(D)$

$ = EOF

$Follow(X)$

$Follow(X)$ = קבוצת האסימונים שיכולים להופיע מייד אחרי משהו שנגזר מ-X



- תמיד מתקיים $\$ \in Follow(S)$
- אם $Z \rightarrow \alpha \boldsymbol{X} t\beta$ אז $t \in Follow(X)$
- אם $Z \rightarrow \alpha \boldsymbol{X} Y\beta$ אז $First(Y) \subseteq Follow(X)$
- אם $Z \rightarrow \alpha \boldsymbol{X}$

  אז $Follow(Z) \subseteq Follow(X)$

$Follow(X)$

- אם $Z \to \alpha \boldsymbol{X} Y_1 Y_2 \cdots Y_k t \beta$ כאשר $Y_1, Y_2, \cdots, Y_k$ אפיסים. אז:
  - לכל $1 \le i \le k$ $First(Y_i) \subseteq Follow(X)$
  - $t \in Follow(X)$

- אם $Z \to \alpha \boldsymbol{X} Y_1 Y_2 \cdots Y_k Y \beta$ כאשר $Y_1, Y_2, \cdots, Y_k$ אפיסים. אז:
  - לכל $1 \le i \le k$ $First(Y_i) \subseteq Follow(X)$
  - $First(Y) \subseteq Follow(X)$

$Follow(X)$

- אם $Z \to \alpha \boldsymbol{X} Y_1 Y_2 \cdots Y_k$ כאשר $Y_1, Y_2, \cdots, Y_k$ אפיסים. אז:
  - לכל $1 \le i \le k$ $First(Y_i) \subseteq Follow(X)$
  - $Follow(Z) \subseteq Follow(X)$

אלגוריתם לחישוב קבוצות $Follow$

1. איתחול:
   - לכל משתנה X: $Follow(X) = \emptyset$
   - $Follow(S) = \{\$\}$

2. עבור על כללי הגזירה ובצע (בהתאם לצורה של כל כלל):
   1. $Z \rightarrow \alpha X t \beta$: הוסף את t ל- $Follow(X)$
   2. $Z \rightarrow \alpha X Y \beta$: הוסף את $First(Y)$ ל- $Follow(X)$
   3. $Z \rightarrow \alpha X$: הוסף את $Follow(Z)$ ל- $Follow(X)$

< המשך בשקף הבא >

4. $Z \rightarrow \alpha X Y_1 Y_2 \cdots Y_k t \beta$ כאשר $Y_1, Y_2, \cdots, Y_k$ אפיסים:
   - לכל $1 \leq i \leq k$: הוסף את $First(Y_i)$ ל- $Follow(X)$
   - הוסף את t ל- $Follow(X)$
5. $Z \rightarrow \alpha X Y_1 Y_2 \cdots Y_k Y \beta$ כאשר $Y_1, Y_2, \cdots, Y_k$ אפיסים:
   - לכל $1 \leq i \leq k$: הוסף את $First(Y_i)$ ל- $Follow(X)$
   - הוסף את $First(Y)$ ל- $Follow(X)$
6. $Z \rightarrow \alpha X Y_1 Y_2 \cdots Y_k$ כאשר $Y_1, Y_2, \cdots, Y_k$ אפיסים:
   - לכל $1 \leq i \leq k$: הוסף את $First(Y_i)$ ל- $Follow(X)$
   - הוסף את $Follow(Z)$ ל- $Follow(X)$

מקרים 1,2,3 הם מקרים פרטיים של 4,5,6 עבור K=0

3. חזור על 2 כל עוד היה שינוי

דוגמא

$$S \to \overset{1}{ab} \mid \overset{2}{CB} \mid \overset{3}{Dc}$$

$$B \to \overset{4}{bB} \mid \overset{5}{D}$$

$$C \to \overset{6}{cC} \mid \overset{7}{\varepsilon}$$

$$D \to \overset{8}{dDCe} \mid \overset{9}{ec}$$

	Nullable	First	Follow
S	$-$	a , c, b, d, e	$\$$
B	$-$	b, d, e	$\$$
C	$+$	c	b, d, e
D	$-$	d , e	$c, e, \$$

האם האלגוריתם הזה סופי??

- האם מספר האיטרציות סופי? כמה איטרציות לכל היותר יהיו??
- חוזרים על שלב 2 כל עוד היה שינוי כלשהו.
- בכל שינוי, מתווסף לפחות אסימון אחד אל אחת מקבוצות ה-$Follow$
- יש $|V|$ קבוצות $Follow$ וגודלה של כל אחת הוא לכל היותר $|T|$
- מספר האיטרציות חסום ע"י $|T| \times |V|$. כלומר, סופי.