דקדוקי LR(K)

הערה: כל הגזירות החל מהרצאה זאת הן ימניות ביותר.

הגדרה: דקדוק שמתקבל ע"י הכנסת סימן תחילי חדש וגזירה $S' \to S$ נקרא **דקדוק מורחב**.

לדוגמה, אם $(1)\, S \to SaSb$ $(2)\, S \to \varepsilon$ הם כללי הגזירה של הדקדוק שלנו, אז כללי הגזירה של הדקדוק המורחב

$(0)\, S' \to S$

שיתקבל הם הכללים הבאים: $(1)\, S \to SaSb$

$(2)\, S \to \varepsilon$

השפה שהדקדוק הזה גוזר היא שפת ביטויי סוגרים מאוזנים.

השפה של שני הדקדוקים זהה לחלוטין. היתרון של הדקדוק המורחב הוא שאנחנו יודעים בוודאות שהמשתנה ההתחלתי מופיע בדיוק בכלל גזירה אחד, ורק בצד שמאל של כלל הגזירה.

נתאים לדקדוק מהדוגמה **טבלת גזירה**:

מצב	פעולות הגזירה			המצב הבא		
	a	b	ε - סוף קלט	S	a	b
T_0	R_2		R_2	T_1		
T_1	$Shift$		$Accept$		T_2	
T_2	R_2	R_2		T_3		
T_3	$Shift$	$Shift$			T_4	T_5
T_4	R_2	R_2		T_6		
T_5	R_1		R_1			
T_6	$Shift$	$Shift$			T_4	T_7
T_7	R_1	R_1				

מטרת הטבלה היא לבדוק, בהינתן מילה, האם היא שייכת לשפה שהדקדוק גוזר.

מתחילים כאשר במחסנית נמצא המצב ההתחלתי T_0 ומצביע לקלט שהוא כל המילה.

בכל שלב מסתכלים על המצב שבראש המחסנית ועל האות הראשונה בקלט. מסתכלים על הכניסה המתאימה בטבלה, בצד של פעולות הגזירה.
אם רואים שם R_i אז מבצעים את כלל הגזירה i הפוך - כלומר מוציאים מהמחסנית את המצבים והאותיות מהסוף להתחלה, המתאימות **לצד הימני** של כלל הגזירה. לבסוף מכניסים לשם את **הצד השמאלי** של כלל הגזירה.
אם רואים שם $Shift$ אז מכניסים את התו שבתחילת הקלט למחסנית, ומקדמים את המצביע לתו הבא בקלט.

לאחר מכן מסתכלים שוב בראש המחסנית (כעת יש שם משתנה או טרמינל ולא מצב), ומסתכלים בצד של הצעד הבא בטבלה, עבור ערך זה, ועבור המצב שמתחתיו במחסנית. מצליבים ומכניסים למחסנית את הערך המתאים מהטבלה.

מסיימים כאשר נתקלים בטבלה בערך $Accept$. אם נתקלים בשלב כלשהו בתא ריק בטבלה, זה סימן שהמילה לא שייכת לדקדוק.

לדוגמה, גזירת המילה: $aabb$:

$$S' \Rightarrow_{R_0} S \Rightarrow_{R_1} SaSb \Rightarrow_{R_1} SaSaSbb \Rightarrow_{R_2} \Rightarrow SaSabb \Rightarrow_{R_2} Saabb \Rightarrow_{R_2} aabb$$

