Граматични дървета

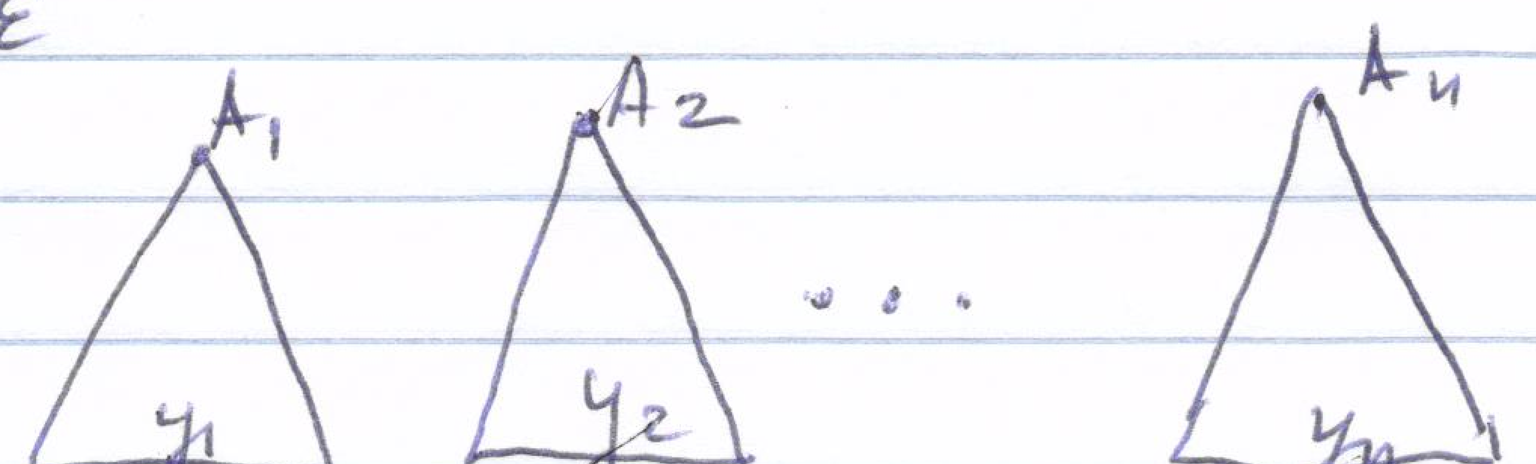
Да разгледаме следния пример на КСГ:
$G=(V,\Sigma,R,S)$, където $\Sigma=\{(,)\}$, $V=\{S\}\cup\Sigma$, $R=\{S\to\varepsilon, S\to SS, S\to(S)\}$ и два извода
$S\Rightarrow SS\Rightarrow (S)S\Rightarrow (S)(S)\Rightarrow (S)((S))\Rightarrow (S)(())\Rightarrow ()(())$ и
$S\Rightarrow SS\Rightarrow (S)S\Rightarrow ()S\Rightarrow ()(S)\Rightarrow ()((S))\Rightarrow ()(())$, с които по различен начин и различна последователност на използването на правилата от S получаваме една и съща дума $()(())$, която се поражда за $L(G)$. Така по различен начин получаваме един и същи резултат. Ще дадем следната дефиниция на граматично дърво.

Д. Нека $G=(V,\Sigma,R,S)$ е произволна КСГ. Индуктивно определяме понятието граматично дърво

