GRAMÁTICAS E LINGUAGENS R*S(k)

Sérgio de Mello Schneider

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS (D.Sc)

Aprovada por:

Prof. Dr. José Lucas Mourão Rangel Netto - Presidente

Dr. Arnaldo Vieira Moura

Prof. Dr. Paulo Augusto Silva Veloso

Prof. Dr. Sérgio E. Rodrigues de Carvalho

Profa. Dra. Sueli Mendes

RIO DE JANEIRO - RJ - BRASIL
Setembro de 1987

FICHA CATALOGRÁFICA

SCHNEIDER, SÉRGIO DE MELLO

Gramáticas e Linguagens R*S(k). (Rio de Janeiro) 1987

VI, 29,7 cm 79 p. (COPPE/UFRJ, D.Sc. Engenharia de Sistemas de Computação, 1987)

Tese - Universidade Federal do Rio de Janeiro, COPPE.

1. Construção de Compiladores I. COPPE/UFRJ II. Título (série)

AGRADECIMENTOS

Ao Prof. José Lucas Rangel, bem como aos demais membros da banca, por diversas sugestões e observações ao longo desse trabalho.

Aos colegas da Universidade Federal de São Carlos, pela liberação dentro do programa do PICD/CAPES.

Aos colegas da Universidade Federal do Rio de Janeiro, pelo incentivo ao longo desse tempo todo.

À Kátia Mara Daud, por tanta paciência e concentração na datilografia deste texto.

À Marilena, ao Marcelo e ao Bruno, pelo carinho com que me cercaram esse tempo todo.

Resumo da Tese Apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc)

GRAMÁTICAS E LINGUAGENS R*S(k)

Sérgio de Mello Schneider

Setembro de 1987

Orientador: Prof. Dr. José Lucas Mourão Rangel Netto
Programa: Engenharia de Sistemas e Computação

Esta tese apresenta um novo método de análise sintática ascendente, da esquerda para a direita, para gramáticas livres de contexto. O modelo matemático do analisador é um autômato com pilha que permite mudanças de configuração após a consulta ao estado no topo da pilha, à cadeia "lookahead" e, (em caso de ações de redução), ao estado descoberto após a retirada ao lado direito da produção. São provadas relações entre gramáticas LR(k) e R*S(k) e entre linguagens LR(k) e R*S(k). São fornecidos os algoritmos que geram as funções que alteram as configurações do autômato analisador R*S a partir de uma gramática dada. É fornecido um teste de consistência desses autômatos que, se satisfeito, indica que o autômato em questão é determinístico.

Abstract of Thesis presented to COPPE/UFRJ as partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc)

R*S(k) GRAMMARS AND LANGUAGES

Sérgio de Mello Schneider

September, 1987

Chairman : Prof. Dr. José Lucas Mourão Rangel Netto
Department: Engenharia de Sistemas e Computação

This thesis presents a new bottom-up, left-to-right parsing method for context free grammars. The mathematical model of the parser is a pushdown automaton, which allows changes of configuration after consulting the state on top of stack, the lookahead string and (in cases of reduction moves) the state uncovered after removal of the "handle". The relationship between LR(k) and R*S(k) grammars, between LR(k) and R*S(k) languages is proved. Algorithms are given, which produce the parsing functions of an automaton built for a given grammar. A consistency test is presented, which indicates whether a given R*S parsing automaton is deterministic or not.

Í N D I C E

CAPÍTULO I - INTRODUÇÃO 01

I.1 - Notações e Definições 03

CAPÍTULO II- TEORIA GERAL DE ANÁLISE R*S 09

II.1 - Análise R*S 09
II.2 - Condição LR(k) e Condição R*S(k). 10
II.3 - Autômato Analisador R*S(k) Geral 17
II.4 - Propriedades Fundamentais da Análise R*S 20

CAPÍTULO III - AUTÔMATOS R*S CANÔNICOS 40

III.1 - Introdução 40
III.2 - Construção de Autômatos R*S Canônicos 40

CAPÍTULO IV - CONSISTÊNCIA E GERAÇÃO DE AUTÔMATOS R*S(1) CANÔNICOS 59

IV.1 - Consistência dos Autômatos R*S(1) Canônicos 59
IV.2 - Algoritmos para a construção de um autômato analisador R*S(1) Canônico 70

CAPÍTULO V - CONCLUSÕES E OBSERVAÇÕES 74

REFERÊNCIAS BIBLIOGRÁFICAS 76

CAPÍTULO 1

INTRODUÇÃO

Este trabalho descreve um novo tipo de analisador sintático empilha-reduz, ascendente, para gramáticas livres de contexto, em que se procura simultaneamente acelerar o processo de análise e diminuir o tamanho das tabelas do analisador através dos seguintes mecanismos:

1) as decisões sobre as ações de redução levam em consideração o estado descoberto pelo desempilhamento da produção envolvida;

2) reduções por produções simples (produções do tipo $A \to B$ onde A e B são não-terminais) são pré-calculadas e não gastam efetivamente nenhum tempo;

3) cada passo do processo de análise é decidido a partir do conteúdo da pilha de análise e do conhecimento da cadeia **lookahead** de símbolos terminais, a qual durante todo o processo de análise tem um valor pré-fixado k, maior ou igual a zero. Ao fim de uma redução por produção não simples, seguida por uma sequência de reduções simples, o primeiro símbolo de cadeia **lookahead** é automaticamente empilhado.

O nome R*S deriva da ação característica do método que é a de produzir um empilhamento (S ou **shift**) após uma sequência de zero ou mais reduções (R*).

O analisador R*S pode emitir a sequência de análise (**parse**) completa, permitindo que as ações semânticas sejam associadas mesmo às produções simples. Mantém ainda a propriedade dos prefixos corretos e a implementação do analisador pode ser feita de forma que opere com características de um analisador com a propriedade de detecção imediata de erros, qualquer que seja a gramática analisável pelo método.

A descrição formal do método R*S será baseada no fato de que o conjunto das possíveis configurações (formadas pelos conteúdos de pilha do analisador mais os k símbolos da cadeia de entrada) é regular no caso da análise R*S, e que os autômatos analisadores são autômatos finitos formados por estados que são por sua vez formados por conjuntos de ítens. Esta técnica já foi explorada por KNUTH em seu artigo original sobre LR(15), usada depois por DEREMER em (9), por BACKHOUSE em (6) e por S. HEILBRUNNER, (10) e (11); ela permite uma descrição concisa e formal de métodos de análise sintática sem retirar do texto a clareza e a simplicidade das idéias.

O empilhamento automático do primeiro símbolo da cadeia que serviu de **lookahead** para uma seqüência de reduções, mais a redução múltipla por uma regra não-simples seguida por várias reduções de regras simples tornam esse método diferente dos trabalhos anteriores sobre análise sintática, de modo geral, e diferente também dos métodos de compactação de tabelas LR que operam por eliminação das reduções por produção simples.

A organização deste trabalho é a seguinte: no capítulo II veremos uma Teoria Geral R*S, na qual são demonstrados alguns resultados válidos para analisadores R*S. No capítulo III descrevemos os autômatos canônicos e sua construção para $k > 0$. No capítulo IV apresentamos um teste de consistência que indica se um autômato analisador R*S(1) canônico é ou não determinístico, bem como a prova de que esse teste é completo. Vemos também os algoritmos que geram esse autômato a partir de uma gramática dada. No capítulo V tratamos das conclusões e sugestões para a continuidade deste trabalho. Após, temos as referências bibliográficas do texto.

I.1 - NOTAÇÕES E DEFINIÇÕES

Ao longo deste trabalho G representará uma gramática livre de contexto, descrita pela quádrupla (N,Σ,P,S), conforme a notação usual. Se dissermos que G é uma gramática estendida, descrita pela quádrupla (N,Σ,P,S') é porque G foi obtida a partir de uma gramática original $G_0 = (N_0,\Sigma_0,P_0,S)$ onde $\Sigma = \Sigma_0 \cup \{\$\}$ (sendo \$ um símbolo terminal não pertencente a Σ_0), $N = N_0 \cup \{S'\}$ (sendo S' um novo símbolo não-terminal inicial, não pertencente a N_0), $P = P_0 \cup \{S' \to \$S\}$ onde $S' \to \$S$ é uma nova produção. Desta forma S' nunca aparecerá do lado direito de nenhuma produção, aparecendo do lado esquerdo de uma única produção, a produção inicial.

Utilizaremos conceitos básicos e a notação usual da análise LR como apresentados em análise LR como apresentados em AHO/ULLMAN (1),(2) e (3), os quais suporemos familiares. Faremos algumas simples modificações, descritas a seguir:

a) $\pi,\pi_1,\pi_2,$ ... representarão elementos em P^*

b) u,v representarão elementos de Σ^k

c) $\mathbb{P}(A)$ representará o conjunto dos subconjuntos de A.

Uma outra mudança que fazemos se refere à ordem de aparecimento das produções na fita de saída produzida pelo autômato analisador. É usual representarmos uma sequência de análise (**parse**) de um autômato LR como produzindo a sequência de produção da derivação mais à direita, em ordem reversa. No presente trabalho mudamos esta convenção de maneira que a sequência emitida é a sequência (em ordem direta) de produção da derivação mais à direita da sentença analisada. Para isso representaremos a concatenação da regra $A \to \beta$ com a sequência π de regras por $A \to \beta\ \pi$.

Os quantificadores serão omitidos sempre que isso não impedir a interpretação correta do texto.

Além disso usaremos R^* para indicar o fecho transitivo-reflexivo de uma relação binária R qualquer e R^+ para indicar o fecho transitivo. Os símbolos $\uparrow$ e $\downarrow$ representarão **"não-definido"** e **"definido"**, respectivamente.

Definição I.1: um símbolo não-terminal A de uma gramática é dito estéril se não existe $x \in \Sigma^*$ tal que $A \Rightarrow^* x$.

Definição I.2: X é dito inacessível se não tivermos $S' \Rightarrow^* \alpha X \beta$ em G.

Definição I.3: dizemos que uma gramática G é reduzida se ela não possui símbolos estéreis ou inacessíveis.

Definição I.4: uma produção de G é dita simples se ela é da forma $A \rightarrow B$ onde A e B são não-terminais.

Obs: $S' \rightarrow \$S$ não é uma produção simples, por construção.

Definição I.5: dizemos que ψ deriva φ numa gramática $G = (N, \Sigma, P, S)$ e escreveremos $\psi \Rightarrow \varphi$ se $\psi = \delta_1 A \delta_2$, $\varphi = \delta_1 \beta \delta_2$ e existe em P a produção $A \rightarrow \beta$. Se essa produção for simples, indicaremos esse caso particular com a notação $\psi \Rightarrow_s \varphi$.

Quando necessário ou conveniente usaremos $\Rightarrow^n$ ou $\Rightarrow^\pi$ para indicar derivações em n passos, ou usando as regras de $\pi \in P^*$.

Definição I.6: Dizemos que G tem ambiguidade nas derivações simples (ou nas regras simples) se tivermos

$$A \Rightarrow_s^{\pi_1} B$$

$$A \Rightarrow_s^{\pi_2} B$$

com $\pi_1 \neq \pi_2$.

Equivalentemente, podemos usar as condições

a) $A \Rightarrow_s^+ A$ para algum $A \in N$

b) existem A,B,C,C' e $D \in N$ tais $A \Rightarrow_s^+ D$ através de duas sequências de derivações distintas, isto é, $A \Rightarrow_s^* B \Rightarrow_s C \Rightarrow_s^* D$, e, $A \Rightarrow_s^* B \Rightarrow_s C' \Rightarrow_s^* D$ com $C \neq C'$.

Note que tanto a) quanto b) são decidíveis de maneira simples e sua ocorrência indica ambiguidade da gramática G.

Definição I.7: para não-terminais A e B quaisquer,

$$s(A,B) = \{\sigma \in P^* \mid A \Rightarrow_s^\sigma B\}$$

Note que se G não tem ambiguidade nas produções simples então $|s(A,B)| \leq 1$ em qualquer caso e $s(A,A) = \{\varepsilon\}$.

Durante todo o texto usaremos somente derivações mais à direita, representada pelo símbolo $\Rightarrow$ em lugar de símbolos como $\underset{rm}{\Rightarrow}$ usados tradicionalmente. Quando uma derivação não for mais à direita, será feita menção a esse fato.

Definição I.8: Seja G uma gramática livre de contexto. Se $S \Rightarrow^* \alpha$ então α é dita uma forma sentencial de G.

Definição I.9: seja $\alpha A\gamma \in V^*$. Se ocorre que

$$\alpha A\gamma \Rightarrow_s^* \alpha B\gamma \Rightarrow \alpha\beta\gamma$$

com $\beta \notin N$ (ou seja, $B \rightarrow \beta$ é uma produção não simples) então dizemos que

$$\alpha A\gamma \equiv> \alpha\beta\gamma$$

Observe que todas as derivações são mais à direita.

Veremos posteriormente que os autômatos R*S terão suas configurações sempre associadas às formas sentenciais obtidas através dessa relação.

Definição I.10: seja $\varphi \in V^*$. Definimos $FIRST_k(\varphi)$ como sendo

$$FIRST_k(\varphi) = \{w \mid \varphi \Rightarrow^* w\alpha \text{ e } |w| = k\}$$

Definição I.11: seja $A \in N$. Então definiremos

$$FOLLOW_k(A) = \{w \mid S \Rightarrow \alpha A\beta \text{ e } w \in FIRST_k(\beta)\}.$$

Observação: em princípio, as derivações a que nos referimos nas duas definições anteriores (I.10 e I.11) são derivações quaisquer, isto é, não são necessariamente derivações mais à direita.

Definição I.12: Sejam $x, y \in \Sigma^*$. Dizemos que $x \equiv_k y$ se $FIRST_k(x) = FIRST_k(y)$. Aqui também simplificaremos a representação de $\equiv_k$ para $\equiv$ uma vez que está implícito um valor de k constante e fixado durante todo o processo.

Definição I.13: Seja $A \rightarrow \alpha\beta$ uma produção de P. Seja $x \in \Sigma^*$. Então definiremos $[A \rightarrow \alpha.\beta, x]$ como um ítem de tamanho k de G, onde $k = |x|$.

Seja $[A \rightarrow \alpha.\beta, x]$ um ítem, tal que $|\beta| = 0$. Dizemos nesse caso que o ítem $[A \rightarrow \alpha.\beta, x]$ é um ítem completo e será representado na forma $[A \rightarrow \alpha., x]$. Um ítem da forma $[S' \rightarrow \$.S, \$^k]$ será dito um ítem inicial. Da mesma forma o ítem $[S' \rightarrow \$S., \$^k]$ será dito ítem final.

Definição I.14: o conjunto dos ítens simples completos de tamanho k de G é dado por:

$$ISC_k(G) = \{[A \rightarrow B., u] \mid A \rightarrow B \text{ é uma produção simples}\}.$$

Definição I.15: o conjunto dos ítens LR(k) de uma gramática G é dado por

$$ILR_k(G) = \{[A\to\alpha.\beta,u] \mid A\to\alpha\beta \in P \text{ e } u \in \Sigma^k\}$$

Definição I.16: o conjunto dos ítens R*S(k) de uma gramática G é dado por

$$IR{*}S_k(G) = \{[A\to\alpha.\beta,u] \mid A\to\alpha\beta \in P,\ u \in \Sigma^k,$$

$$\text{se } \alpha\beta = X,\ X \in N, \text{ então } \beta \neq \varepsilon\}$$

Definição I.17: sejam p e q dois conjuntos de ítens. Dizemos que fecho(q) = p se p for o menor conjunto de ítens satisfazendo

1. $p \supseteq q$

2. $[A\to\alpha.B\beta,u] \in q$
   implica $[B\to.\gamma,v] \in p$ para todo
   $v \in FIRST(\beta u)$ e $B\to\gamma \in P$.

