ÜBER DIE

EFFIZIENZ VON LR(k)-ANALYSATOREN

von Herbert Kopp

A 74/13

ÜBER DIE EFFIZIENZ VON LR(k) - ANALYSATOREN

von Herbert Kopp

Summary

For a phrase structure grammar $\mathcal{G}$ defining a formal language $L \subset \mathcal{V}^*$ the analysis problem requires to find to a given word $w \in \mathcal{V}^*$ the derivations according to $\mathcal{G}$ if there exists one, or to state, that $w \notin L$ otherwise.

A simple procedure to solve this problem for contextfree grammars is the LR(k) - algorithm [3]. For LR(k) - grammars it solves the problem in linear time. But if $\mathcal{G}$ is not of type LR(k), then deadlocks can come up and therefore the algorithm is supposed to be unefficient in this case.

In the following we develop a method to estimate the probability of deadlocks and the average analysis time. It can be shown, that the LR(k) - algorithm is efficient also for grammars, that are not of type LR(k).

Einleitung

Ist $\mathcal{G}$ eine Chomsky-Grammatik, die eine formale Sprache $L \subset \mathcal{T}^*$ definiert, dann besteht das Analyseproblem für $\mathcal{G}$ darin, zu $w \in \mathcal{T}^*$ zu entscheiden, ob $w \in L$ ist, und in diesem Fall die Ableitungen für w zu finden.

Ein einfaches Verfahren, das das Analyseproblem für die Klasse der kontextfreien Grammatiken löst, ist der LR(k)-Algorithmus [3]. Für Grammatiken vom Typ LR(k) kann man damit das Analyseproblem sogar in Linearzeit lösen.

Ist nun $\mathcal{G}$ nicht vom Typ LR(k), dann können während der Analyse Sackgassen auftreten, was zu der Vermutung führt, daß der Algorithmus für solche Grammatiken nicht mehr sehr effizient ist. Um seine Brauchbarkeit in solchen Fällen beurteilen zu können, muß man die folgenden Fragen klären:

- Wie häufig treten bei der Analyse eines Wortes $w \in \mathcal{T}^*$ überhaupt Sackgassen auf?
- Welche Analysezeit benötigt der LR(k)-Algorithmus für $w \in \mathcal{T}^*$?

In der folgenden Arbeit entwickeln wir Methoden, die es erlauben, abzuschätzen, wie groß die Wahrscheinlichkeit für das Auftreten von Sackgassen bei der Analyse eines Wortes $w \in \mathcal{T}^n$ ist, und wie lange die Analyse eines solchen Wortes durchschnittlich dauert.

Daraus kann man entnehmen, daß das LR(k)-Verfahren durchaus auch für Grammatiken effizient sein kann, die nicht vom Typ LR(k) sind.

I. Backgassen bei der Analyse nach dem LR(0) - Verfahren

Präzisierung des Problems und Lösungsansatz

Wir betrachten eine Chomsky-reduzierte [1] Grammatik

$$\mathcal{G} = \langle \mathcal{N}, \mathcal{T}, \mathcal{P}, \alpha \rangle$$

Darin bezeichnet
- $\mathcal{N}$ das Variablenalphabet
- $\mathcal{T}$ das Terminalalphabet
- $\mathcal{P}$ das Produktionensystem
- α das Axiom .

Mit $\mathcal{T}^*$ bezeichnen wir das freie Monoid über dem Alphabet $\mathcal{T}$ und ε ist das neutrale Element darin. Es sei $\mathcal{T}^+ = \mathcal{T} - \{\varepsilon\}$. Ist $f = x \rightarrow y$ eine Produktionsregel aus $\mathcal{P}$, dann bezeichnen wir mit $Q(f) = x$ und mit $Z(f) = y$ die linke bzw. die rechte Seite der Produktion f.

Zu $f \in \mathcal{P}$ betrachten wir nun die Menge

$$R(f) = \left\{ wq \;\middle|\; \begin{array}{l} w \in (\mathcal{N} \cup \mathcal{T})^*, \quad Z(f) = q \quad \text{und es gibt} \\ \text{eine kanonische Reduktion:} \\ wqv \leftarrow wQ(f)v \overset{*}{\leftarrow} \alpha \end{array} \right\}$$

Dabei sind $\leftarrow$ und $\overset{*}{\leftarrow}$ die Relationen, die aus den Relationen $\rightarrow$ und $\overset{*}{\rightarrow}$ durch Spiegelung entstehen. Zum Begriff der kanonischen Reduktion vergleiche man etwa [1] .

Man kann zeigen, daß die Mengen $R(f)$ regulär sind. Mit ihrer Hilfe kann man bekanntlich LR(0)-Grammatiken wie folgt charakterisieren [1] :

Satz: Eine kontextfreie, Chomsky-reduzierte Grammatik $\mathcal{G}$ ist genau dann vom Typ LR(0), wenn für alle Produktionen $f, g \in \mathcal{P}$ gilt:

$$R(f)\cdot \mathcal{V}^* \cap R(g) = \emptyset \qquad \text{falls } f \neq g$$
$$R(f)\cdot \mathcal{V}^+ \cap R(f) = \emptyset$$

Bezeichnung: Ist $wq \in R(f)$, $q = Z(f)$, dann bezeichnen wir q als ein Handle des Wortes wqv, wobei $v \in \mathcal{V}^*$ ist.

Die Mengen $R(f)$ enthalten also diejenigen Anfangswörter, die der LR(0)-Algorithmus gelesen haben muß, um ein Handle zu entdecken. Die Aussage des Satzes ist, daß bei der Analyse stets höchstens ein Handle in jedem Schritt gefunden werden kann.

Ist nun $\mathcal{G}$ nicht vom Typ LR(0), dann kann es vorkommen, daß im Verlauf der Analyse ein Wort w mit mehreren Handles erzeugt wird. Es ist dann $w = w_1 w_2$ mit $w_2 \in \mathcal{V}^*$ und

$$w_1 \in \mathcal{M}(\mathcal{P}) = \bigcup_{\substack{f,g \in \mathcal{P} \\ f \neq g}} \left[R(f)\cdot \mathcal{V}^* \cap R(g)\right] \cup \bigcup_{f \in \mathcal{P}} \left[R(f)\cdot \mathcal{V}^+ \cap R(f)\right]$$

In diesem Fall kann der LR(0)-Algorithmus in eine Sackgasse geraten; dann nämlich, wenn von den in Frage kommenden Handles ein falsches ausgewählt wird.

