1. Контекстносвободна граматика, дърво на синтактичен анализ, контекстносвободен език.

Определение 8. Безконтекстна граматика е четворка от вида

$$G = (V, \Sigma, R, S),$$

където

- V е крайно множество от *променливи*;
- Σ е крайно множество от *букви*, $\Sigma \cap V = \emptyset$;
- $R \subseteq V \times (V \cup \Sigma)^{\star}$, крайно множество от *правила*;
- $S \in V$ е началната променлива.

При дадена граматика G, за правилата на граматиката обикновено ще пишем $A \to \alpha$ вместо $(A, \alpha) \in R$. Ще въведем и релация между думи $\alpha, \beta \in (V \cup \Sigma)^{\star}$, която ще казва, че думата β œ получаава от α като приложим правло от граматиката. За две думи $u, v \in (V \cup \Sigma)^{\star}$ ще пишем $u \to_G v$, ако съществуват думи $x, y \in (\Sigma \cup V)^{\star}$, $A \in V$, правило $A \to \alpha$ и $u = xAy$, $v = x\alpha y$. С $\to_G^{\star}$ ще означаваме рефлексивното и транзитивно затваряне на релацията $\to_G$.

Езикът породен от граматиката G е множеството от думи

$$\mathcal{L}(G) = \{\alpha \in \Sigma^{\star} \mid S \to_G^{\star} \alpha\}.$$

Една граматика $G = (V, \Sigma, R, S)$ се нарича **безконтекстна**, ако имаме ограничението, че $R \subseteq V \times (V \cup \Sigma)^\star$. Да повторим дефиницията на релацията $\alpha \Rightarrow_G \beta$ от Раздел 3.1 в частния случай, когато граматиката е безконтекстна.

$$\frac{(A, \alpha) \in R \qquad \lambda, \rho \in (V \cup \Sigma)^\star}{\lambda A \rho \Rightarrow_G \lambda \alpha \rho}$$

Нека официално да обявим, че един език L се нарича **безконтекстен**, ако съществува безконтекстна граматика G, за която $L = \mathcal{L}(G) = \{\omega \in \Sigma^\star \mid S \overset{\star}{\Rightarrow} \omega\}$.

Като частен случай на *Твърдение* 3.4 получаваме следното свойство.

В [PL98] дефиницията е различна. Там $\Sigma \subseteq V$. На англ. *context-free grammar*. Други срещани наименования на български са *контекстносвободна*, *контекстнонезависима*. Тук всички правила са от вида $A \to \alpha$, където $\alpha \in (V \cup \Sigma)^\star$. В частност имаме, че:

Синтактично (parse) дърво на извод (за тип 2)

Едно наредено дърво на извод, което описва (за тип 2) $S \overset{*}{\Rightarrow} w$ независимо от ред на заместванията.

Конструкция на извода

$S = x_1 \Rightarrow x_2 \Rightarrow \cdots \Rightarrow x_n = x \in \Sigma^*$:

Корен S.

Ако на стъпка i правим заместването $A \to z = z_1, \ldots, z_k$.

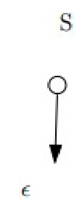
$\to$ възлите наследници на A са $z_1, \ldots, z_k$.

Наблюдение: Листата са буквите на x.

