שפות חסרות הקשר - הגדרות

דקדוק חסר הקשר G הוא רביעייה $G=(V,T,P,S)$

V ו T הם קבוצות **סופיות זרות** של **משתנים וטרמינלים** בהתאמה.

P הוא קבוצה **סופית** של גזירות כאשר גזירה היא ביטוי מהצורה $A\to\alpha$ כאשר $A\in V$ ו $\alpha\in(V\cup T)^*$

$S\in V$ נקרא סימן התחלה.

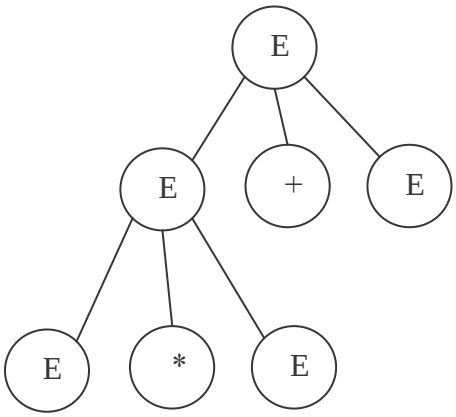
דוגמה: $G=(V,T,P,S)$

$$V=\{E\}$$
$$T=\{+,*,(,),x,y\}$$
$$S=E$$

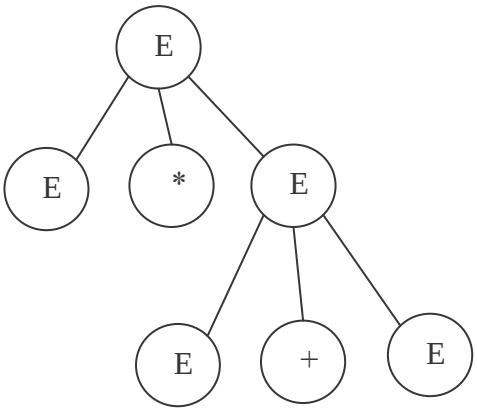
$$P=\begin{array}{l} E\to E+E \\ E\to E*E \\ E\to(E) \\ E\to x \\ E\to y \end{array}$$

זהו דקדוק חסר הקשר **לא חד משמעי** כי ניתן לגזור את אותו הביטוי בכמה דרכים שונות.

1. $E\Rightarrow E+E\Rightarrow E*E+E$



2. $E\Rightarrow E*E\Rightarrow E*E+E$



גזירה שמאלית ביותר: בכל פעם מפעילים כלל גזירה על המשתנה השמאלי ביותר בתבנית הפסוקית.
(כמו בגזירה הראשונה)

גזירה ימנית ביותר: בכל פעם מפעילים כלל גזירה על המשתנה הימני ביותר בתבנית הפסוקית.
(כמו בגזירה השניה)

דקדוק חסר הקשר נקרא **<u>ליניארי שמאלי</u>** אם כל הגזירות שלו הן מהצורה $A \to Ba$ או $A \to a$ כאשר $a \in T$, $A, B \in V$ וכמו כן מותר הכלל $S \to \varepsilon$.

דקדוק חסר הקשר נקרא **<u>ליניארי ימני</u>** אם כל הגזירות שלו הן מהצורה $A \to aB$ או $A \to a$ כאשר $a \in T$, $A, B \in V$ וכמו כן מותר הכלל $S \to \varepsilon$.

<u>משפט</u>: L שפה רגולארית אם ורק אם L מוגדרת ע"י דקדוק ליניארי שמאלי (או ימני), כלומר קיים דקדוק חסר הקשר ליניארי שמאלי (או ימני) G, כך ש $L = L(G)$.

נשים לב שהשפה של דקדוק שמכיל רק גזירות מהצורה $A \to wB$ וגם $A \to Bw$ וגם $A \to w$ אינה בהכרח שפה רגולארית.