Es ist nun vor allem von Interesse, zu wissen, wie wahrscheinlich es ist, daß man bei der Analyse von $w \in \mathcal{V}^n$ zu einem Wort $u = u_1 u_2$ gelangt mit $u_2 \in \mathcal{V}^*$ und $u_1 \in \mathcal{M}(\mathcal{P})$.

Dieses Problem soll nun möglichst unabhängig von der speziellen Form der Produktionen $f \in \mathcal{P}$ beantwortet werden. Es liegt die Vermutung nahe, daß ein Zusammenhang besteht zwischen der Anzahl der Wörter in $\mathcal{M}(\mathcal{P})$ und der gesuchten Wahrscheinlichkeit. An einer einfachen Vorüberlegung wird dies bereits deutlich:

Es sei $\mathfrak{M}(\mathfrak{P}) \subset \mathfrak{T}^*$. Dann kann der LR(0)-Algorithmus höchstens im ersten Reduktionsschritt mehr als ein Handle vorfinden. In diesem Fall gilt für die Wahrscheinlichkeit q(n), daß man bei der Analyse von $u \in \mathfrak{T}^n$ zu einem Wort w mit mehr als einem Handle gelangt:

$$q(n) = \frac{|\mathfrak{M}(\mathfrak{P}) \cdot \mathfrak{T}^* \cap \mathfrak{T}^n|}{|\mathfrak{T}^n|}$$

Hier kann man also die Berechnung von q(n) zurückführen auf das Problem, zu berechnen, wieviele Wörter der Länge n es in einer regulären Sprache gibt. Im allgemeinen sind die Zusammenhänge nicht mehr so einfach. Man kann aber auf ähnliche Weise auch für beliebige kontextfreie Grammatiken brauchbare Aussagen gewinnen.

Sackgassen bei der Analyse kontextfreier Grammatiken

Wir betrachten eine Grammatik $\mathfrak{G}$ mit den oben angegebenen Eigenschaften und verlangen zusätzlich, daß die Produktionen $f \in \mathfrak{P}$ von der Form sind $f = X \rightarrow uv$ mit $X \in \mathfrak{N}$ und $u, v \in \mathfrak{N} \cup \mathfrak{T}$.

Diese Einschränkungen dienen nur der Vereinfachung der folgenden Überlegungen. Sie haben keine grundsätzliche Bedeutung für die erzielten Ergebnisse.

Bezeichnung: $\mathfrak{P} = \{f_1, \ldots, f_r\}$ sei das Produktionensystem von $\mathfrak{G}$. Eine kanonische Reduktionsfolge $u_1 \leftarrow u_2 \leftarrow \ldots \leftarrow u_m$ heißt normiert, wenn sie die folgenden Bedingungen erfüllt:

i) Ist $u_i = a_i b_i c_i$ und $u_{i+1} = a_i X_i c_i$ für ein $i \in [1 : m-1]$ und ist $f_{j_i} = X_i \rightarrow b_i$, dann gilt für alle $f_k \in \mathfrak{P}$ mit $a_i b_i \in R(f_k)$: $k \geq j_i$

ii) Falls $a_i b_i = gh$ ist, und $g \in R(f)$ für ein $f \in \mathcal{P}$, dann ist $h = \varepsilon$.

Anschaulich heißt eine kanonische Reduktionsfolge also normiert, wenn der LR(0)-Algorithmus, der sie erzeugt, in jedem Schritt das erste mögliche Handle reduziert und zwar mit derjenigen Produktion, die von allen in Frage kommenden den kleinsten Index besitzt. Man sieht:

Hilfssatz: Zu jedem $w \in (\mathcal{N} \cup \mathcal{T})^*$ gibt es genau eine normierte, kanonische Reduktionsfolge $w = w_1 \leftarrow w_2 \leftarrow \cdots \leftarrow w_s$ von maximaler Länge.

Wir untersuchen zunächst die folgende Problemstellung: Gegeben sei $w \in (\mathcal{N} \cup \mathcal{T})^n$ so daß es dazu ein $u_1 \in \mathcal{T}^*$ gibt und eine kanonische Reduktionsfolge $u_1 \leftarrow u_2 \leftarrow \cdots \leftarrow u_s$ mit $u_s = w$. Gesucht ist

$$p(n) = \left\{ \begin{array}{l} \text{Wahrscheinlichkeit dafür, daß der} \\ \text{LR(0)-Algorithmus von } w \text{ ausgehend nicht} \\ \text{zu einem } w' \in \mathcal{M}(\mathcal{P}) \cdot \mathcal{T}^* \text{ gelangt.} \end{array} \right.$$

Wir präzisieren diese Fragestellung weiter, indem wir einen geeigneten Wahrscheinlichkeitsraum $\Omega(n)$ definieren. Es sei

$$\Omega(n) = \left\{ \omega = (w_1, \ldots w_s) \;\middle|\; \begin{array}{l} \omega \text{ hat die Eigenschaften} \\ \text{i), ii) iii), iv), v)} \end{array} \right\}$$

i) Es ist $s \geq 1$, $|w_1| = n$ und für $i \in [1 : s]$ ist $w_i \in (\mathcal{N} \cup \mathcal{T})^*$.

ii) $w_1 \leftarrow \cdots \leftarrow w_s$ ist eine normierte, kanonische Reduktion und es gibt ein $u \in \mathcal{T}^*$ und eine kanonische Reduktion $u \leftarrow \cdots \leftarrow w_1$.

iii) $w_s \notin \bigcup_{f \in \mathcal{P}} R(f) \cdot \mathcal{Z}^*$, d.h. w_s ist nicht weiter reduzierbar.

iv) Zu jedem $w \in (\mathcal{N} \cup \mathcal{Z})^*$ mit $|w| = n$, zu dem es ein $u \in \mathcal{Z}^*$ gibt und eine kanonische Reduktion $u \xleftarrow{*} w$, existiert ein $\omega \in \Omega(n)$ mit $\omega = (w, \ldots)$.

Erläuterung: $\Omega(n)$ enthält zu jedem $w \in (\mathcal{N} \cup \mathcal{Z})^n$, das aus einem terminalen Wort u durch eine kanonische Reduktion hervorgeht, genau eine kanonische Reduktionsfolge und zwar die eindeutig bestimmte, die mit w beginnt, normiert ist und maximale Länge hat.

Es ist nun klar, daß der LR(0)-Algorithmus genau dann bei der Analyse von w ein w' mit mehreren Handles erzeugt, wenn die mit w beginnende Reduktionsfolge in $\Omega(n)$ die Gestalt hat: $(w_1, \ldots, w_i, \ldots, w_s)$ mit $w_1 = w$ und $w_i \in \mathcal{M}(\mathcal{P}) \cdot \mathcal{Z}^*$.