Definição I.18: seja I um ítem qualquer da forma $[A\to\alpha.\beta,u]$. Definimos então

$$\text{núcleo}([A\to\alpha.\beta,u]) = [A\to\alpha.\beta,\varepsilon].$$

A representação de ítens da forma $[A\to\alpha.\beta,\varepsilon]$ será simplificada para $[A\to\alpha.\beta]$

Estendemos a definição para um conjunto de ítens q:

$$\text{núcleo}(q) = \{[A\to\alpha.\beta] \mid [A\to\alpha.\beta,u] \in q\}$$

Definição I.19: Conjunto das produções não simples de R, $R \subseteq P$, denota-se e é definido por

$$CPNS(R) = \{A\to\alpha \in R \mid A\to\alpha \text{ não é uma produção simples}\}$$

Definição I.20: definimos uma sequência de produções não simples associada à sequência (de produções quaisquer) $\pi \in P^*$ como sendo SPNS(π), onde SPNS é o homomorfismo de P^* em P^* dado por:

$$\text{SPNS}(p) = \varepsilon \text{ se } p \text{ é uma produção simples}$$
$$\text{SPNS}(p) = p \text{ se } p \text{ é uma produção não simples}$$

Definição I.21: o contexto de redução R*S de uma produção $A \to \beta \in P$ é dado por:

a) se $A \to \beta$ é uma produção simples, então

$$CR^*S(A \to \beta) = \emptyset$$

b) se $A \to \beta$ é uma produção não-simples, então

$$CR^*S(A \to \beta) = \{\alpha\beta u \mid S'\$^k \Rightarrow^* \alpha A u w \Rightarrow \alpha\beta u w\}$$

O conceito de contexto de redução será estendido também para conjuntos de produções conforme a regra a seguir:

$$CR^*S_k(P) = \bigcup_{p \in P} CR^*S_k(p)$$

Não havendo confusão omitiremos o valor k e simplificaremos $CR^*S_k(p)$ para $CR^*S(p)$, e também omitiremos P de $CR^*S(P)$ para CR^*S.

CAPÍTULO II

TEORIA GERAL DE ANÁLISE R*S

O método R*S de análise admite três variantes. No entanto, várias propriedades desse método são comuns, seus resultados valem para uma ou mais variantes. Neste capítulo serão dadas definições e provados resultados gerais.

II.1 - ANÁLISE R*S

A análise R*S é do tipo empilha-reduz, ascendente e sua implementação usa um autômato com pilha para armazenamento das informações sobre a porção já lida da cadeia de entrada até aquele determinado instante. As decisões do analisador são tomadas a partir do conhecimento de todo o conteúdo da pilha e mais k símbolos da cadeia de entrada ($k \geq 0$). Como resultado, o autômato analisador emite a sequência de análise (parse) da porção processada. Em capítulos seguintes veremos que o conhecimento de todo o conteúdo da pilha pode ser substituído pelo conhecimento do conteúdo de apenas uma parte da pilha.

No que se segue suporemos que a gramática $G=(N,\Sigma,P,S')$ será sempre uma gramática livre de contexto, reduzida, estendida com $L(G) \neq \emptyset$. Uma sentença x que desejamos saber se pertence ou não a L(G) será submetida ao autômato analisador na forma $x\$^k$. Garantimos desta maneira que durante todo o processo de análise dessa sentença haverá sempre uma cadeia **lookahead** de k símbolos.

O valor k é constante durante todo o processo de análise. Omitiremos sua notação quando julgarmos que isso não traz confusão para a interpretação do texto.

II.2 - CONDIÇÃO LR(k) E CONDIÇÃO R*S(k)

Definição II.1: retirada de HOPCROFT/ULLMAN (13). Seja G uma gramática estendida. Diremos que G satisfaz à condição LR(k) se

$$S'\$^k \Rightarrow^* \alpha B w \Rightarrow \alpha\beta w$$

$$S'\$^k \Rightarrow^* \gamma D x \Rightarrow \gamma\delta x = \alpha\beta y$$

$$FIRST(w) = FIRST(y)$$

implica que $\alpha B y = \gamma D x$, isto é: $\alpha = \gamma$, $B = D$, $x = y$.

Definição II.2: G_1 é dita uma gramática LR(k) se sua gramática estendida satisfaz à condição LR(k).

Definição II.3: Seja G uma gramática estendida. Diremos que G satisfaz à condição R*S(k) se

$$S'\$^k \equiv>^* \alpha A w \overset{\pi_1}{\equiv>} \alpha\beta w$$

$$S'\$^k \equiv>^* \gamma C x \overset{\pi_2}{\equiv>} \gamma\delta x = \alpha\beta y$$

$$FIRST(w) = FIRST(y)$$

implica que $\alpha A y = \gamma C x$ (isto é, $\alpha = \gamma$, $A = C$, $Y = x$) e $\pi_1 = \pi_2$.

Definição II.4: G_1 é dita uma gramática R*S(k) se sua gramática estendida satisfaz à condição R*S(k).

Fato: Se G satisfaz a condição R*S(k) então G não tem ambigüidade nas derivações simples.

Se G tem ambigüidade nas derivações simples é porque é possível encontrar dois não terminais A e B tais que

$$A \overset{\pi_1}{\Rightarrow_s} B \text{ e } A \overset{\pi_2}{\Rightarrow_s} B,\ \pi_1 \neq \pi_2.$$

Portanto, se $B \overset{\pi_3}{\Rrightarrow} \beta$

$$S'\$^k \Rrightarrow^* \alpha A w \overset{\pi_1\pi_3}{\Rrightarrow} \alpha\beta w$$

$$S'\$^k \Rrightarrow^* \alpha A w \overset{\pi_2\pi_3}{\Rrightarrow} \alpha\beta w$$

com $\pi_1\pi_3 \neq \pi_2\pi_3$, não satisfazendo assim a condição R*S(k).

Fato: Se G é uma gramática LR(k) então

G não é ambígua. (prova em HOPCROFT/ULLMAN(13), pág. 182).

Fato: Se G é uma gramática LR(k), G não é ambígua e portanto G não tem ambigüidade nas produções simples.

Teorema II.5: G satisfaz à condição LR(k) sse G satisfaz à condição R*S(k).

Demonstração:

Parte 1: ($\Rightarrow$) Suponhamos que G satisfaça a condição LR(k), e portanto não tem ambigüidade nas regras simples. Supondo, preparatório para a prova por contradição, que a condição R*S(k) é violada, temos:

$$S'\$^k \Rrightarrow^* \alpha A w \overset{\pi_1}{\Rrightarrow} \alpha\beta w$$

$$S'\$^k \Rrightarrow^* \gamma C x \overset{\pi_2}{\Rrightarrow} \alpha\beta y$$

$$\text{FIRST}(w) = \text{FIRST}(y)$$

$$\pi_1 \neq \pi_2$$

Sejam

$$\pi_1 = \sigma_1\ B\to\beta,\ \text{com } s(A,B) = \{\sigma_1\}$$

$$\pi_2 = \sigma_2\ D\to\delta,\ \text{com } s(C,D) = \{\sigma_2\}$$

Então

$$S'\$^k \Rightarrow^* \alpha A w \underset{s}{\Rightarrow^*} \alpha B w$$

$$S'\$^k \Rightarrow^* \gamma C x \underset{s}{\Rightarrow^*} \gamma D x \Rightarrow \gamma\delta x = \alpha\beta y$$

Pela condição LR, $\alpha B y = \gamma D x$, $\delta = \beta$ e $B = D$.

Portanto

$$\pi_1 = \sigma_1 \quad B\to\beta \qquad\qquad s(A,B) = \{\sigma_1\}$$

$$\pi_2 = \sigma_2 \quad B\to\beta \qquad\qquad s(C,D) = \{\sigma_2\}$$

e devemos ter $\sigma_1 \neq \sigma_2$, ou, equivalentemente, $A \neq C$.

Dependendo da estrutura de σ_1 e σ_2 há dois casos a considerar.

Caso 1:

$$\sigma_1 = \sigma_3 E_1 \to F\ \sigma_5$$

$$\sigma_2 = \sigma_4 E_2 \to F\ \sigma_5$$

com $s(A,E_1) = \{\sigma_3\}$, $s(F,B) = \{\sigma_5\}$, $s(C,E_2) = \{\sigma_4\}$ e $E_1 \neq E_2$

Temos

$$S'\$^k \Rightarrow^* \alpha E_1 w \Rightarrow \alpha F w$$

$$S'\$^k \Rightarrow^* \alpha E_2 w \Rightarrow \alpha F y$$

$$\text{FIRST}(w) = \text{FIRST}(y)$$

Aplicando a condição LR, $E_1 = E_2$!

Caso 2:

$$\sigma_1 = \sigma_3 \; E \to C \; \sigma_2 \quad \text{com } s(A,E) = \{\sigma_3\}$$

Temos

$$S'\$^k \Rightarrow^* \alpha E w \Rightarrow \alpha C w$$

$$S'\$^k \Rightarrow^* \alpha C y$$

Se $\alpha Cy = S'\$^k$ então $S' = C$ e $E \to S' \in P$, o que não pode ocorrer. Portanto

$$S'\$^k \Rightarrow^* \xi F z \Rightarrow \alpha C y$$

sendo a regra do último passo uma regra não simples, como na derivação original. Aplicando a condição LR,

$$\alpha E y = \xi F z$$

o que faz com que a regra do último passo seja $F \to C$!

A condição R*S também é violada quando $\pi_1 = \pi_2$ e $\alpha Ay \neq \gamma Cx$.

Ou seja:

$$S'\$^k \Rightarrow^* \alpha A w \overset{\pi_1}{\Rightarrow} \alpha \beta w$$

$$S'\$^k \Rightarrow^* \gamma C x \overset{\pi_2}{\Rightarrow} \gamma \delta x = \alpha \beta y$$

$$\text{FIRST}(w) = \text{FIRST}(y)$$

Supondo violada a condição R*S, teríamos

$$\alpha A y \neq \gamma C x \text{ temos } A = C,\ B = D \text{ e } \beta = \delta$$

Então

$$S'\$^k \equiv>^* \alpha Aw \Rightarrow_s^* \alpha Bw \Rightarrow \alpha\beta w$$

$$S'\$^k \equiv>^* \gamma Cx \Rightarrow_s^* \gamma Dx \Rightarrow \gamma\delta x = \alpha\beta y$$

e supondo válida a condição LR, teríamos $\alpha By = \gamma Dx$ e portanto $\alpha = \gamma$, $B = D$ e $y = x$ e portanto, como $A = C$, também $\alpha Ay = \gamma Cx$. Contradição.

Parte 2: ($\Leftarrow$) Suponhamos que G satisfaz a condição R*S(k) e suponhamos que

$$\left.\begin{array}{l} S'\$^k \Rightarrow^* \alpha Bw \Rightarrow \alpha\beta w \\ S'\$^k \Rightarrow^* \gamma Dx \Rightarrow \gamma\delta x = \alpha\beta y \end{array}\right\} \quad (*)$$

$$\text{FIRST}(w) = \text{FIRST}(y)$$

Caso 1: Suponhamos que as regras não simples $B\to\beta$ e $D\to\delta$ foram aplicadas nos passos finais das derivações (*) acima. Ou seja, que

$$S'\$^k \equiv>^* \alpha Aw \Rightarrow_s^* \alpha Bw \Rightarrow \alpha\beta w$$

$$S'\$^k \equiv>^* \gamma Cx \Rightarrow_s^* \gamma Dx \Rightarrow \gamma\delta x = \alpha\beta y$$

Pela condição R*S(k) $\alpha Ay = \gamma Cx$ e $\sigma_1 \; B\to\beta = \sigma_2 \; D\to\delta$ de forma que $B = D$, $\alpha = \gamma$, $y = x$, $s(A,B) = \{\sigma_1\}$, $s(C,D) = \{\sigma_2\}$.

Caso 2: Suponhamos que as regras simples $B\to E$ e $D\to F$ foram usadas nos passos finais das derivações$^{(*)}$ acima. Temos:

$$S'\$^k \Rightarrow^* \alpha Bw \Rightarrow \alpha Ew$$

$$S'\$^k \Rightarrow^* \gamma Dx \Rightarrow \gamma Fx = \alpha Ey$$

e mais, pela condição R*S(k),

$$\alpha = \gamma, \ x = y, \ F = E,$$

ou seja:

$$S'\$^k \Rightarrow^* \alpha Bw \Rightarrow \alpha Ew$$

$$S'\$^k \Rightarrow^* \alpha Dy \Rightarrow \alpha Ey$$

o que, prosseguindo com $E \Rightarrow^*_S H \Rightarrow \varphi$ (possível, pois, G é reduzida) dá

$$S'\$^k \equiv>^* \alpha Aw \Rightarrow^*_S \alpha Bw \Rightarrow \alpha Ew \Rightarrow^*_S \alpha Hw \Rightarrow \alpha\varphi w$$

$$S'\$^k \equiv>^* \alpha Cy \Rightarrow^*_S \alpha Dy \Rightarrow \alpha Ey \Rightarrow^*_S \alpha Hy \Rightarrow \alpha\varphi y$$

Aplicando a condição R*S(k), temos $\alpha Ay = \alpha Cy$ e $\pi_1 = \sigma_1 \ B\to E \ \sigma_2 \ H\to\varphi = \pi_2 = \sigma_3 \ D\to E \ \sigma_2 \ H\to\varphi$ e, em particular $B = D$ e portanto $\alpha = \gamma$, $B = D$ e $y = x$, $s(A,B) = \{\sigma_1\}$, $s(E,H) = \{\sigma_2\}$, $s(C,D) = \{\sigma_3\}$.

Nota: Os casos 1 e 2 acima são os únicos que podem ocorrer. Suponhamos que possa ocorrer conflito entre duas regras, uma simples e outra não simples, ou seja $D\to E$ e $B\to\beta$. Veremos que esta condição se reduz ao caso 1 já analisado.

$$S'\$^k \Rightarrow^* \alpha Bw \Rightarrow \alpha\beta w$$

$$S'\$^k \Rightarrow^* \gamma Dx \Rightarrow \gamma Ex = \alpha\beta y$$

$\alpha\beta y = \gamma Ex$ força $x = zy$ para algum z. Portanto

$$\alpha\beta y = \gamma Ezy$$

$$\alpha\beta w = \gamma Ezw$$

Continuando com $E \Rightarrow^*_S F \Rightarrow \delta$ (G é reduzida),

$$S'\$^k \equiv>^* \alpha\beta w = \gamma Ezw \Rightarrow^*_S \gamma Fzw \Rightarrow \gamma\delta zw$$

$$S'\$^k \equiv\!>^* \gamma Cx \Rightarrow_S^* \gamma Dx \Rightarrow_S \gamma Ex \Rightarrow_S^* \gamma Fx \Rightarrow \gamma\delta x$$

Reescrevendo e identificando, temos:

$$S'\$^k \equiv\!>^* \gamma Exw \overset{\pi_1}{\equiv\!>} \gamma Ezw \qquad \pi_1 = s(E,F)\ F\to\delta$$

$$S'\$^k \equiv\!>^* \gamma Cx \overset{\pi_2}{\equiv\!>} \gamma\delta x \qquad \pi_2 = s(C,D)\ D\to E\ s(E,F)\ F\to\delta$$

e pela condição R*S(k) $\gamma Ex = \gamma Cx$ e $\pi_1 = \pi_2$.

Isto completa a prova do Teorema II.5.

Corolário: Toda gramática LR(k) é R*S(k) e vice-versa.

Corolário: Toda linguagem LR(k) é R*S(k) e vice-versa.

