Diplomarbeit

Grad der Mehrdeutigkeit kontextfreier Grammatiken und Sprachen

vorgelegt von

Mohamed NAJI

November 1998

Fachbereich Informatik

Professur für Programmiersprachen und Compiler

Betreuer: Prof. Dr. Detlef Wotschke

Erklärung

Hiermit erkläre ich, die vorliegende Diplomarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben.

Frankfurt am Main, im November 1998

Mohamed NAJI

Danksagung und Widmung

An dieser Stelle möchte ich mich bei allen Freunden, die mich bei der Erstellung dieser Arbeit auf verschiedene Art und Weise unterstützt haben, bedanken. Besonderer Dank gebührt vor allem Herrn Prof. Dr. Detlef Wotschke; Herrn Prof. Dr. Rainer Kemp; Dipl. Inf. Christian Herzog und Dipl. Inf. Klaus Wich für die vielen wertvollen Tipps und Hinweise.

Diese Diplomarbeit widme ich meiner Familie, insbesondere meinem verstorbenen Vater.

Frankfurt am Main, im November 1998

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Zusammenfassung

In dieser Arbeit definieren wir ein Maß für den Grad der Mehrdeutigkeit (degree of ambiguity da) kontextfreier Grammatiken und Sprachen als die Anzahl der Ableitungsbäume in Abhängigkeit von der Länge n eines Wortes. Wir zeigen, dass es weder Sprachen noch zyklenfreie Grammatiken gibt, deren Mehrdeutigkeitsgrad stärker als $2^{\Theta(n)}$ wächst (wie z B. $\Theta(n^n)$). Aus [10] ist es außerdem bekannt, dass es keine Grammatiken (und somit keine Sprachen) gibt, deren Mehrdeutigkeit stärker als polynomiell, aber schwächer als exponentiell wächst (wie z. B. $\Theta(2^{\sqrt{n}})$. Deshalb untersuchen wir in dieser Arbeit hauptsächlich konstant mehrdeutige, polynomiell mehrdeutige und exponentiell mehrdeutige Grammatiken und Sprachen.

Für jede feste, ganze Zahl $k \in \mathbb{N}$ hat Maurer [8] die Existenz einer k–deutigen kontextfreien Sprache nachgewiesen. Durch Verwendung einer einfacheren Sprache, nämlich der Sprache $L_k := \{a^m b_1^{m_1} b_2^{m_2} \ldots b_k^{m_k} | m, m_1, m_2, \ldots, m_k \geq 1, \exists\ i\ mit\ m = m_i\}$, und mit Hilfe von Ogden's Lemma[1] erhalten wir einen wesentlich kürzeren Beweis.

Ferner zeigen wir die Existenz exponentiell mehrdeutiger Sprachen. Wir zeigen, dass die Sprache L^* – wobei $L = \{a^i b^i c^j | i, j \geq 1\} \cup \{a^i b^j c^i | i, j \geq 1\}$–

[1] Als Maurer den Beweis ausführte, existierte Ogden's Lemma noch nicht.

exponentiell mehrdeutig ist, indem wir beweisen, dass das Wort $(a^{h+h!}b^{h+h!}c^{h+h!})^k$ mindestens 2^k Ableitungen in jeder Grammatik G für L^* hat, wobei k aus $\mathbb{N}$ ist und h die Konstante aus OGDEN's Lemma für G ist. Für beliebig kleines c aus $\mathbb{R}_+$ entwerfen wir eine Grammatik G_c für L^*, so dass $da_{G_c} \leq 2^{cn}$ gilt. Somit gilt, dass die Sprache L^* zwar exponentiell mehrdeutig ist, aber es gibt kein festes c aus $\mathbb{R}_+$, so dass L^* 2^{cn}–deutig ist.

Wir geben polynomiell mehrdeutige Grammatiken an und zeigen die Existenz von polynomiell mehrdeutigen Sprachen, indem wir mit Hilfe von OGDEN's Lemma beweisen, dass die Anzahl der Ableitungsbäume eines Wortes der Länge n in jeder Grammatik für die Sprache L^k in der Größenordnung von $\Omega(n^k)$ liegt, wobei k eine Konstante aus $\mathbb{N}$ ist, und $L := \{a^m b^{m_1} c b^{m_2} c \dots b^{m_p} c | p \in \mathbb{N}; m, m_1, m_2, \dots, m_p \in \mathbb{N}; \exists i \in \{1, 2, \dots, p\}\ mit\ m = m_i\}$ gilt. Durch Angabe einer $O(n^k)$–deutigen Grammatik zeigen wir schließlich, dass L^k polynomiell vom Grad k mehrdeutig ist. Außerdem entwerfen wir für jedes feste d aus $\mathbb{R}_+$ eine Grammatik G_d für L, so dass $da_{G_d} \leq dn$ dn für genügend großes n ist.

Kapitel 1

Einleitung

In dieser Arbeit beschäftigen wir uns mit dem Grad der Mehrdeutigkeit kontextfreier Grammatiken und Sprachen. Wir meinen mit Grammatiken bzw. Sprachen – falls nicht anders gesagt – die kontextfreien Grammatiken bzw. Sprachen.

Eine kontextfreie Grammatik CFG besteht aus einer Menge von Variablen, einer Menge von Terminalsymbolen und einer Menge von Produktionen. Eine Produktion ist ein Paar, das aus einer Variablen und einer Zeichenkette besteht, wobei die Zeichenkette Variablen und Terminale enthält, die aber auch leer sein kann. Die Variable wird als die linke, die Zeichenkette als die rechte Seite dieser Produktion bezeichnet. Solch eine Produktion erlaubt in beliebigen Zeichenketten, in denen die Variable ihrer linken Seite vorkommt, die Ersetzung durch die entsprechende rechte Seite ohne Berücksichtigung des Kontext, d. h. die Umgebung dieser Variablen in der gegebenen Zeichenkette, daher die Bezeichnung „kontextfrei". Man sagt dann, dass aus der gegebenen Zeichenkette durch diese Ersetzung die entsprechende neue Zeichenkette abgeleitet werden kann. Ausgehend von einer festgelegten Startvariablen wendet man solche Ableitungsschritte solange an, bis in dem abgeleiteten Wort nur noch Terminalsymbole vorkommen. Die Menge der so ableitbaren Wörter aus Terminalsymbolen ist die von der Grammatik G erzeugte Sprache L(G). Die Menge der von CFG's erzeugten Sprachen heißt die Menge der kontextfreien Sprachen CFL.

Eine Ableitung ist eine Folge von Ableitungsschritten. In jedem Ableitungsschritt sind immer zwei Entscheidungen zu treffen. Zunächst ist das Nichtterminal zu bestimmen, das ersetzt werden soll. Danach ist eine der Alternativen auszuwählen, durch die dieses Nichtterminal ersetzt werden soll. Unter einer Linksableitung verstehen wir eine Ableitung, in der in jedem Ableitungsschritt die am weitesten links stehende Variable abgeleitet wird. Wir kommen nun zu graphischen Darstellungen von Ableitungen. Da bei kontextfreien Grammatiken die linken Seiten aller Regeln genau eine Variable enthalten, lassen sich kontextfreie Ableitungen gut durch sogenannte Ableitungsbäume ohne Berücksichtigung der Ableitungsreihenfolge darstellen. An der Wurzel steht das Startsymbol. Jeder innere Knoten ist mit einer Variable markiert, während an den Blättern Terminalzeichen oder das leere Wort stehen. Wenn ein innerer Knoten mit A markiert ist und seine Söhne von links nach rechts mit den Symbolen X_1, X_2,. . . , X_k markiert sind, dann muss $A \longrightarrow X_1X_2 \dots X_k$ eine Produktion der Grammatik sein. Wegen der Kontextfreiheit der Grammatik ist die Reihenfolge, in der wir Variable ableiten, unerheblich. Deshalb gehört zu jeder Ableitung genau ein Ableitungsbaum, während zu einem Ableitungsbaum verschiedene Ableitungen des gleichen Wortes gehören können. Aus jedem Ableitungsbaum ist genau eine Linksableitung erhältlich. Es ist sinnvoll, Ableitungen als Bäume darzustellen, denn sie geben den Wörtern einer Sprache eine Struktur, die für Anwendungen, wie Compilierung von Programmiersprachen, sehr nützlich ist.

Eine Grammatik heißt eindeutig, falls jedes Wort höchstens einen Ableitungsbaum in dieser Grammatik hat. Eine Grammatik, in der es ein Wort mit mehr als einem Ableitungsbaum gibt, heißt mehrdeutig. Eine Sprache, für die jede Grammatik mehrdeutig ist, heißt mehrdeutige[1] Sprache. Eine Sprache heißt eindeutig, wenn sie nicht mehrdeutig ist. Die Mehrdeutigkeit von Grammatiken spielt in der Praxis eine Wichtige Rolle. So wird beispielsweise die Semantik einer Programmiersprache dem Ableitungsbaum entnommen. Wäre dieser Baum und damit die Grammatik nicht eindeutig, so könnte ein Programm mehrere Bedeutungen haben. Außerdem haben wir für eindeutige Grammatiken die Möglichkeit, für ein Wort den eindeutigen Ableitungsbaum (z. B. durch das bottom–up oder top–down Verfahren) zu erzeugen. Bei der richtigen Konstruktion gibt es keine Wahlmöglichkeiten, während es bei mehrdeutigen Grammatiken verschiedene richtige Entscheidungen geben kann. Die Intuition sagt also, dass das Wortproblem für eindeutige Grammatiken vielleicht einfacher zu lösen ist als für mehrdeutige Grammatiken. In der Tat ist der erforderliche Aufwand zur Lösung des Wortproblems für einige bekannte Parsing–Algorithmen (wie z. B. EARLEY's Algorithmus[2]) bezüglich einer kontextfreien Sprache für mehrdeutige Grammatiken bekanntlich größer als für eindeutige Grammatiken. Daraus resultiert der Wunsch, Mehrdeutigkeiten zu vermeiden.

Wir werden die Mehrdeutigkeit einer Grammatik (bzw. Sprache) als

[1] In der Literatur spricht man meistens von "'inh"arent"' mehrdeutigen Sprachen. In dieser Arbeit verzichten wir auf die Benutzung des Wortes inhärent.

[2] Näheres siehe [9, Seite 321]

die Anzahl der verschiedenen Ableitungsbäume in Abhängigkeit der Länge der Wörter definieren. Das so definierte Mehrdeutigkeitsmaß (degree of ambiguity da) ist also eine Funktion da(n) von der Länge n eines Wortes und erlaubt eine feinere Klassifizierung der kontextfreien Grammatiken (bzw. Sprachen) als die oben beschriebene Unterteilung in eindeutige und mehrdeutige Grammatiken (bzw. Sprachen). Wir können kontextfreie Grammatiken (bzw. Sprachen) bezüglich ihrer Mehrdeutigkeitsgrade (Mehrdeutigkeitsfunktion) in konstant mehrdeutige, polynomiell mehrdeutige und exponentiell mehrdeutige Grammatiken (bzw. Sprachen) klassifizieren. Diese feinere Klassifikation der Grammatiken und Sprachen nach der Mehrdeutigkeitsfunktion ist in der Literatur kaum zu finden. Eine Arbeit von WICH [10] hat sich zwar mit einer ähnlichen feineren Klassifikation beschäftigt, aber nur für die Grammatiken und nicht für die Sprachen, was wesentlich schwieriger und weiterführender ist. In einer anderen Arbeit von MAURER [8] wurde für jedes k die Existenz einer k–deutigen Sprache nachgewiesen. Der Beweis war allerdings sehr aufwendig und lang und die angegebene Sprache ziemlich kompliziert. Wir liefern einen kürzeren und eleganteren Beweis und betrachten einfachere k–deutige Sprachen.

Wir werden beweisen, dass es weder Sprachen noch zyklenfreie Grammatiken gibt, deren Mehrdeutigkeitsgrad stärker als $2^{\Theta(n)}$ wächst (wie z. B. $\Theta(n^n)$), und dass die Mehrdeutigkeit einer Grammatik genau dann für einzelne Wörter unendlich ist, wenn die Grammatik (nützliche) Zyklen enthält (siehe Satz 2.6.1 auf die Seite 20 sowie Bemerkung 2.7.1 auf die Seite 22). Aus [10] ist bekannt, dass es keine Grammatiken (und somit keine Sprachen) gibt, deren Mehrdeutigkeit stärker als polynomiell, aber schwächer als exponentiell wächst (wie z. B. $\Theta(2^{\sqrt{n}})$. Deshalb untersuchen wir in dieser Arbeit hauptsächlich konstant mehrdeutige, polynomiell mehrdeutige und exponentiell mehrdeutige Grammatiken und Sprachen.

Die vorliegende Arbeit ist wie folgt geteilt: Nach dieser Einleitung werden wir im Kapitel 2 einige formale Definitionen und (bekannte) grundlegende Sätze über kontextfreie Grammatiken und Sprachen auflisten, sowie Schreibweisen, Abkürzungen und Notationen festlegen, die wir für diese Arbeit brauchen.

Kapitel 3 behandelt die konstant mehrdeutigen Grammatiken und Sprachen. MAURER hat in [8] gezeigt, dass es für jede Konstante $k \in \mathbb{N}$ eine k–deutige Sprache gibt. Diese Tatsache beweisen wir allerdings durch Verwendung der "‘einfacheren"’ Sprache $L_k := \{a^m b_1^{m_1} b_2^{m_2} \ldots b_k^{m_k} | m, m_1, m_2, \ldots, m_k \geq 1, \exists\ i\ mit\ m = m_i\}$ und einer "‘eleganteren"’ Beweisidee (OGDEN's Lemma[3]) als in [8]. Ferner geben wir eine k–deutige Sprache über einem fixen (von k unabhängigen) Alphabet an. Diese Sprache ist $\widetilde{L_k} := \{a^m b^{m_1} c b^{m_2} c \ldots b^{m_k} c | m, m_1, m_2, \ldots, m_k \geq 1, \exists\ i\ mit\ m = m_i\}$.

Im Kapitel 4 untersuchen wir die exponentiell mehrdeutigen Grammatiken und Sprachen. Die Angabe einer exponentiell mehrdeutigen Grammatik erweist sich als leicht. Wesentlich schwieriger ist es zu zeigen, dass es exponentiell mehrdeutige Sprachen gibt. Wir betrachten die Sprache L^* – wobei $L = \{a^i b^i c^j | i, j \geq 1\} \cup \{a^i b^j c^i | i, j \geq 1\}$:

- Wir zeigen, dass die Sprache L^* exponentiell mehrdeutig ist, indem wir beweisen, dass das Wort $(a^{h+h!} b^{h+h!} c^{h+h!})^k$ mindestens 2^k Ableitungen

[3] Als MAURER den Beweis ausführte, existierte **Ogden**'s Lemma noch nicht.

in jeder Grammatik G für L^* hat, wobei k aus $\mathbb{N}$ und h die Konstante aus **Ogden**'s Lemma für G ist.

- Für jedes beliebig kleine c aus $\mathbb{R}_+$ entwerfen wir eine Grammatik G_c für L^*, so dass die Mehrdeutigkeit von G_c nach oben durch 2^{cn} beschränkt ist. Somit gilt, dass die Sprache L^* zwar exponentiell mehrdeutig ist, aber es gibt kein festes c aus $\mathbb{R}_+$, so dass L^* 2^{cn}–deutig ist (siehe Satz 4.3.1 auf die Seite 48).

Wir schließen Kapitel 4 mit dem offenen Problem, ob es eine exponentiell mehrdeutige Sprache gibt, die für irgend ein festes c aus $\mathbb{R}_+$ 2^{cn}–deutig ist.

Im Kapitel 5 beschäftigen wir uns mit polynomiell mehrdeutigen Sprachen und Grammatiken. Zuerst geben wir polynomiell mehrdeutige Grammatiken an. Danach zeigen wir die Existenz einer linear mehrdeutigen Sprache, indem wir beweisen, dass $L := \{a^m b^{m_1} c b^{m_2} c \ldots b^{m_p} c | p \in \mathbb{N}; m, m_1, m_2, \ldots, m_p \in \mathbb{N}; \exists i \in \{1, 2, \ldots, p\} \ mit \ m = m_i\}$ linear mehrdeutig ist. Ferner zeigen wir, dass für jedes $k \in \mathbb{N}$ L^k polynomiell vom Grad k mehrdeutig ist. Dieses Ergebnis scheint neu zu sein, denn wir konnten es nirgendwo in der Literatur finden. Außerdem entwerfen wir für jedes feste d aus $\mathbb{R}_+$ eine Grammatik G_d für L, so dass $da_{G_d}(n) \leq dn$ für genügend großes n ist.

Kapitel 6 enthält Vorschläge und Ideen für weiterführende Untersuchungen des Grades der Mehrdeutigkeit kontextfreier Grammatiken und Sprachen.

Kapitel 2

Grundlagen

In diesem Kapitel werden wir einige grundlegende Definitionen und Sätze über kontextfreie Grammatiken und Sprachen auflisten, die in dieser Arbeit später benutzt werden. Ist ein Begriff nicht explizit definiert, so gilt seine Definition aus [9], denn wir können in dieser Arbeit nicht alles definieren.

2.1 Notationen

Wir benutzen folgende Standardnotationen und Begriffe siehe Tabelle 2.1 auf die Seite 10).