גזירה	קלט	מחסנית
	$aabb$	T_0
2	מסתכלים בטבלה בצד של פעולות הגזירה ורואים שעבור הזוג (T_0, a) הערך המתאים הוא R_2, ומבצעים את כלל גזירה מספר 2 הפוך, כלומר מחליפים ε ב S. לכן יש כעת במחסנית T_0S. כעת מסתכלים בצד של הצעד הבא עבור הזוג (T_0, S) ורואים שם את T_1 ולכן מכניסים אותו גם למחסנית.	
	$aabb$	T_0ST_1
	מסתכלים בטבלה בצד של פעולות הגזירה ורואים שעבור הזוג (T_1, a) הערך המתאים הוא $Shift$ ולכן מכניסים את a למחסנית ומקדמים את המצביע לתו הבא בקלט. כעת מסתכלים בצד של הצעד הבא עבור הזוג (T_1, a) ורואים שם T_2 ולכן מכניסים גם אותו למחסנית.	
	abb	$T_0ST_1aT_2$
2	עבור הזוג (T_2, a) הערך בטבלה בצד של פעולות הגזירה הוא R_2 ולכן מבצעים את כלל גזירה 2: $S \to \varepsilon$ הפוך, כלומר מכניסים S למחסנית. לאחר מכן עבור הזוג (T_2, S) רואים את הערך S_3 ומכניסים גם אותו למחסנית.	
	abb	$T_0ST_1aT_2ST_3$
	עבור הזוג (T_3, a) הערך בטבלה בצד של פעולות הגזירה הוא $Shift$ ולכן מקדמים את המצביע לתו הבא בקלט ומכניסים a למחסנית. בצד של הצעד הבא הערך הוא T_4 ולכן מכניסים גם אותו למחסנית.	
	bb	$T_0ST_1aT_2ST_3aT_4$
2	עבור הזוג (T_4, b) הערך בטבלה בצד של פעולות הגזירה הוא R_2 ולכן מכניסים למחסנית S ועבור הזוג (T_4, S) הערך בצד של הצעד הבא הוא T_6 ולכן מכניסים גם אותו.	
	bb	$T_0ST_1aT_2ST_3aT_4ST_6$
	עבור הזוג (T_6, b) הערך בטבלה בצד של פעולות הגזירה הוא $Shift$ ולכן דוחפים b למחסנית. הערך בטבלה עבור הזוג (T_6, b) בצד של הצעד הבא הוא T_7.	
	b	$T_0ST_1aT_2ST_3aT_4ST_6bT_7$
1	עבור הזוג (T_7, b) הערך בטבלה בצד של פעולות הגזירה הוא R_1 ולכן מבצעים את כלל הגזירה $S \to SaSb$ הפוך. כלומר, מסלקים מהמחסנית את התווים הללו ואת כל המצבים שבדרך ונשארים עם $T_0ST_1aT_2$. כעת דוחפים את הצד השמאלי של כלל הגזירה ונשארים במחסנית עם $T_0ST_1aT_2S$. עבור הזוג (T_2, S) הערך בטבלה בצד של המצב הבא הוא T_3 ולכן דוחפים גם אותו למחסנית.	
	b	$T_0ST_1aT_2ST_3$
	עבור הזוג (T_3, b) הערך בטבלה בצד של פעולות הגזירה הוא $Shift$ ולכן מכניסים את b למחסנית ומקדמים את המצביע לתו הבא בקלט. עבור (T_3, b) הערך בטבלה בצד של הצעד הוא T_5 ולכן דוחפים גם אותו למחסנית.	
	סוף הקלט ε	$T_0ST_1aT_2ST_3bT_5$
1	עבור הזוג (T_5, ε) הערך בטבלה בצד של פעולות הגזירה הוא R_1 ולכן מסלקים מהמחסנית את צד ימין של כלל גזירה זה ונשארים עם T_0. מכניסים במקום זה את צד שמאל של כלל הגזירה, S, ולאחר מכן, עבור הזוג (T_0, S) הערך בצד של הצעד הבא הוא T_1	
	סוף הקלט ε	T_0ST_1
הערך בטבלה עבור (T_1, ε) הוא $Accept$ ולכן מקבלים את המילה. סדר הגזירה הוא בדיוק הפוך מרשימת כללי הגזירה שהשתמשנו בהם באלגוריתם, הרשומים מימין. זוהי **גזירה ימנית ביותר**.		

יהי $G=(V,T,P,S)$ דקדוק חסר הקשר ותהי $w=\sigma_1\sigma_2...\sigma_n\in T^*$ מילה של טרמינלים.