II.3 - AUTÔMATO ANALISADOR R*S(k) GERAL

Definição II.6: O autômato analisador R*S(k) geral de uma gramática estendida G é definido como $AG_k(G) = (G, ação)$ onde ação é a menor relação

$$ação: V^* \times \Sigma^k \rightarrow \mathbb{P}(\{E,H\} \cup (\{RE,RNE\} \times P^*))$$

tal que:

(1) se $\varphi ux \in CR^*S$, com $x \neq \varepsilon$
então $E \in ação(\varphi,u)$

(2) se $\alpha\beta u \in CR^*S(B\rightarrow\beta)$ e
$A \Rightarrow_S^{\sigma} B$ e
$\alpha Aux \in CR^*S$, com $x \neq \varepsilon$
então $(RE, \sigma\, B\rightarrow\beta) \in ação(\alpha\beta,u)$

(3) se $\alpha\beta u \in CR^*S(B\rightarrow\beta)$ e
$A \Rightarrow_S^{\sigma} B$ e
$\alpha Au \in CR^*S$
então $(RNE, \sigma B\rightarrow\beta) \in ação(\alpha\beta,u)$

(4) $H \in ação(\$S,\$^k)$

Os símbolos E,H,RE,RNE são abreviações para "empilhamento", "parada", "redução com empilhamento" e "redução sem empilhamento", respectivamente. Se k e G estiverem subentendidos, faremos referência a AG em lugar de $AG_k(G)$.

Definição II.7: configuração de um autômato analisador R*S geral é uma terna $(\varphi,x,\pi) \in V^* \times \Sigma^* \times P^*$ e descreve o autômato num determinado instante. φ representa o conteúdo da pilham x representa a porção não lida da cadeia de entrada e π representa a sequência de reduções ocorridas até aquele instante.

Dizemos que $(\$,x\$^k,\varepsilon)$ é uma **configuração inicial**, que $(\$S,\$^k,\pi)$ é uma **configuração final.**

Definição II.8: mudança de configuração é relação binária definida em $V^* \times \Sigma^* \times P^*$, representada pelo símbolo $\vdash$ e definida como segue:

(1) $(\varphi,ax,\pi) \vdash (\varphi a,x,\pi)$
se $E \in$ ação $(\varphi,\mathrm{FIRST}(ax))$

(2) $(\alpha\beta,ax,\pi) \vdash (\alpha Aa,x,\sigma\ B\to\beta\ \pi)$
se $(RE,\ \sigma\ B\to\beta) \in$ ação$(\alpha\beta,\mathrm{FIRST}(ax))$, $\sigma \in s(A,B)$

(3) $(\alpha\beta,x,\pi) \vdash (\alpha A,x,\ \sigma\ B\to\beta\ \pi)$
se $(RNE,\ \sigma\ B\to\beta) \in$ ação$(\alpha\beta,\mathrm{FIRST}(x))$, $\sigma \in s(A,B)$

Dizemos que (φ,y,π) é uma configuração alcançável se $(\$,xy,\varepsilon) \vdash^* (\varphi,y,\pi)$.

Observação: na Definição II.8, nenhuma mudança de configuração é induzida pela ação de parada H.

Definição II.9: linguagem aceita (reconhecida) por um analisador R*S geral AG

$$L(AG) = \{\$x \in \Sigma^* \mid (\$,x\$^k,\varepsilon) \vdash^* (\$,S,\$^k,\pi)\}$$

Definição II.10: sejam $C_0, \ldots, C_n$ configurações de um autômato analisador R*S geral, tais que

$$C_0 \vdash C_1 \vdash C_2 \vdash \ldots \vdash C_n$$

Então notaremos

$$C_0 \vdash^n C_n$$

para indicarmos que a configuração C_0 se transforma na configu ração C_n após n mudanças de configuração.

II.4 - PROPRIEDADES FUNDAMENTAIS DA ANÁLISE R*S

Os autômatos analisadores têm propriedades muito importantes e gerais. Nesta seção veremos algumas delas.

<u>Teorema (Fundamental) II.11</u>: Seja G uma gramática estendida. Seja AG o autômato analisador R*S geral construído a partir de G. Então L(AG) = L(G).

Demonstração: decorre dos Lemas II.12, II.13 e II.14 a seguir.

<u>Lema II.12</u>: Seja $\varphi ux = \varphi yv \in CR{*}S$.

Então $(\varphi, uxz, \pi) \vdash^{*} (\varphi y, vz, \pi)$

Demonstração: basta observar que pela definição do AG, a relação ação contém o valor E se o contexto u examinado é um prefixo próprio de um elemento de CR*S e que, além disso, a ação E permite uma mudança de configuração por empilhamento. Nas condições do Lema (e do Teorema) isto se mantém até que $|x|$ símbolos sejam empilhados, ou seja, até que y seja empilhada. Note que se $|x| = 0$ então as configurações são iguais.

<u>Lema II.13</u>: se $\$x \in L(G)$ então $\$x \in L(AG)$

Demonstração: se $\$x \in L(G)$ existe uma derivação direita $S'\$^k \Rightarrow^* \$x\$^k$. Como a última regra aplicada é não-simples vamos reescrever $S'\$^k \equiv>^+ \$x\$^k$ identificando os passos intermediários.

$$S'\$^k = \alpha_0 \equiv> \alpha_1 \equiv> \dots \alpha_{n-1} \equiv> \alpha_n = \$x\$^k$$

Vamos definir configurações do AG, $C_1, \dots, C_n$ da forma seguinte:

se $\alpha_{i-1} = \gamma_i A_i \; i$

e $\alpha_i \; = \gamma_i \beta_i w_i$

e π_i é tal que $\alpha_i \overset{\pi_i}{\equiv>} \$x\$^k$

então $C_i = (\gamma_i\beta_i, w_i, \pi_i)$

Vamos definir ainda $D_o = (\$, x\$^k, \varepsilon)$ a configuração inicial para $\$x$. Como $\alpha_o = S'\$^k$ e $\alpha_1 = \$S\$^k$ temos $\gamma_1 = \varepsilon$, $\beta_1 = \$S$, $w_1 = \$^k$ e portanto

$$C_1 = (\$S, \$^k, \pi_1)$$

com π_1 tal que $\$S\$^k \overset{\pi_1}{\equiv>} \$x\$^k$

(Note que w_1 é a análise direita da derivação de x na gramática não estendida).

Falta provar que:

(1) $D_o \vdash^* C_n$

(2) $C_i \vdash^* C_{i-1}$ para i = n, n-1, ..., 2.

Para provarmos (1) observamos que

$$D_o = (\$, x\$^k, \varepsilon)$$

$$\alpha_{n-1} = \gamma_n A_n w_n \Rightarrow_S^* \gamma_n B_n w_n \Rightarrow \alpha_n$$

$$\alpha_n = \gamma_n \beta_n w_n = \$x\$^k \overset{\varepsilon}{\equiv>} \$x\$^k$$

$$C_n = (\gamma_n\beta_n, w_n, \varepsilon)$$

Como $\gamma_n\beta_n$ FIRST$(w_n) \in$ CR*S$(B_n \text{-} \beta_n)$ e $\gamma_n\beta_n w_n = \$x\$^k$ então, pelo Lema II.12,

$$D_o \vdash^* C_n$$

Para provarmos (2), temos

$$\alpha_{i-1} = \gamma_i A_i w_i \overset{*}{\underset{s}{\Rightarrow}} \gamma_i B_i w_i \Rightarrow \alpha_i$$

$$\alpha_i = \gamma_i \beta_i w_i \overset{\pi_i}{\underset{\equiv}{\Rightarrow}} \$x\$^k$$

$$C_i = (\gamma_i \beta_i, w_i, \pi_i)$$

Por outro lado

$$\alpha_{i-2} = \gamma_{i-1} A_{i-1} w_{i-1} \overset{*}{\underset{s}{\Rightarrow}} \gamma_{i-1} B_{i-1} w_{i-1} \Rightarrow \alpha_{i-1}$$

$$\alpha_{i-1} = \gamma_{i-1} \beta_{i-1} w_{i-1} \overset{\pi_{i-1}}{\underset{\equiv}{\Rightarrow}} \$x\$^k$$

$$C_{i-1} = (\gamma_{i-1} \beta_{i-1}, w_{i-1}, \pi_{i-1})$$

Observando que $\gamma_i \beta_i$ FIRST$(w_i) \in CR^*S\ (B_i \to \beta_i)$ podemos ter duas situações possíveis:

2.1 : $\gamma_i A_i = \gamma_{i-1} \beta_{i-1}, w_i = w_{i-1}$

e $\gamma_i A_i$ FIRST$(w_i) \in CR^*S\ (B_{i-1} \to \beta_{i-1})$

Portanto ação$(\gamma_i \beta_i$, FIRST$(w_i)) \ni (RNE, \sigma_i\ B_i \to \beta_i)$,

$\sigma_i \in s(A_i, B_i)$ e

$$C_i = (\gamma_i \beta_i, w_i, \pi_i) \vdash (\gamma_i A_i, w_i, \sigma_i\ B_i \to \beta_i\ \pi_i) =$$

$$= (\gamma_{i-1} \beta_{i-1}, w_{i-1}, \pi_{i-1}) = C_{i-1}$$

2.2 -

$\gamma_i A_i$ é um prefixo próprio de $\gamma_{i-1} \beta_{i-1}$ e $\gamma_i A_i$FIRST(w_i) é um prefixo próprio de $\gamma_{i-1} \beta_{i-1}$ FIRST$(w_{i-1}) \in CR^*S(B_{i-1} \to \beta_{i-1})$

Temos ação($\gamma_i\beta_i$, FIRST(w_i)) $\ni$ (RE,σ_i $B_i \to \beta_i$) e $C_i = (\gamma_i\beta_i, w_i, \pi_i) \vdash (\gamma_i A_i a_i, \hat{w}_i, \sigma_i\ B_i \to \beta_i\ \pi_i) \vdash^{*}_{i}$

$\vdash (\gamma_{i-1}\beta_{i-1}, w_{i-1}, \pi_{i-1}) = C_{i-1}$

Portanto, temos:

$$D_0 \vdash^* C_n \vdash^* \dots \vdash^* C_1$$

ou

$$D_0 \vdash^* C_1$$

e $$(\$, x\$^k, \varepsilon) \vdash^* (\$S, \$^k, \pi_1)$$

onde π_1 é dado por $S \overset{\pi_1}{\Rightarrow} x$ é a sequência de análise à direita de x na gramática original.

Concluindo, $\$x \in L(AG)$ como desejado.

<u>Lema II.14</u>: (ainda dentro das condições do Teorema II.11) se $\$x \in L(AG)$ então $\$x \in L(G)$.

Demonstração: como $\$x \in L(AG)$, podemos afirmar que $(\$, x\$^k, \varepsilon) \vdash^* (\$S, \$^k, \pi)$ para algum $\pi \in P^*$. Seja $D_0 = (\$, x\$^k, \varepsilon)$. Nas configurações C_i atingidas no Lema anterior, vamos distin - guir aquelas em que o valor de ação empregado para a transição de configuração **não foi** E (em outras palavras, aquelas em que a 3ª componente da configuração é distinta da 3ª componente da configuração seguinte). Sejam (nessa ordem) C_1, C_2, ..., C_n essas configurações distinguidas.

Temos, do fato de que $\$x \in L(AG)$, que

$$(\$, x\$^k, \varepsilon) = D_0 \vdash^* C_1 \vdash^+ C_2 \vdash \dots \vdash^+ C_n =$$

$$= (\$S, \$^k, \pi)$$

onde $C_i = (\varphi_i, w_i, \pi_i)$ para $1 \leq i \leq n$. Observe que AG pode não ser determinístico, mas as configurações C_i são aquelas que conduzem ao reconhecimento de $\$x$.

Vamos definir agora :

$$\alpha_i = \varphi_i w_i \text{ para } 1 \leq i \leq n$$

Vamos mostrar que

$$S'\$^k \Rightarrow \$S\$^k = \alpha_n \equiv> \ldots \equiv> \alpha_1 = \$x\$^k$$

ou de maneira mais geral

$$\alpha_i \overset{\pi}{\equiv>} \$x\$^k \qquad 1 \leq i \leq n$$

Isto será suficiente para provar este Lema, já que $C_n = (\$S, \$^k, \pi)$ e $\alpha_n = \$S\$^k$.

A demonstração é por indução em i.

Base da indução: $i = 1$

$$C_1 = (\varphi_1, w_1, \pi_1)$$

foi obtido por uma sucessão em empilhamentos a partir de D_0. Portanto

$$\alpha_1 = \varphi_1 w_1 = \$x\$^k \text{ e } \pi_1 = \varepsilon$$

e logo

$$\alpha_1 \overset{\varepsilon}{\equiv>} \$x\$^k$$

Indução: nos outros casos, $C_i = (\varphi_i, w_i, \pi_i)$. Por construção, a transição de C_i para C_{i+1} não é de empilhamento podendo ser de um dos dois casos de redução a seguir:

caso 1: redução com empilhamento. Então

$$C_i = (\varphi_i, w_i, \pi_i)$$

$$\varphi_i = \gamma_i \beta_i$$

$$u = \mathrm{FIRST}(w_i)$$

$$w_i = a\hat{w}_i, \ \sigma_i \in s(A_i, B_i),$$

$(RE, \sigma_i\ B_i \to \beta_i) \in$ ação$(\gamma_i\beta_i, u)$ é ação usada.

Temos, pela definição de ação

$$\gamma_i \beta_i u \in CR^*S\ (B_i \to \beta_i)$$

$$\gamma_i A_i uz \in CR^*S \text{ para } z \neq \varepsilon$$

$$A_i \overset{\sigma_i}{\underset{s}{\Rightarrow}} B_i$$

e $$C_i \vdash (\gamma_i A_i a,\ \hat{w}_i, \sigma_i\ B_i \to \beta_i\ \pi_i) \vdash^* C_{i+1}$$

As últimas transições até $C_{i+1} = (\varphi_{i+1} w_{i+1}, \pi_{i+1})$ são de empilhamento.

$$\alpha_{i+1} = \sigma_{i+1} w_{i+1} = \gamma_i A_i a\hat{w}_i = \gamma_i A_i w_i$$

$$\pi_{i+1} = \sigma_i\ B_i \to \beta_i\ \pi_i \text{ e}$$

$$\alpha_{i+1} \overset{\sigma_i\ B_i \to \beta_i}{\equiv>} \alpha_i \overset{\pi_i}{\equiv>} \$x\$^k$$

ou

$$\alpha_{i+1} \overset{\pi_{i+1}}{\equiv>} \$x\$^k$$

caso 2: redução sem empilhamento. Então

$C_i = (\varphi_i, w_i, \pi_i)$

$\varphi_i = \gamma_i \beta_i$

$u = \text{FIRST}(w_i)$

$(\text{RNE}, \sigma_i\ B_i \to \beta_i) \in \text{ação}(\gamma_i \beta_i, u)$ com $\sigma_i \in s(A_i, B_i)$ é a ação aplicável.

Então pela definição de ação

$\gamma_i \beta_i u \in CR{*}S(B_i \to \beta_i)$

$\gamma_i A_i \in CR{*}S$

$A_i \overset{\sigma_i}{\Rightarrow_s} B_i$

$C_i \vdash (\gamma_i A_i, w_i, \sigma_i\ B_i \to \beta_i\ \pi_i) \Vdash^{*} C_{i+1}$

onde C_{i+1} é a configuração seguinte a C_i na sequência de configurações que reconhece \$x. O raciocínio prossegue como no caso anterior, e nos permite concluir que

$$S'\$^k \equiv> \alpha_n = \$S\$^k \overset{\pi_n}{\equiv>} \$x\$^k$$

e portanto $\$x \in L(G)$

cqd (do Lema II.14)

Corolário

$$\$S\$^k \overset{\pi}{\equiv>} \$x\$^k \quad \text{implica} \quad S \overset{\pi}{\Rightarrow} x$$

ou seja, o terceiro elemento na configuração final é sempre uma sequência de análise à direita da sentença x na **gramática original**.

Observação: a ocorrência de ambigüidade nas derivações simples não invalida o resultado do Lema II.14.

Definição II.15: seja G uma gramática. Seja AG(G,ação) o autômato analisador R*S geral construído a partir de G. Dizemos que AG(G) é determinístico se $|ação(\varphi,u)| \leq 1$.

Observação: Note que se alguma ação de empilhamento ou de redução for definida numa configuração final $(\$S,\$^k,\pi)$, teremos um conflito dessa ação com $H \in ação(\$S,\$^k)$.