Notation	erklärung
ε	das leere Wort
$\lvert w\rvert$	Lange des Wortes w (es gill $\lvert\varepsilon\rvert$=0)
x^q	$\underbrace{xx\dots x}_{q\ mal}$ = q fache Konkatenation vom Wort x
w^R	die Spiegelung des Wortes w.
$\#_a(w)$	die Anzahl von a's im Wort w.
$\mathbb{N}$	Menge der natürlichen Zahlen (ohne Null).
$\mathbb{N}_0$	Menge der natürlichen Zahlen (mit Null).
$\mathbb{R}_+$	Menge der positiven reellen Zahlen (ohne Null).
$\bar{\cup}$	disjunkte Vereinigung.
$:\Leftrightarrow$	ist definiert als.
$:=$	ist definiert als.
$\lvert P\rvert$	Kardinalität der Menge P.
$A \Longrightarrow B$	aus A folgt B.
q. e. d.	Ende des Beweises.
o. B. d. A.	ohne Beschränkung der Allgemeinheit.
n!	n Fakultät =$1 * 2 * \dots * (n-1) * n$; 0!=1.
$\sum_{i=1}^{n} a_i$	Die Summe $a_1 + a_2 + \dots + a_n$.
$\prod_{i=1}^{n} a_i$	Das Produkt $a_1 * a_2 * \dots * a_n$.
$\binom{n}{k}$	$\frac{n!}{k!(n-k)!}$ der Binominalkoeffizient n über k
$\lfloor x\rfloor$	untere Gaußklammer.
$\lceil x\rceil$	obere Gaußklammer.

Tabelle 2.1: Notationen

- Für eine endliche oder unendliche Menge L definieren wir:

  * $L^n := \{w_1w_2\dots w_n \mid w_i \in L, \forall i \in \{1,2,\dots,n\}\}$ für $n \in \mathbb{N}$

  * $L^0 := \{\varepsilon\}$
  * $L^* := \bigcup_{n=0}^{\infty} L^n$: der KLEENEsche Abschluß von L.
  * $L^+ := \bigcup_{n=1}^{\infty} L^n$
  * $L^{\neq k} := \bigcup_{\substack{n=0 \\ n\neq k}}^{\infty} L^n$

- Sei w ein Wort. Für die einelementige Menge $\{w\}$ definieren wir:
  * $w^* := \{w\}^* = \{\varepsilon, w, ww, www, \dots\}$
  * $w^+ := \{w\}^+ = \{w, ww, www, \dots\}$.

- Seien y, w Wörter. $(y \in w)$: ($\exists$ Wörter x, z so das w=xyz).

- Außerdem werden wir folgende asymptotische Notationen brauchen: Seien $f, g : \mathbb{N} \to \mathbb{R}_+$ zwei Funktionen
  * $g = O(f) :\Leftrightarrow (\exists c \in \mathbb{R}_+, \exists n_o \in \mathbb{N})$ *so dass* $(\forall n \geq n_0)$ *gilt* $(g(n) \leq cf(n))$
  * $g = \Omega(f) :\Leftrightarrow (\exists c \in \mathbb{R}_+, \exists n_o \in \mathbb{N})$ *so dass* $(\forall n \geq n_0)$ *gilt* $(g(n) \geq cf(n))$
  * $g = \Theta(f) :\Leftrightarrow g = O(f)$ *und* $g = \Omega(f)$

2.2 Definition einer CFG

Eine kontextfreie Grammatik CFG ist ein 4–Tupel, G=(N, Σ, P, S), mit:

- N ist eine endliche nicht leere Menge von Nichtterminalen bzw. Variablen.

- Σ ist eine endliche nicht leere Menge von Terminalen. Wir nehmen an, das Σ und N disjunkt sind.

- P ist eine endliche Menge von Produktionen der Form $A \rightarrow \alpha$ mit $A \in N$ und $\alpha \in (N \cup \Sigma)^*$. Es empfiehlt sich, die Produktionen durchzunumerieren.

- $S \in N$ ist das Startsymbol.

Konventionen Mit $V := N \cup \Sigma$ bezeichnen wir die Menge des Gesamtalphabets. Wenn nichts anderes definiert wird, dann sind:

- Kleinbuchstaben vom Anfang des Alphabets (a, b, c, d, e) Terminale aus Σ,

- Kleinbuchstaben vom Ende des Alphabets (t, u, v, w, x, y, z) Wörter aus Σ^*,

- Grossbuchstaben vom Ende des Alphabets (X, Y, Z) Symbole aus V,

- Grossbuchstaben vom Anfang des Alphabets (A, B, C,. . . ,U, W) Nichtterminale aus N, S ist – falls nicht anders gesagt – das Startsymbol,

- Griechische Buchstaben (α, β , γ , δ) Elemente aus V^*.

- Statt der Schreibweise $A \rightarrow \alpha_1$, $A \rightarrow \alpha_2$,. . . , $A \rightarrow \alpha_k$ schreiben wir

$$A \rightarrow \alpha_1|\alpha_2 \ldots |\alpha_k \ bzw. \ A \rightarrow \left| \begin{array}{l} \alpha_1 \\ \alpha_2 \\ \ldots \\ \ldots \\ \alpha_k \end{array} \right.$$

- Diese letzte Schreibweise benutzen wir vor allem, wenn wir alle möglichen Ableitungen, die aus A erhältlich sind, darstellen wollen. Dies werden wir bei der Darstellung vom Ableitungsschema einiger Grammatiken brauchen.

2.3 CFG und Ableitungen

Sei G=(N, Σ, P, S) eine CFG.

- Auf V^*NV^* definieren wir eine Ableitungsrelation $\Longrightarrow$ mit $(\alpha_1 A \alpha_2 \Longrightarrow \alpha_1 \alpha \alpha_2) :\Leftrightarrow ((A \rightarrow \alpha) \in P)$.
  $\overset{i}{\Longrightarrow}$ ist die i-fache, $\overset{+}{\Longrightarrow}$ die transitive, $\overset{*}{\Longrightarrow}$ die reflexive und transitive Hülle von $\Longrightarrow$, wobei $i \in \mathbb{N}_0$ (insbesondere gilt: $A \overset{0}{\Longrightarrow} A$, $\forall\ A \in N$).
  Wenn nicht klar ist, um welche Grammatik G es sich handelt, schreiben wir $\underset{G}{\Longrightarrow}$ statt $\Longrightarrow$.

- Falls $(A \rightarrow \alpha) \in P$, dann nennt man den Übergang $\alpha_1 A \alpha_2 \Longrightarrow \alpha_1 \alpha \alpha_2$ einen Ableitungsschritt.

- Sind die Produktionen von P durchnumeriert, so ist $\xrightarrow{j}$ (bzw. $\xrightarrow{n*j}$) die einmalige (bzw. n–malige) Anwendung der Produktion mit der Nummer j.

- $\xrightarrow{*j}$ (bzw. $\xrightarrow{*j*k}$) ist die mehrmalige Anwendung der Produktion mit der Nummer j (bzw. die mehrmalige Anwendung der Produktionen mit den Nummern j und/oder k).

- Eine Ableitung ist eine Folge von Ableitungsschritten. Unter einer Linksableitung verstehen wir eine Ableitung, in der in jedem Ableitungsschritt das linkeste Nichtterminal (Variable) abgeleitet wird. Formal definieren wir eine Linksableitungsrelation $\overset{lm}{\Longrightarrow}$ mit: $(wA\beta \overset{lm}{\Longrightarrow} w\alpha\beta) :\Leftrightarrow ((A \rightarrow \alpha) \in P$ und $w \in \Sigma^*)$.

- Mit $\mathbf{j_1 j_2 \ldots j_n}$ bezeichnen wir die Linksableitung, in der man die Produktionen mit den Nummern j_1, j_2,...,j_n in dieser Reihenfolge anwendet.

- Ein geordneter markierter Baum[1] ist ein Ableitungsbaum für G, wenn folgendes gilt:

  * Jeder Knoten hat ein Symbol aus $V \cup \{\varepsilon\}$ als Markierung.
  * Die Markierung der Wurzel ist das Startsymbol S.
  * Wenn ein innerer Knoten die Markierung A hat, dann muss A aus N sein.
  * Wenn der Knoten n die Markierung A hat und dessen Söhne n_1, n_2,..., n_k von links her mit den Markierungen X_1, X_2,..., X_k versehen sind, dann muss $A \rightarrow X_1 X_2 \ldots X_k$ eine Produktion in P sein.
  * Wenn der Knoten n die Markierung ε hat, dann ist n ein Blatt und der einzige Sohn seines Vaterknotens.

- Unter einem A–Pumpbaum verstehen wir einen Ableitungsbaum mit der Wurzel A und einem ausgezeichneten Vorkommen von A als echten[2] inneren Knoten.

- Wenn wir die Markierungen der Blätter von links nach rechts lesen, so erhalten wir die sogenannte Front des Ableitungsbaums.

- Die von der CFG G erzeugte Sprache ist L(G):= $\{w \in \Sigma^* | S \stackrel{*}{\Longrightarrow} w\}$.

- Eine Sprache L heißt kontextfrei (CFL), wenn es eine CFG G mit L=L(G) gibt.

---

[1]Die Kenntnis des Begriffes Baum (siehe [11] bzw. [5, Seite 3]) wird vorausgesetzt.

[2]Ein echter innerer Knoten ist ein innerer Knoten, der nicht die Wurzel ist.

- Eine Variable A heißt nützlich, falls gilt: $(\exists \alpha,\ \beta \in V^*;\ w \in \Sigma^*)$ so dass $(S \overset{*}{\Longrightarrow} \alpha A \beta \overset{*}{\Longrightarrow} w)$.

- Ein Nichtterminal A heißt zyklisch, falls $A \overset{+}{\Longrightarrow} A$ gilt.

- Eine Grammatik enthält (nützliche) Zyklen, falls sie ein (nützliches) zyklisches Nichtterminal besitzt.

- Eine Grammatik heißt zyklenfrei, falls sie keine Ableitungen der Form $A \overset{+}{\Longrightarrow} A$ für irgendein Nichtterminal A hat.

Bemerkung 2.3.1

- *Es ist sinnvoll, Ableitungen als Bäume darzustellen. Allerdings ist dabei zu beachten, dass zu jeder Ableitung genau ein Ableitungsbaum gehört, während zu einem Ableitungsbaum verschiedene Ableitungen des gleichen Wortes gehören können. Aus jedem Ableitungsbaum ist genau eine Linksableitung erhältlich.*

- *Wir definieren die kontextfreien Sprachen CFL aus gutem Grund über die CFG, denn die Ableitungsbäume spielen eine zentrale Rolle in dieser Arbeit und lassen sich bekanntlich besser über CFG definieren. Man könnte eine Sprache auch durch Kellerautomaten PDA und die Mehrdeutigkeit als die Anzahl der akzeptierenden Berechnungen[3] definieren, aber diese beiden Definitionen für die Mehrdeutigkeit erweisen sich als äquivalent. Denn die bekannten Konstruktionen zur Umwandlung eines PDA in eine äquivalente CFG und umgekehrt erhalten den Grad der Mehrdeutigkeit.*

Beispiel 2.3.1

[3] Näheres siehe [4].

Betrachten wir die Grammatik G mit den Produktionen:

1: $S \rightarrow AB$;

2: $S \rightarrow SS$;

3: $S \rightarrow \varepsilon$;

4: $A \rightarrow a$ und

5: $B \rightarrow b$.

In Abbildung 2.1 auf die Seite 16 sind Ableitungsbäume für G dargestellt. Zum Ableitungsbaum in Abbildung 2.1b gibt es zwei verschiedene Ableitungen:

$S \xrightarrow{1} AB \xrightarrow{4} aB \xrightarrow{5} ab$, also **145**.

$S \xrightarrow{1} AB \xrightarrow{5} Ab \xrightarrow{4} ab$, also **154**.

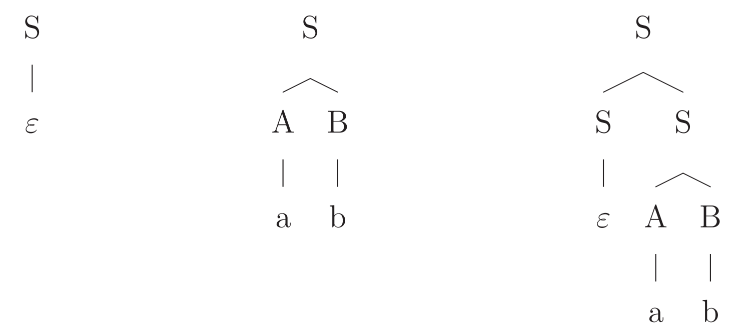

Abbildung 2.1: Ableitungsbäume

2.4 Ogden's Lemma

Satz 2.4.1 (Ogden's Lemma) *(*OGDEN*'s Lemma, [9])*

Für jede CFG G=(N, Σ, P, S) gibt es eine Konstante h=h(G) aus $\mathbb{N}$*, so dass für jedes Wort* $z \in L(G)$ *mit* $|z| \geq h$ *folgende Aussage gilt: Wenn wir in z mindestens h Positionen markieren, dann hat z eine Zerlegung uvwxy mit:*

- *In w tritt mindestens eine der h markierten Positionen von z auf,*

- ***Entweder*** *treten sowohl in u als auch in v markierte Positionen von z auf* ***oder*** *es treten sowohl in x als auch in y markierte Positionen von z auf,*

- *In vwx treten höchstens h der markierten Positionen von z auf,*

- $\exists A \in N$ *mit*
  $S \overset{+}{\Longrightarrow} uAy \overset{+}{\Longrightarrow} uvAxy \overset{+}{\Longrightarrow} \ldots \overset{+}{\Longrightarrow} uv^qAx^qy \overset{+}{\Longrightarrow} uv^qwx^qy \in L(G)$ *für alle* $q \in \mathbb{N}_0$ *(insbesondere für q=0 gilt $S \overset{+}{\Longrightarrow} uAy \overset{+}{\Longrightarrow} uwy \in L(G)$).*

Beweis Der Beweis wird in [9, Seite 193, theorem 2.24], [7, Seite 157] sowie [12, Seite 78] geliefert.
q. e. d.

Bemerkung 2.4.1
- *Die Konstante h aus* Ogden*'s Lemma hängt von der Grammatik G (nicht von der Sprache $L(G)$) ab und wird meistens in der Grösenordnung von $t^{2|N|+3}$ gewählt, wobei*
  t*:=max*$(max\{|\alpha|$ *mit* $(A \rightarrow \alpha) \in P\}; 2)$.

- *Punkt (4) in* OGDEN*'s Lemma (Satz 2.4.1) auf die Seite 16 besagt, dass jeder Ableitungsbaum von z=$uvwxy$ in G einen A–Pumpbaum mit der Front vwx hat, der q mal "'gepumpt"' werden kann, um einen Ableitungsbaum für uv^qwx^qy in G zu erhalten. siehe Abbildung 2.2 auf die Seite 18*

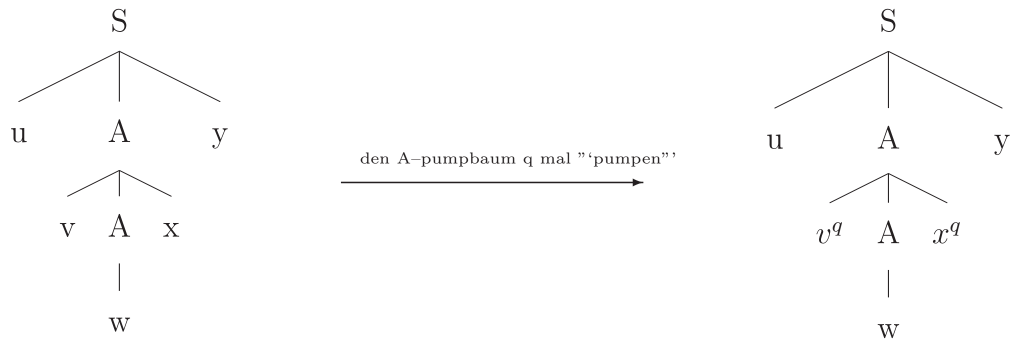

Abbildung 2.2: "'Pumpen"' eines A–Pumpbaumes

2.5 Mehrdeutigkeitsgrad und CFG

Seien $k \in \mathbb{N}$, G=(N, Σ, P, S) eine CFG und L(G) die von G erzeugte Sprache.

- Der Mehrdeutigkeitsgrad (degree of ambiguity) $da_G(w)$ eines Wortes $w \in \Sigma^*$ ist definiert als da_G(w):=Anzahl der Linksableitungen (Ableitungsbäume) für w in G.

- Der Mehrdeutigkeitsgrad einer CFG G ist definiert als da_G(n):=sup$\{da_G(w) | w \in \Sigma^*$ und $|w| \leq n\}$.

- G ist mindestens k–deutig :⇔ Es gibt ein Wort, das mindestens k Ableitungsbäume in G hat.

- G ist höchstens k-deutig :⇔ jedes Wort hat höchstens k Ableitungsbäume in G.

- G ist k–deutig :⇔ G ist mindestens k-deutig und G ist höchstens k–deutig.

- Mit CFG ($da(n) \leq k$) bezeichnen wir die Menge aller k-deutigen CFG.

- G ist polynomiell vom Grad k mehrdeutig $:\Leftrightarrow da_G(n) = \Theta(n^k)$.
- Mit CFG($da(n) = \Theta(n^k)$) bezeichnen wir die Menge aller polynomiell vom Grad k mehrdeutigen Grammatiken
- G ist exponentiell mehrdeutig $:\Leftrightarrow da_G(n) = 2^{\Theta(n)}$.
- Mit CFG($da(n) = 2^{\Theta(n)}$) bezeichnen wir die Menge aller exponentiell mehrdeutigen Grammatiken.