Синтактично дърво на извод

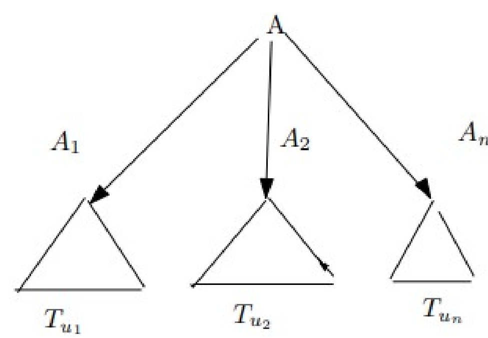
Дърво с резултат a

Дърво с резултат ε



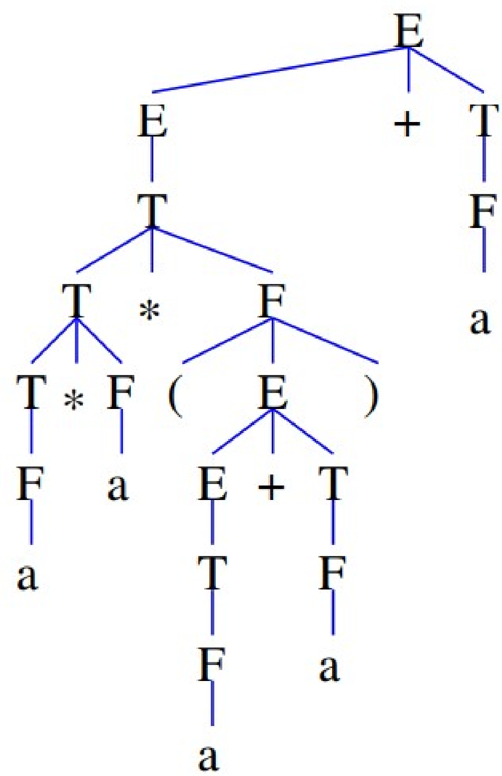
Дърво с резултат $u_1u_2\dots u_n$

$A \to A_1A_2\dots A_n$



Пример

$\color{red}{E} \Rightarrow$	$E \to E + T$
$\Rightarrow \color{red}{E} + T$	$E \to T$
$\Rightarrow \color{red}{T} + T$	$T \to T * F$
$\Rightarrow \color{red}{T} * F + T$	$T \to T * F$
$\Rightarrow \color{red}{T} * F * F + T$	$T \to F$
$\Rightarrow \color{red}{F} * F * F + T$	$F \to a$
$\Rightarrow a * \color{red}{F} * F + T$	$F \to a$
$\Rightarrow a * a * \color{red}{F} + T$	$F \to (E)$
$\Rightarrow a * a * (\color{red}{E}) + T$	$E \to E + T$
$\Rightarrow a * a * (\color{red}{E} + T) + T$	$E \to T$
$\Rightarrow a * a * (\color{red}{T} + T) + T$	$T \to F$
$\Rightarrow a * a * (\color{red}{F} + T) + T$	$F \to a$
$\Rightarrow a * a * (a + \color{red}{T}) + T$	$T \to F$
$\Rightarrow a * a * (a + \color{red}{F}) + T$	$F \to a$
$\Rightarrow a * a * (a + a) + \color{red}{T}$	$T \to F$
$\Rightarrow a * a * (a + a) + \color{red}{F}$	$F \to a$
$\Rightarrow a * a * (a + a) + a$	



Най-ляв извод

На всяка стъпка в извода:

заместваме най-лявата променлива

Пример: от предната стр.

1-1 релация най-ляв извод $\leftrightarrow$ синтактично дърво

Наблюдение (Твърдение) (за тип 2)

$x \in L(G) \Leftrightarrow \exists$извод за x

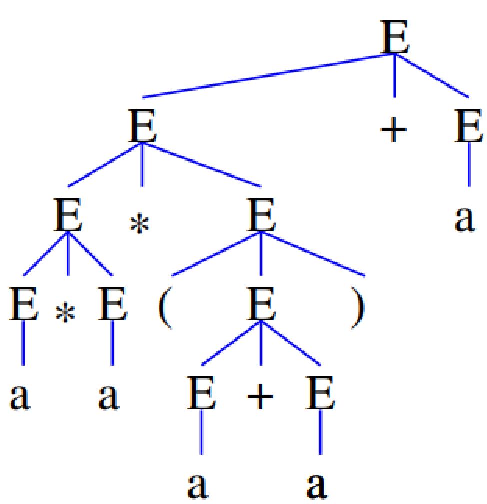
$\Leftrightarrow \exists$ синтактично дърво за извода на x по листата

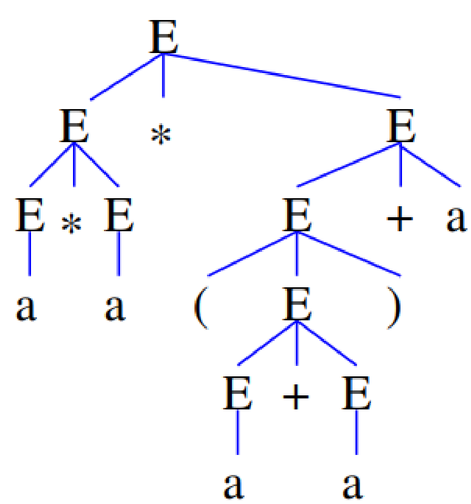
$\Leftrightarrow \exists$ най-ляв извод за x

Пример за нееднозначни изводи

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

$$
\begin{aligned}
P = \{&E \to E + E,\\
&E \to E * E,\\
&E \to a,\\
&E \to (E)\}
\end{aligned}
$$





2. Доказателство на теоремите за затвореност на контекстносвободните езици.

Затвореност на CFG относно $\cup$

Да разгледаме

$G_1 = (V_1, \Sigma, P_1, S_1)$,

$G_2 = (V_2, \Sigma, P_2, S_2)$,

Нека $V_1 \cap V_2 = \emptyset$ и

$G = (\{S\} \cup V_1 \cup V_2, \Sigma, \{S \to S_1, S \to S_2\} \cup P_1 \cup P_2)$.

Очевидно имаме

$L(G) = L(G_1) \cup L(G_2)$.

Затвореност на CFG относно $\cdot$

Да разгледаме
$G_1 = (V_1, \Sigma, P_1, S_1)$,
$G_2 = (V_2, \Sigma, P_2, S_2)$,
Нека $V_1 \cap V_2 = \emptyset$ и
$G = (\{S\} \cup V_1 \cup V_2, \Sigma, \{S \to S_1 S_2\} \cup P_1 \cup P_2)$.

Ясно е, че
$L(G) = L(G_1) \cdot L(G_2)$.

Затвореност на CFG относно $*$

Да разгледаме
$G_1 = (V_1, \Sigma, P_1, S_1)$
и нека S_1 не участва в десните страни на P. И
$G = (\{S\} \cup V_1, \Sigma, \{S \to \varepsilon, S \to S_1, S_1 \to S_1 S_1\} \cup P_1 \setminus \{S_1 \to \varepsilon\})$.

Тогава
$L(G) = L(G_1)^*$.

3. Недетерминиран стеков автомат.

Недетерминиран стеков автомат е седморка от вида

$$P = \langle Q, \Sigma, \Gamma, \sharp, \Delta, q_{\mathtt{start}}, q_{\mathtt{accept}} \rangle,$$

където:

- Q е крайно множество от състояния;
- Σ е крайна входна азбука;
- Γ е крайна стекова азбука;
- $\sharp \in \Gamma$ е символ за дъно на стека;
- $\Delta : Q \times \Sigma_\varepsilon \times \Gamma \to \mathscr{P}(Q \times \Gamma^{\leq 2})$ е функция на преходите;
- $q_{\mathtt{start}} \in Q$ е начално състояние;
- $q_{\mathtt{accept}} \in Q$ е заключителното състояние.

