év3

Kriterien für zustandsbeschränkte Grammatiken nach Earley

Diplomarbeit

eingereicht bei:
Prof. Dr. R. Kemp
SADS am Fachbereich Informatik
J.-W.-Goethe Universität Frankfurt

vorgelegt von:
Michael Klöckner
Eschersheimer Landstr. 218
6000 Frankfurt am Main 1

22. November 2001

Hiermit versichere ich, Michael Klöckner, die Diplomarbeit

Kriterien für zustandsbeschränkte Grammatiken nach Earley

selbständig und ohne andere als die hier angegebene Hilfe angefertigt zu haben. Alle Quellen und Hilfsmittel sind vollständig im Litaraturverzeichnis aufgelistet.

Ort, Datum.............................. Unterschrift..............................

Ich möchte mich für die anregende Betreuung bei Herrn R. Kemp bedanken, der mir bei komplizierten Sachverhalten die entscheidenden Hinweise gab. Mein Dank gilt ebenso meiner Freundin Caroline Geiger, die als Romanistin dennoch nicht müde wurde meinen Diskussionen zu folgen. Nicht zuletzt bedanke ich mich bei meiner Mutter, die mir das Studium und damit diese Arbeit ermöglichte.

Inhaltsverzeichnis

Kapitel 1

Einführung und Definitionen

1.1 Einführung

Die vorliegende Arbeit befaßt sich mit dem Laufzeitverhalten einer Parsing-Methode für allgemeine kontextfreie Grammatiken, wie sie von J. Earley[1] 1968 veröffentlicht wurde. Das Earley-Parse-Verfahren (**EPV**) ist neben dem Cocke-Younger-Kasami-Algorithmus[2] (**CYKA**) eine allgemeine Lösung des Wortproblems für kontextfreie Grammatiken. In einigen Darstellungen[3] wird die Nähe des **EPV** zur Methode der Berechnung der 'recognition matrix' im **CYKA** deutlich. So zeigt Valiant[4], daß jede kontextfreie Grammatik in der Zeit $O(n^{2.81})$ zu erkennen ist, indem er den **CYKA** zurückführt auf Methoden der Multiplikation allgemeiner und boolscher Matrizen und die Berechnung ihrer transitiven Hüllen. Wir sind jedoch an der unteren Schranke des **EPV** interessiert, die für zustandsbeschränkte[5] Grammatiken bei $O(n)$ liegt und folgen in der vorliegenden Diplomarbeit der Darstellung von Earley. Wir werden uns mit den Bedingungen befassen, die verhindern, daß eine Grammatik in linearer Zeit mittels des **EPV** zu erkennen ist.

[1] in [Ear]

[2] siehe [You] und [Kas] oder [Aho] S. 315

[3] in [Rév] S. 143 und [Arb] S. 72

[4] in [Har] S. 442

[5] siehe Definition *1.12* S. 12

Die Arbeit ist wie folgt aufgebaut: Kapitel 1 beginnt mit einer Einführung und beinhaltet Definitionen, die Beschreibung des **EPV**-Algorithmus, sowie die Motivation der Arbeit. Um die Ergebnisse auf allgemeine kontextfreie Grammatiken übertragen zu können, stellt Kapitel 2 den Zusammenhang zwischen Grammatiken in allgemeiner und Chomsky-Normal-Form dar. Kapitel 3 befaßt sich mit den Kriterien nichtzustandsbeschränkter Grammatiken und das letzte Kapitel erläutert und bewertet die gewonnenen Ergebnisse.

1.2 Definitionen

In diesem Abschnitt sollen die grundlegenden Definitionen zusammengestellt werden.

Definition 1.1: Eine *kontextfreie Grammatik* (CFG) ist ein 4-Tupel $\mathbf{G}{:=}(V, T, S, P)$, wobei

- V eine endliche Menge von Nonterminals,
- T eine endliche Menge von Terminals mit $V \cap T = \emptyset$,
- S ein ausgewähltes Startsysmbol aus V und
- P eine endliche Teilmenge von $V \times (V \cup T)^\star$ ist.

Statt $f = (A, \alpha) \in P$ schreiben wir $f : A \to \alpha \in P$. Wir werden kleine Buchstaben vom Anfang (bzw. Ende) des Alphabetes für Elemente aus T (bzw. $T^\star$) verwenden. Große Buchstaben vom Anfang (bzw. Ende) des Alphabetes bezeichnen Elemente aus V (bzw. $V^\star$). Griechische Buchstaben bezeichnen Elemente aus $(V \cup T)^\star$ und ϵ das leere Wort. Die Kleinbuchstaben i, j, k, l, n, r, s, t bezeichnen Elemente aus $\mathcal{N}$,der Menge der natürlichen Zahlen.

Definition 1.2: Sei $\mathbf{G} := (V, T, S, P)$ eine CFG. Die *Ableitungs-Relation* $\Rightarrow \;\subseteq (V \cup T)^\star \times (V \cup T)^\star$ ist definiert durch:

$$\alpha \Rightarrow \alpha' \quad \leftrightarrow \quad \alpha = \alpha_1 B \alpha_2, \quad \alpha' = \alpha_1 \beta \alpha_2, \quad f : B \to \beta \in P.$$

Ist α_1 (bzw. α_2) $\in T^\star$, so schreiben wir $\alpha \Rightarrow_l \alpha'$ (bzw. $\alpha \Rightarrow_r \alpha'$). Ein Wort $\alpha \in (V \cup T)^\star$ heißt *Satzform* (bzw. *Links-Satzform*, bzw. *Rechts-Satzform*), falls $S \Rightarrow^\star \alpha$ (bzw. $S \Rightarrow_l^\star \alpha$; bzw. $S \Rightarrow_r^\star$

α); Hierbei sind $\Rightarrow^\star$ (bzw. $\Rightarrow_l^\star$; bzw. $\Rightarrow_r^\star$) die reflexiv-transitiven Hüllen der Ableitungs-Relationen $\Rightarrow$ (bzw. $\Rightarrow_l$; bzw. $\Rightarrow_r$). Ist $S \Rightarrow \alpha_1 \Rightarrow \alpha_2 \Rightarrow ... \Rightarrow \alpha_n = \alpha$ (bzw. $S \Rightarrow_l \alpha_1 \Rightarrow_l \alpha_2 \Rightarrow_l ... \Rightarrow_l \alpha_n = \alpha$; bzw. $S \Rightarrow_r \alpha_1 \Rightarrow_r \alpha_2 \Rightarrow_r ... \Rightarrow_r \alpha_n = \alpha$) dann heißt $< S, \alpha_1, \alpha_2, ..., \alpha_n >$ die *Ableitung* (bzw. *Links-Ableitung*; bzw. *Rechts-Ableitung*) von α aus S; n ist die Länge der Ableitung.

Sei im folgenden $\mathbf{G}$:=(V, T, S, P) stets eine CFG.

Definition 1.3: Die von $\mathbf{G}$ erzeugte Sprache ist gegeben durch:
$L(\mathbf{G}) := \{w \in T^\star \,|\, S \Rightarrow^\star w\}$.
Zwei Grammatiken $\mathbf{G}$ und $\mathbf{G}'$ heißen *äquivalent* $\leftrightarrow L(\mathbf{G}) = L(\mathbf{G}')$.

Definition 1.4: $\mathbf{G}$ heißt *reduziert* $\leftrightarrow (\forall \alpha \in (V \cup T))(\exists \gamma, \delta \in (V \cup T)^\star)$ $(S \Rightarrow^\star \gamma\alpha\delta \wedge \alpha \Rightarrow^\star w \in T^\star) \vee P = \emptyset$.

Definition 1.5: $\mathbf{G}$ heißt *kettenregelfrei* $\leftrightarrow (\forall f : A \to \alpha \in P)(\alpha \notin V)$.
Falls das leere Wort in $L(\mathbf{G})$ liegt, so enthält $\mathbf{G}$ die zusätzlichen Produktionen $f_0 : S' \to S$ und $f_1 : S' \to \epsilon$, wobei das neue Startsymbol S' in keiner anderen Produktion vorkommt.

Definition 1.6: $\mathbf{G}$ heißt *ϵ-frei* $\leftrightarrow (\forall f : A \to \alpha \in P)(\alpha \neq \epsilon)$.
Falls das leere Wort in $L(\mathbf{G})$ liegt, so enthält $\mathbf{G}$ die zusätzlichen Produktionen $f_0 : S' \to S$ und $f_1 : S' \to \epsilon$, wobei das neue Startsymbol S' in keiner anderen Produktion vorkommt.

Definition 1.7: $\mathbf{G}$ hat Chomsky-Normal-Form (CNF) $\leftrightarrow$
$(\forall f : A \to \alpha \in P)(\alpha \in T \setminus \{\epsilon\} \cup V^2)$.
Falls das leere Wort in $L(\mathbf{G})$ liegt, so enthält $\mathbf{G}$ die zusätzlichen Produktionen $f_0 : S' \to S$ und $f_1 : S' \to \epsilon$, wobei das neue Startsymbol S' in keiner anderen Produktion vorkommt.

Definition 1.8: Ein Nonterminal $A \in V$ heißt :

$$\begin{array}{llllll}
\textit{rechtsrekursiv} & \leftrightarrow & \exists\gamma & \in (V \cup T)^\star & \text{mit } A \Rightarrow^\star \gamma A \,, \\
\textit{echt rekursiv} & \leftrightarrow & \exists\gamma, \delta & \in (V \cup T)^\star & \text{mit } A \Rightarrow^\star \gamma A \delta \,, \\
\textit{linksrekursiv} & \leftrightarrow & \exists\delta & \in (V \cup T)^\star & \text{mit } A \Rightarrow^\star A \delta \,.
\end{array}$$

Definition 1.9: $\mathbf{G}$ heißt *eindeutig*, falls jedes Wort $w \in L(\mathbf{G})$ nur eine Linksableitung hat; sonst heißt $\mathbf{G}$ *mehrdeutig*.

Definition 1.10: Mit $|w|$ bezeichnen wir die *Länge* von $w \in T^\star$. Das leere Wort ϵ hat die Länge 0.

1.3 Beschreibung des Earley-Parse-Verfahrens

Informell beschrieben arbeitet das **EPV** wie folgt: Als Eingabe erhält es eine CFG und ein Wort. Für jedes Terminalzeichen im Wort wird eine Liste aufgebaut, die unter Verwendung der vorangehenden Listen den Stand der Ableitung beschreibt. Die Elemente der Listen nennen wir Items. Sie bestehen aus den Produktionen der Grammatik, angereichert mit einem Punkt als Markierung und einem Zeiger. Die Markierung gibt an, wie weit die rechte Seite der Produktion bereits erfolgreich abgearbeitet ist, und der Zeiger verweist auf die Liste, in der die Abarbeitung begann.

Definition 1.11: Ein Tupel $[A \to \alpha\boldsymbol{.}\beta, i]$ mit $\alpha, \beta \in (V \cup T)^\star, A \in V$, $i \in \mathcal{N}, f : A \to \alpha\beta \in P$ heißt *Item*.
Eine Menge von Items heißt *Itemliste*. Sie wird im folgenden mit I_j bezeichnet; die Menge aller Itemlisten bezüglich **G** bezeichnen wir mit $J^{\mathbf{G}}$. Für Itemlisten $I_i, I_j \in J^{\mathbf{G}}$ sind nach dem **EPV** die folgenden Operationen definiert:

1. Die Abbildung *Scanner*: $T \times J^{\mathbf{G}} \to J^{\mathbf{G}}$ vermöge:
   $Scanner(a, I_j) := \{[A \to \alpha a\boldsymbol{.}\beta, i] \mid [A \to \alpha\boldsymbol{.}a\beta, i] \in I_{j-1} \wedge j > 0\}$.
2. Die Abbildung *Predictor*: $J^{\mathbf{G}} \to J^{\mathbf{G}}$ vermöge:
   $Predictor(I_j) := \{[B \to \boldsymbol{.}\eta, j] \mid [A \to \alpha\boldsymbol{.}B\beta, i] \in I_j \wedge f : B \to \eta \in P\}$.
3. Die Abbildung *Completor*: $J^{\mathbf{G}} \to J^{\mathbf{G}}$ vermöge:
   $Completor(I_j) := \{[A \to \alpha B\boldsymbol{.}\beta, k] \mid [B \to \eta\boldsymbol{.}, i] \in I_j \wedge [A \to \alpha\boldsymbol{.}B\beta, k] \in I_i\}$.

Als Eingabe erhält das **EPV** eine kontextfreie Grammatik **G** und ein Wort $w \in T^\star$, um mit den drei Operationen *Scanner*, *Predictor* und *Completor* für jedes Terminal des Wortes $w = a_1 ... a_n$ eine Liste I_j für $j \in [0 : n]$ aufzubauen. Das **EPV** arbeitet wie folgt:

Algorithmus **EPV**[6]:

BEGIN
FOR j:= 0 TO n DO
    BEGIN
    IF j = 0 THEN
        $I_j := \{[S \rightarrow \bullet\alpha, 0] | f : S \rightarrow \alpha \in P\}$;
    ELSE
        $I_j := \mathit{Scanner}(a_j, I_j)$;
    $\tilde{I}_j := \emptyset$;
    WHILE $\tilde{I}_j \neq I_j$ *DO*
        BEGIN $\tilde{I}_j := I_j$;
            $I_j := I_j \cup \mathit{Completor}(I_j)$;
            $I_j := I_j \cup \mathit{Predictor}(I_j)$;
        END;
    END;
IF $[S \rightarrow \alpha\bullet, 0] \in I_n$ THEN $w \in L(\mathbf{G})$;
    ELSE $w \notin L(\mathbf{G})$;
END.

Die drei definierten Abbildungen haben folgende praktische Anschauung:
1. Die Operation *Scanner* sucht in der vorangehenden Liste I_{j-1} alle Items mit dem Punkt vor dem zu lesenden Terminal und verschiebt den Punkt über dieses Terminal.
2. Die Operation *Predictor* sucht in Liste I_j alle Items mit dem Punkt vor einem Nonterminal und nimmt für jede Produktion mit diesem Nonterminal auf der linken Seite ein Item als möglichen Nachfolger in die Liste I_j auf.
3. Die Operation *Completor* durchsucht die Liste I_j nach abgearbeiteten Produktionen - gekennzeichnet durch den Markierungspunkt am Ende - und sucht in der Liste, auf die der Zeiger verweist, nach möglichen Vorgängern - gekennzeichnet durch den Punkt vor dem Nonterminalzeichen, welches im ursprünglichen Item auf der linken Seite der Produktion stand. Das als Vorgänger in Betracht kommende Item wird übernommen, indem der Punkt

[6] zur Korrektheit siehe [Rév] und [Arb]

über das entsprechende Nonterminal verschoben wird.
Im folgenden bezeichnen wir mit $I_j(w)$ die vom **EPV** bezüglich $w \in T^\star$ aufgebauten Itemlisten. Die vorliegende Arbeit benutzt vor allem die Bedingungen, unter denen ein Item in einer Liste enthalten ist. Wir werden diese fundamentale Eigenschaft des **EPV** im folgenden Satz zeigen:

Satz 1.1 *Sei* **G** *eine CFG und seien* $J^{\mathbf{G}}$ *die bezüglich* **G** *vom* **EPV** *aufgebauten Itemlisten. Für ein Wort* $w = a_1...a_n \in T^\star$ *ist das Item* $[A \rightarrow \alpha \bullet \beta, i]$ *in Liste* $I_j(w) \in J^{\mathbf{G}} \quad \leftrightarrow$

(a) $(\exists \gamma, \delta \in (V \cup T)^\star)\ (S \Rightarrow^\star \gamma A \delta)$,

(b) $\gamma \Rightarrow^\star a_1...a_i$,

(c) $f : A \rightarrow \alpha\beta \in P$,

(d) $\alpha \Rightarrow^\star a_{i+1}...a_j$.

Beweis:[7]
"$\Rightarrow$" Wir beweisen diesen Teil durch Induktion über die Anzahl der Items, die in Liste $I_0(w)$, $I_1(w)$,...,$I_j(w)$ aufgenommen wurden, bevor $[A \rightarrow \alpha \bullet \beta, i]$ hinzu kam.

Induktionsverankerung: Jedes Item in Liste $I_0(w)$ hat die Form $[S \rightarrow \bullet \beta, 0]$ falls $f : S \rightarrow \beta \in P$ ist. Also ist $i = 0$, $\alpha = \epsilon$ und $S = A$, und für $\gamma = \delta = \epsilon$ gilt $S \Rightarrow^\star \gamma A \delta$ und somit:

(a) $(\exists \gamma, \delta \in (V \cup T)^\star)(S \Rightarrow^\star \gamma A \delta)$,

(b) $\gamma = \epsilon \Rightarrow^\star a_1...a_0 = \epsilon$,

(c) $f : A \rightarrow \alpha\beta \in P$,

(d) $\alpha = \epsilon \Rightarrow^\star a_{i+1}...a_j = \epsilon$.

Induktionsschritt: Nehmen wir an, $I_0(w)$ sei konstruiert und die Aussage gelte für alle Items in $I_i(w)$, $i \leq j$. Das Item $[A \rightarrow \alpha \bullet \beta, i]$ wird von einer der drei Operationen *Scanner*,*Predictor* oder *Completor* in Liste $I_j(w)$ aufgenommen.

[7] Beweis nach [Aho] S. 324 und [Kem]

Angenommen $[A \to \alpha \bullet \beta, i]$ wurde mittels des *Scanner* in Liste $I_j(w)$ aufgenommen: Also ist $\alpha = \alpha' a_j$ und $[A \to \alpha' \bullet a_j \beta, i]$ ist in Liste $I_{j-1}(w)$. Wegen der Induktionsannahme gilt $\alpha' \Rightarrow^\star a_{i+1}...a_{j-1}$ und es gibt Satzformen γ' und δ', für die $S \Rightarrow^\star \gamma' A \delta'$ gilt. Somit folgt $\alpha = \alpha' a_j \Rightarrow^\star a_{i+1}...a_j$ und mit $\gamma = \gamma'$ und $\delta = \delta'$ gilt die Aussage.

Angenommen $[A \to \alpha \bullet \beta, i]$ wurde mittels des *Predictor* in Liste $I_j(w)$ aufgenommen: Also ist $\alpha = \epsilon$ und $i = j$, und in Liste $I_j(w)$ gibt es ein Item $[B \to \eta \bullet A \eta', i]$. Wegen der Induktionsannahme folgt: es gibt Satzformen γ' und δ' und $S \Rightarrow^\star \gamma' B \delta'$ und $\eta \Rightarrow^\star a_{i+1}...a_j$ und $\alpha = \epsilon \Rightarrow^\star a_{j+1}...a_j = \epsilon$. Für $\gamma = \gamma' \eta$ und $\delta = \eta' \delta'$ folgt die Aussage.

Angenommen $[A \to \alpha \bullet \beta, i]$ wurde mittels des *Completor* in Liste $I_j(w)$ aufgenommen: Also ist $\alpha = \alpha' B$ für ein $B \in V$, und für ein $k \in [i : j]$ ist das Item $[A \to \alpha' \bullet B \beta, i]$ in Liste $I_k(w)$. Als Auslöser für den *Completor* muß ein Item $[B \to \eta \bullet, k]$ in $I_j(w)$ sein. Wegen der Induktionsannahme folgt $\eta \Rightarrow^\star a_{k+1}...a_j$ und $\alpha' \Rightarrow^\star a_{i+1}...a_k$ und somit gilt $\alpha = \alpha' B \Rightarrow^\star a_k...a_i...a_j$. Wegen der Induktionsannahme gibt es außerdem Satzformen γ' und δ', für die gilt $S \Rightarrow^\star \gamma' A \delta'$. Mit $\gamma = \gamma'$ und $\delta = \delta'$ folgt die Aussage und dieser Teil des Satzes ist bewiesen.

"$\Leftarrow$" Für die Rückrichtung des Beweises fassen wir die Voraussetzungen zusammen und definieren als Hilfsmittel eine korrekte Konfiguration :

> Eine Konfiguration $[\alpha, \beta, \gamma, \delta, A, i, j]$ heißt *korrekt* falls gilt:
> - $S \Rightarrow^\star \gamma A \delta$,
> - $\gamma \Rightarrow^\star a_1...a_i$,
> - $\alpha \Rightarrow^\star a_{i+1}...a_j$,
> - $f : A \to \alpha\beta \in P$.
>
> Ist $\chi = [\alpha, \beta, \gamma, \delta, A, i, j]$ eine korrekte Konfiguration und
> - $l_1(\chi) =$ die Länge einer kürzesten Ableitung $S \Rightarrow^\star \gamma A \delta$,
> - $l_2(\chi) =$ die Länge einer kürzesten Ableitung $\gamma \Rightarrow^\star a_1...a_i$,
> - $l_3(\chi) =$ die Länge einer kürzesten Ableitung $\alpha \Rightarrow^\star a_{i+1}...a_j$,
>
> dann bezeichnet $l(\chi) := l_1(\chi) + 2\,[l_2(\chi) + l_3(\chi) + j]$ den *Rang* von χ.

Mit diesen Definitionen beweisen wir folgende Behauptung mittels Induktion über den Rang einer Konfiguration:

Behauptung: Ist $[\alpha, \beta, \gamma, \delta, A, i, j]$ korrekt, so folgt:
$$[A \rightarrow \alpha \bullet \beta, i] \in I_j(w).$$

Induktionsverankerung: ei $l(\chi) = 0$. Also ist $l_1(\chi) = l_2(\chi) = l_3(\chi) = j = i = 0$. Wegen $l_3(\chi) = 0$ und $j = i = 0$ folgt $\alpha = \epsilon$. Wegen $l_2(\chi) = 0$ und $i = 0$ folgt $\gamma = \epsilon$. Wegen $l_1(\chi) = 0$ folgt $A = S$ und $\delta = \epsilon$. Also hat χ die Form $[\epsilon, \beta, \epsilon, \epsilon, S, 0, 0]$ und es bleibt zu zeigen, daß $[S \rightarrow \bullet \beta, 0]$ in Liste $I_0(w)$ ist. Dies folgt aber aus der Definition des **EPV**.

Induktionsschritt : Gelte die Aussage für alle korrekten Konfigurationen χ' mit Rang $l(\chi') \in \;]0 : r[$. Wir betrachten die korrekte Konfiguration $\chi = [\alpha, \beta, \gamma, \delta, A, i, j]$ in den drei Fällen, in denen α mit einem Terminal endet, mit einem Nonterminal endet, oder gleich ϵ ist.

<u>*Fall 1*: $\alpha = \alpha' a,\ a \in T$.</u>
Da $\alpha' a \Rightarrow^\star a_1 ... a_j$ ist, folgt $a = a_j$ und $\chi = [\alpha' a_j, \beta, \delta, \gamma, A, i, j]$ und wir wissen :

$$S \Rightarrow^\star \gamma A \delta\,,$$
$$\gamma \Rightarrow^\star a_1 ... a_i\,,$$
$$\alpha' \Rightarrow^\star a_{i+1} ... a_{j-1}\,,$$
$$f : A \rightarrow \alpha' a_j \beta \in P\,.$$

Also ist $\chi' = [\alpha', a_j\beta, \gamma, \delta, A, i, j-1]$ eine korrekte Konfiguration. Da

$$l_1(\chi') = l_1(\chi) \wedge l_2(\chi') = l_2(\chi) \wedge l_3(\chi') < l_3(\chi)$$

hat χ' den Rang

$$\begin{aligned} l(\chi') &< l_1(\chi) + 2\,[l_2(\chi) + l_3(\chi) + j - 1] \\ &= l(\chi) - 2 = r - 2 < r\,. \end{aligned}$$

Nach Induktionsvoraussetzung ist das Item $[A \rightarrow \alpha' \bullet a_j\beta, i]$ in Liste $I_{j-1}(w)$ und $Scanner(a, I_j(w))$ liefert das Item $[A \rightarrow \alpha \bullet \beta, i]$ in Liste $I_j(w)$.