Bemerkung 2.5.1

- *Für k=1 erhalten wir die bekannte Klasse der eindeutigen Grammatiken UCFG.*
- *Mit CFG(Eigenschaft) bezeichnen wir die Menge aller Grammatiken, die die Bedingung "'Eigenschaft"' erfüllen. Beispielsweise bezeichnet CFG($|N| \leq 4$) die Menge aller Grammatiken, die höchstens 4 Nichtterminale besitzen.*
- *Ist eine CFG G mindestens k–deutig für jedes $k \in \mathbb{N}$, so bezeichnen wir G als ∞–deutig bzw. als mehrdeutig vom Grad unendlich.*
- *Eine Grammatik G heißt für einzelne Wörter mehrdeutig vom Grad unendlich, falls es ein Wort w gibt, mit $da_G(w) = \infty$.*

Beispiel 2.5.1

- Die CFG G=($\{S\}, \{a, b\}, \{\mathbf{1} : S \to aSb,\ \mathbf{2} : S \to ab\}, S$) ist eindeutig, denn es gilt L(G)=$\{a^i b^i \ mit\ i \geq 1\}$. Für jedes $i \geq 1$ ist $\mathbf{1^{i-1}2}$ die Linksableitung des eindeutigen Ableitungsbaumes für $a^i b^i$

$$da_G(n) = \begin{cases} 0 & , n = 0 \\ 1 & , sonst \end{cases}$$

- Für die CFG G=($\{S\}, \{a\}, \{\mathbf{1} : S \rightarrow SS,\ \mathbf{2} : S \rightarrow a\}, S$) gilt $L(G) = \{a^n | n \geq 1\}$. Mit Hilfe der Rekursionsgleichung:

$$da_G(a) = 1 \tag{2.5.1}$$

$$da_G(a^n) = \sum_{i=1}^{n-1} da_G(a^i) da_G(a^{n-i}),\ \text{für } n \geq 2 \tag{2.5.2}$$

  kann man (etwa mit Hilfe von erzeugenden Funktionen) zeigen: $da_G(a^n) = \frac{1}{n}\binom{2(n-1)}{n-1}$ für $n \geq 1$. Damit ist da(n) exponentiell in n. Die Grammatik G ist somit exponentiell (unendlich) mehrdeutig.

- Für die Grammatik G=($\{S\}, \{a\}, \{\mathbf{1} : S \rightarrow S,\ \mathbf{2} : S \rightarrow a\}, S$) gilt, dass für jedes k aus $\mathbb{N}_0$ $\mathbf{1^k 2}$ eine Linksableitung eines Ableitungsbaumes für das Wort $a \in L(G)$ ist. Das heißt, $da_G(a) = \infty$. Also ist G für einzelne Wörter mehrdeutig vom Grad unendlich.

2.6 Mehrdeutigkeitsgrad von zyklenfreien Grammatiken

Es ist leicht einzusehen, dass der Mehrdeutigkeitsgrad einer Grammatik für einzelne Wörter unendlich ist, falls sie (nützliche) Zyklen enthält. Für die zyklenfreie Grammatiken zeigen wir folgenden Satz:

Satz 2.6.1 *Es gibt keine zyklenfreie Grammatiken, deren Mehrdeutigkeitsgrad größer als $2^{\Theta(n)}$ ist (wie z. B. $\Theta(n^n)$).*

Beweis Sei G=(N, Σ, P, S) eine zyklenfreie CFG. Aus S kann man:

- in einem einzigen Linksableitungsschritt höchstens auf $|P|$ viele (verschiedene) Weisen ableiten,

- in zwei Linksableitungsschritten höchstens auf $|P|^2$ viele (verschiedene) Weisen ableiten,
- ...,
- in i Linksableitungsschritten auf höchstens $|P|^i$ viele (verschiedene) Weisen ableiten.

Das heißt, die Anzahl der Ableitungsbäume, die man in höchstens i Linksableitungsschritten[4] erhält, ist durch $|P|^i$ beschränkt

Andererseits folgt aus [3, Seite 130, Theorem 4.1], dass für jede zyklenfreie Grammatik Konstanten a und b existieren, so dass $(A \overset{i}{\Longrightarrow} w) \Longrightarrow (i \leq a|w| + b)$

Insgesamt erhalten wir, dass für jede zyklenfreie Grammatik die Anzahl der Ableitungsbäume eines Wortes w durch $|P|^{a|w|+b}$ beschränkt ist

Mit $n := |w|$ erhalten wir $|P|^{a|w|+b} = 2^{(an+b)\log|P|} = 2^{\Theta(n)}$, wobei log der Logarithmus zur Basis 2 ist.

q. e. d.

2.7 Mehrdeutigkeitsgrad und CFL

Sei $k \in \mathbb{N}$.

- Eine CFL L ist k–deutig :⇔ jede Grammatik, die L erzeugt ist mindestens k-deutig und es gibt eine höchstens k-deutige Grammatik, die L erzeugt.

[4] Wir erinnern uns, dass es zu jedem Ableitungsbaum genau eine Linksableitung gibt und umgekehrt (siehe Bemerkung 2.3.1 auf Seite 15).

- Für k=1 sprechen wir von eindeutigen Sprachen (UCFL).
- Mit CFL(da(n)=k) bezeichnen wir die Menge aller k-deutigen CFL.
- Eine CFL L ist polynomiell vom Grad k mehrdeutig :⇔ für jede CFG G mit L=L(G) gilt $G \in CFG(da(n) = \Omega(n^k))$ und es gibt eine Grammatik aus $CFG(da(n) = O(n^k))$ für L.
- Mit $CFL(da(n) = \Theta(n^k))$ bezeichnen wir die Menge aller polynomiell vom Grad k mehrdeutigen CFL.
- Eine CFL L ist exponentiell mehrdeutig :⇔ für jede CFG G mit L=L(G) gilt $G \in CFG(da(n) = 2^{\Omega(n)})$ und es gibt eine Grammatik aus $CFG(da(n) = 2^{O(n)})$ für L.
- Mit $CFL(da(n) = 2^{\Theta(n)})$ bezeichnen wir die Menge aller exponentiell mehrdeutigen CFL.
- Mit CFL(Eigenschaft) bezeichnen wir die Menge aller Sprachen, die die Bedingung "‘Eigenschaft"’ erfüllen.

Bemerkung 2.7.1 *Da man Zyklen aus einer Grammatik effizient entfernen[5] kann, folgt aus Satz 2.6.1 (auf Seite 20), dass es keine mehrdeutigen Sprachen geben kann, deren Mehrdeutigkeitsgrade größer als $2^{\Theta(n)}$ sind (wie z. B. $\Theta(n^n)$). Außerdem wissen wir aus [10], dass es keine Grammatiken (und somit keine Sprachen) gibt, deren Mehrdeutigkeit stärker als polynomiell, aber schwächer als exponentiell wächst (wie z. B. $\Theta(2^{\sqrt{n}})$). Deshalb untersuchen wir hauptsächlich konstant mehrdeutige, polynomiell mehrdeutige und exponentiell mehrdeutige Grammatiken und Sprachen, womit wir uns in den nächsten drei Kapiteln beschäftigen werden.*

[5] Näheres siehe [9]

Kapitel 3

Konstant mehrdeutige Grammatiken und Sprachen

3.1 Motivation

In diesem Kapitel beschäftigen wir uns mit der Frage, ob es für eine (k+1)–deutige CFG eine äquivalente k-deutige CFG gibt, wobei k eine Konstante aus ist. Zuerst betrachten wir die kontextfreie Grammatik $G = (\{S\}, \{(,)\}, \{S \rightarrow SS|(S)|\varepsilon\}, S)$ Diese Grammatik erzeugt alle wohlgeformten Klammerausdrücke und ist mehrdeutig vom Grad unendlich. Sie ist allerdings äquivalent zur eindeutigen Grammatik $\widetilde{G} = (\{S\}, \{(,)\}, \{S \rightarrow (S)S|\varepsilon\}, S)$ Es ist uns also gelungen, für die mehrdeutige Grammatik G eine äquivalente eindeutige Grammatik $\widetilde{G}$ zu finden. Dies geht in allgemein nicht, denn Maurer hat in [8] für jedes k aus die Existenz einer k–deutigen Sprache nachgewiesen. Diese bekannte Tatsache beweisen wir eleganter: wir benutzen eine ”‘einfachere”’ Sprache und eine ”‘bessere”’ Beweisidee (Ogden's Lemma[1]). Wir zeigen

[1] Als Maurer den Beweis ausführte, existierte Ogden's Lemma noch nicht.

3.2 Existenz k-deutige Sprachen

Satz 3.2.1 *Sei k eine Konstante aus* $\mathbb{N}$ *. Es gibt k–deutige Sprachen.*

Beweis Sei k eine Konstante aus $\mathbb{N}$. Wir betrachten die Sprache $L_k := \{a^m b_1^{m_1} b_2^{m_2} \dots b_k^{m_k} | m, m_1, m_2, \dots, m_k \geq 1, \exists\ i\ mit\ m = m_i\}$. Für k=1 erhalten wir die (bekannte) eindeutige Sprache $L_1 := \{a^m b_1^m | m \geq 1\}$. Seien $k \geq 2$ und G=(N, Σ, P, S) eine CFG für L_k. Sei h die Konstante aus OGDEN's Lemma (Satz 2.4.1 auf Seite 16) für die Grammatik G. Wir betrachten die Wörter z_i aus L_k der Form

$$z_i := a^h b_1^{h_1} b_2^{h_2} \dots b_k^{h_k}\ mit \quad h_j := \begin{cases} h & , \text{ für } j = i \\ h + h! & , \ sonst \end{cases} \quad , \text{für } \ i = 1, \dots, k$$

und markieren alle a's. Sei nun $z_i = u_i v_i w_i x_i y_i$ die nach OGDEN's Lemma (Satz 2.4.1 auf Seite 16) existierende Zerlegung. Zuerst zeigen wir

Lemma 3.2.1 *Für die nach* OGDEN*'s Lemma (Satz 2.4.1 auf Seite 16) existierende Zerlegung* $z_i = u_i v_i w_i x_i y_i$ *gilt:*

$$\begin{aligned}
u_i &= a^{r_i} & 1 &\leq r_i \leq h-2, \\
v_i &= a^{s_i} & 1 &\leq s_i \leq h-2, \\
w_i &= a^{h-s_i-r_i} b_1^{h+h!} \dots b_{i-1}^{h+h!} b_i^{t_i} & 0 &\leq t_i \leq h-1, \\
x_i &= b_i^{s_i} \\
y_i &= b_i^{h-s_i-t_i} b_{i+1}^{h+h!} \dots b_k^{h+h!}.
\end{aligned}$$

Beweis (Lemma 3.2.1) Aus OGDEN's Lemma folgt: **Entweder** treten sowohl in u_i als auch in v_i markierte Positionen von z_i auf **oder** es treten sowohl in x_i als auch in y_i markierte Positionen von z_i auf.

Wenn sowohl x_i als auch y_i markierte Positionen haben, dann ist $x_i \in a^+$, denn w_i hat gemäß OGDEN's Lemma mindestens eine markierte Position und x_i kommt zwischen w_i und y_i in z_i vor.
Da v_i in z_i links von w_i vorkommt, gilt $v_i \in a^*$.
Daraus folgt, dass $v_i x_i = a^s$ mit $1 \leq s \leq h-2$. Damit ist aber $\overline{z} := u_i v_i^2 w_i x_i^2 y_i = a^{h+s} b_1^{h+h!} \ldots b_{i-1}^{h+h!} b_i^h b_{i+1}^{h+h!} \ldots b_k^{h+h!} \notin L_k$. Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma. Damit folgt: x_i und y_i haben **nicht** beide markierte Positionen. D. h. sowohl u_i als auch v_i müssen markierte Positionen haben. Da w_i mindestens eine markierte Position hat, folgt $v_i \in a^+$. Damit ist $v_i = a^s$, wobei $1 \leq s \leq h-2$.
Damit $\overline{z} := u_i v_i^2 w_i x_i^2 y_i \in L_k$ ist, müssen im Wort $\overline{z}$ die Terminale a, b_1,...,b_k in dieser Reihenfolge vorkommen. Damit enthält x_i nur Terminale eines Typs. Das heißt, $x_i \in a^* \cup b_1^* \cup \cdots \cup b_k^*$. Damit kommen für x_i folgende Fälle in Frage (Siehe Tabelle 3.1 auf die Seite 25):

	a^h	$b_1^{h+h!} \ldots b_{i-1}^{h+h!}$	b_i^h	$b_{i+1}^{h+h!} \ldots b_k^{h+h!}$	Widerspruch
Fall 1	$u_i v_i w_i x_i$	y_i			$u_i v_i^2 w_i x_i^2 y_i \notin L_k$
Fall 2	$u_i v_i$	x_i			$u_i v_i^2 w_i x_i^2 y_i \notin L_k$
Fall 3	$u_i v_i$			x_i	$u_i v_i^2 w_i x_i^2 y_i \notin L_k$
Fall 4	$u_i v_i$		x_i		kein Widerspruch

Tabelle 3.1: Zerlegung von z_i in $u_i v_i w_i x_i y_i$

Fall 1 $x_i = \varepsilon$ oder $x_i \in a^+$.
Hier gilt $\overline{z} := u_i v_i^2 w_i x_i^2 y_i = a^{h+s+|x_i|} b_1^{h+h!} \ldots b_{i-1}^{h+h!} b_i^h b_{i+1}^{h+h!} \ldots b_k^{h+h!}$.
Da u_i und w_i jeweils mindestens eine markierte Position (also mindestens ein

a) haben, folgt $s+|x_i| \leq h-2$. Daraus folgt $h < h+s+|x_i| < h+h!$. Damit ist $\overline{z} := u_i v_i^2 w_i x_i^2 y_i = a^{h+s+|x_i|} b_1^{h+h!} \ldots b_{i-1}^{h+h!} b_i^h b_{i+1}^{h+h!} \ldots b_k^{h+h!} \notin L_k$. Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Fall 2 $\quad x_i \in b_r^+$ mit $1 \leq r \leq i-1$

Es gilt $\overline{z} := u_i v_i^2 w_i x_i^2 y_i = a^{h+s} b_1^{h+h!} \ldots b_r^{h+s+|x_i|} \ldots b_{i-1}^{h+h!} b_i^h b_{i+1}^{h+h!} \ldots b_k^{h+h!} \notin L_k$, denn $1 \leq s \leq h-2$. Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Fall 3 $\quad x_i \in b_r^+$ mit $i+1 \leq r \leq k$

Es gilt für das Wort $\overline{z} := u_i v_i^2 w_i x_i^2 y_i = a^{h+s} b_1^{h+h!} \ldots b_{i-1}^{h+h!} b_i^h b_{i+1}^{h+h!} \ldots b_r^{h+h!+|x_i|} \ldots b_k^{h+h!} \notin L_k$. Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Fall 4 $\quad x_i \in b_i^+$ d.h $x_i = b_i^t$ mit $1 \leq t \leq h$

Die Fälle 1, 2 und 3 können nicht eintreten. Es muss also Fall 4 eintreten.
Für das Wort $\overline{z} := u_i v_i^2 w_i x_i^2 y_i$
gilt:
$\#_a(\overline{z}) = h + s$
$\#_{b_j}(\overline{z}) = h + h!$ für $j \neq i$
$\#_{b_i}(\overline{z}) = h + t$
Da aber $\overline{z} := u_i v_i^2 w_i x_i^2 y_i$ in L_k sein muss (wegen OGDEN's Lemma), erhalten wir t=s. Bisher haben wir gezeigt: Für jede Zerlegung von z_i in $u_i v_i w_i x_i y_i$, die die Bedingungen vom OGDEN's Lemma erfüllt, muss gelten:

$v_i = a^s$ und $x_i = b_i^s$ mit $1 \leq s \leq h-2$.

Daraus folgt:

$$
\begin{aligned}
u_i &= a^{r_i} && 1 \le r_i \le h-2,\\
v_i &= a^{s_i} && 1 \le s_i \le h-2,\\
w_i &= a^{h-s_i-r_i} b_1^{h+h!} \dots b_{i-1}^{h+h!} b_i^{t_i} && 0 \le t_i \le h-1,\\
x_i &= b_i^{s_i}\\
y_i &= b_i^{h-s_i-t_i} b_{i+1}^{h+h!} \dots b_k^{h+h!}.
\end{aligned}
$$

q. e. d.