На англ. *Push-down automaton*. В този курс няма да разглеждаме детерминирани стекови автомати. Когато кажем стеков автомат, ще имаме предвид недетерминиран стеков автомат. Означаваме $\Sigma_\varepsilon \stackrel{\text{деф}}{=} \Sigma \cup \{\varepsilon\}$ и $\Gamma^{\leq 2} \stackrel{\text{деф}}{=} \{\varepsilon\} \cup \Gamma \cup \Gamma^2$.

Обикновено се взима Δ функцията да има сигнатура $\Delta : Q \times \Sigma_\varepsilon \times \Gamma \to \mathscr{P}(Q \times \Gamma^\star)$. Дефиницията на стеков автомат има много вариации, всички еквивалентни помежду си.

Конфигурация (или моментно описание) на изчислението със стеков автомат представлява тройка от вида $(q, \alpha, \gamma) \in Q \times \Sigma^\star \times \Gamma^\star$, т.е. автоматът се намира в състояние q, думата, която остава да се прочете е α, а съдържанието на стека е думата γ. Удобно е да въведем бинарната релация $\vdash_P$ над $Q \times \Sigma^\star \times \Gamma^\star$, която ще ни казва как моментното описание на автомата P се променя след изпълнение на една стъпка.

На англ. Instanteneous description

Пример 3.14. За езика $L = \{a^n b^n \mid n \in \mathbb{N}\}$, да разгледаме $P = \langle Q, \Sigma, \Gamma, \sharp, \Delta, q_{\mathtt{start}}, q_{\mathtt{accept}} \rangle$, където

- $Q \stackrel{\text{деф}}{=} \{q, p, f\}$;
- $q_{\mathtt{start}} \stackrel{\text{деф}}{=} q$ и $q_{\mathtt{accept}} \stackrel{\text{деф}}{=} f$;
- $\Sigma \stackrel{\text{деф}}{=} \{a, b\}$ и $\Gamma \stackrel{\text{деф}}{=} \{\sharp, a\}$;
- Релацията на преходите Δ има следната дефиниция:

Тук получаваме детерминистичен стеков автомат.

(1) $\Delta(q, a, \sharp) \stackrel{\text{деф}}{=} \{(q, a\sharp)\}$;

(2) $\Delta(q, a, a) \stackrel{\text{деф}}{=} \{(q, aa)\}$; // трупаме a-та в стека

(3) $\Delta(q, \varepsilon, \sharp) \stackrel{\text{деф}}{=} \{(f, \varepsilon)\}$; // трябва да разпознаем и думата ε

(4) $\Delta(q, b, a) \stackrel{\text{деф}}{=} \{(p, \varepsilon)\}$; // Започваме да четем само b-та

(5) $\Delta(p, b, a) \stackrel{\text{деф}}{=} \{(p, \varepsilon)\}$; // Чистим a-тата от стека

(6) $\Delta(p, \varepsilon, \sharp) \stackrel{\text{деф}}{=} \{(f, \varepsilon)\}$.

(7) За всички останали тройки (r, x, y), нека $\Delta(r, x, y) \stackrel{\text{деф}}{=} \emptyset$.

Да видим как думата a^2b^2 се разпознава от стековия автомат P:

$$\begin{aligned}
(q, a^2b^2, \sharp) &\vdash_P (q, ab^2, a\sharp) && \text{// правило } (1)\\
&\vdash_P (q, b^2, aa\sharp) && \text{// правило } (2)\\
&\vdash_P (p, b, a\sharp) && \text{// правило } (4)\\
&\vdash_P (p, \varepsilon, \sharp) && \text{// правило } (5)\\
&\vdash_P (f, \varepsilon, \varepsilon) && \text{// правило } (6)
\end{aligned}$$

Получихме, че $(q_{\mathtt{start}}, a^2b^2, \sharp) \vdash_P^\star (q_{\mathtt{accept}}, \varepsilon, \varepsilon)$, откъдето следва, че $a^2b^2 \in \mathcal{L}(P)$.

4. Изпълнение в недетерминиран стеков автомат.

?

5. Език, разпознаван от недетерминиран стеков автомат.

Пример 3.15. Езикът $L = \{\, \omega\omega^{\mathtt{rev}} \mid \omega \in \{a,b\}^\star \,\}$ се разпознава от стеков автомат

$$P = \langle Q, \Sigma, \Gamma, \sharp, \Delta, q_{\mathtt{start}}, q_{\mathtt{accept}} \rangle,$$

където:

- $Q \stackrel{\text{деф}}{=} \{q, p, f\}$ и $q_{\mathtt{start}} \stackrel{\text{деф}}{=} q$, $q_{\mathtt{accept}} \stackrel{\text{деф}}{=} f$;
- $\Sigma \stackrel{\text{деф}}{=} \{a, b\}$, $\Gamma \stackrel{\text{деф}}{=} \{a, b, \sharp\}$;
- Функцията на преходите Δ има следната дефиниция:

За всички липсващи твойки в дефиницията на Δ приемаме, че Δ връща $\emptyset$

(1) $\Delta(q, x, \sharp) \stackrel{\text{деф}}{=} \{(q, x\sharp)\}$, където $x \in \{a, b\}$;

(2) $\Delta(q, \varepsilon, \sharp) \stackrel{\text{деф}}{=} \{(q, \varepsilon)\}$;

(3) $\Delta(q, x, x) \stackrel{\text{деф}}{=} \{(q, xx), (p, \varepsilon)\}$, където $x \in \{a, b\}$;

(4) $\Delta(q, a, b) \stackrel{\text{деф}}{=} \{(q, ab)\}$;

(5) $\Delta(q, b, a) \stackrel{\text{деф}}{=} \{(q, ba)\}$;

(6) $\Delta(p, x, x) \stackrel{\text{деф}}{=} \{(p, \varepsilon)\}$, където $x \in \{a, b\}$;