Lema II.16: Se G, reduzida, tem ambigüidade nas produções simples então AG(G) não é determinístico.

Demonstração: como G tem ambigüidade nas produções simples, podemos encontrar não-terminais A e B tais que

$$A \overset{\sigma_1}{\Rightarrow_s} B, \quad A \overset{\sigma_2}{\Rightarrow_s} B, \quad \sigma_1 \neq \sigma_2$$

Seja a derivação

$$S'\$^k \equiv>^* \alpha Ax \Rightarrow_s^* \alpha Bx \Rightarrow \alpha\beta x$$

Caso 1: $S'\$^k = \alpha Ax$. Temos então $\alpha = \varepsilon$, $S' = A$, $\$^k = x$. Como a única regra em S' é $S' \to \$S$ então $\beta = \$S$, $A = B = S'$ e $\sigma_1 = \sigma_2 = \varepsilon$, não ocorrendo portanto ambiguidade na primeira derivação a partir de $S'\$^k$.

Caso 2: $Ax \neq S'\$^k$. Explicaremos então a última regra de derivação

$$S'\$^k \overset{+}{\equiv>} \gamma Cw \Rightarrow_s^* \gamma Dw \Rightarrow \gamma\delta w = \alpha Ax, \qquad v \equiv w,\ u \equiv x$$

Temos $|x| \geq |w|$ e $|\gamma\delta v| \leq |\alpha Au|$ e $\gamma\delta v \in CR{*}S(D{\to}\delta)$

Por outro lado, $\alpha Au = \gamma\delta vz \in CR{*}S$ com $z \in \Sigma^*$

Como temos também que $\alpha\beta u \in CR^*S(B\to\beta)$ e $A \Rightarrow_S^* B$ então $(X,\pi) \in$ ação$(\alpha\beta,u)$, $X \in \{RE,RNE\}$.

X será RNE ou RE dependendo de $z = \varepsilon$ ou $z \neq \varepsilon$. Já o valor de π será $\sigma_1\ B\to\beta$ ou $\sigma_2 B\to\beta$, $\sigma_1 \neq \sigma_2$, o que caracteriza o não-determinismo, pois

$$\{(X,\sigma_1 B\to\beta),(X,\sigma_2 B\to\beta)\} \subseteq \text{ação}(\alpha\beta,u)$$

<u>Lema II.17</u>: Se numa gramática G, tivermos

$$S'\$^k \overset{+}{\equiv>} \alpha Auw \overset{\sigma B\to\beta}{\equiv>} \alpha\beta uw \;,\quad \sigma \in s(A,B)$$

então (a) $\alpha Au\hat{w} \in CR^*S$ para algum $\hat{w} \in \Sigma^*$

(b) $(X,\sigma\ B\to\beta) \in$ ação$(\alpha\beta,u)$, onde X = RNE ou RE, dependendo de $\hat{w} = \varepsilon$ ou não em (a).

Demonstração:

Caso(a): explicitando o último passo de derivação de $S'\$^k \overset{+}{\equiv>} \alpha Auw$, temos:

$$S'\$^k \overset{*}{\equiv>} \Psi Evz \overset{\sigma_1 F\to\varphi}{\equiv>} \Psi\varphi vz = \alpha Auw \overset{\sigma_2 B\to\beta}{\equiv>} \alpha\beta uw,$$

$$\sigma_1 \in s(E,F),\ \sigma_2 \in s(A,B),$$

e como vz não contém não-terminais, $|vz| \leq |uw|$, e $\Psi\varphi v = \alpha Au\hat{w}$ onde $\hat{w}$ é um prefixo de w de comprimento $|w| - |z|$. Assim, $\alpha Au\hat{w} = \Psi\varphi v \in CR^*S(F\to\varphi)$, completando a prova de (a).

Caso(b): da hipótese do Lema temos

$\alpha\beta u \in CR^*S(B\to\beta)$

$A \Rightarrow_S^* B$

$\alpha Au\hat{w} \in CR^*S$

Se $\hat{w} = \varepsilon$, $(RNE, s(A,B)B \to \beta) \in$ ação$(\alpha\beta, u)$

Se $\hat{w} \neq \varepsilon$, $(RE, s(A,B)B \to \beta) \in$ ação$(\alpha\beta, u)$

Teorema II.18: seja G uma gramática livre de contexto, estendida e seja AG(G) o seu autômato analisador R*S geral. Nestas condições G satisfaz a condição R*S(k) se e somente se AG(G) é determinístico.

Demonstração: Parte 1: Se G é R*S(k) então AG(G) é determinístico.

Satisfeita a condição R*S(k) não teremos em G ambigüidade nas regras simples. Abusaremos da notação usando s(A,B) para representar seu único elemento, se este existir: $\pi = s(A,B)B \to \beta$, que abrevia $\pi = \sigma B \to \beta$, $s(A,B) = \{\sigma\}$.

Vamos supor que o autômato AG(G) não seja determinístico. Então há três casos possíveis de conflito nas transições do autômato.

Caso 1: o autômato tem conflito empilha/reduz.

$E \in$ ação(φ, u)

$(X, s(A,B)\ B \to \beta) \in$ ação(φ, u), $X = RE$ ou $X = RNE$

Temos $\varphi = \alpha\beta$, $\alpha\beta u \in CR{*}S(B \to \beta)$ e

$$S'\$^k \overset{*}{\Rightarrow} \alpha A w \overset{s(A,B)B \to \beta}{\Rightarrow} \alpha\beta w \text{ com } u = FIRST(w)$$

Por outro lado, da ação de empilhamento temos:

$\alpha\beta uz \in CR{*}S(D \to \delta)$ para $z \neq \varepsilon$ e $\alpha\beta uz = \gamma\delta v$

$$S'\$^k \overset{*}{\Rightarrow} \gamma C x \overset{s(C,D)\ D \to \delta}{\Rightarrow} \gamma\delta x = \alpha\beta y \text{ com } u = FIRST(y).$$

Pela condição R*S(k)

$\alpha Ay = \gamma Cx$ (e portanto $\alpha = \gamma$, $A = C$ e $x = y$) e

$s(A,B)\ B\to\beta = s(C,D)\ D\to\delta$

e portanto $z = \varepsilon$! Contradição.

Caso 2: temos duas ações distintas de redução (conflito reduz/reduz)

$(X, s(A,B)\ B\to\beta) \in \text{ação}(\varphi, u)$

$(Y, s(C,D)\ D\to\delta) \in \text{ação}(\varphi, u)$, onde X e Y podem ser RE ou RNE.

Temos

$S'\$^k \equiv>^* \alpha Aw \equiv> \alpha\beta w$

$S'\$^k \equiv>^* \gamma Cx \equiv> \gamma\delta x$

com $\varphi = \alpha\beta = \gamma\delta$ e $u = FIRST(w) = FIRST(x)$

Pela condição R*S(k),

$\alpha Ax = \gamma Cx$

$s(A,B)\ B\to\beta = s(C,D)\ D\to\delta$

Então o conflito pode ser expresso por

$(X, s(A,B)\ B\to\beta) \in \text{ação}(\varphi, u)$

$(Y, s(A,B)\ B\to\beta) \in \text{ação}(\varphi, u)$

e as ações são distintas se $X \neq Y$. Sejam X = RE e Y = RNE. Nestas condições

$$\alpha Auz \in CR^*S$$

$$\alpha Au \in CR^*S$$

ou seja,

$$\alpha Auz \in CR^*S(B_1 \to \beta_1)$$

$$\alpha Au \in CR^*S(B_2 \to \beta_2)$$

o que nos leva a um conflito empilha/reduz, excluído pela análise do caso (1) acima.

Caso 3: temos uma ação de redução e uma ação de parada (conflito reduz/para).

Temos H $\in$ ação (φ, u) e portanto $\varphi = \$S$, $u = \$^k$.

Temos $(X, s(A,B)B \to \beta) \in$ ação(φ, u), onde X é RE ou RNE, e φ é da forma $\alpha\beta$.

De $\alpha\beta = \varphi = \$S$, temos $\alpha = \$S$, $\beta = \varepsilon$. (Com efeito, o caso $\beta = S$ pode ser excluído porque $B \to \beta$ seria uma regra simples; o caso $\beta = \$S$ exigiria, pela construção da gramática estendida, $A = B = S'$, e também $\alpha Avz = S'\$^k z \in CR^*S$, o que é excluído pela definição de CR^*S).

Nesta situação

$$S'\$^k \Rightarrow^* \$SA\$^k \Rightarrow \$S\$^k$$

Como também

$$S'\$^k \Rightarrow \$S\$^k$$

o que contradiz o fato de que G é R^*S e portanto não-ambígua.

Note que a possibilidade de um conflito para/empilha pode ser excluída, uma vez que H $\in$ ação(φ,u) nos leva a $\varphi = \$S$, $u = \$^k$, e E $\in$ ação(φ,u) obrigaria $\varphi uz = \$S\$^k \in CR*S$ com $z \neq \varepsilon$, o que não é possível pela construção de G e pela definição de CR*S.

Parte 2: Se AG(G) é determinístico então G é R*S(k).

Observamos, de início, que G não pode ter ambigüidade nas produções simples, pois isto implicaria no não determinismo de AG(G), conforme o Lema II.16.

Vamos supor que G não tenha ambigüidade nas produções simples e que não seja R*S(k). Vamos provar que, nestas condições, AG(G) não é determinístico.

Se G não satisfaz a condição R*S, então, explicitando as cadeias lookahead, temos

$$(D1)\quad S'\$^k \equiv>^* \alpha Auw \overset{\pi_1}{\equiv>} \alpha\beta uw$$

$$(D2)\quad S'\$^k \equiv>^* \gamma Cvx \overset{\pi_2}{\equiv>} \gamma\delta vx = \alpha\beta uy$$

onde $\pi_1 = s(A,B)\ B\to\beta$ e $\pi_2 = s(C,D)\ D\to\delta$,

implica que $\alpha Auy \neq \gamma Cvx$ ou $\pi_1 \neq \pi_2$

1. Suponhamos $\pi_1 \neq \pi_2$, $S'\$^k \equiv>^* \alpha Auw$ em 0 passos.

Temos $S'\$^k = \alpha Auw$ com $\alpha = \varepsilon$, $S' = A = B$, $\$^k = u$, $w = \varepsilon$, $\beta = \$S$ e $\pi_1 = S'\to\$S$, o que reduz (D1) a

$$(D1)\quad S'\$^k \overset{S'\to\$S}{\equiv>} \$S\$^k$$

De (D2) $S'\$^k \equiv>^* \alpha\beta uy = \$S\$^k y$ e $y = \varepsilon$. Portanto

$$(D2)\quad S'\$^k \equiv>^* \gamma Cvz \overset{s(C,D)\ D\to\delta}{\equiv>} \$S\$^k$$

Note agora que $\delta = \varepsilon$, uma vez que $\delta = S$ faria $D \to \delta$ ser uma produção simples, e $\delta = \$S$ faria $D \to \delta = S' \to \$S$, que não pode ser usada senão no primeiro passo da derivação.

Como $\pi_1 \neq \pi_2$, $S'\$^k \equiv>^+ \gamma Cvx$ e, pelo Lema II.17

$(X,\pi_2) \in$ ação$(\$S,\$^k)$ para algum $X \in \{RE,RNE\}$ o que, juntamente com $H \in$ ação$(\$S,\$^k)$ caracteriza AG como não-determinístico.

2. $\pi_1 \neq \pi_2$, $S'\$^k \equiv>^* \gamma Cvx$ em 0 passos. A prova se dá de maneira semelhante ao caso anterior.

3. Suponha $\pi_1 \neq \pi_2$, $S'\$^k \equiv>^+ \alpha Auw$, $S'\$^k \equiv>^+ \gamma Cvx$.

Temos

(D1) $S'\$^k \equiv>^+ \alpha Auw \overset{\pi_1}{\equiv>} \alpha\beta uw$

(D2) $S'\$^k \equiv>^+ \gamma Cvx \overset{\pi_2}{\equiv>} \gamma\delta vx = \alpha\beta uy$

Pela Lema II.17

$(X,\pi_1) \in$ ação $(\alpha\beta,u)$ para algum X

$(Y,\pi_2) \in$ ação $(\gamma\delta,v)$ para algum Y

Para caracterizar o não-determinismo, precisamos avaliar 3 situações distintas.

3.1: Suponhamos $|\alpha\beta| = |\gamma\delta|$, ou seja, $\alpha\beta u = \gamma\delta v$. Neste caso, (X,π_1) e (Y,π_2) pertencem a ação $(\alpha\beta,u) =$ ação$(\gamma\delta,v)$ e AG(G) não é determinístico, já que $\pi_1 \neq \pi_2$.

3.2: Suponhamos $|\alpha\beta| < |\gamma\delta|$. De $\gamma\delta vx = \alpha\beta uy$, tiramos que $\gamma\delta v = \alpha\beta uz$, onde z é um prefixo adequado de y.

Temos, de $(Y,\pi_2) \in$ ação$(\gamma\delta,v)$, que $\gamma\delta v \in CR{*}S(D \to \delta)$, e portanto, $\alpha\beta uz = \gamma\delta v \in CR^*S$, donde $E \in$ ação$(\alpha\beta, u)$, caracterizando-se assim o não-determinismo de AG, uma vez que $(X, \pi_1) \in$ ação$(\alpha\beta, u)$.

3.3: Suponhamos que $|\alpha\beta| > |\gamma\delta|$. A prova se dá de maneira semelhante ao caso 3.2.

4. Suponhamos que $\pi_1 = \pi_2$ e $\alpha Auy \neq \gamma Cvx$

De $\pi_1 = \pi_2$ tiramos que $A = C$, $B = D$, $\beta = \delta$. Portanto

$$\gamma\delta vx = \gamma\beta vx = \alpha\beta uy$$

$$\gamma Cvx = \gamma Avx \neq \alpha Auy$$

Se $\gamma = \alpha$ então $\gamma\beta vx = \alpha\beta vx$ e $x = y$, $u = v$ e portanto $\gamma Avx = \alpha Auy$, o que contraria a hipótese.

Se $\gamma \neq \alpha$ temos 2 situações a considerar.

$|\gamma| < |\alpha|$ e $|\gamma\delta| < |\alpha\beta|$. Ver 3.3.

$|\gamma| > |\alpha|$. Ver caso 3.2.

<u>Definição II.19</u>: dizemos que um método de análise sintática possui a propriedade dos prefixos corretos se o primeiro elemento do terno que representa a configuração do autômato (que representa o que também se chama de conteúdo da pilha de análise) for (ou representar) um prefixo de uma forma sentencial da gramática para a qual o analisador foi construído.

<u>Teorema II.20</u>: seja AG um autômato analisador $R^*S(k)$ geral construído a partir de G, com $k \geq 1$. Então AG tem a propriedade dos prefixos corretos.

Demonstração: o teorema vale por construção. Seja (φ, ux, π) uma configuração qualquer de AG. a qual supomos seja tal que φ forme um prefixo correto de G. Então, qualquer mudança de configuração válida de AG levará a uma nova configuração cujo conteúdo de pilha será um prefixo correto de G.

Assim:

a) se empilhamento $\in$ ação(φ,u) é porque φux $\in$ CR*S para x $\neq$ ε e portanto φa é um prefixo correto de G para ay = ux.

b) se redução sem empilhamento $\in$ ação($\alpha\beta$,u) é porque αAu $\in$ CR*S e portanto αA é um prefixo correto de G.

c) se redução com empilhamento $\in$ ação($\alpha\beta$,u) é porque αAux $\in$ CR*S para x $\neq$ ε e portanto Ay é um prefixo correto de G para yv = ux.

Logo, toda a mudança de configuração a partir de uma configuração de prefixo correto, leva a outra de prefixo correto. Considerando que a configuração inicial $(\$,S\$^k,\varepsilon)$ já contém na pilha um prefixo correto, \$, concluimos que toda configuração (φ,y,π) tal que $(\$,x\$^k,\varepsilon) \vdash^* (\varphi,y,\pi)$ deixa um prefixo correto na pilha.