Wir setzen nun den Beweis von Satz 3.2.1 fort.
Aus Satz 2.4.1 (auf Seite 16) folgt, dass es ein A_i aus N geben muss, so dass $S \overset{+}{\Longrightarrow} u_i A_i y_i \overset{+}{\Longrightarrow} u_i v_i A_i x_i y_i \overset{+}{\Longrightarrow} u_i v_i w_i x_i y_i = z_i$

Damit hat jeder Ableitungsbaum B_i mit der Front z_i in G einen A_i–Pumpbaum, welcher die Anzahl der Symbole a's und b_i's in z_i gleichmäßig erhöht. Somit sieht der Ableitungsbaum B_i wie folgt aus (siehe Abbildung 3.1 auf die Seite 28):

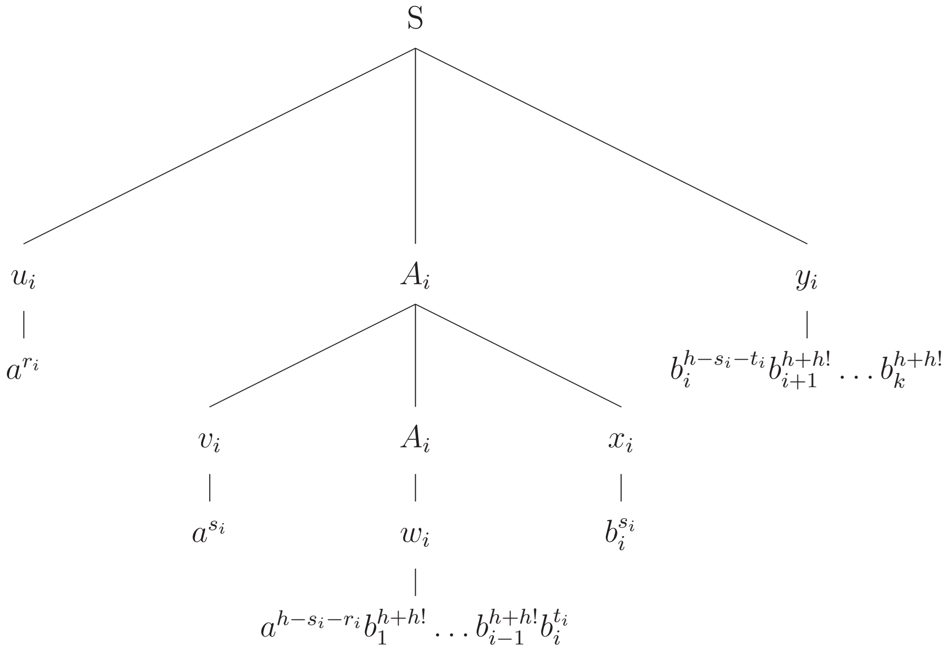

Abbildung 3.1: Ableitungsbaum B_i mit A_i–Pumpbaum für $z_i := a^h b_1^{h+h!} \dots b_{i-1}^{h+h!} b_i^h b_{i+1}^{h+h!} \dots b_k^{h+h!}$

Pumpen wir den A_i–Pumpbaum (des Baumes B_i) $q_i := \frac{h!}{s_i} + 1$ mal, erhalten wir einen Ableitungsbaum T_i in G für das Wort $z := a^{h+h!} b_1^{h+h!} b_2^{h+h!} \dots b_k^{h+h!}$. Der Ableitungsbaum T_i ist in der Abbildung 3.2 auf die Seite 29 dargestellt.

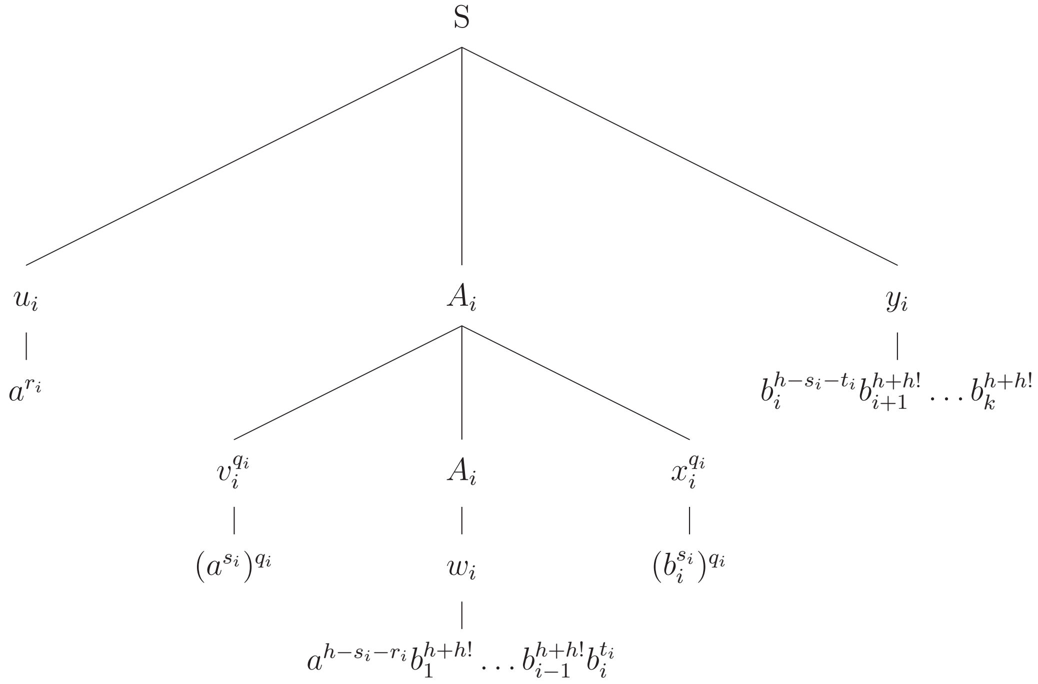

Abbildung 3.2: Ableitungsbaum T_i für z:=$a^{h+h!} b_1^{h+h!} b_2^{h+h!} \dots b_k^{h+h!}$

Verwenden wir die gleiche Argumentation für alle Wörter z_1, z_2,..., z_k, dann erhalten wir k Ableitungsbäume T_1, T_2,..., T_k für das Wort $z := a^{h+h!} b_1^{h+h!} b_2^{h+h!} \dots b_k^{h+h!}$ in G. Als nächstes beweisen wir, dass diese k Ableitungsbäume paarweise verschieden sind, indem wir für $i \neq j \in \{1, 2, \dots, k\}$ zeigen, dass die Ableitungsbäume T_i und T_j für das Wort $z := a^{h+h!} b_1^{h+h!} b_2^{h+h!} \dots b_k^{h+h!}$ verschieden sind. Nehmen wir nun an, dass die Ableitungsbäume T_i und T_j für das Wort $z := a^{h+h!} b_1^{h+h!} b_2^{h+h!} \dots b_k^{h+h!}$ gleich wären, also gleich einem Ableitungsbaum T. Dieser Ableitungsbaum T müsste sowohl einen A_i– als auch einen A_j–Pumpbaum haben. Wir können ausschließen, dass der Knoten A_i links vom Knoten A_j im Baum T vorkommt, sonst erhielten wir ein Wort in L_k, in dem b_i's vor a's vorkämen (siehe Abbildung 3.3 auf die Seite 30). Analog gilt es für A_i rechts von A_j. Es bleiben somit zwei Fälle zu untersuchen: Der Knoten A_j ist ein Nachfolger des Knotens A_i im Baum T bzw. der Knoten A_i ist ein Nachfolger des Knotens A_j im Baum T (siehe Abbildung 3.4 auf die Seite 31). Diese beiden Fälle sind ähnlich, deshalb untersuchen wir nur den letzten Fall.

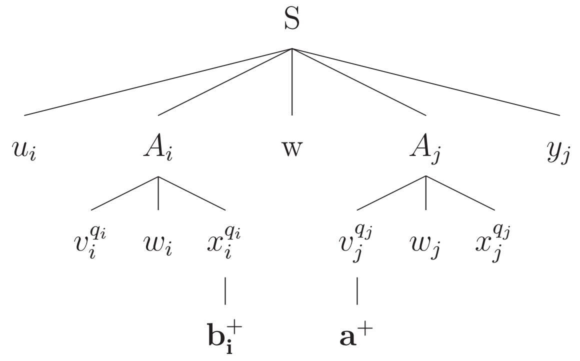

Abbildung 3.3: A_i kommt im Baum T links von A_j vor

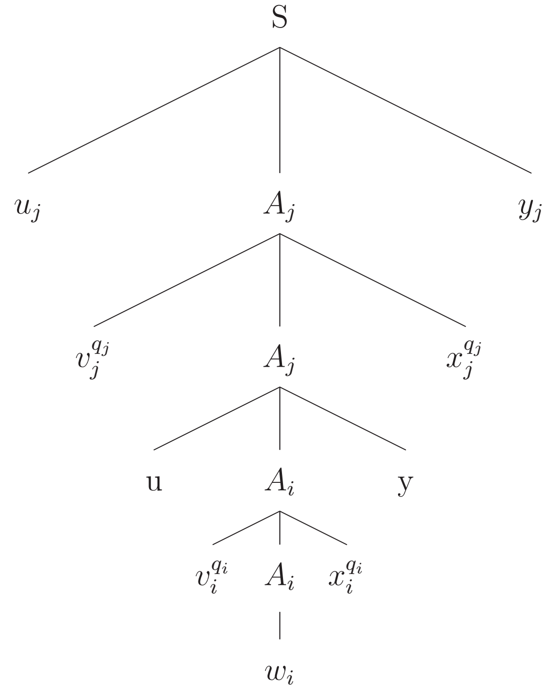

Abbildung 3.4: A_i ist Nachfolger von A_j im Baum T für z=$a^{h+h!}b_1^{h+h!}b_2^{h+h!}\dots b_k^{h+h!}$

Hier erhalten wir:

$$\begin{aligned} S &\stackrel{+}{\Longrightarrow} u_jA_jy_j \\ &\stackrel{+}{\Longrightarrow} u_jv_j^{q_j}A_jx_j^{q_j}y_j \\ &\stackrel{+}{\Longrightarrow} u_jv_j^{q_j}uA_iyx_j^{q_j}y_j \\ &\stackrel{+}{\Longrightarrow} u_jv_j^{q_j}uv_i^{q_i}w_ix_i^{q_i}yx_j^{q_j}y_j \\ &= z \in L_k \end{aligned}$$

Es gilt: $\#_a(z) = \#_{b_r}(z) = h + h! \;\; \forall r \in \{1, \dots, i, \dots, j, \dots, k\}$

Pumpen wir den A_i–Pumpbaum q_i+1 mal und den A_j–Pumpbaum q_j+1 mal, dann erhalten wir :

$$\begin{aligned} S &\overset{+}{\Longrightarrow} u_jA_jy_j \\ &\overset{+}{\Longrightarrow} u_jv_j^{q_j+1}A_jx_j^{q_j+1}y_j \\ &\overset{+}{\Longrightarrow} u_jv_j^{q_j+1}uA_iyx_j^{q_j+1}y_j \\ &\overset{+}{\Longrightarrow} u_jv_j^{q_j+1}uv_i^{q_i+1}w_ix_i^{q_i+1}yx_j^{q_j+1}y_j \\ &:= \tilde{z} \in L_k \end{aligned}$$

wobei:

$$\begin{aligned} \#_a(\tilde{z}) &= \#_a(z) + |v_j| + |v_i| = h + h! + |v_j| + |v_i| \\ \#_{b_i}(\tilde{z}) &= \#_{b_i}(z) + |x_i| = h + h! + |v_i| < h + h! + |v_j| + |v_i| \\ \#_{b_j}(\tilde{z}) &= \#_{b_j}(z) + |x_j| = h + h! + |v_j| < h + h! + |v_j| + |v_i| \\ \#_{b_r}(\tilde{z}) &= \#_{b_r}(z) = h + h! < h + h! + |v_j| + |v_i| \ \forall\ r \in \{i, j\} \end{aligned}$$

Es gibt also kein $r \in \{1, 2, \ldots, k\}$, so dass $\#_a(\tilde{z}) = \#_{b_r}(\tilde{z})$. Das ist ein Widerspruch zu $u_jv_j^{q_j+1}uv_i^{q_i+1}w_ix_i^{q_i+1}yx_j^{q_j+1}y_j = \tilde{z} \in L_k$.

Insgesamt haben wir bewiesen, dass für jede Grammatik G für L_k $da_G(n) \geq k$ gilt. Damit ist jede Grammatik für L_k mindestens k–deutig. Ob es aber eine k–deutige Grammatik für L_k gibt, wissen wir noch nicht.

3.3 Eine k–deutige Grammatik für L_k

Wir geben nun eine k–deutige Grammatik für L_k an. Dazu betrachten wir die Grammatik G_k:=(N, Σ, P, S) mit:

- $N := \{S\}\dot{\cup}\{S_1, S_2, \ldots, S_k\}\dot{\cup}\{A_1, A_2, \ldots, A_k\}\dot{\cup}\{B_1, B_2, \ldots, B_k\}\dot{\cup}\{T_1, T_2, \ldots, T_k\}$,
- $\Sigma := \{a, b_1, b_2, \ldots, b_k\}$,
- Die Regeln von P sind:

(1) $S \rightarrow S_1|S_2 \ldots |S_k$

(2) $S_i \to aT_ib_iB_{i+1}$ für $1 \leq i \leq k-1$

(3) $S_k \to aT_kb_k$

(4) $T_i \to aT_ib_i$ für $1 \leq i \leq k$

(5) $T_1 \to \varepsilon$

(6) $T_i \to A_{i-1}$ für $2 \leq i \leq k$

(7) $A_i \to A_ib_i|A_{i-1}b_i$ für $2 \leq i \leq k$

(8) $A_1 \to A_1b_1|b_1$

(9) $B_i \to b_iB_i|b_iB_{i+1}$ für $1 \leq i \leq k-1$

(10) $B_k \to b_kB_k|b_k$

Zuerst machen wir eine Beobachtung. Aus A_i können nur die Wörter $b_1^{m_1}b_2^{m_2}\dots b_i^{m_i}$ abgeleitet werden, aus B_i können nur die Wörter $b_i^{m_i}b_{i+1}^{m_{i+1}}\dots b_k^{m_k}$ abgeleitet werden.

Um aus dem Startsymbol S abzuleiten, müssen wir mit einer Produktion vom Typ **(1)** anfangen. Dabei haben wir die Wahl zwischen k Produktionen. Durch die Entscheidung für eine (oder mehrere) Produktion(en) vom Typ **(1)** legen wir unsere Vorgehensweise für die weitere Ableitung fest (siehe Abbildung 3.5 auf die Seite 34).

$$S \xrightarrow{1} \left| \begin{array}{l} S_1 \xrightarrow{2} aT_1b_1B_2 \xrightarrow{(m-1)*4} a^mT_1b_1^mB_2 \xrightarrow{5} a^mb_1^mB_2 \xrightarrow{*9*10} a^mb_1^mb_2^+\dots b_k^+ \\ \dots \\ \dots \\ S_i \xrightarrow{2} aT_ib_iB_{i+1} \xrightarrow{(m-1)*4} a^mT_ib_i^mB_{i+1} \xrightarrow{6} a^mA_{i-1}b_i^mB_{i+1} \\ \quad \xrightarrow{*7*8} a^mb_1^+\dots b_{i-1}^+b_i^mB_{i+1} \xrightarrow{*9*10} a^mb_i^+\dots b_{i-1}^+b_i^mb_{i+1}^+\dots b_k^+ \;; 2\le i\le k-1 \\ \dots \\ \dots \\ S_k \xrightarrow{3} aT_kb_k \xrightarrow{(m-1)*4} a^mT_kb_2^m \xrightarrow{6} a^mA_{k-1}b_k^m \xrightarrow{*7*8} a^mb_1^+\dots b_{k-1}^+b_k^m \end{array}\right.$$

Abbildung 3.5: Ableitungsschema der Grammatik G_k

Da es insgesamt k Produktionen vom Typ **(1)** gibt, kann es für jedes Wort $w \in L(G_k)$ höchstens k Linksableitungen geben. Wie man leicht erkennen kann, gibt es für das Wort $a^mb_1^mb_2^m\dots b_k^m$ genau k Linksableitungen. Das heißt, G_k ist k–deutig.
Insgesamt haben wir gezeigt, dass die Sprache L_k k–deutig ist.

q. e. d.

Bemerkung 3.3.1 *Es ist möglich eine k–deutige Sprache über einem Alphabet anzugeben, so dass von k unabhängig ist, d. h. mit fixem Alphabet* Σ. *Ein Beispiel dafür ist die Sprache* $\widetilde{L_k} := \{a^mb^{m_1}cb^{m_2}c\dots b^{m_k}c | m, m_1, m_2, \dots, m_k \geq 1, \exists\ i\ mit\ m = m_i\}$

Kapitel 4

Exponentiell mehrdeutige Grammatiken und Sprachen

4.1 Motivation

Es ist nicht schwer, eine exponentiell mehrdeutige Grammatik anzugeben, denn für die Grammatik $G_1 = (\{S\}, \{a, \mathbf{1}: S \to aS, \mathbf{2}: S \to aSa, \mathbf{3}: S \to a\}, S)$ gilt :

$L(G_1) = \{a^n | n \in \mathbb{N}\}$,

$da_{G_1}(a) = 1$, $da_{G_1}(aa) = 2$, $da_{G_1}(aaa) = 3$,

$$da_{G_1}(a^i) = \underbrace{da_{G_1}(a^{i-1})}_{falls\ man\ mit\ \mathbf{1}\ beginnt} + \underbrace{da_{G_1}(a^{i-2})}_{falls\ man\ mit\ \mathbf{2}\ beginnt} \quad \text{für } i \geq 3.$$

Das heißt, $da_{G_1}(a^i)$ sind die Fibonacci Zahlen Fib(i), also exponentiell in i.

Diese exponentiell mehrdeutige Grammatik G_1 ist allerdings äquivalent zur eindeutigen Grammatik $G_2 = (\{S\}, \{a\}, \{S \to aS|a\}, S)$. Damit ist die Sprache $L(G_1)$ nicht exponentiell mehrdeutig. Dieses negative Beispiel hindert uns nicht daran, die Existenz von exponentiell mehrdeutigen Sprachen nachzuweisen.