Als σ-Algebra auf dem Stichprobenraum $\Omega(n)$ wählen wir die Potenzmenge von $\Omega(n)$ und als Wahrscheinlichkeitsmaß darauf die Gleichverteilung. Für die Wahrscheinlichkeit eines Elementarereignisses $\omega \in \Omega(n)$ gilt also stets:

$$\frac{1}{|\mathcal{Z}^n|} \geq \sigma(\omega) \geq \frac{1}{\left| \bigcup_{f \in \mathcal{P}} R(f) \mathcal{Z}^* \cap (\mathcal{N} \cup \mathcal{Z})^n \right| + |\mathcal{Z}^n|}$$

es sei nun

$$z(n) = \begin{cases} \text{Anzahl der } \omega \in \Omega(n) \text{ von der Form} \\ \omega = (w_1, \ldots, w_s), \text{ für die ein } i \text{ existiert,} \\ i \in [1 : s] \text{ mit } w_i \in \mathcal{M}(\mathcal{P}) \cdot \mathcal{Z}^* \end{cases}$$

Wegen $$p(n) = 1 - \frac{z(n)}{|\Omega(n)|}$$

können wir die Berechnung der gesuchten Größe $p(n)$ zurückführen auf die Berechnung der Größen $z(n)$ und $|\Omega(n)|$.

Abschätzung von $z(n)$: Ist $(w_1,\ldots,w_s) \in \Omega(n)$ und für ein $i \in [1:s]$ $w_i \in \mathcal{M}(\mathcal{P}) \cdot \mathcal{Y}^*$, dann gilt entweder i) oder ii):

i) Es gibt ein $i \in [2:s]$ mit $w_i \in \mathcal{M}(\mathcal{P}) \cdot \mathcal{Y}^*$

ii) $w_1 \in \mathcal{M}(\mathcal{P}) \cdot \mathcal{Y}^*$ und $w_i \notin \mathcal{M}(\mathcal{P}) \cdot \mathcal{Y}^*$ für alle $i \in [2:s]$

Wir benutzen im folgenden noch einige Bezeichnungen: Es sind

$$A(n) = \left\{ (w_1,\ldots,w_s) \in \Omega(n) \;\middle|\; \begin{array}{l} w_1 \in \mathcal{M}(\mathcal{P}) \cdot \mathcal{Y}^* \text{ und für} \\ i \in [2:s] \text{ ist } w_i \notin \mathcal{M}(\mathcal{P}) \cdot \mathcal{Y}^* \end{array} \right\}$$

$$\alpha(n) = |A(n)|$$
$$T = |\mathcal{Y}|$$
$$N = |\mathcal{N}|$$

Ist $k(x) = \left\{ w \in (\mathcal{N} \cup \mathcal{Y})^2 \;\middle|\; x \longrightarrow w \right\}$ für alle $x \in \mathcal{N}$, dann setzen wir fest:

$$k = \max_{x \in \mathcal{N}} |k(x)|$$

Da es nun höchstens k Möglichkeiten gibt, um eine Folge $\omega' \in \Omega(n)$ zu einer Folge $\omega \in \Omega(n)$ fortzusetzen, erhalten wir für $z(n)$ die folgende Rekursionsformel:

$$z(n+1) \leq \alpha(n+1) + z(n) \cdot k$$

Durch Auflösen dieser Rekursion mit dem Startwert $z(1) = 0$ erhält man

$$z(n) \leq \sum_{i=2}^{n} \alpha(i) \cdot k^{n-i}$$

Man überlegt sich leicht, daß man ganz analoge Abschätzungen herleiten kann, wenn die Produktionsregeln nicht von der speziellen Form sind, die oben vorausgesetzt wurde.

Die Größe k entnimmt man leicht aus dem Produktionensystem, während zur Bestimmung von $\alpha(i)$ kein einfaches Verfahren zur Verfügung steht. Es genügt jedoch oft, daß man die Beziehung

$$\alpha(i) \leq \left| \mathfrak{M}(\mathfrak{P}) \cdot \mathfrak{Y}^* \cap (\mathfrak{N} \cup \mathfrak{Y})^i \right|$$

benutzt, wobei man die reguläre Menge $\mathfrak{M}(\mathfrak{P})$ und damit auch $\left| \mathfrak{M}(\mathfrak{P}) \cdot \mathfrak{Y}^* \cap (\mathfrak{N} \cup \mathfrak{Y})^i \right|$ effektiv aus dem Produktionensystem gewinnen kann.

<u>Abschätzung von $|\Omega(n)|$:</u> Zu $f \in \mathfrak{P}$ betrachten wir die reguläre Menge

$$S(f) = \left\{ wX \mid X = Q(f),\ wZ(f) \in R(f) \right\}$$

Wie $R(f)$ kann man auch $S(f)$ direkt aus $\mathfrak{P}$ ableiten. Man sieht, daß $w \in (\mathfrak{N} \cup \mathfrak{Y})^n$ genau dann aus einem $u \in \mathfrak{Y}^*$ durch eine kanonische Reduktion hervorgeht, wenn i) oder ii) erfüllt sind:

i) $w \in \mathfrak{Y}^n$

ii) $w = uv$ mit $v \in \mathfrak{Y}^*$ und $u \in \bigcup_{f \in \mathfrak{P}} S(f)$

Damit können wir $|\Omega(n)|$ berechnen als die Anzahl der Wörter der Länge n in der regulären Menge

$$\bigcup_{f \in \mathfrak{P}} S(f) \cdot \mathfrak{Y}^* \cup \mathfrak{Y}^*$$

Es gilt daher

$$p(n) \geq 1 - \frac{1}{|\Omega(n)|} \cdot \sum_{i=2}^{n} \alpha(i) k^{n-i}$$

Wir können nun auf die ursprüngliche Fragestellung zurückkommen und die Wahrscheinlichkeit

$$q(n) = \begin{cases} \text{Wahrscheinlichkeit dafür, daß bei der LR(0)-} \\ \text{Analyse von } w \in \mathcal{Y}^n \text{ ein } w' \in \mathcal{M}(\mathcal{P}) \cdot \mathcal{Y}^* \text{ ensteht.} \end{cases}$$

berechnen. Dazu sei

$$s(n) = \begin{cases} \text{Anzahl der } \omega = (w_1, \ldots, w_s) \in \Omega(n) \text{ mit } w_1 \in \mathcal{Y}^* \\ \text{für die ein } i \in [1 : s] \text{ existiert mit } w_i \in \mathcal{M}(\mathcal{P}) \cdot \mathcal{Y}^* \end{cases}$$