Observação: No caso k = 0, a cadeia lookahead é sempre ε. Quando ação(φ,εx,π) $\ni$ empilhamento, o símbolo a ser empilhado é, até aquele instante, desconhecido, ou seja, não foi lido. Ocorre assim o empilhamento do primeiro símbolo encontrado, podendo assim deixar na pilha um prefixo não correto. Não obstante, esse autômato acusa erro na transição seguinte, continuando a valer L(AG) = L(G).

<u>Teorema II.21</u>: Seja G satisfazendo a condição R*S. Seja AG um autômato analisador R*S geral construído a partir de G. Então

$$(\$,w\$^k,\varepsilon) \vdash^j (\$S,\$^k,\pi) \text{ implica } j \leq |SPNS(\pi)| + |w|.$$

Em outras palavras a análise de uma sentença pertencente a L(G) tem complexidade linear no tempo.

Demonstração: a prova decorre dos seguintes argumentos:

1) pelo Lema II.14

$$(\$, xy, \varepsilon) \vdash^{m} (\varphi, y, \pi) \text{ implica } \varphi y \overset{\pi}{\Rightarrow} \$xy$$

2) o número máximo de ações de redução (com ou sem empilhamento) na análise de w tal que

$$S \overset{\pi}{\Rightarrow} w \quad \text{é } |\pi|$$

Em HEILBRUNNER (11) temos ainda que, sendo G não ambígua,

$$S \overset{\pi}{\Rightarrow} w \quad \text{implica} \quad |\pi| \leq c_1 \; |w|$$

Portanto o número máximo de mudanças de configuração do autômato analisador é dada pela soma do número máximo de reduções somado com o número máximo de empilhamentos, ou seja

$|SPNS(\pi)|$ reduções

$|w|$ empilhamentos

ou seja

$$j \leq |SPNS(\pi)| + |w| \leq |\pi| + |w| \leq$$

$$\leq c_1|w| + |w| \leq c_2|w|$$

que confirma a linearidade (no tempo) da análise de w.

cqd

Teorema II.22: seja G uma gramática, seja AG o autômato analisador R*S geral construído a partir de G. Seja xy $\in$ Σ^* tal que

$$(\$,xy\$^k,\varepsilon) \vdash^* (\varphi,y\$^k,\pi) \vdash \text{erro}$$

Então existe z $\in \Sigma^*$ tal que

$$(\$,xz\$^k,\varepsilon) \vdash^* (\varphi,z\$^k,\pi) \vdash^* (\$S,\$^k,\pi'\pi).$$

Demonstração:

Visto que $(\$,xy\$^k,\varepsilon) \vdash^* (\varphi,y\$^k,\pi)$ podemos afirmar, pelo Lema II.14 que existe $\beta \in V^*$ tal que

$$S' \Rightarrow \$S \Rightarrow^* \varphi\beta$$

Seja z $\in \Sigma^*$ tal que $\beta \Rightarrow^* z$. Isto é válido, visto que G é reduzida.

Então

$$S' \Rightarrow \$S \Rightarrow^* \varphi\beta \Rightarrow^* \$xz$$

Portanto, pelo tema II.13

$$(\$,xz\$^k,\varepsilon) \vdash^* (\varphi,z\$^k,\pi) \vdash^* (\$S,\$^k,\pi'\pi)$$

cqd

Teorema II.23: seja AG(G) onde G satisfaz a condição R*S(k). Então para qualquer x $\in \Sigma^*$.

ou $(\$,x\$^k,\varepsilon) \vdash (\$S,\$^k,\pi)$

ou $(\$,x\$^k,\varepsilon) \vdash^* (\varphi,uw,\pi)$ com ação $(\varphi,u) = \emptyset$

e $\varphi uw \neq \$S\$^k$

(Em outras palavras, para AG(G) nas condições acima, para qualquer $x \in \Sigma^*$, ou o autômato pára e aceita x, ou o autômato pára e rejeita x).

Caso 1: $x \in L(G)$

Então, pelo Lema II.13, a configuração final é atingida.

Caso 2: $x \notin L(G)$

Seja $x = yaz$ tal que

$\nexists\ z' \in \Sigma^*$ com $yaz' \in L(G)$

$\exists\ z'' \in \Sigma^*$ com $yz'' \in L(G)$

(y é o maior prefixo de x, comum a L(G)).

Nestas condições:

$(\$, yaz\$^k \varepsilon) \vdash^* (\varphi, uz\$^k, \pi)$

$\varphi u = \psi a$

$\psi \overset{\pi}{\Rightarrow} \$y$

$ação(\varphi, u) = \emptyset$ porque $\psi aw \notin CR^*S,\ \forall\ w \in \Sigma^*$.

Vamos acompanhar as mudanças de configuração de AG(G) durante a análise das sentenças $yaz \notin L(G)$ e $ybz'' \in L(G)$.

y	b	z"	$\in L(G)$
y	a	z	$\notin L(G)$

As mudanças de configuração dessas duas sentenças podem ser descritas por

$$(\$, yz''\$^k, \varepsilon) \vdash^n C_1 \vdash (\varphi, vz''\$^k, \pi)$$

$$(\$, yaz\$^k, \varepsilon) \vdash^n C_2 \vdash (\varphi, uz\$^k, \pi) = \text{erro}$$

de tal forma que $C_1 = C_2$, indicando dessa forma os passos do autômato enquanto as cadeias lookahead ainda se encontram totalmente dentro do trecho comum y. A passagem $C_2 \vdash (\varphi, uz\$^k, \pi)$ indica a mudança de configuração em que o símbolo **a** passa a fazer parte da cadeia lookahead Seja $\varphi u = \psi a$, $\psi \Rightarrow^* \$y$. $\psi aw \notin CR^*S$, $\forall\ w \in \Sigma^*$ porque $\varphi aw \Rightarrow^* \$yaw \notin L(G)$ por hipótese. Logo, em$(\varphi, uz\$^k, \pi)$ o autômato tem ação $(\varphi, u) = \emptyset$, portanto pára e não aceita x.

No caso de yz", a análise prossegue até a configuração final $(\$S, \$^k, \pi)$ e pára, aceitando a cadeia lida, conforme nos garante o Teorema II.11.

Note que no caso k = 0, a presença do símbolo "errado" a em $\varphi u = \varphi$ implica efetivamente no seu empilhamento, e é após este que o erro é anunciado, já que não há continuações possíveis. Esta é, entretanto, a primeira ocasião em que o símbolo pode ser considerado.

CAPÍTULO III

AUTÔMATOS R*S CANÔNICOS

III.1 - INTRODUÇÃO

Vimos no capítulo II como se constrói o autômato analisador R*S geral a partir de uma gramática dada. Vimos também as propriedades principais desse autômato, não necessariamente determinístico. E definimos a relação **ação** que determinava as mudanças de configuração durante o processo de análise. Essa relação era definida de tal forma que seus argumentos, φ e u, representavam todo o conteúdo da pilha e a cadeia lookahead , respectivamente.

No presente capítulo trataremos de construir um autômato analisador de interesse prático, e que portanto deverá ser necessariamente determinístico. Para tanto faremos uso do conceito de estados, entidades que servirão para indicar o "status" do processo de análise, e que substituirão os símbolos que eram colocados na pilha.

Ao autômato, que deverá reproduzir todas as ações do modelo proposto no capítulo II denominaremos **canônico**. O estudo das propriedades principais do autômato analisador **R*S** canônico é parte também do presente capítulo.

Neste capítulo e nos seguintes, consideraremos $k > 0$, abandonando o caso $k = 0$, em virtude de sua menor aplicação prática. O caso prático de maior importância é certamente o caso $k = 1$, que será o único considerado a partir do capítulo IV.

<u>Definição III.1</u>: chamaremos de estado qualquer conjunto não vazio de ítens. Dependendo da classe de ítens considerada, falaremos em estado LR(k), R*S(k), etc.

Definição III.2: seja $q \subseteq$ IR*S. Definimos a função de transição δ_c e notamos

$$\delta_c(q,X) = \text{fecho}\ (\{[A\to\alpha X.\beta,u]\,|\,[A\to\alpha.X\beta,u] \in q \wedge$$

$$[A\to\alpha X.\beta,u] \text{ é um ítem R*S(k)}\})$$

Quando $\delta_c(q,X) = \emptyset$ diremos que $\delta_c(q,X)\uparrow$.

Definição III.3: coleção canônica de estados R*S. Q_c é uma coleção de estados R*S construídos a partir de uma gramática estendida G se Q_c é o menor conjunto **C** tal que

1. $q_{oc} \in \mathbf{C}$ onde $q_{oc} = \text{fecho}(\{[S'\to\$.S,\$^k]\})$

2. Se $p \in \mathbf{C}$ e se $\delta_c(p,X)\downarrow$ então $\delta_c(p,X) \in \mathbf{C}$ para $X \in N \cup \Sigma$.

Definição III.4: Definimos as funções $\gamma_c:V^*\to Q_c$ e $\lambda_c:V^*\ Q_c^*$ da forma seguinte:

1. $\nu_c(\$) = q_{oc}$

   $\nu_c(\varphi X) = \delta_c(\nu_c(\varphi),X)$

2. $\lambda_c(\$) = q_{oc}$

   $\lambda_c(\varphi X) = \lambda_c(\varphi)\delta_c(\nu_c(\varphi),X)$

sendo os valores das duas funções considerados indefinidos nos demais casos.

Notas:

- se estendermos, como habitualmente, δ_c aceitando cadeias em seu segundo argumento, temos

$$\nu_c(\$\varphi) = \delta_c(q_o,\varphi)$$

- $\nu_c(\varphi)$ é sempre o último estado de $\lambda_c(\varphi)$

- se $\nu_c(\varphi_1\varphi_2)\downarrow$ então $\nu_c(\varphi_1)\downarrow$ (para $\varphi_1 \neq \varepsilon$)

- se $\lambda_c(\varphi_1\varphi_2)\downarrow$ então $\lambda_c(\varphi_1)\downarrow$ (para $\varphi_1 \neq \varepsilon$)

- como cada estado da coleção canônica é alcançável através de um único símbolo, a função (parcial) λ_c é 1:1.

<u>Teorema III.5</u>: Um ítem R*S $[A\to\beta.,u] \in \nu_c(\varphi)$ sse $\varphi u \in$ CR*S$(A\to\beta)$.

Demonstração: ver os Lemas III.6 e III.7 a seguir.

<u>Lema III.6</u>: Se $[A\to\beta.\beta,u] \in \nu_c(\varphi)$ e $A\to\alpha.\beta$ é uma regra não simples então $\varphi\beta u \in$ CR*S$(A\to\alpha\beta)$.

Demonstração: por indução no comprimento de φ

Base da indução

1) para $\varphi = \$$, $\nu_c(\varphi) = q_{oc}$

$[S'\to\$.S,\$^k] \in q_{oc}$ $\therefore$ $\varphi = \$$, $\beta = S$, $u = \$^k$

$S'\$^k \Rightarrow^* S'\$^k \Rightarrow \$S\$^k$ $\therefore$ pela definição de CR*S

$\$S\$^k \in$ CR*S$(S'\to\$S)$

2) Vamos supor que um ítem da forma $[A\to.\beta,u] \in q_{oc}$. Note que os ítens iniciais de q_{oc} (só há um) foram considerados acima.

Então os seguintes ítens estão em q_{oc}

$[S'\to\$.S,\$^k]$

$[A_1 \to .A_2\beta_1, u_1]$ $\quad u_1 = \$^k, \quad A_1 = S$

$[A_2 \to .A_3\beta_2, u_2]$ $\quad u_2 \in FIRST(\beta_1 u_1)$

.
.
.

$[A_{n-1} \to .A_n\beta_{n-1}, u_{n-1}]$ $\quad u_{n-1} \in FIRST(\beta_{n-2}u_{n-2})$

$[A_n \to .\beta_n, u_n] = [A \to .\beta, u]$ $\quad u = u_n \in FIRST(\beta_{n-1}u_{n-1})$

Podemos escrever também que, se todos estes ítens estão em q_{oc}, que

$$S \Rightarrow^* A_n\beta_{n-1} \cdots \beta_1 = A\beta_{n-1} \cdots \beta_1 \Rightarrow \beta_n\beta_{n-1} \cdots \beta_1 \Rightarrow$$
$$= \beta\beta_{n-1} \cdots \beta_1$$

Então (as derivações não precisam ser mais à direita)

$$\$S\$^k \Rightarrow^* \$S\beta_{n-1} \cdots \beta_1\$^k \Rightarrow \$\beta\beta_{n-1} \cdots \beta_1\$^k$$

$$\$S\$^k \Rightarrow^* \$Auw \equiv> \$\beta uw$$

onde supusemos que $\beta_{n-1} \cdots \beta_1\$^k \Rightarrow^* uw$

Seja $j = \min\{i | \beta_i \cdots \beta_{n-1} = \varepsilon\}$ se $i \leq n-1$
senão $j = n$

Para esse valor j vale $A_j \Rightarrow_s^* A_n$

E podemos escrever

$$\$S\$^k \equiv>^* \$A_j uw \equiv> \$\beta uw$$

o que nos garante que $\$\beta u \in CR^*S\ (A \to \beta)$ completando a prova da base de indução.

Indução: seja $\varphi = \psi X$, $q = \nu_c(\varphi)$ $p = \nu_c(\psi)$

Temos

$$q = \nu_c(\varphi) = \nu_c(\psi X) = \delta_c(\nu_c(\psi), X) = \delta_c(p, X)$$

Supondo que o Lema vale para p, vamos procurar provar que ele vale para q.

Por hipótese, o ítem $[A \to \alpha . \beta, u] \in q = \nu_c(\varphi)$.

Vamos considerar os seguintes casos:

Caso 1: $\alpha = \gamma Y$, $Y \in V = N \cup \Sigma$

Temos $[A \to \gamma Y . \beta, u] \in q = \delta_c(p, X)$. Logo $Y = X$ e $[A \to \gamma . X\beta, u] \in p$.

Pela hipótese de indução, já que $A \to \alpha\beta$ não é simples,

$$\psi X \beta u \in CR^*S(A \to \gamma X \beta)$$

ou seja, que

$$\varphi \beta u \in CR^*S(A \to \alpha\beta).$$

Caso 2: $\alpha = \varepsilon$

Então o ítem $[A \to . \beta, u]$ apareceu na operação fecho(q). Então encontraremos em q ítens da forma

$$[C \to \gamma Y . B_1 \delta, v]$$

$$[B_1 \to . B_2 \beta_1, u_1]$$

$$\vdots$$

$$[B_{n-1}\rightarrow .B_n\beta_{n-1},u_{n-1}]$$

$$[B_n\rightarrow .\beta_n,u_n]$$

de forma que $X = Y$, $[B_n\rightarrow .\beta_n,u_n] = [A\rightarrow .\beta,u]$ e

$$u_1 \in \mathrm{FIRST}(\delta v)$$

$$u_2 \in \mathrm{FIRST}(\beta_1 u_1)$$

$$\vdots$$

$$u_n \in \mathrm{FIRST}(\beta_{n-1}u_{n-1})$$

Como em p aparece o ítem $[C\rightarrow\gamma .XB_1\delta,v]$, pela hipótese de indução

$$\varphi B_1\delta v = \psi XB_1\delta v \in CR^*S(C\rightarrow\gamma XB_1\delta)$$

ou seja

$$S'\$^k \Rightarrow^* \varphi B_1\delta w \Rightarrow^* \varphi B_n\beta_{n-1} \ldots \beta_1\delta w \Rightarrow \varphi\beta_n\ {}_{n-1} \ldots\beta_1\delta w$$

(com derivações nem sempre direitas), onde $v = \mathrm{FIRST}(w)$.

Temos $u = u_n \in \mathrm{FIRST}(\beta_{n-1} \ldots \beta_1\delta w)$

Escolhendo x tal que $\mathrm{FIRST}(x) = u$ sabendo que

$$\beta_{n-1} \ldots \beta_1\delta w \Rightarrow^* x$$

$$S'\$^k \Rightarrow^* \varphi B_n x \Rightarrow \varphi\beta_n x$$

ou

$$S'\$^k \Rightarrow^* \varphi Ax \Rightarrow \varphi\beta x$$

e $\varphi\beta\mathrm{FIRST}(x) = \varphi\beta u \in CR^*S(A\rightarrow\beta)$ (visto que $A\rightarrow\beta$ não é produção simples.