<u>*Fall 2*: $\alpha = \alpha' B, B \in V$.</u>

χ hat also die Form $[\alpha' B, \beta, \gamma, \delta, A, i, j]$ und wir wissen:

$$S \Rightarrow^\star \gamma A \delta\,,$$
$$\gamma \Rightarrow^\star a_1 ... a_i\,,$$
$$f : A \to \alpha' B \beta \in P\,.$$

Wegen $\alpha = \alpha' B \Rightarrow^\star a_{i+1} ... a_j$ folgt:

$$(\exists k \in [i:j])(\alpha' \Rightarrow^\star a_{i+1} ... a_k \wedge B \Rightarrow^\star a_{k+1} ... a_j)\,.$$

Also ist $\chi' = [\alpha', B\beta, \gamma, \delta, A, i, k]$ eine korrekte Konfiguration. Da:

$$l_1(\chi') = l_1(\chi) \wedge l_2(\chi') = l_2(\chi) \wedge l_3(\chi') < l_3(\chi)$$

und da $k \leq j$ ist, folgt für den Rang von χ':

$$\begin{aligned} l(\chi') &< l_1(\chi) + 2\,[l_2(\chi) + l_3(\chi) + k] \\ &\leq l(\chi) \end{aligned}$$

und somit:

$$\begin{aligned} l(\chi') &< l_1(\chi) + 2\,[l_2(\chi) + l_3(\chi) + j] \\ &= l(\chi) = r\,. \end{aligned}$$

Nach Induktionsvoraussetzung ist das Item $[A \to \alpha' \bullet B\beta, i] \in I_k(w)$. Sei $B \Rightarrow \eta$ der erste Ableitungsschritt in einer Ableitung $B \Rightarrow^\star a_{k+1} ... a_j$. Wir wissen:

$$\begin{aligned} S \Rightarrow^\star \gamma A \delta \wedge f : A \to \alpha' B \beta \in P \quad &\rightsquigarrow \quad S \Rightarrow^\star \gamma \alpha' B \beta \delta\,, \\ \gamma \Rightarrow^\star a_1 ... a_i \wedge \alpha' \Rightarrow^\star a_{i+1} ... a_k \quad &\rightsquigarrow \quad \gamma\alpha' \Rightarrow^\star a_1 ... a_i ... a_k\,, \\ \eta \Rightarrow^\star a_{k+1} ... a_j\,. \end{aligned}$$

Also ist $\chi'' = [\eta, \epsilon, \gamma\alpha', \beta\delta, B, k, j]$ eine korrekte Konfiguration. Wir wissen:

$$S \Rightarrow^\star \gamma A \delta \Rightarrow \gamma \alpha' B \beta \delta \rightsquigarrow l_1(\chi'') \leq l_1(\chi) + 1\,.$$

Sei nun n_1 die Länge einer kürzesten Ableitung von $\alpha' \Rightarrow^\star a_{i+1} ... a_k$ und n_2 die Länge einer kürzesten Ableitung von $B \Rightarrow^\star a_{k+1} ... a_j$. Damit gilt:

$$\begin{aligned} l_3(\chi) &= n_1 + n_2\,, \\ \gamma \Rightarrow^\star a_1 ... a_i \wedge \alpha' \Rightarrow^\star a_{i+1} ... a_k \rightsquigarrow l_2(\chi'') &= l_2(\chi) + n_1\,, \\ B \Rightarrow \eta \Rightarrow^\star a_{k+1} ... a_j \rightsquigarrow l_3(\chi'') &= n_2 - 1\,. \end{aligned}$$

Also ist

$$\begin{aligned}
l(\chi'') &= l_1(\chi'') + 2\,[l_2(\chi'') + l_3(\chi'') + j] \\
&\leq l_1(\chi) + 1 + 2\,[l_2(\chi) + n_1 + n_2 - 1 + j] \\
&= l_1(\chi) + 1 + 2\,[l_2(\chi) + l_3(\chi) + j - 1] = l(\chi) - 1 \\
&= r - 1 < r\,.
\end{aligned}$$

Nach Induktionsvoraussetzung ist das Item $[B \to \eta\,\bullet\,, k] \in I_j(w)$. Zusammen mit dem Item $[A \to \alpha'\bullet B\beta, i] \in I_k(w)$ liefert *Completor* $(I_j(w))$ das Item $[A \to \alpha' B\,\bullet\,\beta, i]$ in Liste $I_j(w)$.

<u>*Fall 3*: $\alpha = \epsilon$.</u>
Es gilt : $\alpha = \epsilon \Rightarrow^\star a_{j+1}...a_j$. Daraus folgt $l_3(\chi) = 0 \wedge i = j$. Da $r > 0$ gilt, können wir folgern, daß die Ableitung $S \Rightarrow^\star \gamma A\delta$ mindestens die Länge 1 hat. Anderenfalls wäre $l_1(\chi) = 0$ und $\gamma = \epsilon$ und somit wäre $l_2(\chi) = i = 0$. Wegen $\alpha = \epsilon$ folgte $l_3(\chi) = 0$ und wegen $i = j = 0$ wäre der Rang $l(\chi) = 0$. Der Fall $r = 0$ wurde aber bereits in der Induktionsverankerung gezeigt. Also gilt:

$$(\exists \gamma', \gamma'', \delta', \delta'' \in (V \cup T)^\star)\ (S \Rightarrow^\star \gamma' B \delta' \Rightarrow \gamma'\gamma'' A \delta'' \delta')$$

Wir setzen $\gamma := \gamma'\gamma''$ und $\delta := \delta''\delta'$ und erhalten:

$$\begin{aligned}
&S \Rightarrow^\star \gamma' B \delta' \Rightarrow \gamma'\gamma'' A \delta''\delta' = \gamma A \delta \wedge \gamma = \gamma'\gamma'' \Rightarrow^\star a_1...a_i \\
&\rightsquigarrow (\exists k \in [1:i] = [1:j])\ (\gamma' \Rightarrow^\star a_1...a_k \wedge \gamma'' \Rightarrow^\star a_{k+1}...a_j).
\end{aligned}$$

Wir wissen außerdem: $f' : B \to \gamma'' A \delta'' \in P$. $\chi' = [\gamma'', A\delta'', \gamma', \delta', B, k, j]$ ist also eine korrekte Konfiguration. Sei n_1 (bzw. n_2) die Länge einer kürzesten Ableitung $\gamma' \Rightarrow^\star a_1...a_k$ (bzw. $\gamma'' \Rightarrow^\star a_{k+1}...a_j$). Da

$$l_1(\chi') = l_1(\chi) - 1 \wedge l_2(\chi') = n_1 \wedge l_3(\chi') = n_2 \wedge l_2(\chi) = n_1 + n_2$$

berechnet sich der Rang $l(\chi')$ zu:

$$\begin{aligned}
l(\chi') &= l_1(\chi) - 1 + 2\,[n_1 + n_2 + j] \\
&\leq l(\chi) - 1 < r\,.
\end{aligned}$$

Nach Induktionsvoraussetzung ist das Item $[B \to \gamma''\bullet A\delta'', k] \in I_j(w)$. Die Operation *Predictor* $(I_j(w))$ liefert alle Items $[A \to \bullet\,\beta, j]$ in Liste $I_j(w)$. Damit ist **Satz 1.1** bewiesen. $\diamond$

Wir erhalten als Spezialfall von **Satz 1.1** folgendes Lemma:

Lemma 1.1 *Seien $I_j(w)$ die vom* **EPV** *bezüglich $w \in T^n$ aufgebauten Itemlisten. Es gilt:* $w \in L(\mathbf{G}) \leftrightarrow [S \to \alpha\bullet, 0] \in I_n(w)$.

Betrachten wir ein Beispiel:

Beispiel 1.1 Sei $\mathbf{G}_1 := (V, T, S, P)$ wobei gilt:

$$V := \{S, A, B\}, \qquad T := \{a, b\}, \qquad P := \{\ f_1 : S \to SA,$$
$$f_2 : S \to SB, \quad f_3 : A \to a, \quad f_4 : B \to a, \quad f_5 : S \to a\ \}.$$

Die Sprache $L(\mathbf{G}_1)$ betseht aus Wörtern der Form $a^k, k \in \mathcal{N}$. Betrachten wir die Itemlisten, die das **EPV** für das Wort $w_k := a^k$ für $k = 7$ aufbaut:

Liste $I_0(a^7)$	Liste $I_1(a^7)$	Liste $I_2(a^7)$	Liste $I_3(a^7)$
$[S \to \bullet SA, 0]$	$[S \to a\bullet, 0]$	$[A \to a\bullet, 1]$	$[A \to a\bullet, 2]$
$[S \to \bullet SB, 0]$	$[S \to S\bullet A, 0]$	$[B \to a\bullet, 1]$	$[B \to a\bullet, 2]$
$[S \to \bullet a, 0]$	$[S \to S\bullet B, 0]$	$[S \to SA\bullet, 0]$	$[S \to SA\bullet, 0]$
	$[A \to \bullet a, 1]$	$[S \to SB\bullet, 0]$	$[S \to SB\bullet, 0]$
	$[B \to \bullet a, 1]$	$[S \to S\bullet A, 0]$	$[S \to S\bullet A, 0]$
		$[S \to S\bullet B, 0]$	$[S \to S\bullet B, 0]$
		$[A \to \bullet a, 2]$	$[A \to \bullet a, 3]$
		$[B \to \bullet a, 2]$	$[B \to \bullet a, 3]$

Liste $I_4(a^7)$	Liste $I_5(a^7)$	Liste $I_6(a^7)$	Liste $I_7(a^7)$
$[A \to a\bullet, 3]$	$[A \to a\bullet, 4]$	$[A \to a\bullet, 5]$	$[A \to a\bullet, 6]$
$[B \to a\bullet, 3]$	$[B \to a\bullet, 4]$	$[B \to a\bullet, 5]$	$[B \to a\bullet, 6]$
$[S \to SA\bullet, 0]$	$[S \to SA\bullet, 0]$	$[S \to SA\bullet, 0]$	$[S \to SA\bullet, 0]$
$[S \to SB\bullet, 0]$	$[S \to SB\bullet, 0]$	$[S \to SB\bullet, 0]$	$[S \to SB\bullet, 0]$
$[S \to S\bullet A, 0]$	$[S \to S\bullet A, 0]$	$[S \to S\bullet A, 0]$	$[S \to S\bullet A, 0]$
$[S \to S\bullet B, 0]$	$[S \to S\bullet B, 0]$	$[S \to S\bullet B, 0]$	$[S \to S\bullet B, 0]$
$[B \to \bullet a, 4]$	$[B \to \bullet a, 5]$	$[B \to \bullet a, 6]$	$[B \to \bullet a, 7]$
$[A \to \bullet a, 4]$	$[A \to \bullet a, 5]$	$[A \to \bullet a, 6]$	$[A \to \bullet a, 7]$

Da die Items $[S \to SA\bullet, 0]$ und $[S \to SB\bullet, 0]$ in Liste $I_7(a^7)$ sind, ist das

Wort $a^7 \in \mathrm{L}(\mathbf{G})$.

Ausschlaggebend für das Laufzeitverhalten des **EPV** ist die maximale Anzahl der Items pro Liste, denn die Operation *Completor* muß Itemlisten vollständig nach möglichen Vorgängern durchsuchen. Wir definieren daher Grammatiken, deren Itemlisten beschränkt sind wie folgt:

Definition 1.12: Eine Grammatik **G** heißt *zustandsbeschränkt*, falls gilt:

$$(\exists n \in \mathcal{N})\,(\forall w \in L(\mathbf{G}))\,(\forall j \in [0 : |w|])\,(|I_j(w)| \leq n)\,.$$

Anderenfalls heißt **G** *nichtzustandsbeschränkt.*

Mit **NIB** bezeichnen wir die Menge aller kontextfreien Grammatiken, die *nichtzustandsbeschränkt* sind.

1.4 Motivation

Mit der Laufzeit des **EPV** meinen wir die Anzahl der Schritte, die, bei einer geeigneten Implementierung der Listenoperationen, nötig sind, um zu entscheiden, ob das Eingabewort w von der Grammatik **G** erzeugt wird. Für zustandsbeschränkte Grammatiken benötigt das **EPV** konstant viele Schritte, um ein Item in eine Liste einzuordnen[8].

> Die Laufzeit, die das **EPV** für **zustandsbeschränkte Grammatiken** $\mathbf{G} \notin$ **NIB** benötigt, hängt nur von der Länge des Eingabewortes w ab und wächst somit **linear**. (1.1)

Da es für jede Grammatik **G** nur k viele linke Itemkomponenten $[A \to \alpha\,.\,\beta,$ gibt, kann Liste $I_j(w)$ auf maximal $k(j+1)$ Items anwachsen. Falls die Zeit, die je Item in Liste $I_j(w)$ verwendet wird, konstant ist, so ergibt sich für **nichtzustandsbeschränkte Grammatiken als untere Schranke**:

[8] zum Beweis siehe [Ear] S. 98

Sei c_1 die Zeit, die pro Item in jeder Liste verwendet wird, so ist die Laufzeit des **EPV** für ein Wort der Länge n: $c_1 \sum_{j=0}^{n} k(j+1) = \frac{1}{2} k c_1 (n+1)(n+2) \leq c_2 n^2 \,, c_2 \in \mathcal{N}\,.$ (1.2)

Aho[9] zeigt, daß ein Item $[A \to \alpha\bullet\beta, i]$ in Liste $I_j(w)$ auf genau eine Weise gelangt, falls **G** eindeutig ist und $\alpha \neq \epsilon$ gilt. Somit lassen sich eindeutige kontextfreie Grammatik in $O(n^2)$ Schritten erkennen. Earley[10] weist darauf hin, daß dies sogar für einige endlich mehrdeutige Grammatiken gilt. Falls die Grammatik **G** sowohl mehrdeutig als auch nichtzustandsbeschränkt ist, sodaß für jedes der $O(j)$ Items in Liste $I_j(w)$ auch $O(j)$ Schritte zum Einfügen verwendet werden, benötigt das **EPV** **als obere Schranke** die Zeit $O(n^3)$, um ein Eingabewort der Länge n zu erkennen.

Im allgemeinen Fall benötigt das **EPV** $O(n^3)$ Schritte um das Wortproblem für ein Wort der Länge n zu lösen. (1.3)

Die Klasse der zustandsbeschänkten kontextfreien Grammtiken ist bisher kaum erforscht. Bekannt[11] ist, daß **LR**(k) Grammatiken bei k Zeichen look-ahead in linearer Zeit zu analysieren sind. Andererseits gibt es unendlich mehrdeutige Grammtiken, die zustandsbeschränkt sind. $\mathbf{G}_1$ aus **Beispiel 1.1** ist sicher unendlich mehrdeutig, denn das Wort a^k hat 2^k Linksableitungen. Die Liste $I_k^{\mathbf{G}_1}(a^k)$ enthält jedoch nur die folgenden acht Items[12]:

[9] in [Aho] S. 325

[10] in [Ear] S. 99

[11] siehe [Aho] S. 398

[12] Beweis durch Induktion über die Länge k des Wortes w_k

Liste $I_k(a^k)$
$[A \to a\bullet, k-1]$
$[B \to a\bullet, k-1]$
$[S \to SA\bullet, 0]$
$[S \to SB\bullet, 0]$
$[S \to S\bullet A, 0]$
$[S \to S\bullet \mathrm{B}, 0]$
$[A \to \bullet a, k]$
$[B \to \bullet a, k]$

Das Ziel dieser Arbeit ist es, die Klasse der zustandsbeschränkten kontextfreien Grammatiken zu beschreiben. Hierzu werden wir versuchen, das Kompliment der zustandsbeschränkten Grammatiken, also die Menge **NIB** zu bestimmen. Wir werden nichttriviale Eigenschaften sammeln, die das Wachstum der Listen bedingen und deshalb ein lineares Laufzeitverhalten verhindern. Wir können die Menge **NIB** allerdings nicht vollständig charakterisieren und somit nicht alle zustandsbeschränkten Grammatiken klassifizieren. Jedoch werden wir einige Kriterien für Grammatiken der Menge **NIB** vorstellen.

Kapitel 2

Normalform und Itemlisten

2.1 Allgemeines

Die Beweise der Kriterien nichtzustandsbeschränkter Grammatiken in Kapitel 3 beziehen sich auf reduzierte Grammatiken in CNF. Um die gewonnenen Ergebnisse auf allgemeine kontextfreie Grammatiken übertragen zu können, ist es von Interesse, die bei den Umformungen allgemeiner Grammatiken in CNF entstehenden Auswirkungen auf das **EPV** genauer zu betrachten. In diesem Kapitel werden wir den Zusammenhang von Itemlisten einer allgemeinen kontextfreien Grammatik $\mathbf{G}$ und einer äquivalenten reduzierten Grammatik $\mathbf{G}'$ in CNF beschreiben. Das Ergebnis dieses Kapitels läßt sich wie folgt formulieren:

Satz 2.1 *Seien $I_j^{\mathbf{G}}(w)$ und $I_j^{CNF}(w)$ die vom* **EPV** *bezüglich $w \in T^\star$ aufgebauten Itemlisten der Grammatik* $\mathbf{G}$ *und einer äquivalenten reduzierten Grammatik in CNF. Es gilt:*

$$(\exists c \in \mathcal{N})\,(\forall w \in L(\mathbf{G})\,)\,(\forall j \in [\,0: \,|w|\,]\,)\,(\,|\,I_j^{CNF}(w)\,| \,\leq\, c|\,I_j^{\mathbf{G}}(w)\,|\,)\,.$$

Wir betrachten zunächst die Operation *Predictor*. Sie erzeugt maximal so viele Items, wie die Menge P Produktionen enthält. Dies geschieht dann, wenn sich in Liste I_j für jedes Nonterminal $B \in V$ mindestens ein Item $[A \to \alpha \bullet B\beta, i]$ befindet. Die Operation *Predictor* kann also kein Wachstum der Listen hervorrufen. Falls für ein $c_1 \in \mathcal{N}$ gilt: $|P^{CNF}| \leq c_1|P^{\mathbf{G}}|$, so folgt für die Abschätzung des Mehraufwandes, der bei den Umformungen von einer CFG **G** in eine äquivalente reduzierte Grammatik in CNF entsteht:

$$|Predictor\,(I_j^{CNF}(w))| \leq c_1 |Predictor\,(I_j^{\mathbf{G}}(w))|\,. \tag{2.1}$$

Bezüglich der Operation Scanner wissen wir, daß sie maximal so viele Items in Liste I_j erzeugt, wie sich in der Vorgängerliste I_{j-1} befanden. Dies geschieht dann, wenn sich in I_{j-1} nur Items $[A \to \alpha \bullet a_j\beta, i]$ befinden. Die Oparation *Scanner* kann ebenfalls kein Wachstum der Listen erzeugen. Falls für ein $c_2 \in \mathcal{N}$ gilt:

$$|Completor\,(I_j^{CNF}(w))| \leq c_2 |Completor\,(I_j^{\mathbf{G}}(w))|\,, \tag{2.2}$$

so folgt mit der Aussage (2.1) der **Satz 2.1** . Wir müßen also die Operation *Completor* und die Menge P unter den Umformungen genaueren betrachten, um **Satz 2.1** zu beweisen.

2.2 Überflüßige Nonterminals

In diesem Abschnitt werden wir zu **G** eine äquivalente CFG konstruieren, die keine überflüßigen Nonterminalzeichen enthält.

Lemma 2.1 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine CFG. Es gibt eine CFG* $\mathbf{G}' := (V', T, S, P')$ *mit* $L(\mathbf{G}) = L(\mathbf{G}') \wedge (\forall A \in V')\,(A \Rightarrow^\star w \in T^\star)$.

Beweis:[1] Wir konstruieren die Hilfsmenge $\mathcal{U}$:

1. $\mathcal{U}_1 := \{A \in V \mid f : A \to w \in P, w \in T^\star\}\,.$
2. $\mathcal{U}_{i+1} := \mathcal{U}_i \cup \{A \in V \mid f : A \to \alpha \in P, \alpha \in (\mathcal{U}_i \cup T)^\star\}\,.$

[1] Beweis nach [Sud] S. 107

Da $\mathcal{U}_i \subseteq \mathcal{U}_{i+1} \subseteq V$ ist, gilt:

$$(\exists l \in \mathcal{N})\,(\forall j \in \mathcal{N})\,(\mathcal{U}_{l+j} = \mathcal{U}_l)\,.$$

Wir setzen $\mathcal{U} := \mathcal{U}_l$. Es folgt:

$$(\forall A \in \mathcal{U})\ (A \Rightarrow^\star w \in T^\star)\,.$$

Behauptung: $(\forall A \in V)\,(A \Rightarrow^\star w \in T^\star) \rightsquigarrow A \in \mathcal{U}$.

Beweis: durch Induktion über die Länge k der Ableitung $A \Rightarrow^\star w \in T^\star$.

$\underline{k := 1\ \ }$ Also ist mit 1. $f : A \to w \in P$. Gelte die Behauptung für Ableitungen der Länge kleiner k:

$\underline{(k-1) \to k\ \ }$ Sei $A \Rightarrow^\star \alpha_1\alpha_2 \cdots \alpha_n \Rightarrow^\star w$ eine Ableitung der Länge k. Wir nehmen an $w := w_1 w_2 \cdots w_n$ mit $\alpha_i \Rightarrow^\star w_i$ für $i \in [1 : n]$. Falls α_i ein Nonterminal ist, befindet sich α_i bereits in $\mathcal{U}$, da die Ableitung $\alpha_i \Rightarrow^\star w_i$ von einer Länge kleiner k ist; nach Vorschrift 2. wird dann aber auch A in $\mathcal{U}$ aufgenommen und die Behauptung ist bewiesen.

Wir konstruieren $\mathbf{G}' := (V', T, S, P')$ wie folgt :

1. $V' := \mathcal{U}$.

2. $P' := \{f : A \to \alpha \in P \mid A, \alpha \in (V' \cup T)^\star\}$.

Es folgt: $L(\mathbf{G}') \subseteq L(\mathbf{G})$. Falls es ein Wort $w \in L(\mathbf{G})$ gibt, das nicht in $L(\mathbf{G}')$ ist, so muß an der Ableitung von w ein Nonterminal aus $V \setminus V'$ oder eine Produktion aus $P \setminus P'$, also ebenfalls ein Nonterminal aus $V \setminus V'$ beteiligt gewesen sein. Dann gäbe es ein Nonterminal $A \notin \mathcal{U}$ mit $A \Rightarrow^\star w \in T^\star$; ein Widerspruch zur Behauptung. Es folgt $L(\mathbf{G}) \subseteq L(\mathbf{G}')$ und das Lemma ist bewiesen. $\diamond$

Durch die Konstruktion nach **Lemma 2.1** wird der Grammatik $\mathbf{G}'$ keine Produktion und kein Nonterminalzeichen hinzugefügt; also gilt:

$$|\,V'\,| \leq |\,V\,| \quad \text{und} \quad |\,P'\,| \leq |\,P\,|\,. \tag{2.3}$$

Somit folgt für die Itemlisten:

$$(\forall w \in L(\mathbf{G}))\,(\forall j \in [\,0 : \,|w|\,])\,(\,|\,I_j^{\mathbf{G}'}(w)\,| \;\le\; |\,I_j^{\mathbf{G}}(w)\,|\,)\,. \tag{2.4}$$

Im weiteren Verlauf der Arbeit gehen wir davon aus, daß die Grammtik **G** bereits keine überflüssigen Nonterminals enthält.