Mit diesen Bezeichnungen erhalten wir

$$q(n) = 1 - \frac{s(n)}{T^n} \geq 1 - \frac{z(n)}{T^n}$$

$$\text{also} \qquad q(n) \geq 1 - \frac{1}{T^n} \cdot \sum_{i=2}^{n} \alpha(i) k^{n-i}$$

Bemerkung:

$\mathcal{G}$ sei nun eine lineare Grammatik, d.h. unter den obigen Voraussetzungen sind alle Produktionen von der Form:

$$X \longrightarrow aY$$

oder $\quad X \longrightarrow Ya$

oder $\quad X \longrightarrow ab \qquad$ mit $X, Y \in \mathcal{N}$ und $a, b \in \mathcal{Y}$

Da $\mathcal{G}$ Chomsky-reduziert ist, kann man jedes Wort $w_1 Z w_2$ mit $w_1, w_2 \in \mathcal{Y}^*$ und $Z \in \mathcal{N}$ aus einem geeigneten Wort $u \in \mathcal{Y}^*$ gewinnen, und wir erhalten somit

$$|\Omega(n)| = N \cdot T^{n-1} \cdot n$$

und

$$p(n) \geq 1 - \frac{1}{n \cdot N \cdot T^{n-1}} \cdot \sum_{i=2}^{n} \alpha(i) k^{n-i}$$

Beispiel 1 :

Wir betrachten die Grammatik $\mathcal{G} = \langle \mathcal{N}, \mathcal{T}, \mathcal{P}, S \rangle$ mit :

$$\mathcal{N} = \{ A, A_1, A_2, B, B_1, S \}$$
$$\mathcal{T} = \{ a, b, c \}$$
$$\mathcal{P} = \{ f_1, \ldots, f_9 \}$$

dabei gilt :

$f_1 = S \rightarrow Ac$ $\quad$ $f_6 = A_2 \rightarrow b$

$f_2 = S \rightarrow aB$ $\quad$ $f_7 = B \rightarrow B_1 b$

$f_3 = A \rightarrow aA_1$ $\quad$ $f_8 = B_1 \rightarrow aB$

$f_4 = A_1 \rightarrow A_2 b$ $\quad$ $f_9 = B \rightarrow b$

$f_5 = A_2 \rightarrow Ab$

Es ist $\mathcal{L}(\mathcal{G}) = \{ a^n b^{2n} c \mid n \geq 1 \} \cup \{ a^n b^n \mid n \geq 1 \}$

$\mathcal{G}$ ist für kein k vom Typ LR(k). Wir berechnen $\mathcal{M}(\mathcal{P})$:

$R(f_1) = \{ Ac \}$

$R(f_2) = \{ aB \}$

$R(f_3) = \{ a^n A_1 \mid n \geq 1 \}$

$R(f_4) = \{ a^n A_2 b \mid n \geq 1 \}$

$R(f_5) = \{ a^n A b \mid n \geq 1 \}$

$R(f_6) = \{ a^n b \mid n \geq 1 \}$

$R(f_7) = \{ a^n B_1 b \mid n \geq 1 \}$

$R(f_8) = \{ a^n B \mid n \geq 2 \}$

$R(f_9) = \{ a^n b \mid n \geq 1 \}$

$\Rightarrow \quad \mathcal{M}(\mathcal{P}) = \{ a^n b \mid n \geq 1 \}$

In diesem Fall ist $z(n) = s(n) = \left| \mathfrak{M}(\mathfrak{P}) \cdot \mathfrak{T}^* \cap \mathfrak{T}^n \right|$

Für $n \geq 2$ folgt somit:

$$q(n) = 1 - \frac{z(n)}{T^n} = 1 - \frac{1}{3^n} \cdot \sum_{i=0}^{n-2} 3^i$$

$$\underline{q(n) = \frac{5}{6} + \frac{1}{2 \cdot 3^n}}$$

Beispiel 2:

Wir betrachten die folgende kontextfreie Grammatik $\mathfrak{G} = \langle \mathfrak{N}, \mathfrak{T}, \mathfrak{P}, \alpha \rangle$ mit:

$$\mathfrak{N} = \{\alpha, \tau, \varphi\}$$
$$\mathfrak{T} = \{V, (,), +, x\}$$
$$\mathfrak{P} = \{f_1, \ldots, f_6\}$$

dabei sind

$$f_1 = \varphi \rightarrow V$$
$$f_2 = \varphi \rightarrow (\alpha)$$
$$f_3 = \tau \rightarrow \varphi$$
$$f_4 = \tau \rightarrow \tau x \varphi$$
$$f_5 = \alpha \rightarrow \tau$$
$$f_6 = \alpha \rightarrow \alpha + \tau$$

Grammatiken dieser Art werden in Programmiersprachen verwendet, um arithmetische Ausdrücke zu definieren, die unvollständige Klammerung und Vorrangregeln zwischen den Operatoren zulassen.

$\mathfrak{G}$ hat nicht die in II verlangte Normalform. Um jedoch an einem Beispiel vorzuführen, wie sich die obigen Überlegungen auch in anderen Fällen anwenden lassen, legen wir diese Form der Produktionen zugrunde und modifizieren die obigen Abschätzungen entsprechend. Zunächst berechnen wir $\mathfrak{M}(\mathfrak{P})$:

Es seien $Y = \{\varepsilon, \alpha+, \tau x, \alpha+\tau x\}$

$B = \{y_1(y_2(\ldots(y_s(\mid s \geq 2,\ y_i \in Y\} \cup \{\varepsilon\}$

Mit ε haben wir dabei das neutrale Element in $(\mathcal{M} \cup \mathcal{T})^*$ bezeichnet. Damit ist

$$R(f_1) = B \cdot Y \cdot \{V\}$$
$$R(f_2) = B \cdot Y \cdot \{(\alpha)\}$$
$$R(f_3) = B \cdot \{\varphi,\ \alpha+\varphi\}$$
$$R(f_4) = B \cdot \{\tau x \varphi,\ \alpha+\tau x \varphi\}$$
$$R(f_5) = B \cdot \{\tau\}$$
$$R(f_6) = B \cdot \{\alpha+\tau\}$$

$$\Rightarrow \quad \mathcal{M}(\mathcal{P}) = B \cdot \{\tau x V,\ \alpha+\tau x V\}$$

Berechnung von $z(n)$:

Es sei $\beta(n) = \mid B \cap (\mathcal{M} \cup \mathcal{T})^n \mid$

$\mu(n) = \mid \mathcal{M}(\mathcal{P}) \cap (\mathcal{M} \cup \mathcal{T})^n \mid$