דוגמה: הדקדוק $G = (V, T, P, S)$ כאשר P הוא אוסף כללי הגזירה הבאים:

$$S \to aB \mid bA$$
$$A \to a \mid aS \mid bAA$$
$$B \to b \mid bS \mid aBB$$

נקבל ש $L(G) = \{ w \mid |w|_a = |w|_b \}$ כלומר כל המילים שמספר ה a-ים בהן זהה למספר ה b-ים.
ניתן להוכיח זאת באינדוקציה. (על אורך הגזירה???)

$S \Rightarrow^* \alpha$ - נוכיח ש $|w|_{a,A} = |w|_{b,B}$

$A \Rightarrow^* \alpha$ - נוכיח ש $|w|_{a,A} = |w|_{b,B} + 1$

$B \Rightarrow^* \alpha$ - נוכיח ש $|w|_{a,A} + 1 = |w|_{b,B}$

(זוהי לא שפה רגולארית - נחתוך אותה עם השפה הרגולארית $a^* b^*$ ונקבל את $a^n b^n$ שכידוע לא רגולארית). מכיוון שחיתוך של שפות רגולאריות מפיק שפה רגולארית, נקבל ש $L(G)$ לא רגולארית.

<u>סימונים</u>:

A, B, C, D מסמנים משתנים מ V.
S מסמן סימן התחלה.
a, b, c מסמנים טרמינלים מ T.
X, Y, Z מסמנים משתנים או טרמינלים.
u, v, w, x, y, z מסמנים מילים של טרמינלים.
α, β, γ מסמנים מילים של משתנים וטרמינלים.

<u>טיפוסי דקדוקים:</u>

הגדרה: **<u>דקדוק מטיפוס 0</u>** הוא רביעייה $G=(V,T,P,S)$ כאשר הגזירות הן מהצורה $\alpha \to \beta$ כאשר α חייב להכיל סימן מ V.

$\gamma\alpha\delta \Rightarrow \gamma\beta\delta$ אם ורק אם $\alpha \to \beta \in P$

כאשר $\Rightarrow^*$ הוא סגור רפלקסיבי וטרנזיטיבי של $\Rightarrow$, כלומר $\alpha \Rightarrow^* \beta$ אם ורק אם קיימת סדרת מילים $\alpha = \alpha_0, \alpha_1, \ldots \alpha_n = \beta$ כאשר לכל i ניתן לגזור $\alpha_i \Rightarrow \alpha_{i+1}$.

עבור הסדרה הנ"ל נכתוב $\alpha_0 \Rightarrow^n \alpha_n$.

נגדיר את $L(G)$ - השפה מטיפוס 0 המוגדרת ע"י G:

$L(G)=\{w \in T^* \mid S \Rightarrow^* w\}$ (כל המילים שניתן לגזור מ S).

α נקראת **<u>תבנית פסוקית</u>** של G אם $S \Rightarrow^* \alpha$.

ε מסמן את המילה הריקה.

$|\alpha|$ מסמן את האורך (מספר הסימנים) של המילה α. לדוגמה: $|\varepsilon| = 0$

דוגמה:

$$G=(V,T,P,S)$$
$$V=\{A,B,C,D,E,S\}$$
$$T=\{a\}$$

$$P: \begin{array}{l} S \to ACaB \\ Ca \to aaC \\ CB \to DB \\ aD \to Da \\ AD \to AC \\ CB \to E \\ aE \to Ea \\ AE \to \varepsilon \end{array}$$

זוהי שפה תלוית הקשר הכוללת את כל המילים מהצורה $a^{(2^n)}$: $L(G)=\{a^{(2^n)} \mid n=1,2,3,\ldots\}$

דוגמה לגזירה:

$$S \Rightarrow ACaB \Rightarrow AaaCB \Rightarrow AaaDB \Rightarrow AaDaB \Rightarrow ADaaB \Rightarrow ACaaB \Rightarrow AaaCaB$$
$$\Rightarrow AaaaaCB \Rightarrow AaaaaE \Rightarrow AaaaEa \Rightarrow AaaEaa \Rightarrow AaEaaa \Rightarrow AEaaaa \Rightarrow a^4$$

הגזירה מתבצעת באיטרציות - בכל איטרציה מתחילים ב A ואחריו C ואחריו רצף של a-ים ובסוף B (באיטרציה הראשונה יש a יחיד).