Im folgenden Abschnitt werden wir zu **G** eine äquivalente CFG konstruieren, die keine überflüßigen Terminalzeichen enthält.

2.3 Überflüßige Terminals

Lemma 2.2 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine CFG. Es gibt eine CFG* $\mathbf{G}' := (V', T', S, P')$ *mit* $L(\mathbf{G}) = L(\mathbf{G}') \wedge (\forall \alpha \in (V' \cup T'))\ (S \Rightarrow^\star \gamma\alpha\delta)$.

Beweis:[2] Wir konstruieren V', T' und P' wie folgt:

1. Nehme S in V' auf.

2. $(\forall A \in V')\ (f : A \to \alpha_1\alpha_2 \cdots \alpha_n \in P)\ (i \in [\,1 : n\,])\ (\alpha_i \in (V \cup T))$:

   Falls $\alpha_i \in V$, so nehme α_i in V' auf.

   Falls $\alpha_i \in T$, so nehme α_i in T' auf.

3. $P' := \{f : A \to \alpha \in P \mid A, \alpha \in (V' \cup T')^\star\}$.

Da P nur endlich viele Produktionen hat, bricht Schritt 2. und 3. ab. Nach Konstruktion von V', T' und P' ist jedes Zeichen aus $(V' \cup T')$ in einer Ableitung vom Startsymbol S aus zu erreichen. Es gilt für die Sprachen $L(\mathbf{G}') \subseteq L(\mathbf{G})$ und wir wissen :

$$(\forall \alpha \in (V \cup T))\ (S \Rightarrow^\star \gamma\alpha\delta \Rightarrow^\star w \in T^\star) \rightsquigarrow (\alpha \in (V' \cup T')^\star)\,.$$

Also waren zur Erzeugung von $L(\mathbf{G})$ nur Produktionen aus P' beteiligt; somit folgt $L(\mathbf{G}) \subseteq L(\mathbf{G}')$ und das **Lemma 2.2** ist bewiesen. $\diamond$

[2]Beweis nach [Sud] S. 109

Weder V' noch T' noch P' sind gewachsen und es folgt:

$$|\,V'\,| \;\leq\; |\,V\,| \qquad \text{und} \qquad |\,P'\,| \;\leq\; |\,P\,|\,. \tag{2.5}$$

Für die Itemlisten von $\mathbf{G}'$ gilt also:

$$(\forall w \in L(\mathbf{G})\,)\,(\forall j \in [\,0 :\, |w|\,]\,)\,(\,|\,I_j^{\mathbf{G}'}(w)\,| \;\leq\; |\,I_j^{\mathbf{G}}(w)\,|\,)\,. \tag{2.6}$$

Im weiteren Verlauf der Arbeit gehen wir davon aus, daß die Grammtik $\mathbf{G}$ bereits keine überflüssigen Terminals enthält.

2.4 ϵ-Freiheit

In diesem Abschnitt werden wir zu $\mathbf{G}$ eine äquivalente CFG konstruieren, die keine ϵ-Produktionen enthält.

Lemma 2.3 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine CFG. Es gibt eine CFG* $\mathbf{G}' := (V, T, S, P')$ *ohne* ϵ*-Regeln, für die gilt:* $L(\mathbf{G}') = L(\mathbf{G}) \setminus \{\epsilon\}$.

Beweis:[3] Wir konstruieren die Hilfsmenge $\mathcal{U}$:

1. $\mathcal{U}_1 \;:=\; \{A \in V \mid f : A \to \epsilon \in P\}\,.$
2. $\mathcal{U}_{i+1} \;:=\; \mathcal{U}_i \cup \{A \in V | f : A \to \alpha_1\alpha_2 \cdots \alpha_n \in P, \alpha_k \in \mathcal{U}_i, 1 \leq\, k \leq n\}\,.$

Da $\mathcal{U}_i \subseteq \mathcal{U}_{i+1} \subseteq V$ ist, gilt:

$$(\exists l \in \mathcal{N})\,(\forall j \in \mathcal{N})\,(\mathcal{U}_{l+j} = \mathcal{U}_l)\,.$$

Wir setzen $\mathcal{U} := \mathcal{U}_l$ und wissen:

$$(\forall A \in V)\,(A \in \mathcal{U} \leftrightarrow A \Rightarrow^\star \epsilon) \wedge (\epsilon \in L(\mathbf{G}) \leftrightarrow S \in \mathcal{U})\,.$$

Wir konstruieren P' :

> Für jede Produktion $f : A \to \alpha \in P$ nehmen wir alle Produktionen $f' : A \to \alpha'$ in P' auf, falls man aus α die Satzform α' erhält, indem man aus α beliebig viele Vorkommen von Elementen aus $\mathcal{U}$ streicht und $\alpha' \neq \epsilon$ gilt. (2.7)

[3] Beweis nach [Sud] S. 97 und [Sal] S. 54

Es folgt: $L(\mathbf{G}') \subseteq L(\mathbf{G})$. Da man jede Ableitung für ein Wort $w \in L(\mathbf{G})$ und $w \neq \epsilon$ durch Produktionen aus P' simulieren kann, folgt für die beiden Sprachen $L(\mathbf{G}) \setminus \{\epsilon\} \subseteq L(\mathbf{G}')$ und das Lemma ist bewiesen. $\diamond$

Falls $\epsilon \in L(\mathbf{G})$ ist, so konstruieren wir $\mathbf{G}'$ wie oben und fügen zu P' die Produktionen $f_0 : S' \to S$ und $f_1 : S' \to \epsilon$ und zu V' das neue Startsymbol S' hinzu, wobei S' in keiner anderen Produktion vorkommt.

Betrachten wir die Auswirkungen auf die Größe von $\mathbf{G}'$. Im schlechtesten Fall sind alle Produktionen von der Form: $f : A \to \alpha_1 B_1 \alpha_2 B_2 \cdots \alpha_r B_r \alpha_{r+1}$ mit $B_l \in \mathcal{U}$ für $l \in [1 : r]$ und $\alpha_l \in (V \setminus \mathcal{U} \cup T)^\star$ für $l \in [1 : r+1]$. Sei r die maximale Anzahl von Elementen aus $\mathcal{U}$ in den rechten Seiten der Produktionen von P. Nach (2.7) besteht die Möglichkeit beliebig viele der B_l zu streichen. Sei $\mathcal{M}_f$ die Menge der für $f \in P$ nach (2.7) erzeugten Produktionen f'. Also ist die Kardinalität:

$$\begin{aligned} |\mathcal{M}_f| &= \binom{r}{0} + \binom{r}{1} + \cdots + \binom{r}{r-1} + \binom{r}{r} \\ &= \sum_{l=0}^{r} \binom{r}{l} = 2^r \,. \end{aligned}$$

für jede Produktionen $f : A \to \alpha_1 B_1 \alpha_2 \cdots B_r \alpha_{r+1} \in P$ mit $B_l \in \mathcal{U}$ und mindestens ein $\alpha_l \neq \epsilon$. Die Menge P' enthält also maximal $2^r\,|\,P\,|$ viele Produktionen, falls $\epsilon \notin L(\mathbf{G})$ ist und sonst 2 Regeln mehr. Die Menge der Nonterminals wird nur um das neue Sartsymbol S' vergrößert und es gilt:

$$|\,P'\,| \leq 2^r\,|\,P\,| + 2 \leq 2^{r+1}\,|\,P\,| \quad \text{und} \quad |\,V'\,| \leq |\,V\,| + 1\,. \tag{2.8}$$

Betrachten wir die Auswirkungen der Umformungen nach **Lemma 1.3**, bei denen die ϵ-Regeln entfernt wurden, auf die Itemlisten. Sei das Terminalwort w erzeugt worden mit der Ableitung:

$$S \Rightarrow^\star \gamma_1 A \delta_1 \Rightarrow \gamma_1 \alpha_1 B_1 \alpha_2 \cdots B_r \alpha_{r+1} \delta_1 \Rightarrow^\star a_1 \cdots a_n = w \in T^\star\,.$$

Wir betrachten nur Regeln $f : A \to \alpha \in P$, für die Satzformen γ_1 und δ_1 existieren, sodaß gilt:

$$S \Rightarrow^\star \gamma_1 A \delta_1 \,, \tag{2.9}$$

$$\gamma_1 \Rightarrow^\star p_1 = a_1 \cdots a_s \in T^\star \quad \wedge \quad \delta_1 \Rightarrow^\star q_1 \in T^\star \,. \tag{2.10}$$

Weil **G** reduziert ist, gibt es für die Produktion

$$f : A \to \alpha_1 B_1 \alpha_2 \cdots B_r \alpha_{r+1} \in P,\, B_l \in \mathcal{U},\, 1 \leq l \leq r+1 \tag{2.11}$$

Wörter $p_l \in T^\star, 1 \leq l \leq r+1$ und Zahlen $t_l \in [s:n]$ mit:

$$\alpha_1 B_1 \cdots B_{l-1} \alpha_l \Rightarrow^\star p_l = a_{s+1} \cdots a_{t_l} \in T^\star,\, 1 \leq s \leq t_l \leq n \,. \tag{2.12}$$

Sei nun $f\epsilon : B_l \to \epsilon \in P$ eine ϵ-Regel in einer Ableitung

$$\gamma_1 A \delta_1 \Rightarrow \gamma_1 \alpha_1 B_1 \alpha_2 \cdots B_r \alpha_{r+1} \delta_1 \Rightarrow^\star w \,, l \in [1:r] \,.$$

Wir wissen

$$\begin{aligned}
S \overset{(2.9)}{\Rightarrow^\star} \gamma_1 A \delta_1 \overset{(2.10)}{\Rightarrow^\star}\; & p_1 A q_1 \\
\overset{(2.11)}{\Rightarrow}\; & p_1 \alpha_1 B_1 \alpha_2 \cdots B_r \alpha_{r+1} q_1 \\
\overset{(2.12)}{\Rightarrow^\star}\; & p_1 p_l B_l \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1} q_1 \\
\overset{(2.12)}{\Rightarrow^\star}\; & p_1 p_{r+1} q_1 = w \\
=\; & a_1 \cdots a_s \cdots a_{t_{r+1}} \cdots a_n,\, a_i \in T,\, 1 \leq i \leq n \,.
\end{aligned}$$

Betrachten wir zunächst die Itemlisten, die das **EPV** für das Wort w bezüglich **G** aufbaut. Mit den Voraussetzungen (2.9) bis (2.12) und wegen **Satz 1.1** folgern wir für alle $l \in [1:r]$:

$$[A \to \alpha_1 B_1 \cdots \alpha_l \bullet B_l \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}, s] \in I^{\mathbf{G}}_{t_l}(w) \,. \tag{2.13}$$

Hierzu setzen wir in **Satz 1.1**: $\gamma := \gamma_1$, $\delta := \delta_1$, $\alpha := \alpha_1 B_1 \alpha_2 \cdots \alpha_l B_l$, $\beta := \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}$, $i := s$ und $j := t_l$. Mit dieser Wahl folgt für alle $l \in$

$[1:r]$ und $1 \leq s \leq t_l \leq n$:

$$\begin{aligned}
&\text{(a)} \quad S \overset{(2.9)}{\Rightarrow^\star} \gamma_1 A \delta_1 = \gamma A \delta \,, \\
&\text{(b)} \quad \gamma \overset{(2.10)}{\Rightarrow^\star} p_1 = a_1 \cdots a_s = a_1 \cdots a_i \,, \\
&\text{(c)} \quad f : A \to \alpha_1 B_1 \alpha_2 \cdots B_r \alpha_{r+1} \in P,\ B_l \in \mathcal{U} \,, \\
&\text{(d)} \quad \alpha = \alpha_1 B_1 \alpha_2 \cdots B_{l-1} \alpha_l \overset{(2.12)}{\Rightarrow^\star} p_l \\
&\qquad\qquad = a_{s+1} \cdots a_{t_l} = a_{i+1} \cdots a_j \,.
\end{aligned}$$

Behauptung (2.13) ist bewiesen. Mit den gleichen Vorraussetzungen (2.9) bis (2.12) und **Satz 1.1** folgern wir aber ebenfalls für alle $l \in [1:r]$:

$$[B_l \to \epsilon \bullet, t_l] \in I^{\mathbf{G}}_{t_l}(w) \,. \tag{2.14}$$

Hierfür setzen wir $\gamma := \gamma_1 \alpha_1 B_1 \cdots B_{l-1} \alpha_l$, $\delta := \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1} \delta_1$, $\alpha := \beta := \epsilon$, $i := j := t_l$. Mit dieser Wahl folgt für alle $l \in [1:r]$ und $1 \leq s \leq t_l \leq n$:

$$\begin{aligned}
&\text{(a)} \quad S \overset{(2.9)}{\Rightarrow^\star} \gamma_1 A \delta_1 \\
&\qquad\qquad \overset{(2.11)}{\Rightarrow} \gamma_1 \alpha_1 B_1 \cdots B_{l-1} \alpha_l B_l \alpha_{l+1} \cdots B_r \alpha_{r+1} \delta_1 = \gamma B_l \delta \,, \\
&\text{(b)} \quad \gamma = \gamma_1 \alpha_1 B_1 \cdots B_{l-1} \alpha_l \\
&\qquad\qquad \overset{(2.10)}{\Rightarrow^\star} p_1 \alpha_1 B_1 \cdots B_{l-1} \alpha_l \\
&\qquad\qquad \overset{(2.12)}{\Rightarrow^\star} p_1 p_l = a_1 \cdots a_s p_l \\
&\qquad\qquad = a_1 \cdots a_s a_{s+1} \cdots a_{t_l} = a_1 \cdots a_s a_{s+1} \cdots a_j \,, \\
&\text{(c)} \quad f\epsilon : B_l \to \epsilon \in P,\ B_l \in \mathcal{U} \,, \\
&\text{(d)} \quad \alpha = \epsilon \Rightarrow^\star a_{t_l+1} \cdots a_{t_l} = \epsilon \,.
\end{aligned}$$

Nun sucht die Operation *Completor* wegen (2.14) in Liste $I^{\mathbf{G}}_{t_l}(w)$ nach Vorgängern für das Item $[B \to \epsilon \bullet, t_l]$ und findet mit (2.13) das Item $[A \to \alpha_1 B_1 \cdots \alpha_l \bullet B_l \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}, s]$. Dieses wird übernommen als $[A \to \alpha_1 B_1 \cdots \alpha_l B_l \bullet \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}, s]$. In Liste $I^{\mathbf{G}}_{t_l}(w)$ befinden sich also neben $[B_l \to \epsilon \bullet, t_l]$ und $[A \to \alpha_1 B_1 \cdots \alpha_l \bullet B_l \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}, s]$ auch das Item $[A \to \alpha_1 B_1 \cdots \alpha_l B_l \bullet \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}, s]$, falls die ϵ-Regel $f_\epsilon : B_l \to \epsilon \in P$ in einer Ableitung des Wortes w bezüglich $\mathbf{G}$ angewandt wurde. Also gilt für alle $l \in [1:r]$ und $1 \le s \le t_l \le n$:

$$[A \to \alpha_1 B_1 \cdots \alpha_l B_l \bullet \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}, s] \in I_{t_l}^{\mathbf{G}}(w)\,. \tag{2.15}$$

Betrachten wir nun die Itemlisten, die das **EPV** für w bezüglich der Grammatik $\mathbf{G}'$ aufbaut. Wegen (2.11) und (2.7) gilt für alle $l \in [1:r]$

$$f' : A \to \beta_1 \beta_2 \in P'\,,$$

wobei $\beta_1 := \alpha_1 B_1 \alpha_2 \cdots B_{l-1} \alpha_l$ ist und β_2 nach (2.7) aus $\alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}$ entstanden ist. Fassen wir die Aussagen (2.9) bis (2.12) in

(a) $S \overset{(2.9)}{\Rightarrow^\star} \gamma_1 A \delta_1\,,$

(b) $\gamma_1 \overset{(2.10)}{\Rightarrow^\star} p_1 = a_1 \cdots a_s = a_1 \cdots a_i\,,$

(c) $f' : A \to \beta_1 \beta_2 \in P'\,,$

(d) $\beta_1 = \alpha_1 B_1 \cdots B_{l-1} \alpha_l \overset{(2.12)}{\Rightarrow^\star} p_l = a_{s+1} \cdots a_{t_l} = a_{i+1} \cdots a_j$

zusammen und setzen in **Satz 1.1** $\gamma := \gamma_1$, $\delta := \delta_1$, $\alpha := \beta_1$, $\beta := \beta_2$, $i := s$ und $j := t_l$, so erhalten wir die Bedingungen, daß für alle $l \in [1:r]$ und $1 \le s \le t_l \le n$ gilt:

$$[A \to \beta_1 \bullet \beta_2, s] \in I_{t_l}^{\mathbf{G}'}(w)\,. \tag{2.16}$$

Sind in $I_{t_l}^{\mathbf{G}}(w)$ die drei Items $[A \to \alpha_1 B_1 \cdots \alpha_l \bullet B_l \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}, s]$, $[A \to \alpha_1 B_1 \cdots \alpha_l B_l \bullet \alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}, s]$ und $[B_l \to \epsilon \bullet, t_l]$, so befinden sich mit (2.16) in Liste $I_{t_l}^{\mathbf{G}'}(w)$ alle Items $[A \to \beta_1 \bullet \beta_2, s]$. Hierbei ist $\beta_1 := \alpha_1 B_1 \alpha_2 \cdots B_{l-1} \alpha_l$ und β_2 nach (2.7) aus $\alpha_{l+1} B_{l+1} \cdots B_r \alpha_{r+1}$ entstanden. Die maximale Anzahl von 2^{r+1} Items wird erreicht, falls $\beta_1 = \alpha_1 = \epsilon$ ist. Dann liefert *Predictor*$(I_s^{\mathbf{G}'}(w))$ alle 2^{r+1} Items $[A \to \bullet \beta_2, s]$ in Liste $I_s^{\mathbf{G}'}(w)$. Für die Itemlisten der Grammatik $\mathbf{G}'$ folgt mit (2.8) bis (2.16):

$$(\forall w \in L(\mathbf{G}))\,(\forall j \in [0:|w|])\,(\,|\,I_j^{\mathbf{G}'}(w)\,| \le 2^{r+1}\,|\,I_j^{\mathbf{G}}(w)\,|\,)\,. \tag{2.17}$$

Im weiteren Verlauf der Arbeit gehen wir davon aus, daß die Grammtik $\mathbf{G}$ bereits ϵ-frei ist, falls $L(\mathbf{G})$ ϵ-frei ist, und ansonsten P' als einzige ϵ-Regel die Produktion $f_1 : S' \to \epsilon \in P'$ enthält, wobei das neue Startsymbol S' nur noch in $f_0 : S' \to S \in P'$ vorkommt.

2.5 Kettenregel-Freiheit

In diesem Abschnitt werden wir zu $\mathbf{G}$ eine äquivalente CFG konstruieren, die keine Kettenregeln enthält.

Lemma 2.4 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine CFG. Es gibt eine CFG* $\mathbf{G}' := (V, T, S, P')$ *ohne Kettenregeln, für die gilt:* $L(\mathbf{G}') = L(\mathbf{G}) \setminus \{\epsilon\}$.

Beweis:[4] Wir konstruieren die Hilfsmengen $\mathcal{U}_A$ für alle $A \in V$:

$$
\begin{array}{rrcl}
1. & \mathcal{U}_{A,1} & := & \{A\}\,. \\
2. & \mathcal{U}_{A,i+1} & := & \mathcal{U}_{A,i} \cup \{B \in V \mid f : C \to B \in P, C \in \mathcal{U}_{A,i}\}\,.
\end{array}
$$

Da $\mathcal{U}_{A,i} \subseteq \mathcal{U}_{A,i+1} \subseteq V$ ist, folgt:

$$(\exists l \in \mathcal{N})\,(\forall j \in \mathcal{N})\,(\mathcal{U}_{A,l} = \mathcal{U}_{A,l+j})\,.$$

Wir definieren $\mathcal{U}_A := \mathcal{U}_{A,l}$. Es folgt:

$$B \in \mathcal{U}_A \;\leftrightarrow\; A \Rightarrow^\star B\,.$$

Wir konstruieren P' wie folgt:

$$
\begin{array}{rcl}
P'_A & := & \{f' : A \to \beta \mid B \in \mathcal{U}_A,\; f : B \to \beta \in P,\; \beta \notin V \cup \{\epsilon\}\}\,. \\
P' & := & \bigcup\limits_{A \in V} P'_A \cup \{f : A \to \alpha \in P \mid \alpha \notin V \cup \{\epsilon\}\}\,.
\end{array}
$$

Es folgt: $L(\mathbf{G}') \subseteq L(\mathbf{G})$ und $\mathbf{G}'$ ist kettenregelfrei. Durch Induktion über die Länge der Ableitung von $w \in L(\mathbf{G})$ läßt sich zeigen, daß die Sprache

[4]Beweis nach [Sud] S. 104 und [Sal] S. 55

$L(\mathbf{G}) \setminus \{\epsilon\} \subseteq L(\mathbf{G}')$ ist, und das Lemma ist bewiesen. $\diamond$

Falls $\epsilon \in L(\mathbf{G})$ ist, so nehmen wir zu P' die Kettenregel $f_0 : S' \to S$ und die ϵ-Regel $f_1 : S' \to \epsilon$ hinzu, wobei das neue Startsymbol S' in keiner anderen Produktion vorkommt. Durch die Konstruktion von P' nach **Lemma 2.4** wächst P' an. Im schlechtesten Fall sind alle $A \in V$ in allen $\mathcal{U}_{A'}$ enthalten. Sei r die maximale Anzahl der Produktionen $f : B \to \beta \in P$ mit linker Seite B und $B \in \mathcal{U}_A$; für jedes $A \in V$ werden somit maximal $r\,(\,|V| - 1)$ viele neue Produktionen erzeugt, also $|P'| \leq r(\,|V| - 1)|V|$. Da $r \leq |P|$ gilt, ist: $|P'| \leq |P|\,|V|^2 := g_1\,|P|$. Die Menge der Nonterminals wude nur um das neue Startsymbol S' vergrößert,und es ist:

$$|\,P'\,| \leq g_1\,|\,P\,| \quad \text{und} \quad |\,V'\,| \leq |\,V\,| + 1\,. \tag{2.18}$$

Betrachten wir die Auswirkung der Umformung nach **Lemma 2.4** auf die Itemlisten. Da $\mathbf{G}$ bereits reduziert und ϵ-frei war, gilt für alle $A \in V$:

$$(\exists \gamma, \delta \in V^\star)(S \Rightarrow^\star \gamma A \delta)\,, \tag{2.19}$$

$$(\gamma \Rightarrow^\star p_1 \in T^\star \quad \wedge \quad \delta \Rightarrow^\star q_1 \in T^\star)\,. \tag{2.20}$$

Seien nun in P die beiden Produktionen

$$f_2 : A \to B \quad \wedge \quad f_3 : B \to \alpha\beta\,, \tag{2.21}$$

und gelte für f_3:

$$\alpha \Rightarrow^\star p \in T^\star \quad \wedge \quad \beta \Rightarrow^\star q \in T^\star\,. \tag{2.22}$$

Wir wissen

$$S \overset{(2.19)}{\Rightarrow^\star} \gamma A \delta \overset{(2.20)}{\Rightarrow^\star} p_1 A q_1 \overset{f_2}{\Rightarrow} p_1 B q_1 \overset{f_3}{\Rightarrow} p_1 \alpha\beta q_1$$
$$\overset{(2.22)}{\Rightarrow^\star} p_1 p q q_1 := w = a_1 \cdots a_n \in L(\mathbf{G}),\ a_i \in T,\ 1 \leq i \leq n\,.$$

Seien für $1 \leq i < j \leq n$ die Längen $|p_1| = i$ und $|p| = j - i$, und gelten die Aussagen (2.19) bis (2.22) so, daß $f_2 : A \to B \in P$ eine Kettenregel in einer

Ableitung des Wortes w bezüglich $\mathbf{G}$ ist. Wir folgern für die Itemlisten, die das **EPV** für das Wort w bezüglich $\mathbf{G}$ aufbaut mit **Satz 1.1**:

$$[B \to \alpha \bullet \beta, i] \in I_j^{\mathbf{G}}(w)\,. \tag{2.23}$$

Denn es ist:

(a) $S \overset{(2.19)}{\Rightarrow^\star} \gamma A \delta \overset{(2.21)}{\Rightarrow} \gamma B \delta\,,$

(b) $\gamma \overset{(2.20)}{\Rightarrow^\star} p_1 = a_1 \cdots a_i\,,$

(c) $f_3 : B \to \alpha\beta \in P\,,$

(d) $\alpha \overset{(2.22)}{\Rightarrow^\star} p = a_{i+1} \cdots a_j\,.$

und Aussage (2.23) ist bewiesen.