Nun ist aber $B = B \cdot \{(, \alpha+(, \tau x(, \alpha+\tau x(\} \cup \{\varepsilon\}$

und daher gilt für $n \geq 5$:

$$\beta(n) = \beta(n-1) + 2\beta(n-3) + \beta(n-5)$$

und $\beta(0) = \beta(1) = \beta(2) = 1$

$\beta(3) = 3$

$\beta(4) = 5$

Die allgemeine relle Lösung der Differenzengleichung

$$\beta(n+5) - \beta(n+4) - 2\beta(n+2) - \beta(n) = 0$$

ist von der Form $\beta(n) = q^n \cdot \text{const}$, wobei q die einzige reelle Nullstelle des Polynoms

$$p(x) = x^5 - x^4 - 2x^2 - 1$$

ist. Man rechnet nach, daß gilt:

$$\frac{7}{4} = \frac{168}{96} < q < \frac{169}{96}$$

Unter Beachtung der Anfangsbedingungen erhält man für $n \geq 5$ die folgende Abschätzung für $\beta(n)$:

$$0{,}4731 \cdot \left(\frac{168}{96}\right)^n < \beta(n) < 0{,}4875 \cdot \left(\frac{169}{96}\right)^n$$

Nun gilt aber für $n \geq 5$: $\mu(n) = \beta(n-3) + \beta(n-5)$ und daraus erhält man für $n \geq 10$:

$$0{,}1171 \cdot \left(\frac{168}{96}\right)^n < \mu(n) < 0{,}1182 \cdot \left(\frac{169}{96}\right)^n$$

Die Anfangswerte für $\mu(n)$ berechnet man leicht direkt:

$$\mu(0) = \mu(1) = \mu(2) = 0$$

$$\mu(3) = \mu(4) = 1$$

$$\mu(5) = 2$$

$$\mu(6) = 4$$

$$\mu(7) = 6$$

$$\mu(8) = 11$$

$$\mu(9) = 20$$

Mit $k(x) = \left\{ u \in (\mathfrak{N} \cup \mathfrak{T})^3 \mid x \xrightarrow[*]{} u \right\}$ für $x \in \mathfrak{N}$ erhalten wir:

$$\max \left\{ |k(\alpha)|, |k(\varphi)|, |k(\tau)| \right\} = 6$$

Damit erhalten wir für $z(n)$ die folgende Rekursion:

$$z(n+1) \leq \alpha(n+1) + 6z(n-1)$$

also ist

$$z(2n+1) \leq \sum_{i=2}^{n} \alpha(2i+1)\cdot 6^{n-i}$$

$$z(2n) \leq \sum_{i=2}^{n} \alpha(2i)\cdot 6^{n-i}$$

Nun ist $\alpha(i) \leq |\mathfrak{M}(\mathfrak{P})\cdot \mathfrak{Z}^* \cap (\mathfrak{N}\cup\mathfrak{Z})^i|$

d.h. $$\alpha(i) \leq \sum_{j=0}^{i} \mu(j)\cdot 5^{i-j}$$

für $i \geq 10$ folgt: $$\alpha(i) \leq \left[0{,}0106 - 0{,}0642\cdot\left(\frac{169}{480}\right)^i\right]\cdot 5^i$$

für $i < 10$ errechnet man leicht direkt:

$\alpha(0) = \alpha(1) = \alpha(2) = 0$ $\qquad$ $\alpha(6) < 164$

$\alpha(3) = 1$ $\qquad$ $\alpha(7) < 826$

$\alpha(4) = 6$ $\qquad$ $\alpha(8) < 4141$

$\alpha(5) < 32$ $\qquad$ $\alpha(9) < 20725$

Das ergibt für $n \geq 5$ die folgende Abschätzung für $z(2n)$:

$$z(2n) < 25^n\cdot 0{,}0140 - 6^n\cdot 0{,}0939 - 3^n\cdot 0{,}0642$$

Die Bestimmung von $z(n)$ für ungerade n und für $n = 2,4,6,8$ bereitet keine Schwierigkeiten und wird daher übergangen.

Berechnung von $|\mathfrak{Q}(n)|$:

Nach II gilt:

$$|\mathfrak{Q}(n)| = \left|\left[\bigcup_{f\in\mathfrak{P}} S(f)\cdot\mathfrak{Z}^*\cup\mathfrak{Z}^*\right]\cap(\mathfrak{N}\cup\mathfrak{Z})^n\right|$$

Nun ist
$$S(f_1) = S(f_2) = B \cdot Y \cdot \{\varphi\}$$
$$= B \cdot \{\varphi\ ,\ \alpha+\varphi\ ,\ \tau \times \varphi\ ,\ \alpha+\tau \times \varphi\}$$
$$S(f_3) = S(f_4) = B \cdot \{\tau\ ,\ \alpha+\tau\}$$
$$S(f_5) = S(f_6) = B \cdot \{\alpha\}$$

Bezeichnet man mit $\eta(i)$ die Anzahl der Wörter der Länge i in der Sprache

$$\bigcup_{f \in \mathcal{P}} S(f)$$

dann gilt für $i \geq 5$:

$$\eta(i) = 3\beta(i-1) + 3\beta(i-3) + \beta(i-5)$$

und somit
$$1{,}1048 \cdot \left(\frac{168}{96}\right)^i < \eta(i) < 1{,}1257 \cdot \left(\frac{169}{96}\right)^i$$

für $i \geq 10$. Außerdem ist leicht zu sehen, daß für $i < 10$ gilt:

$\eta(0) = 0$	$\eta(5) = 19$
$\eta(1) = \eta(2) = 3$	$\eta(6) = 34$
$\eta(3) = 6$	$\eta(7) = 61$
$\eta(4) = 12$	$\eta(8) = 90$
	$\eta(9) = 180$

Wegen $|\Omega(n)| = \sum_{i=0}^{n} \eta(i) \cdot 5^{n-i} + 5^n$ erhalten wir daraus:

$$|\Omega(n)| > 5^n \cdot \left[1{,}7966 - 0{,}5949\left(\frac{7}{20}\right)\right]^n$$

Dabei ist $n \geq 10$ vorausgesetzt. Der Fall $n < 10$ bereitet keine prinzipiellen Schwierigkeiten. Wir verzichten daher auf die Durchrechnung.

Aus den Abschätzungen für $|\Omega(n)|$ und $z(n)$ kann man nun die gesuchten Wahrscheinlichkeiten $p(n)$ und $q(n)$ berechnen.