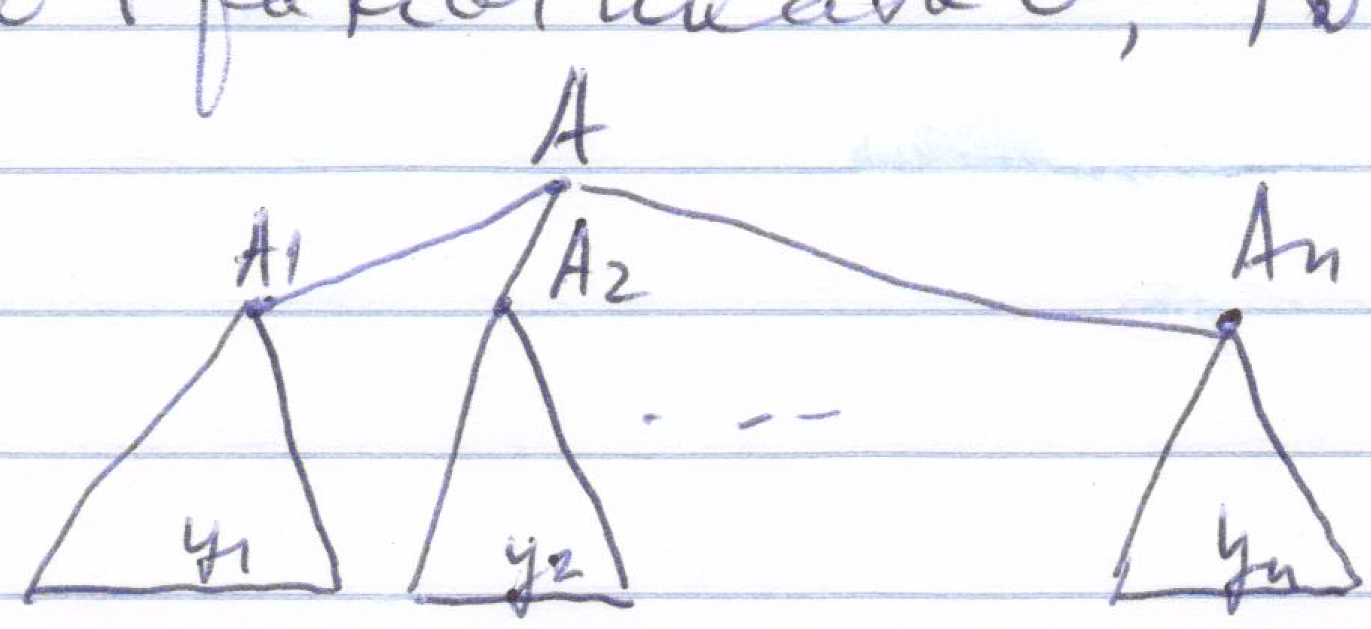
1. За всяко $a\in\Sigma$, $\bullet a$ е граматично дърво с корен a и листо a (тривиалното дърво)
2. Ако $A\to\varepsilon$ е правило в G, където $A\in V\setminus\Sigma$, то

е граматично дърво с корен A и листо ε

3. Ако



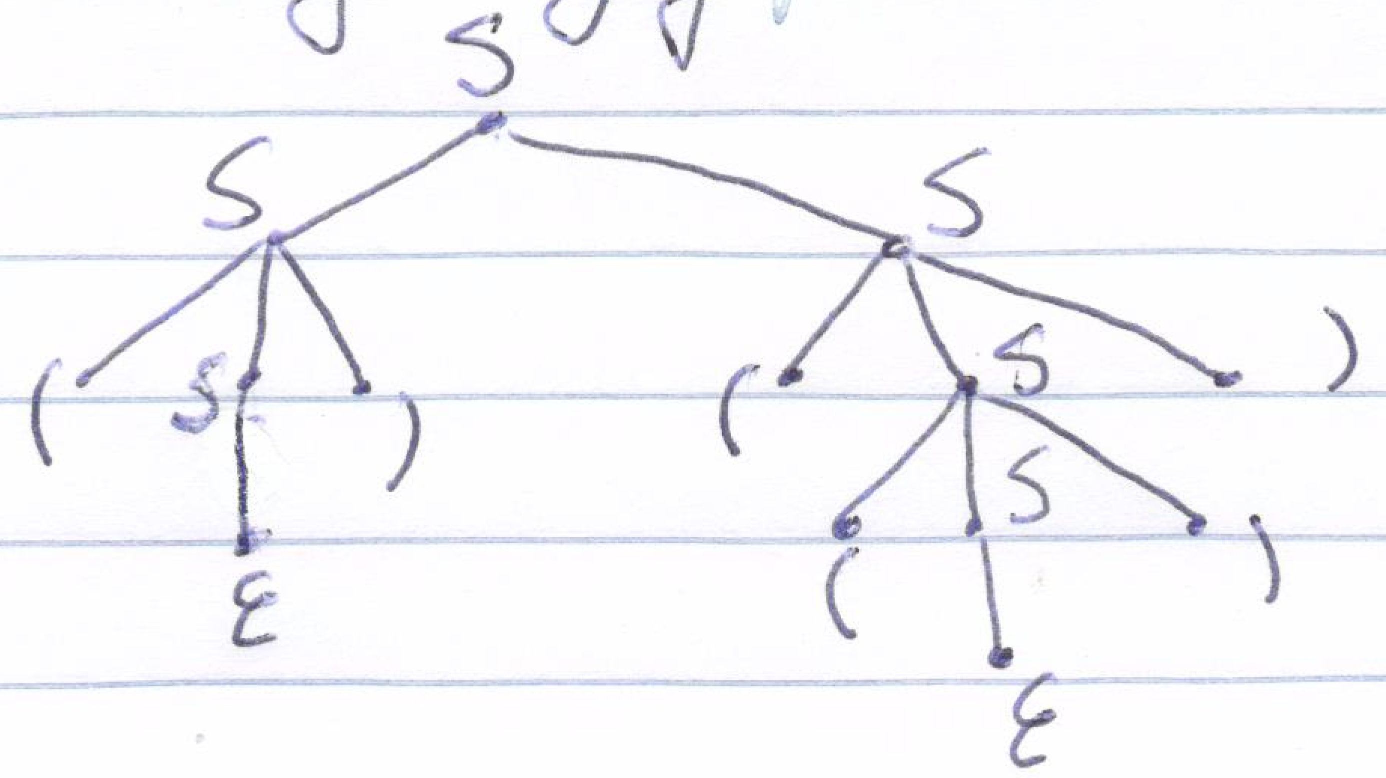
са граматични дървета с корени $A_1,\dots,A_n$ съответно и $y_1,\dots,y_n$ са думи в азбуката Σ, получени съответно от дърветата с корени $A_1,\dots,A_n$ съответно и $A\to A_1A_2\dots A_n$ е правило в граматиката G, то