Betrachten wir nun die Itemlisten für das Wort w bezüglich $\mathbf{G}'$. Mit (2.19) bis (2.22) und **Satz 1.1** folgern wir:

$$[A \to \alpha \bullet \beta, i] \in I_j^{\mathbf{G}'}(w)\,. \tag{2.24}$$

Denn es ist:

(a) $S \overset{(2.19)}{\Rightarrow^\star} \gamma A \delta\,,$

(b) $\gamma \overset{(2.20)}{\Rightarrow^\star} p_1 = a_1 \cdots a_i\,,$

(c) $f_3' : A \to \alpha\beta \in P'$ (nach Konstr. von P') ,

(d) $\alpha \overset{(2.22)}{\Rightarrow^\star} p = a_{i+1} \cdots a_j\,,$

und Aussage (2.24) ist bewiesen. Für jedes Item $[B \to \alpha \bullet \beta, i]$ in Liste $I_j^{\mathbf{G}}(w)$ befindet sich genau ein Item $[A \to \alpha \bullet \beta, i]$ in Liste $I_j^{\mathbf{G}'}(w)$, falls die Kettenregel $f_2 : A \to B \in P$ der Produktion $f_3 : B \to \alpha\beta \in P$ voraus ging. Bezüglich $\mathbf{G}$ wird zusätzlich das Item $[A \to B \bullet, i]$ von der Operation *Completor* in die Liste $I_j^{\mathbf{G}}(w)$ geholt. Weiter sucht *Completor* sowohl für $\mathbf{G}$ als auch für $\mathbf{G}'$ in den Itemlisten mit Index i nach möglichen Vorgängern und findet jeweils alle Items $[C \to \nu \bullet A\mu, k]$. Für $\mathbf{G}$ werden aber zusätzlich alle Items $[\hat{C} \to \hat{\nu} \bullet B\hat{\mu}, \hat{k}]$ in $I_i^{\mathbf{G}}(w)$ gefunden. Es gilt für die Itemlisten:

$$(\forall w \in L(\mathbf{G}))\,(\forall j \in [\,0 : |\,w\,|\,])(\,|\,I_j^{\mathbf{G}'}(w)\,| \;\leq g_1\,|\,I_j^{\mathbf{G}}(w)\,|\,)\,. \tag{2.25}$$

Fassen wir die bisherigen Ergebnisse zusammen: Seien $I_j^{\mathbf{G}}(w)$ und $I_j^{\mathbf{G}'}(w)$ die vom **EPV** bezüglich der Grammatik $\mathbf{G}$ und einer äquivalenten, reduzierten, ϵ-freien und kettenregelfreien Grammatik $\mathbf{G}'$ aufgebauten Itemlisten für das Wort w. Es gilt:

$$(\forall w \in L(\mathbf{G}))\,(\forall j \in [\,0 : |w|\,])(\,|I_j^{\mathbf{G}'}(w)| \;\leq 2^{r+1} g_1 |I_j^{\mathbf{G}}(w)|\,)\,. \tag{2.26}$$

Im weiteren Verlauf der Arbeit gehen wir davon aus, daß die Grammtik $\mathbf{G}$ bereits kettenregel- und ϵ-regelfrei ist, falls $L(\mathbf{G})$ ϵ-frei ist, und ansonsten P nur die Kettenregel $f_0 : S' \to S$ und die ϵ-Regel $f_1 : S' \to \epsilon$ enthält, wobei das neue Startsymbol S' in keiner anderen Produktion vorkommt.

2.6 Normalform

In diesem Abschnitt werden wir zu $\mathbf{G}$ eine äquivalente CFG konstruieren, die Chomsky-Normal-Form hat.

Satz 2.2 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine CFG. Es gibt eine reduzierte CFG* $\mathbf{G}^{CNF}$ *in CNF, für die gilt:* $L(\mathbf{G}) = L(\mathbf{G}^{CNF})$.

Beweis:[5] Nach **Lemma 2.1** bis **2.4** konstruieren wir aus $\mathbf{G}$ eine Grammatik $\mathbf{G}_1$, die reduziert, kettenregelfrei und ϵ-frei ist. Alle Produktionen aus P_1 sind von der Form:

$$(\forall f : A \to \alpha \in P_1)\;(\alpha \in T_1 \vee (\alpha \in (V_1 \cup T_1)^n, n \geq 2))\,.$$

Wir konstruieren $\mathbf{G}_2 := (V_2, T_1, S_1, P_2)$ wie folgt:

[5] Beweis nach [Sud] S. 113 und [Sal] S. 56

1. Wir ersetzen jede Produktion $f : A \to \alpha_1 b_1 \alpha_2 \cdots b_n \alpha_{n+1} \in P_1$ mit $b_i \in T_1$ für $i \in [\,1 : n\,]$ und $\alpha_i \in V_1^\star$ für $i \in [\,1 : n+1\,]$ in P_2 durch die Produktion $f : A \to \alpha_1 H_1^{(f)} \alpha_2 \cdots H_n^{(f)} \alpha_{n+1}$ und die Produktionen $f_i : H_i^{(f)} \to b_i$ für $i \in [\,1 : n\,]$, wobei die $H_i^{(f)} \notin V_1$ sind.

2. Wir nehmen für alle $f \in P_1$ alle $H_i^{(f)}$ zu V_2 dazu.

Alle Produktionen in P_2 sind nun von der Form:

$$(\forall f : A \to \alpha \in P_2)\,(\alpha \in T_1 \cup V_2^n,\, n \geq 2)\,.$$

Es gilt: $L(\mathbf{G}_1) = L(\mathbf{G}_2)$. Aus $\mathbf{G}_2$ konstruieren wir $\mathbf{G}_3 := (V_3, T_1, S_1, P_3)$:

3. Jede Produktion $f : A \to B_1 B_2 \cdots B_r \in P_2$ mit $r > 2$ wird in P_3 ersetzt durch die Folge von Produktionen $\bar{f}_i : \bar{H}_{i-1}^{(f)} \to B_i \bar{H}_i^{(f)}$ für $i \in [\,1 : r-1\,]$, wobei $\bar{H}_0^{(f)} := A$ und $\bar{H}_{r-1}^{(f)} := B_r$ gilt und die $\bar{H}_i^{(f)} \notin V_2$ sind.

4. Wir nehmen zu V_3 alle $\bar{H}_i^{(f)}$ aus 3. dazu.

Es folgt: $L(\mathbf{G}_3) = L(\mathbf{G}_2)$ und mit $\mathbf{G}^{CNF} := \mathbf{G}_3$ ist **Satz 2.2** bewiesen. $\diamond$

Betrachten wir die Auswirkungen auf die Größe von $\mathbf{G}^{CNF}$. Sei r die Länge der längsten rechten Seite aller Produktionen aus P_1. P_2 wächst mit 1. auf maximal $|\,P_2\,| \,\leq\, (r+1)\,|\,P_1\,|$ an, falls die längste rechte Seite aus r Terminals besteht. Wegen 3. werden maximal $(r-1)\,|\,P_2\,|$ viele Produktionen in P_3 erzeugt. Die Menge der Nonterminals V_3 erhält mit 2. r neue Hilfszeichen je Produktion in P_1 und mit 4. $r-2$ neue Hilfszeichen je Produktion in P_2. Also folgt:

$$|\,P_3\,| \leq (r-1)|\,P_2\,| \leq r^2|\,P_1\,| \quad \text{und} \quad |\,V_3\,| \,\leq\, |\,V_1\,| + r^2|\,P_1\,|\,. \tag{2.27}$$

Betrachten wir die Auswirkungen der Umformungen nach **Satz 2.2** auf die Itemlisten, und gehen davon aus, daß $\mathbf{G}$ bereits reduziert, kettenregel- und

ϵ-frei war. Also gilt für alle $A \in V_1$:

$$(\exists \gamma_1, \delta_1 \in (T_1 \cup V_1)^\star)\,(S \Rightarrow^\star \gamma_1 A \delta_1)\,, \tag{2.28}$$

$$\gamma_1 \Rightarrow^\star a_1 \cdots a_s \in T_1^\star \quad \wedge \quad \delta_1 \Rightarrow^\star q_1 \in T_1^\star\,. \tag{2.29}$$

Sei nun die Produktion $f \in P_1$ mit:

$$f : A \to \alpha_1 \cdots \alpha_r \in P_1,\ \alpha_l \in (T_1 \cup V_1),\ 1 \leq l \leq r\,, \tag{2.30}$$

$$\alpha_1 \cdots \alpha_l \Rightarrow^\star a_{s+1} \cdots a_{t_l} \in T^\star, 1 \leq s \leq t_l\,. \tag{2.31}$$

Wir wissen:

$$\begin{aligned} S \overset{(2.28)}{\Rightarrow^\star} \gamma_1 A \delta_1 \overset{(2.29)}{\Rightarrow^\star}\ & a_1 \cdots a_s A q_1 \\ \overset{(2.30)}{\Rightarrow}\ & a_1 \cdots a_s \alpha_1 \cdots \alpha_r q_1 \\ \overset{(2.31)}{\Rightarrow^\star}\ & a_1 \cdots a_s a_{s+1} \cdots a_{t_r} q_1 \\ =\ & a_1 \cdots a_s a_{s+1} \cdots a_{t_r} \cdots a_n = w \in T_1^\star\,. \end{aligned}$$

Wir betrachten die Listen, die das **EPV** für $w \in L(\mathbf{G}_1)$ bezüglich $\mathbf{G}_1$ aufbaut und zeigen für alle $l \in [1:r]$ und $1 \leq s \leq t_r < n$:

$$[A \to \alpha_1 \cdots \alpha_l \bullet \alpha_{l+1} \cdots \alpha_r, s] \in I_{t_l}^{\mathbf{G}_1}(w)\,. \tag{2.32}$$

Wir zeigen die Behauptung durch Nachweis der Bedingungen von **Satz 1.1** und setzen hierfür: $\gamma := \gamma_1$, $\delta := \delta_1$, $\alpha := \alpha_1 \cdots \alpha_l$, $\beta := \alpha_{l+1} \cdots \alpha_r$, $i := s$ und $j := t_l$. Mit dieser Wahl folgt für alle $l \in [1:r]$ und $1 \leq s \leq t_r < n$:

(a) $S \overset{(2.28)}{\Rightarrow^\star} \gamma_1 A \delta_1\,,$

(b) $\gamma = \gamma_1 \overset{(2.29)}{\Rightarrow^\star} a_1 \cdots a_s = a_1 \cdots a_i\,,$

(c) $f : A \to \alpha_1 \cdots \alpha_r \in P_1,\ \alpha_l \in (T_1 \cup V_1)\,,$

(d) $\alpha = \alpha_1 \cdots \alpha_l \overset{(2.31)}{\Rightarrow^\star} a_{s+1} \cdots a_{t_l} = a_{i+1} \cdots a_j\,.$

Betrachten wir nun die Itemlisten für w bezüglich $\mathbf{G}^{CNF}$. Mit (2.30) und 1. folgt:

$$f : A \to H_1^{(f)} H_2^{(f)} \cdots H_r^{(f)} \in P_2,\ 1 \leq l \leq r\,, \tag{2.33}$$

wobei $H_l^{(f)} = \alpha_l$ ist, falls $\alpha_l \in V_1$ ist und $H_l^{(f)} \in V_2 \setminus V_1$ ist, falls $\alpha_l \in T_1$ ist. Gleichfalls folgt mit 1. :

$$f_l : H_l^{(f)} \to \alpha_l \in P_2,\ 1 \leq l \leq r\,, \tag{2.34}$$

falls $\alpha_l \in T_1$ ist. Mit (2.33), (2.34) und 3. folgt nun:

$$\bar{f}_l : \bar{H}_{l-1}^{(f)} \to H_l^{(f)} \bar{H}_l^{(f)} \in P_3,\ l \in [1 : r-1],\ r > 2\,, \tag{2.35}$$

wobei $\bar{H}_0^{(f)} = A$ und $\bar{H}_{r-1}^{(f)} = H_r^{(f)}$ gilt und alle $\bar{H}_l^{(f)} \in V_3 \setminus V_2$ sind. Somit wissen wir:

$$\begin{aligned} A &\overset{(2.35)}{\Rightarrow} H_1^{(f)} \bar{H}_1^{(f)} \overset{(2.35)}{\Rightarrow} H_1^{(f)} H_2^{(f)} \bar{H}_2^{(f)} \\ &\qquad\overset{(2.35)}{\Rightarrow^\star} H_1^{(f)} H_2^{(f)} \cdots H_{r-2}^{(f)} \bar{H}_{r-2}^{(f)} \\ &\qquad\overset{(2.35)}{\Rightarrow} H_1^{(f)} H_2^{(f)} \cdots H_{r-2}^{(f)} H_{r-1}^{(f)} H_r^{(f)}\,. \end{aligned}$$

Für alle $H_l^{(f)} \in V_2$ gibt es Produktionen der Form (2.34), falls $\alpha_l \in T_1$ ist. Also ist $H_l^{(f)} = \alpha_l$, falls α_l ein Nonterminal war, oder die Produktion $f_l : H_l^{(f)} \to \alpha_l$ liegt in P_2, falls α_l ein Terminal war. Somit folgt:

$$A \overset{(2.35)}{\Rightarrow^\star} H_1^{(f)} H_2^{(f)} \cdots H_r^{(f)} \overset{(2.34)}{\Rightarrow^\star} \alpha_1 \alpha_2 \cdots \alpha_r\,.$$

Mit den Aussagen (2.28) bis (2.35) zeigen wir für $r > 2$ und $l \in [1 : r-1]$, mit $1 \leq s \leq t_l < n$ und $t_0 = s$:

$$[\bar{H}_{l-1}^{(f)} \to H_l^{(f)} \bullet \bar{H}_l^{(f)}, t_{l-1}] \in I_{t_l}^{CNF}(w)\,. \tag{2.36}$$

Wir zeigen die Behauptung durch Nachweis der Bedingungen von **Satz 1.1** und setzen hierfür: $\gamma := \gamma_1 H_1^{(f)} H_2^{(f)} \cdots H_{l-1}^{(f)}$, $\delta := \delta_1$, $\alpha := H_l^{(f)}$, $\beta := \bar{H}_l^{(f)}$, $i := t_{l-1}$ und $j := t_l$. Mit dieser Wahl folgt für $r > 2$ und alle $l \in [1 : r-1]$, $1 \leq s \leq t_l < n$ und $t_0 = s$:

(a) $$\begin{aligned} S &\overset{(2.28)}{\Rightarrow^\star} \gamma_1 A \delta_1 \\ &\overset{(2.35)}{\Rightarrow^\star} \gamma_1 H_1^{(f)} H_2^{(f)} \cdots H_{l-1}^{(f)} \bar{H}_{l-1}^{(f)} \delta_1 = \gamma \bar{H}_{l-1}^{(f)} \delta\,, \end{aligned}$$

$$
\begin{aligned}
&\text{(b)} \quad \gamma = \gamma_1 H_1^{(f)} H_2^{(f)} \cdots H_{l-1}^{(f)} \overset{(2.34)}{\Rightarrow^\star} \gamma_1 \alpha_1 \cdots \alpha_{l-1} \\
&\qquad \overset{(2.29)}{\Rightarrow^\star} a_1 \cdots a_s \alpha_1 \cdots \alpha_{l-1} \\
&\qquad \overset{(2.31)}{\Rightarrow^\star} a_1 \cdots a_s a_{s+1} \cdots a_{t_{l-1}} = a_1 \cdots a_i \,, \\
&\text{(c)} \quad \bar{f}_l : \bar{H}_{l-1}^{(f)} \to H_l^{(f)} \bar{H}_l^{(f)} \in P_3 \text{ (nach Konstr.)} \;, \\
&\text{(d)} \quad \alpha = H_l^{(f)} \overset{(2.34)}{\Rightarrow^\star} \alpha_l \overset{(2.31)}{\Rightarrow^\star} a_{t_{l-1}+1} \cdots a_{t_l} = a_{i+1} \cdots a_j \,.
\end{aligned}
$$

Damit ist Behauptung (2.36) bewiesen. Für $r > 2$ und alle $l \in [1 : r-1]$ befindet sich in allen Listen $I_{t_l}^{CNF}(w)$ je ein Item $[\bar{H}_{l-1}^{(f)} \to H_l^{(f)} \bullet \bar{H}_l^{(f)}, t_{l-1}]$. Mit den Aussagen (2.28) bis (2.35) zeigen wir gleichfalls für alle $r > 2$ und $1 \le s \le t_r < n$:

$$
[\bar{H}_{r-2}^{(f)} \to H_{r-1}^{(f)} H_r^{(f)} \bullet, t_{r-2}] \in I_{t_r}^{CNF}(w) \,. \tag{2.37}
$$

Verglichen mit **G** entspricht (2.37) der Situation, in der die Produktion $f : A \to \alpha_1 \cdots \alpha_r$ vollständig abgearbeitet ist. Wir zeigen die Behauptung (2.37) durch Nachweis der Bedingungen von **Satz 1.1** und setzen hierfür: $\gamma := \gamma_1 H_1^{(f)} H_2^{(f)} \cdots H_{r-2}^{(f)}$, $\delta := \delta_1$, $\alpha := H_{r-1}^{(f)} H_r^{(f)}$, $\beta := \epsilon$, $i := t_{r-2}$ und $j := t_r$ und erhalten:

$$
\begin{aligned}
&\text{(a)} \quad S \overset{(2.28)}{\Rightarrow^\star} \gamma_1 A \delta_1 \\
&\qquad \overset{(2.35)}{\Rightarrow^\star} \gamma_1 H_1^{(f)} H_2^{(f)} \cdots H_{r-2}^{(f)} \bar{H}_{r-2}^{(f)} \delta_1 = \gamma \bar{H}_{r-2}^{(f)} \delta \,, \\
&\text{(b)} \quad \gamma \overset{(2.34)}{\Rightarrow^\star} \gamma_1 \alpha_1 \cdots \alpha_{r-2} \\
&\qquad \overset{(2.29)}{\Rightarrow^\star} a_1 \cdots a_s \alpha_1 \cdots \alpha_{r-2} \\
&\qquad \overset{(2.31)}{\Rightarrow^\star} a_1 \cdots a_s a_{s+1} \cdots a_{t_{r-2}} = a_1 \cdots a_i \,, \\
&\text{(c)} \quad \bar{f}_{r-1} : \bar{H}_{r-2}^{(f)} \to H_{r-1}^{(f)} H_r^{(f)} \in P_3 \text{ (nach Konstr.)} \;, \\
&\text{(d)} \quad \alpha = H_{r-1}^{(f)} H_r^{(f)} \overset{(2.34)}{\Rightarrow^\star} \alpha_{r-1} \alpha_r \\
&\qquad \overset{(2.31)}{\Rightarrow^\star} a_{t_{r-2}+1} \cdots a_{t_r} = a_{i+1} \cdots a_j \,.
\end{aligned}
$$

Somit ist Behauptung (2.37) bewiesen. Die Operation *Completor* holt also aus $I^{CNF}_{t_{r-1}}(w)$ das Item $[\bar{H}^{(f)}_{r-2} \rightarrow H^{(f)}_{r-1}H^{(f)}_r \bullet, t_{r-2}]$ in die Liste $I^{CNF}_{t_r}(w)$ und sucht in Liste $I^{CNF}_{t_{r-2}}(w)$ nach Items mit dem Punkt vor dem Nonterminal $\bar{H}^{(f)}_{r-2}$. Mit Aussage (2.36) befindet sich nun in allen Listen $I^{CNF}_{t_l}(w)$ für $l \in [1 : r-1]$ je ein Item $[\bar{H}^{(f)}_{l-1} \rightarrow H^{(f)}_l \bullet \bar{H}^{(f)}_l, t_{l-1}]$. Diese Items werden nacheinander in den Listen $I^{CNF}_{t_l}(w)$ gefunden und in Liste $I^{CNF}_{t_r}(w)$ gebracht. Durch Induktion über r läß sich für alle $l \in [1 : r-1]$ und $1 \leq s \leq t_l < n$ zeigen:

$$(\forall l \in [r-1])\,([\bar{H}^{(f)}_{l-1} \rightarrow H^{(f)}_l \bar{H}^{(f)}_l \bullet, t_{l-1}] \in I^{CNF}_{t_r}(w))\,, \tag{2.38}$$

wobei $t_0 = s$, $\bar{H}^{(f)}_0 = A$ und $\bar{H}^{(f)}_{r-1} = H^{(f)}_r$ ist. Dies sind $r-1$ viele Items. Wir möchten hier die Listen $I^{CNF}_{t_l}(w)$ in schematischer Weise darstellen um eine Anschauung zu geben. Wir zeigen deshalb nur die Items, die aus den Aussagen (2.28) bis (2.38) folgen. Es ist $l \in [1 : r-1]$, $t_0 = s$, $\bar{H}^{(f)}_0 = A$ und $\bar{H}^{(f)}_{r-1} = H^{(f)}_r$.