ה C- הולך ימינה עד לקצה השני של רצף ה a-ים, ובדרך מכפיל פי 2 את המספר שלהם. כאשר ה C מגיע לסוף ונתקל ב B יש שתי אפשרויות:

1. להמשיך בגזירה - הוא הופך ל D, שחוזר חזרה שמאלה עד שהוא נתקל ב A ואז הופך שוב ל C, ועוברים לאיטרציה הבאה.
2. לסיים את הגזירה - ה C ו ה B ביחד הופכים ל E, שהולך שמאלה עד שהוא נתקל ב A, ואז ביחד הם הופכים ל ε ומסתיימת הגזירה.

<u>דקדוק בעל הקשר</u> הוא דקדוק שהגזירות שלו הן מהצורה: $\alpha \to \beta$ כך ש $|\alpha| \le |\beta|$ או $S \to \varepsilon$.

בנוסף לא קיים כלל גזירה $A \to \alpha S \beta$, כלומר S לא מופיע בצד ימין של כלל גזירה.

הדקדוק הנ"ל איננו תלוי הקשר כי מופיעות בו הגזירות $CB \to E$ ו $AE \to \varepsilon$

<u>ההיררכיה של חומסקי:</u>

1. שפות מטיפוס 0.
2. שפות בעלות הקשר.
3. שפות חסרות הקשר.
4. שפות רגולאריות.

רוב הקורס יתעסק בשפות חסרות הקשר.

הכרעה	NP=P? האם במודל זה חישובים דטרמיניסטים שקולים לחישובים לא דטרמיניסטים	אוטומט מקבל / סוג הזיכרון	דקדוק	שפה
ניתנת למניה רקורסיבית	כן	מכונת טיורינג	דקדוק מטיפוס 0	שפות מטיפוס 0
ניתנת להכרעה	?	אוטומט חסום ליניארי	דקדוק בעל הקשר	שפות בעלות הקשר
ניתנת להכרעה	לא	אוטומט עם מחסנית	דקדוק חסר הקשר	שפות חסרות הקשר
ניתנת להכרעה	כן	אוטומט סופי	דקדוק ליניארי שמאלי	שפות רגולאריות

נסתכל על השפה: $L=\left\{ww \mid w\in\{a,b\}^*\right\}$. זוהי לא שפה חסרת הקשר.

ומה עם המשלים שלה, $\overline{L}$?

נסתכל על השפה: $\left\{a^l b^m c^n \mid l=m \vee m=n\right\}$ היא שפה חסרת הקשר כי היא איחוד של השפות חסרות ההקשר הבאות: $L_2=\left\{a^n b^n c^*\right\}$ ו $L_1=\left\{a^* b^n c^n\right\}$. (לשתיהן ניתן ליצור אוטומט מחסנית).

החיתוך של 2 השפות $a^n b^n c^*$ ו $a^* b^n c^n$ הוא $a^n b^n c^n$ שאיננו חסר הקשר.

טענה: $\overline{L_1}\cup\overline{L_2}$ היא שפה חסרת הקשר לא דטרמיניסטית. (מה זה שפה דטרמיניסטית?)

הוכחה: השפות $L_2=a^n b^n c^*$ ו $L_1=a^* b^n c^n$ הן שפות דטרמיניסטיות חסרות הקשר. לכן השפות המשלימות להן הן חסרות הקשר. לכן $\overline{L_1}\cup\overline{L_2}$ היא שפה חסרת הקשר.

<u>משפט</u>: תהי $L = L(G)$ שפה מטיפוס 0 עבור דקדוק G. אזי L מתקבלת (לא בהכרח מוכרעת) ע"י מכונת טיורינג.

<u>הוכחה</u>: נבנה מ"ט לא דטרמיניסטית בעלת שני סרטים אשר בהינתן מילת קלט w היא תקבל אותה אם ורק אם $w \in L(G)$.