(7) $\Delta(p, \varepsilon, \sharp) \stackrel{\text{деф}}{=} \{(f, \varepsilon)\}$;

Основното наблюдение, което трябва да направим за да разберем конструкцията на автомата е, че всяка дума от вида $\omega\omega^{\mathtt{rev}}$ може да се запише като $\omega_1 aa\omega_1^{\mathtt{rev}}$ или $\omega_1 bb\omega_1^{\mathtt{rev}}$. Да видим защо P разпознава думата $abaaba$ с празен стек. Започваме по следния начин:

$$\begin{aligned}
(q, abaaba, \sharp) &\vdash_P (q, baaba, a\sharp) && \text{// правило } (1)\\
&\vdash_P (q, aaba, ba\sharp) && \text{// правило } (5)\\
&\vdash_P (q, aba, aba\sharp). && \text{// правило } (4)
\end{aligned}$$

Сега можем да направим два избора как да продължим. Състоянието p служи за маркер, което ни казва, че вече сме започнали да четем $\omega^{\mathtt{rev}}$. Поради тази причина, продължаваме така:

$$\begin{aligned}
(q, aba, aba\sharp) &\vdash_P (p, ba, ba\sharp) && \text{// правило } (3)\\
&\vdash_P (p, a, a\sharp) && \text{// правило } (6)\\
&\vdash_P (p, \varepsilon, \sharp) && \text{// правило } (6)\\
&\vdash_P (f, \varepsilon, \varepsilon). && \text{// правило } (7)
\end{aligned}$$

Да проиграем още един пример. Да видим защо думата aba не се извежда от автомата.

$$\begin{aligned}
(q, aba, \sharp) &\vdash_P (q, ba, a\sharp) && \text{// правило } (1)\\
&\vdash_P (q, a, ba\sharp) && \text{// правило } (5)\\
&\vdash_P (q, \varepsilon, aba\sharp). && \text{// правило } (4)
\end{aligned}$$

От последното моментно описание на автомата нямаме нито един преход, следователно думата aba не се разпознава от P.

6. Доказателство на теорема за свеждане на контекстносвободна граматика към еквивалентен недетерминиран стеков автомат.

Лема 3.12. За всяка безконтекстна граматика G, съществува стеков автомат P, такъв че $\mathcal{L}(G) = \mathcal{L}(P)$.

Доказателството на лемата следва до голяма степен [PL98, стр. 136].

Доказателство. Нека е дадена безконтекстната граматика $G = \langle V, \Sigma, R, S\rangle$ в нормална форма на Чомски. Нашата цел е да построим стеков автомат

Тук е важно да използваме най-ляв извод в граматика.

$$P = \langle Q, \Sigma, \Gamma, \sharp, \Delta, q_{\mathtt{start}}, q_{\mathtt{accept}}\rangle,$$

който разпознава езика $\mathcal{L}(G)$.

- $Q = \{q_{\mathtt{start}}, p, q_{\mathtt{accept}}\}$;
- $\Gamma = \Sigma \cup V \cup \{\sharp\}$;
- Релацията на преходите Δ дефинираме по следния начин:

Понеже граматиката е в нормална форма на Чомски, то $|\alpha| \leq 2$ и удовлетворяваме дефиницията на Δ.

(1) $\Delta(q_{\mathtt{start}}, \varepsilon, \sharp) = \{(p, S\sharp)\}$;

(2) $\Delta(p, \varepsilon, A) = \{(p, \gamma) \mid A \to_G \gamma\}$, за всяка променлива $A \in V$;

(3) $\Delta(p, a, a) = \{(p, \varepsilon)\}$, за всяка буква $a \in \Sigma$;

(4) $\Delta(p, \varepsilon, \sharp) = \{(q_{\mathtt{accept}}, \varepsilon)\}$.

Ще докажем, че за всяка променлива $A \in V$, за всяка дума $\alpha \in \Sigma^\star$, и $\delta \in (V \cup \Sigma)^\star$, то е изпълнено, че:

(а) ако $A \overset{\star}{\Rightarrow}_{\mathtt{left}} \alpha$, то $(p, \alpha, A) \vdash_P^\star (p, \varepsilon, \varepsilon)$;

(б) ако $(p, \alpha, \delta) \vdash_P^\star (p, \varepsilon, \varepsilon)$, то $\delta \overset{\star}{\Rightarrow}_{\mathtt{left}} \alpha$.

Ако приемем, че (а) и (б) са изпълнени, тогава, ако вземем $\delta = S$ и $A = S$, то ще получим, че

$$\begin{aligned}
\alpha \in \mathcal{L}(G) &\Leftrightarrow S \overset{\star}{\Rightarrow}_{\mathtt{left}} \alpha \\
&\Leftrightarrow (p, \alpha, S) \vdash_P^\star (p, \varepsilon, \varepsilon) && \text{// от (а) и (б)} \\
&\Leftrightarrow (q_{\mathtt{start}}, \alpha, \sharp) \vdash_P^\star (q_{\mathtt{accept}}, \varepsilon, \varepsilon) && \text{// от деф. на } P \\
&\Leftrightarrow \alpha \in \mathcal{L}(P).
\end{aligned}$$

Сега преминаваме към доказателствата на двете твърдения.

Доказателството на (а) ще проведем с пълна индукция по дължината ℓ на извода $A \overset{\ell}{\Rightarrow}_{\mathtt{left}} \alpha$ за $\ell \geq 1$.

Очевидно е, че не е възможно да имаме $A \overset{0}{\Rightarrow}_{\mathtt{left}} \alpha$.