Lema III.7: se $\varphi\beta u \in CR^*S(A\to\alpha\beta)$ e $A\to\alpha\beta$ não é uma produção simples, então $[A\to\alpha.\beta,u] \in \nu_c(\varphi)$.

Demonstração: será feita por indução no comprimento de φ.

Base de indução: $|\varphi| = 1$, $\varphi = \$$, $\nu_c(\varphi) = q_{oc}$

$\$\beta u \in CR^*S(A\to\alpha\beta)$ por hipótese, e portanto, ou $\alpha = \varepsilon$ ou $\alpha = \$$.

Caso 1: $\alpha = \varepsilon$

$$S'\$^k \Rightarrow \$S\$^k \Rightarrow^* \$Auw \Rightarrow \$\beta uw$$

A transição $\$S\$^k \Rightarrow^* \$Auw$ é obtida por uma sequência de produções do tipo:

$$A_1 \to A_2\beta_1 \qquad (\text{com } S = A_1)$$

$$A_2 \to A_3\beta_2$$

$$\vdots$$

$$A_{n-1} \to A_n\beta_{n-1}$$

$$A_n \to \beta_n \quad (\text{com } A_n\to\beta_n = A\to\beta)$$

Desta forma

$$\beta_{n-1} \ldots \beta_1\$^k \Rightarrow^* uw$$

onde $\beta_i \Rightarrow^* w_i$ e $uw = w_{n-1} \ldots w_1\$^k$

Seja

$$u_i = FIRST(w_{i-1} \ldots w_1\$^k) \qquad (\text{com } u_n = u)$$

Então

$$[A_1 \to . A_2 \beta_1, u_1] \in q_{oc}$$
$$\vdots$$
$$[A_n \to . \beta_n, u_n] \in q_{oc}$$

ou seja, $[A \to . \beta, u] \in q_{oc}$ como desejado.

Caso 2: se $\alpha = \$$ e $\beta = S$, então o ítem procurado que pertence a q_{oc} é $[S' \to \$.S, \$^k]$.

Passo de indução:

Seja $\varphi = \psi X$, $q = \nu_c(\varphi)$, $p = \nu_c(\psi)$

$$q = \nu_c(\varphi) = \nu_c(\psi X) = \delta_c(\nu_c(\psi), X) = \delta_c(p, X)$$

Por hipótese $\varphi\beta u \in CR{*}S(A \to \alpha\beta)$

$$S'\$^k \Rightarrow^* \gamma A w \Rightarrow \gamma\alpha\beta w \qquad \text{com } u = FIRST(w), \ \varphi = \gamma\alpha$$

Caso 1: $\alpha = \rho Y$, $\varphi = \gamma\rho Y = \psi X$ com $\gamma\rho = \psi$ e $X = Y$

Pela hipótese de indução, como $\psi X \beta u \in CR{*}S(A \to \alpha\beta)$ temos $[A \to \rho . X\beta, u] \in \nu_c(\psi) = p$ e portanto $[A \to \rho X . \beta, u] \in \delta_c(p, X) = q$

ou $[A \to \alpha . \beta, u] \in q$.

Caso 2:

Vamos tomar $\varphi = \psi X$, tendo como consequência que $q = \nu_c(\varphi)$, $p = \nu_c(\psi)$ $q = \delta_c(p, X)$. Seja $\alpha = \varepsilon$.

Se $\varphi\beta u \in CR{*}S(A \to \beta)$, então basta provar que

$$[A \to . \beta, u] \in \nu_c(\varphi)$$

Ora, temos que

$$S'\$^k \Rightarrow^* \varphi A w \Rightarrow \varphi \beta w$$

Vamos supor que X tenha sido introduzido na forma sentencial pela regra $C \rightarrow \gamma_1 X B \gamma_2$.

$$S'\$^k \Rightarrow^* \rho C x \Rightarrow \rho\gamma_1 X B \gamma_2 x \Rightarrow^* \psi X A w \Rightarrow \psi X \beta w$$

onde

$$\rho\gamma_1 = \psi$$

$$B\gamma_2 x \Rightarrow^* A w$$

Suponha ainda $B\gamma_2 x \Rightarrow^* Aw$ pelas seguintes produções

$$A_1 \rightarrow A_2\beta_1 \qquad (\text{com } B = A_1)$$
$$\vdots$$
$$A_{n-1} \rightarrow A_n\beta_{n-1}$$

$$A_n \rightarrow \beta_n \qquad (\text{com } A_n{\rightarrow}\beta_n = A{\rightarrow}\beta)$$

Temos que $\beta_{n-1} \ldots \beta_1\gamma_2 x \Rightarrow^* w$

de tal forma que

$$\beta_i \Rightarrow^* w_i$$

$$\gamma_2 \Rightarrow^* y_2$$

$$w = w_{n-1} \ldots w_1 y_2 x$$

Ora, pela hipótese de indução

$$[C{\rightarrow}\gamma_1 . X B\gamma_2, v] \in p = \nu_c(\psi)$$

$$[C{\rightarrow}\gamma_1 X . B\gamma_2, v] \in q = \nu_c(\psi X), \text{ com } v = \text{FIRST}(x).$$

Vamos definir agora

$$u_i = FIRST(w_{i-1} \ldots w_1 y_2 x)$$

então

$$[A_1 \rightarrow .A_2 \beta_1 . u_1] \in q$$

$$[A_{n-1} \rightarrow .A_n \beta_{n-1}, u_{n-1}] \in q$$

$$[A_n \rightarrow .\beta_n . u] \in q$$

ou seja, $[A \rightarrow .\beta, u] \in q$ cqd

III.2 ÷ CONSTRUÇÃO DE AUTÔMATOS R*S CANÔNICOS

Estamos agora em condições de construir um autômato com pilha, com estados formados a partir de conjuntos de ítens, que reproduza fielmente as mudanças de configuração do autômato R*S geral. Nossa tarefa fica facilitada pelo Teorema III.5 que diz que ítens compõem os estados do autômato analisador canônico.

Definição III.8: O autômato R*S canônico $AC_k(G)$, construído a partir de uma gramática estendida $G = (N,\Sigma,P,S')$, é dado por

$$AC_k(G) = (Q_c,P,\Sigma,e_c,d_c,r_c,q_{oc},q_{fc})$$

onde

- Q_c é a coleção de estados R*S de G

- q_{oc} = fecho$(\{[S'\to\$.S,\$^k]\})$ é o estado inicial

- $q_{fc} = \delta_c(q_{oc},S)$ é o estado final, isto é, o que contém o ítem final $[S'\to\$S.,\$^k]$.

- a função $e_c:Q_c \times \Sigma^k \to Q_c$ é definida por $e_c(\nu_c(\varphi),u) = \nu_c(\varphi a)$ onde $u = ay$ e $E \in$ ação (φ,u) - ver Teorema III.9.

- a relação $d_c:Q_c \times \Sigma^k \to Z$ (onde Z representa o conjunto dos inteiros não negativos) definida como a menor relação tal que:

  se $(X,\sigma\ B\to\beta) \in$ ação$(\alpha\beta,u)$, $X \in \{RE,RNE\}$,
  então $|\beta| \in d_c(\nu_c(\alpha\beta),u)$

- a relação $r_c:Q_c \times \Sigma^k \times Q_c \to (Q_c \cup (Q_c \times Q_c)) \times P^*$ definida como a menor relação tal que

1. Se $(RE,\sigma\ B\to\beta) \in$ ação$(\alpha\beta,u)$, $\sigma \in s(A,B)$, $u = ay$ então $((\nu_c(\alpha A),\nu_c(\alpha Aa)),\alpha\ B\to\beta) \in$
$\in r_c(\nu_c(\alpha),u,\nu_c(\alpha\beta))$

2. se $(RNE,\sigma\ B\to\beta) \in$ ação$(\alpha\beta,u)$, $\sigma \in s(A,B)$ então$(\nu_c(\alpha A),\ \sigma\ B\to\beta) \in r_c(\nu_c(\alpha),u,\nu_c(\alpha\beta))$

Nossa intenção é de que as relações r_c e d_c sejam funções nos casos práticos, quando G satisfaz a condição R*S(k) e AG é determinístico. Isto será estabelecido em resultados a seguir onde também será mostrado que a função e_c está bem definida. No que se segue, $AC_k(G)$ será abreviado para AC, subentendendo-se k e G.

<u>Teorema III.9</u>: Nas condições da Definição III.8, e_c é uma função bem definida.

Demonstração: cumpre provar que a escolha de φ, na definição III.8,

(a) não altera o valor atribuído a $e_c(\nu_c(\varphi),u)$

(b) não altera a presença ou não da ação de empilhamento em ação (φ,u)

Sejam φ_1 e φ_2 tais que $\nu_c(\varphi_1) = \nu_c(\varphi_2) = q$, $u = ay$.

Para provarmos (a), observamos que

$$e_c(\nu_c(\varphi_1),u) = \nu_c(\varphi_1 a) = \delta_c(\nu_c(\varphi_1),a) = \delta_c(q,a) =$$

$$= \delta_c(\nu_c(\varphi_2),a) = \nu_c(\varphi_2 a) = e_c(\nu(\varphi_2),u)$$

Para provarmos (b), suporemos que $E \in$ ação(φ_1,u). Então $\varphi_1 uy \in CR{*}S(B\to\beta)$ com $y \neq \varepsilon$ para alguma regra $B\to\beta$

$$\varphi_1 uy = \varphi_1 zv = \alpha_1 \beta v$$

Portanto $[B \to \beta.,v] \in \nu_c(\alpha_1\beta) = \nu_c(\varphi_1 z) = \nu_c(\varphi_2 z)$ e $\varphi_2 uy = \varphi_2 zv = \alpha_2\beta v$.

Temos então que $\varphi_2 uy \in CR^*S(B\to\beta)$ com $y \neq \varepsilon$ e portanto $E \in$ ação(φ_2,u).

<u>Teorema III.10</u>: Sejam AG(G,ação) e AC$(Q_c,\Sigma,P,e_c,d_c,r_c,q_{oc},q_{fc})$ construídos a partir de G. Se AG é determinístico então d_c e r_c definem funções, isto é, para q, u e p quaisquer

(a) $|d_c(p,u)| \leq 1$

(b) $|r_c(q,u,p)| \leq 1$

Demonstração: (a) Suponha i_1 e $i_2 \in d_c(p,u)$, com $i_1 > i_2$. Temos então:

$$\nu_c(\alpha_1\beta_1) = \nu_c(\alpha_2\beta_2) = p$$

$$|\beta_1| = i_1 > |\beta_2| = i_2$$

$$(X_1,\sigma_1 \quad B_1\to\beta_1) \in \text{ação}(\alpha_1\beta_1,u)$$

$$(X_2,\sigma_2 \quad B_2\to\beta_2) \in \text{ação}(\alpha_2\beta_2,u)$$

$$\sigma_1 \in s(A_1,B_1)$$

$$\sigma_2 \in s(A_2,B_2), \qquad X_1,X_2 \in \{RE,RNE\}$$

Pela definição de ação, $\alpha_1\beta_1 u \in CR^*S(B_1\to\beta_1)$

$\alpha_2\beta_2 u \in CR^*S(B_2\to\beta_2)$

Pelo Teorema III.5, $[B_2\to\beta_2.,u] \in \nu_c(\alpha_2\beta_2) = p = \nu_c(\alpha_1\beta_1)$. Assim $\alpha_1\beta_1 = \alpha_3\beta_2$, para algum α_3 e $\alpha_1\beta_1 = \alpha_3\beta_2$ e $\alpha_3\beta_2$ CR*S $(B_2\to\beta_2)$. Consequentemente

$$(X_3,\sigma_3 \; B_2\to\beta_2) \in \text{ação}(\alpha_3\beta_2,u) = \text{ação}(\alpha_1\beta_1,u)$$

e AG não é determinístico, o que prova (a) por contradição.

(b) Temos vários casos a considerar, dependendo dos valores das primeiras componentes de $r_c(q,u,p)$.

(b_1) $((r_1,s_1),\pi_1) \in r_c(q,u,p)$
$((r_2,s_2),\pi_2) \in r_c(q,u,p)$
$\pi_1 \neq \pi_2$

Temos, escrevendo u = ay, que:

$q = \nu_c(\alpha_1) = \nu_c(\alpha_2)$

$p = \nu_c(\alpha_1\beta_1) = \nu_c(\alpha_2\beta_2)$

$r_1 = \nu_c(\alpha_1 A_1)$ $\quad\quad$ $r_2 = \nu_c(\alpha_2 A_2)$

$s_1 = \nu_c(\alpha_1 A_1 a)$ $\quad\quad$ $s_2 = \nu_c(\alpha_2 A_2 a)$

$\pi_1 = \sigma_1 \; B_1 \to \beta_1$ $\quad\quad$ $\pi_2 = \pi_2 \; B_2 \to \beta_2$

$(RE,\pi_1) \in$ ação $(\alpha_1\beta_1,u)$ $(RE,\pi_2) \in$ ação$(\alpha_2\beta_2,u)$

Se tivermos $|\beta_1| \neq |\beta_2|$, então deverá ocorrer

$|\beta_1| \in d_c(p,u)$

$|\beta_2| \in d_c(p,u)$

contrariando a prova feita em (a).

Por outro lado, $|\beta_1| = |\beta_2|$ nos leva a $\beta_1 = \beta_2$ e

$(RE,\sigma_2 \; B_2 \to \beta_2) \in$ ação$(\alpha_2\beta_2,u)$

$\alpha_2\beta_2 u \in CR{*}S\ (B_2 \to \beta_2)$

$[B_2 \to \beta_2 .,u] \in \nu_c(\alpha_2\beta_2) = p = \nu_c(\alpha_1\beta_1)$

Assim, $\alpha_1\beta_1$ tem um sufixo β_2 com comprimento $|\beta_1|$ e $\beta_1 = \beta_2$. Escreveremos $\beta = \beta_1 = \beta_2$.

De $(RE,\sigma_2\ B_2 \to \beta_2) \in$ ação$(\alpha_2\beta,u)$ temos também $\alpha_2 A_2 uz \in CR^*S$ com $z \neq \varepsilon$ e portanto

$$S'\$^k \Rightarrow^* \alpha_2 A_2 uzy_2 \text{ para algum } y_2.$$

O caso $S'\$^k = \alpha_2 A_2 uzy_2$ pode ser abandonado porque $z \neq \varepsilon$. A derivação acima tem um ou mais passos e podemos explicitar o passo que introduz A_2:

$$S'\$^k \Rightarrow^* \rho_2 Cvx_2 \overset{\pi}{\Rightarrow} \alpha_2\xi A_2\gamma vx_2 \Rightarrow^* \rho_2\xi A_2 wvx_2 = $$
$$= \alpha_2 A_2 uzy_2$$

com $\pi = \sigma\ D \to \xi A_2\gamma$, $\sigma \in s(C,D)$, $\alpha_2 = \rho_2\xi$, $\gamma \Rightarrow^* w$.

$u \in FIRST(\gamma v)$, $wvx_2 = uzy_2$.

Mesmo sem explicitar qual é a regra cujo contexto de redução $\alpha_2 A_2 uz$ pertence, sabemos que ou se trata da regra $D \to \xi A_2\gamma$, ou da última regra usada na derivação (direita) de w a partir de γ. Concluimos então que $|uz| \leq |wv|$, ou seja z, uz é um prefixo de wv.

Temos

$$\rho_2\xi A_2\gamma v \in CR^*S(D \to \xi A_2\gamma)$$

e
$$[D \to \xi . A_2\gamma, v] \in \nu_c(\rho_2\xi) = \nu_c(\alpha_2) = \nu_c(\alpha_1) = q$$

Logo $\alpha_1 = \rho_1\xi$ e
$\rho_1\xi A_2\gamma v \in CR^*S\ (D \to \xi A_2\gamma)$.