Für $n \geq 5$ gilt:

$$p(2n) > 1 - \frac{25^n \cdot 0{,}0140 - 6^n \cdot 0{,}0939 - 3^n \cdot 0{,}0642}{25^n \cdot \left[1{,}7966 - 0{,}5949\left(\frac{7}{20}\right)^n\right]}$$

$$p(2n) > 0{,}9922 + 0{,}5 \cdot \left(\frac{6}{25}\right)^n + 0{,}03 \cdot \left(\frac{3}{25}\right)^n$$

Außerdem erhalten wir für $n \geq 5$:

$$q(2n) > 0{,}9860 + 0{,}09 \cdot \left(\frac{6}{25}\right)^n + 0{,}06 \cdot \left(\frac{3}{25}\right)^n$$

Bemerkung: Die Wahrscheinlichkeiten $p(n)$ und $q(n)$ sind im allgemeinen verschieden. $p(n)$ gibt die Wahrscheinlichkeit für das Auftreten von Sackgassen an, wenn man in einer kanonischen Reduktion bei einem Wort der Länge n startet. $q(n)$ gibt die Wahrscheinlichkeit für das Auftreten von Sackgassen an, wenn man bei einem terminalen Wort der Länge n startet.

Einen einfachen Zusammenhang zwischen diesen beiden Wahrscheinlichkeiten scheint es nicht zu geben. Das hängt vermutlich daran, daß beim Erzeugen einer formalen Sprache neben dem Ableitungsprozeß noch ein weiterer, wesentlich verschiedener hinzukommt, nämlich die Auswahl terminaler Wörter.

II. Mittlerer Zeitbedarf für die LR(0)-Analyse

Die bisherigen Untersuchungen über das Auftreten von Sackgassen bei der LR(0)-Analyse haben gezeigt, daß man für viele Grammatiken durchaus erwarten kann, daß Sackgassen bei der Analyse sehr selten auftreten.

Um ein zuverlässiges Kriterium für die Brauchbarkeit des LR(0)-Verfahrens zu haben, untersuchen wir nun, wie lange die Lösung des Analyseproblems für eine gegebene Grammatik $\mathcal{G}$ durchschnittlich dauert.

Wir setzen dazu voraus, daß $\mathcal{G}$ Chomsky-reduziert ist und alle Produktionen von der Form sind

$$X \longrightarrow w$$

mit $X \in \mathcal{N}$ und $w \in (\mathcal{N} \cup \mathcal{T})^2$. Diese Einschränkung hat keine grundsätzliche Bedeutung für das folgende. Sie vereinfacht nur die Rechnung.

Weiter verwenden wir die folgenden Bezeichnungen:

- Eine R-Folge der Ordnung n zu $\mathcal{G}$ ist eine Folge

$$\varrho = (w_n, \ldots, w_r)$$

wobei $w_i \in (\mathcal{N} \cup \mathcal{T})^i$ ist für $i \in [m:n]$ und $w_n \in \mathcal{T}^*$, $w_n \longleftarrow \cdots \longleftarrow w_r$ eine kanonische Reduktionsfolge und w_r nicht weiter reduzierbar ist.

- Wir sagen w komme auf der m-ten Ebene in der R-Folge ϱ vor, wenn $m \in [n:r]$ ist und $w = w_m$.

- Ist $L \subset (\mathcal{N} \cup \mathcal{T})^*$ eine formale Sprache, dann bezeichnen wir mit $[L]_r$ die Anzahl der Elemente $w \in L$ mit $|w| = r$.

Wir untersuchen zunächst das folgende Teilproblem:

Gegeben sei eine Grammatik $\mathcal{G}$ wie oben. Gesucht ist die Anzahl der Reduktionsschritte, die ein LR(0)-Automat ausführen muß, um alle R-Folgen der Ordnung n zu erzeugen.

Dazu komme $w \in (\mathcal{N} \cup \mathcal{T})^m$ in einer R-Folge der Ordnung n zu $\mathcal{G}$ auf der m-ten Ebene vor. Dann kann man w genau dann mit der Produktion $f \in \mathcal{P}$ reduzieren, wenn $w \in R(f) \cdot \mathcal{T}^*$ ist. Dies ist eventuell auf mehrere Arten möglich und zwar auf

$$\sum_{r=2}^{m} \left| \left[R(f) \cap (\mathcal{N} \cup \mathcal{T})^r \right] \cdot \mathcal{T}^* \cap \{w\} \right|$$

verschiedene Weisen. Variiert man nun über alle R-Folgen der Ornung n, dann kann man auf der Ebene m die Produktion f auf höchstens

$$\sum_{r=2}^{m} \left[R(f) \right]_r \cdot T^{m-r}$$

verschiedene Arten anwenden. Jede dieser Anwendungsmöglichkeiten kann nun mehrfach benutzt werden. Um abzuschätzen, wie oft dies möglich ist, ordnen wir den R-Folgen der Ordnung n einen Graphen zu.

Seine Knoten seien die Wörter $w \in (\mathcal{N} \cup \mathcal{T})^*$, die in den R-Folgen der Ordnung n vorkommen.

Der Graph besitzt genau dann eine Kante $u \longrightarrow v$, wenn es eine R-Folge der Ordnung n gibt, in der v auf der m-ten Ebene und u auf der (m+1)-ten Ebene vorkommen.

Es sei wie früher für $x \in \mathcal{N}$ definiert: $k(x) = \{u \mid x \longrightarrow u\}$ und $k = \max\{|k(x)| \mid x \in \mathcal{N}\}$.

Da die R-Folgen kanonische Reduktionsfolgen sind, ist klar, daß höchstens k Kanten auf ein und denselben Knoten gerichtet sind.