Нека $\ell = 1$. Този случай е лесен. Понеже граматиката е в нормална форма на Чомски, тук единствената възможност е да имаме извод $A \overset{1}{\Rightarrow}_{\texttt{left}} a$, за някоя $a \in \Sigma$. Тогава, според дефиницията на стековия автомат P, имаме следното изчисление:

$$(p, a, A) \vdash_P (p, a, a) \vdash_P (p, \varepsilon, \varepsilon).$$

Нека $\ell > 1$. Тогава изводът може да се разбие така:

$$\dfrac{\dfrac{A \to_G BC}{A \Rightarrow_{\texttt{left}} BC} \quad BC \overset{\ell-1}{\Rightarrow}_{\texttt{left}} \alpha}{A \overset{\ell}{\Rightarrow}_{\texttt{left}} \alpha}$$

Сега прилагаме *Твърдение 3.6* и *Лема 3.11* и получаваме, че можем да разбием извода $BC \overset{\ell-1}{\Rightarrow}_{\texttt{left}} \alpha$ така:

$$B \overset{\ell_1}{\Rightarrow}_{\texttt{left}} \alpha_1$$
$$C \overset{\ell_2}{\Rightarrow}_{\texttt{left}} \alpha_2,$$

където $\alpha = \alpha_1 \cdot \alpha_2$ и $\ell - 1 = \ell_1 + \ell_2$. Понеже $\ell_1 < \ell$ и $\ell_2 < \ell$, от индукционното предположение получаваме изчислението:

$$\dfrac{\dfrac{\dfrac{A \to_G BC}{\Delta(p, \varepsilon, A) \ni (p, BC)}}{(p, \alpha_1\alpha_2, A) \vdash_P (p, \alpha_1\alpha_2, BC)} \quad \textbf{(и.п.)} \quad \dfrac{\dfrac{B \overset{\ell_1}{\Rightarrow}_{\texttt{left}} \alpha_1}{(p, \alpha_1, B) \vdash_P^\star (p, \varepsilon, \varepsilon)} \quad \dfrac{C \overset{\ell_2}{\Rightarrow}_{\texttt{left}} \alpha_2}{(p, \alpha_2, C) \vdash_P^\star (p, \varepsilon, \varepsilon)}}{(p, \alpha_1\alpha_2, BC) \vdash_P^\star (p, \varepsilon, \varepsilon)}}{(p, \underbrace{\alpha_1\alpha_2}_{\alpha}, A) \vdash_P^\star (p, \varepsilon, \varepsilon)}$$

За (б), индукция по броя на стъпките ℓ в изчислението на стековия автомат.

Нека $\ell = 0$ и $(p, \alpha, \delta) \vdash_P^0 (p, \varepsilon, \varepsilon)$. Ясно е, че единствената възможност е $\alpha = \varepsilon$ и $\delta = \varepsilon$. Тогава $\varepsilon \overset{\star}{\Rightarrow}_{\mathtt{left}} \varepsilon$.

Нека $\ell > 0$ и $(p, \alpha, \delta) \vdash_P^\ell (p, \varepsilon, \varepsilon)$. Имаме два варианта за първата стъпка в това изчисление.

Започваме със случая, който не зависи от правилата на граматиката. Това означава, че първата стъпка от изчислението $(p, \alpha, \delta) \vdash_P^\ell (p, \varepsilon, \varepsilon)$ е направена защото $\Delta(p, a, a) \ni (p, \varepsilon)$, за някоя буква a. Тогава със сигурност можем да представим думите α и δ като $\alpha = a\beta$ и $\delta = a\rho$, за някои β и ρ, и да разбием изчислението по следния начин:

$$\dfrac{\dfrac{\Delta(p, a, a) \ni (p, \varepsilon)}{(p, a\beta, a\rho) \vdash_P (p, \beta, \rho)} \qquad (p, \beta, \rho) \vdash_P^{\ell-1} (p, \varepsilon, \varepsilon)}{(p, \underbrace{a\beta}_{\alpha}, \underbrace{a\rho}_{\delta}) \vdash_P^\ell (p, \varepsilon, \varepsilon)}$$

Сега можем да приложим индукционното предположение и да получим извода:

$$\dfrac{a \in \Sigma \qquad \dfrac{(p, \beta, \rho) \vdash_P^{\ell-1} (p, \varepsilon, \varepsilon)}{\rho \overset{\star}{\Rightarrow}_{\mathtt{left}} \beta} \text{ (и.п.)}}{\underbrace{a\rho}_{\delta} \overset{\star}{\Rightarrow}_{\mathtt{left}} \underbrace{a\beta}_{\alpha}} \text{ (\textit{Твърдение 3.16})}$$

Вторият случай зависи от правилата на граматиката. Това означава, че първата стъпка от изчислението $(p, \alpha, \delta) \vdash_P^\ell (p, \varepsilon, \varepsilon)$ е направена защото $\Delta(p, \varepsilon, A) \ni (p, \gamma)$. Според конструкцията на стековия автомат, това означава, че със сигурност имаме правилото $A \to_G \gamma$, а думата δ може да се представи като $\delta = A\rho$, за някое ρ, и изчислението може да се разбие по следния начин:

Понеже граматиката е в нормална форма на Чомски, то $1 \leq |\gamma| \leq 2$.

$$\dfrac{\dfrac{\Delta(p, \varepsilon, A) \ni (p, \gamma)}{(p, \alpha, A\rho) \vdash_P (p, \alpha, \gamma\rho)} \qquad (p, \alpha, \gamma\rho) \vdash_P^{\ell-1} (p, \varepsilon, \varepsilon)}{(p, \alpha, \underbrace{A\rho}_{\delta}) \vdash_P^\ell (p, \varepsilon, \varepsilon)}$$

Сега прилагаме индукционното предположение и получаваме извода:

$$\dfrac{\dfrac{A \to_G \gamma}{A\rho \Rightarrow_{\mathtt{left}} \gamma\rho} \qquad \dfrac{(p, \alpha, \gamma\rho) \vdash_P^{\ell-1} (p, \varepsilon, \varepsilon)}{\gamma\rho \overset{\star}{\Rightarrow}_{\mathtt{left}} \alpha} \text{ (и.п.)}}{\underbrace{A\rho}_{\delta} \overset{\star}{\Rightarrow}_{\mathtt{left}} \alpha}$$

□

7. Формулировка и доказателство на лемата за разрастване за контекстносвободни езици.