Liste $I_{t_0}(w)$	Liste $I_{t_1}(w)$	Liste $I_{t_2}(w)$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$[H^{(f)}_1 \rightarrow \alpha_1 \bullet, t_0]$	$[H^{(f)}_2 \rightarrow \alpha_2 \bullet, t_1]$
$\vdots$	$[\bar{H}^{(f)}_0 \rightarrow H^{(f)}_1 \bullet \bar{H}^{(f)}_1, t_0]$	$[\bar{H}^{(f)}_1 \rightarrow H^{(f)}_2 \bullet \bar{H}^{(f)}_2, t_1]$
$[\bar{H}^{(f)}_0 \rightarrow \bullet H^{(f)}_1 \bar{H}^{(f)}_1, t_0]$	$[\bar{H}^{(f)}_1 \rightarrow \bullet H^{(f)}_2 \bar{H}^{(f)}_2, t_1]$	$[\bar{H}^{(f)}_2 \rightarrow \bullet H^{(f)}_3 \bar{H}^{(f)}_3, t_2]$
$[H^{(f)}_1 \rightarrow \bullet \alpha_1, t_0]$	$[H^{(f)}_2 \rightarrow \bullet \alpha_2, t_1]$	$[H^{(f)}_3 \rightarrow \bullet \alpha_3, t_2]$

Liste $I_{t_l}(w)$	Liste $I_{t_{r-1}}(w)$
$\vdots$	$\vdots$
$[H^{(f)}_l \rightarrow \alpha_l \bullet, t_{l-1}]$	$\vdots$
$[\bar{H}^{(f)}_{l-1} \rightarrow H^{(f)}_l \bullet \bar{H}^{(f)}_l, t_{l-1}]$	$[H^{(f)}_{r-1} \rightarrow \alpha_{r-1} \bullet, t_{r-2}]$
$[\bar{H}^{(f)}_l \rightarrow \bullet H^{(f)}_{l+1} \bar{H}^{(f)}_{l+1}, t_l]$	$[\bar{H}^{(f)}_{r-2} \rightarrow H^{(f)}_{r-1} \bullet \bar{H}^{(f)}_{r-1}, t_{r-2}]$
$[H^{(f)}_{l+1} \rightarrow \bullet \alpha_{l+1}, t_l]$	$[\bar{H}^{(f)}_{r-1} \rightarrow \bullet \alpha_r, t_{r-1}]$

Liste $I_{t_r}(w)$	
$\vdots$	$[\bar{H}^{(f)}_{l-1} \to H^{(f)}_l \bar{H}^{(f)}_l \bullet, t_l]$
$[\bar{H}^{(f)}_{r-1} \to \alpha_r \bullet, t_{r-1}]$	$\vdots$
$[\bar{H}^{(f)}_{r-2} \to H^{(f)}_{r-1} \bar{H}^{(f)}_{r-1} \bullet, t_{r-2}]$	$[\bar{H}^{(f)}_{1} \to H^{(f)}_2 \bar{H}^{(f)}_2 \bullet, t_1]$
$[\bar{H}^{(f)}_{r-3} \to H^{(f)}_{r-2} \bar{H}^{(f)}_{r-2} \bullet, t_{r-3}]$	$[\bar{H}^{(f)}_{0} \to H^{(f)}_1 \bar{H}^{(f)}_1 \bullet, t_0]$
$\vdots$	$\vdots$

Fassen wir das letzte Ergebnis in einer Aussage zusammen: Sei r die längste rechte Seite aller Produktionen $f : A \to \alpha_1 \cdots \alpha_n \in P_1$, so folgt für alle $i \in [\,0 : k\,]$ und $k \leq r$: Für jedes Item $[A \to \alpha_1 \cdots \alpha_i \bullet \alpha_{i+1} \cdots \alpha_k, s] \in I^{\mathbf{G}_1}_{t_i}(w)$ sind mit (2.27) und (2.38) maximal r^2 viele Items in $I^{CNF}_{t_i}(w)$. Also gilt:

$$(\forall w \in L(\mathbf{G}_1))\,(\forall j \in [\,0 : |w|\,])\,(\,|\,I^{CNF}_j(w)\,| \;\leq r^2\,|\,I^{\mathbf{G}_1}_j(w)\,|\,) \tag{2.39}$$

Die Umformungen nach **Lemma 2.1** bis **2.4** und **Satz 2.2** verändern die Itemlisten maximal um eine von der ursprünglichen Grammatik abhängige Konstante. Bei einigen Umformungen wächst zwar die Größe der Grammatik, nicht aber die Größe der Listen. Die ausschlaggebenden Faktoren sind g_1 und die Länge r der längsten rechten Seite der Produktionen. Es gilt für die Konstante aus **Satz 2.1**: $c := 2^{r+1} g_1 r^2$ und somit folgt:

$$(\forall w \in L(\mathbf{G}))\,(\forall j \in [\,0 : |w|\,])\,(\,|I^{CNF}_j(w)| \;\leq 2^{r+1} g_1 r^2\,|I^{\mathbf{G}}_j(w)|\,)\,. \tag{2.40}$$

und **Satz 2.1** ist bewiesen. $\diamond$

Wir können die in Kapitel 3 gewonnen Ergebnisse also auf allgemeine kontextfreie Grammatiken übertragen.

Kapitel 3

Nichtzustandsbeschränkte Grammatiken

In diesem Kapitel werden wir hinreichende Kriterien vorstellen, die ein unbeschränktes Wachstum der Itemlisten zur Folge haben. Alle Grammatiken, die mindestens einem dieser Kriterien genügen, sind also Elemente der Menge **NIB**.

3.1 Rechtsrekursion

Satz 3.1 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine reduzierte CFG in CNF und enthalte* $\mathbf{G}$ *ein rechtsrekursives Nonterminal* A. *Es gilt:* $\mathbf{G} \in \mathbf{NIB}$.

Beweis: Weil $\mathbf{G}$ das rechtsrekursive Nonterminal A enthält und in ϵ-freier und reduzierter Form vorliegt, gilt:

$$(\exists \gamma_1, \delta_1 \in (V \cup T)^\star)\,(S \Rightarrow^\star \gamma_1 A \delta_1)\,, \tag{3.1}$$

$$(\exists p_1, q_1 \in T^\star)\,(\gamma_1 \Rightarrow^\star p_1 \,\wedge\, \delta_1 \Rightarrow^\star q_1)\,, \tag{3.2}$$

$$(\exists \gamma_A \in V^+)\,(A \Rightarrow^\star \gamma_A A)\,, \tag{3.3}$$

$$(\exists q \in T^\star)\,(A \Rightarrow^\star q)\,. \tag{3.4}$$

Sei nun $l := |\gamma_A|$. Da **G** in CNF ist, hat (3.3) die folgende Form:

$$A \Rightarrow B_1 B_1' \Rightarrow B_1 B_2 B_2' \Rightarrow^\star \cdots \Rightarrow^\star B_1 B_2 \cdots B_l A\,, \tag{3.5}$$

wobei $f_i : B_{i-1}' \to B_i B_i' \in P$ Produktionen sind, für die gilt: $B_i, B_i' \in V$, $1 \leq i \leq l$, $B_0' = B_l' = A$. Aufgrund der ϵ-Freiheit und Reduziertheit von **G** erhalten wir weiter:

$$(\forall i \in [1:l]\,)\,(\forall B_i \in V)\,(\exists m_i \in T^+)\,(B_i \Rightarrow^\star m_i)\,.$$

Wir setzen $m := m_1 \cdots m_l \in T^+$ und erhalten:

$$\gamma_A = B_1 \cdots B_l \Rightarrow^\star m\,. \tag{3.6}$$

Durch k-maliges Wiederholen von (3.3) mit (3.6) folgt:

$$A \Rightarrow^\star (\gamma_A)^k A \Rightarrow^\star m^k A\,, k \in \mathcal{N}\,. \tag{3.7}$$

Betrachten wir nun das Wort $w_k := p_1 m^k q q_1$, $k \in \mathcal{N}_0$. Es ist

$$S \overset{(3.1)}{\Rightarrow^\star} \gamma_1 A \delta_1 \overset{(3.2)}{\Rightarrow^\star} p_1 A q_1 \overset{(3.3)}{\Rightarrow^\star} p_1 (\gamma_A)^k A q_1 \overset{(3.7)}{\Rightarrow^\star} p_1 m^k A q_1 \overset{(3.4)}{\Rightarrow^\star} p_1 m^k q q_1\,,$$

d.h.

$$(\forall k \in \mathcal{N}_0)\,(w_k \in L(\mathbf{G})\,)\,.$$

Sei nun $k \in \mathcal{N}_0$ und $w_k := a_1 a_2 \cdots a_n$, $a_i \in T$, $1 \leq i \leq n$. Wir betrachten die Itemlisten, die das **EPV** für das Wort w_k aufbaut und zeigen:

$$(\forall s \in [0:k-1])\,([A \to B_1 B_1' \bullet, |p_1 m^s|] \in I^{\mathbf{G}}_{|p_1 m^k q|}(w_k))\,. \tag{3.8}$$

Beweis: Wir zeigen die Behauptung (3.8) durch Nachweis der Bedingungen von **Satz 1.1**. Hierzu setzen wir: $\gamma := \gamma_1 (\gamma_A)^s\,, \delta := \delta_1\,, \alpha := B_1 B_1'\,, \beta := \epsilon\,, i := |p_1 m^s|$ und $j := |p_1 m^k q|$. Mit dieser Wahl folgt für alle $s \in [0:k-1]$:

(a) $S \overset{(3.1)}{\Rightarrow^\star} \gamma_1 A \delta_1 \overset{(3.3)}{\Rightarrow^\star} \gamma_1 (\gamma_A)^s A \delta_1 = \gamma A \delta\,,$

(b) $\gamma = \gamma_1 (\gamma_A)^s \overset{(3.2)}{\Rightarrow^\star} p_1 (\gamma_A)^s \overset{(3.6)}{\Rightarrow^\star} p_1 m^s = a_1 \cdots a_{|p_1 m^s|} = a_1 \cdots a_i\,,$

(c) $f_1 : A \to B_1B'_1 = f : A \to \alpha\beta \in P$ (nach Konstr.),

(d) $$\begin{aligned} \alpha &= B_1B'_1 \\ &\overset{(3.5)}{\Rightarrow^\star} B_1B_2\cdots B_lB'_l \overset{(3.6)}{=} \gamma_A A \\ &\overset{(3.3)}{\Rightarrow^\star} (\gamma_A)^{k-s}A \\ &\overset{(3.6)}{\Rightarrow^\star} m^{k-s}A \overset{(3.4)}{\Rightarrow^\star} m^{k-s}q \\ &= a_{|p_1m^s|+1}\cdots a_{|p_1m^kq|} = a_{i+1}\cdots a_j\,. \end{aligned}$$

Damit ist Behauptung (3.8) bewiesen. In der Itemliste $I^{\mathbf{G}}_{|p_1m^kq|}(w_k)$ befinden sich bei Anwendung des **EPV** auf w_k somit mindestens k Items. Da für jedes $k \in \mathcal{N}$ das Wort $w_k \in L(\mathbf{G})$ liegt, ist $\mathbf{G} \in \mathbf{NIB}$ und **Satz 3.1** ist bewiesen. $\diamond$

Die anschauliche Interpretation von **Satz 3.1** ist wie folgt: Bei der Ableitung der Satzform $A \Rightarrow^\star \gamma_A A$ nach (3.5) erzeugt das **EPV** nur Items, deren Zeiger jeweils von Liste $I^{\mathbf{G}}_{|p_1pm_1\cdots m_i|}(w_k)$ in die Liste $I^{\mathbf{G}}_{|p_1pm_1\cdots m_{i-1}|}(w_k)$ zurückzeigen. Diese Items werden alle von der Operation *Completor* nach erfolgreicher Abarbeitung von $B_1B_2\cdots B_lB'_l$ in die Liste $I^{\mathbf{G}}_{|p_1pm_1\cdots m_lm'_l|}(w_k)$ geliefert. Da $B'_l = A$ ist, kann sich dieser Vorgang beliebig oft wiederholen.

Wir gehen bei den weiteren Betrachtungen davon aus, daß $\mathbf{G}$ kein rechtsrekursives Nonterminal enthält.

3.2 Linksrekursion

Satz 3.2 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine reduzierte CFG in CNF mit:*

1. $(\exists A, B, C \in V)\,(f_1 : C \to AB \in P)$,

2. $(\exists (m, p, q) \in T^+ \times T^\star \times T^\star)\,(\forall s, t \in \mathcal{N})\,(A \Rightarrow^\star pm^s \wedge B \Rightarrow^\star m^tq)\,.$

Es gilt: $\mathbf{G} \in \mathbf{NIB}$.

Beweis: Wir betrachten zunächst die Ableitung $B \Rightarrow^\star m^t q$ nach Voraussetzung *2*. Da es nur endlich viele Produktionen in P mit linker Seite B gibt, existiert eine unendliche, streng wachsende Folge natürlicher Zahlen $t_1, t_2, \cdots, t_{l+1} > t_l \geq l \geq 1$ für die gilt:

$$B \Rightarrow \nu \Rightarrow^\star m^{t_l} q,\ t_l \geq l \geq 1\,. \tag{3.9}$$

Die Produktion $f : B \to \nu \in P$ ist also mögliche Startregel in allen Ableitungen $B \Rightarrow^\star m^{t_l} q$. Da $\mathbf{G}$ in ϵ-freier und reduzierter Form vorliegt, gilt:

$$(\exists \gamma_1, \delta_1 \in (V \cup T)^\star)\,(S \Rightarrow^\star \gamma_1 C \delta_1)\,, \tag{3.10}$$

$$(\exists p_1, q_1 \in T^\star)\,(\gamma_1 \Rightarrow^\star p_1 \,\wedge\, \delta_1 \Rightarrow^\star q_1)\,. \tag{3.11}$$

Betrachten wir nun das Wort $w_k := p_1 p m^k q q_1$, $k := s + t$. Es ist

$$S \overset{(3.10)}{\Rightarrow^\star} \gamma_1 C \delta_1 \overset{(3.11)}{\Rightarrow^\star} p_1 C q_1 \overset{\text{Vor. } 1.}{\Rightarrow} p_1 A B q_1 \overset{\text{Vor. } 2.}{\Rightarrow^\star} p_1 p m^k q q_1 \in T^\star\,,$$

d.h.

$$(\forall k \in \mathcal{N} \setminus \{1\})\,(w_k \in L(\mathbf{G}))\,.$$

Sei nun $k \in \mathcal{N} \setminus \{1\}$ und $w_k := a_1 \cdots a_n$, $a_i \in T$, $1 \leq i \leq n$. Wir betrachten die Listen, die das **EPV** für das Wort w_k aufbaut. Wir zeigen:

$$(\forall l \in [1 : l_r]\,)\,([B \to \nu\,\bullet, |p_1 p m^{k-t_l}|] \in I^{\mathbf{G}}_{|p_1 p m^k q|}(w_k))\,. \tag{3.12}$$

Beweis: Wir zeigen die Behauptung (3.12) durch Nachweis der Bedingungen von **Satz 1.1**. Hierzu setzen wir $\gamma := \gamma_1 A$, $\delta := \delta_1$, $\alpha := \nu$, $\beta := \epsilon$, $i := |p_1 p m^{k-t_l}|$ und $j := |p_1 p m^k q|$. Mit dieser Wahl folgt für alle $l \in [1 : l_r]$:

a) $S \overset{(3.10)}{\Rightarrow^\star} \gamma_1 C \delta_1 \overset{\text{Vor. } 1.}{\Rightarrow} \gamma_1 A B \delta_1 = \gamma B \delta\,,$

b) $\gamma = \gamma_1 A \overset{(3.11)}{\Rightarrow^\star} p_1 A \overset{\text{Vor. } 2.}{\Rightarrow^\star} p_1 p m^{k-t_l}$

$\quad = a_1 \cdots a_{|p_1 p m^{k-t_l}|} = a_1 \cdots a_i\,,$

c) $f : B \to \nu = f : B \to \alpha\beta \in P$ (nach Konstr.),

d) $\alpha = \nu \overset{(3.9)}{\Rightarrow^\star} m^{t_l} q$

$$= a_{|p_1 p m^{k-t_l}|+1} \cdots a_{|p_1 p m^k q|}$$

$$= a_{i+1} \cdots a_j \, .$$

Damit ist Behauptung (3.12) bewiesen. In Liste $I^{\mathbf{G}}_{|p_1 p m^k q|}(w_k)$ befinden sich also bei Anwendung des **EPV** auf w_k mindestens l_r Items, wobei l_r maximal ist mit $t_{l_r} < k$. Wegen (3.9) ist aber $t_{l_r} < k$ für wachsende k eine unbeschränkte Teilfolge von $t_1, t_2, \cdots$. Da für jedes $k \in \mathcal{N} \setminus \{1\}$ jedes Wort $w_k \in L(\mathbf{G})$ liegt, ist **G** nichtzustandtsbeschränkt, also $\mathbf{G} \in \mathbf{NIB}$ und der Satz ist bewiesen. ◇

In einer Ableitung des Wortes w_k kann das Teilwort $pm^k q$ nicht eindeutig auf die Nonterminals A und B aufgeteilt werden. Dies genau hat das Wachstum der Listen zur Folge. Wir gehen in den weiteren Betrachtungen davon aus, daß die Grammatik **G** die Voraussetzung *2.* aus **Satz 3.2** nicht erfüllt. In diesem Satz nahmen wir an, daß die Teilwörter m die gleiche Form haben. Wie der folgende Satz zeigt, wachsen die Itemlisten auch dann, wenn die Wörter m in Voraussetzung *2.* variieren.

Satz 3.3 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine reduzierte CFG in CNF mit*

1. $(\exists A, B, C \in V)\,(f_1 : C \to AB \in P)$,

2. $(\forall n \in \mathcal{N})(\exists k \geq n)(\exists (p, q) \in T^\star \times T^\star)(\forall s \in [1:k]\,\exists m_s \in T^+)\,.$

   $(\forall s \in [1 : k-1])$

   2.I $A \Rightarrow^\star pm_1 \cdots m_{k-s}\,,$

   2.II $B \Rightarrow^\star m_{k-s+1} \cdots m_k q\,.$

Es gilt: $\mathbf{G} \in \mathbf{NIB}$.

Beweis: Wir betrachten zunächst die Ableitung $B \Rightarrow^\star m_{k-s+1} \cdots m_k q$ nach Voraussetzung *2.II* . Sei P_B die Menge der Produktionen $f : B \to \nu \in P$. Jede Ableitung *2.II* beginnt mit einer Produktion $f \in P_B$. Für verschiedene $s \in [1 : k-1]$ sind dies möglicherweise verschiedene Regeln $f \in P_B$. Da *2.II* für wachsende $k \in \mathcal{N}$ und alle $s \in [1 : k-1]$ gilt, existiert eine unendliche, streng wachsende Folge narürlicher Zahlen $k_1, k_2, \cdots, k_{l+1} > k_l \geq l \geq 1$ für die gilt:

$$(\forall r \in [1 : l])\,(\exists s_r \in [1 : k_l - 1])(B \Rightarrow \nu \Rightarrow^\star m_{k_l - s_r + 1} \cdots m_{k_l} q)\,. \quad (3.13)$$

Die Produktion $f : B \to \nu \in P$ ist also mögliche Startregel in allen Ableitungen $B \Rightarrow^\star m_{k_l - s_r + 1} \cdots m_{k_l} q$. Da $\mathbf{G}$ in ϵ-freier und reduzierter Form vorliegt, gilt:

$$(\exists \gamma_1, \delta_1 \in (V \cup T)^\star)\,(S \Rightarrow^\star \gamma_1 C \delta_1)\,, \quad (3.14)$$

$$(\exists p_1, q_1 \in T^\star)\,(\gamma_1 \Rightarrow^\star p_1 \wedge \delta_1 \Rightarrow^\star q_1)\,. \quad (3.15)$$

Betrachten wir das Wort $w_k := p_1 p m_1 \cdots m_k q q_1$. Es ist

$$S \overset{(3.14)}{\Rightarrow^\star} \gamma_1 C \delta_1 \overset{(3.15)}{\Rightarrow^\star} p_1 C q_1 \overset{\text{Vor. } 1.}{\Rightarrow} p_1 A B q_1 \overset{\text{Vor. } 2.}{\Rightarrow^\star} p_1 p m_1 \cdots m_k q q_1 \in T^\star,$$

d.h.

$$(\forall n \in \mathcal{N}\,\exists k \geq n)\,(w_k \in L(\mathbf{G}))\,.$$

Sei nun $k_l \in \mathcal{N} \setminus \{1\}$ und $w_{k_l} := a_1 \cdots a_n$, $a_i \in T$, $1 \leq i \leq n$. Wir betrachten die Listen, die das **EPV** für das Wort w_{k_l} aufbaut und zeigen:

$$(\forall r \in [1:l])\,(\exists s_r \in [1:k_l-1])$$
$$([B \to \nu\,\bullet, |p_1pm_1 \cdots m_{k_l-s_r}|] \in I^{\mathbf{G}}_{|p_1pm_1\cdots m_{k_l}q|}(w_{k_l})). \tag{3.16}$$

Beweis: Wir zeigen die Behauptung (3.16) durch Nachweis der Bedingungen von **Satz 1.1**. Hierzu setzen wir $\gamma := \gamma_1 A$, $\delta := \delta_1$, $\alpha := \nu$, $\beta := \epsilon$, $i := |p_1pm_1 \cdots m_{k_l-s_r}|$ und $j := |p_1pm_1 \cdots m_{k_l}q|$. Mit dieser Wahl folgt für alle $r \in [1:l]$ und $s_r \in [1:k_l-1]$:

a) $S \overset{(3.14)}{\Rightarrow^\star} \gamma_1 C \delta_1 \overset{\text{Vor. }1.}{\Rightarrow^\star} \gamma_1 A B \delta_1 = \gamma B \delta\,,$

b) $\gamma \overset{(3.15)}{\Rightarrow^\star} p_1 A \overset{\text{Vor. }2.I}{\Rightarrow^\star} p_1pm_1 \cdots m_{k_l-s_r}$

$= a_1 \cdots a_{|p_1pm_1\cdots m_{k_l-s_r}|}$

$= a_1 \cdots a_i\,,$

c) $f : B \to \nu = f : B \to \alpha\beta \in P$ (nach Konstr.),

d) $\alpha = \nu \overset{(3.13)}{\Rightarrow^\star} m_{k_l-s_r+1} \cdots m_{k_l} q$

$= a_{|p_1pm_1\cdots m_{k_l-s_r}|+1} \cdots a_{|p_1pm_1\cdots m_{k_l}q|}$

$= a_{i+1} \cdots a_j\,.$

Damit ist Behauptung (3.16) bewiesen. In Liste $I^{\mathbf{G}}_{|p_1pm_1\cdots m_{k_l}q|}(w_{k_l})$ befinden sich also bei Anwendung des **EPV** auf w_{k_l} mindestens $l-1$ Items. Da für jedes $n \in \mathcal{N}$ ein Wort w_{k_l} mit $k_l \geq n \in L(\mathbf{G})$ liegt, ist **G** nichtzustandtsbeschränkt, also $\mathbf{G} \in$ **NIB** und **Satz 3.3** ist bewiesen. $\diamond$

In einer Ableitung des Wortes w_k kann das Teilwort $pm_1 \cdots m_k q$ nicht eindeutig auf die Nonterminals A und B aufgeteilt werden. Dies genau hat das Wachstum der Listen zur Folge. Wir gehen in den weiteren Betrachtungen davon aus, daß **G** keine der Voraussetzungen *2.* aus **Satz 3.3** erfüllt. Im folgenden Abschnitt betrachten wir Grammatiken mit echt rekursiven Nonterminals.

3.3 Echte Rekursion

Satz 3.4 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine reduzierte CFG in CNF und enthalte* $\mathbf{G}$ *das echt rekursive Nonterminal* C *mit:*

1. $C \Rightarrow^\star \nu C \mu\,,$

2. $(\exists m \in T^+)\,(\nu \Rightarrow^\star m \,\wedge\, C \Rightarrow^\star m \,\wedge\, \mu \Rightarrow^\star m)\,.$

Es gilt: $\mathbf{G} \in \mathbf{NIB}$.