е граматично дърво.

Тук думата y_i се получава от дървото с корен A_i, y_i представлява думата, получена от всички листа на дървото, получена в същия ред, в който се намират в дървото с

корен A_i (отляво надясно). В такъв случай казваме, че дървото с корен A_i извежда думата y_i. А окончателното дърво с корен A извежда думата $y_1y_2\ldots y_n$.
Сега да разгледаме граматично дърво за граматиката $G=(V,\Sigma,R,S)$, с която започнахме днешните разглеждания, което извежда думата ()(()):



Обърнете внимание, че в граматичните дървета всички възли, които не са листа, имаме само нетерминали. Възможностите да получаваме една и съща дума по различни начини ни подсказва, че може би те са "еквивалентни" в някакъв смисъл. Ще дадем следната дефиниция:

D. Нека $G=(V,\Sigma,R,S)$ е КСГ и нека $D = x_1 \Rightarrow x_2 \Rightarrow \ldots \Rightarrow x_n$ и $D' = x_1' \Rightarrow x_2' \Rightarrow \ldots \Rightarrow x_n'$ са два извода в G, където $x_i, x_i' \in V^*$, $i=1,\ldots,n$; $x_1, x_1' \in V\setminus\Sigma$ и $x_n, x_n' \in \Sigma^*$. Казваме, че D предхожда D' и пишем $D \prec D'$, ако $n>2$ и съществува естествено число k, $1<k<n$ такова, че са изпълнени следните 5 условия:

(1) За всяко i, $1\le i\le n$, $i\ne k$ имаме $x_i = x_i'$;
(2) $x_{k-1}=x'_{k-1}=uAvBw$, където $u,v,w\in V^*$ и $A,B\in V\setminus\Sigma$;
(3) $x_k = uyvBw$, където $A\to y$ е правило от G;
(4) $x_k' = uAvzw$, където $B\to z$ е правило от G;
(5) $x_{k+1}=x'_{k+1}=uyvzw$.

С други думи един извод предхожда друг, ако в изводите те се различават само на едно място и в този извод, който предхожда другия, по-напред заместваме нетерминала, който е по-вляво.

Връщайки се към примера, от началото на темата за граматичните дървета нека

$D_1 = S \Rightarrow SS \Rightarrow (S)S \Rightarrow ()S \Rightarrow ()(S) \Rightarrow ()((S)) \Rightarrow ()(())$

$D_2 = S \Rightarrow SS \Rightarrow (S)S \Rightarrow (S)(S) \Rightarrow ()(S) \Rightarrow ()((S)) \Rightarrow ()(())$

$D_3 = S \Rightarrow SS \Rightarrow (S)S \Rightarrow (S)(S) \Rightarrow (S)((S)) \Rightarrow ()((S)) \Rightarrow ()(())$

Очевидно е, че $D_1 < D_2 < D_3$, но D_1 не предхожда D_3, т.е. тази релация $<$ не е транзитивна.

Д. Нека D и D' са два извода на една и съща дума в граматиката G. Казваме, че те са подобни ако (D, D') принадлежи на рефлексивното, симетричното и транзитивно затваряне на релацията $<$. Следователно, релацията подобни изрази е релация на еквивалентност. Да видим кои са останалите изводи на думата $()(())$. Това са

$D_4 = S \Rightarrow SS \Rightarrow (S)S \Rightarrow (S)(S) \Rightarrow (S)((S)) \Rightarrow (S)(()) \Rightarrow ()(())$

$D_5 = S \Rightarrow SS \Rightarrow S(S) \Rightarrow (S)(S) \Rightarrow (S)((S)) \Rightarrow (S)(()) \Rightarrow ()(())$