Лема 3.5 (за покачването (безконтекстни езици)). За всеки безконтекстен език L съществува $p > 0$, такова че ако $\alpha \in L$, за която $|\alpha| \geq p$, то съществува разбиване на думата на пет части, $\alpha = xyuvw$, за което е изпълнено:

1) $|yv| \geq 1$,

2) $|yuv| \leq p$, и

3) $(\forall i \in \mathbb{N})[xy^iuv^iw \in L]$.

[Sip12, стр. 125], [HU79, стр. 125], [HMU01, стр. 275], [Koz97, стр. 148].
Ще казваме, че p е константа на покачването. Тук отново да напомним, че $0 \in \mathbb{N}$ и $xy^0uv^0w = xuw$.

Доказателство. Нека G е безконтекстна граматика за езика L. Да положим

$$p \overset{\text{деф}}{=} b^{|V|+1}, \text{ където } b \overset{\text{деф}}{=} \max\{\, |\gamma| \mid A \to \gamma \text{ е правило в } G \,\}.$$

Числото b е максималната разклоненост на всяко дърво на извод за дума изводима от граматиката G.

Ще покажем, че този избор на p е удачен за удовлетворяването на условията на лемата. Ще наричаме p константа на покачването за граматиката G. Да разгледаме произволна дума $\alpha \in \mathcal{L}(G)$, за която $|\alpha| \geq p$. Понеже $\mathcal{L}(G)$ е безкраен език, то със сигурност можем да намерим такава дума. Щом $S \overset{\star}{\lhd} \alpha$, според *Твърдение 3.14*, винаги имаме $S \overset{\ell}{\lhd} \alpha$ за $\ell \geq |V| + 1$. Нека измежду всички синтактични дървета $P = (T, \lambda)$ за извода $S \overset{\ell}{\lhd} \alpha$ сме избрали такова с *минимален* брой елементи в T. Да фиксираме *максимален* път π в T, т.е. дума $\pi \in T$ и $|\pi| = \ell$. Чрез функцията λ пътя π определя редицата $X_0, X_1, \ldots, X_\ell$, като $X_i = \lambda(\rho)$, където $\rho \prec \pi$ и $|\rho| = i$. Ясно е, че $X_i \in V$ за $i < \ell$ и $X_\ell \in \Sigma \cup \{\varepsilon\}$. Щом $\ell \geq |V| + 1$, то това означава, че по този път π се срещат $\ell \geq |V| + 1$ на брой променливи. От принципа на Дирихле следва, че трябва да има повторения на променливи по този път. Във вече фиксираното синтактично дърво P можем да изберем това срещане на двойка повтарящи се променливи по пътя π, както на *Фигура 3.8a*, което е възможно най-близо до края на пътя. Това означава, че можем да разбием извода $S \overset{\ell}{\lhd} \alpha$ като $S \overset{\star}{\lhd} xAw$ и $A \overset{\leq|V|+1}{\lhd} \gamma$, където $\alpha = x\gamma w$. Сега според нашия избор, изводът $A \overset{\leq|V|+1}{\lhd} \gamma$ може да се разбие като $A \overset{\leq|V|+1}{\lhd} yAv$ и $A \overset{\leq|V|}{\lhd} u$, където $\gamma = yuv$. Да обобщим. Можем да разбием думата α на пет части като $\alpha = xyuvw$ с изводите:

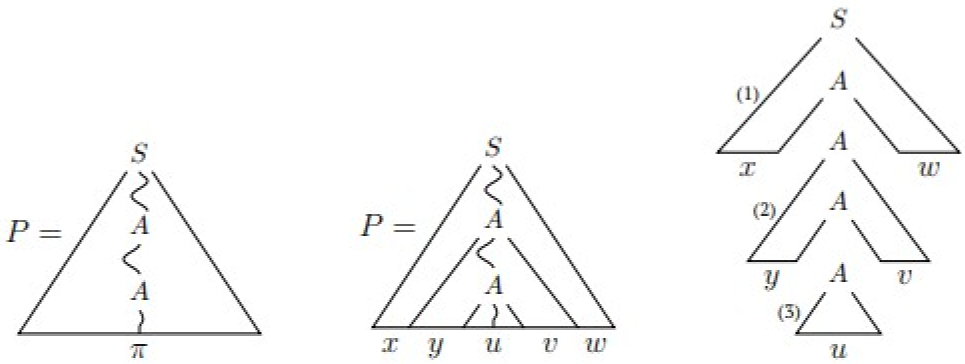
Важното е да вземем дума $\alpha \in \mathcal{L}(G)$, за която $|\alpha| > b^{|V|}$.

Принципът на Дирихле е известен също и като принципа на чекмеджетата.

(1) $S \overset{\star}{\lhd} xAw$,

(2) $A \overset{\leq|V|+1}{\lhd} yuv$, защото в дървото P можем да изберем първата двойка повтарящи се променливи, които срещнем отзад напред по пътя π. Тук сме означили тази повтаряща се променлива с .

(3) $A \overset{\leq|V|}{\lhd} u$, като тук вече няма повтарящи се променливи по останалата част от пътя π.



(а) Първата двойка повтарящи се променливи, които намерим като тръгнем отдолу нагоре по пътя π.