Beweis: Weil $\mathbf{G}$ in CNF vorliegt und reduziert ist, folgt:

$$(\exists \gamma_1, \delta_1 \in (V \cup T)^\star)\,(S \Rightarrow^\star \gamma_1 C \delta_1)\,, \tag{3.17}$$

$$(\exists p_1, q_1 \in T^\star)\,(\gamma_1 \Rightarrow^\star p_1 \wedge \delta_1 \Rightarrow^\star q_1)\,. \tag{3.18}$$

Wir betrachten das Wort $w_k := p_1 m^{2k+1} q_1$. Weil $\mathbf{G}$ das echt rekursive Nonterminal C enthält, ist

$$S \overset{(3.17)}{\Rightarrow^\star} \gamma_1 C \delta_1 \overset{(3.18)}{\Rightarrow^\star} p_1 C q_1 \overset{\text{Vor. }1.}{\Rightarrow^\star} p_1 \nu^k C \mu^k q_1 \overset{\text{Vor. }2.}{\Rightarrow^\star} p_1 m^{2k+1} q_1\,,$$

d.h

$$(\forall k \in \mathcal{N}_0)\,(w_k \in L(\mathbf{G}))\,.$$

Um aus C die Satzform $\nu C \mu$ abzuleiten, gibt es zwei Möglichkeiten für $X, Y \in V$:

Fall 1: $C \Rightarrow^\star \nu X \,\wedge\, f : X \to CY \in P \,\wedge\, Y \Rightarrow^\star \mu\,.$

Fall 2: $C \Rightarrow^\star Y \mu \,\wedge\, f : Y \to XC \in P \,\wedge\, X \Rightarrow^\star \nu\,.$

Sei $w_k := a_1 \cdots a_n$, $a_i \in T$, $1 \le i \le n$. Wir betrachten die Itemlisten, die das **EPV** für das Wort $w_k \in L(\mathbf{G})$ aufbaut, für beide Fälle und zeigen:

Fall 1: $(\forall s \in [1:k])\,([X \to CY\,\bullet, |p_1 m^{2k-2s+1}|] \in I^{\mathbf{G}}_{|p_1 m^{2k+1}|}(w_k))$. (3.19)

Beweis : Wir zeigen die Behauptung (3.19) durch Nachweis der Bedingungen von **Satz 1.1**. Hierzu setzen wir: $\gamma := \gamma_1 \nu^{2k-2s+1}$, $\delta := \mu^{2k-2s} \delta_1$, $\alpha :=$

CY, $\beta := \epsilon$, $i := |p_1 m^{2k-2s+1}|$ und $j := |p_1 m^{2k+1}|$. Mit dieser Wahl folgt für alle $s \in [1:k]$:

$$
\begin{aligned}
&\text{(a)} \quad S \overset{(3.17)}{\Rightarrow^\star} \gamma_1 C \delta_1 \\
&\qquad \overset{\text{Vor. }\mathit{1}.}{\Rightarrow^\star} \gamma_1 \nu^{2k-2s} C \mu^{2k-2s} \delta_1 \\
&\qquad \overset{\text{Fall 1:}}{\Rightarrow^\star} \gamma_1 \nu^{2k-2s} \nu X \mu^{2k-2s} \delta_1 \\
&\qquad = \gamma_1 \nu^{2k-2s+1} X \mu^{2k-2s} \delta_1 = \gamma X \delta\,, \\
&\text{(b)} \quad \gamma = \gamma_1 \nu^{2k-2s+1} \overset{(3.18)}{\Rightarrow^\star} p_1 \nu^{2k-2s+1} \\
&\qquad \overset{\text{Vor. }\mathit{2}.}{\Rightarrow^\star} p_1 m^{2k-2s+1} \\
&\qquad = a_1 \cdots a_{|p_1 m^{2k-2s+1}|} = a_1 \cdots a_i\,, \\
&\text{(c)} \quad f : X \to CY = f : X \to \alpha\beta \in P \ \ (\text{nach Konstr.})\,, \\
&\text{(d)} \quad \alpha = CY \overset{\text{Vor. }\mathit{1}.}{\Rightarrow^\star} \nu^{s-1} C \mu^{s-1} Y \\
&\qquad \overset{\text{Vor. }\mathit{2}.}{\Rightarrow^\star} m^{s-1} m m^{s-1} Y = m^{2s-1} Y \\
&\qquad \overset{\text{Fall 1:}}{\Rightarrow^\star} m^{2s-1} \mu \overset{\text{Vor. }\mathit{2}.}{\Rightarrow^\star} m^{2s-1} m = m^{2s} \\
&\qquad = a_{|p_1 m^{2k-2s+1}|+1} \cdots a_{|p_1 m^{2k+1}|} = a_{i+1} \cdots a_j\,.
\end{aligned}
$$

Damit ist Behauptung (3.19) bewiesen.

Im anderen Fall zeigen wir:

$$\textbf{Fall 2}: \ (\forall s \in [1:k])\,([Y \to XC\,\bullet\,, |p_1 m^{2k-2s+1}|] \in I^{\mathbf{G}}_{|p_1 m^{2k+1}|}(w_k)). \quad (3.20)$$

Hierzu setzen wir: $\gamma := \gamma_1 \nu^{2k-2s+1}$, $\delta := \mu^{2k-2s+2}\delta_1$, $\alpha := XC$, $\beta := \epsilon$, $i := |p_1 m^{2k-2s+1}|$ und $j := |p_1 m^{2k+1}|$. Mit dieser Wahl folgt für alle $s \in [1:k]$:

$$
\begin{aligned}
&\text{(a)} \quad S \overset{(3.17)}{\Rightarrow^\star} \gamma_1 C \delta_1 \overset{\text{Vor. }\mathit{1}.}{\Rightarrow^\star} \gamma_1 \nu^{2k-2s+1} C \mu^{2k-2s+1} \delta_1 \\
&\qquad \overset{\text{Fall 2:}}{\Rightarrow^\star} \gamma_1 \nu^{2k-2s+1} Y \mu \mu^{2k-2s+1} \delta_1 \\
&\qquad = \gamma_1 \nu^{2k-2s+1} Y \mu^{2k-2s+2} \delta_1 = \gamma Y \delta\,, \\
&\text{(b)} \quad \gamma = \gamma_1 \nu^{2k-2s+1} \overset{(3.18)}{\Rightarrow^\star} p_1 \nu^{2k-2s+1} \overset{\text{Vor. }\mathit{2}.}{\Rightarrow^\star} p_1 m^{2k-2s+1} \\
&\qquad = a_1 \cdots a_{|p_1 m^{2k-2s+1}|} = a_1 \cdots a_i\,,
\end{aligned}
$$

$$\begin{aligned}
&\text{(c)} \quad f : Y \to XC = f : Y \to \alpha\beta \in P \;\; \text{(nach Konstr.)}\,,\\
&\text{(d)} \quad \alpha = XC \overset{\text{Vor. }\mathit{1}.}{\Rightarrow^\star} X\nu^{s-1}C\mu^{s-1} \overset{\text{Vor. }\mathit{2}.}{\Rightarrow^\star} Xm^{s-1}mm^{s-1}\\
&\qquad = Xm^{2s-1} \overset{\text{Fall 2:}}{\Rightarrow^\star} \nu m^{2s-1} \overset{\text{Vor. }\mathit{2}.}{\Rightarrow^\star} mm^{2s-1} = m^{2s}\\
&\qquad = a_{|p_1 m^{2k-2s+1}|+1} \cdots a_{|p_1 m^{2k+1}|} = a_{i+1} \cdots a_j \,.
\end{aligned}$$

Damit ist Behauptung (3.20) bewiesen. In beiden Fällen befinden sich in Liste $I^{\mathbf{G}}_{|p_1 m^{2k+1}|}(w_k)$ bei Anwendung des **EPV** auf w_k mindestens k Items. Da für jedes $k \in \mathcal{N}_0$ das Wort $w_k \in L(\mathbf{G})$ liegt, ist **G** nichtzustandsbeschränkt. Somit gilt $\mathbf{G} \in$ **NIB** und **Satz 3.4** ist bewiesen. $\diamond$

Wir gehen im Folgenden davon aus, daß die Sprache $L(\nu) \cap L(C) \cap L(\mu)$ für jedes echt rekursive Nonterminal C der Grammatik **G** leer ist.

3.4 Echte Rekursion und Linksrekursion

In diesem Abschnitt betrachten wir Grammatiken, die ein Nonterminal C enthalten, das sowohl echt rekursiv als auch linksrekursiv ist. Wir nehmen an, daß C entweder mit der Satzform ν oder der Satzform μ ein gemeinsames Terminalwort ableitet, die Sprache $L(\nu) \cap L(C) \cap L(\mu)$ aber leer ist.

Satz 3.5 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine reduzierte CFG in CNF und enthalte* **G** *das echt rekursive Nonterminal* C *mit:*

1. $C \Rightarrow^\star \nu C \mu$ *,*

2. $(\exists (m, q) \in T^+ \times T^\star)\, (\forall t \in \mathcal{N})\, (\nu \Rightarrow^\star m \,\wedge\, C \Rightarrow^\star m^t q)$ *.*

Es gilt: $\mathbf{G} \in$ **NIB**.

Beweis: Wir betrachten zunächst die Ableitung $C \Rightarrow^\star m^t q$ nach Voraussetzung *2.* . Da es nur endlich viele Produktionen in P mit linker Seite C gibt, existiert eine unendliche, streng wachsende Folge natürlicher Zahlen

$t_1, t_2, \cdots, t_{l+1} > t_l \geq l \geq 1$, für die gilt:

$$C \Rightarrow \eta \Rightarrow^\star m^{t_l} q,\ t_l \geq l \geq 1\,. \tag{3.21}$$

Die Produktion $f : C \to \eta \in P$ ist also mögliche Startregel in allen Ableitungen $C \Rightarrow^\star m^{t_l} q$. Da $\mathbf{G}$ in ϵ-freier und reduzierter Form vorliegt, gilt:

$$(\exists \gamma_1, \delta_1 \in (V \cup T)^\star)\,(S \Rightarrow^\star \gamma_1 C \delta_1)\,, \tag{3.22}$$

$$(\exists p_1, q_1 \in T^\star)\,(\gamma_1 \Rightarrow^\star p_1 \,\wedge\, \delta_1 \Rightarrow^\star q_1)\,, \tag{3.23}$$

$$(\exists u \in T^+)\,(\mu \Rightarrow^\star u)\,. \tag{3.24}$$

Betrachten wir nun das Wort $w_{k_s} := p_1 m^{k_s} q u^s q_1$, $k_s := s + t$. Es ist

$$\begin{aligned}
& S \overset{(3.22)}{\Rightarrow^\star} \gamma_1 C \delta_1 \overset{(3.23)}{\Rightarrow^\star} p_1 C q_1 \overset{\text{Vor. } \mathit{1}.}{\Rightarrow^\star} p_1 \nu^s C \mu^s q_1 \\
& \overset{\text{Vor. } \mathit{2}.}{\Rightarrow^\star} p_1 m^s C \mu^s q_1 \overset{(3.24)}{\Rightarrow^\star} p_1 m^s C u^s q_1 \overset{\text{Vor. } \mathit{2}.}{\Rightarrow^\star} p_1 m^s m^t q u^s q_1 \\
& = p_1 m^{k_s} q u^s q_1,\ 1 \leq s < k_s\,,
\end{aligned}$$

d.h.

$$(\forall k \in \mathcal{N} \setminus \{1\})\,(\forall s \in [1 : k-1])(w_{k_s} \in L(\mathbf{G}))\,.$$

Sei nun $k_s \in \mathcal{N} \setminus \{1\}$ und $w_{k_s} := a_1 \cdots a_n$, $a_i \in T$, $1 \leq i \leq n$, $1 \leq s < k$. Wir betrachten die Listen, die das **EPV** für das Wort w_{k_s} aufbaut und zeigen:

$$(\forall l \in [1 : l_r]\,)\,([C \to \eta\,\bullet, |p_1 m^{k_s - t_l}|] \in I^{\mathbf{G}}_{|p_1 m^{k_s} q|}(w_{k_s}))\,. \tag{3.25}$$

Beweis: Wir zeigen die Behauptung (3.25) durch Nachweis der Bedingungen von **Satz 1.1**. Hierzu setzen wir $\gamma := \gamma_1 \nu^{k_s - t_l} C$, $\delta := \mu^{k_s - t_l} \delta_1$, $\alpha := \eta$, $\beta := \epsilon$, $i := |p_1 m^{k_s - t_l}|$ und $j := |p_1 m^{k_s} q|$. Mit dieser Wahl folgt:

a) $S \overset{(3.22)}{\Rightarrow^\star} \gamma_1 C \delta_1 \overset{\text{Vor. } \mathit{1}.}{\Rightarrow} \gamma_1 \nu^{k_s - t_l} C \mu^{k_s - t_l} \delta_1 = \gamma C \delta\,,$

b) $\gamma \overset{(3.23)}{\Rightarrow^\star} p_1 \nu^{k_s - t_l} \overset{\text{Vor. } \mathit{2}.}{\Rightarrow^\star} p_1 m^{k_s - t_l}$

$$= a_1 \cdots a_{|p_1 m^{k_s - t_l}|} = a_1 \cdots a_i\,,$$

c) $f : C \to \eta = f : C \to \alpha\beta \in P$ (nach Konstr.),

d) $\alpha = \eta \overset{(3.21)}{\Rightarrow^\star} m^{t_l} q = a_{|p_1 m^{k_s - t_l}|+1} \cdots a_{|p_1 m^{k_s} q|}$

$$= a_{i+1} \cdots a_j\,.$$

Damit ist Behauptung (3.25) bewiesen. In Liste $I^{\mathbf{G}}_{|p_1 m^{k_s} q|}(w_{k_s})$ befinden sich also bei Anwendung des **EPV** auf w_{k_s} mindestens l_r Items, wobei l_r maximal ist mit $t_{l_r} < k_s$. Wegen (3.21) ist aber $t_{l_r} < k$ für wachsende k eine unbeschränkte Teilfolge von $t_1, t_2, \cdots$. Da für jedes $k \in \mathcal{N} \setminus \{1\}$ und jedes $s \in [1 : k-1]$ jedes Wort $w_{k_s} \in L(\mathrm{G})$ liegt, ist **G** nichtzustandtsbeschränkt, also $\mathbf{G} \in \mathbf{NIB}$ und **Satz 3.5** ist bewiesen. $\diamond$

Satz 3.6 *Sei* $\mathbf{G} := (V, T, S, P)$ *eine reduzierte CFG in CNF und enthalte* **G** *das echt rekursive Nonterminal* C *mit:*

1. $C \Rightarrow^\star \nu C \mu\,,$

2. $(\exists (m,p) \in T^+ \times T^\star)\,(\forall s \in \mathcal{N})\,(\mu \Rightarrow^\star m \wedge C \Rightarrow^\star p m^s)\,.$

Es gilt: $\mathbf{G} \in \mathbf{NIB}$.

Beweis: Weil **G** in reduzierter Form und in CNF vorliegt, folgt:

$$(\exists \gamma_1, \delta_1 \in (V \cup T)^\star)\,(S \Rightarrow^\star \gamma_1 C \delta_1)\,, \tag{3.26}$$

$$(\exists p_1, q_1 \in T^\star)\,(\gamma_1 \Rightarrow^\star p_1 \wedge \delta_1 \Rightarrow^\star q_1)\,, \tag{3.27}$$

$$(\exists u \in T^+)\,(\nu \Rightarrow^\star u)\,. \tag{3.28}$$

Betrachten wir das Wort $w_k := p_1 u^k p m^{2k} q_1$. Weil **G** das echt rekursive Nonterminal C enthält, ist

$$\begin{aligned} & S \overset{(3.26)}{\Rightarrow^\star} \gamma_1 C \delta_1 \overset{(3.27)}{\Rightarrow^\star} p_1 C q_1 \overset{\text{Vor. } 1.}{\Rightarrow^\star} p_1 \nu^k C \mu^k q_1 \\ & \overset{(3.28)}{\Rightarrow^\star} p_1 u^k C \mu^k q_1 \overset{\text{Vor. } 2.}{\Rightarrow^\star} p_1 u^k C m^k q_1 \overset{\text{Vor. } 2.}{\Rightarrow^\star} p_1 u^k p m^k m^k q_1 = w_k\,, \end{aligned}$$

d.h

$$(\forall k \in \mathcal{N} \setminus \{1\})\,(w_k \in L(\mathbf{G}))\,.$$

Um aus C die Satzform $\nu C\mu$ abzuleiten, gibt es zwei Möglichkeiten für $X, Y \in V$:

Fall 1: $C \Rightarrow^\star \nu X \wedge f : X \to CY \in P \wedge Y \Rightarrow^\star \mu\,.$

Fall 2: $C \Rightarrow^\star Y\mu \wedge f : Y \to XC \in P \wedge X \Rightarrow^\star \nu\,.$

Sei $w_k := a_1 \cdots a_n$, $a_i \in T$, $1 \le i \le n$. Wir betrachten die Itemlisten, die das **EPV** für das Wort w_k aufbaut, für die beiden Fälle getrennt und zeigen:

$$\textbf{Fall 1}: (\forall s \in [1:k])\,([X \to C\bullet Y, |p_1u^{k-s+1}|] \in I^{\mathbf{G}}_{|p_1u^kpm^k|}(w_k))\,. \tag{3.29}$$

Wir beweisen Behauptung (3.29) durch Nachweis der Bedingungen von **Satz 1.1**. Hierzu setzen wir: $\gamma := \gamma_1\nu^{k-s+1}$, $\delta := \mu^{k-s}\delta_1$, $\alpha := C$, $\beta := Y$, $i := |p_1u^{k-s+1}|$ und $j := |p_1u^kpm^k|$. Mit dieser Wahl folgt für alle $s \in [1:k]$:

(a) $$S \overset{(3.26)}{\Rightarrow^\star} \gamma_1C\delta_1 \overset{\text{Vor. } \mathit{1}.}{\Rightarrow^\star} \gamma_1\nu^{k-s}C\mu^{k-s}\delta_1 \overset{\text{Fall 1:}}{\Rightarrow^\star} \gamma_1\nu^{k-s}\nu X\mu^{k-s}\delta_1 = \gamma_1\nu^{k-s+1}X\mu^{k-s}\delta_1 = \gamma X\delta\,,$$

(b) $$\gamma = \gamma_1\nu^{k-s+1} \overset{(3.27)}{\Rightarrow^\star} p_1\nu^{k-s+1} \overset{(3.28)}{\Rightarrow^\star} p_1u^{k-s+1} = a_1\cdots a_{|p_1u^{k-s+1}|} = a_1\cdots a_i\,,$$

(c) $f : X \to CY = f : X \to \alpha\beta \in P$ (nach Konstr.),

(d) $$\alpha = C \overset{\text{Vor. } \mathit{1}.}{\Rightarrow^\star} \nu^{s-1}C\mu^{s-1} \overset{(3.28)}{\Rightarrow^\star} u^{s-1}C\mu^{s-1} \overset{\text{Vor. } \mathit{2}.}{\Rightarrow^\star} u^{s-1}Cm^{s-1} \overset{\text{Vor. } \mathit{2}.}{\Rightarrow^\star} u^{s-1}pm^{k-s+1}m^{s-1} = u^{s-1}pm^k = a_{|p_1u^{k-s+1}|+1}\cdots a_{|p_1u^kpm^k|} = a_{i+1}\cdots a_j\,.$$

Damit ist Behauptung (3.29) bewiesen. Im anderen Fall zeigen wir:

$$\textbf{Fall 2}: (\forall s \in [1:k])\,([Y \to XC\bullet, |p_1u^{k-s}|] \in I^{\mathbf{G}}_{|p_1u^kpm^k|}(w_k))\,. \tag{3.30}$$

Wir setzen: $\gamma := \gamma_1\nu^{k-s}$, $\delta := \mu^{k-s+1}\delta_1$, $\alpha = XC$, $\beta := \epsilon$, $i := |p_1u^{k-s}|$ und $j := |p_1u^kpm^k|$. Mit dieser Wahl folgt für alle $s \in [1:k]$:

(a) $$S \overset{(3.26)}{\Rightarrow^\star} \gamma_1C\delta_1 \overset{\text{Vor. } \mathit{1}.}{\Rightarrow^\star} \gamma_1\nu^{k-s}C\mu^{k-s}\delta_1$$

$$\overset{\text{Fall 2:}}{\Rightarrow^\star} \gamma_1 \nu^{k-s} Y \mu \mu^{k-s} \delta_1$$

$$= \gamma_1 \nu^{k-s} Y \mu^{k-s+1} \delta_1 = \gamma Y \delta\,,$$

(b) $\gamma = \gamma_1 \nu^{k-s} \overset{(3.27)}{\Rightarrow^\star} p_1 \nu^{k-s} \overset{(3.28)}{\Rightarrow^\star} p_1 u^{k-s}$

$$= a_1 \cdots a_{|p_1 u^{k-s}|} = a_1 \cdots a_i\,,$$

(c) $f : Y \to XC = f : Y \to \alpha\beta \in P$ (nach Konstr.) ,

(d) $\alpha = XC \overset{\text{Vor. }2.}{\Rightarrow^\star} X\nu^{s-1} C \mu^{s-1} \overset{(3.28)}{\Rightarrow^\star} X u^{s-1} C \mu^{s-1}$

$$\overset{\text{Vor. }2.}{\Rightarrow^\star} X u^{s-1} C m^{s-1} \overset{\text{Fall 2:}}{\Rightarrow^\star} \nu u^{s-1} C m^{s-1}$$

$$\overset{\text{Fall 2:}}{\Rightarrow^\star} u u^{s-1} C m^{s-1} = u^s C m^{s-1}$$

$$\overset{\text{Vor. }2.}{\Rightarrow^\star} u^s p m^{k-s+1} m^{s-1} = u^s p m^k$$

$$= a_{|p_1 u^{k-s}|+1} \cdots a_{|p_1 u^k p m^k|} = a_{i+1} \cdots a_j\,.$$

Damit ist Behauptung (3.30) bewiesen. In beiden Fällen befinden sich in Liste $I^{\mathbf{G}}_{|p_1 u^k p m^k|}(w_k)$ bei Anwendung des **EPV** auf das Wort w_k mindestens k Items. Da für jedes $k \in \mathcal{N} \setminus \{1\}$ jedes Wort $w_k \in L(\mathbf{G})$ liegt, ist **G** nichtzustandtsbeschränkt, also $\mathbf{G} \in$ **NIB** und **Satz 3.6** ist bewiesen. $\diamond$

Wie bereits am Ende von Kapitel 2 bemerkt, können die hier gewonnenen Ergebnisse auf allgemeine kontextfreie Grammatiken übertragen werden. Im letzten Kapitel wollen wir die Sätze miteinander vergleichen und anhand von Beispielen veranschaulichen. Von Interesse ist hierbei die Entscheidbarkeit und die Frage der gegenseitigen Unabhängigkeit der in **Satz 3.1** bis **3.6** bewiesenen Kriterien.

Kapitel 4

Bewertung der Ergebnisse

4.1 Allgemeines

Im letzten Kapitel haben wir einige Eigenschaften kontextfreier Grammatiken und ihrer Sprachen bestimmt, die zur Folge haben, daß die Anwendung des **EPV** das Wortproblem nicht unter $0(n^2)$ Schritten löst[1]. Wir wollen in diesem Kapitel die Ergebnisse in praktischer Hinsicht untersuchen. Hierzu gehört die Frage nach der Entscheidbarkeit der in **Satz 3.1** bis **3.6** bewiesenen Eigenschaften, sowie deren Unabhängigkeit voneinander.

