

מבוא

- הזכרנו שלוש דרכים להגדרת שפה:
 - הדרך הישירה, ע"י שימוש בכלים מתמטיים לסימון קבוצות
 - הגדרת מכונה ש מזהה את השפה: עד כה היו אלה אוטומטים סופיים
 - סימון המגדיר נוהל פורמלי לייצור המילים בשפה "מחולל שפות".
- עד כה הודגמו בשתי הדרכים הראשונות.
 - לביטויים רגולריים יש דמיון לסוג השלישי.

אוניברסיטת חיפה - חורף 2009

דקדוקים

תורת הקומפילציה
אהרון נץ

מבוסס על השקפים של עומר ביהם
שמבוססים על שקפי הרצאה מהקורס אוטומטים ושפות פורמאליות
בטכניון, פרופ' שמואל זקס

אוניברסיטת חיפה - חורף 2009

דקדוק שכתוב

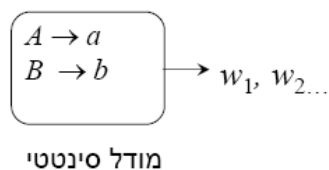
- כל צעד בנוהל ליצירת מילים הינו שכתוב חלק מהמילה, עפ"י קבוצת כללי שכתוב מוגדרים היטב.
- התהליך מפסיק כאשר מגיעים למילה **טרמינלית**, כלומר מילה שלא ניתן לשכתב אותה עוד.

אוניברסיטת חיפה - חורף 2009

מבוא



- דקדוק – מודל סינטטי



אוניברסיטת חיפה - חורף 2009

דקדוק שכתוב – דוגמא

- נהוג לציין מספרים עשרוניים עם פסיקים המפרידים בין כל שלוש ספרות עוקבות (מימין), כגון:
1,984 ■
345 ■
314,159,265 ■
- מהו הדקדוק היוצר שפה זו מעל הא"ב?
 $\Sigma^* = \{0,1, \dots, 9\} \cup \{,\}$

אוניברסיטת חיפה - חורף 2009

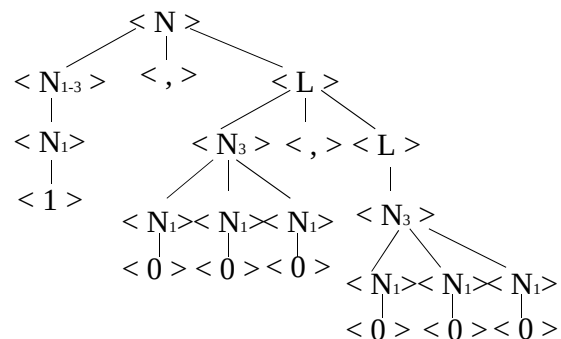
דקדוק שכתוב – דוגמא

- $N \rightarrow N_{1-3} | N_{1-3}, L$
- $L \rightarrow N_3 | L, N_3$
- $N_{1-3} \rightarrow N_1 | N_2 | N_3$
- $N_3 \rightarrow N_1 N_1 N_1$
- $N_2 \rightarrow N_1 N_1$
- $N_1 \rightarrow 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9$

אוניברסיטת חיפה - חורף 2009

דקדוק שכתוב – עץ גזירה

- $N \Rightarrow^* 1,000,000$



אוניברסיטת חיפה - חורף 2009

דקדוק שכתוב - ביטויים אריתמטיים

- נציג דקדוק היוצר ביטויים אריתמטיים בסגנון המקובל בשפות תכנות רבות, כגון:

$$a/z - c, (a+b)*c$$

- הא"ב אשר מעליו תוגדר השפה הוא

$$\Sigma = \{a, b, \dots, z\} \cup \{+, -, *, /, (,)\}$$

אוניברסיטת חיפה - חורף 2009

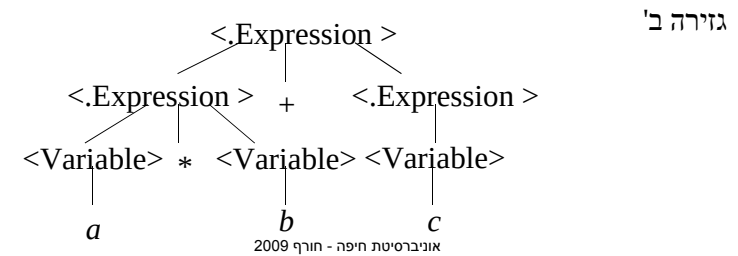
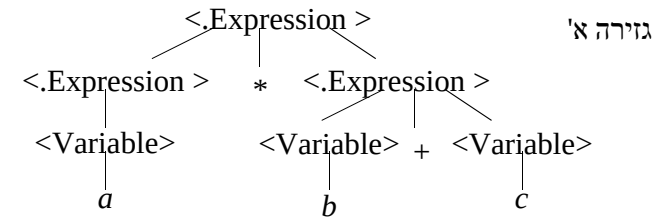
דקדוק שכתוב - ביטויים אריתמטיים