4.2 Existenz exponentiell mehrdeutige Sprachen

Satz 4.2.1 *Es gibt exponentiell mehrdeutige Sprachen.*

Beweis Sei $L = \{a^i b^i c^j | i, j \geq 1\} \cup \{a^i b^j c^i | i, j \geq 1\}$. Wir zeigen, dass die Sprache L^* exponentiell mehrdeutig ist. Sei G=(N, Σ, P, S) eine CFG für L^*. Im ersten Teil des Beweises benutzen wir das Beweisprinzip von OGDEN's Lemma wie in [9, Seite193, Theorem 2.24], [7, Seite 157] bzw. [12, Seite 78]. Dazu sei h die Konstante aus OGDEN's Lemma für die Grammatik G. Wir betrachten die Wörter aus $\{a^h b^h c^{h+h!}, a^h b^{h+h!} c^h\}^k$, wobei $k \in \mathbb{N}$. Sei $z = z_1 z_2 \dots z_k$, mit $z_i \in \{a^h b^h c^{h+h!}, a^h b^{h+h!} c^h\}$ $\forall i \in \{1, \dots, k\}$. Da $z \in L^*$ ist, gibt es in G einen Ableitungsbaum B(z) mit der Wurzel S und der Front z. Wir markieren alle a's in jedem z_i. Für jedes $i \in \{1, 2, \dots, k\}$ ist die Anzahl der markierten Positionen in z_i mindestens h, deshalb können wir –gemäß dem Konstruktionsverfahren im Beweis von OGDEN's Lemma wie in [9, Seite193, Theorem 2.24], [7, Seite 157] bzw. [12, Seite 78]– einen Pfad π_i im Baum B(z) von der Wurzel S zu einem Blatt in z_i konstruieren, dabei betrachten wir **nur** die a's von z_i als markiert. Der Pfad π_i enthält unter seinem letzten $|N|+1$ z_i-Verzweigungsknoten[1] mindestens zwei, die mit der gleichen Variable A_i markiert sind, so dass folgendes gilt:

1. $S \stackrel{+}{\Longrightarrow} \hat{u}_i A_i \hat{y}_i \stackrel{+}{\Longrightarrow} \hat{u}_i v_i A_i x_i \hat{y}_i \stackrel{+}{\Longrightarrow} \hat{u}_i v_i w_i x_i \hat{y}_i = z$

[1]Ein z_i–Verzweigungsknoten ist ein Knoten, der mindestens zwei Söhne mit markierten Nachkommen in z_i hat.

2. In w_i tritt mindestens eine der h markierten Positionen von z_i auf.

3. **Entweder** treten sowohl in $\hat{u}_i$ als auch in v_i markierte Positionen von z_i auf **oder** es treten sowohl in x_i als auch in $\hat{y}_i$ markierte Positionen von z_i auf.

4. In $v_i w_i x_i$ treten höchstens h der markierten Positionen von z_i auf,

5. $$\begin{aligned} S &\overset{+}{\Longrightarrow} \hat{u}_i A_i \hat{y}_i \\ &\overset{+}{\Longrightarrow} \hat{u}_i v_i A_i x_i \hat{y}_i \\ &\overset{+}{\Longrightarrow} \dots \\ &\overset{+}{\Longrightarrow} \hat{u}_i v_i^q A_i x_i^q \hat{y}_i \\ &\overset{+}{\Longrightarrow} \hat{u}_i v_i^q w_i x_i^q \hat{y}_i \in L^* \textit{ wobei } q \in \mathbb{N}_0. \end{aligned}$$

   (insbesondere für q=0 gilt $S \overset{+}{\Longrightarrow} \hat{u}_i A_i \hat{y}_i \overset{+}{\Longrightarrow} \hat{u}_i w_i \hat{y}_i$)

Der Pfad π_i und die Zerlegung $z = \hat{u}_i v_i w_i x_i \hat{y}_i$ sind in der Abbildung 4.1 auf die Seite 37 illustriert.

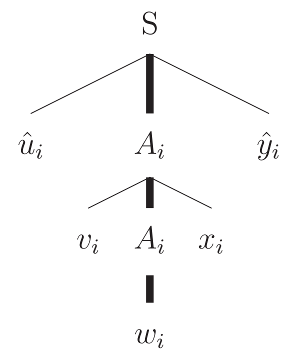

Abbildung 4.1: Illustration des Pfades π_i und die Zerlegung z= $\hat{u}_i v_i w_i x_i \hat{y}_i$

Als nächstes zeigen wir das

Lemma 4.2.1 *Für die nach* OGDEN*'s Lemma existierende Zerlegung* $z = \hat{u}_i v_i w_i x_i \hat{y}_i$ *gilt:*

$$
\begin{array}{lll}
falls\ \mathbf{z_i = a^h b^h c^{h+h!}}: & \hat{u}_i = z_1 \dots z_{i-1} u_i & u_i = a^{r_i}\ und\ 1 \le r_i \le h-2, \\
& v_i = a^{s_i} & 1 \le s_i \le h-2, \\
& w_i = a^{h-r_i-s_i} b^{h-s_i-t_i} & 0 \le t_i \le h-1, \\
& x_i = b^{s_i}, & \\
& \hat{y}_i = y_i z_{i+1} \dots z_k & y_i = b^{t_i} c^{h+h!}.
\end{array}
$$

$$
\begin{array}{lll}
falls\ \mathbf{z_i = a^h b^{h+h!} c^h}: & \hat{u}_i = z_1 \dots z_{i-1} u_i & u_i = a^{r_i}\ und\ 1 \le r_i \le h-2, \\
& v_i = a^{s_i} & 1 \le s_i \le h-2, \\
& w_i = a^{h-r_i-s_i} b^{h+h!} c^{t_i} & 0 \le t_i \le h-1, \\
& x_i = c^{s_i}, & \\
& \hat{y}_i = y_i z_{i+1} \dots z_k & y_i = c^{h-t_i-s_i}.
\end{array}
$$

Beweis (Lemma 4.2.1)
Wir behandeln nur den Fall $z_i = a^h b^h c^{h+h!}$ (der Fall $z_i = a^h b^{h+h!} c^h$ wird analog behandelt.). Es gilt:
Entweder treten sowohl in $\hat{u}_i$ als auch in v_i markierte Positionen von z_i auf **oder** es treten sowohl in x_i als auch in $\hat{y}_i$ markierte Positionen von z_i auf.
Wenn sowohl in x_i als auch in $\hat{y}_i$ markierte Positionen von z_i auftreten, dann ist $x_i \in a^+$ und $x_i \in z_i$, denn w_i hat gemäß OGDEN's Lemma mindestens eine markierte Position von z_i und x_i kommt zwischen w_i und $\hat{y}_i$ in z vor.
Sei $x_i = a^s \in z_i$ mit $1 \le s \le h-2$.
Daraus folgt: w_i endet mit einem a aus z_i und $\hat{y}_i = a^t b^h c^{h+h!} z_{i+1} \dots z_k$, wobei $1 \le t \le h-2$. Für v_i kommen folgende Möglichkeiten in Frage: ($v_i \in a^*$ und $v_i \in z_i$) oder (v_i beginnt mit einem Zeichen aus $z_1 \dots z_{i-1}$).
Ist $v_i \in a^*$ und $v_i \in z_i$, dann ist $\hat{u}_i v_i^2 w_i x_i^2 \hat{y}_i = z_1 \dots z_{i-1} a^{h+s+|v_i|} b^h c^{h+h!} z_{i+1} \dots z_k \notin L^*$, weil $a^{h+s+|v_i|} b^h c^{h+h!} \notin L$, denn

$1 < s + |v_i| < h - 2$. Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.
Beginnt v_i mit einem Zeichen aus $z_1 \dots z_{i-1}$, so ist $v_i w_i = vca^{h-s-t}$, wobei das Terminal c das letzte Zeichen c in z_{i-1} ist. Dann gilt:

$$\begin{array}{rcl} \hat{u}_i v_i^2 w_i x_i^2 \hat{y}_i & = & \hat{u}_i v_i v_i w_i x_i x_i \hat{y}_i \\ & = & \hat{u}_i v_i vca^{h-s-t} a^s a^s a^t b^h c^{h+h!} z_{i+1} \dots z_k \\ & = & \hat{u}_i v_i vca^{h+s} b^h c^{h+h!} z_{i+1} \dots z_k \notin L^*, \\ & weil & a^{h+s} b^h c^{h+h!} \notin L. \end{array}$$

Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Damit folgt: x_i und $\hat{y}_i$ haben **nicht** beide markierte Positionen von z_i. D. h. sowohl $\hat{u}_i$ als auch v_i müssen markierte Positionen von z_i haben. Da w_i mindestens eine markierte Position von z_i hat und v_i zwischen $\hat{u}_i$ und w_i in z vorkommt, folgt $v_i = a^s \in z_i$, wobei $1 \leq s \leq h - 2$.
Daraus folgt: w_i beginnt mit einem a aus z_i und $\hat{u}_i = z_1 \dots z_{i-1} a^r$, wobei $1 \leq r \leq h - 2$ ist. Wir behandeln zunächst zwei mögliche Fälle für x_i:

- Ist $x_i = a^*$ und $x_i \in z_i$, so ist $\hat{u}_i v_i^2 w_i x_i^2 \hat{y}_i = z_1 \dots z_{i-1} a^{h+s+|x_i|} b^h c^{h+h!} z_{i+1} \dots z_k \notin L^*$, weil $a^{h+s+|x_i|} b^h c^{h+h!} \notin L$. Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

- Endet x_i mit einem Zeichen aus $z_{i+1} \dots z_k$, so ist $w_i x_i = a^{h-r-s} b^h c^{h+h!} ax$, wobei das a das erste a in z_{i+1} ist. Dann gilt aber

$$\begin{array}{rcl} \hat{u}_i v_i^2 w_i x_i^2 \hat{y}_i & = & \hat{u}_i v_i v_i w_i x_i x_i \hat{y}_i \\ & = & z_1 \dots z_{i-1} a^r a^s a^s a^{h-r-s} b^h c^{h+h!} axx_i \hat{y}_i \\ & = & z_1 \dots z_{i-1} a^{h+s} b^h c^{h+h!} axx_i \hat{y}_i \notin L^*, \end{array}$$

wegen dem Teilwort $a^{h+s}b^hc^{h+h!}$

Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Damit gilt: $x_i \in z_i$, $x_i \neq \varepsilon$ und $x_i \notin a^+$.
Ist $x_i \in a^+b^+$ oder $x_i \in b^+c^+$, so ist $x_ix_i \in a^+b^+a^+b^+$ oder $x_ix_i \in b^+c^+b^+c^+$. Damit ist $\hat{u}_iv_i^2w_ix_i^2\hat{y}_i \notin L^*$. Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Ist $x_i \in a^+b^+c^+$, so ist $w_i = a^t$, $x_i = a^{h-r-s-t}b^hc^\lambda$ und $\hat{y}_i = c^{h+h!-\lambda}z_{i+1}\dots z_k$, wobei $1 \leq \lambda \leq h+h!$. Hier gilt

$$\begin{aligned}
\hat{u}_iv_i^2w_ix_i^2\hat{y}_i &= \hat{u}_iv_iv_iw_ix_ix_i\hat{y}_i \\
&= z_1\dots z_{i-1}a^ra^sa^sa^ta^{h-r-s}b^hc^\lambda c^{h+h!-\lambda}z_{i+1}\dots z_k \\
&= z_1\dots z_{i-1}a^{h+s}b^hc^\lambda a^{h-r-s-t}b^hc^{h+h!}z_{i+1}\dots z_k \notin L^*,
\end{aligned}$$

weil $a^{h-r-s-t}b^hc^{h+h!} \notin L$ (beachte $a^{h+s}b^hc^\lambda \in L$ für $h+s=\lambda$). Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Bisher haben wir gezeigt: ($x_i \in b^+$ oder $x_i \in c^+$) und $x_i \in z_i$.

Ist ($x_i \in c^+$) und ($x_i \in z_i$), so ist $\hat{u}_iv_i^2w_ix_i^2\hat{y}_i = z_1\dots z_{i-1}a^{h+s}b^hc^{h+h!+|x_i|}z_{i+1}\dots z_k \notin L^*$. Daraus folgt ($x_i \in b^h$) und ($x_i \in z_i$). Da außerdem gelten muss, dass das Wort $\hat{u}_iv_i^2w_ix_i^2\hat{y}_i$ in L^* sein muss, folgt $|v_i| = |x_i| = s$.
Damit erhalten wir.

$$\begin{array}{ll}
\hat{u}_i = z_1\dots z_{i-1}u_i & u_i = a^{r_i} \textit{ und } 1 \leq r_i \leq h-2, \\
v_i = a^{s_i} & 1 \leq s_i \leq h-2, \\
w_i = a^{h-r_i-s_i}b^{h-s_i-t_i} & 0 \leq t_i \leq h-1, \\
x_i = b^{s_i}, & \\
\hat{y}_i = y_iz_{i+1}\dots z_k & y_i = b^{t_i}c^{h+h!}.
\end{array}$$

q.e.d.

Wir setzen nun den Beweis von Satz 4.2.1 Seite 36 auf fort. Wegen $A_i \stackrel{+}{\Longrightarrow} v_iA_ix_i$ und $v_iw_ix_i \in z_i$ hat der Ableitungsbaum B(z) einen A_i–Pumpbaum, dessen Front $v_iw_ix_i$ vollständig in z_i enthalten ist. Diese Tatsache gilt für jedes $i \in \{1, 2, \ldots, k\}$. Daraus folgt, dass der Ableitungsbaum B(z) k nebeneinander–stehende A_1–, A_2–,. . . , A_k–Pumpbäume hat. Damit sieht der Ableitungsbaum B(z) mit der Front z wie folgt aus (siehe Abbildung 4.2 auf die Seite 41):

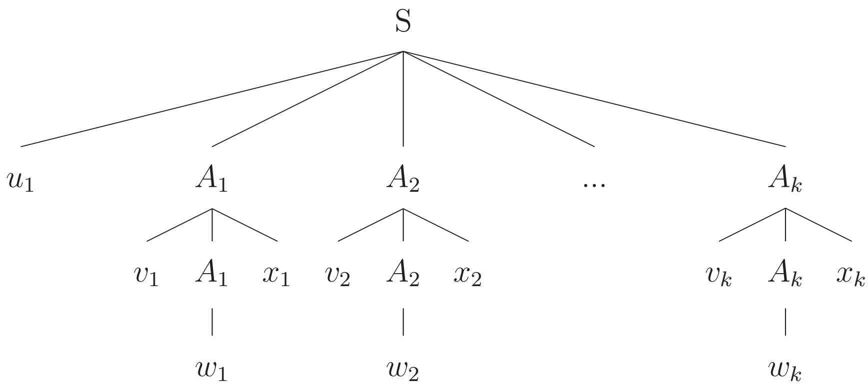

Abbildung 4.2: Ein Ableitungsbaum B(z) für ein z aus $\{a^hb^hc^{h+h!}, a^hb^{h+h!}c^h\}^k$

Aus dem Ableitungsbaum B(z) erhalten wir eine Ableitung für das Wort z, die folgendermaßen aussieht:

$$S \stackrel{+}{\Longrightarrow} u_1A_1y_1u_2A_2y_2u_3 \ldots y_{k-1}u_kA_ky_k \tag{4.2.1}$$

wobei $A_i \stackrel{+}{\Longrightarrow} v_iA_ix_i$, und $A_i \stackrel{+}{\Longrightarrow} w_i$ für alle i aus $\{1, 2, \ldots, k\}$.

Pumpen wir nun jeden A_i–Pumpbaum des Baumes B(z) $q_i := \frac{h!}{s_i} + 1$ mal, so erhalten wir einen Ableitungsbaum T(z) mit der Front $(a^{h+h!}b^{h+h!}c^{h+h!})^k$ (siehe Abbildung 4.3 auf die Seite 42).

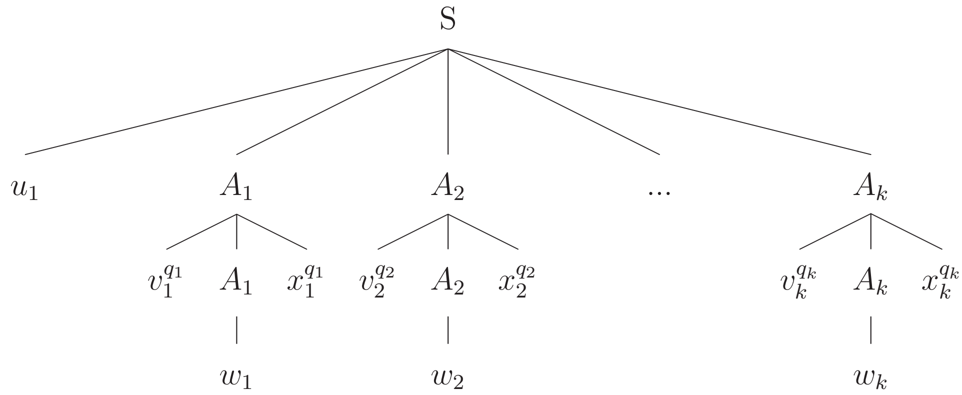

Abbildung 4.3: Jeden A_i–Pumpbaum von B(z) pumpen ergibt einen Ableitungsbaum T(z) für das Wort $(a^{h+h!}b^{h+h!}c^{h+h!})^k$

Aus dem Ableitungsbaum T(z) für das Wort $(a^{h+h!}b^{h+h!}c^{h+h!})^k$ erhält man folgende Ableitung:

$$
\begin{array}{lcl}
S & \overset{+}{\Longrightarrow} & u_1 v_1^{q_1} A_1 x_1^{q_1} y_1 u_2 v_2^{q_2} A_2 x_2^{q_2} y_2 u_3 \ldots y_{k-1} u_k v_k^{q_k} A_k x_k^{q_k} y_k \\
 & \overset{+}{\Longrightarrow} & (a^{h+h!}b^{h+h!}c^{h+h!})^k \in L^* \\
A_i & \overset{+}{\Longrightarrow} & v_i A_i x_i \\
A_i & \overset{+}{\Longrightarrow} & w_i
\end{array}
$$

$$\forall i \in \{1, 2, \ldots, k\}. \tag{4.2.2}$$

Hierbei ist zu beachten, dass jeder A_i–Pumpbaum die Anzahl der Symbole in v_i und x_i von z_i gleichmäßig erhöht.