Man sieht: der LR(0)-Algorithmus benutzt jede Anwendungsmöglichkeit der Produktion f auf der m-ten Ebene höchstens k^{n-m} mal. Es gibt daher höchstens

$$\sum_{r=2}^{m} \left[R(f)\right]_r \cdot T^{m-r} \cdot k^{n-m}$$

Möglichkeiten, ein Wort auf der m-ten Ebene mit der Produktion f zu reduzieren. Summiert man nun noch über alle in Frage kommenden Produktionen und alle möglichen Ebenen in den R-Folgen der n-ten Ordnung, dann erhält man eine obere Schranke für Anzahl S(n) der insgesamt auszuführenden Analyseschritte:

$$S(n) \leq \sum_{f \in \mathcal{P}} \sum_{m=2}^{n} \sum_{r=2}^{m} \left[R(f)\right]_r \cdot T^{m-r} \cdot k^{n-m}$$

Dies kann man leicht umformen zu

$$S(n) \leq \sum_{f \in \mathcal{P}} k^n \sum_{r=2}^{n} \sum_{m=r}^{n} \frac{\left[R(f)\right]_r}{T^r} \left(\frac{T}{k}\right)^m$$

$$S(n) \leq \frac{1}{T-k} \sum_{f \in \mathcal{P}} \left\{ T^{n+1} \sum_{r=2}^{n} \frac{\left[R(f)\right]_r}{T^r} - k^{n+1} \sum_{r=2}^{n} \frac{\left[R(f)\right]_r}{k^r} \right\}$$

Damit erhalten wir auch eine obere Schranke für die mittlere Anzahl der Analyseschritte, die für $w \in \mathcal{L}^n$ höchstens erforderlich ist: Es gilt

$$\circledast \qquad \overline{t(n)} = \frac{S(n)}{T^n} < \frac{T}{T-k} \sum_{f \in \mathcal{P}} \left\{ \sum_{r=2}^{n} \frac{\left[R(f)\right]_r}{T^r} - \left(\frac{k}{T}\right)^{n+1} \cdot \sum_{r=2}^{n} \frac{\left[R(f)\right]_r}{k^r} \right\}$$

Daraus erhält man nun einige interessante Folgerungen:

Satz: Es sei $\tau : \mathbb{N} \longrightarrow \mathbb{R}_+$ eine Abbildung mit

$$\frac{[R(f)]_r}{T^r} \leq \tau(n) \quad \text{für alle } f \in \mathcal{P} \text{ und für alle } r \in [0:n]$$

Dann gilt:

i) falls $k < T$ ist: $\overline{t(n)} < \frac{T \cdot |\mathcal{P}|}{T-k} \cdot n \cdot \tau(n)$

ii) falls $k = T$ ist: $\overline{t(n)} < \frac{|\mathcal{P}|}{2} \cdot (n^2 - 3n + 3) \cdot \tau(n)$

Beweis:

i) Wegen $k < T$ erhält man aus ⊛:

$$\overline{t(n)} < \frac{T}{T-k} \sum_{f \in \mathcal{P}} \sum_{r=2}^{n} \frac{[R(f)]_r}{T^r} \left\{1 - \left(\frac{k}{T}\right)^{n+1}\right\}$$

$$\overline{t(n)} < \frac{|\mathcal{P}| \cdot T}{T-k} \cdot \tau(n) \cdot (n-1)$$

ii) Wegen $k = T$ erhält man aus ⊛:

$$\overline{t(n)} < \sum_{f \in \mathcal{P}} \sum_{r=2}^{n} \frac{[R(f)]_r}{T^r} \cdot (n-r)$$

$$\overline{t(n)} < |\mathcal{P}| \cdot \tau(n) \sum_{r=2}^{n} (n-r)$$

$$\overline{t(n)} < |\mathcal{P}| \cdot \tau(n) \cdot \frac{(n-2)(n-1)}{2}$$

Korollar: $\mathcal{G}$ sei eine metalineare Grammatik und $k < T$.
Dann gilt:

$$\overline{t(n)} < \frac{2 \cdot N \cdot |\mathcal{P}|}{T - k} \cdot (n-1)$$

Beweis: Für alle Produktionen $f \in \mathcal{P}$, die nicht Startproduktion sind, ist $Z(f) \in \mathcal{Y} \cdot \mathcal{N} \cup \mathcal{N} \cdot \mathcal{Y}$
Somit gilt:

$$R(f) \subset \mathcal{Y}^* \cdot \mathcal{N} \cup \mathcal{Y}^* \cdot \mathcal{N} \cdot \mathcal{Y}$$

und für alle $r \in \mathbb{N}$ folgt daraus

$$\frac{\left[R(f)\right]_r}{T^r} \leq \frac{2 \cdot T^{r-1} \cdot N}{T^r} = \frac{2 \cdot N}{T}$$

Setzt man dies in die Abschätzung i) des obigen Satzes ein, so erhält man:

$$\overline{t(n)} < \frac{2 \cdot N \cdot |\mathcal{P}|}{T - k} \cdot (n-1)$$

Beispiel 3: Wir betrachten die Grammatik $\mathcal{G} = \langle \mathcal{N}, \mathcal{T}, \mathcal{P}, X \rangle$ mit:

$$\mathcal{N} = \{X, Y\}$$
$$\mathcal{T} = \{a, b, c\}$$
$$\mathcal{P} = \{f_1, f_2, f_3, f_4\}$$

Dabei gilt: $f_1 = X \rightarrow aY$

$f_2 = X \rightarrow cc$

$f_3 = Y \rightarrow Xb$

$f_4 = Y \rightarrow aY$

Man sieht: $\mathcal{L}(\mathcal{G}) = \{ a^n ccb^m \mid n \geq m \geq 0 \}$

Außerdem ist die Grammatik $\mathcal{G}$ nicht eindeutig und somit für kein k vom Typ LR(k). Die Mengen R(f) sind:

$$R(f_1) = \{ a^n Y \mid n \geq 1 \}$$
$$R(f_2) = \{ a^n cc \mid n \geq 1 \}$$
$$R(f_3) = \{ a^n Xb \mid n \geq 1 \}$$
$$R(f_4) = \{ a^n Y \mid n \geq 2 \}$$

$$\Rightarrow \quad \mathcal{M}(\mathcal{P}) = \{ a^n Y \mid n \geq 2 \}$$

Zunächst kann man daraus die Wahrscheinlichkeiten p(n) und q(n) berechnen: Es ist

$$A(n) = \{ (w_1, \ldots, w_s) \in \mathcal{L}(n) \mid w_1 = aaYu \text{ mit } u \in \mathcal{T}^* \}$$
$$\alpha(n) = T^{n-3} = 3^{n-3} \quad \text{für } n \geq 3$$
$$\alpha(2) = 0$$
$$k = 2$$

Daraus erhalten wir die gesuchten Abschätzungen für $p(n)$ und $q(n)$:

Es ist: $$p(n) > 1 - \frac{1}{2 \cdot n \cdot 3^{n-1}} \cdot \sum_{i=3}^{n} 3^{i-3} \cdot 2^{n-i} \qquad (\text{für } n \geq 2)$$

$$\Rightarrow \qquad \underline{p(n) > 1 - \frac{1}{2n}\left[\frac{1}{3} - \frac{3}{4}\left(\frac{2}{3}\right)^n\right]}$$

und $$q(n) > 1 - \frac{1}{3^n} \sum_{i=3}^{n} 3^{i-3} \cdot 2^{n-i}$$

$$\Rightarrow \qquad \underline{q(n) > \frac{8}{9} + \frac{1}{4}\left(\frac{2}{3}\right)^n}$$