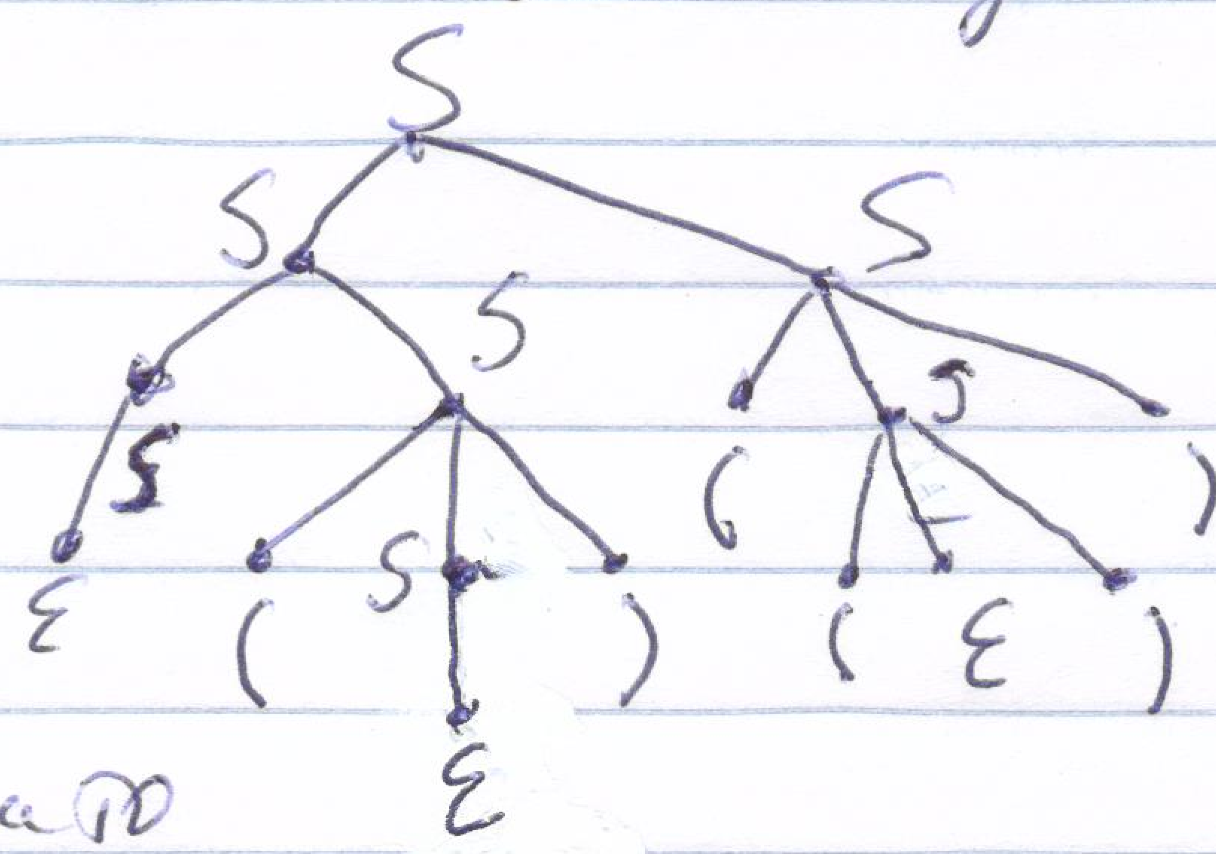
$D_6 = S \Rightarrow SS \Rightarrow S(S) \Rightarrow S((S)) \Rightarrow (S)((S)) \Rightarrow (S)(()) \Rightarrow ()(())$

$D_7 = S \Rightarrow SS \Rightarrow S(S) \Rightarrow S((S)) \Rightarrow S(()) \Rightarrow (S)(()) \Rightarrow ()(())$

и може би още? Колко са и кои са те? Каква е релацията $<$ м/у някои от тях?

Да обърнем внимание, че $<$ можем да го разширим до строга наредба, като вземем само транзитивното затваряне на $<$. Ако във всеки клас на еквивалентност относно подобие, разгледаме "най-малкия" елемент относно транзитивното затваряне на $<$, разглеждано като строга наредба, то той се нарича най-ляв извод, а най-големия елемент относно посочената строга наредба се нарича най-десен извод. Така ние пишем $x \overset{L}{\Rightarrow} y$ т.т.к. $x = wA\beta$, $y = w\alpha\beta$, където $w \in \Sigma^*$, $\alpha, \beta \in V^*$, $A \in V \setminus \Sigma$ и $A \to \alpha$ е правило в граматиката G, т.е. когато искаме да намерим $x \overset{L}{\Rightarrow} y$ ние търсим най-левия нетерминал и прилагаме правило за този нетерминал. По този начин ако $x_1 \overset{L}{\Rightarrow} x_2 \overset{L}{\Rightarrow} \dots \overset{L}{\Rightarrow} x_n$ е най-левия извод на всяка стъпка пишем $x_1 \overset{L}{\Rightarrow}{}^* x_n$.