נבנה את המכונה באופן הבא:
הסרט הראשון מכילה מילה w (ששייכת או לא שייכת ל L).
הסרט השני מכיל תבניות פסוקיות של G. בהתחלת החישוב הסרט השני מכיל את S.
נקרא לתבנית הפסוקית על הסרט השני α.

<u>אופן הפעולה של המכונה:</u>

1. המכונה בחורת באופן לא דטרמיניסטי תת מילה לא ריקה של התבנית הפסוקית α. נקרא לה δ.
2. המכונה בוחרת באופן לא דטרמיניסטי כלל גזירה $\beta \to \gamma$
(אין בעיה לבצע את זה כי יש מספר סופי של כללי גזירה).
3. אם $\delta = \beta$ המכונה מחליפה δ ב γ.
(אם $|\beta| \neq |\gamma|$ המכונה מזיזה את הקצוות של המילה המתקבלת.)
4. המכונה משווה את התבנית הפסוקית שנתקבלה על הסרט השני עם w. אם יש שוויון המכונה מקבלת את ω. אחרת היא חוזרת לשלב מספר 1.

משפט: תהי L שפה שניתנת למניה רקורסיבית, כלומר $L = L(M)$ עבור מ"ט M.
אזי קיים דקדוק G מטיפוס 0 כך ש $L = L(G)$.

הוכחה:
תהי M המ"ט $M = (Q, \Sigma, \Gamma, \delta, q_0, B, F)$ כפי שנלמד בחישוביות.
נניח שהראש הקורא / כותב לא נשאר במקום, כלומר: $\delta : Q \times \Gamma \Rightarrow Q \times \Gamma \times \{L, R\}$
(במקום $\delta : Q \times \Gamma \Rightarrow Q \times \Gamma \times \{L, R, S\}$)
B - הסימן הריק.

נבנה את הדקדוק באופן הבא:
$T = \Sigma$ - הטרמינלים יהיו האותיות של מילות הקלט של המכונה.

$$V = \left\{ \left\{ \left\{ \Sigma \cup \{\varepsilon\} \right\} \times \Gamma \right\} \cup \{A_1, A_2, A_3\} \cup Q \right\}$$

כלומר יהיו שלושה סוגי משתנים:
1. זוגות שהצד השמאלי שלהם הוא אות קלט או ε, והצד הימני הוא אות "עבודה" של המכונה.
2. אחד משלושה משתנים A_1, A_2, A_3
3. מצבי המכונה M.

$S = A_1$ - נגדיר את המשתנה ההתחלתי של G להיות A_1.
הדקדוק יוצר שני העתקים של מילה ב Σ^* ואחר כך עושה סימולציה של M על העתק אחד. אם M מקבלת, אזי G הופך את ההעתק השני למילה של טרמינלים.

כללי הגזירה הם:

יצירת שני העתקים:
$A_1 \to q_0 A_2$
$A_2 \to [a, a] A_2$ לכל $a \in \Sigma$. (עד כאן הכל משתנים - אין שימוש בטרמינלים)

מקום חישוב - נניח כי הסרט אינסופי בכיוון אחד.

$A_2 \to A_3$

$A_3 \to [\varepsilon, B] A_3$ (מגדילים בתו אחד את הסרט השני)

$A_3 \to \varepsilon$ (סוף הסרט)

סימולציה של M:

$q[a,X] \to [a,Y]p$ אם $\delta(q,X) = (p,Y,R)$ עבור $a \in \Sigma \cup \{\varepsilon\}$, $X,Y \in \Gamma$, $q \in Q$

$[b,Z]q[a,X] \Rightarrow p[b,Z][a,Y]$ אם $\delta(q,X) = (p,Y,L)$ עבור

$a \in \Sigma \cup \{\varepsilon\}$, $X,Y,Z \in \Gamma$, $q \in Q$

הפיכת ההעתק השני למילת הטרמינלים.