נגדיר: $FIRST_k(w)$ באופן הבא:

אם $k\le n$ אז $FIRST_k(w)=\sigma_1\sigma_2...\sigma_k$

אם $k>n$ אז $FIRST_k(w)=w$

כלומר, $FIRST_k(w)$ היא מילה המורכבת מ k האותיות הראשונות של w

או w עצמה, אם אין ל w k אותיות.

עבור $\alpha\in(V\cup T)^*$ מילה של משתנים ושל טרמינלים, נגדיר:

$$FIRST_k(\alpha)=\{FIRST_k(w)\mid \alpha\Rightarrow_R^* w\}$$

כלומר, אוסף כל ה $FIRST_k(w)$ כך שניתן לגזור את w בגזירה ימנית ביותר מ α.

$(0)\ S'\to S$

עבור הדקדוק מהדוגמה הקודמת $(1)\ S\to SaSb$, נקבל:

$(2)\ S\to\varepsilon$

$FIRST_1(\varepsilon)=\varepsilon$

$FIRST_1(a)=a$

$FIRST_1(b)=b$

$FIRST_1(S')=FIRST_1(S)=\{a,\varepsilon\}$

אם $G=(V,T,P,S)$ דקדוק חסר הקשר ו $G'=(V\cup\{S'\},T,P\cup\{S'\to S\},S')$ הוא הדקדוק המורחב המתאים לו, אז נאמר ש $G=(V,T,P,S)$ הוא דקדוק $LR(k)$

אם לכל $A,B\in V$, $\alpha,\beta,\gamma\in(V\cup T)^*$, $w,x,y\in T^*$ מתקיים:

אם שלושת התנאים הבאים מתקיימים:

1. $S'\Rightarrow_R^*\alpha Aw\Rightarrow_R\alpha\beta w$
2. $S'\Rightarrow_R^*\gamma Bx\Rightarrow_R\alpha\beta y$
3. $FIRST_k(w)=FIRST_k(y)$

אזי מתקיים: $\alpha Ay=\gamma Bx$.

מכיוון שבמילים x,y אין משתנים ו A,B הם משתנים, נובע ש $y=x$. מכיוון ש A,B הם המשתנים האחרונים במילים $\alpha Ay,\gamma Bx$ נובע ש $A=B$. לכן נובע ש $\alpha=\gamma$.

כלומר $\alpha=\gamma,\quad A=B,\quad y=x$.

דוגמה: הדקדוק $S\to Sa\mid a$ אינו דקדוק $LR(0)$. הוכחה:

1. $S'\Rightarrow_R^*\underbrace{\varepsilon}_{\alpha}\underbrace{S}_{A}\underbrace{\varepsilon}_{w}\Rightarrow\underbrace{\varepsilon}_{\alpha}\underbrace{S}_{\beta}\underbrace{\varepsilon}_{w}$
2. $S'\Rightarrow_R^*\underbrace{\varepsilon}_{\gamma}\underbrace{S}_{B}\underbrace{\varepsilon}_{x}\Rightarrow\underbrace{\varepsilon}_{\alpha}\underbrace{S}_{\beta}\underbrace{a}_{y}$
3. $FIRST_0(\underbrace{\varepsilon}_{w})=FIRST_0(\underbrace{a}_{y})=\varepsilon$

אבל: $\alpha Ay=\varepsilon S'a=S'a\ne S=\varepsilon S\varepsilon=\gamma Bx$

דקדוק זה הוא כן דקדוק $LR(1)$. לדוגמה:

1. $S' \Rightarrow^4 \underbrace{\varepsilon}_{\alpha}\underbrace{S}_{A}\underbrace{aaa}_{w} \Rightarrow \underbrace{\varepsilon}_{\alpha}\underbrace{Sa}_{\beta}\underbrace{aaa}_{w}$
2. $S' \Rightarrow^3 \underbrace{\varepsilon}_{\gamma}\underbrace{S}_{B}\underbrace{aa}_{x} \Rightarrow \underbrace{\varepsilon}_{\alpha}\underbrace{Sa}_{\beta}\underbrace{aa}_{y}$
3. $FIRST_1\underbrace{(aaa)}_{w} = a = FIRST_1\underbrace{(aa)}_{y}$