- (1) $\langle \text{Expression} \rangle \rightarrow \langle \text{Variable} \rangle$
- (2) $\langle \text{Expression} \rangle \rightarrow \langle \text{Expression} \rangle + \langle \text{Expression} \rangle$
- (3) $\langle \text{Expression} \rangle \rightarrow \langle \text{Expression} \rangle - \langle \text{Expression} \rangle$
- (4) $\langle \text{Expression} \rangle \rightarrow \langle \text{Expression} \rangle * \langle \text{Expression} \rangle$
- (5) $\langle \text{Expression} \rangle \rightarrow \langle \text{Expression} \rangle / \langle \text{Expression} \rangle$
- (6) $\langle \text{Expression} \rangle \rightarrow (\langle \text{Expression} \rangle)$
- (7) $\langle \text{Variable} \rangle \rightarrow a \mid b \mid \dots \mid z$

אוניברסיטת חיפה - חורף 2009

דקדוק שכתוב - דו-משמעות

לעיתים קיימים עצי גזירה שונים עבור אותה מילה. לדוגמא, עבור $a*b+c$



דקדוק שכתוב - הגדרה

- דקדוק G מוגדר ע"י הרביעייה $G = (V, T, P, S)$
- V - קבוצה סופית לא ריקה של משתנים
- T - קבוצה סופית לא ריקה של טרמינלים
- $S \in V$ - משתנה תחילי
- P - קבוצה סופית של כללי שכתוב מהצורה $\alpha \rightarrow \beta: \alpha \in (V \cup T)^+ \text{ and } \beta \in (V \cup T)^*$

אוניברסיטת חיפה - חורף 2009

דקדוק שכתוב - סימונים

- אותיות לטיניות גדולות מתחילת הא"ב כגון C, B, A לציון סימנים קבועים מתוך קבוצת המשתנים הדקדוקיים V .
- אותיות לטיניות קטנות מתחילת הא"ב כגון c, b, a לציון סימנים קבועים מתוך קבוצת הטרמינלים T .
- אותיות יווניות קטנות כגון $\alpha, \beta, \gamma, \xi$ לציון מילים מ $(V \cup T)^*$ כלומר סדרות של טרמינלים ומשתנים דקדוקיים
- אותיות לטיניות גדולות מסוף הא"ב כגון Z, Y, X לציון משתנים מתוך $(V \cup T)$
- אותיות לטיניות קטנות מסוף הא"ב כגון z, y, x לציון איברים מתוך T^* , כלומר מילים טרמינליות

אוניברסיטת חיפה - חורף 2009

דקדוק שכתוב - גזירה

- בהינתן מילה, ניתן לבדוק האם היא שייכת לשפה שהזקדוק מגדיר, ע"י ביצוע גזירה הפוכה, כלומר להתחיל מהמילה ולהפעיל חוקי גזירה רק בכיוון ההפוך, מימין לשמאל.
- מילה כלשהי שייכת לשפה אם ע"י גזירה הפוכה נגיע ל-S, המשתנה ההתחלתי

גזירה הפוכה - דוגמא

- יהי $G = (V, T, P, S)$ דקדוק. השפה הנוצרת ע"י הדקדוק $L(G)$, מוגדרת ע"י $L(G) = \{x \in T^* \mid S \Rightarrow^* x\}$
- מילה $\alpha \in (T \cup V)^*$ נקראת תבנית פסוקית אם $S \Rightarrow^* \alpha$
- α אינה חייבת להיות מילה טרמינלית!
- שני דקדוקים יקראו שקולים אם $L(G1) = L(G2)$

שפה של דקדוק - דוגמא

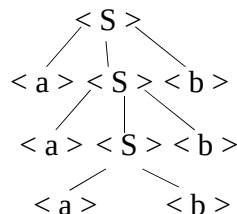
■ השפה $L = \{a^n b^n \mid n \geq 1\}$ אינה רגולרית.