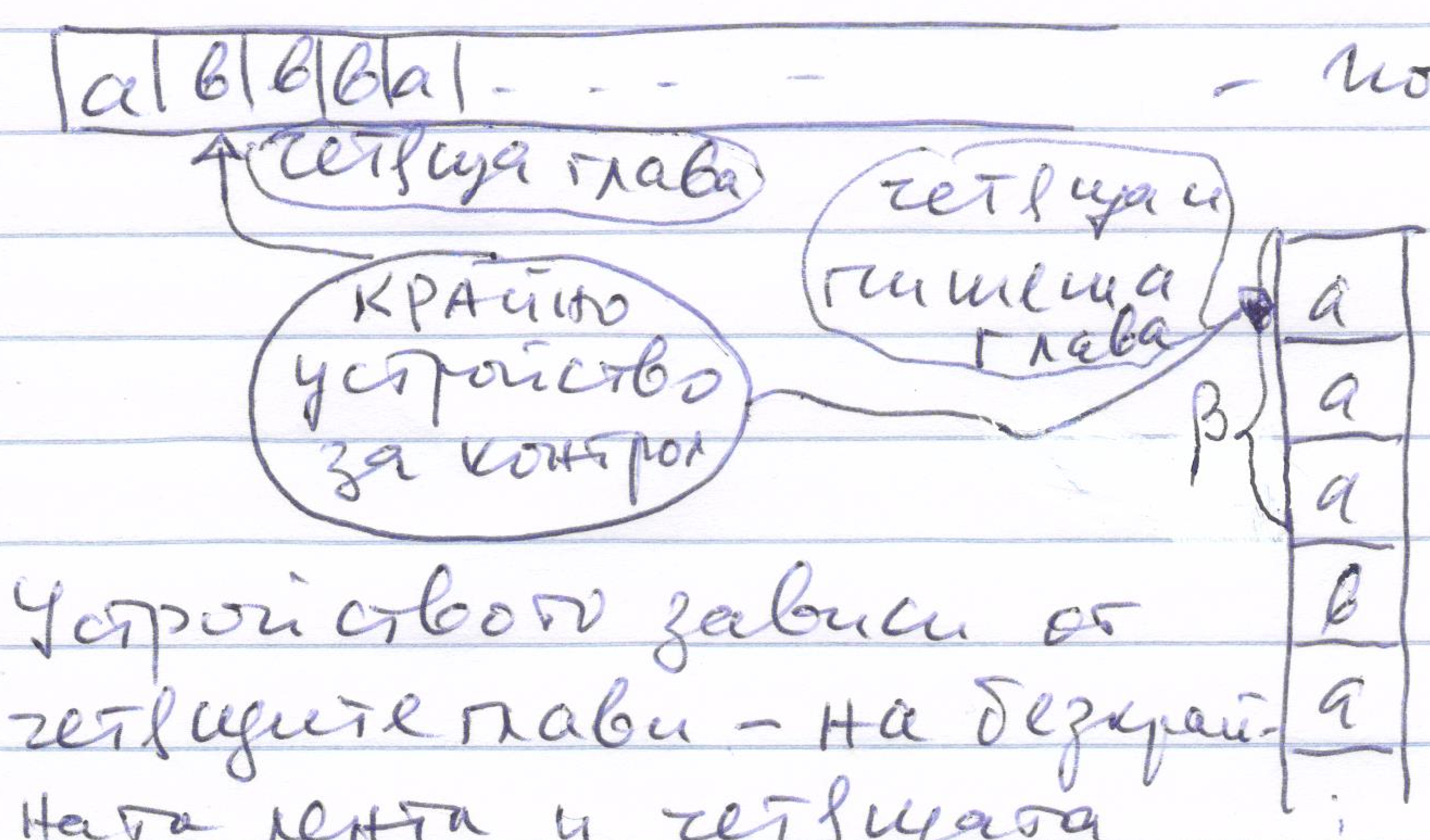
Аналогично се определя $x \overset{R}{\Rightarrow} y$ за най-десния извод за една стъпка и $x_1 \overset{R}{\Rightarrow}^{*} x_n$ за много стъпки най-десния извод.
От всичко казано по-горе ние можем да резюмираме полученото в следното
твърдение. Нека $G=(V,\Sigma,R,S)$ е КСГ и $A \in V\setminus\Sigma$, а $w \in \Sigma^{*}$. Тогава следните твърдения са еквивалентни: а) $A \Rightarrow^{*} w$;
б) Съществува граматично дърво с корен A и извод w
в) Съществува най-ляв извод $A \overset{L}{\Rightarrow}^{*} w$;
г) Съществува най-десен извод $A \overset{R}{\Rightarrow}^{*} w$.
Да обърна внимание, че не е задължително всички изводи в едно граматично дърво от даден нетерминал да отговарят на един и същ клас подобие при извод на фиксирана дума. Например, за разглежданата КСГ G можем да построим грам. д-во



