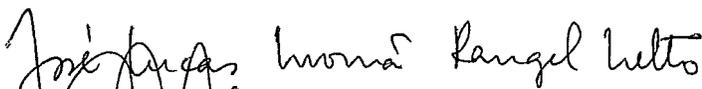


GRAMÁTICAS E LINGUAGENS R\*5(k)

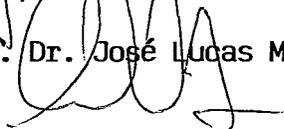
Sérgio de Mello Schneider

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)

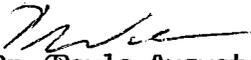
Aprovada por:



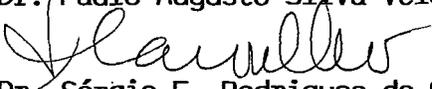
Prof. Dr. José Lucas Mourão Rangel Netto - Presidente



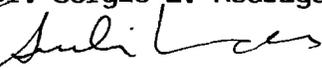
Dr. Arnaldo Vieira Moura



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

<u>CAPÍTULO I</u> - INTRODUÇÃO . . . . .	01
I.1 - Notações e Definições . . . . .	03
<u>CAPÍTULO II</u> - 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
<u>CAPÍTULO III</u> - AUTÔMATOS R*S CANÔNICOS . . . . .	40
III.1 - Introdução . . . . .	40
III.2 - Construção de Autômatos R*S Canônicos . . . . .	40
<u>CAPÍTULO IV</u> - 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 analisa dor R*S(1) Canônico . . . . .	70
<u>CAPÍTULO V</u> - 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  $A \rightarrow B$  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 itens. 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  $(N, \Sigma, P, S)$ , conforme a notação usual. Se dissermos que  $G$  é uma gramática estendida, descrita pela quádrupla  $(N, \Sigma, P, S')$  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  $\Sigma_0$ ),  $N = N_0 \cup \{S'\}$  (sendo  $S'$  um novo símbolo não-terminal inicial, não pertencente a  $N_0$ ),  $P = P_0 \cup \{S' \rightarrow \$S\}$  onde  $S' \rightarrow \$S$  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)  $\pi, \pi_1, \pi_2, \dots$  representarão elementos em  $P^*$
- b)  $u, v$  representarão elementos de  $\Sigma^k$
- 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  $A \rightarrow \beta$  com a sequência  $\pi$  de regras por  $A \rightarrow \beta \pi$ .

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  $\dagger$  e  $\ddagger$  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  $\psi$  deriva  $\varphi$  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  $s(A, A) = \{ \epsilon \}$ .

Durante todo o texto usaremos somente derivações mais à direita, representada pelo símbolo  $\Rightarrow$  em lugar de símbolos como  $\Rightarrow_{rm}$  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  $\alpha$  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

$$\alpha A \gamma \Rightarrow \alpha \beta \gamma$$

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 ítems 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 itens LR(k) de uma gramática G é dado por

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

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

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

se  $\alpha\beta = X$ ,  $X \in N$ , então  $\beta \neq \epsilon\}$

Definição I.17: sejam p e q dois conjuntos de itens. Dizemos que  $\text{fecho}(q) = p$  se p for o menor conjunto de itens satisfazendo

1.  $p \supseteq q$
2.  $[A \rightarrow \alpha \cdot B\beta, u] \in q$   
implica  $[B \rightarrow \cdot \gamma, v] \in p$  para todo  
 $v \in \text{FIRST}(\beta u)$  e  $B \rightarrow \gamma \in P$ .

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

$$\text{núcleo}([A \rightarrow \alpha \cdot \beta, u]) = [A \rightarrow \alpha \cdot \beta, \epsilon].$$

A representação de itens da forma  $[A \rightarrow \alpha \cdot \beta, \epsilon]$  será simplificada para  $[A \rightarrow \alpha \cdot \beta]$

Estendemos a definição para um conjunto de itens q:

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

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

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

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

$$\begin{aligned} SPNS(p) &= \epsilon \text{ se } p \text{ é uma produção simples} \\ SPNS(p) &= p \text{ se } p \text{ é uma produção não simples} \end{aligned}$$

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

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

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

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

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

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 da as 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 toma das a partir do conhecimento de todo o conteúdo da pilha e mais  $k$  símbolos da cadeia de entrada ( $k \geq 0$ ). Como resultado, o autô<sup>u</sup>mato 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 conheci<sup>-</sup>mento 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

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

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

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

implica que  $\alpha \beta y = \gamma \delta x$ , 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

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

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

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

implica que  $\alpha \beta y = \gamma \delta x$  (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

$$A \stackrel{\pi_1}{\Rightarrow}_S B \text{ e } A \stackrel{\pi_2}{\Rightarrow}_S B, \pi_1 \neq \pi_2.$$

Portanto, se  $B \stackrel{\pi_3}{\equiv} \beta$

$$S' \$^k \equiv^* \alpha A w \stackrel{\pi_1 \pi_3}{\equiv} \alpha \beta w$$

$$S' \$^k \equiv^* \alpha A w \stackrel{\pi_2 \pi_3}{\equiv} \alpha \beta w$$

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:

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

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

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

$$\pi_1 \neq \pi_2$$

Sejam

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

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

Então

$$S' \$^k \Rightarrow^* \alpha A w \xrightarrow[S]{\Rightarrow^*} \alpha B w$$

$$S' \$^k \Rightarrow^* \gamma C x \xrightarrow[S]{\Rightarrow^*} \gamma D x \Rightarrow \gamma \delta x = \alpha \beta y$$

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

Portanto

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

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

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

Dependendo da estrutura de  $\sigma_1$  e  $\sigma_2$  há dois casos a considerar.

Caso 1:

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

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

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:

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

Temos

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

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

Se  $\alpha C y = S' \$^k$  então  $S' = C$  e  $E \rightarrow S' \in P$ , 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  $F \rightarrow C$ !

A condição  $R^*S$  também é violada quando  $\pi_1 = \pi_2$  e  $\alpha A y \neq \gamma C x$ .

Ou seja:

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

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

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

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

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

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

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

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

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

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

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

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

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

Pela condição  $R^*S(k)$   $\alpha A y = \gamma C x$  e  $\sigma_1$   $B \rightarrow \beta = \sigma_2$   $D \rightarrow \delta$  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  $B \rightarrow E$  e  $D \rightarrow F$  foram usadas nos passos finais das derivações (\*) acima. Temos:

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

$$S' \$^k \Rightarrow^* \gamma D x \Rightarrow \gamma F x = \alpha E y$$

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

$$S' \$^k \Rightarrow^* \alpha Aw \Rightarrow_S^* \alpha Bw \Rightarrow \alpha Ew \Rightarrow_S^* \alpha Hw \Rightarrow \alpha \varphi w$$

$$S' \$^k \Rightarrow^* \alpha Cy \Rightarrow_S^* \alpha Dy \Rightarrow \alpha Ey \Rightarrow_S^* \alpha Hy \Rightarrow \alpha \varphi y$$

Aplicando a condição  $R^*S(k)$ , temos  $\alpha Ay = \alpha Cy$  e  $\pi_1 = \sigma_1$   $B \rightarrow E$   $\sigma_2$   $H \rightarrow \varphi$   $= \pi_2 = \sigma_3$   $D \rightarrow E$   $\sigma_2$   $H \rightarrow \varphi$  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  $D \rightarrow E$  e  $B \rightarrow \beta$ . 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),