Wegen $k < T$ erhalten wir weiter aus dem Korollar:

$$\overline{t(n)} < \frac{2 \cdot 2 \cdot 4}{3-2} \cdot (n-1) = 16 \cdot (n-1)$$

Legt man die Abschätzung des Satzes selbst zugrunde, dann gilt sogar wegen

$$\frac{[R(f)]_r}{T^r} \leq \frac{2}{3}$$

$$\overline{t(n)} < \frac{3 \cdot 4}{3-2} \cdot (n-1) \cdot \frac{2}{3} = 8(n-1)$$

Beispiel 4: Wir wollen nun noch die mittlere Analysezeit für die in Beispiel 2 definierte Ausdruckssprache berechnen. Da die Produktionen in $\mathcal{P}$ nicht die verlangte Normalform besitzen muß die Abschätzung für $S(n)$ geeignet modifiziert werden:

Auf der m-ten Ebene können wir wie früher genau auf

$$\sum_{f\in\mathcal{P}} \sum_{r=2}^{m} \left[R(f)\right]_r \cdot T^{m-r}$$

verschiedene Weisen mit den Produktionen $f \in \mathcal{P}$ reduzieren. Jede dieser Möglichkeiten kann nun noch in mehreren R-Folgen angewandt werden, in unserem Fall jedoch in höchstens

$$6^{\frac{n-m}{2}}$$

verschiedenen R-Folgen der Ordnung n . Somit erhalten wir:

$$S(n) < \sum_{m=3}^{n} \sum_{r=1}^{m} \sum_{f\in\mathcal{P}} \left[R(f)\right]_r \cdot T^{m-r} \cdot k^{\frac{n-m}{2}}$$

$$S(n) < \frac{1}{T-\sqrt{k}} \sum_{f\in\mathcal{P}} \left\{ T^{n+1} \sum_{r=1}^{n} \frac{\left[R(f)\right]_r}{T^r} - (\sqrt{k})^{n+1} \sum_{r=1}^{n} \frac{\left[R(f)\right]_r}{(\sqrt{k})^r} \right\}$$

Nun ist aber $\sqrt{k} < T$ und daher folgt

$$S(n) < \frac{1}{T-\sqrt{k}} \sum_{f\in\mathcal{P}} \sum_{r=1}^{n} \frac{\left[R(f)\right]_r}{T^r} \left\{ 1 - \left(\frac{\sqrt{k}}{T}\right)^{n+1} \right\} \cdot T^{n+1}$$

Aus der Definition der Mengen $R(f_i)$ in Beispiel 2 entnehmen wir leicht:

$$\left[R(f_1)\right]_r = \beta(r-3) + 2\beta(r-5) + \beta(r-7)$$
$$\left[R(f_2)\right]_r = \beta(r-1) + 2\beta(r-3) + \beta(r-5)$$
$$\left[R(f_3)\right]_r = \beta(r-1) + \beta(r-3)$$

$$\left[R(f_4)\right]_r = \beta(r-3) + \beta(r-5)$$
$$\left[R(f_5)\right]_r = \beta(r-1)$$
$$\left[R(f_6)\right]_r = \beta(r-3)$$

Wegen $\overline{t(n)} = S(n) : T^n$ erhalten wir daraus:

$$\overline{t(n)} < \frac{T}{T-\sqrt{k}} \cdot \sum_{r=1}^{n} \frac{1}{T^r} \left\{3\beta(r-1) + 6\beta(r-3) + \beta(r-7) + 4\beta(r-5)\right\} \cdot \left\{1 - \left(\frac{\sqrt{k}}{T}\right)^{n+1}\right\}$$

Nun ist aber $\beta(r) < 0{,}4875\left(\frac{169}{96}\right)^r$ und $\sqrt{k} < \frac{5}{2}$ und somit erhalten wir aus der obigen Ungleichung durch Einsetzen dieser Werte

$$\overline{t(n)} < 2$$

Bemerkung: Das Beispiel zeigt, daß das LR(0)-Verfahren im Mittel bereits nach 2 Schritten die Entscheidung bringen kann. Das heißt, daß $w \in \mathcal{Y}^n - \mathcal{L}(\mathcal{G})$ im allgemeinen sehr schnell ausgeschieden wird.

Eine andere Fragestellung wäre, wie lange die Analyse im Mittel dauert, wenn man nur Worte $w \in \mathcal{L}(\mathcal{G})$ vorgibt. Bezeichnen wir diese Analysezeit mit $\overline{\tau(n)}$, dann gilt offenbar

$$\overline{\tau(n)} < \frac{S(n)}{|\mathcal{L}(\mathcal{G}) \cap \mathcal{Y}^n|}$$

Es ist jedoch klar, daß diese Abschätzung im allgemeinen nicht scharf genug sein kann, insbesondere dann, wenn $\mathcal{L}(\mathcal{G})$ eine Sprache von geringer Dichte ist.

III. Übertragung auf das LR(k)-Verfahren

Die obigen Untersuchungen basieren auf einer Charakterisierung von LR(0)-Grammatiken durch die regulären Mengen $R(f)$ für $f \in \mathcal{P}$. In [1, Seite 204] wird eine analoge Charakterisierung von LR(k)-Grammatiken hergeleitet, die anstelle der Mengen $R(f)$ durch eine endliche Anzahl regulärer Mengen $P(f,t)$ ersetzt, wobei $t \in \mathcal{T}^k$ ist. Wie die $R(f)$ kann man auch die Mengen $P(f,t)$ effektiv aus dem Produktionensystem gewinnen. Damit lassen sich die hier für das LR(0)-Verfahren gewonnenen Ergebnisse direkt auch für das LR(k)-Verfahren formulieren.

Herrn Professor Dr.G.Hotz, Herrn Dr.G.Kaufholz sowie Herrn Dr.R.Kemp danke ich für zahlreiche Anregungen und Gespräche bei der Entstehung dieser Arbeit.

Literatur

[1] G.Hotz, V.Claus : Automatentheorie und formale Sprachen III
B.I.-Hochschulskripten 823/823a

[2] R.Kemp : LR(k)-Analysatoren
Dissertation an der Universität des
Saarlandes 1973

[3] D.E.Knuth : On the translation of languages from
left to right.
Information an Control 8 , 607-639 (1965)

Dr.Herbert Kopp
Institut für Angewandte Mathematik und Informatik
der Universität des Saarlandes
66 Saarbrücken 11
Im Stadtwald
Bundesrepublik Deutschland