което разглеждано като
извод може да бъде представено така:
$S \Rightarrow SS \Rightarrow SSS \Rightarrow SS \Rightarrow (S)S \Rightarrow ()S \Rightarrow ()(S) \Rightarrow$
$()((S)) \Rightarrow ()(())$

Стекови автомати

Нашата цел е да построим модел на разпознавател на КСЕ. Знаем, че не всеки КСЕ се разпознава от КДА (КНА). Затова ще разширим КДА до така наречените стекови автомати. Един език, който не се разпознава от КДА, но е КСЕ е следният език $L = \{ww^R \mid w \in \Sigma^*\}$ за някоя азбука Σ. Този език е КСЕ, защото той се поражда с правилата $\{S \to \varepsilon, S \to aSa, S \to bSb\}$ за $G = (V, \Sigma, R, S)$, където $\Sigma = \{a, b\}$, $V = \{S\} \cup \Sigma$, R са горните правила. Сега ще опишем изчислителния модел на стеков автомат. Интуитивно, той представлява следното:



Устройството зависи от четящите глави — на безкрайната лента и четящата глава на стека, която не се мести, тя е неподвижна.

Точната дефиниция на стеков автомат е следната:

Д. Стеков автомат се нарича шестицата $M = (K, \Sigma, \Gamma, \Delta, s, F)$, където K е крайна азбука от състояния, Σ - крайна входна (основна) азбука, Γ - крайно множество от стекови символи (стекова азбука), $F \subseteq K$, F - множество на заключителните състояния, s - начален символ, $s \in K$ и $\Delta \subseteq (K \times (\Sigma \cup \{\varepsilon\}) \times \Gamma^*) \times (K \times \Gamma^*)$ е релация на преходите, като всеки елемент на Δ се нарича правило (за преход) от Δ (от M).

Интуитивно, ако четящата прочете буквата b в/у безкрайната лента и още думата β от стека (четейки отгоре надолу) и се намира в състояние q, и има пра-

вило $((q, b, \beta), (p, \alpha))$ от Δ, то главата въу лентата се премества с една клетка надясно, ако $b \in \Sigma$ и остава неподвижна ако $b = \varepsilon$, у стройството преминава в състояние p и заместваме думата β от стека с думата α, така че α да е начало на думата, намираща се в стека, прочетена отгоре надолу. Например, ако имаме правило $((p, u, a), (q, \varepsilon)) \in \Delta$ и $a \in \Gamma$, то изтриваме a от стека, а ако $((p, u, \varepsilon)(q, a)) \in \Delta$, то добавяме a в стека.
Както и при крайните автомати ще използваме понятието конфигурация, което има нов смисъл тук.
Д. Конфигурация се нарича всеки елемент на $K \times \Sigma^* \times \Gamma^*$. Ако $(p, x, \alpha) \in K \times \Sigma^* \times \Gamma^*$, то $p \in K$, $x \in \Sigma^*$ и $\alpha \in \Gamma^*$. Всичко това се отнася винаги за конкретен стеков автомат $M = (K, \Sigma, \Gamma, \Delta, s, F)$, като за p ще си мислим като състоянието, в което се намира M, x е непрочетана част от думата, с която е запълнена лентата от самото начало, а α е думата, с която е запълнен стека. Както и при КДА (НКА) ще дефинираме $\vdash_M$ както следва:
Д. Нека $M = (K, \Sigma, \Gamma, \Delta, s, F)$ е стеков автомат и (p, x, α), $(q, y, \zeta) \in K \times \Sigma^* \times \Gamma^*$, т.е. те са конфигурации. Казваме, че от (p, x, α) за една стъпка се извежда (q, y, ζ) и означаваме $(p, x, \alpha) \vdash_M (q, y, \zeta)$ точно тогава, когато същ. правило $((p, a, \beta), (q, \gamma)) \in \Delta$ такова че $x = ay$, $\alpha = \beta\eta$ и $\zeta = \gamma\eta$ за някое $\eta \in \Gamma^*$.
С $\vdash_M^*$ означаваме рефлексивното и транзитивно затваряне на релацията $\vdash_M$ в множеството на конфигурациите.
Д. Казваме, че стековият автомат $M = (K, \Sigma, \Gamma, \Delta, s, F)$ разпознава думата $w \in \Sigma^*$ т.т.к. $(s, w, \varepsilon) \vdash_M^* (p, \varepsilon, \varepsilon)$ за някое $p \in F$. Всички конфигурации от вида $(p, \varepsilon, \varepsilon)$ такива, че $p \in F$ се наричат още заключителни.
Горната дефиниция можем да я изкажем още по следния начин: M разпознава (приема) думата $w \in \Sigma^*$ т.т.к. съществува редица от конфигурации $C_0, C_1, \ldots, C_n$ $(n > 0)$