■ נראה כי ניתן ליצור אותה בעזרת דקדוק

■ $G = (\{S\}, \{a,b\}, P, S)$ כללי P הם

■ (1) $S \rightarrow aSb$

■ (2) $S \rightarrow ab$



$S \Rightarrow^* aaabbbb$

אוניברסיטת חיפה - חורף 2009

ההיררכיה של חומסקי

מס'	שם	צורת כללי השכתוב
0	דקדוק כללי	$\alpha \rightarrow \beta, \alpha \neq \varepsilon$
1	דקדוק תלוי-הקשר	$\alpha A \beta \rightarrow \alpha \gamma \beta, \gamma \neq \varepsilon$ $S \rightarrow \varepsilon$
2	דקדוק חסר-הקשר	$A \rightarrow \alpha$ $S \rightarrow \varepsilon$
3	דקדוק רגולרי	ליניארי שמאלי $A \rightarrow a, A \rightarrow Ba$ $S \rightarrow \varepsilon$ ליניארי ימני $A \rightarrow a, A \rightarrow aB$ $S \rightarrow \varepsilon$

אוניברסיטת חיפה - חורף 2009

דקדוקים רגולריים - הערות

■ שקולים לשפות רגולריות

■ אפשר להוכיח כי לכל אס"ד M קיים דקדוק היוצר את $L(M)$

■ לכל שפה רגולרית L קיים דקדוק ליניארי (ימני וגם שמאלי) כך

$L(G) = L$ ש:

■ קיימת שקילות בין דקדוק ליניארי ימני לדקדוק ליניארי שמאלי

אוניברסיטת חיפה - חורף 2009

דקדוקים חסרי הקשר

■ דקדוק חסר-הקשר (דח"ה) הוא דקדוק בעל כללים

מהצורה $A \rightarrow \alpha$ כאשר $A \in V$ ו $\alpha \in (V \cup T)^*$

אוניברסיטת חיפה - חורף 2009

גזירות קנוניות

- ע"י הצגת עץ גזירה אנו מתמצתים את כל הסדרים האפשריים בגזירה מסוימת. אך עם זאת, מבנהו המסתעף לא תמיד נוח, לדוגמא בהוכחות באינדוקציה.
- לכל סדרת גזירה מתאים עץ גזירה יחיד. עם זאת, לאותו עץ גזירה עשויות להתאים מספר סדרות גזירה שונות.
- נרצה להתאים גזירה סטנדרטית (קנונית) יחידה לכל עץ גזירה של דקדוק חסר הקשר, הקובעת חד-משמעית את סדר גזירות המשתנים.

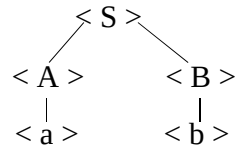
אוניברסיטת חיפה - חורף 2009

עצי גזירה - מוטיבציה

- המוטיבציה העיקרית לדיון בעצי גזירה היא הרצון לתפוס גזירות בסדר משתנים שונה כשקולות
- לדוגמא, בדקדוק שכלליו $A \rightarrow BC, B \rightarrow b, C \rightarrow c$ ניתן לגזור את bc בשתי בדרכים הבאות:

$$A \Rightarrow BC \Rightarrow bC \Rightarrow bc \quad \square$$

$$A \Rightarrow BC \Rightarrow Bc \Rightarrow bc \quad \square$$



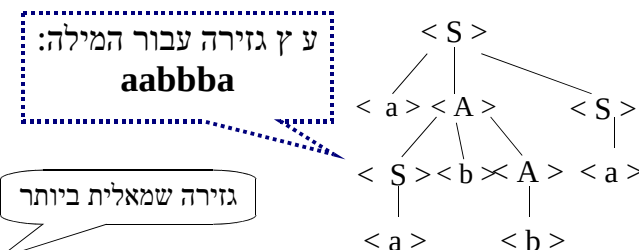
- נרצה לראות את שתי הגזירות כשקולות
- נוכל לעשות זאת ע"י הצגת עץ הגזירה המתאים:

אוניברסיטת חיפה - חורף 2009

גזירה קנונית - דוגמא

- נתבונן בדקדוק L בעל הכללים הבאים:

- (1) $S \rightarrow aAS$
- (2) $S \rightarrow a$
- (3) $A \rightarrow SbA$
- (4) $A \rightarrow b$



גזירה שמאלית ביותר

$$S \xRightarrow{(1)} aAS \xRightarrow{(3)} aSbAS \xRightarrow{(2)} aabAS \xRightarrow{(4)} aabbS \xRightarrow{(2)} aabba$$

$$S \xRightarrow{(1)} aAS \xRightarrow{(2)} aAa \xRightarrow{(3)} aSbAa \xRightarrow{(4)} aSbbS \xRightarrow{(2)} aabba$$