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

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

Reescrevendo e identificando, temos:

$$S' \$^k \Rightarrow^* \gamma Exw \stackrel{\pi_1}{\Rightarrow} \gamma Ezw \quad \pi_1 = s(E, F) F \rightarrow \delta$$

$$S' \$^k \Rightarrow^* \gamma Cx \stackrel{\pi_2}{\Rightarrow} \gamma \delta x \quad \pi_2 = s(C, D) D \rightarrow E \quad s(E, F) F \rightarrow \delta$$

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  $\varphi x \in CR^*S$ , com  $x \neq \epsilon$   
então  $E \in ação(\varphi, u)$
- (2) se  $\alpha\beta u \in CR^*S(B \rightarrow \beta)$  e  
 $A \xrightarrow[S]{\sigma} B$  e  
 $\alpha Aux \in CR^*S$ , com  $x \neq \epsilon$   
então  $(RE, \sigma B \rightarrow \beta) \in ação(\alpha\beta, u)$
- (3) se  $\alpha\beta u \in CR^*S(B \rightarrow \beta)$  e  
 $A \xrightarrow[S]{\sigma} B$  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.  $\varphi$  representa o conteúdo da pilha  $x$  representa a porção não lida da cadeia de entrada e  $\pi$  representa a sequência de reduções ocorridas até aquele instante.

Dizemos que  $(\$, x\$, \epsilon)$  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  $E \in \text{ação}(\varphi, \text{FIRST}(ax))$
- (2)  $(\alpha\beta, ax, \pi) \vdash (\alpha Aa, x, \sigma B \rightarrow \beta \pi)$   
se  $(RE, \sigma B \rightarrow \beta) \in \text{ação}(\alpha\beta, \text{FIRST}(ax))$ ,  $\sigma \in s(A, B)$
- (3)  $(\alpha\beta, x, \pi) \vdash (\alpha A, x, \sigma B \rightarrow \beta \pi)$   
se  $(RNE, \sigma B \rightarrow \beta) \in \text{ação}(\alpha\beta, \text{FIRST}(x))$ ,  $\sigma \in s(A, B)$

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

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

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

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

$$C_0 \vdash C_1 \vdash C_2 \vdash \dots \vdash C_n$$

Então notaremos

$$C_0 \vdash^n C_n$$

para indicarmos que a configuração  $C_0$  se transforma na configuração  $C_n$  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.

Teorema (Fundamental) II.11: 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.

Lema II.12: Seja  $\varphi ux = \varphi yv \in CR*S$ .

Então  $(\varphi, uxz, \pi) \vdash^* (\varphi y, v z, \pi)$

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.

Lema II.13: 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  $S'\$^k \Rightarrow^+ \$x\$^k$  identificando os passos intermediários.

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

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

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

$$\text{e } \alpha_i = \gamma_i \beta_i w_i$$

$$\text{e } \pi_i \text{ é tal que } \alpha_i \stackrel{\pi_i}{\equiv} \$x\$^k$$

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

Vamos definir ainda  $D_0 = (\$, x\$^k, \epsilon)$  a configuração inicial para  $\$x$ . Como  $\alpha_0 = S'\$^k$  e  $\alpha_1 = \$S\$^k$  temos  $\gamma_1 = \epsilon$ ,  $\beta_1 = \$S$ ,  $w_1 = \$^k$  e portanto

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

$$\text{com } \pi_1 \text{ tal que } \$S\$^k \stackrel{\pi_1}{\equiv} \$x\$^k$$

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

Falta provar que:

$$(1) D_0 \stackrel{*}{|-} C_n$$

$$(2) C_i \stackrel{*}{|-} C_{i-1} \text{ para } i = n, n-1, \dots, 2.$$

Para provarmos (1) observamos que

$$D_0 = (\$, x\$^k, \epsilon)$$

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

$$\alpha_n = \gamma_n \beta_n w_n = \$x\$^k \stackrel{\epsilon}{\equiv} \$x\$^k$$

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

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

$$D_0 \vdash^* C_n$$

Para provarmos (2), temos

$$\alpha_{i-1} = \gamma_i A_i w_i \Rightarrow_S^* \gamma_i B_i w_i \Rightarrow \alpha_i$$

$$\alpha_i = \gamma_i \beta_i w_i \stackrel{\pi_i}{\equiv} \text{\$}\text{\$}^k$$

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

Por outro lado

$$\alpha_{i-2} = \gamma_{i-1} A_{i-1} w_{i-1} \Rightarrow_S^* \gamma_{i-1} B_{i-1} w_{i-1} \Rightarrow \alpha_{i-1}$$

$$\alpha_{i-1} = \gamma_{i-1} \beta_{i-1} w_{i-1} \stackrel{\pi_{i-1}}{\equiv} \text{\$}\text{\$}^k$$

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

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

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

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

Portanto ação  $(\gamma_i \beta_i, \text{FIRST}(w_i)) \ni (\text{RNE}, \sigma_i B_i \rightarrow \beta_i)$ ,

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

$$\begin{aligned} C_i &= (\gamma_i \beta_i, w_i, \pi_i) \vdash (\gamma_i A_i, w_i, \sigma_i B_i \rightarrow \beta_i, \pi_i) = \\ &= (\gamma_{i-1} \beta_{i-1}, w_{i-1}, \pi_{i-1}) = C_{i-1} \end{aligned}$$

2.2 -

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

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

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

Portanto, temos:

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

ou

$$D_0 \vdash^* C_1$$

$$e \quad (\$, x \$^k, \epsilon) \vdash^* (\$, \$^k, \pi_1)$$

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

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

Lema II.14: (ainda dentro das condições do Teorema II.11) se  $\$x \in L(\text{AG})$  então  $\$x \in L(G)$ .

Demonstração: como  $\$x \in L(\text{AG})$ , podemos afirmar que  $(\$, x \$^k, \epsilon) \vdash^* (\$, \$^k, \pi)$  para algum  $\pi \in P^*$ . Seja  $D_0 = (\$, x \$^k, \epsilon)$ . Nas configurações  $C_i$  atingidas no Lema anterior, vamos distinguir 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)  $C_1, C_2, \dots, C_n$  essas configurações distinguidas.

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

$$\begin{aligned} (\$, x \$^k, \epsilon) &= D_0 \vdash^* C_1 \vdash^+ C_2 \vdash \dots \vdash^+ C_n = \\ &= (\$, \$^k, \pi) \end{aligned}$$

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

$$S' \$^k \Rightarrow \$S\$^k = \alpha_n \Rightarrow \dots \Rightarrow \alpha_1 = \$x\$^k$$

ou de maneira mais geral

$$\alpha_i \stackrel{\pi}{\Rightarrow} \$x\$^k \quad 1 \leq i \leq n$$

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

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

e logo

$$\alpha_1 \stackrel{\epsilon}{\Rightarrow} \$x\$^k$$

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

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

$$\varphi_i = \gamma_i \beta_i$$

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

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

$(RE, \sigma_i B_i \rightarrow \beta_i) \in \text{ação}(\gamma_i \beta_i, u)$  é ação usada.