такива, че $C_0 \vdash_M C_1 \vdash_M \dots \vdash_M C_n$, където $C_0 = (s, w, \varepsilon)$ а C_n е заключителна конфигурация.
Обърнете внимание, че в (s, w, ε), която понякога се нарича начална конфигурация и в заключителната конфигурация стека е празен. За всяка редица $C_0, C_1, \dots, C_n$ от конфигурации такива, че $C_0 \vdash_M C_1 \vdash_M \dots \vdash_M C_n$ се казва, че е изчисление (извод) с дължина n относно M. С $L(M)$ се означава $L(M) = \{w \mid w \in \Sigma^*$ и w се разпознава от $M\} = \{w \mid w \in \Sigma^*$ и $(s, w, \varepsilon) \vdash_M^* (f, \varepsilon, \varepsilon)$ за някое $f \in F\}$.
Пример 1. Ще построим стеков автомат, който разпознава езикът $L = \{ww^R \mid w \in \Sigma^*\}$. Нека $M = (K, \Sigma, \Gamma, \Delta, s, F)$, където $K = \{s, f\}$, $\Sigma = \{a, b\} = \Gamma$, $F = \{f\}$, s - начално състояние а Δ се състой от следните правила:
(1) $((s, a, \varepsilon), (s, a))$, (2) $((s, b, \varepsilon), (s, b))$, (3) $((s, \varepsilon, \varepsilon), (f, \varepsilon))$
(4) $((f, a, a)(f, \varepsilon))$, (5) $((f, b, b), (f, \varepsilon))$
Да проверим, че думата abaaba се разпознава от M.
Наистина $(s, abaaba, \varepsilon) \vdash_M (s, baaba, a) \vdash_M (s, aaba, ba) \vdash_M$ $\vdash_M (s, aba, aba) \vdash (f, aba, aba) \vdash_M (f, ba, ba) \vdash_M (f, a, a) \vdash_M (f, \varepsilon, \varepsilon)$
т.е. думата abaaba се разпознава от M.
Когато е ясно кой е автомата M ще използваме стандартно $\vdash$ и $\vdash^*$ вместо $\vdash_M$ и $\vdash_M^*$. По-горе обаче не е ясно кое правило се прилага. Затова се използват таблици като следната:

Състояние	Непрочетена част от входа	Съдържание на стека	Използвано правило
s	abaaba	ε	—
s	baaba	a	(1)
s	aaba	ba	(2)
s	aba	aba	(1)
f	aba	aba	(3)
f	ba	ba	(4)
f	a	a	(5)
f	ε	ε	(4) думата се приема

Да обърнем внимание, че при извода по-горе можехме да продължим $(s, aba, aba) \vdash (s, ba, aaba) \vdash (s, a, baaba) \vdash (s, \varepsilon, abaaba)$

т.е. тук стекови автомати отново не са детерминирани.
Пример 2. Ще построим стеков автомат, който разпознава езикът $L = \{w \mid w \in \{a,b\}^*$ и в w има равен брой a-та и b-та$\}$
Нека $M = (K, \Sigma, \Gamma, \Delta, s, F)$, където $K = \{s, q, f\}$, $\Sigma = \{a, b\}$, $\Gamma = \Sigma \cup \{c\}$, $F = \{f\}$ и Δ се състои от правилата:
(1) $((s, \varepsilon, \varepsilon)(q, c))$; (2) $((q, a, c)(q, ac))$; (3) $((q, a, a), (q, aa))$;
(4) $((q, a, b), (q, \varepsilon))$; (5) $((q, b, c), (q, bc))$; (6) $((q, b, b), (q, bb))$;
(7) $((q, b, a), (q, \varepsilon))$; (8) $((q, \varepsilon, c)(f, \varepsilon))$.
Да разгледаме думата $abbbaa$. Тогава
$(s, abbbaa, \varepsilon) \vdash (q, abbbaa, c) \vdash (q, bbbaa, ac) \vdash (q, bbaa, c)$
$\vdash (q, baa, bc) \vdash (q, aa, bbc) \vdash q(a, bc) \vdash (q, \varepsilon, c) \vdash (f, \varepsilon, \varepsilon)$
Напишете сами тези изводи в таблица, както по-горе в Пример 1, като посочите кои правила са приложени.
Забележка. Всеки краен недетерминиран автомат $M = (K, \Sigma, \Delta, s, F)$ може да се разглежда като стеков автомат $M' = (K, \Sigma, \phi, \Delta', s, F)$, където $\Delta' = \{((p, u, \varepsilon), (q, \varepsilon)) \mid (p, u, q) \in \Delta\}$, т.е. нямаме стекова азбука и никакви промени не се извършват в стека. Той (стека) просто изпълнява фиктивна роля.