ואכן מתקיים: $\alpha Ay = \varepsilon Saa = \gamma Bx$ (זוהי כמובן לא הוכחה לכך שהדקדוק הזה הוא $LR(1)$)

דוגמה: הדקדוק שכללי הגזירה שלו הם:

$$S \to C \mid D$$
$$C \to aC \mid b$$
$$D \to aD \mid c$$

הוא דקדוק $LR(1)$.

הוכחה:

לכל גזירה ימנית ביותר של הדקדוק המורחב יש את אחת משתי הצורות הבאות:

$$S' \Rightarrow S \Rightarrow C \Rightarrow^i a^iC \Rightarrow a^ib$$
$$S' \Rightarrow S \Rightarrow D \Rightarrow^i a^iD \Rightarrow a^ic$$

אם :

1. $S' \Rightarrow_R^* \alpha Aw \Rightarrow_R \alpha\beta w$
2. $S' \Rightarrow_R^* \gamma Bx \Rightarrow_R \alpha\beta y$

אזי: $x = w = \varepsilon$.

זאת מכיוון שכאשר אנחנו נמצאים גוזרים אחת מהמילים $\alpha Aw, \gamma Bx$ אז A, B הם משתנים, ובדקדוק שלנו, אין דרך להגיע למילה שיש בה משתנה ואחר כך מספר טרמינלים שונה מאפס, ובתתי המילים w, x אין משתנים.

לכן, אם :

3. $FIRST_1(y) = FIRST_1(w)$ אז $w = \varepsilon$ ולכן $FIRST_1(w) = \varepsilon$. לכן $FIRST_1(y) = \varepsilon$ והמילה היחידה שמקיימת זאת היא ε ולכן $y = \varepsilon$. כלומר $y = x$.

לכן נשאר להוכיח שמתקיים: $\gamma = \alpha$ ו $A = B$.

נניח כי $\alpha\beta$ שבגזירה $S' \Rightarrow_R^* \gamma Bx \Rightarrow_R \alpha\beta y$ מתקבלת מ γB ע"י הגזירה $B \to \delta$.

כלומר הגזירות הן:

1. $S' \Rightarrow_R^* \alpha A \Rightarrow_R \alpha\beta$
2. $S' \Rightarrow_R^* \gamma B \Rightarrow_R \alpha\beta = \gamma\delta$

נבדיל בין האפשרויות הבאות: $A = D$ $A = C$ $A = S$ $A = S'$.

אם $A = S'$ אז $\alpha = \varepsilon$ ו $\beta = S$

(כי לא ניתן לגזור את S' ולכן הגזירה ו $S' \Rightarrow_R^* \alpha A$ היא בעצם 0 גזירות).

באופן דומה עבור $S' \Rightarrow_R^* \gamma B \Rightarrow_R \alpha\beta = S$ נקבל ש $B = S'$ כי זו הדרך היחידה לגזור את S ו $\gamma = \varepsilon$ כנדרש.

אם $A=S$ אזי $\alpha=\varepsilon$ ו $\beta=C$ או $\beta=D$ כי אלו המילים ש S יודע לגזור.
בשני המקרים יש רק דרך אחת לגזור C או D ולכן נקבל ש $\gamma B=S$ כלומר $\gamma=\varepsilon=\alpha$ ו $B=S$.

אם $A=C$ אז $\beta=aC$ או $\beta=b$ כי זה מה ש A יודע לגזור.
אם $\beta=aC$ אזי $B=C$ אז הדרך היחידה לגזור מ $\gamma B\Rightarrow\alpha\beta=\alpha aC$ כאשר מפתחים את B פעם אחת בלבד היא אם $B=C$. לכן $\gamma=\alpha$.
אם $\beta=b$ אזי $\gamma B\Rightarrow\alpha\beta=\alpha b$ ושוב מתחייב ש $B=C$ ולכן $\gamma=\alpha$.