Temos, pela definição de ação

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

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

$$A_i \stackrel{\sigma_i}{\equiv} B_i$$

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

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

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

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

$$\alpha_{i+1} \stackrel{\sigma_i B_i \rightarrow \beta_i}{\equiv} \alpha_i \stackrel{\pi_i}{\equiv} \$x\$^k$$

ou

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

caso 2: redução sem empilhamento. Então

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

$$\varphi_i = \gamma_i \beta_i$$

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

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

Então pela definição de ação

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

$$\gamma_i A_i \in CR^*S$$

$$A_i \xrightarrow[S]{\sigma_i} B_i$$

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

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

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

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

cqd (do Lema II.14)

Corolário

$$\$S\$^k \stackrel{\pi}{\equiv} \$x\$^k \text{ implica } S \stackrel{\pi}{\Rightarrow} x$$

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  $AG(G, \text{ação})$  o autômato analisador  $R^*S$  geral construído a partir de  $G$ . Dizemos que  $AG(G)$  é determinístico se  $|\text{ação}(\varphi, u)| \leq 1$ .

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  $H \in \text{ação}(\$S, \$^k)$ .

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

$$A \xrightarrow[S]{\sigma_1} B, \quad A \xrightarrow[S]{\sigma_2} B, \quad \sigma_1 \neq \sigma_2$$

Seja a derivação

$$S' \$^k \Rightarrow^* \alpha A x \xrightarrow[S]{*} \alpha B x \Rightarrow \alpha \beta x$$

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

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

$$S' \$^k \xRightarrow{+} \gamma C w \xrightarrow[S]{*} \gamma D w \Rightarrow \gamma \delta w = \alpha A x, \quad v \equiv w, \quad u \equiv x$$

Temos  $|x| \geq |w|$  e  $|\gamma \delta v| \leq |\alpha A u|$  e  $\gamma \delta v \in CR^*S(D \rightarrow \delta)$

Por outro lado,  $\alpha A u = \gamma \delta v z \in CR^*S$  com  $z \in \Sigma^*$

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

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

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

Lema II.17: Se numa gramática G, tivermos

$$S' \$^k \Rightarrow^+ \alpha A u w \stackrel{\sigma B \rightarrow \beta}{\Rightarrow} \alpha \beta u w, \quad \sigma \in s(A, B)$$

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

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

Demonstração:

Caso(a): explicitando o último passo de derivação de  $S' \$^k \Rightarrow^+ \alpha A u w$ , temos:

$$S' \$^k \stackrel{*}{\Rightarrow} \Psi E v z \stackrel{\sigma_1 F \rightarrow \varphi}{\Rightarrow} \Psi \varphi v z = \alpha A u w \stackrel{\sigma_2 B \rightarrow \beta}{\Rightarrow} \alpha \beta u w,$$

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

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

Caso(b): da hipótese do Lema temos

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

$$A \Rightarrow_S^* B$$

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

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

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

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:  $\pi = s(A,B)B \rightarrow \beta$ , que abrevia  $\pi = \sigma B \rightarrow \beta$ ,  $s(A,B) = \{\sigma\}$ .

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.

$E \in \text{ação}(\varphi, u)$

$(X, s(A,B) B \rightarrow \beta) \in \text{ação}(\varphi, u)$ ,  $X = RE$  ou  $X = RNE$

Temos  $\varphi = \alpha\beta$ ,  $\alpha\beta u \in CR^*S(B \rightarrow \beta)$  e

$S' \$^k \Rightarrow^* \alpha A w \quad \begin{matrix} s(A,B) B \rightarrow \beta \\ \Rightarrow \end{matrix} \alpha \beta w$  com  $u = \text{FIRST}(w)$

Por outro lado, da ação de empilhamento temos:

$\alpha\beta u z \in CR^*S(D \rightarrow \delta)$  para  $z \neq \epsilon$  e  $\alpha\beta u z = \gamma\delta v$

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

Pela condição  $R*S(k)$

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

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

e portanto  $z = \epsilon!$  Contradição.

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

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

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

Temos

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

$$S' \$^k \Rightarrow^* \gamma Cx \Rightarrow \gamma \delta x$$

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

Pela condição  $R*S(k)$ ,

$$\alpha Ax = \gamma Cx$$

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

Então o conflito pode ser expresso por

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

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

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,

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

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

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  $(\varphi, u)$  e portanto  $\varphi = \$S$ ,  $u = \$^k$ .

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

De  $\alpha\beta = \varphi = \$S$ , temos  $\alpha = \$S$ ,  $\beta = \epsilon$ . (Com efeito, o caso  $\beta = S$  pode ser excluído porque  $B \rightarrow \beta$  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  $H \in \text{ação}(\varphi, u)$  nos leva a  $\varphi = \$S$ ,  $u = \$^k$ , e  $E \in \text{ação}(\varphi, u)$  obrigaria  $\varphi z = \$S\$^k \in CR^*S$  com  $z \neq \epsilon$ , 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

$$(D1) S'\$^k \Rightarrow^* \alpha A u w \stackrel{\pi_1}{\equiv} \alpha \beta u w$$

$$(D2) S'\$^k \Rightarrow^* \gamma C v x \stackrel{\pi_2}{\equiv} \gamma \delta v x = \alpha \beta u y$$

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

implica que  $\alpha A u y \neq \gamma C v x$  ou  $\pi_1 \neq \pi_2$

1. Suponhamos  $\pi_1 \neq \pi_2$ ,  $S'\$^k \Rightarrow^* \alpha A u w$  em 0 passos.

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

$$(D1) S'\$^k \stackrel{S' \rightarrow \$S}{\equiv} \$S\$^k$$

De (D2)  $S'\$^k \Rightarrow^* \alpha \beta u y = \$S\$^k y$  e  $y = \epsilon$ . Portanto

$$(D2) S'\$^k \Rightarrow^* \gamma C v z \stackrel{s(C, D) D \rightarrow \delta}{\equiv} \$S\$^k$$

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

Como  $\pi_1 \neq \pi_2$ ,  $S' \$^k \Rightarrow^+ \gamma C v x$  e, pelo Lema II.17

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

2.  $\pi_1 \neq \pi_2$ ,  $S' \$^k \Rightarrow^* \gamma C v x$  em 0 passos. A prova se dá de maneira semelhante ao caso anterior.

3. Suponha  $\pi_1 \neq \pi_2$ ,  $S' \$^k \Rightarrow^+ \alpha A u w$ ,  $S' \$^k \Rightarrow^+ \gamma C v x$ .

Temos

$$(D1) S' \$^k \Rightarrow^+ \alpha A u w \stackrel{\pi_1}{\Rightarrow} \alpha \beta u w$$

$$(D2) S' \$^k \Rightarrow^+ \gamma C v x \stackrel{\pi_2}{\Rightarrow} \gamma \delta v x = \alpha \beta u y$$

Pela Lema II.17

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

$(Y, \pi_2) \in \text{ação}(\gamma \delta, v)$  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,  $(X, \pi_1)$  e  $(Y, \pi_2)$  pertencem a  $\text{ação}(\alpha \beta, u) = \text{ação}(\gamma \delta, v)$  e  $AG(G)$  não é determinístico, já que  $\pi_1 \neq \pi_2$ .

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

Temos, de  $(Y, \pi_2) \in \text{ação}(\gamma \delta, v)$ , que  $\gamma \delta v \in CR^*S(D \rightarrow \delta)$ , e

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

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  $\alpha A u y \neq \gamma C v x$

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

$$\gamma\delta v x = \gamma\beta v x = \alpha\beta u y$$

$$\gamma C v x = \gamma A v x \neq \alpha A u y$$

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

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

$$|\gamma| < |\alpha| \quad \text{e} \quad |\gamma\delta| < |\alpha\beta|. \quad \text{Ver 3.3.}$$

$$|\gamma| > |\alpha|. \quad \text{Ver caso 3.2.}$$

Definição II.19: 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.