Стекови автомати и КСГ

Целта ни в тази тема е да докажем следната Теорема. Класът на езиците, разпознавани от стеков автомат съвпада с класът на КСЕ.
Д. Един език L се разпознава от стеков автомат, ако съществува стеков автомат M, такъв че $L = L(M)$.
Горната теорема ще я разделим на две твърдения
Твърдение 1. Всеки контекстно свободен език се разпознава от стеков автомат.
Д-во. Нека $G = (V, \Sigma, R, S)$ е КСГ. Ще построим стеков автомат M, такъв че $L(M) = L(G)$. Нека $M = (\{p, q\}, \Sigma, V, \Delta, p, \{q\})$ където Δ се състои от следните правила за преход. Всъщност ние ще ги опишем като типове правила. Те са три типа:

(1) $((p,\varepsilon,\varepsilon),(q,S))$ само едно единствено правило;
(2) $((q,\varepsilon,A),(q,x))$ за всяко правило $A \to x$ от R (много на брой правила; точно толкова на брой, колкото са правилата в R)
(3) $((q,a,a),(q,\varepsilon))$ за всяко $a \in \Sigma$ (много правила; точно толкова на брой, колкото са буквите в Σ).
Построеният стеков автомат M в известен смисъл имитира КСГ G в стека. Не съвсем разбира се, а само частично, докато в началото на стека се появи терминал. После изтрива началото на стека, сравнявайки го с думата, която искаме да разпознаем. Причината е, че няма как да заменим нетерминал в средата на стека.
За да добием представа как работи съответния стеков автомат M, ще разгледаме следния пример
Пример 3. Нека $G=(V,\Sigma,R,S)$, където $V=\{S\}\cup\Sigma$, $\Sigma=\{a,b\}$, $R=\{S\to aSa, S\to bSb, S\to\varepsilon\}$. Тази КСГ поражда езикът $L=\{ww^R \mid w\in\Sigma^*\}$. Тогава съответния стеков автомат е $M=(\{p,q\},\Sigma,V,\Delta,p,\{q\})$ където $\Delta=$

$$
\begin{array}{ll}
\{((p,\varepsilon,\varepsilon),(q,S)), & (1)\\
((q,\varepsilon,S),(q,aSa)), & (2.1)\\
((q,\varepsilon,S),(q,bSb)), & (2.2)\\
((q,\varepsilon,S),(q,\varepsilon)), & (2.3)\\
((q,a,a),(q,\varepsilon)), & (3.1)\\
((q,b,b),(q,\varepsilon))\} & (3.2)
\end{array}
$$

Да проверим, че думата $bbaabb$ се разпознава от стековия автомат M. Наистина, $(p,bbaabb,\varepsilon)\vdash_M (q,bbaabb,S)\vdash_M (q,bbaabb,bSb)\vdash_M (q,baabb,Sb)\vdash_M (q,baabb,bSbb)\vdash_M (q,aabb,Sbb)\vdash_M (q,aabb,aSabb)\vdash_M (q,abb,Sabb)\vdash_M (q,abb,abb)\vdash_M (q,bb,bb)\vdash_M (q,b,b)\vdash_M (q,\varepsilon,\varepsilon)$, т.е. думата се разпознава от M. Добре е, този извод да го напишете таблично и да видите на всяка стъпка кое правило се прилага.
За да докажем твърдението е достатъчно да докажем следната лема.

Лема. Нека $w \in \Sigma^*$ и $\alpha \in ((V \setminus \Sigma)V^*) \cup \{\varepsilon\}$. Тогава

$$S \overset{L}{\underset{G}{\Longrightarrow^*}} w\alpha \text{ т.т.к. } (q, w, S) \vdash_M^* (q, \varepsilon, \alpha).$$

Да обър на внимание, че $\overset{L}{\underset{G}{\Longrightarrow^*}}$ означава най-ляв извод в G. Преди да докажем лемата ще видим, че от нея следва твърдението. Да допуснем, че лемата е доказана. Тогава $w \in L(G) \iff S \overset{L}{\underset{G}{\Longrightarrow^*}} w \iff (p, w, \varepsilon) \vdash_M (q, w, S) \vdash_M^* (q, \varepsilon, \varepsilon)$ и $q \in F$ (Тук $\alpha = \varepsilon$) $\iff w \in L(M)$, с което $L(G) = L(M)$. С други думи, ако докажем лемата и твърдението е доказано