(б) Разбиваме дървото на три части.

(в) Получаваме три отделни синтактични дървета.

Сега остава да проверим условието на лемата:

- Да допуснем, че $|yv| = 0$. Това означава, че $\alpha = xuw$ и имаме извода:

$$\frac{S \overset{\star}{\lhd} xAw \quad A \overset{\star}{\lhd} u}{S \overset{\star}{\lhd} \underbrace{xuw}_{\alpha},} \ (\textit{Твърдение } 3.12)$$

  за който съществува дърво на извод с по-малко на брой възли отколкото P, защото махаме средната част, която съдържа поне един възел. Това е противоречие с избора на P като синтактично дърво за $S \overset{\star}{\lhd} \alpha$ с минимален брой възли. Заключаваме, че $|yv| \geq 1$.

- Понеже имаме извода $A \overset{\leq |V|+1}{\lhd} yuv$, то от *Твърдение 3.14* следва, че $|yuv| \leq b^{|V|+1} = p$.

- За произволно $i \in \mathbb{N}$ имаме извода:

$$\frac{\dfrac{\dfrac{A \overset{\star}{\lhd} yAv}{A \overset{\star}{\lhd} y^iAv^i} \ (\textit{Твърдение } 3.13) \quad A \overset{\star}{\lhd} u}{A \overset{\star}{\lhd} y^iuv^i} \ (\textit{Твърдение } 3.12) \qquad S \overset{\star}{\lhd} xAw}{S \overset{\star}{\lhd} xy^iuv^iw} \ (\textit{Твърдение } 3.12)$$

Оттук заключаваме, че $(\forall i \in \mathbb{N})[xy^iuv^iw \in \mathcal{L}(G)]$.

$\square$

- Цеци учебник

Теорема 13.1. *По всяка конт.-свободна граматика G можем да построим стеков автомат M, такъв че $L(G) = L_s(M)$.*

Доказателство. Нека е дадена к.-с. граамтика $G =< V, \Sigma, R, s >$ в НФЧ.

Построяваме стеков автомат $M =< Q = \{q\}, \Sigma, \Gamma = \Sigma \cup V, \delta, s = q, \# = s, F = \emptyset >$, който ще разпознава $L(G)$.

I. $\delta(q, \epsilon, A) \ni (q, \alpha)$, ако $A \Rightarrow \alpha \in P$

II. $\delta(q, a, a) \ni (q, \epsilon)$, $a \in \Sigma$

За да покажем, че $L(M) = L(G)$, ще докажем лема:

Лема 13.1. *Ако $w \in \Sigma^*, \alpha \in \{\epsilon\} \cup V(\Sigma \cup V)^*$, то*

$s \implies^* w\alpha$ *(ляв извод)* $\iff (q, w, S) \vdash^* (q, \epsilon, \alpha)$

Следствие от Лемата: $\alpha = \epsilon \; \forall \; w \in \Sigma^*$

$w \in L(G) \iff S \implies^* w$ *(ляв извод)* $\iff (q, w, s) \vdash^* (q, \epsilon, \epsilon) \iff w \in L(M)$

Доказателство. на лемата

$(\implies)$ Нека $S \xrightarrow{n} w\alpha$. $w \in \Sigma^*$, $\alpha \in \{\epsilon\} \cup V(V \cup \Sigma)^*$

Тогава има извод $u_0 = S \implies u_1 \ldots \implies u_n = w\alpha$.

С индукция по n (дължината на извода) ще покажем, че $(q, w, S) \vdash^* (q, \epsilon, \alpha)$

$S \implies {}^*w\alpha \longrightarrow (q, w, S) \vdash^* (q, \epsilon, \alpha)$:

1) n = 0:

$u_0 = S = w\alpha \longrightarrow w = \epsilon \& \alpha = S \longrightarrow (q, w, S) \vdash^* (q, \epsilon, \alpha)$

$(q, w, s) \vdash^* (q, \epsilon, \alpha)$

$(q, w = \epsilon, s) \vdash^* (q, \epsilon, \alpha = s)$ ✓

$s \equiv w\alpha \; w = \epsilon \;\&\; \alpha = s$

2) Нека твърдението е вярно за n.

$S \xrightarrow{n+1} w\alpha$ най-ляв извод

$s \xrightarrow{n} xAB \implies w\alpha, \;\; x \in \Sigma^*, A \in V, B \in (\Sigma \cup V)^*$

$A \longrightarrow \gamma \in P$

$xAB \implies x\gamma B = w\alpha$

Сл.

$w = xy$ за някое $y \in \Sigma^*$

$x\gamma B = xy\alpha \longrightarrow \sigma B = y\alpha$

За $s \xrightarrow{n} x(= w)AB(\alpha)$ по И.П.

$(q, x, s) \vdash^* (q, \epsilon, AB)$

Същият преход:

$(q, xy, S) \vdash^* (q, y, AB) \vdash (q, xy, S) \vdash (q, \epsilon, \alpha)$ $y \in \Sigma^*$

y пъти преходи от II. тип

(II $(q, a, a) \vdash (q, \epsilon, \epsilon)$

$A \longrightarrow \gamma \in P$

($\Longleftarrow$)

Нека $(q, w, s) \vdash^* (q, \epsilon, \alpha)$

Индукция по броя n на преходите от тип I ($q, \epsilon, A) \vdash (q, \epsilon, \gamma),\ A \longrightarrow \gamma \in P$)

1) n = 0

S не е терминал $\longrightarrow$ няма преходи

$w = \epsilon, \alpha = S$

$S \xrightarrow{*} w(= \epsilon)\alpha(= S)$ т.е. $S \xrightarrow{*} S$ ✓

2) $n \ \longrightarrow \ n+1$ n пъти тип I. $(\star)(q, w, S) \vdash (q, y, AB) \vdash^{n+1} (q, y, \gamma B) \vdash^* (q, \epsilon, \alpha)$

преходи от тип II.