המקרה של $A=D$ דומה מאוד למקרה של $A=C$.

לכן בכל מקרה, אם שלושת התנאים הנ"ל מתקיימים אז $\alpha Ay=\gamma Bx$.
לכן זהו דקדוק $LR(1)$.

$$S\to Ab\mid Bc$$

דוגמה: הדקדוק: $A\to Aa\mid\varepsilon$ איננו דקדוק $LR(k)$ עבור כל $k\ge 0$.

$$B\to Ba\mid\varepsilon$$

הוכחה: נניח בשלילה כי הדקדוק הנ"ל הינו דקדוק $LR(k)$.
נתבונן בשתי הגזירות הבאות:

1. $S'\Rightarrow S\Rightarrow^{k+1}\underbrace{\underbrace{A}_{\alpha A}\underbrace{a^kb}_{w}}\Rightarrow\underbrace{\underbrace{a^k}_{\alpha\beta}\underbrace{b}_{w}}$
2. $S'\Rightarrow S\Rightarrow Bc\Rightarrow^{k}\underbrace{\underbrace{B}_{\gamma B}\underbrace{a^kc}_{x}}\Rightarrow\underbrace{\underbrace{a^k}_{\alpha\beta}\underbrace{c}_{y}}$

נקבל:

3. $FIRST_k(y)=FIRST_k(w)=a^k$

אבל $A\ne B$. לכן זהו לא דקדוק $LR(k)$.

למה: יהי $G'=(V,T,P,S')$ דקדוק מורחב שאינו דקדוק $LR(k)$. אזי קיימות שתי גזירות:

1. $S'\Rightarrow_R^* \alpha Aw\Rightarrow\alpha\beta w$
2. $S'\Rightarrow_R^* \gamma Bx\Rightarrow\alpha\beta y=\gamma\delta x$

(מכיוון שבגזירה האחרונה המילים γ,x לא הולכות לאיבוד, נוכל לסמן $\alpha\beta y=\gamma\delta x$)
כך ש: $FIRST_k(y)=FIRST_k(w)$ וגם $|\gamma\delta|\ge|\alpha\beta|$ אך $\gamma Bx\ne\alpha Ay$.

הוכחה: על פי ההגדרה של דקדוק $LR(k)$ ומכיוון שהדקדוק G' איננו דקדוק $LR(k)$, מתחייב שקיימות שתי גזירות הנ"ל, וגם $FIRST_k(y)=FIRST_k(w)$ אבל $\gamma Bx\ne\alpha Ay$.
לכן נשאר להוכיח ש $|\gamma\delta|\ge|\alpha\beta|$.
נניח ש $|\gamma\delta|<|\alpha\beta|$ ונמצא זוג גזירות אחר שמקיים את התנאים הנ"ל.

מכיוון ש $\gamma\delta x = \alpha\beta y$ וגם $|\gamma\delta| < |\alpha\beta|$ קיים $z \in T^+$ כך ש $\alpha\beta = \gamma\delta z$
(כלומר z היא השארית של $\alpha\beta$)
זה יראה כך:

α	β	y

γ	δ		x

z

לכן הגזירות הן:

1. $S' \Rightarrow_R^* \alpha A w \Rightarrow \alpha\beta w = \gamma\delta z w$
2. $S' \Rightarrow_R^* \gamma B x \Rightarrow \alpha\beta y = \gamma\delta x$

ומתקיים $x = zy$ (ולכן $FIRST_k(x) = FIRST_k(zy)$)
מכיוון ש $FIRST_k(y) = FIRST_k(w)$ זה אומר ש: $FIRST_k(x) = FIRST_k(zy) = FIRST_k(zw)$