Portanto

$$S'\$^k \Rightarrow^* \rho_1\xi A_2\gamma vx_1 \Rightarrow^* \rho_1\xi A_2 wvzx_1 = \alpha_1 A_2 uzy_1$$

para algum y_1, Note que uz é um prefixo de wv e portanto um prefixo de wvx_1.

Consequentemente, $\alpha_1 A_2 uz \in CR^*S$.

Além disso,

$[B_2 \to \beta.,u] \in \nu_c(\alpha_1\beta)$,

$\alpha_1\beta u \in CR^*S(B_2 \to \beta)$,

$A_2 \overset{\sigma_2}{\Rightarrow}_s B_2$

$\pi_2 = \sigma_2 \; B_2 \to \beta$

e portanto $(RE,\pi_2) \in$ ação$(\alpha_1\beta,u)$, conflitando com $(RE,\pi_1) \in$ ação$(\alpha_1\beta,u)$. AG não pode ser determinístico e a prova está concluída.

Note adicionalmente que se tivermos $\pi_1 = \pi_2$ teremos também $r_1 = r_2$ e $s_1 = s_2$, desaparecendo o conflito.

(b2) $(r_1,\pi_1) \in r_c(q,u,p)$

$((r_2,s_2),\pi_2) \in r_c(q,u,p)$

A demonstração é semelhante à do caso (b1), não havendo aqui a necessidade de assumir $\pi_1 \neq \pi_2$, uma vez que $\pi_1 = \pi_2 = \pi$, nos leva a

$(RNE,\pi) \in$ ação$(\alpha_1\beta,u)$

$(RE,\pi) \in$ ação$(\alpha_1\beta,u)$

indicando não-determinismo em AG.

(b3) $(r_1,\pi_1) \in r_c(q,u,p)$
$(r_2,\pi_2) \in r_c(q,u,p)$
$\pi_1 \neq \pi_2$

Sendo $u = ay$, temos

$$q = \nu_c(\alpha_1) = \nu_c(\alpha_2)$$

$$p = \nu_c(\alpha_2\beta_1) - \nu_c(\alpha_2\beta_2)$$

$$r_1 = \nu_c(\alpha_1 A_1) \qquad r_2 = \nu_c(\alpha_2 A_2)$$

$$\pi_1 = \sigma_1 \;\; B_1 \to \beta_1 \qquad \pi_2 = \sigma_2 \;\; B_2 \to \beta_2$$

$$\sigma_1 \in s(A_1,B_1) \qquad \sigma_2 \in s(A_2,B_2)$$

$$(RNE,\pi_1) \in \text{ação}(\alpha_1\beta_1,u) \;\; (RNE,\pi_2) \in \text{ação}(\alpha_2\beta_2,u)$$

Se $|\beta_1| \neq |\beta_2|$ então d_c não é função, contrariando a prova (a).

Se $|\beta_1| = |\beta_2|$ então $\beta_1 = \beta_2 = \beta$.

De $(RNE,\sigma_2 \;\; B_2 \to \beta) \in \text{ação}(\alpha_2\beta,u)$ temos também

$$\alpha_2 A_2 u \in CR{*}S$$

e portanto

$$S'\$^k \Rightarrow^* \alpha_2 A_2 u y_2$$

(b) (3.1) $S'\$^k \Rightarrow^0 \alpha_2 A_2 u y_2$

O caso $S'\$^k = \alpha_2 A_2 u y_2$ nos permite concluir que

$$\alpha_2 = \varepsilon$$

$$S' = A_2$$

$u = \$^k$

$y_2 = \varepsilon$

$B_2 = S'$

$\beta = \$S$

$\alpha_2\beta = \$S$

e portanto ação$(\alpha_2\beta, u)$ = ação$(\$S, \$^k) \ni H$ e portanto AG não é determinístico.

(b) (3.2) $S'\$^k \Rightarrow^+ \alpha_2 A_2 u y_2$

Neste caso, explicitando o passo que introduz A_2 temos:

$$S'\$^k \Rightarrow^* \rho_2 C v x_2 \overset{\pi}{\Rightarrow} \rho_2 \xi A_2 \gamma v x_2 \Rightarrow^* \rho_2 \xi A_2 w v x_2 = \alpha_2 A_2 u y_2$$

com $\pi = \sigma\, D \to \xi A_2 \gamma$, $\sigma \in s(C,D)$, $\alpha_2 = \rho_2 \xi$, $\gamma \Rightarrow^* w$, $u \in \mathrm{FIRST}(\gamma v)$, $wvx_2 = uy_2$.

$\alpha_2 A_2 u \in CR^*S$ e sabemos que

$\alpha_2 A_2 u \in CR^*S(D \to \xi A_2 \gamma)$ ou $\alpha_2 A_2 u \in CR^*S(\hat{p})$ onde $\hat{p}$ é a última regra usada na derivação direita de w a partir de γ. Temos $uy_2 = wvx_2$ e que $|u| \leq |wv|$, necessariamente.

Temos

$\rho_2 \xi A_2 \gamma v \in CR^*S(D \to \xi A_2 \gamma)$

$[D \to \xi . A_2 \gamma, v] \in \nu_c(\rho_2 \xi) = \nu_c(\alpha_2) = \nu_c(\alpha_1) = q$

Logo $\alpha_1 = \rho_1 \xi$ e

$\rho_1 \xi A_2 \gamma v \in CR^*S\ (D \to \xi A_2 \gamma)$

Portanto

$S'\$^k \equiv>^* \rho_1\xi A_2\gamma v x_1 \equiv>^* \rho_1\xi A_2 w v x_1 = \alpha_1 A_2 u y_1$ para algum y_1, sendo u prefixo de wv. Consequentemente $\alpha_1 A_2 u \in CR{*}S$.

Assim

$[B_2 \to \beta.,u] \in \nu_c(\alpha_1\beta)$

$\alpha_1\beta u \in CR{*}S(B_2 \to \beta)$

$\sigma_2 \in s(A_2, B_2)$

$\pi_2 = \sigma_2 \; B_2 \to \beta$

e portanto $(RNE,\pi_2) \in$ ação$(\alpha_1\beta,u)$, conflitando com $(RNE,\pi_1) \in$ ação$(\alpha_1\beta,u)$, o que torna AG não determinístico.

Também neste caso, $\pi_1 = \pi_2$ implica em $r_1 = r_2$ e no desaparecimento do conflito.

CAPÍTULO IV

CONSISTÊNCIA E GERAÇÃO DE AUTÔMATOS R*S(1) CANÔNICOS

Neste capítulo apresentaremos as condições que precisam ser satisfeitas para que um autômato analisador R*S(1) canônico seja determinístico. Veremos também os algoritmos de geração do conjunto de estados Q_c, da função de transição δ_c e das (possivelmente) funções e_c, d_c e r_c. Note que k = 1 em todo este capítulo.

IV.1 - CONSISTÊNCIA DOS AUTÔMATOS R*S(1) CANÔNICOS

Definição IV.1: dizemos que AC satisfaz o Teste de Consistência (TC) se nenhuma das condições (a), (b), (c) ou (d) abaixo são satisfeitas

(a) existe $p \in Q_c$ tal que

$[B \to \beta.,a] \in p$

$[D \to \alpha.a\gamma,b] \in p$

(b) existe $p \in Q_c$ tal que

$[B_1 \to \gamma\beta.,a] \in p$

$[B_2 \to \beta.,a] \in p, \quad \gamma \neq \varepsilon$

(c) existem p q $\in Q_c$ tais que

(1) $[B_1 \to \beta.,a] \in p$

$[B_2 \to \beta.,a] \in p$

(2) $\delta_c(q,\beta) = p$

(3) $A_1 \overset{\sigma_1}{\Rightarrow_s} B_1$

$A_2 \overset{\sigma_2}{\Rightarrow_s} B_2$

$\sigma_1 B_1 \to \beta \neq \sigma_2 B_2 \to \beta$

(4) $[D_1 \to \xi_1.A_1\gamma_1,b] \in q \qquad \xi_1\gamma_1 \neq \varepsilon$

$[D_2 \to \xi_2.A_2\gamma_2,c] \in q \qquad \xi_2\gamma_2 \neq \varepsilon$

$a \in FIRST(\gamma_1 b)$
$a \in FIRST(\gamma_2 c)$

(d) $[B\to\beta.,\$] \in q_{fc}$ com $B\to\beta \neq S'\to\$S$

<u>Teorema IV.2</u>: Seja $AC = (Q_c,\Sigma,P,e_c,d_c,r_c,q_{oc},q_{fc})$ um autômato analisador R*S canônico construído a partir de uma gramática G. AC satisfaz TC se e somente se AC é determinístico.

Demonstração: é consequência dos Lemas IV.3 e IV.4 a seguir e do Corolário do Lema III.16.

<u>Lema IV.3</u>: nas mesmas condições do Teorema IV.2, se AC satisfaz TC então AG é determinístico.

Demonstração: se AG não é determinístico então AC não satisfaz TC.

Caso 1: $H \in$ ação (φ,a)
$E \in$ ação (φ,a)

Temos $\varphi a = \$S\$$. $\$S\$z \in CR{*}S$ com $z \neq \varepsilon$, o que não pode ocorrer pela construção da gramática estendida.

Caso 2: $H \in$ ação (φ,a)
$(X,\pi) \in$ ação(φ,a), $X \in \{RE,RNE\}$, $\pi = \sigma B\to\beta$, $\sigma \in s(A,B)$

Temos $\varphi a = \$S\$$
$\$S\$ \in CR{*}S\ (B\to\beta)$

e pelo Lema III.7

$[B\to\beta.,\$] \in \nu_c(\$S)$

satisfazendo o TC, condição d, e portanto, violando o TC.

Caso 3: $E \in$ ação(φ, a)
$(X, \pi) \in$ ação(φ, a), $X \in \{RE, RNE\}$, $\pi = \sigma B \to \beta$, $\sigma \in s(A, B)$.

De $E \in$ ação(φ, a) temos que $\varphi a z \in CR^*S$, $z \neq \varepsilon$.

$$S'\$^k \equiv>^+ \varphi a z w$$

Vamos supor que a ocorrência do símbolo a tenha sido introduzida através da produção $D \to \alpha a \gamma$

$$S'\$^k \equiv>^* \xi C b x \overset{\pi_1}{\equiv>} \xi \alpha a \gamma b x \equiv>^* \varphi a z w$$

com $\pi_1 = \sigma_1 \; D \to \alpha a \gamma$, $\sigma_1 \in s(C, D)$, $\xi\alpha = \varphi$

Temos

$$\xi \alpha a \gamma b \in CR^*S(D \to \alpha a \gamma)$$

e pelo Lema III.7.

$$[D \to \alpha . a \gamma, b] \in \nu_c(\varphi)$$

Por outro lado, de $(X, \pi) \in$ ação (φ, a),

$$\varphi a \in CR^*S(B \to \beta)$$

e pelo Lema III.7,

$$[B \to \beta ., a] \in \nu_c(\varphi)$$

Os resultados acima mostram que a condição a do TC é satisfeita, e portanto AC não satisfaz TC.

Caso 4: $(X, \pi_1) \in$ ação(φ, a)
$(Y, \pi_2) \in$ ação(φ, a), $X, Y \in \{RE, RNE\}$
$\pi_1 = \sigma_1 \; B_1 \to \beta_1$, $\sigma_1 \in s(A_1, B_1)$
$\pi_2 = \sigma_2 \; B_2 \to \beta_2$, $\sigma_2 \in s(A_2, B_2)$
$\beta_1 \neq \beta_2$

Temos:

$$\varphi a \in CR^*S\ (B_1 \to \beta_1)$$

$$\varphi a \in CR^*S\ (B_2 \to \beta_2)$$

Logo $\varphi = \alpha_1\beta_1 = \alpha_2\beta_2$

Note que $|\beta_1| = |\beta_2|$ implica em $\beta_1 = \beta_2$, e portanto $|\beta_1| \neq |\beta_2|$. Vamos assumir $|\beta_1| > |\beta_2|$. Temos então $\beta_2 = \beta$, $\beta_1 = \gamma\beta$ com $\gamma \neq \varepsilon$, $\alpha_2 = \alpha_1\gamma$, $\varphi = \alpha_1\gamma\beta$. Como $\alpha_1\gamma\beta a \in CR^*S\ (B_1 \to \gamma\beta)$ e $\alpha_1\gamma\beta a \in CR^*S\ (B_2 \to \beta)$, pelo Lema III.7, temos

$$[B_1 \to \gamma\beta.,a] \in \nu_c(\varphi)$$

$$[B_2 \to \beta.,a] \in \nu_c(\varphi)$$

Assim, $\nu_c(\varphi)$ satisfaz a condição b do TC, e portanto AC não satisfaz TC.

Caso 5: $(X,\pi_1) \in$ ação(φ,a)
$(Y,\pi_2) \in$ ação(φ,a)
$X,Y \in \{RE,RNE\}$
$\pi_1 = \sigma_1 \quad B_1 \to \beta, \quad \sigma_1 \in s(A_1,B_1)$
$\pi_2 = \sigma_2 \quad B_2 \to \beta, \quad \sigma_2 \in s(A_2,B_2)$
$(X,\pi_1) \neq (Y,\pi_2)$

Temos:

$$\varphi a \in CR^*S\ (B_1 \to \beta)$$

$$\varphi a \in CR^*S\ (B_2 \to \beta)$$

$$\varphi = \alpha\beta$$

Caso 5.1: $\pi_1 = \pi_2$, $X \neq Y$. Sem perda de generalidade vamos adotar $X = RE$ e $Y = RNE$.

Temos $\alpha Aaz \in CR^*S$ com $z \neq \varepsilon$, $E \in$ ação$(\alpha A,a)$.

Temos também $\alpha Aa \in CR^*S$, de forma que

$$S'\$ \equiv>^+ \alpha Aaw$$

Há dois casos a considerar:

Caso 5.1.1: $S'\$ \equiv> \$S\$ = \alpha Aaw$ e $H \in$ ação$(\$S,\$)$. Já vimos no caso 1 que isto não pode ocorrer.

Caso 5.1.2: $S'\$ \equiv>^+ \xi Caw \overset{\pi}{\equiv>} \xi\gamma Aaw$ com $\pi = \sigma\ D\to\gamma A$, $\sigma \in s(C,D)$.

Pelo Lema II.17

$$(Z,\pi) \in \text{ação}(\alpha A,a)$$

Este resultado, em combinação com $E \in$ ação$(\alpha A,a)$, como visto no caso 2 acima, caracteriza uma violação do TC.

Caso 5.2: $\pi_1 \neq \pi_2$

Temos, de $\varphi a \in CR^*S$ $(B_1\to\beta)$,

$$[B_1\to\beta.,a] \in \nu_c(\varphi)$$

De $\varphi a \in CR^*S$ $(B_2\to\beta)$,

$$[B_2\to\beta.,a] \in \nu_c(\varphi).$$

Da definição de ν_c,

$$\delta_c(\nu_c(\varphi),\beta) = \nu_c(\alpha\beta) = \nu_c(\varphi)$$

Temos também

$$A_1 \overset{\sigma_1}{\Rightarrow_s} B_1$$

$$A_2 \overset{\sigma_2}{\Rightarrow_s} B_2$$

$\pi_1 = \sigma_1 \; B_1{\to}\beta \qquad \pi_2 = \sigma_2 \; B_2{\to}\beta$

Por outro lado, de

$(X, \pi_1) \in$ ação(φ, a),

$\pi_1 = \sigma_1 \quad B_1{\to}\beta$,

$\varphi = \alpha\beta$,

temos $\alpha A_1 az \in CR^*S$ com $|z| \geq 0$, dependendo do valor de X. Ou seja,

$$S'\$^k \equiv>^+ \alpha A_1 azw$$

Explicitando o passo da derivação que introduz A_1 através de uma regra do tipo $D_1 \to \xi_1 A_1 \gamma_1$,

$$S'\$ \equiv>^* \psi_1 C_1 by \equiv> \psi_1 \xi_1 A_1 \gamma_1 by \equiv>^* \alpha A_1 azw$$

com $C_1 \Rightarrow_s^* D_1$, $\alpha = \psi_1 \xi_1$ e $\gamma_1 by \equiv>^* azw$, de forma que $a \in \mathrm{FIRST}(\gamma_1 b)$

Temos

$$\alpha A_1 \gamma_1 b \in CR^*S \; (D_1{\to}\xi_1 A_1 \gamma_1)$$

de forma que $[D_1{\to}\xi_1 . A_1 \gamma_1, b) \in \nu_c(\alpha)$.