Teorema II.20: 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  $(\varphi, ux, \pi)$  uma configuração qualquer de AG. a qual supomos seja tal que  $\varphi$  forme um prefixo correto de G. Então, qualquer mudan-

ça 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( $\varphi, u$ ) é porque  $\varphi ux \in CR^*S$  para  $x \neq \epsilon$  e portanto  $\varphi a$  é um prefixo correto de G para  $ay = ux$ .

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

c) se redução com empilhamento  $\in$  ação( $\alpha\beta, u$ ) é porque  $\alpha Aux \in CR^*S$  para  $x \neq \epsilon$  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  $(\$, S\$, \epsilon)$  já contém na pilha um prefixo correto,  $\$,$  concluímos que toda configuração  $(\varphi, y, \pi)$  tal que  $(\$, x\$, \epsilon) \vdash^* (\varphi, y, \pi)$  deixa um prefixo correto na pilha.

Observação: No caso  $k = 0$ , a cadeia lookahead é sempre  $\epsilon$ . Quando ação( $\varphi, \epsilon x, \pi$ )  $\ni$  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  $L(AG) = L(G)$ .

Teorema II.21: Seja G satisfazendo a condição  $R^*S$ . Seja AG um autômato analisador  $R^*S$  geral construído a partir de G. Então

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

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

$$(\$ , xy, \epsilon) \stackrel{m}{\vdash} (\varphi, y, \pi) \text{ implica } \varphi y \stackrel{\pi}{\Rightarrow} \$xy$$

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

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

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

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

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

$$|\text{SPNS}(\pi)| \text{ reduções}$$

$$|w| \text{ empilhamentos}$$

ou seja

$$\begin{aligned} j &\leq |\text{SPNS}(\pi)| + |w| \leq |\pi| + |w| \leq \\ &\leq c_1 |w| + |w| \leq c_2 |w| \end{aligned}$$

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

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

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

Então existe  $z \in \Sigma^*$  tal que

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

Demonstração:

Visto que  $(\$ , xy\$^k , \epsilon) \vdash^* (\varphi , y\$^k , \pi)$  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

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

cqd

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

$$\text{ou } (\$ , x\$^k , \epsilon) \vdash (\$S , \$^k , \pi)$$

$$\text{ou } (\$ , x\$^k , \epsilon) \vdash^* (\varphi , uw , \pi) \text{ com ação } (\varphi , u) = \emptyset$$

$$\text{e } \varphi uw \neq \$S\$^k$$

(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

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

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

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

Nestas condições:

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

$$\varphi u = \psi a$$

$$\psi \xrightarrow{\pi} \$y$$

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

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

y	b	z''	$\in L(G)$
---	---	-----	------------

y	a	z	$\notin L(G)$
---	---	---	---------------

As mudanças de configuração dessas duas sentenças podem ser descritas por

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

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

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  $C_2 \vdash (\varphi , uz^k \$^k , \pi)$  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  $(\varphi , uz^k \$^k , \pi)$  o autômato tem ação  $(\varphi , u) = \emptyset$ , portanto pára e não aceita  $x$ .

No caso de  $yz^k$ , 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,  $\varphi$  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.

Definição III.1: chamaremos de estado qualquer conjunto não vazio de itens. Dependendo da classe de itens considerada, falaremos em estado LR(k), R\*S(k), etc.

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

$$\delta_c(q, X) = \text{fecho} (\{ [A \rightarrow \alpha X \beta, u] \mid [A \rightarrow \alpha \cdot X \beta, u] \in q \wedge [A \rightarrow \alpha X \beta, u] \text{ é um item } R^*S(k) \})$$

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.  $q_{oc} \in C$  onde  $q_{oc} = \text{fecho}(\{ [S' \rightarrow \$ \cdot S, \$^k] \})$
2. Se  $p \in C$  e se  $\delta_c(p, X) \downarrow$  então  $\delta_c(p, X) \in C$  para  $X \in N \cup \Sigma$ .

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

1.  $v_c(\$) = q_{oc}$   
 $v_c(\varphi X) = \delta_c(v_c(\varphi), X)$
2.  $\lambda_c(\$) = q_{oc}$   
 $\lambda_c(\varphi X) = \lambda_c(\varphi) \delta_c(v_c(\varphi), X)$

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

Notas:

- se estendermos, como habitualmente,  $\delta_c$  aceitando cadeias em seu segundo argumento, temos

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

- $v_c(\varphi)$  é sempre o último estado de  $\lambda_c(\varphi)$
- se  $v_c(\varphi_1\varphi_2) \downarrow$  então  $v_c(\varphi_1) \downarrow$  (para  $\varphi_1 \neq \epsilon$ )
- se  $\lambda_c(\varphi_1\varphi_2) \downarrow$  então  $\lambda_c(\varphi_1) \downarrow$  (para  $\varphi_1 \neq \epsilon$ )
- como cada estado da coleção canônica é alcançável através de um único símbolo, a função (parcial)  $\lambda_c$  é 1:1.

Teorema III.5: Um item  $R^*S [A \rightarrow \beta., u] \in v_c(\varphi)$  sse  $\varphi u \in CR^*S(A \rightarrow \beta)$ .

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

Lema III.6: Se  $[A \rightarrow \beta.\beta, u] \in v_c(\varphi)$  e  $A \rightarrow \alpha.\beta$  é uma regra não simples então  $\varphi\beta u \in CR^*S(A \rightarrow \alpha\beta)$ .

Demonstração: por indução no comprimento de  $\varphi$

Base da indução

1) para  $\varphi = \$$ ,  $v_c(\varphi) = q_{0c}$

$[S' \rightarrow \$ . S, \$^k] \in q_{0c} \quad \therefore \varphi = \$, \beta = S, u = \$^k$

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

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

2) Vamos supor que um item da forma  $[A \rightarrow \beta.u] \in q_{0c}$ .

Note que os itens iniciais de  $q_{0c}$  (só há um) foram considerados acima.

Então os seguintes itens estão em  $q_{0c}$

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

$$[A_1 \rightarrow \cdot A_2 \beta_1, u_1] \quad u_1 = \$^k, \quad A_1 = S$$

$$[A_2 \rightarrow \cdot A_3 \beta_2, u_2] \quad u_2 \in \text{FIRST}(\beta_1 u_1)$$

\cdot  
\cdot  
\cdot

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

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

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

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

Então (as derivações não precisam ser mais à direita)

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

$$SS\$^k \Rightarrow^* \$Auw \Rightarrow \$\beta uw$$

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

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

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

E podemos escrever

$$SS\$^k \Rightarrow^* \$A_j uw \Rightarrow \$\beta uw$$

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

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

Temos

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

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

Por hipótese, o item  $[A \rightarrow \alpha. \beta, u] \in q = v_c(\varphi)$ .

Vamos considerar os seguintes casos:

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

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

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

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

ou seja, que

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

Caso 2:  $\alpha = \epsilon$

Então o item  $[A \rightarrow. \beta, u]$  apareceu na operação fecho( $q$ ).  
Então encontraremos em  $q$  itens da forma

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

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

·  
·  
·

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

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

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

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

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

.

.

.

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

Como em  $p$  aparece o item  $[C \rightarrow \gamma \cdot X B_1 \delta, v]$ , pela hipótese de indução

$$\varphi B_1 \delta v = \psi X B_1 \delta v \in \text{CR}^* S(C \rightarrow \gamma X B_1 \delta)$$

ou seja

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

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

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

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

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

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

ou

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

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

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

Demonstração: será feita por indução no comprimento de  $\varphi$ .