Seien nun z, $\tilde{z}$ aus $\{a^h b^h c^{h+h!}, a^h b^{h+h!} c^h\}^k$ mit $z \neq \tilde{z}$. Wir zeigen, dass die Ableitungsbäume T(z) und T($\tilde{z}$) verschieden sind. Es gilt : $z = z_1 z_2 \ldots z_k$ und $\tilde{z} = \tilde{z}_1 \tilde{z}_2 \ldots \tilde{z}_k$.

$z \neq \tilde{z} \Rightarrow$ es gibt ein $i \in \{1, \ldots, k\}$ mit $z_i \neq \tilde{z}_i$.

O. B. d. A. sei $z_i = a^h b^h c^{h+h!}$ und $\tilde{z}_i = a^h b^{h+h!} c^h$.

Der Ableitungsbaum T(z) hat gemäß (4.2.2) folgende Ableitung:

$$\begin{aligned}
S &\overset{+}{\Longrightarrow} (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}u_iv_i^{q_i}A_ix_i^{q_i}y_i(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \\
&\overset{+}{\Longrightarrow} (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}u_iv_i^{q_i}w_ix_i^{q_i}y_i(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \\
&:= (a^{h+h!}b^{h+h!}c^{h+h!})^{k} \in L^*.
\end{aligned}$$

Analog hat der Ableitungsbaum T($\tilde{z}$) folgende Ableitung:

$$\begin{aligned}
S &\overset{+}{\Longrightarrow} (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\tilde{u}_i\tilde{v}_i^{\tilde{q}_i}\tilde{A}_i\tilde{x}_i^{\tilde{q}_i}\tilde{y}_i(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \\
&\overset{+}{\Longrightarrow} (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\tilde{u}_i\tilde{v}_i^{\tilde{q}_i}\tilde{w}_i\tilde{x}_i^{\tilde{q}_i}\tilde{y}_i(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \\
&:= (a^{h+h!}b^{h+h!}c^{h+h!})^{k} \in L^*.
\end{aligned}$$

Nehmen wir an, dass die Ableitungsbäume T(z) und T($\tilde{z}$) für das Wort $(a^{h+h!}b^{h+h!}c^{h+h!})^k$ gleich wären, also gleich einem Ableitungsbaum T(z,$\tilde{z}$). Dieser Ableitungsbaum T(z,$\tilde{z}$) mit der Front $(a^{h+h!}b^{h+h!}c^{h+h!})^k$ müsste sowohl einen A_i– als auch einen $\tilde{A}_i$–Pumpbaum haben. Damit können wir folgende Fälle unterscheiden:

1. Der Knoten A_i kommt links vom Knoten $\tilde{A}_i$ im T(z,$\tilde{z}$) vor (siehe Abbildung 4.4 auf die Seite 44)

2. Der Knoten A_i kommt rechts vom Knoten $\tilde{A}_i$ im Baum T(z,$\tilde{z}$) vor.

3. Der Knoten A_i ist ein Nachfolger des Knotens $\tilde{A}_i$ (siehe Abbildung 4.5 auf die Seite 45)

4. Der Knoten $\tilde{A}_i$ ist ein Nachfolger des Knotens A_i

Der 1. und 2. Fall (bzw. 3. und 4.) sind ähnlich, deshalb behandeln wir nur den 1. sowie den 3. Fall.

Für den 1. Fall sieht der Baum T(z,$\tilde{z}$) wie folgt aus (siehe Abbildung 4.4 auf die Seite 44):

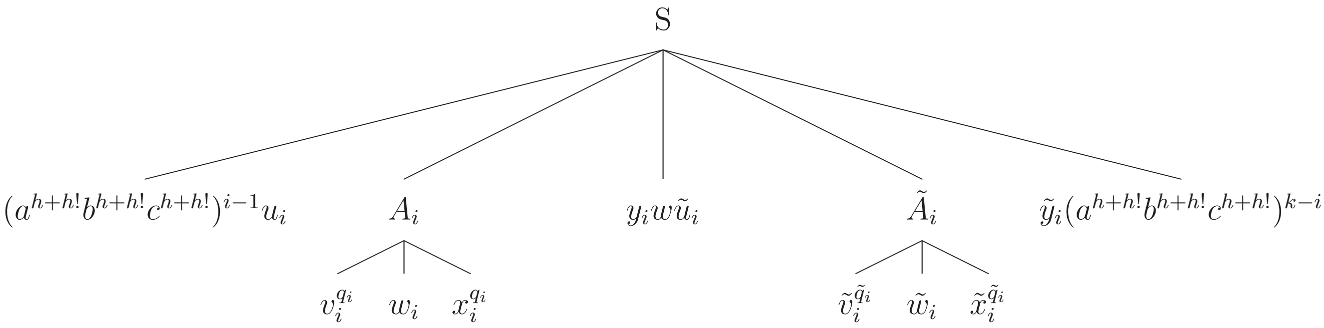

Abbildung 4.4: Der Knoten A_i kommt links vom Knoten $\tilde{A}_i$ im $T(z, \tilde{z})$ vor

Hier hätte die Front des Baumes T(z,$\tilde{z}$) mindestens (k+1) Teilwörter der Form $a^{h+h!}b^{h+h!}c^{h+h!}$, im Widerspruch zur Tatsache, dass die Front gleich dem Wort $(a^{h+h!}b^{h+h!}c^{h+h!})^k$ ist.

Für den 3. Fall sieht der Baum T(z,$\tilde{z}$) wie folgt aus (siehe Abbildung 4.5 auf die Seite 45):

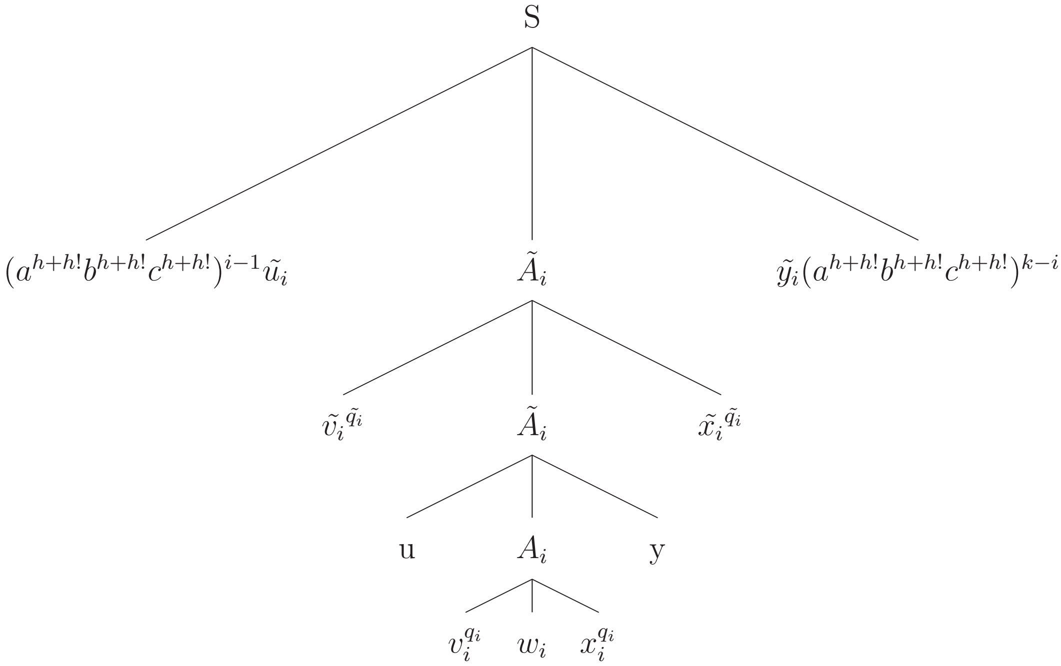

Abbildung 4.5: Der Knoten A_i ist ein Nachfolger des Knotens $\tilde{A}_i$

Hier erhalten wir:

$$\begin{aligned}
S &\overset{+}{\Longrightarrow} (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\tilde{u}_i\tilde{v}_i^{\tilde{q}_i}\tilde{A}_i\tilde{x}_i^{\tilde{q}_i}\tilde{y}_i(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i}\\
&\overset{+}{\Longrightarrow} (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\tilde{u}_i\tilde{v}_i^{\tilde{q}_i}uA_iy\tilde{x}_i^{\tilde{q}_i}\tilde{y}_i(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i}\\
&\overset{+}{\Longrightarrow} (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\tilde{u}_i\tilde{v}_i^{\tilde{q}_i}uv_i^{q_i}A_ix_i^{q_i}y\tilde{x}_i^{\tilde{q}_i}\tilde{y}_i(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i}\\
&\overset{+}{\Longrightarrow} (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\underbrace{\tilde{u}_i\tilde{v}_i^{\tilde{q}_i}uv_i^{q_i}w_ix_i^{q_i}y\tilde{x}_i^{\tilde{q}_i}\tilde{y}_i}_{t_1}(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i}\\
&= (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}t_1(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \in L^*.
\end{aligned}$$

Da die Front des Baumes T(z,$\tilde{z}$) dem Wort $(a^{h+h!}b^{h+h!}c^{h+h!})^k$ entspricht, folgt $t_1 = a^{h+h!}b^{h+h!}c^{h+h!}$.

Im Baum T(z,$\tilde{z}$) können wir aber auch den A_i–Pumpbaum $q_i + 1$ mal und den $\tilde{A}_i$–Pumpbaum $\tilde{q}_i + 1$ mal pumpen. Dann gilt aber:

$$\begin{array}{rcl} S & \stackrel{+}{\Longrightarrow} & (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\tilde{u_i}\tilde{v_i}^{\tilde{q_i}+1}\tilde{A}_i\tilde{x_i}^{\tilde{q_i}+1}\tilde{y_i}(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \\ & \stackrel{+}{\Longrightarrow} & (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\tilde{u_i}\tilde{v_i}^{\tilde{q_i}+1}uA_iy\tilde{x_i}^{\tilde{q_i}+1}\tilde{y_i}(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \\ & \stackrel{+}{\Longrightarrow} & (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\tilde{u_i}\tilde{v_i}^{\tilde{q_i}+1}uv_i^{q_i+1}A_ix_i^{q_i+1}y\tilde{x_i}^{\tilde{q_i}+1}\tilde{y_i}(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \\ & \stackrel{+}{\Longrightarrow} & (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}\underbrace{\tilde{u_i}\tilde{v_i}^{\tilde{q_i}+1}uv_i^{q_i+1}w_ix_i^{q_i+1}y\tilde{x_i}^{\tilde{q_i}+1}\tilde{y_i}}_{t_2}(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \\ & = & (a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}t_2(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \in L^*. \end{array}$$

wobei

$$\begin{array}{rcl} \#_a(t_2) & = & \#_a(t_1) + |\tilde{v_i}| + |v_i| = h + h! + |\tilde{v_i}| + |v_i| \\ \#_b(t_2) & = & \#_a(t_1) + |x_i| = h + h! + |v_i| \\ \#_c(t_2) & = & \#_a(t_1) + |\tilde{x_i}| = h + h! + |\tilde{v_i}| \end{array}$$

Das heißt, $\#_a(t_2) \neq \#_b(t_2)$ und $\#_a(t_2) \neq \#_c(t_2)$ Damit gilt $t_2 \notin L$ Dies steht im Widerspruch zu $(a^{h+h!}b^{h+h!}c^{h+h!})^{i-1}t_2(a^{h+h!}b^{h+h!}c^{h+h!})^{k-i} \in L^*$.

Insgesamt können wir schließen, dass T(z) und T($\tilde{z}$) verschieden sind. Da es insgesamt 2^k verschiedene Wörter $\{a^hb^hc^{h+h!}, a^hb^{h+h!}c^h\}^k$ gibt, hat das Wort $(a^{h+h!}b^{h+h!}c^{h+h!})^k$ mindestens 2^k verschiedene Ableitungsbäume in jeder Grammatik G für L^*.

Für jede Grammatik G von L^* und $k \in \mathbb{N}$ haben wir gezeigt :

$$da_G((a^{h+h!}b^{h+h!}c^{h+h!})^k) \geq 2^k. \tag{4.2.3}$$

Aus $n := |(a^{h+h!}b^{h+h!}c^{h+h!})^k| = 3k(h + h!)$ folgt .

$$k = \frac{n}{3(h + h!)} \tag{4.2.4}$$

Kombiniert man (4.2.3) und (4.2.4), kann man folgern:

Für jede Grammatik G von L^* und $k \in \mathbb{N}$ gilt :

$$da_G(n) \geq 2^{\frac{n}{3(h+h!)}} \tag{4.2.5}$$

d.h.

$$da_G(n) \geq 2^{cn}, \ wobei \ c := \frac{n}{3(h+h!)} \tag{4.2.6}$$

Somit gilt, dass jede Grammatik für L^* $2^{\Omega(n)}$–deutig ist. Da es keine mehrdeutigen Sprachen geben kann, deren Mehrdeutigkeitsgrade größer als $2^{\Theta(n)}$ sind (siehe Bemerkung 2.7.1 auf die Seite 22), können wir daraus schließen, dass L^* exponentiell mehrdeutig ist.

q. e. d.

Bemerkung 4.2.1 HARRISON *betrachtete in [2, Seiten 241–242] die Sprache* $(L_0 \cup L_1)^n$*, wobei* $L_0 := \{a^i b^i c^j | i,j \geq 1\}$ *und* $L_1 := \{a^i b^j c^j | i,j \geq 1\}$ *ist. Zuerst zeigte er, dass für jedes* $w_{i+1} \in (L_0 \cup L_1)^{k_i}$*,* $i = 1, 2, k_1, k_2 \geq 0$ *das Wort* $w_2 a^{3p} b^{3p} c^{3p} w_3$ *mindestens 2 Ableitungen in jeder Grammatik für* $(L_0 \cup L_1)^n$ *hat, wobei* $p = p_0!$ *und* p_0 *die Konstante aus* OGDEN*'s Lemma für die Grammatik* $\widetilde{G}$ *für* $(L_0 \cup L_1)^n$ *ist. Ferner behauptete er, dass das Wort* $(a^{3p} b^{3p} c^{3p})^n$ *mindestens* 2^n *Ableitungen in jeder Grammatik für* $(L_0 \cup L_1)^n$ *hat. Dies begründete er damit, dass jedes der n Teilwörter* $a^{3p} b^{3p} c^{3p}$ *sich auf mindestens zwei verschiedene Weisen ableiten lässt. Diese Argumentation lässt sich aber ohne weiteres auch auf die Sprache* $L_0^n \cup L_1^n$ *anwenden. Man würde damit zeigen, dass* $L_0^n \cup L_1^n$ 2^n*–deutig ist, was nicht korrekt ist, denn* $L_0^n \cup L_1^n$ *ist 2–deutig. Harrison hat versäumt zu zeigen, dass die zwei Ableitungen für* $w_2 a^{3p} b^{3p} c^{3p} w_3$ *sich von den zwei Ableitungen* $\widetilde{w_2} a^{3p} b^{3p} c^{3p} \widetilde{w_3}$ *unterscheiden, falls* $w_2 \neq \widetilde{w_2}$ *oder* $w_3 \neq \widetilde{w_3}$ *. Damit hat Harrison nicht bewiesen, dass eine Ableitung für* $(a^{3p} b^{3p} c^{3p})^n$ *unter der Betrachtung eines Teilwortes* $a^{3p} b^{3p} c^{3p}$ *unabhängig von einer Ableitung unter Betrachtung eines anderen Teilwortes* $a^{3p} b^{3p} c^{3p}$ *ist. Der Beweis in [2] ist also unvollständig. In dieser Arbeit betrachteten wir dagegen die* 2^k *Ableitungsbäume der* 2^k *Wörter der Form* $\{a^h b^h c^{h+h!}, a^h b^{h+h!} c^h\}^k$ *in jeder Grammatik G für die Sprache* L^**,* *wobei h die Konstante aus* OGDEN*'s Lemma für die Grammatik G für die Sprache L^* ist. Wir haben bewiesen, dass diese 2^k Ableitungsbäume sich zu 2^k verschiedenen Ableitungsbäumen für das Wort $a^{h+h!}b^{h+h!}c^{h+h!}$ aufpumpen lassen. Wäre der Beweis von Harrison dafür, dass $(L_0 \cup L_1)^n$ 2^n–deutig ist, vollständig, so könnten wir mit viel weniger Aufwand zeigen, dass L^* exponentiell mehrdeutig ist.*

4.3 Die Sprache L^* ist nicht 2^{cn}–deutig

Mit Hilfe von OGDEN's Lemma (Satz 2.4.1auf Seite 16) haben wir gezeigt, dass der Mehrdeutigkeitsgrad jeder Grammatik G für die Sprache L^* **mindestens** 2^{cn} ist, wobei n die Länge eines Wortes aus L^* und $c = \frac{1}{3(h+h!)}$ ist (siehe Gleichung (4.2.6)). Wir bemerken, dass c von der Konstante h aus OGDEN's Lemma (also von der Grammatik) abhängt. Es stellt sich aber die Frage, ob wir eine (exponentielle) untere Schranke 2^{cn} für die Mehrdeutigkeit der Sprache L^* finden, die von keiner Grammatik abhängt. Anders ausgedrückt: Können wir ein c aus $\mathbb{R}_+$ finden, so dass L^* 2^{cn}–deutig ist? Wir zeigen, dass dies nicht möglich ist, indem wir für jedes gegebene c aus $\mathbb{R}_+$ eine Grammatik G für L^* entwerfen, so dass jedes Wort der Länge n aus L^* höchstens 2^{cn} Ableitungsbäume in G hat. Wir beweisen den Satz

Satz 4.3.1 *Die Sprache L^* ist zwar exponentiell mehrdeutig, aber es gibt kein c aus $\mathbb{R}_+$, so dass L^* 2^{cn}–deutig ist.*

Beweis Die Beweisidee ist denkbar einfach. Zunächst entwerfen wir eine Grammatik $\widetilde{G}$ für L, so dass für jedes w aus L gilt:

$$da_{\widetilde{G}}(w) \begin{cases} = 1 & , falls\ |w| \leq k \\ \leq 2 & , sonst \end{cases} \tag{4.3.1}$$

wobei k eine Konstante aus $\mathbb{N}$ ist, die wir später geeignet festlegen werden.