הסבר: אם $k < |z|$ אז $FIRST_k(zy) = FIRST_k(z) = FIRST_k(zw)$
אחרת: $FIRST_k(y) = FIRST_k(w)$ ולכן לכל $i < k$ מתקיים $FIRST_i(y) = FIRST_i(w)$ ולכן:

$$FIRST_k(zy) = z \cdot FIRST_{k-|z|}(y) = z \cdot FIRST_{k-|z|}(w) = FIRST_k(zw)$$

נסתכל שוב על הגזירות:

1. $S' \Rightarrow_R^* \underbrace{\gamma}_{\alpha'} \underbrace{B}_{A'} \underbrace{x}_{w'} \Rightarrow \underbrace{\gamma}_{\alpha'} \underbrace{\delta}_{\beta'} \underbrace{x}_{w'}$
2. $S' \Rightarrow_R^* \underbrace{\alpha}_{\gamma'} \underbrace{A}_{B'} \underbrace{w}_{x'} \Rightarrow \underbrace{\gamma}_{\alpha'} \underbrace{\delta}_{\beta'} \underbrace{zw}_{y'} = \underbrace{\alpha}_{\gamma'} \underbrace{\beta}_{\delta'} \underbrace{w}_{x'}$

$FIRST_k(w') = FIRST_k(y')$

אם התנאי של $LR(k)$ היה מתקיים אזי היינו מקבלים ש $\gamma' B' x' = \alpha A w = \gamma B z w = \alpha' A' y'$.
כלומר $\alpha A = \gamma B z$. זה כמובן לא יכול להתקיים כי $A \in V$ (משתנה) ו $z \in T^+$ (לא מכיל משתנים)
אולם $T^+ \cap V = \phi$. לכן $\gamma' B' x' = \alpha A w \neq \gamma B z w = \alpha' A' y'$

הנחנו $|\gamma\delta| < |\alpha\beta|$ כלומר, $|\alpha'\beta'| < |\gamma'\delta'|$. בפרט $|\gamma'\delta'| \geq |\alpha'\beta'|$

מסקנה - מצאנו זוג גזירות:

1. $S' \Rightarrow_R^* \alpha' A' w' \Rightarrow \alpha'\beta' w'$
2. $S' \Rightarrow_R^* \gamma' B' x' \Rightarrow \alpha'\beta' y' = \gamma'\delta' x'$

כך ש $FIRST_k(w') = FIRST_k(y')$ וגם $|\gamma'\delta'| \geq |\alpha'\beta'|$ וגם $\gamma' B' x' \neq \alpha' A' y'$
וזה מה שרצינו להוכיח.

ε FREE FIRST FUNCTION

הגדרה: יהי $G=(V,T,P,S)$ דקדוק חסר הקשר ותהי $\alpha \in (T \cup V)^*$ מילה של משתנים וטרמינלים.

נגדיר את הפונקציה $EFF_k^G(\alpha)$ באופן הבא:

1. אם α לא מתחיל במשתנה אזי $EFF_k(\alpha)=\{FIRST_k(\alpha)\}$.
2. אם α כן מתחיל במשתנה אזי:

$$EFF_k(\alpha)=\{FIRST_k(w) \mid \alpha \Rightarrow_R^* \beta \Rightarrow_R w \wedge \beta \notin \{Cw \mid C \in V\}\}$$

כלומר, כל ה $FIRST_k(w)$ של מילים w שניתן לגזור מ α שבגזירתם המשתנה הראשון לא מוחלף ב בסימן הריק ε.

דוגמה:

$$S \to SaSb \mid \varepsilon$$
$$S' \to S$$

$$EFF_1(\varepsilon)=\varepsilon$$
$$EFF_1(a)=a$$
$$EFF_1(b)=b$$
$$EFF_1(S')=EFF_1(S)=\phi$$

הסבר: נניח לדוגמה שנרצה לגזור את ab. מכיוון שזו גזירה ימנית ביותר, בשלב האחד לפני האחרון נישאר עם Sab כעת, אנחנו מחליפים את S שהוא המשתנה הראשון ב ε ולכן המילה ab לא שייכת ל $EFF_1(S)$ או ל $EFF_1(S')$.