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

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

Caso 1:  $\alpha = \epsilon$

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

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

$$A_1 \rightarrow A_2\beta_1 \quad (\text{com } S = A_1)$$

$$A_2 \rightarrow A_3\beta_2$$

.

.

.

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

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

Desta forma

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

onde

$$\beta_i \Rightarrow^* w_i \quad \text{e} \quad uw = w_{n-1} \dots w_1 \$^k$$

Seja

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

Então

$$[A_1 \rightarrow \cdot A_2 \beta_1, u_1] \in q_{oc}$$

.

.

.

$$[A_n \rightarrow \cdot \beta_n, u_n] \in q_{oc}$$

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

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

Passo de indução:

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

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

Por hipótese  $\varphi \beta u \in CR^*S(A \rightarrow \alpha \beta)$

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

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

Pela hipótese de indução, como  $\psi X \beta u \in CR^*S(A \rightarrow \alpha \beta)$  temos  $[A \rightarrow \rho \cdot X \beta, u] \in v_c(\psi) = p$  e portanto  $[A \rightarrow \rho X \cdot \beta, u] \in \delta_c(p, X) = q$

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

Caso 2:

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

Se  $\varphi \beta u \in CR^*S(A \rightarrow \beta)$ , então basta provar que

$$[A \rightarrow \cdot \beta, u] \in v_c(\varphi)$$

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  $B \gamma_2^x \Rightarrow^* A w$  pelas seguintes produções

$$A_1 \rightarrow A_2 \beta_1 \quad (\text{com } B = A_1)$$

$$\vdots$$

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

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

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

de tal forma que

$$\beta_i \Rightarrow^* w_i$$

$$\gamma_2 \Rightarrow^* y_2$$

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

Ora, pela hipótese de indução

$$[C \rightarrow \gamma_1 \cdot X B \gamma_2, v] \in p = v_c(\psi)$$

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

Vamos definir agora

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

então

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

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

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

ou seja,

$$[A \rightarrow \cdot \beta, u] \in q$$

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 itens, que reproduza fielmente as mudanças de configuração do autômato R\*S geral. Nossa tarefa fica facilitada pelo Teorema III.5 que diz que itens 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
- $q_{oc} = \text{fecho}(\{[S' \rightarrow \$ \cdot S, \$^k]\})$  é o estado inicial
- $q_{fc} = \delta_c(q_{oc}, S)$  é o estado final, isto é, o que contém o item final  $[S' \rightarrow \$ \cdot S, \$^k]$ .
- a função  $e_c: Q_c \times \Sigma^k \rightarrow Q_c$  é definida por  $e_c(v_c(\varphi), u) = v_c(\varphi a)$  onde  $u = ay$  e  $E \in \text{ação}(\varphi, u)$  - ver Teorema III.9.
- a relação  $d_c: Q_c \times \Sigma^k \rightarrow Z$  (onde Z representa o conjunto dos inteiros não negativos) definida como a menor relação tal que:  
  
se  $(X, \sigma B \rightarrow \beta) \in \text{ação}(\alpha\beta, u)$ ,  $X \in \{RE, RNE\}$ ,  
então  $|\beta| \in d_c(v_c(\alpha\beta), u)$
- a relação  $r_c: Q_c \times \Sigma^k \times Q_c \rightarrow (Q_c \cup (Q_c \times Q_c)) \times P^*$  definida como a menor relação tal que

1. Se  $(RE, \sigma B \rightarrow \beta) \in \text{ação}(\alpha\beta, u)$ ,  $\sigma \in s(A, B)$ ,  $u = ay$   
então  $((v_c(\alpha A), v_c(\alpha Aa)), \alpha B \rightarrow \beta) \in$   
 $\in r_c(v_c(\alpha), u, v_c(\alpha\beta))$
2. se  $(RNE, \sigma B \rightarrow \beta) \in \text{ação}(\alpha\beta, u)$ ,  $\sigma \in s(A, B)$   
então  $(v_c(\alpha A), \sigma B \rightarrow \beta) \in r_c(v_c(\alpha), u, v_c(\alpha\beta))$

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  $G$ .

Teorema III.9: Nas condições da Definição III.8,  $e_c$  é uma função bem definida.

Demonstração: cumpre provar que a escolha de  $\varphi$ , na definição III.8,

- (a) não altera o valor atribuído a  $e_c(v_c(\varphi), u)$
- (b) não altera a presença ou não da ação de empilhamento em ação  $(\varphi, u)$

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

Para provarmos (a), observamos que

$$\begin{aligned} e_c(v_c(\varphi_1), u) &= v_c(\varphi_1 a) = \delta_c(v_c(\varphi_1), a) = \delta_c(q, a) = \\ &= \delta_c(v_c(\varphi_2), a) = v_c(\varphi_2 a) = e_c(v_c(\varphi_2), u) \end{aligned}$$

Para provarmos (b), suporemos que  $E \in \text{ação}(\varphi_1, u)$ . Então  $\varphi_1 u y \in CR^*S(B \rightarrow \beta)$  com  $y \neq \epsilon$  para alguma regra  $B \rightarrow \beta$

$$\varphi_1 u y = \varphi_1 z v = \alpha_1 \beta v$$

Portanto  $[B \rightarrow \beta, v] \in v_c(\alpha_1 \beta) = v_c(\varphi_1 z) = v_c(\varphi_2 z)$  e  $\varphi_2 u y = \varphi_2 z v = \alpha_2 \beta v$ .

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

Teorema III.10: 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:

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

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

$$(X_1, \sigma_1 \ B_1 \rightarrow \beta_1) \in ação(\alpha_1 \beta_1, u)$$

$$(X_2, \sigma_2 \ B_2 \rightarrow \beta_2) \in ação(\alpha_2 \beta_2, u)$$

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

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

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

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

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

$$(X_3, \sigma_3 \ B_2 \rightarrow \beta_2) \in ação(\alpha_3 \beta_2, u) = ação(\alpha_1 \beta_1, u)$$

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

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

$$\begin{aligned} (b_1) \quad & ((r_1, s_1), \pi_1) \in r_c(q, u, p) \\ & ((r_2, s_2), \pi_2) \in r_c(q, u, p) \\ & \pi_1 \neq \pi_2 \end{aligned}$$

Temos, escrevendo  $u = ay$ , que:

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

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

$$r_1 = v_c(\alpha_1 A_1)$$

$$r_2 = v_c(\alpha_2 A_2)$$

$$s_1 = v_c(\alpha_1 A_1 a)$$

$$s_2 = v_c(\alpha_2 A_2 a)$$

$$\pi_1 = \sigma_1 B_1 \rightarrow \beta_1$$

$$\pi_2 = \pi_2 B_2 \rightarrow \beta_2$$

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

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

$$(RE, \sigma_2 B_2 \rightarrow \beta_2) \in \text{ação}(\alpha_2\beta_2, u)$$

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

$$[B_2 \rightarrow \beta_2 \cdot, u] \in v_c(\alpha_2\beta_2) = p = v_c(\alpha_1\beta_1)$$

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

De  $(RE, \sigma_2 B_2 \rightarrow \beta_2) \in \text{ação}(\alpha_2\beta, u)$  temos também  $\alpha_2 A_2 uz \in CR^*S$  com  $z \neq \epsilon$  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  $z \neq \epsilon$ . A derivação acima tem um ou mais passos e podemos explicitar o passo que introduz  $A_2$ :