4.2 Rechtsrekursion

Satz 3.1 besagt, daß jede CFG, die ein rechtsrekursives Nonterminal enthält, aus der Menge **NIB** ist.

Für jede CFG **G** ist die Menge der
rechtsrekursiven Nonterminals entscheidbar. (4.1)

Hierzu konstruieren wir die Hilfsmengen $\mathcal{U}_A$ für jedes $A \in V$ wie folgt:

1. $\mathcal{U}_{A,0} \quad := \{B \in V \mid f : A \to \alpha B \in P,\, \alpha \in (V \cup T)^\star\}$

[1] siehe Aussage (1.2) und (1.3) in Kapitel 1.

2. $\mathcal{U}_{A,i+1} := \mathcal{U}_{A,i} \cup \{B \in V \mid f : C \to \alpha B \in P,\, \alpha \in (V \cup T)^\star,\, C \in \mathcal{U}_{A,i}\}$.

3. Wir setzen $\mathcal{U}_A$ auf $\mathcal{U}_{A,i}$, falls $\mathcal{U}_{A,i} = \mathcal{U}_{A,i+1}$ ist.

Das beschriebene Verfahren bricht nach endlich vielen Schritten ab, und nach Konstruktion von $\mathcal{U}_A$ ist $B \in \mathcal{U}_A$ genau dann wenn eine Ableitung $A \Rightarrow^\star \gamma B$ existiert, und somit ist A rechtsrekursiv falls $A \in \mathcal{U}_A$ liegt.

Beispiel 4.2 Sei $\mathbf{G}_1 := (\{S, A\}, \{a, b\}, S, \{f_1 : S \to AS,$
$f_2 : S \to a,\ f_3 : A \to b\})$.

Die Sprache $L(\mathbf{G}_1)$ besteht aus Wörtern der Form $b^\star a$, und $\mathbf{G}_1$ enthält weder ein links- noch ein echt rekursives Nonterminal, sondern das rechtsrekusive Startsymbol S und erfüllt somit nur die Bedingungen von **Satz 3.1**.

Betrachten wir die Itemlisten, die das **EPV** für das Wort $w_6 := b^6 a$ aufbaut:

Liste $I_0(b^6a)$	Liste $I_1(b^6a)$	Liste $I_2(b^6a)$	Liste $I_3(b^6a)$
$[S \to \bullet AS, 0]$	$[A \to b \bullet, 0]$	$[A \to b \bullet, 1]$	$[A \to b \bullet, 2]$
$[S \to \bullet a, 0]$	$[S \to A \bullet S, 0]$	$[S \to A \bullet S, 1]$	$[S \to A \bullet S, 2]$
$[A \to \bullet b, 0]$	$[S \to \bullet AS, 1]$	$[S \to \bullet AS, 2]$	$[S \to \bullet AS, 3]$
	$[S \to \bullet a, 1]$	$[S \to \bullet a, 2]$	$[S \to \bullet a, 3]$
	$[A \to \bullet b, 1]$	$[A \to \bullet b, 2]$	$[A \to \bullet b, 3]$

Liste $I_4(b^6a)$	Liste $I_5(b^6a)$	Liste $I_6(b^6a)$	Liste $I_7(b^6a)$
$[A \to b \bullet, 3]$	$[A \to b \bullet, 4]$	$[A \to b \bullet, 5]$	$[S \to a \bullet, 6]$
$[S \to A \bullet S, 3]$	$[S \to A \bullet S, 4]$	$[S \to A \bullet S, 5]$	$[S \to AS \bullet, 5]$
$[S \to \bullet AS, 4]$	$[S \to \bullet AS, 5]$	$[S \to \bullet AS, 6]$	$[S \to AS \bullet, 4]$
$[S \to \bullet a, 4]$	$[S \to \bullet a, 5]$	$[S \to \bullet a, 6]$	$[S \to AS \bullet, 3]$
$[A \to \bullet b, 4]$	$[A \to \bullet b, 5]$	$[A \to \bullet b, 6]$	$[S \to AS \bullet, 2]$
			$[S \to AS \bullet, 1]$
			$[S \to AS \bullet, 0]$

Da sich das Item $[S \to AS \bullet, 0]$ in $I_7^{\mathbf{G}_1}(b^6 a)$ befindet, ist das Wort $w_6 = b^6 a$

in der Sprache $L(\mathbf{G}_1)$. In den Listen $I_1(b^k a)$ bis $I_{|b^k|}(b^k a)$ verweist die zweite Komponente der Items nur eine Liste zurück, so daß *Completor*$(I_{|b^k a|})$ alle k Listen $I_{|b^k|}(b^k a)$ bis $I_1(b^k a)$ nacheinander besuchen muß, um von dort jeweils ein Item $[S \to AS\bullet, s]$ für $s \in [0 : k-1]$ in Liste $I_{|ba^k|}$ zu erzeugen.

> Das Kriterium von **Satz 3.1** ist entscheidbar und von den anderen Kriterien unabhängig. (4.2)

4.3 Linksrekursion

Satz 3.2 besagt, daß eine CFG $\mathbf{G} \in \mathbf{NIB}$ ist, falls sie eine Produktion $f : C \to AB \in P$ enthält und A das Wort pm^s und B das Wort $m^t q$ für alle $s, t \in \mathcal{N}$ erzeugt.

Beispiel 4.3 Sei $\mathbf{G}_2 := (\{S, A, B, X, Y\}, \{a, b\}, S, P_2)$ mit $P_2 :=$

$$\begin{array}{lllll} \{ & f_1 : S \to AB, & f_2 : A \to AX, & f_3 : B \to BY, & f_4 : A \to a, \\ & f_5 : X \to b, & f_6 : B \to b, & f_7 : Y \to b, & f_8 : Y \to a \ \} \end{array}$$

Es ist $L(\mathbf{G}_2) = ab^+(a \cup b)^\star = ab(a \cup b)^\star$. Betrachten wir die Listen, die das **EPV** für das Wort $w_k := ab^k$ und $k := 6$ erzeugt:

Liste $I_0(ab^6)$	Liste $I_1(ab^6)$	Liste $I_2(ab^6)$	Liste $I_3(ab^6)$
$[S \to \bullet AB, 0]$	$[A \to a\bullet, 0]$	$[B \to b\bullet, 1]$	$[Y \to b\bullet, 2]$
$[A \to \bullet AX, 0]$	$[S \to A\bullet B, 0]$	$[X \to b\bullet, 1]$	$[B \to b\bullet, 2]$
$[A \to \bullet a, 0]$	$[A \to A\bullet X, 0]$	$[S \to AB\bullet, 0]$	$[X \to b\bullet, 2]$
	$[B \to \bullet BY, 1]$	$[B \to B\bullet Y, 1]$	$[B \to BY\bullet, 1]$
	$[B \to \bullet b, 1]$	$[A \to AX\bullet, 0]$	$[S \to AB\bullet, 0]$
	$[X \to \bullet b, 1]$	$[S \to A\bullet B, 0]$	$[B \to B\bullet Y, 2]$
		$[A \to A\bullet X, 0]$	$[A \to AX\bullet, 0]$
		$[Y \to \bullet a, 2]$	$[B \to B\bullet Y, 1]$
		$[Y \to \bullet b, 2]$	$[S \to A\bullet B, 0]$
		$[B \to \bullet BY, 2]$	$[A \to A\bullet X, 0]$
		$[B \to \bullet b, 2]$	$[Y \to \bullet a, 3]$
		$[X \to \bullet b, 2]$	$[Y \to \bullet b, 3]$
			$[B \to \bullet BY, 3]$
			$[B \to \bullet b, 3]$
			$[X \to \bullet b, 3]$

Liste $I_4(ab^6)$	Liste $I_5(ab^6)$	Liste $I_6(ab^6)$	Liste $I_7(ab^6)$
$[Y \to b\,\bullet, 3]$	$[Y \to b\,\bullet, 4]$	$[Y \to b\,\bullet, 5]$	$[Y \to b\,\bullet, 6]$
$[B \to b\,\bullet, 3]$	$[B \to b\,\bullet, 4]$	$[B \to b\,\bullet, 5]$	$[B \to b\,\bullet, 6]$
$[X \to b\,\bullet, 3]$	$[X \to b\,\bullet, 4]$	$[X \to b\,\bullet, 5]$	$[X \to b\,\bullet, 6]$
$[B \to BY\,\bullet, 2]$	$[B \to BY\,\bullet, 3]$	$[B \to BY\,\bullet, 4]$	$[B \to BY\,\bullet, 5]$
$[B \to BY\,\bullet, 1]$	$[B \to BY\,\bullet, 2]$	$[B \to BY\,\bullet, 3]$	$[B \to BY\,\bullet, 4]$
$[S \to AB\,\bullet, 0]$	$[B \to BY\,\bullet, 1]$	$[B \to BY\,\bullet, 2]$	$[B \to BY\,\bullet, 3]$
$[B \to B\,\bullet Y, 3]$	$[S \to AB\,\bullet, 0]$	$[B \to BY\,\bullet, 1]$	$[B \to BY\,\bullet, 2]$
$[A \to AX\,\bullet, 0]$	$[B \to B\,\bullet Y, 4]$	$[S \to AB\,\bullet, 0]$	$[B \to BY\,\bullet, 1]$
$[B \to B\,\bullet Y, 2]$	$[A \to AX\,\bullet, 0]$	$[B \to B\,\bullet Y, 5]$	$[S \to AB\,\bullet, 0]$
$[B \to B\,\bullet Y, 1]$	$[B \to B\,\bullet Y, 3]$	$[A \to AX\,\bullet, 0]$	$[B \to B\,\bullet Y, 6]$
$[S \to A\,\bullet B, 0]$	$[B \to B\,\bullet Y, 2]$	$[B \to B\,\bullet Y, 4]$	$[A \to AX\,\bullet, 0]$
$[A \to A\,\bullet X, 0]$	$[B \to B\,\bullet Y, 1]$	$[B \to B\,\bullet Y, 3]$	$[B \to B\,\bullet Y, 5]$
$[Y \to \bullet\, a, 4]$	$[S \to A\,\bullet B, 0]$	$[B \to B\,\bullet Y, 2]$	$[B \to B\,\bullet Y, 4]$
$[Y \to \bullet\, b, 4]$	$[A \to A\,\bullet X, 0]$	$[B \to B\,\bullet Y, 1]$	$[B \to B\,\bullet Y, 3]$
$[B \to \bullet\, BY, 4]$	$[Y \to \bullet\, a, 5]$	$[S \to A\,\bullet B, 0]$	$[B \to B\,\bullet Y, 2]$
$[B \to \bullet\, b, 4]$	$[Y \to \bullet\, b, 5]$	$[A \to A\,\bullet X, 0]$	$[B \to B\,\bullet Y, 1]$
$[X \to \bullet\, b, 4]$	$[B \to \bullet\, BY, 5]$	$[Y \to \bullet\, a, 6]$	$[S \to A\,\bullet B, 0]$
	$[B \to \bullet\, b, 5]$	$[Y \to \bullet\, b, 6]$	$[A \to A\,\bullet X, 0]$
	$[X \to \bullet\, b, 5]$	$[B \to \bullet\, BY, 6]$	$[Y \to \bullet\, a, 7]$
		$[B \to \bullet\, b, 6]$	$[Y \to \bullet\, b, 7]$
		$[X \to \bullet\, b, 6]$	$[B \to \bullet\, BY, 7]$
			$[B \to \bullet\, b, 7]$
			$[X \to \bullet\, b, 7]$

Die Items $[B \to B\,\bullet Y, t]$ und $[B \to BY\,\bullet, t]$ befinden sich mindestens $k-1$ mal in den Itemlitsten $I^{\mathbf{G}_2}_{|ab^k|}(ab^k)$ für $t \in [1 : k-1]$.

Grammtik $\mathbf{G}_2$ enthält die zwei linksrekursiven Nonterminals A und B, die aus S in einem Schritt in $S \overset{f_1}{\Rightarrow} AB$ abgeleitet werden. $L(A) = ab^+$ und $L(B) = b(a \cup b)^\star$. $L(B)$ enthält b^+, so daß die Bedingungen von **Satz 3.2** für $p := a$, $m := b$ und $q := \epsilon$ erfüllt sind. $\mathbf{G}_2$ enthält weder ein rechtsrekursives noch ein echtrekursives Nonterminal und kann den **Sätzen 3.1** und **3.4** bis **3.6** nicht entsprechen. Für $p := a$, $m_i := b^i$ und $q := \epsilon$ erfüllt $\mathbf{G}_2$ aber auch die Bedingungen von **Satz 3.3**. Dies gilt für jede CFG die **Satz 3.2** erfüllt mit $\bar{p} := p$, $\bar{m}_i := m^i$ und $\bar{q} := q$.

Ob eine CFG ein linksrekursives Nonterminal enthält, ist für jedes $A \in V$ entscheidbar[2]. Damit die Nonterminals A und B die Voraussetzungen von **Satz 3.2** erfüllen, darf $L(A) \cap L(B)$ nicht leer sein, falls $p = q = \epsilon$ gilt. Ob der Durchschnitt zweier kontextfreier Sprachen leer oder unendlich ist, ist unentscheidbar[3].

Das Kriterium von **Satz 3.2** ist unentscheidbar.
Aus ihm folgt das von **Satz 3.3**. (4.3)

Satz 3.3 ist nun eine Erweiterung von **Satz 3.2**, insofern, als die Form des Wortes pm^kq auch $pm_1...m_i...m_kq$ sein kann, und die m_i sich möglicherweise voneinander unterscheiden.

Beispiel 4.4

Sei $\mathbf{G}_3 := (\{S, A, B, C, D, E, F, U, V, X, Y\}, \{a, b\}, S, P_3)$
mit $P_3 :=$

$$\begin{array}{lllll} \{ & f_1 : S \to UV, & f_2 : U \to UX, & f_3 : V \to VY, & f_4 : U \to a, \\ & f_5 : V \to a, & f_6 : X \to CD, & f_7 : Y \to EF, & f_8 : C \to AB, \\ & f_9 : D \to BA, & f_{10}: E \to BB, & f_{11}: F \to AA, & f_{12}: A \to a, \\ & f_{13}: B \to b\}\,. & & & \end{array}$$

Betrachten wir die von $\mathbf{G}_3$ erzeugte Sprache $L(\mathbf{G}_3)$. Für $i, j \geq 0$ folgt:

$$S \overset{f_1}{\Rightarrow^\star} UV \overset{(f_2)^i}{\Rightarrow^\star} UX^iV \overset{(f_3)^j}{\Rightarrow^\star} UX^iVY^j\,. \tag{4.4}$$

Wir wissen weiter:

$$X \overset{f_6}{\Rightarrow^\star} CD \overset{f_8}{\Rightarrow^\star} ABD \overset{f_9}{\Rightarrow^\star} ABBA \overset{f_{12}}{\Rightarrow^\star} aBBa \overset{f_{13}}{\Rightarrow^\star} abba := u\,, \tag{4.5}$$

[2]durch Konstruktion der Menge $\mathcal{U}_A$ analog den rechtsrekusiven Nonterminals.

[3]siehe [Sal] S. 279

$$Y \overset{f_7}{\Rightarrow^\star} EF \overset{f_{10}}{\Rightarrow^\star} BBF \overset{f_{11}}{\Rightarrow^\star} BBAA \overset{f_{13}}{\Rightarrow^\star} bbAA \overset{f_{12}}{\Rightarrow^\star} bbaa := v\,. \tag{4.6}$$

Mit (4.5) und (4.6) folgt für UX^i und VY^j:

$$UX^i \overset{f_4}{\Rightarrow^\star} aX^i \overset{f_6}{\Rightarrow^\star} a(CD)^i \overset{(4.5)}{\Rightarrow^\star} a(abba)^i \overset{(4.5)}{=} au^i\,, \tag{4.7}$$

$$VY^j \overset{f_5}{\Rightarrow^\star} aY^j \overset{f_7}{\Rightarrow^\star} a(EF)^j \overset{(4.6)}{\Rightarrow^\star} a(bbaa)^j \overset{(4.6)}{=} av^j\,. \tag{4.8}$$

Also folgt für die Sprache $L(\mathbf{G}_3)$:

$$\begin{aligned}&S \overset{(4.4)}{\Rightarrow^\star} UX^iVY^j \overset{(4.7)}{\Rightarrow^\star} a(abba)^iVY^j\\ &\overset{(4.8)}{\Rightarrow^\star} a(abba)^ia(bbaa)^j = au^iav^j \in L(\mathbf{G}_3)\,.\end{aligned}$$

Da die einzigen Ableitungen bezüglich $\mathbf{G}_3$ in (4.4) bis (4.8) bestehen, sind die Wörter der Sprache $L(\mathbf{G}_3)$ von der Form au^iav^j, $i, j \geq 0$. Wir geben die Itemlisten nicht an, sondern verweisen nur darauf, daß Liste $I^{\mathbf{G}_3}_{|au^iav^j|}(au^iav^j)$ $i+j$ Items der Form $[V \to VY\,\bullet, |au^t|]$ enthält für $t \in [1 : i+j]$. Da $\mathbf{G}_3$ weder ein rechtsrekursives, noch ein echt rekursives Nonterminal besitzt, kann sie weder den Bedingungen von **Satz 3.1**, noch denen der **Sätze 3.4** bis **3.6** genügen. Für $p := a, q := \epsilon$ und $m_1 := m_2 := \cdots := m_i := u$ und $m_{i+1} := a$ und $m_{i+2} := m_{i+3} := \cdots := m_k := v$ erfüllt $\mathbf{G}_3$ aber die Bedingungen von **Satz 3.3**.

Andererseits ist:

$$av^j = a(bbaa)^j = (abba)^ja = u^ja$$

und somit folgt:

$$au^iav^j = au^iu^ja\,.$$

Also erfüllen die Nonterminals U und V für $\bar{p} := \epsilon$, $\bar{m} := u$ und $\bar{q} := a$ auch die Bedingungen von **Satz 3.2**.

Wir vermuten, falls eine CFG $\mathbf{G}$ die Bedingungen von **Satz 3.3** erfüllt, so genügt sie auch den Bedingungen von **Satz 3.2** für geeignete $\bar{p}, \bar{q}, \bar{m} \in T^\star$. Wir werden diese Vermutung im folgenden beweisen.

Beweis: Das Iterations- oder Pumping-Lemma[4] besagt nun, daß es für jede kontextfreie Sprache $\mathcal{L}$ eine feste Zahl $r(\mathcal{L})$ gibt, so daß für alle Wörter $w \in \mathcal{L}$ mit Länge $|w| \geq r(\mathcal{L})$ gilt:

1. $(w := w_1 w_2 w_3 w_4 w_5\,, w_i \in T^\star,\ 1 \leq i \leq 5)$
2. $(\forall l \in \mathcal{N})\,(w_1 w_2^l w_3 w_4^l w_5 \in \mathcal{L})$
3. $(w_1 w_2 w_3 \neq \epsilon \vee w_3 w_4 w_5 \neq \epsilon) \wedge (|w_2 w_3 w_4| \leq r(\mathcal{L})\,)\,.$

Sei $\mathcal{L} := L(\mathbf{G})$ und $w := p_1 p m_1 \cdots m_k q q_1 \in L(\mathbf{G})$. Wir wählen k so, daß

$$pm_1 \cdots m_{k-1} m_k q = u_1^s u_2 u_3^s m_k q = pm_1 v_1^t v_2 v_3^t,\ s, t \geq 1 \tag{4.9}$$

ist, mit den Bedingungen

$$u_i, v_i \in T^\star,\ 1 \leq i \leq 3 \wedge |u_1 u_2 u_3| \leq r(\mathcal{L}) \wedge |v_1 v_2 v_3| \leq r(\mathcal{L})\,.$$

Dies ist mit **Satz 3.3** möglich wegen

$$\begin{aligned} &S \overset{(3.14)}{\Rightarrow^\star} \gamma C \delta \overset{\text{Vor. }1.}{\Rightarrow} \gamma A B \delta \Rightarrow^\star w_1 A w_5 \\ &\Rightarrow^\star w_1 p m_1 \cdots m_{k-1} w_5 = w_1 w_2^s w_3 w_4^s w_5 \end{aligned}$$

und wegen

$$\begin{aligned} &S \overset{(3.14)}{\Rightarrow^\star} \gamma C \delta \overset{\text{Vor. }1.}{\Rightarrow} \gamma A B \delta \Rightarrow^\star \bar{w}_1 B \bar{w}_5 \\ &\Rightarrow^\star \bar{w}_1 m_2 \cdots m_k q \bar{w}_5 = \bar{w}_1 \bar{w}_2^t \bar{w}_3 \bar{w}_4^t \bar{w}_5\,. \end{aligned}$$

Also ist

$$w_2^s w_3 w_4^s = u_1^s u_2 u_3^s \overset{(4.9)}{=} pm_1 \cdots m_{k-1}\,, \tag{4.10}$$

$$\bar{w}_2^t \bar{w}_3 \bar{w}_4^t = v_1^t v_2 v_3^t \overset{(4.9)}{=} m_2 \cdots m_k q\,. \tag{4.11}$$

[4] siehe [Har] S. 190, oder [Ber] S. 33

Sowohl A als auch B sind linksrekursiv, also ist

$$u_1 = \epsilon \quad \wedge \quad u_3 \neq \epsilon \quad \wedge \quad v_1 = \epsilon \quad \wedge \quad v_3 \neq \epsilon \,,$$

und somit

$$u_2 u_3^s m_k q \overset{(4.10)}{=} p m_1 \cdots m_k q \overset{(4.11)}{=} p m_1 v_2 v_3^t \,. \tag{4.12}$$

Sei oBdA **G** eine reduzierte CFG in CNF und sei

$$|v_3| \geq |m_k q| \quad \text{also} \quad v_3 = z m_k q \,, z \in T^\star \,, \tag{4.13}$$

$$|u_3| \geq |m_{k-1}| \quad \text{also} \quad u_3 = z' m_{k-1} \,, z' \in T^\star \,. \tag{4.14}$$

Anderenfalls wäre $zv_3 = m_k q$ und $z'u_3 = m_{k-1}$, und der Beweis verliefe analog. Wegen (4.12) und (4.13) endet m_{k-1} auf z, also folgt:

$$u_3 \overset{(4.14)}{=} z' m_{k-1} = z' x z,\ x \in T^\star \,. \tag{4.15}$$

Betrachten wir das Teilwort $p m_1 \cdots m_k q$:

$$\begin{aligned}
p m_1 \cdots m_k q &\overset{(4.12)}{=} u_2 u_3^s m_k q \\
&\overset{(4.15)}{=} u_2 (z' x z)^s m_k q \\
&= u_2 z' x (z z' x)^{s-1} z m_k q \\
&\overset{(4.13)}{=} u_2 z' x (z z' x)^{s-1} v_3 \\
&\overset{(4.12)}{=} p m_1 v_2 (z m_k q)^{t-1} v_3 \,.
\end{aligned}$$

Also ist:

$$z' x = m_k q \,, \tag{4.16}$$

$$u_3 = m_k q z \,, \tag{4.17}$$

$$v_3 = z z' x \,, \tag{4.18}$$

und somit

$$u_3^s \overset{(4.15)}{=} (z' x z)^s = z' x (z z' x)^{s-1} z \overset{(4.18)}{=} z' x v_3^{s-1} z \overset{(4.16)}{=} m_k q v_3^{s-1} z \,. \tag{4.19}$$

In allen Listen $I^{\mathbf{G}}_{|p_1 u_2 u_3^s|}$ befinden sich alle Items $[B \to \bullet\nu, |p_1 u_2 u_3^s|]$, falls die Produktion $f : B \to \nu \in P$ ist. Denn es gilt für alle $s \in \mathcal{N}$:

$$S \overset{(3.14)}{\Rightarrow^\star} p_1 C q_1 \overset{\text{Vor. }1.}{\Rightarrow} p_1 A B q_1 \overset{(4.12)}{\Rightarrow^\star} p_1 u_2 u_3^s B q_1\,.$$

Somit beginnt hinter dem Teilwort $p_1 u_2 u_3^s$ die Abarbeitung des von B erzeugten Anschlußwortes $v_2 v_3^t$. Da B als linksrekursives Nonterminal die Satzform $B \Rightarrow^\star v_2 v_3^t$ erzeugt und $\mathbf{G}$ in CNF vorliegt, folgt:

$$B \Rightarrow^\star B v_3^t\,. \tag{4.20}$$

Mit den Voraussetzungen von **Satz 3.3** wissen wir:

$$B \Rightarrow^\star m_k q\,. \tag{4.21}$$

Betrachten wir nun das Wort $\quad p_1 u_2 u_3^{s+l} m_k q q_1 \in L(\mathbf{G}),\ l \geq 1.$

$$\begin{aligned} & p_1 u_2 u_3^{s+l} m_k q q_1 \\ & \quad = \quad p_1 u_2 u_3^s u_3^l m_k q \\ & \quad \overset{(4.19)}{=} \quad p_1 u_2 u_3^s m_k q v_3^{l-1} z m_k q \\ & \quad \overset{(4.13)}{=} \quad p_1 u_2 u_3^s m_k q v_3^l\,. \end{aligned}$$

Wir wissen:

$$\begin{aligned} B & \overset{(4.20)}{\Rightarrow^\star} B v_3^l \overset{(4.21)}{\Rightarrow^\star} m_k q v_3^l \\ & \overset{(4.13)}{=} \quad m_k q (z m_k q)^l \\ & = \quad m_k q z (m_k q z)^{l-1} m_k q \\ & = \quad (m_k q z)^l m_k q \\ & \overset{(4.17)}{=} \quad u_3^l m_k q\,. \end{aligned}$$

Für $\bar{p} := p_1 u_2 u_3^s$, $\bar{m} := u_3$, $\bar{q} := m_k q$ erfüllt $\mathbf{G}$ die Bedingungen von **Satz 3.2**, denn es ist $A \Rightarrow^\star p_1 u_2 u_3^s \bar{m}^l = p_1 p m_1 \cdots m_{k-1} \bar{m}^l$ und $B \Rightarrow^\star m_k q v_3^t = u_3^t m_k q = \bar{m}^t m_k q$ und die Behauptung ist bewiesen. $\diamond$

Satz 3.2 und **3.3** sind also zwei Sätze, die den selben Sachverhalt kennzeichnen und somit, falls sie für eine CFG **G** gelten, immer beide zutreffen. Sie charakterisieren die Sprache $L(\mathbf{G})$ von verschiedenen Seiten und ergeben unterschiedliche Kriterien, um zu zeigen, daß $\mathbf{G} \in \mathbf{NIB}$ liegt. Bezüglich der Entscheidbarkeit gilt das gleiche Argument wie in (4.3).