גזירה ימנית ביותר

אוניברסיטת חיפה - חורף 2009

גזירות קנוניות - הגדרה

- תהי $\alpha \Rightarrow \beta$ גזירה בדח"ה, כלומר, $\alpha = \delta_1 A \delta_2$, $\beta = \delta_1 \gamma \delta_2$, עבור משתנה A ותבניות פסוקיות δ_1, δ_2 כלשהן

$$\square \text{ הגזירה } \alpha \Rightarrow \beta \text{ תקרא שמאלית ביותר אם } \delta_1 \in T^*$$

$$\square \text{ הגזירה } \alpha \Rightarrow \beta \text{ תקרא ימנית ביותר אם } \delta_2 \in T^*$$

- סדרת הגזירה $\alpha \Rightarrow^* \beta$ נקראת שמאלית ביותר אם כל גזירה בה היא שמאלית ביותר.
- סדרת הגזירה $\alpha \Rightarrow^* \beta$ נקראת ימנית ביותר אם כל גזירה בה היא ימנית ביותר.
- לכל עץ גזירה בדח"ה G , מתאימה לכל היותר גזירה שמאלית (ימנית) ביותר אחת!

אוניברסיטת חיפה - חורף 2009

חד-משמעיות ורב-משמעיות

- נתון הדקדוק עבור ביטויים אריתמטיים:

$$G = (\{E\}, \{a, +, *, (,)\}, P, E)$$

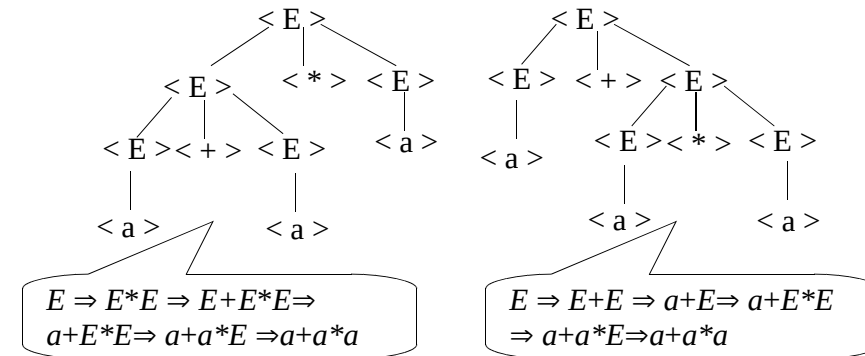
כאשר כללי הגזירה הם:

$$E \rightarrow E + E \mid E * E \mid (E) \mid a$$

- לדקדוק זה תכונה לא רצויה והיא שלאותה מילה יכולים להיות מספר עצי גזירה שונים (פתרונות מימושיים יוסברו בתרגול על yacc/bison)

רב משמעיות - דוגמא

- להלן דוגמא של שני עצי גזירה שונים עבור המילה $a + a * a$ יחד עם הגזירות השמאליות ביותר המתאימות.



רב משמעיות - הגדרה

- דקדוק G יקרא רב משמעי אם "מ קיימת מילה $w \in L(G)$ שיש לה שני עצי גזירה שונים.
- אחרת הדקדוק יקרא חד משמעי
- נשים לב כי תכונת הרב משמעיות מתייחסת לדקדוק, ולא לשפה.
- עם זאת, קיימות שפות אשר כל דקדוק היוצר אותן הוא רב משמעי. שפות אלו נקראות שפות רב משמעיות.

דקדוקים רגולריים ושקילות לאס"ד

- דקדוק G יקרא רגולרי אם כל החוקים בו הם ליניאריים ימניים

$$S \rightarrow \varepsilon, A \rightarrow B\alpha, A \rightarrow \alpha$$

ליניאריים שמאליים

$$S \rightarrow \varepsilon, A \rightarrow \alpha B, A \rightarrow \alpha$$

דקדוקים רגולריים דוגמא

■ חוקי דקדוק G

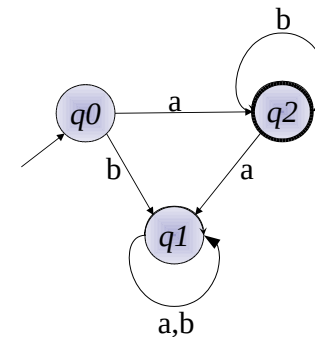
- $S \rightarrow aB \mid a$
- $A \rightarrow aB \mid a$
- $B \rightarrow bA$