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

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

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

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

Temos

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

$$e \quad [D \rightarrow \xi A_2 \gamma, v] \in v_c(\rho_2 \xi) = v_c(\alpha_2) = v_c(\alpha_1) = q$$

$$\text{Logo } \alpha_1 = \rho_1 \xi \quad e \\ \rho_1 \xi A_2 \gamma v \in CR^*S(D \rightarrow \xi A_2 \gamma).$$

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,

$$[B_{2 \rightarrow \beta}, u] \in v_c(\alpha_1 \beta),$$

$$\alpha_1 \beta u \in CR^*S(B_{2 \rightarrow \beta}),$$

$$A_2 \xrightarrow[S]{\sigma_2} B_2$$

$$\pi_2 = \sigma_2 B_{2 \rightarrow \beta}$$

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.

$$\begin{aligned} (b3) \quad & (r_1, \pi_1) \in r_c(q, u, p) \\ & (r_2, \pi_2) \in r_c(q, u, p) \\ & \pi_1 \neq \pi_2 \end{aligned}$$

Sendo  $u = ay$ , temos

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

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

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

$$\pi_1 = \sigma_1 \quad B_1 \rightarrow \beta_1 \qquad \pi_2 = \sigma_2 \quad B_2 \rightarrow \beta_2$$

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

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

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

$$\text{Se } |\beta_1| = |\beta_2| \text{ então } \beta_1 = \beta_2 = \beta.$$

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

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

e portanto

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

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

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

$$\alpha_2 = \varepsilon$$

$$S' = A_2$$

$$u = \$^k$$

$$y_2 = \epsilon$$

$$B_2 = S'$$

$$\beta = \$S$$

$$\alpha_2 \beta = \$S$$

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

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

Neste caso, explicitando o passo que introduz  $A_2$  temos:

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

com  $\pi = \sigma D \rightarrow \xi A_2 \gamma$ ,  $\sigma \in s(C, D)$ ,  $\alpha_2 = \rho_2 \xi$ ,  $\gamma \equiv^* w$ ,  $u \in \text{FIRST}(\gamma v)$ ,  $w v x_2 = u y_2$ .

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

$\alpha_2 A_2 u \in CR^*S(D \rightarrow \xi A_2 \gamma)$  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  $\gamma$ . Temos  $u y_2 = w v x_2$  e que  $|u| \leq |wv|$ , necessariamente.

Temos

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

$$[D \rightarrow \xi \cdot A_2 \gamma, v] \in v_c(\rho_2 \xi) = v_c(\alpha_2) = v_c(\alpha_1) = q$$

Logo  $\alpha_1 = \rho_1 \xi$  e

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

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

$$[B_{2 \rightarrow \beta}, u] \in v_c(\alpha_1 \beta)$$

$$\alpha_1 \beta u \in CR^*S(B_{2 \rightarrow \beta})$$

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

$$\pi_2 = \sigma_2 B_{2 \rightarrow \beta}$$

e portanto  $(RNE, \pi_2) \in \text{ação}(\alpha_1 \beta, u)$ , conflitando com  $(RNE, \pi_1) \in \text{ação}(\alpha_1 \beta, u)$ , 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  $\delta_c$  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  
     $[B \rightarrow \beta., a] \in p$   
     $[D \rightarrow \alpha.a\gamma, b] \in p$

(b) existe  $p \in Q_c$  tal que  
     $[B_1 \rightarrow \gamma\beta., a] \in p$   
     $[B_2 \rightarrow \beta., a] \in p, \quad \gamma \neq \epsilon$

(c) existem  $p, q \in Q_c$  tais que  
    (1)  $[B_1 \rightarrow \beta., a] \in p$   
         $[B_2 \rightarrow \beta., a] \in p$   
    (2)  $\delta_c(q, \beta) = p$

(3)  $A_1 \xrightarrow{\sigma_1} B_1$   
     $A_2 \xrightarrow{\sigma_2} B_2$

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

(4)  $[D_1 \rightarrow \xi_1.A_1\gamma_1, b] \in q \quad \xi_1\gamma_1 \neq \epsilon$   
     $[D_2 \rightarrow \xi_2.A_2\gamma_2, c] \in q \quad \xi_2\gamma_2 \neq \epsilon$

$$\begin{aligned} a &\in \text{FIRST}(\gamma_1 b) \\ a &\in \text{FIRST}(\gamma_2 c) \end{aligned}$$

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

Teorema IV.2: 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.

Lema IV.3: 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:  $H \in \text{ação}(\varphi, a)$   
 $E \in \text{ação}(\varphi, a)$

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

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

Temos  $\varphi a = \$S\$$   
 $\$S\$ \in CR^*S (B \rightarrow \beta)$

e pelo Lema III.7

$$[B \rightarrow \beta., \$] \in v_c(\$S)$$

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

Caso 3:  $E \in \text{ação}(\varphi, a)$

$(X, \pi) \in \text{ação}(\varphi, a)$ ,  $X \in \{RE, RNE\}$ ,  $\pi = \sigma B \rightarrow \beta$ ,  $\sigma \in s(A, B)$ .

De  $E \in \text{ação}(\varphi, a)$  temos que  $\varphi az \in CR^*S$ ,  $z \neq \epsilon$ .

$$S' \$^k \equiv >^+ \varphi azw$$

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

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

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

Temos

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

e pelo Lema III.7.

$$[D \rightarrow \alpha . a \gamma, b] \in v_c(\varphi)$$

Por outro lado, de  $(X, \pi) \in \text{ação}(\varphi, a)$ ,

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

e pelo Lema III.7,

$$[B \rightarrow \beta . , a] \in v_c(\varphi)$$

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

Caso 4:  $(X, \pi_1) \in \text{ação}(\varphi, a)$

$(Y, \pi_2) \in \text{ação}(\varphi, a)$ ,  $X, Y \in \{RE, RNE\}$

$\pi_1 = \sigma_1 B_1 \rightarrow \beta_1$ ,  $\sigma_1 \in s(A_1, B_1)$

$\pi_2 = \sigma_2 B_2 \rightarrow \beta_2$ ,  $\sigma_2 \in s(A_2, B_2)$

$\beta_1 \neq \beta_2$

Temos:

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

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

$$\text{Logo } \varphi = \alpha_1 \beta_1 = \alpha_2 \beta_2$$

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  $\alpha_1\gamma\beta a \in CR^*S (B_1 \rightarrow \gamma\beta)$  e  $\alpha_1\gamma\beta a \in CR^*S (B_2 \rightarrow \beta)$ , pelo Lema III.7, temos

$$[B_1 \rightarrow \gamma\beta., a] \in v_c(\varphi)$$

$$[B_2 \rightarrow \beta., a] \in v_c(\varphi)$$

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

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

Temos:

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

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

$$\varphi = \alpha\beta$$

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  $z \neq \epsilon$ ,  $E \in \text{ação}(\alpha A, a)$ .