Das Kriterium von **Satz 3.3** ist unentscheidbar.
Aus ihm folgt das von **Satz 3.2**. (4.22)

4.4 Echte Rekursion

Satz 3.4 besagt, daß eine CFG **G** nichtzustandsbeschränkt ist, falls sie ein echt rekursives Nonterminal C enthält, welches $C \Rightarrow^\star \nu C \mu$ erzeugt, und es ein Wort $m \in T^+$ gibt, welches im Durchschnitt von allen drei Sprachen $L(\nu) \cap L(C) \cap L(\mu)$ liegt.

Beispiel 4.5 Sei $\mathbf{G}_{4,0} := (\{S, A, B, C, X\}, \{a\}, S, P_{4,0})$ mit $P_{4,0} :=$

$$\{ \quad f_1\colon S \to AX, \quad f_2\colon X \to CB, \quad f_3\colon C \to AX, \quad f_4\colon A \to a, \quad f_5\colon B \to a, \quad f_6\colon C \to a \}.$$

Es ist $L(\mathbf{G}_{4,0}) = a^{2k+1}$, $k \geq 1$,. $\mathbf{G}_{4,0}$ enthält das echt rekursive Nonterminal C, welches in die Satzformen $C \Rightarrow^\star A^i C B^i = \mathcal{P}$ und $C \Rightarrow^\star A^{i+1} X B^i = \mathcal{P}'$ und das Wort a abgeleitet werden kann, sowie das echt rekursive Nonterminal X, welches in die Satzformen $X \Rightarrow^\star A^j X B^j = \mathcal{Q}$ und die Satzformen $X \Rightarrow^\star A^j C B^{j+1} = \mathcal{Q}'$ abgeleitet werden kann. Die Ableitungen für $\mathcal{P}'$ und $\mathcal{Q}'$ ergeben nacheinander ausgeführt die Satzform $A^{i+j+1} C B^{i+j+1} = \mathcal{R}$, welche aus dem Nonterminal C erzeugt werden kann. Aus X läßt sich nun aber $X \Rightarrow^\star A^{i+j+1} X B^{i+j+1} = \mathcal{R}'$ ableiten, indem die Ableitungen für $\mathcal{Q}'$ und $\mathcal{P}'$ hintereinander ausgeführt werden. Jede Satzform $\mathcal{R}$ läßt sich für $i, j \in \mathcal{N}$

also $(k-1)$-mal darstellen als

$$C \Rightarrow^\star A^{i+1}XB^i \Rightarrow^\star A^{i+j+1}CB^{i+j+1}\,, i+j+1 := k\,.$$

Das gleiche gilt für die Satzform $\mathcal{R}'$ und das Nonterminal X. Da $\{m\} \in L(A) \cap L(B) \cap L(C)$ ist, erzeugt A^{i+j+1} für $i+j = 2k$ das gleiche Wort m^{2k+1} wie A^kCB^k. Mit wachsendem k wächst also die Anzahl der Satzformen $\mathcal{R}$ und $\mathcal{R}'$, die ein Prefix von m^{2k+1} ableiten. Dies hat das Wachstum der Itemlisten zur Folge. Da die Nonterminals $A, B \in V_{4,0}$ beide nur das Terminalzeichen a erzeugen und $\mathbf{G}_{4,0}$ kein rechts- und kein linksrekursives Nonterminal enthält, erfüllt $\mathbf{G}_{4,0}$ alleine die Voraussetzungen von **Satz 3.4**. Betrachten wir die Itemlisten, die das **EPV** für das Wort $w_k := a^{2k+1}$, $k := 3$ aufbaut.

Liste $I_0(a^7)$	Liste $I_1(a^7)$	Liste $I_2(a^7)$	Liste $I_3(a^7)$
$[S \to \bullet AX, 0]$	$[A \to a\bullet, 0]$	$[C \to a\bullet, 1]$	$[B \to a\bullet, 2]$
$[A \to \bullet a, 0]$	$[S \to A\bullet X, 0]$	$[A \to a\bullet, 1]$	$[C \to a\bullet, 2]$
	$[X \to \bullet CB, 1]$	$[X \to C\bullet B, 1]$	$[A \to a\bullet, 2]$
	$[C \to \bullet AX, 1]$	$[C \to A\bullet X, 1]$	$[X \to CB\bullet, 1]$
	$[C \to \bullet a, 1]$	$[B \to \bullet a, 2]$	$[X \to C\bullet B, 2]$
	$[A \to \bullet a, 1]$	$[X \to \bullet CB, 2]$	$[C \to A\bullet X, 2]$
		$[C \to \bullet AX, 2]$	$[S \to AX\bullet, 0]$
		$[C \to \bullet a, 2]$	$[B \to \bullet a, 3]$
		$[A \to \bullet a, 2]$	$[X \to \bullet CB, 3]$
			$[C \to \bullet AX, 3]$
			$[C \to \bullet a, 3]$
			$[A \to \bullet a, 3]$

Liste $I_4(a^7)$	Liste $I_5(a^7)$	Liste $I_6(a^7)$	Liste $I_7(a^7)$
$[B \to a\bullet, 3]$	$[B \to a\bullet, 4]$	$[B \to a\bullet, 5]$	$[B \to a\bullet, 6]$
$[C \to a\bullet, 3]$	$[C \to a\bullet, 4]$	$[C \to a\bullet, 5]$	$[C \to a\bullet, 6]$
$[A \to a\bullet, 3]$	$[A \to a\bullet, 4]$	$[A \to a\bullet, 5]$	$[A \to a\bullet, 6]$
$[X \to CB\bullet, 2]$	$[X \to CB\bullet, 3]$	$[X \to CB\bullet, 4]$	$[X \to CB\bullet, 5]$
$[X \to C\bullet B, 3]$	$[X \to CB\bullet, 1]$	$[X \to CB\bullet, 2]$	$[X \to CB\bullet, 3]$
$[C \to A\bullet X, 3]$	$[X \to C\bullet B, 4]$	$[X \to C\bullet B, 5]$	$[X \to CB\bullet, 1]$
$[C \to AX\bullet, 1]$	$[C \to A\bullet X, 4]$	$[C \to A\bullet X, 5]$	$[X \to C\bullet B, 6]$
$[X \to C\bullet B, 1]$	$[C \to AX\bullet, 2]$	$[C \to AX\bullet, 3]$	$[C \to A\bullet X, 6]$
$[B \to \bullet a, 4]$	$[S \to AX\bullet, 0]$	$[C \to AX\bullet, 1]$	$[C \to AX\bullet, 4]$
$[X \to \bullet CB, 4]$	$[X \to C\bullet B, 2]$	$[X \to C\bullet B, 3]$	$[C \to AX\bullet, 2]$
$[C \to \bullet AX, 4]$	$[B \to \bullet a, 5]$	$[X \to C\bullet B, 1]$	$[S \to AX\bullet, 0]$
$[C \to \bullet a, 4]$	$[X \to \bullet CB, 5]$	$[B \to \bullet a, 6]$	$[X \to C\bullet B, 4]$
$[A \to \bullet a, 4]$	$[C \to \bullet AX, 5]$	$[X \to \bullet CB, 6]$	$[X \to C\bullet B, 2]$
	$[C \to \bullet a, 5]$	$[C \to \bullet AX, 6]$	$[B \to \bullet a, 7]$
	$[A \to \bullet a, 5]$	$[C \to \bullet a, 6]$	$[X \to \bullet CB, 7]$
		$[A \to \bullet a, 6]$	$[C \to \bullet AX, 7]$
			$[C \to \bullet a, 7]$
			$[A \to \bullet a, 7]$

Da sich das Item $[S \to AX\bullet, 0]$ in Liste $I_7^{\mathbf{G}_{4,0}}(a^7)$ befindet, ist das Wort a^7 in der Sprache $L(\mathbf{G}_{4,0})$ enthalten. Das Item $[X \to CB\bullet, 2t-1]$ kommt k mal in Liste $I_{|a^{2k+1}|}^{\mathbf{G}_{4,0}}(a^{2k+1})$ für $t \in [1:k]$ vor . Ob der Durchschnitt der drei Sprachen $L(\alpha) \cap L(C) \cap L(\beta)$ leer ist, ist für kontextfreie Sprachen unentscheidbar[5].

Das Kriterium von **Satz 3.4** ist unentscheidbar und von den anderen Kriterien unabhängig. (4.23)

Wir möchten hier bemerken, daß tatsächlich der Durchschnitt aller drei Sprachen $L(\alpha) \cap L(C) \cap L(\beta)$ das Wort m enthalten muß, damit **Satz 3.4** gilt. Dies zeigt das folgende Beispiel:

[5]siehe Aussage (4.3)

Beispiel 4.6 Seien die Grammatiken $\mathbf{G}_{4,i}$ für $i \in [1:3]$ $:=(\{S, A, B, C, X\}, \{a, b\}, S, P_{4,i})$ mit

$$P_{4,1} := \begin{array}{lll} \{ \; f_1\colon S \to AX, & f_2\colon X \to CB, & \\ f_3\colon C \to AX, & f_4\colon A \to a, & \\ f_5\colon B \to b, & f_6\colon C \to a & \}, \end{array}$$

$$P_{4,2} := \begin{array}{lll} \{ \; f_1\colon S \to AX, & f_2\colon X \to CB, & \\ f_3\colon C \to AX, & f_4\colon A \to a, & \\ f_5\colon B \to b, & f_6\colon C \to b & \}, \end{array}$$

$$P_{4,3} := \begin{array}{lll} \{ \; f_1\colon S \to AX, & f_2\colon X \to CB, & \\ f_3\colon C \to AX, & f_4\colon A \to a, & \\ f_5\colon B \to a, & f_6\colon C \to b & \}. \end{array}$$

Die vier Sprachen $L(\mathbf{G}_{4,1}) = a^j ab^j$, $L(\mathbf{G}_{4,2}) = a^j bb^j$, $L(\mathbf{G}_{4,3}) = a^j ba^j$ und $L(\mathbf{G}_{4,0}) = a^j aa^j$ werden pro Wort mit genau einem Ableitungsbaum erzeugt. Die Ableitungsbäume weisen für alle vier Grammatiken die gleiche Struktur auf und unterscheiden sich nur in der Markierung der Blätter. Dieser Unterschied genügt, um im vierten Fall ein Wachstum der Listen zu erzeugen. Die drei Sprachen $L(\mathbf{G}_{4,1}) = a^j ab^j$, $L(\mathbf{G}_{4,2}) = a^j bb^j$ und $L(\mathbf{G}_{4,3}) = a^j ba^j$ sind jedoch unter dem **EPV** mit jeweils 9,6 und 6 Items zustandsbeschränkt[6] und somit sind alle drei $\mathbf{G}_{4,i}, \notin$ **NIB** für $i \in [1:3]$.

4.5 Echte und Linksrekursion

Anders verhält es sich, falls das Nonterminal C sowohl echt rekursiv als auch linksrekursiv ist und unendlich oft das Wort m ableitet. Dann genügt es, daß

[6] Beweis durch Induktion über j.

entweder mit der Satzform ν oder μ ein Wort $m \in T^\star$ gemeinsam abgeleitet wird. Dies wurde in **Satz 3.5** und **3.6** bewiesen. Betrachten wir dazu zwei Beispielgrammatiken $\mathbf{G}_5$ und $\mathbf{G}_6$.

Beispiel 4.7 Seien $\mathbf{G}_5 := (V, T, S, P_5)$ und $\mathbf{G}_6 := (V, T, S, P_6)$ wobei $V := \{S, A, B, C, X, Y\}$, $T := \{a, b\}$ und

$P_5 :=$		$P_6 :=$	
$\{\ f_1\colon S \to AX,$	$f_2 : C \to AX,$	$\{\ f_1\colon S \to AX,$	$f_2 : C \to AX,$
$f_3\colon X \to CB,$	$f_4 : C \to DY,$	$f_3\colon X \to CB,$	$f_4 : C \to DY,$
$f_5\colon D \to DY,$	$f_6 : D \to a,$	$f_5\colon D \to DY,$	$f_6 : D \to b,$
$f_7\colon A \to a,$	$f_8 : B \to b,$	$f_7\colon A \to a,$	$f_8 : B \to b,$
$f_9\colon C \to a,$	$f_{10}\colon Y \to a\ \}$	$f_9\colon C \to b,$	$f_{10}\colon Y \to b\,,\}$

Es ist $L(\mathbf{G}_5) = a^i a^+ b^i$ und $L(\mathbf{G}_6) = a^i b^+ b^i$. Ihre Grammatiken erfüllen alleine[7] die Bedingungen von **Satz 3.5** bzw. **3.6**. Deshalb gilt[8]:

Die Kriterien von **Satz 3.5** und **3.6** sind unentscheidbar und von den anderen Kriterien unabhängig. (4.24)

4.6 Schlussbetrachtungen

Wir haben in der vorliegenden Arbeit das **EPV** unter dem Gesichtspunkt des Laufzeitverhaltens betrachtet und hierbei Erkentnisse über kontextfreie Grammatiken und ihre Sprachen gewonnen. Zwar konnte die Menge der zustandsbeschränkten Grammatiken nicht charakterisiert werden. Es wurden dennoch hinreichende Kriterien für nichtzustandsbeschränkte Grammatiken

[7]Man prüft leicht nach, daß $\mathbf{G}_5$ bzw. $\mathbf{G}_6$ die Satzformen A^iCB^i und aus C die Satzformen DY^j erzeugen, wobei DY^j in a^j bzw. b^j ableitbar ist.

[8]In Bezug auf die Entscheibarkeit gilt das selbe Argument wie in Aussage (4.3).

bestimmt. Für alle Sätze außer **Satz 3.2** und **Satz 3.3** konnten wir eine Grammatik angeben, die alleine den Voraussetzungen des entsprechenden Satzes genügte. Also sind alle Eigenschaften bis auf die von **Satz 3.2** und **Satz 3.3** voneinander unabhängig. Aus **Satz 3.2** folgt **Satz 3.3** und umgekehrt.

In praktischer Hinsicht ergibt sich ein ungünstiges Bild: Außer bei der Rechtsrekursion von **Satz 3.1** sind die Voraussetzungen alle unentscheidbar.

4.7 Offene Probleme

Wir möchten diese Arbeit beschließen, indem wir einige noch ungelöste Probleme zusammentragen.

Offen bleibt die Frage der vollständigen Klassifizierung der Menge **NIB**. Allen Sätzen gemeinsam ist die Bedingung der Existenz eines unendlich oft iterierbaren Teilwortes $m \in T^+$ in der Sprache $L(\mathbf{G})$, wobei die Art des Erzeugungsprozesses das Listenwachstum verursacht. Wir möchten an dieser Stelle die Vermutung äußern, daß die **Sätze 3.1** bis **3.6** die Menge **NIB** vollständig charakterisieren. Unsere Vermutung stützt sich auf praktische Erfahrungen im Versuch, ein unendlich oft iterierbares Teilwort $m \in T^+$ mit einer CFG so zu erzeugen, daß der Erzeugungsprozess das Listenwachstum nach sich zieht, ohne Kriterien der **Sätze 3.1** bis **3.6** zu erfüllen.

Da sich die **Sätze 3.1** bis **3.6** auf Grammatiken beziehen, bleibt ebenfalls offen, welche Sprachen überhaupt eine zustandsbeschränkte Grammatik besitzen.

Mehrdeutigkeit, also die Anzahl der Ableitungsbäume, ist offenbar kein Kriterium für das Wachstum der Listen. Dies macht **Beispiel 4.5** und das folgende **Beispiel 4.7** für die Sprache $L(\mathbf{G}_7) = \{a^i b^j c^k \mid i = j \vee j = k\}$. deutlich[9].

Beispiel 4.8 Sei $\mathbf{G}_7 := (V, T, S, P_7)$ mit

$V := \{A, B, C, P, Q, S, U, W, X, Y\}$, $T := \{a, b, c\}$, und

$P_7 :=$

$$
\begin{array}{lllll}
\{ & f_1 : S \to UW, & f_2 : S \to PQ, & f_3 : U \to AX, & f_4 : X \to UB, \\
 & f_5 : U \to AB, & f_6 : W \to WC, & f_7 : P \to PA, & f_8 : Q \to BY, \\
 & f_9 : Y \to QC, & f_{10}: Q \to BC, & f_{11}: P \to a, & f_{12}: A \to a, \\
 & f_{13}: B \to b, & f_{14}: C \to c, & f_{15}: W \to c\,\}\,. &
\end{array}
$$

Die Sprache $L(\mathbf{G}_7)$ ist bekanntlich[10] inhärent unendlich mehrdeutig, unter dem **EPV** jedoch mit 13 Items zustandsbeschränkt[11].

Da das **EPV** als Parsing-Methode korrekt arbeitet, muß es in der Lage sein, für verschiedene Wörter $w_1, w_2 \in L(\mathbf{G})$ mindestens zwei entsprechende Ableitungsbäume zu unterscheiden. Es muß desweiteren für alle Wörter $w_1 := wv$, $w_2 := wv' \in L(\mathbf{G})$ die Listen $I_0(wv)$ bis $I_{|w|}(wv)$ einerseits gleich aufbauen, andererseits so, daß mit ihrer Hilfe die weiteren Listen $I_{|w|+1}(wv)$ bis $I_{|wv|}(wv)$ und die Listen $I_{|w|+1}(wv')$ bis $I_{|wv'|}(wv')$ erzeugt werden können. Somit müssen alle Ableitungsbäume für wv bzw. wv' vom **EPV** in die Listen $I_0(wv)$ bis $I_{|w|}(wv)$ und die weiteren Listen $I_{|w|+1}(wv)$ bis $I_{|wv|}(wv)$ bzw. $I_{|w|+1}(wv')$ bis $I_{|wv'|}(wv')$ zerlegt werden. Einen möglichen Beschreibungsansatz dieses Vorgangs bietet die Theorie der Erzeugendenfunktionen, die ein Abzählen der vom **EPV** zerlegten Ableitungsbäume unter Verwendung geeigneter Markierungsfunktionen und eventuell benötigter Äquivalenzrelationen ermöglicht.

---

[9] $L(UW) = a^i b^i c^j$ und $L(PQ) = a^i b^j c^j$; $L(\mathbf{G}_7) = L(UW) \cup L(PQ)$.

[10] siehe [Har] S. 241

[11] Beweis durch Induktion über $|w| \in L(\mathbf{G}_7)$

Daneben eröffnet sich ein Ansatz durch die von J. Berstel[12] aufgestellte Theorie der iterativen Paare. Alle **Sätze 3.1** bis **3.6** bedingen die Existenz eines Teilwortes pm^kq für geeignete $p, q \in T^\star$ und $m \in T^+$. Die Ausarbeitung der Iterationstheoreme (Pumping-Lemma, Ogdens-Lemma) bietet eine Klassifizierung der kontextfreien Grammatiken und ihrer Sprachen anhand der enthaltenen iterierten Teilwörter. Es wäre zu klären, unter welchen Bedingungen iterative Paare ein Wachstum der Itemlisten zur Folge haben.

[12] in [Ber] Kapitel VIII

Literaturverzeichnis

SudThomas A. Sudkamp, **Languages and Mashines**, Addison-Wesley NY 1988

[0] [1] S.Sippu, E.Soisalon-Soininen **Parsing Theory** Vol. I Springer Berlin 1988

[2] Arto K. Salomaa **Formale Sprachen** Springer Berlin 1988

[3] Gyorgy E. Révész **Introduction to Formal Languages**

[4] Michael A. Harrisson **Introduction to Formal Languages Theory**

[5] J. E. Hopcraft, J. D. Ullman**Introduction to Automata Theory, Languages and Computation** Addison-Wesley Menlo Park California, 1979

[6] M.A. Arbib, A.J. Kfoury Introduction to Formal Language Theory Springer NY 1988

[7] A. V. Aho, J. D. Ullman The Theory of Parsing, Translation and Compiling Vol 1: Parsing; Prentice-Hall London 1972

[8] Comunication of the ACM 1970 Vol 13:2 S 94 - 102