■ $L(G) = (ab)^*a$ לדוגמא $w=ababa$ נגזרת

$$\text{■ } S \xrightarrow{(1)} aB \xrightarrow{(4)} abA \xrightarrow{(3)} abaB \xrightarrow{(4)} ababA \xrightarrow{(5)} ababa$$

אוניברסיטת חיפה - חורף 2009

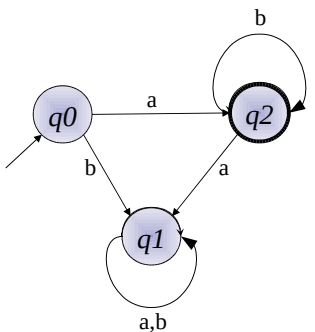
מעבר מאס"ד לדקדוק רגולרי (ימני)



- נתון M המקבל את ab^*
- נדגים בניית דקדוק G_M עבור M
- לכל מצב q_i נבנה כלל גזירה ליניארי ימני מתאים
- המשתנה התחילי יהיה כמובן זה שהוגדר עבור מצב q_0 התחילי

אוניברסיטת חיפה - חורף 2009

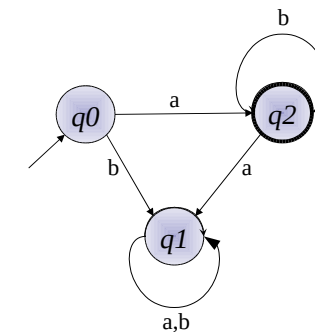
מעבר מאס"ד לדקדוק רגולרי (ימני)



- עבור המעבר $\delta(q_0, b) = q_1$ נבנה כלל גזירה מהצורה
- $V_0 \rightarrow bV_1$
- באופן דומה עבור המעבר $\delta(q_0, a) = q_2$ נבנה כלל גזירה מהצורה
- $V_0 \rightarrow aV_2$
- מכיוון ש q_2 מצב מקבל נוסיף כלל מהצורה
- $V_0 \rightarrow a$

אוניברסיטת חיפה - חורף 2009

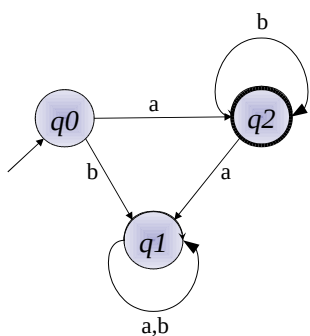
מעבר מאס"ד לדקדוק רגולריים (ימני)



- באופן דומה נקבל כללי גזירה מהצורה
- $V_1 \rightarrow aV_1 \mid bV_1$
- $V_2 \rightarrow bV_2 \mid b \mid aV_1$
- המשתנה התחילי יהיה V_0

אוניברסיטת חיפה - חורף 2009

מעבר מאס"ד לזקדוק רגולריים (ימני)

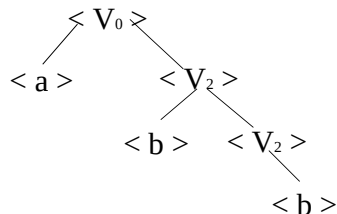


■ דוגמא למעבר מ M ל G_M עבור

והמילה $w = 'abb' \in L(G)$

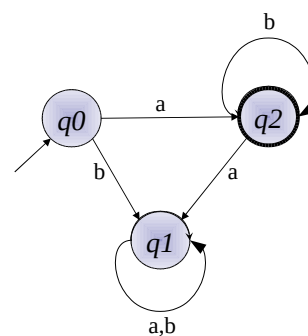
■ $q_0 \xRightarrow{(a)} q_2 \xRightarrow{(b)} q_2 \xRightarrow{(b)} q_2$

■ הגזירה המתאימה תהיה



אוניברסיטת חיפה - חורף 2009

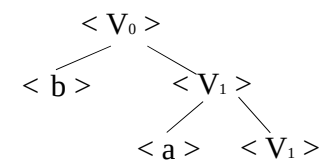
מעבר מאס"ד לזקדוק רגולריים (ימני)



■ והמילה $w = 'ba'$

■ $q_0 \xRightarrow{(b)} q_1 \xRightarrow{(a)} q_2$

■ הגזירה המתאימה תהיה



■ עץ הגזירה אינו של מילה

טרמינלית!

אוניברסיטת חיפה - חורף 2009