$A \longrightarrow \gamma \in P,\ y \in \Sigma^*$

$w = xy$ за някое $x \in \Sigma^*$

y е начало на γB

$\gamma B = y\alpha$

От ($\star$:

n пъти преходи от тип I.

$q, x(= w), S) \vdash (q, \epsilon, AB(= \alpha))$

И.П. $\implies S \implies {}^* xAB \implies x\gamma B = xy\alpha = w\alpha$

$S \implies w\alpha$

- Проверете за правотата на доказателството за всеки случай със записките на проф. Соскова.

:)))

□

8. Примери с доказателства за езици, които не са контекстносвободни.

Следствие 3.3. Нека L е произволен **безкраен** език. Нека също така е изпълнено, че:

Ако L е краен език, то е ясно, че L е безконтекстен.

(∀) за *всяко* естествено число $p \geq 1$,

(∃) можем да намерим дума $\alpha \in L$, $|\alpha| \geq p$, такава че

(∀) за *всяко* разбиване на думата на пет части, $\alpha = xyuvw$, със свойствата $|yv| \geq 1$ и $|yuv| \leq p$,

(∃) можем да намерим $i \in \mathbb{N}$, за което е изпълнено, че $xy^iuv^iw \notin L$.

Тогава L **не** е безконтекстен език.

Пример 3.9. Да видим защо езикът $L = \{a^nb^nc^n \mid n \in \mathbb{N}\}$ не е безконтекстен.

(∀) Разглеждаме произволна константа $p \geq 1$.

(∃) Избираме дума $\alpha \in L$, $|\alpha| \geq p$. В случая, нека $\alpha = a^pb^pc^p$.

(∀) Разглеждаме произволно разбиване на α на пет части $\alpha = xyuvw$, за което $|yuv| \leq p$ и $1 \leq |yv|$.

(∃) За всяко такова разбиване ще посочим i, за което $xy^iuv^iw \notin L$. Знаем, че поне едно от y и v не е празната дума. Имаме няколко случая за y и v.

- y и v са думи съставени от една буква. В този случай получаваме, че xy^2uv^2w има различен брой букви a, b и c.
- y или v е съставена от две букви. В този случай получаваме, че в xy^2uv^2w редът на буквите е нарушен.
- понеже $|yuv| \leq p$, то не е възможно в y или v да се срещат и трите букви.

Оказа се, че във всички възможни случаи за y и v, $xy^2uv^2w \notin L$.

Така от *Следствие* 3.3 следва, че езикът L не е безконтекстен.

Задача 3.10. Докажете, че езикът $L = \{a^i b^j c^k \mid 0 \leq i \leq j \leq k\}$ не е безконтекстен.

Упътване. Прилагаме *Следствие 3.3*.

($\forall$) Разглеждаме произволна константа $p \geq 1$.

($\exists$) Избираме конкретна дума $\alpha \in L$, $|\alpha| \geq p$. В случая, нека $\alpha = a^p b^p c^p$.

($\forall$) Разглеждаме произволно разбиване $xyuvw = \alpha$, за което $|yuv| \leq p$ и $1 \leq |yv|$. Знаем, че поне една от y и v не е празната дума.

($\exists$) Ще посочим конкретно $i \in \mathbb{N}$, за което $xy^i uv^i w \notin L$.

Обърнете внимание, че тук i съществено зависи от вида на разбиването.

- y и v са съставени от една буква. Имаме три случая.
  - i) a не се среща в y и v. Тогава $xy^0 vu^0 w$ съдържа повече a от b или c.
  - ii) b не се среща в y и v.
    - Ако a се среща в y или v, тогава $xy^2 uv^2 w$ съдържа повече a от b.
    - Ако c се среща в y или v, тогава $xy^0 uv^0 w$ съдържа по-малко c от b.
  - iii) c не се среща в y и v. Тогава $xy^2 uv^2 w$ съдържа повече a или b от c.
- y или v е съставена от две букви. Тук разглеждаме $xy^2 uv^2 w$ и съобразяваме, че редът на буквите е нарушен.

9. Доказателство за незатвореността на контекстносвободните езици относно допълнение и сечение.

> **Твърдение 3.15.** Безконтекстните езици **не** са затворени относно операциите сечение и допълнение.

Упътване. Да разгледаме езика $L_0 = \{a^n b^n c^n \mid n \in \mathbb{N}\}$, за който вече знаем от *Пример 3.9*, че не е безконтекстен. Да вземем също така и безконтекстните езици

$$L_1 = \{a^n b^n c^m \mid n, m \in \mathbb{N}\},\ L_2 = \{a^m b^n c^n \mid n, m \in \mathbb{N}\}.$$

- Понеже $L_0 = L_1 \cap L_2$, то заключаваме, че безконтекстните езици не са затворени относно операцията сечение.
- Да допуснем, че безконтекстните езици са затворени относно операцията допълнение. Тогава $\overline{L}_1$ и $\overline{L}_2$ са безконтекстни. Знаем, че безконтекстните езици са затворени относно обединение. Следователно, езикът $L_3 = \overline{L}_1 \cup \overline{L}_2$ също е безконтекстен. Понеже допуснахме, че безконтекстните са затворени относно допълнение, то $\overline{L}_3$ също е безконтекстен. Но тогава получаваме, че езикът

  $$L_0 = L_1 \cap L_2 = \overline{\overline{L}_1 \cup \overline{L}_2} = \overline{L}_3$$

  е безконтекстен, което е противоречие.

$\square$