$[a,X]q \to qaq$ $\quad q[a,X] \to qaq$ $\quad q \to \varepsilon$

עבור: $a \in \Sigma \cup \{\varepsilon\}$, $X \in \Gamma$, $q \in F$

ניתן להוכיח באינדוקציה על מספר הצעדים של M ואורך הגזירה של G ש:

$q_0 a_1, \ldots, a_n \to_M^n X_1 X_2, \ldots, X_{k-1} q X_k \ldots X_s$ אם ורק אם

$q_0 [a_1, a_1], \ldots, [a_n a_n] \Rightarrow^n [a_1, X_1], \ldots [a_{k-1}, X_{k-1}] q [a_k, X_k], \ldots [a_s X_s]$

נראה דקדוק בעל הקשר עבור השפה: $a^{(2^n)}$

המשתנים בדקדוק הם:

$$V = \left\{ \begin{array}{l} [ACaB],[Aa],[ACa],[ADa],[AEa],[Ca],[Da],[Ea],[aCB], \\ [CaB],[aDB],[aE],[DaB],[aB],S \end{array} \right\}$$

יש טרמינל אחד: $T = \{a\}$

כללי הגזירה:

נחליף את כללי הגזירה בדקדוק הקודם שראינו עבור השפה הזאת, שלא היה בעל הקשר:

1. במקום הכלל $S \to ACab$ יופיע הכלל: $S \to [ACaB]$.
2. במקום הכלל $Ca \to aaC$ (הכפלת מספר ה a-ים) יופיעו הכללים הבאים:

$[Ca]a \to aa[Ca]$ $\quad [Ca][aB] \to aa[CaB]$ $\quad [ACa]a \to [Aa]a[Ca]$

$[ACa][aB] \to [Aa]a[CaB]$ $\quad [ACaB] \to [Aa][aCB]$ $\quad [CaB] \to [aDB]$

3. במקום הכלל $CB \to DB$ (התחלת חזרה שמאלה) יופיע הכלל: $[aCB] \to [aDB]$.
4. במקום הכלל $aD \to Da$ (תנועה שמאלה) יופיעו הכללים:

$a[Da] \to [Da]a$ $\quad [aDB] \to [DaB]$ $\quad [Aa][DaB] \to [ADa]$

$[Aa][Da] \to [ADa]a$ $\quad a[DaB] \to [Da][aB]$

5. במקום הכלל $AD \to AC$ (התחלת תנועה ימינה) יופיע הכלל: $[ADa] \to [ACa]$.
6. במקום הכלל $CB \to E$ (איטרציה אחרונה) יופיע הכלל: $[aCB] \to [aE]$.
7. במקום הכלל $aE \to Ea$ (חזרה להתחלה) יופיעו הכללים:

$a[Ea] \to [Ea]a$ $\quad [aE] \to [Ea]$ $\quad [Aa][Ea] \to [AEa]a$.

8. במקום הכלל $AE \to \varepsilon$ (סיום) יופיע הכלל: $[AEa] \to a$

בשיטה הזאת ביטלנו את השימוש בגזירות הבעייתיות $AE \to \varepsilon$ ו $CB \to E$ שאינן מתאימות לדקדוק בעל הקשר.

דוגמה לשינויים באופן הגזירה: (הגזירה העליונה היא הגזירה בדקדוק החדש בעל ההקשר, והגזירה התחתונה היא הגזירה בדקדוק שאינו בעל הקשר):

$$S \to [ACaB] \to [Aa][aCB] \to [Aa][aDB] \to$$
$$S \to ACaB \to AaaCB \to AaaDB \to$$

$$[Aa][DaB] \to [ADa][aB] \to [ACa][aB] \to$$
$$AaDaB \to ADaaB \to ACaaB \to$$

$$[Aa]a[CaB] \to [Aa]aa[aCB] \to [Aa]aa[aE] \to$$
$$AaaCaB \to AaaaaCB \to AaaaaE \to$$

$$[Aa] \to aa[Ea] \to [Aa]a[Ea]a \to [Aa][Ea]aa \to$$
$$AaaaEa \to AaaEaa \to AaEaaa \to$$

$$[AEa]aaa \to aaaa$$
$$AEaaaa \to aaaa$$