De forma semelhante

$$[D_2{\to}\xi_2 . A_2 \gamma_2, c] \in \nu_c(\alpha)$$

com $a \in \mathrm{FIRST}(\gamma_2 c)$

Portanto, a condição c do TC é satisfeita para $p = \nu_c(\varphi)$ e $q = \nu_c(\alpha)$ e desse modo o TC é violado.

<u>Lema IV.4</u>: nas mesmas condições do Teorema IV.2, AG é determinístico então AC satisfaz TC.

Demonstração: será feita provando-se que AC não satisfaz TC implica que AG não é determinístico.

Caso 1: AC não satisfaz a condição (a) do TC. Ou seja, existe $p \in Q_c \mid [B\to\beta.,a] \in p$, $[D\to\alpha.a\gamma,b] \in p$.

Seja φ tal que $\nu_c(\varphi) = p$.

Pelo Lema III.6, de $[B\to\beta.,a] \in \nu_c(\varphi)$, temos

$$\varphi a \in CR^*S\ (B\to\beta)$$

ou seja

$$S'\$ \Rightarrow^* \alpha A a w \overset{\pi}{\Rightarrow} \alpha\beta a w$$

com $\varphi = \alpha\beta$, $\pi = \sigma\ B\to\beta$, $\sigma \in s(A,B)$.

Devemos considerar dois casos:

(1) $S'\$ = \alpha A a w$, $\pi = S'\to\$S$
Temos $H \in$ ação$(\$S,\$) =$ ação(φ,a)

(2) $S'\$ \Rightarrow^+ \alpha A a w$

Pelo Lema II.17 (b), $(X,\pi) \in$ ação$(\alpha\beta,a) =$ ação(φ,a)

Por outro lado, de $[D\to\alpha.a\gamma,b] \in \nu_c(\varphi)$, pelo Lema II.6 $\varphi a\gamma b$ CR*S $(D\to\alpha a\gamma)$.

Temos $S'\$ \equiv>^{+} \varphi a\gamma bx$. Suponha $\gamma \equiv>^{*} y$. Então $S'\$ \equiv>^{+} \varphi aybx$. Explicitando o último passo dessa derivação:

$$S'\$ \equiv>^{*} \psi A_1 c w_1 \equiv> \psi\beta_1 c w_1 = \varphi aybx$$

de forma que $\psi\beta_1 \in CR^*S$. Naturalmente, como a derivação é direita $|\psi\beta_1 c| > |\varphi a|$ e portanto $\varphi az = \psi\beta_1 c \in CR^*S$ para $z \neq \varepsilon$.

Consequentemente,

$$E \in \text{ação}(\varphi,a)$$

e em qualquer dos casos vistos acima, $|\text{ação}(\varphi,a)| > 1$, de forma que AG não é determinístico.

Caso 2: AC não satisfaz a condição (b) do TC. Ou seja, existe $p \in Q_c \mid [B_1 \to \gamma\beta.,a] \in p$, $\gamma \neq \varepsilon$, e $[B_2 \to \beta.,a] \in p$.

Seja $p = \nu_c(\varphi)$

$[B_1 \to \gamma\beta.,a] \in \nu_c(\varphi)$ implica, pelo Lema III.6, que $\varphi a \in CR^*S\ (B_1 \to \gamma\beta)$.

$[B_2 \to \beta.,a] \in \nu_c(\varphi)$ implica, pelo Lema III.6, que $\varphi a \in CR^*S\ (B_2 \to \beta)$.

Logo $\varphi \doteq \alpha\gamma\beta$. Podemos dizer que

$$S'\$ \equiv>^{*} \alpha\gamma A_2 a w_2 \overset{\pi_2}{\equiv>} \alpha\gamma\beta a w_2$$

com $\pi_2 = \sigma_2\ B_2 \to \beta,\ \sigma_2 \in s(A_2,B_2)$

Se $S'\$ = \alpha\gamma A_2 a w_2$, $\gamma = \varepsilon$, o que não pode ocorrer. Logo,

$$S'\$ \equiv>^{+} \alpha\gamma A_2 a w_2 \overset{\pi_2}{\equiv>} \alpha\gamma\beta a w$$

e pelo Lema II.17, $(X,\pi_2) \in \text{ação}(\varphi,a)$ para $X \in \{RE,RNE\}$.

Temos também

$$S'\$ \Rightarrow^* \alpha A_1 a w_1 \overset{\pi_1}{\Rightarrow} \alpha\gamma\beta a w_1$$

com $\quad \pi_1 = \sigma_1 \; B_1 \to \beta, \quad \sigma_1 \in s(A_1, B_1).$

Se tivermos $S'\$ = \alpha A_1 a w_1$, $\pi_1 = S' \to \$S$ e $H \in$ ação(φ, a). Caso contrário,

$$S'\$ \Rightarrow^+ \alpha A_1 a w_1 \overset{\pi_1}{\Rightarrow} \alpha\gamma\beta a w_1$$

e pelo Lema II.17, $(Y, \pi_1) \in$ ação(φ, a).

Em qualquer das situações acima, como $\gamma \neq \varepsilon$ implica em $\pi_1 \neq \pi_2$, ação(φ,a) tem mais de um elemento e portanto AG não é determinístico.

Caso 3: AC não satisfaz a condição (c) dO TC. Então existem em Q_c dois estados p e q tais que

(c1) $[B_1 \to \beta., a] \in p$
$[B_2 \to \beta., a] \in p$

(c2) $\delta_c(q, \beta) = p$

(c3) $A_1 \overset{\sigma_1}{\Rightarrow_s} B_1$

$A_2 \overset{\sigma_2}{\Rightarrow_s} B_2$

$\sigma_1 \; B_1 \to \beta \neq \sigma_2 \; B_2 \to \beta$

(c4) $[D_1 \to \xi_1 . A_1 \gamma_1, b] \in q, \; \xi_1 \gamma_1 \neq \varepsilon$
$[D_2 \to \xi_2 . A_2 \gamma_2, c] \in q, \; \xi_2 \gamma_2 \neq \varepsilon$
$a \in FIRST(\gamma_1 b)$
$a \in FIRST(\gamma_2 c)$

Seja $\alpha \in V^*$ tal que $\nu_c(\alpha) = q$; $\nu_c(\alpha\beta) = p$, portanto

$[B_1 \rightarrow \beta., a] \in \nu_c(\alpha\beta)$

$\alpha\beta a \in CR{*}S\ (B_1 \rightarrow \beta)$

$[D_1 \rightarrow \xi_1.A_1\gamma_1, b] \in \nu_c(\varphi)$, $\alpha = \psi_1\xi_1$ e

$\alpha A_1\gamma_1 b \in CR{*}S\ (D_1 \rightarrow \xi_1 A_1\gamma_1)$.

Podemos então escrever que

$$S'\$ \equiv>^* \psi_1 C_1 b y_1 \equiv> \psi_1\xi_1 A_1\gamma_1 b y_1 = \alpha A_1\gamma_1 b y_1 \equiv>^*$$

$$\equiv>^* \alpha A_1 a x_1 \overset{\pi_1}{\equiv>} \alpha\beta a x_1 = \varphi a x_1 \qquad \text{já que } a \in FIRST(\gamma_1 b)$$

Pelo Lema II.17,

$(X, \pi_1) \in ação(\varphi, a)$, $\pi_1 = \sigma_1\ B_1 \rightarrow \beta$

De forma semelhante

$(Y, \pi_2) \in ação(\varphi, a)$, $\pi_2 = \sigma_2\ B_2 \rightarrow \beta$

e portanto AG não é determinístico.

Caso 4: AC não satisfaz a condição (d) do TC. Então existe $[B \rightarrow \beta., \$] \in q_{fc}$ com $B \rightarrow \beta \neq S' \rightarrow \$S$.

Temos por construção, $\nu_c(\varphi) = q_{fc} \Rightarrow \varphi = \$S$.

Se $[B \rightarrow \beta., \$] \in \nu_c(\$S)$ então $\$S = \alpha\beta$, $\$S\$ \in CR{*}S\ (B \rightarrow \beta)$

Caso 4.1: $\alpha = \varepsilon$, $\beta = \$S$, obriga $B = S'$, contradiz a hipótese.

Caso 4.2: $\alpha = \$$, $\beta = S$ contraria o fato de que ítens simples completo não podem compor estados R*S e desta forma $[B \rightarrow \beta., \$]$ seria um ítem simples completo.

Caso 4.3: α = \$S, β = ε. É o único caso possível. Note que [$B\to.$,\$] $\in \nu_c$(\$S) implica, pelo Lema III.6 que \$S\$ $\in$ CR*S ($B\to\varepsilon$) e portanto

$$S'\$ \Rrightarrow^{+} \$SA\$ \overset{\pi}{\Rrightarrow} \$S\$$$

com $\pi = \sigma\ B\to\varepsilon$, $\sigma \in s(A,B)$. Pelo Lema II.17, $(X,\pi) \in$ ação(\$S,\$) o que mostra, conjuntamente com H $\in$ ação(\$S,\$) que AG não é determinístico.

Corolário: AG é determinístico se e somente se AC satisfaz TC.

É uma consequência dos Lemas IV.3 e IV.4.

IV.2 - ALGORITMOS PARA A CONSTRUÇÃO DE UM AUTÔMATO ANALISADOR R*S(1) CANÔNICO

Algoritmo IV.5: construção do conjunto de estados Q_c e da função de transição δ_c de um autômato R*S(1) canônico.

Entrada: uma gramática $G = (N,\Sigma,P,S')$, estendida

Saída: o conjunto de estados Q_c e a função de transição δ_c.

Obs: deltacanônico (estado, símbolo) é um procedimento que realiza a definição III.2; tabdelta é a tabela da função δ_c.

Algoritmo:

```
   qoc:= fecho({[S'→$S.,$]}); Qc:= {qoc};
marque qoc "não processado";
enquanto há estados "não processados" em Qc

faça

   q := estado não processado em Qc;
   para todo X ∈ V
   faça

        p := deltacanônico(q,X);
        se p ≠ Ø então tabdelta [q,X] := p;
        se p ∉ Qc
        então Qc := Qc ∪ {p};
              marcar p "não processado";
   fim-do-para-todo;
   marcar q "processado"
fim do enquanto
```

Observação: os algoritmos desta seção fazem uso dos Teoremas vistos nas seções anteriores emitindo a mensagem de que a gramática fornecida não é R*S(1) quando da construção de

d_c e r_c, se descobrir que estas relações não são funções. Por esta razão dizemos que estes algorimos produzem as funções d_c e r_c, uma vez que, não sendo funções, além da emissão da mensagem o processamento pode ser suspenso ou continuado com a finalidade de que sejam identificados todos os pontos de conflito.

Definição IV.6: conjunto Q_e de estados de um autômato R*S(1) canônico

$$Q_e = \{p \in Q_c \mid ([A \to \alpha a.\beta, b] \in p) \wedge ([B \to \gamma.c\eta, d] \in p)\}$$

Algoritmo IV.7: construção da função e_c

Entrada: conjunto Q_c de estados do analisador AC e a função δ_c,

Saída: a função e_c

Obs: $e_c : Q_e \times \Sigma \to Q_c$

Algoritmo

```
para todo p ∈ Qe
faça
    para todo a ∈ Σ
    faça
        se δc(p,a)↓
        então  ec(p,a):= δc(p,a)
```

Observação: por construção, δ_c é uma função, e_c é uma restrição de δ_c é também uma função.

Algoritmo IV.8: construção da função d_c

Entrada: o conjunto de estados Q_c e a função e_c

Saída: a função d_c ou a informação que G não é R*S(1)

Algoritmo

para todo $p \in Q_c$
faça
    para todo $[A \to \alpha., a]$ em p
    faça
        se $d_c(p,a)\uparrow$ ou $d_c(p,a) = |\alpha|$
        então $d_c(p,a) := |\alpha|$
        senão mensagem ("G não é R*S (1)");
        se $e_c(p,a)\downarrow$ e $d_c(p,a)\downarrow$
        então mensagem ("G não é R*S (1)")
    fim-do-faça
fim-do-faça

Algoritmo IV.9: construção da função r_c

Entrada: o conjunto Q_c de estados,
a função de transição δ_c
a função d_c,
os pares de não terminais (A,B) tais que $A \overset{*}{\underset{s}{\Rightarrow}} B$

Saída: a função r_c ou a informação de que G não é E*S (1)

Algoritmo:

para todo $q \in Q_c$ tal que $[B \to .\beta, a] \in q$, $\beta \notin N$
faça
    $p := \delta_c(q,\beta)$;
    para todo não-terminal A tal que $[E \to \alpha.A\gamma, b] \in q$
    e tal que $A \overset{*}{\underset{s}{\Rightarrow}} B$ e $a \in FIRST(\gamma b)$
    faça
        se $\delta_c(q,Aa)\downarrow$ e $r_c(q,a,p)\uparrow$
        então (para $p_1 = \delta_c(q,A)$ e $p_2 = \delta_c(q,Aa)$
            $r_c(q,a,p) := ((p_1,p_2), s(A,B)\ B \to \beta)$
        senão se $\delta_c(q,A)\downarrow$ e $r_c(q,a,p)\uparrow$
            então (para $p_1 = \delta_c(q,A)$)
            $r_c(q,a,p) := (p_1, s(A,B) B \to \beta)$

L.13 senão mensagem ("não é R*S (1)")
L.14 fim-do-para-todo
L.15 fim-do-para-todo

Observação: os algoritmos que produzem e_c, d_c e r_c verificam as condições a, b e c impostas no TC. A verificação da condição (d) é trivial.

<u>Condição a</u>: - verificada na linha 8 do Algoritmo IV.8. A ocorrência simultânea de $e_c(p,a)$ e $d_c(p,a)$ indica conflito empilha/reduz no estado de p de AC.

<u>Condição b</u>: é verificada na linha 5 do Algoritmo IV.8. Quando $d_c(p,a)\downarrow$ e $d_c(p,a) \neq |\alpha|$ ocorrer o desvio para a condição "senão" do teste (linha 7 do Algoritmo IV.8), fruto da existência de dois valores distintos, representando duas reduções por regras com diferentes comprimentos de lado direito no mesmo estado p.

<u>Condição c</u>: - observe que ao entrar na região de testes, o Algoritmo IV.9 já fixou diversos valores, a saber

- o estado q
- o ítem completo $[B \to \beta.,a] \in q$, e seu lookahead <u>a</u>
- o estado $p = \delta_c(q,\beta)$
- o não terminal A tal que $A \Rightarrow^*_S B$

A condição "então" da linha 8 atribui valor a $r_c(q,a,p)$, que ainda não foi instaciada, deixando a condição "senão" para os casos que $\delta_c(q,Aa)\uparrow$ ou $r_c(q,a,p)\downarrow$. O teste da linha 10 causa a atribuição de valor a $r_c(q,a,p)$ quando $r_c(q,a,p)\uparrow$ deixando para a condição senão da linha 13 os casos em que

$$(\delta_c(q,Aa)\uparrow, \delta_c(q,A)\uparrow) \text{ ou } r_c(q,a,p)\downarrow$$

indicando assim a tentativa de uma segunda atribuição de valor a $r_c(q,a,p)$, fruto da existência de um segundo não terminal (representado no algoritmo por A) tal que deriva simples B que deriva β detectando assim ambigüidade na gramática que produziu Q_c.