Danach konstruieren wir mit Hilfe der Grammatik $\widetilde{G}$ eine Grammatik G für L^*, die uns zweierlei garantiert:

- Zum einen können die Wörter der Sprache L (und somit L^*), deren Länge kleiner als k ist, eindeutig erzeugt werden. Das leere Wort ε ist zwar in L^*, aber nicht in L, lässt sich aber eindeutig in G erzeugen.
- Zum anderen bestehen die Wörter aus L^*, deren Länge n größer als k ist, aus höchstens $\frac{n}{k}$ Wörtern aus L und haben deshalb höchstens $2^{\frac{n}{k}}$ Ableitungsbäume in G , weil jedes Wort aus L höchstens 2 Ableitungsbäume in $\widetilde{G}$ hat.

Wir führen nun den Beweis formal aus:

Sei $\widetilde{G} := (\{\widetilde{S}, A, B, D, E\}, \{a, b, c\}, P, \widetilde{S})$. Die Regeln von P sind:

(1) $\widetilde{S} \rightarrow w|\varepsilon$, wobei $w \in L$ und $|w| \leq k$

(2) $\widetilde{S} \rightarrow a^i b^i c^{k-2i} A$, wobei $1 \leq i \leq \frac{k}{2}$

$A \rightarrow c|cA$

(3) $\widetilde{S} \rightarrow a^i b^{k-2i} B c^i$, wobei $1 \leq i \leq \frac{k}{2}$

$B \rightarrow b|bB$

(4) $\widetilde{S} \rightarrow a^j D b^j A$, wobei $j = \left\lfloor \frac{k}{2} \right\rfloor$

$D \rightarrow aDb|ab$

(5) $\widetilde{S} \to a^j E c^j$, wobei $j = \left\lfloor \frac{k}{2} \right\rfloor$

$E \to aEc|aBc$

Wir machen folgende Beobachtungen:

- Produktionstyp **(1)** leitet alle Wörter w aus L ab, mit $|w| \leq k$.
- Produktionstyp **(2)** leitet alle Wörter w aus L ab, mit $|w| > k$ und $\#_a(w) = \#_b(w) \leq \frac{k}{2}$.
- Produktionstyp **(3)** leitet alle Wörter w aus L ab, mit $|w| > k$ und $\#_a(w) = \#_c(w) \leq \frac{k}{2}$.
- Produktionstyp **(4)** leitet alle Wörter w aus L ab, mit $|w| > k$ und $\#_a(w) = \#_b(w) > \frac{k}{2}$.
- Produktionstyp **(5)** leitet alle Wörter w aus L ab, mit $|w| > k$ und $\#_a(w) = \#_c(w) > \frac{k}{2}$.

Aus diesen Beobachtungen folgt unmittelbar:

- $L = L(\widetilde{G})$.
- $$da_{\widetilde{G}}(w) = \begin{cases} 2 & , falls\ |w| > k\ und\ \#_a(w) = \#_b(w) = \#_c(w) \\ 1 & , sonst \end{cases}$$

womit die gestellte Bedingung (4.3.1) erfüllt ist.

Die Grammatik $G := (\{S\} \dot\cup \{\widetilde{S}, A, B, D, E\}, \{a, b, c\}, P \dot\cup \{S \to S\widetilde{S}|\varepsilon\}, S)$ erzeugt die Sprache L^*.

Sei nun $w \neq \varepsilon$ ein Wort der Länge n aus L^*. Es gilt $w = w_1 w_2 \ldots w_m$, wobei $w_i \in L$, $\forall i \in \{1, 2, \ldots, m\}$.

$$da_G(w) = \prod_{i=1}^{m} da_{\widetilde{G}}(w_i) \leq 2^{Anzahl\ der\ w_i's\ mit\ |w_i|>k} \tag{4.3.2}$$

Da es in w höchstens $\frac{n}{k}$ w_i's mit $|w_i| > k$ gibt, folgt aus (4.3.2), dass

$$da_G(w) \leq 2^{\frac{n}{k}} \tag{4.3.3}$$

Wir wählen nun k so, dass $\frac{1}{k} \leq c$ gilt, das heißt, $k \geq \frac{1}{c}$. Wir setzen

$$k := \left\lceil \frac{1}{c} \right\rceil \tag{4.3.4}$$

Dann erhalten wir

$$da_G(n) \leq 2^{cn} \tag{4.3.5}$$

q. e. d.

Es ist interessant zu beobachten, dass wir für alle $c \geq 1$ k=1 (wegen 4.3.4) wählen können. Das heißt, dass wir für alle $c \geq$ 1eine einzige Grammatik $\widetilde{G}$ entwerfen können. Die Produktionsregeln dieser Grammatik $\widetilde{G}$ sehen folgendermaßen aus : siehe Tabelle 4.1 auf die Seite 51.

Regel 1	$\widetilde{S} \rightarrow aDbA$
Regel 2	$\widetilde{S} \rightarrow aEc$
Regel 3	$D \rightarrow aDb\|\varepsilon$
Regel 4	$E \rightarrow aEc\|B$
Regel 5	$A \rightarrow cA\|c$
Regel 6	$B \rightarrow b\|bB$

Tabelle 4.1: Produktionsregeln der Grammatik $\widetilde{G}$

Wir möchten die Konstruktion der Grammatik $\widetilde{G}$ verdeutlichen, indem wir die Produktionen für ein gerades k (k=6) und für ein ungerades k (k=7) explizit angeben.

Beispiel 4.3.1 (k=6)

1. $\widetilde{S} \rightarrow abc|abcc|abccc|abcccc|aabbc|aabbcc|$
   $aaabbb|abbc|abbbc|abbbbc|aabcc|aaaccc$

2. $\widetilde{S} \rightarrow abccccA|aabbccA|aaabbbA$
   $A \rightarrow c|cA$

3. $\widetilde{S} \rightarrow abbbbBc|aabbBcc|aaaBccc$
   $B \rightarrow b|bB$

4. $\widetilde{S} \rightarrow aaaDbbbA$
   $D \rightarrow aDb|ab$

5. $\widetilde{S} \rightarrow aaaEccc$
   $E \rightarrow aEc|aBc$

Beispiel 4.3.2 (k=7)

1. $\widetilde{S} \rightarrow abc|abcc|abccc|abcccc|abccccc|aabbc|aabbcc|aabbccc|aaabbbc|$
   $abbc|abbbc|abbbbc|abbbbbc|aabcc|aabbbcc|aaabccc$

2. $\widetilde{S} \rightarrow abcccccA|aabbcccA|aaabbbcA$
   $A \rightarrow c|cA$

3. $\widetilde{S} \rightarrow abbbbbBc|aabbbBcc|aaabBccc$
   $B \rightarrow b|bB$

4. $\widetilde{S} \rightarrow aaaDbbbA$
   $D \rightarrow aDb|ab$

5. $\widetilde{S} \rightarrow aaaEccc$

   $E \rightarrow aEc|aBc$

Bemerkung 4.3.1 *Die Beweisidee von Satz 4.3.1 auf Seite 48, dass man für jedes beliebig gegebene c eine Grammatik für die Sprache L^* konstruieren kann, deren Mehrdeutigkeitsgrad unter 2^{cn} liegt, funktioniert nicht nur für die benutzte Sprache L^*, sondern für jede exponentiell mehrdeutige Sprache L der Form $K^* \cup E$, wobei K eine k–deutige und E eine endliche Sprache ist. Das heißt für jede exponentiell mehrdeutige Sprache L der Form $K^* \cup E$, wobei K eine k–deutige und E eine endliche Sprache ist, gibt es kein festes c aus $\mathbb{R}_+$, so dass L 2^{cn}–deutig ist. Es stellt sich folgendes offenes Problem.*

Offenes Problem Gibt es überhaupt eine exponentiell mehrdeutige Sprache, die für irgendein festes c aus $\mathbb{R}_+$ 2^{cn}–deutig ist?

Kapitel 5

Polynomiell mehrdeutige Grammatiken und Sprachen

5.1 Motivation

In [10] wurde gezeigt, dass es zu jeder Grammatik, die nicht exponentiell mehrdeutig ist, eine polynomielle Schranke für deren Mehrdeutigkeit gibt. Dies hat zur Folge, dass es keine Grammatiken (und somit keine Sprachen) gibt, deren Mehrdeutigkeit stärker als polynomiell, aber schwächer als exponentiell wächst (wie z. B. $\Theta(2^{\sqrt{n}})$). Bisher konnten wir die Existenz von konstant sowie exponentiell mehrdeutigen Grammatiken und Sprachen nachweisen. Wie sieht es mit polynomiell mehrdeutigen Grammatiken und Sprachen aus? Gibt es überhaupt polynomiell mehrdeutige Grammatiken und Sprachen?

Die Existenz von polynomiell mehrdeutigen Grammatiken ist leicht zu zeigen. Wich [10] hat für $k \geq 1$ die $\Theta(n^k)$–deutige Grammatik G mit den folgenden Produktionen angegeben:

(1) $S \to aS$

(2) $S \to A_1$

(3) $A_i \to aA_i$ für $1 \le i \le k$

(4) $A_i \to A_{i+1}$ für $1 \le i \le k-1$

(5) $A_k \to \varepsilon$

Gibt es überhaupt polynomiell mehrdeutige Sprachen? In diesem Kapitel zeigen wir, dass es tatsächlich polynomiell mehrdeutige Sprachen gibt. Zuerst weisen wir die Existenz von linear mehrdeutigen Sprachen nach. Danach geben wir für jedes $k \in \mathbb{N}$ eine polynomiell vom Grad k mehrdeutige Sprache an.

5.2 linear mehrdeutige Sprachen

Satz 5.2.1 *Es gibt linear mehrdeutige Sprachen.*

Beweis Wir betrachten die Sprache $L := \{a^m b^{m_1} c b^{m_2} c \dots b^{m_p} c \,\big|\, p \in \mathbb{N}; m, m_1, m_2, \dots, m_p \in \mathbb{N}; \exists i \in \{1, 2, \dots, p\} \; mit \; m = m_i\}$. Zuerst bemerken wir, dass die Sprache L sozusagen p–deutig (wie die k–deutige Sprache $\widetilde{L}_k$ aus der Bemerkung 3.3.1 vom Kapitel 3 auf die Seite 34) ist, aber p ist nicht konstant wie k im Kapitel 3, sondern eine beliebige Variable aus $\mathbb{N}$. Sei G=(N, Σ, P, S) eine CFG für L. Sei h die Konstante aus Ogden's Lemma (Satz 2.4.1 auf Seite 16) für die Grammatik G. Für jedes $p \in \mathbb{N}$ betrachten wir die Wörter z_i aus L der Form $z_i := a^h (b^{h+h!} c)^{i-1} b^h c (b^{h+h!} c)^{p-i}$, wobei $i \in \{1, 2, \dots, p\}$ und markieren alle a's. Wir zeigen zuerst folgendes:

Lemma 5.2.1 *Sei $z_i = u_i v_i w_i x_i y_i$ die nach* Ogden*'s Lemma (Satz 2.4.1 auf Seite 16) existierende Zerlegung. Es gilt:*

$$
\begin{aligned}
u_i &= a^{r_i} && 1 \le r_i \le h,\\
v_i &= a^{s_i} && 1 \le s_i \le h,\\
w_i &= a^{h-s_i-r_i}(b^{h+h!}c)^{i-1}b^{t_i} && 0 \le t_i \le h,\\
x_i &= b_i^{s_i} && 1 \le s_i \le h,\\
y_i &= b^{h-s_i-t_i}c(b^{h+h!}c)^{p-i}.
\end{aligned}
$$

Beweis (Lemma 5.2.1)

Aus OGDEN's Lemma folgt: **Entweder** treten sowohl in u_i als auch in v_i markierte Positionen von z_i auf **oder** es treten sowohl in x_i als auch in y_i markierte Positionen von z_i auf. Wenn sowohl x_i als auch y_i markierte Positionen (also a's) haben, dann besteht x_i nur aus a's, denn w_i hat gemäß OGDEN's Lemma mindestens eine markierte Position (also ein a) und x_i kommt zwischen w_i und y_i in z_i vor. Damit gilt $x_i \in a^+$. w_i hat gemäß OGDEN's Lemma mindestens eine markierte Position. Da v_i in z_i links von w_i vorkommt, gilt $v_i \in a^*$.

Somit erhalten wir $v_ix_i = a^t$ mit $1 \le t \le h-2$.

Damit ist $\bar{z} := u_iv_i^2w_ix_i^2y_i = a^{h+t}(b^{h+h!}c)^{i-1}b^hc(b^{h+h!}c)^{p-i} \notin L$. Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Es muss also gelten, dass sowohl u_i als auch v_i mindestens eine markierte Position (also mindestens ein a) haben.

Da w_i gemäß OGDEN's Lemma mindestens eine markierte Position hat, besteht v_i nur aus a's. Damit gilt $v_i = a^s$, wobei $1 \le s \le h-2$. Daraus folgt: w_i beginnt mit einem a und $u_i = a^r$ mit $1 \le r \le h-2$.

Ist $x_i = \varepsilon$ oder $x_i \in a^+$,

so ist $\bar{z} := u_iv_i^2w_ix_i^2y_i = a^{h+s+|x_i|}(b^{h+h!}c)^{i-1}b^hc(b^{h+h!}c)^{p-i} \notin L$.

Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Hat x_i ein a (sowie andere Zeichen außer den a's),

so ist $\bar{z} := u_i v_i^2 w_i x_i^2 y_i$ von der Form
$u_i v_i^2 w_i a^+\{b,c\}^+ a^+\{b,c\}^+ y_i$ und somit nicht in L, was im Widerspruch zum OGDEN's Lemma steht.

Damit gilt: x_i besteht also nur aus b's und c's. Wir unterscheiden zunächst zwei Fälle: x_i hat mindestens ein c, x_i hat kein c.

Hat x_i mindestens ein c, so gilt:

$u_i = a^r$

$v_i = a^s$

$w_i = a^{h-r-s} w b^{f-t}$

$x_i = b^t c x b^\lambda$

$y_i = b^{g-\lambda} y$

mit

$w, x, y \in (b^{h+h!}c)^* b^h c (b^{h+h!}c)^* \cup (b^{h+h!}c)^*$;

$f, g \in \{h, h+h!\}$;

$0 \leq t \leq h+h!$ und $0 \leq \lambda \leq h+h!$

- Falls $t+\lambda = 0$, so ist $x_i = cx$. Aber dann hat $\bar{z} := u_i v_i^2 w_i x_i^2 y_i = u_i v_i^2 w_i cxcx y_i$ ein Teilwort $cc \in cxc$ und wäre somit nicht in L, was ein Widerspruch zu OGDEN's Lemma wäre.

- Falls $t+\lambda \neq h+s$, so ist
  $\bar{z} := u_i v_i^2 w_i x_i^2 y_i = a^r a^s a^s a^{h-r-s} w b^{f-t} b^t c x b^\lambda b^t c x b^\lambda b^{g-\lambda} y = a^{h+s} w b^f c x b^{\lambda+t} c x b^g y \notin L.$

  Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

- Falls $t+\lambda = h+s$, so ist
  $\hat{z} := u_i v_i^3 w_i x_i^3 y_i = a^r a^s a^s a^s a^{h-r-s} w b^{f-t} b^t c x b^\lambda b^t c x b^\lambda b^t c x b^\lambda b^{g-\lambda} y = a^{h+2s} w b^f c x b^{\lambda+t} c x b^{\lambda+t} c x b^g y \notin L.$

Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Damit gilt: x_i hat kein c.