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

$$S'\$ \Rightarrow^+ \alpha Aaw$$

Há dois casos a considerar:

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

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

Pelo Lema II.17

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

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

Caso 5.2:  $\pi_1 \neq \pi_2$

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

$$[B_1 \rightarrow \beta., a] \in v_c(\varphi)$$

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

$$[B_2 \rightarrow \beta., a] \in v_c(\varphi).$$

Da definição de  $v_c$ ,

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

Temos também

$$A_1 \xRightarrow[S]{\sigma_1} B_1$$

$$A_2 \xrightarrow[S]{\sigma_2} B_2$$

$$\pi_1 = \sigma_1 B_1 \rightarrow \beta \quad \pi_2 = \sigma_2 B_2 \rightarrow \beta$$

Por outro lado, de

$$(X, \pi_1) \in \text{ação}(\varphi, a),$$

$$\pi_1 = \sigma_1 B_1 \rightarrow \beta,$$

$$\varphi = \alpha\beta,$$

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

$$S' \$^k \Rightarrow^+ \alpha A_1 a z w$$

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

$$S' \$ \Rightarrow^* \psi_1 C_1 b y \Rightarrow \psi_1 \xi_1 A_1 \gamma_1 b y \Rightarrow^* \alpha A_1 a z w$$

com  $C_1 \xrightarrow[S]{*} D_1$ ,  $\alpha = \psi_1 \xi_1$  e  $\gamma_1 b y \Rightarrow^* a z w$ , de forma que  $a \in \text{FIRST}(\gamma_1 b)$

Temos

$$\alpha A_1 \gamma_1 b \in CR^*S (D_1 \rightarrow \xi_1 A_1 \gamma_1)$$

de forma que  $[D_1 \rightarrow \xi_1 \cdot A_1 \gamma_1, b) \in v_c(\alpha)$ .

De forma semelhante

$$[D_2 \rightarrow \xi_2 \cdot A_2 \gamma_2, c] \in v_c(\alpha)$$

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

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

Lema IV.4: 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  $p \in Q_c \mid [B \rightarrow \beta, a] \in p, [D \rightarrow \alpha.a\gamma, b] \in p.$

Seja  $\varphi$  tal que  $v_c(\varphi) = p.$

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

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

ou seja

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

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

Devemos considerar dois casos:

$$(1) S' \$ = \alpha A a w, \pi = S' \rightarrow \$ S$$

$$\text{Temos } H \in \text{ação}(\$S, \$) = \text{ação}(\varphi, a)$$

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

Pelo Lema II.17 (b),  $(X, \pi) \in \text{ação}(\alpha \beta, a) = \text{ação}(\varphi, a)$

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

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

$$S' \$ \equiv >^* \psi A_1 c w_1 \equiv \psi \beta_1 c w_1 = \varphi a y b x$$

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

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  $p \in Q_c \mid [B_1 \rightarrow \gamma \beta., a] \in p, \gamma \neq \epsilon, \text{ e } [B_2 \rightarrow \beta., a] \in p$ .

$$\text{Seja } p = v_c(\varphi)$$

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

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

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

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

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

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

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

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

Temos também

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

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

Se tivermos  $S'\$ = \alpha A_1 a w_1, \pi_1 = S' \rightarrow \$ S$  e  $H \in \text{ação}(\varphi, a).$   
 Caso contrário,

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

e pelo Lema II.17,  $(Y, \pi_1) \in \text{ação}(\varphi, a).$

Em qualquer das situações acima, como  $\gamma \neq \epsilon$  implica em  $\pi_1 \neq \pi_2, \text{ação}(\varphi, a)$  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

$$(c1) [B_1 \rightarrow \beta, a] \in p$$

$$[B_2 \rightarrow \beta, a] \in p$$

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

$$(c3) A_1 \xrightarrow{\sigma_1} B_1$$

$$A_2 \xrightarrow{\sigma_2} B_2$$

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

$$(c4) [D_1 \rightarrow \xi_1 \cdot A_1 \gamma_1, b] \in q, \xi_1 \gamma_1 \neq \epsilon$$

$$[D_2 \rightarrow \xi_2 \cdot A_2 \gamma_2, c] \in q, \xi_2 \gamma_2 \neq \epsilon$$

$$a \in \text{FIRST}(\gamma_1 b)$$

$$a \in \text{FIRST}(\gamma_2 c)$$

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

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

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

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

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

Podemos então escrever que

$$\begin{aligned} S'\$ \equiv >^* \psi_1 C_1 b y_1 \equiv > \psi_1 \xi_1 A_1 \gamma_1 b y_1 = \alpha A_1 \gamma_1 b y_1 \equiv >^* \\ \equiv >^* \alpha A_1 a x_1 \stackrel{\pi_1}{\equiv} > \alpha \beta a x_1 = \varphi a x_1 \quad \text{já que } a \in \text{FIRST}(\gamma_1 b) \end{aligned}$$

Pelo Lema II.17,

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

De forma semelhante

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

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,  $v_c(\varphi) = q_{fC} \Rightarrow \varphi = \$S$ .

Se  $[B \rightarrow \beta., \$] \in v_c(\$S)$  então  $\$S = \alpha\beta$ ,  $\$\$S \in CR^*S(B \rightarrow \beta)$

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

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

Caso 4.3:  $\alpha = \$S$ ,  $\beta = \epsilon$ . É o único caso possível. Note que  $[B \rightarrow \cdot, \$] \in v_c(\$S)$  implica, pelo Lema III.6 que  $\$S\$ \in CR^*S$  ( $B \rightarrow \epsilon$ ) e portanto

$$S' \$ \Rightarrow^+ \$SA\$ \xRightarrow{\pi} \$S\$$$

com  $\pi = \sigma B \rightarrow \epsilon$ ,  $\sigma \in s(A, B)$ . Pelo Lema II.17,  $(X, \pi) \in \text{ação}(\$S, \$)$  o que mostra, conjuntamente com  $H \in \text{ação}(\$S, \$)$  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  $\delta_c$  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  $\delta_c$ .

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

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 ≠ ∅ 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 algoritmos 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

$$Q_e = \{p \in Q_c \mid ([A \rightarrow \alpha a. \beta, b] \in p) \wedge ([B \rightarrow \gamma. c\eta, d] \in p)\}$$

Algoritmo IV.7: construção da função  $e_c$

Entrada: conjunto  $Q_c$  de estados do analisador AC e a função  $\delta_c$ ,

Saída: a função  $e_c$

Obs:  $e_c : Q_e \times \Sigma \rightarrow Q_c$

Algoritmo

```
para todo  $p \in Q_e$ 
  faça
    para todo  $a \in \Sigma$ 
      faça
        se  $\delta_c(p, a) \neq \emptyset$ 
          então  $e_c(p, a) := \delta_c(p, a)$ 
```

Observação: por construção,  $\delta_c$  é uma função,  $e_c$  é uma restrição de  $\delta_c$  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

L.1 para todo  $p \in Q_c$   
L.2 faça  
L.3 para todo  $[A \rightarrow \alpha., a]$  em  $p$   
L.4 faça  
L.5 se  $d_c(p, a) \uparrow$  ou  $d_c(p, a) = |\alpha|$   
L.6 então  $d_c(p, a) := |\alpha|$   
L.7 senão mensagem ("G não é R\*S (1)");  
L.8 se  $e_c(p, a) \uparrow$  e  $d_c(p, a) \uparrow$   
L.9 então mensagem ("G não é R\*S (1)")  
L.10 fim-do-faça  
L.11 fim-do-faça

Algoritmo IV.9: construção da função  $r_c$

Entrada: o conjunto  $Q_c$  de estados,  
a função de transição  $\delta_c$   
a função  $d_c$ ,  
os pares de não terminais  $(A, B)$  tais que  $A \xrightarrow[S]{*} B$

Saída: a função  $r_c$  ou a informação de que G não é  $E*S (1)$

Algoritmo:

L.1 para todo  $q \in Q_c$  tal que  $[B \rightarrow \beta, a] \in q, \beta \notin N$   
L.2 faça  
L.3  $p := \delta_c(q, \beta)$ ;  
L.4 para todo não-terminal  $A$  tal que  $[E \rightarrow \alpha.A\gamma, b] \in q$   
L.5 e tal que  $A \xrightarrow[S]{*} B$  e  $a \in \text{FIRST}(\gamma b)$   
L.6 faça  
L.7 se  $\delta_c(q, Aa) \uparrow$  e  $r_c(q, a, p) \uparrow$   
L.8 então (para  $p_1 = \delta_c(q, A)$  e  $p_2 = \delta_c(q, Aa)$ )  
L.9  $r_c(q, a, p) := ((p_1, p_2), s(A, B) B \rightarrow \beta)$   
L.10 senão se  $\delta_c(q, A) \uparrow$  e  $r_c(q, a, p) \uparrow$   
L.11 então (para  $p_1 = \delta_c(q, A)$ )  
L.12  $r_c(q, a, p) := (p_1, s(A, B) B \rightarrow \beta)$



## CAPÍTULO V

### CONCLUSÕES E OBSERVAÇÕES

Vimos que as gramáticas livres de contexto admitem autômatos analisadores que, a cada mudança de configuração por redução, executam a redução por todas as produções simples que se seguem a uma redução por produção não simples.

O método proposto difere dos que tratam da eliminação das produções simples (AHO/ULLMAN (3), ANDERSON/EVE/HORING (5), BACKHOUSE (6), DEMERS (7), DENKER/DURRE/HEUFT (8), HEILBRUNNER (12), JOLIAT (14), JOLIAT (15), KOSKIMIES/SOISALON-SOININEN (16), LALONDE (17), PAGER (18), PAGER (19), PAGER (20), SOISALON-SOININEN (21), SOISALON-SOININEN (22), TOKUDA (23)) sendo aplicável a qualquer gramática LR(k), mantendo a linearidade na complexidade de no tempo e a capacidade de se atrelar ações semânticas mesmo às produções simples.

Assim como no método LR, o método R\*S admite que sejam definidas variantes "lookahead" e "simples". Essas variantes se mostram de grande interesse visto produzirem autômatos de tamanho mais reduzido como também com grande capacidade de reconhecimento de linguagens.

No que se refere ao tamanho e velocidade do analisador, o método R\*S forneceu bons resultados. Como o método não faz uso de itens simples completos, estados compostos unicamente desses itens podem existir na análise LR mas inexistem na análise R\*S. Assim os autômatos R\*S têm, em geral, um número menor de estados. Exploramos este tema construindo um gerador R\*S "simples" e comparamos as tabelas das funções de análise R\*S com a tabela SLR não compactada. Nos exemplos testados, as tabelas R\*S tinham dimensões menores que as R\*S simples.

Para a continuidade deste trabalho sugerimos (e nos propomos a) dar tratamento formal à análise R\*S "lookahead" e

"simples", estudar a relação entre essas variantes e as análises LALR e SLR, as classes de gramáticas reconhecidas por esses métodos, a geração de algoritmos de recuperação de erros sintáticos adequados à análise R\*S e estudo de métodos de compactação das tabelas de análise.

## REFERÊNCIAS BIBLIOGRÁFICAS

- (1) AHO, A. V. e ULLMAN, J.D. ; The Theory of Parsing, Translation and Compiling. Volume I: Parsing; Prentice-Hall, Inc.; 1972
- (2) AHO, A. V. e ULLMAN, J.D. ; The Theory of Parsing, Translation and Compiling. Volume II: Compiling; Prentice-Hall, Inc.; 1973
- (3) AHO, A.V. e ULLMAN, J.D. ; A Technique for Speeding Up LR(k) Parsers; SIAM Journal of Computing 2:2, 106-127, 1973.
- (4) AHO, A. V. e ULLMAN, J.D. ; Principles of Compiler Design; Addison-Wesley Publishing Co.; 1977.
- (5) ANDERSON, T., EVE, J. e HORNING, J.J. ; Efficient LR(1) Parsers; Acta Informatica 2, 12-39, 1973
- (6) BACKHOUSE, R.C. ; An Alternative Approach to the Improvement of LR Parsers; Acta Informatica 6:3, 277-296, 1976.
- (7) DEMERS, A.J. ; Elimination of Single Productions and Merging of Nonterminal Symbols in LR(1) Grammars; Journal of Computer Languages 1:2, 105-119, 1975.
- (8) DENCKER, PETER; DURRE, KARL e HEUFT, JOHANNES ; Optimization of Parser Tables for Portable Compilers; ACM Transactions on Programming Languages and Systems, Vol. 6, nº 4, Out/84, pag. 546-572.

- (9) DEREMER, F.L. ; Practical Translators for LR(k) Languages ; Tese de Doutorado, M.I.T., Cambridge, Mass, 1969.
- (10) HEILBRUNNER, STEPHAN; Using Item Grammars to Prove LR(k) Theorems; Bericht nº 7701; Hochschule der Bundeswehr, Munchen, 1977.
- (11) HEILBRUNNER, STEPHAN; A parsing Approach to LR Theory; Theoretical Computer Science 15, p. 117-157; North Holland Publishing Company, 1981.
- (12) HEILBRUNNER, STEPHAN; Practical Conditions for Correct Elimination of Chain Reductions from LR Parsers; Bericht nº 8307; Hochschule de Bundeswehr Munchen, 1983.
- (13) HOPCROFT, J.E. e ULLMAN, J.D. ; Formal Languages and Their Relation to Automata, Addison-Wesley, Reading, Mass., 1969.
- (14) JOLIAT, M.J. ; A Practical Minimization of LR(k) Parser Tables; Proceedings of the IFIP Congress 1984; North Holland, Amsterdam, 376-380, 1974.
- (15) JOLIAT, M.J. ; A Simple Technique for Partial Elimination of Unit Productions from LR(k) Parsers; IEEE Transactions on Computers, C-27, 763-764, 1976.
- (16) KNUTH, D.E. ; On the Translation of Languages from Left to right; Information and Control 8:6, 607-639, 1965.

- (16) KOSKIMIES, KAI e SOISALON-SOININEN. . ELJAS; On a Method for Optimizing LR Parsers; International Journal of Computer Mathematics, Section A, Vol. 7, 287-295, 1979.
- (17) LALONDE, WILF R.; On Directly Constructing LR(k) Parsers without Chain Reductions; Conference Record of the Third ACM Symposium on Principles of Programming Languages, Atlanta, Georgia, 127-133, 1976.
- (18) PAGER, DAVID; On Eliminating Unit Productions from LR(k) Parsers; Automata, Languages and Programming; (J. Loecks, editor); Springer Verlag, NY, NY 1974.
- (19) PAGER, DAVID; A Practical Method for Constructing LR(k) Parsers; Acta Informatica 7, 249-268, 1977.
- (20) PAGER, DAVID; Eliminating Unit Productions from LR(k) Parsers; Acta Informatica 9, 31-59, 1977.
- (21) SOISALON-SOININEN, ELJAS; Elimination of Single Productions from LR Parsers in Conjunction with the use of Default Reductions; Proceedings of the Fourth ACM Symposium on Principles of Programming Languages; 183-193, 1977.
- (22) SOISALON-SOININEN, ELJAS; On the Space Optimizing Effect of Eliminating Single Productions form LR Parsers; Acta Informatica 14, 157-174, 1980.

- (23) TOKUDA, TAKEHIRO; Eliminating Unit Productions from LR(k) Parsers Using Minimum Contexts; Acta Informatica 15, 447-470, 1981.