Es gibt genau zwei Möglichkeiten für x_i: $x_i \in b^{h+h}!$ oder $x_i \in b^h \in cb^hc$.

Ist $x_i \in b^{h+h}!$, erhalten wir:

$u_i = a^r$

$v_i = a^s$

$w_i = a^{h-r-s}wb^{h+h!-t-\lambda}$

$x_i = b^t$

$y_i = b^\lambda y$

mit

$w, y \in (b^{h+h!}c)^*b^hc(b^{h+h!}c)^* \cup (b^{h+h!}c)^*$,

$1 \leq t \leq h + h!$ und $0 \leq \lambda \leq h + h!$.

Aber dann ist

$\bar{z} := u_iv_i^2w_ix_i^2y_i = a^ra^sa^sa^{h-r-s}wb^{h+h!-t-\lambda}b^tb^tb^\lambda y = a^{h+s}wb^{h+h!+t}y \notin L$.

Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Es muss also gelten: $x_i \in b^h \in cb^hc$. Hier erhalten wir:

$u_i = a^r$

$v_i = a^s$

$w_i = a^{h-r-s}(b^{h+h!}c)^{i-1}b^{h-t-\lambda}$

$x_i = b^t$

$y_i = b^\lambda(b^{h+h!}c)^{p-i}$

mit

$1 \leq t \leq h$ und $0 \leq \lambda \leq h$.

Da aber außerdem gelten muss, dass $\bar{z}$:=

$u_i v_i^2 w_i x_i^2 y_i = a^r a^s a^s a^{h-r-s} (b^{h+h!}c)^{i-1} b^{h-t-\lambda} b^t b^t b^\lambda (b^{h+h!}c)^{p-i} = a^{h+s}(b^{h+h!}c)^{i-1} b^{h+t} (b^{h+h!}c)^{p-i} \in L$, folgt t=s. Damit erhalten wir:

$$
\begin{aligned}
u_i &= a^{r_i} && 1 \leq r_i \leq h, \\
v_i &= a^{s_i} && 1 \leq s_i \leq h, \\
w_i &= a^{h-s_i-r_i}(b^{h+h!}c)^{i-1} b^{t_i} && 0 \leq t_i \leq h, \\
x_i &= b_i^{s_i} && 1 \leq s_i \leq h, \\
y_i &= b^{h-s_i-t_i} c (b^{h+h!}c)^{p-i}.
\end{aligned}
$$

q. e. d.

Wir setzen nun den Beweis von Satz 5.2.1 fort.

Aus OGDEN's Lemma (Satz 2.4.1 auf Seite 16) folgt, dass es ein A_i aus N geben muss, so dass

$S \overset{+}{\Longrightarrow} u_i A_i y_i \overset{+}{\Longrightarrow} u_i v_i A_i x_i y_i \overset{+}{\Longrightarrow} u_i v_i w_i x_i y_i := z_i$

Damit hat jeder Ableitungsbaum B_i in G mit der Front z_i einen A_i–Pumpbaum, welcher die Anzahl der Symbole in v_i und x_i gleichmäßig erhöht. Der Ableitungsbaum ist in der Abbildung 5.1 auf Seite 60 dargestellt.

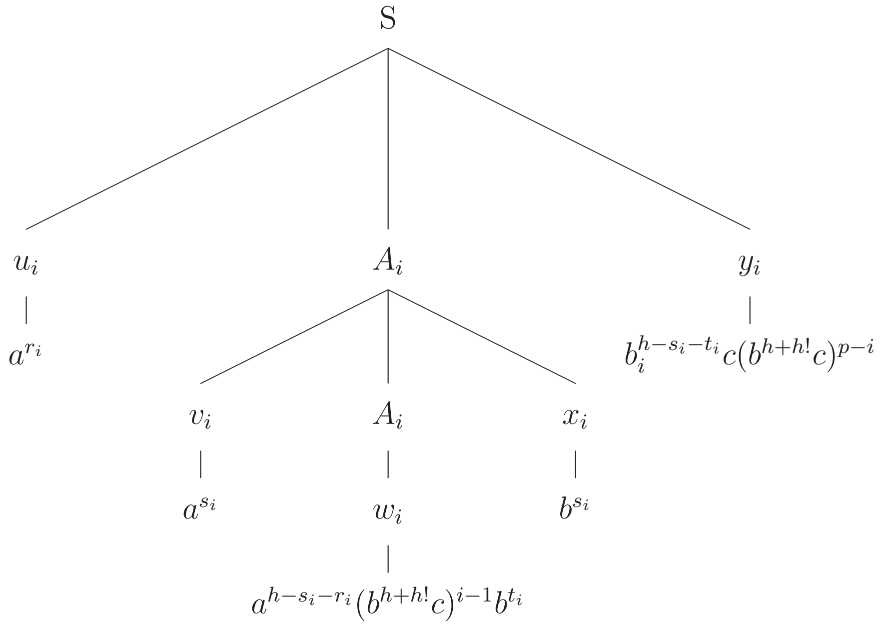

Abbildung 5.1: Ableitungsbaum B(z) für $z_i = a^h(b^{h+h!}c)^{i-1}b^h c(b^{h+h!}c)^{p-i}$

Pumpen wir den A_i–Pumpbaum des Baumes B_i $q_i := \frac{h!}{s_i} + 1$ mal, so erhalten wir einen Ableitungsbaum T_i für das Wort $z := a^{h+h!}(b^{h+h!}c)^p$. Der Ableitungsbaum T_i ist in der Abbildung 5.2 auf die Seite 61 dargestellt.

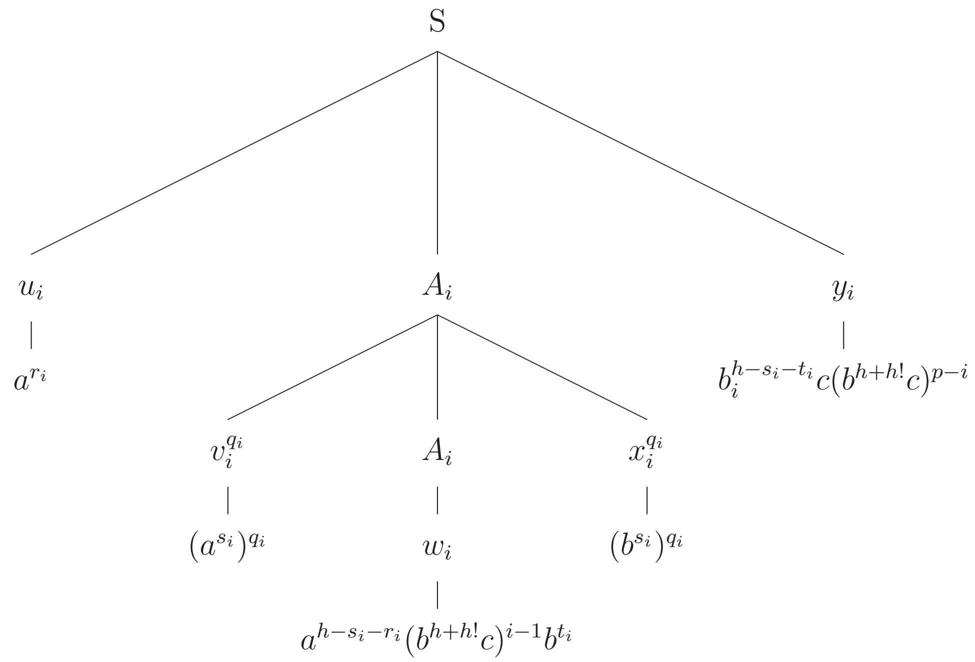

Abbildung 5.2: Ableitungsbaum T(z) für $a^{h+h!}(b^{h+h!}c)^p$

Entsprechend erhalten wir p Ableitungsbäume T_1, T_2,..., T_p. Nun zeigen wir für $i \neq j \in \{1, 2, \ldots, p\}$, dass die Ableitungsbäume T_i und T_j für das Wort z verschieden sind. Nehmen wir nun an, dass die Ableitungsbäume T_i und T_j für das Wort $z := a^{h+h!}(b^{h+h!}c)^p$ gleich wären, also gleich einem Ableitungsbaum T. Dieser Ableitungsbaum T müßte sowohl A_i– als auch A_j–Knoten haben. Wir können ausschließen, dass der Knoten A_i links vom Knoten A_j im Baum T vorkommt, sonst erhielten wir ein Wort in L von der Form $a^+\{b, c\}^+a^+\{b, c\}^+$ (siehe Abbildung 5.3 auf die Seite 62). Analog gilt es für A_i rechts von A_j. Es bleiben somit zwei Fälle zu untersuchen: Der Knoten A_j ist ein Nachfolger des Knotens A_i bzw. der Knoten A_i ist ein Nachfolger des Knotens A_j im Baum T (siehe Abbildung 5.4 auf die Seite 63). Diese beiden Fälle sind ähnlich, wir behandeln nur den letzten Fall.

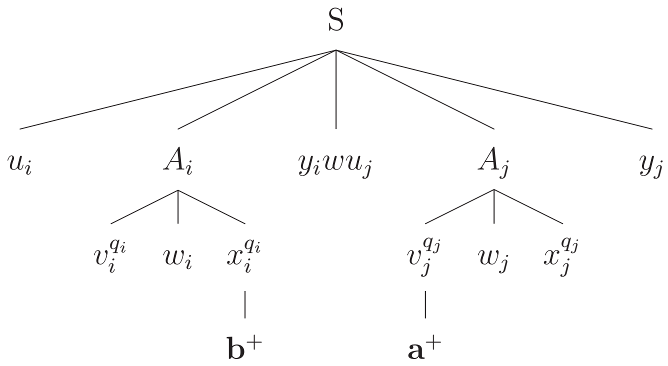

Abbildung 5.3: A_i kommt im Baum T links von A_j vor

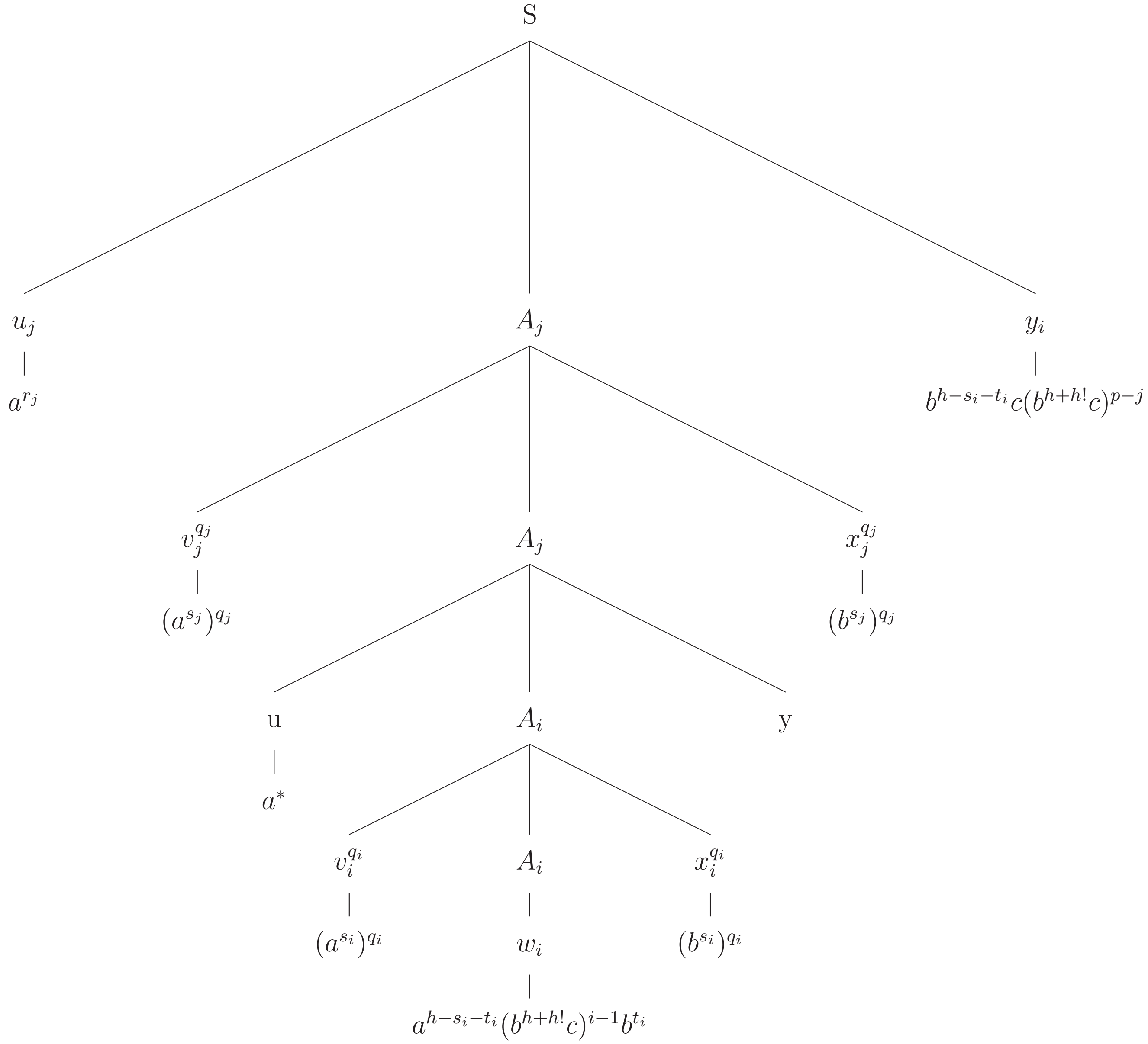

Abbildung 5.4: Ableitungsbaum T(z) für $a^{h+h!}(b^{h+h!}c)^p$ mit A_i– und A_j–Knoten

Hier erhalten wir:

$$\begin{aligned} S &\overset{+}{\Longrightarrow} u_jA_jy_j \\ &\overset{+}{\Longrightarrow} u_jv_j^{q_j}A_jx_j^{q_j}y_j \\ &\overset{+}{\Longrightarrow} u_jv_j^{q_j}uA_jyx_j^{q_j}y_j \\ &\overset{+}{\Longrightarrow} u_jv_j^{q_j}uv_i^{q_i}w_jx_i^{q_i}yx_j^{q_j}y_j \\ &:= z \in L. \end{aligned}$$

Es gilt: $\#_a(z)$ und die Anzahl der b's in jedem b–Block von z sind gleich h+h!, denn es gilt $z = a^{h+h!}(b^{h+h!}c)^p$.

Wir können aber auch den A_i–Pumpbaum $q_i + 1$ mal und den A_j–Pumpbaum $q_j + 1$ mal pumpen. Dann gilt aber:

$$\begin{aligned} S &\overset{+}{\Longrightarrow} u_jA_jy_j \\ &\overset{+}{\Longrightarrow} u_jv_j^{q_j+1}A_jx_j^{q_j+1}y_j \\ &\overset{+}{\Longrightarrow} u_jv_j^{q_j+1}uA_jyx_j^{q_j+1}y_j \\ &\overset{+}{\Longrightarrow} u_jv_j^{q_j+1}uv_i^{q_i+1}w_jx_i^{q_i+1}yx_j^{q_j+1}y_j \\ &:= \tilde{z} \in L. \end{aligned}$$

wobei $\#_a(\tilde{z}) = \#_a(z)+|v_j|+|v_i| = h+h!+|v_j|+|v_i|$, Jeder b–Block in $\tilde{z}$ aus h+h!, $h+h!+|x_i|$ oder $h+h!+|x_j|$ b's besteht. Damit ist $\#_a(\tilde{z})$ ungleich der Anzahl der b's in jedem b–Block in $\tilde{z}$, weil $|v_j| = |x_j| \geq 1$ und $|v_j| = |x_j| \geq 1$. Das ist ein Widerspruch zu $u_jv_j^{q_j+1}uv_i^{q_i+1}w_jx_i^{q_i+1}yx_j^{q_j+1}y_j := \tilde{z} \in L$.

Insgesamt haben wir gezeigt, dass für $i \neq j$ die Ableitungsbäume T_i und T_j für das gleiche Wort z verschieden sind. Somit folgt, dass das Wort $z := a^{h+h!}(b^{h+h!}c)^p$ mindestens p paarweise verschiedene Ableitungsbäume in G hat, d. h. $da_G(z) \geq p$.
Aus $n := |z| = h + h! + (h + h! + 1)p$ folgt $p = \frac{n-h-h!}{h+h!+1}$.
Somit gilt $da_G(n) \geq \frac{n-h-h!}{h+h!+1}$. Das heißt $da_G(n) = \Omega(n)$, denn h ist eine Konstante.

eine Grammatik G für L Wir entwerfen nun eine Grammatik G für L, so dass $da_G(n) = O(n)$. Dazu betrachten wir die Grammatik $G := (\{S, T, A\}, \{a, b, c\}, P, S)$. Die Regeln von P sind:

(1) $S \rightarrow aTbcA$

(2) $S \rightarrow aTbc$

(3) $T \rightarrow aTb$

(4) $T \rightarrow \varepsilon$

(5) $T \rightarrow A$

(6) $A \rightarrow bA|bcA|bc$

Zuerst bemerken wir, dass sich Dank des Regeltyps **(6)** alle Wörter der Form $(b^+c)^+$ aus A ableiten lassen. Aus S kann man nur ableiten, wenn man nach folgendem Ableitungsschema vorgeht (siehe Abbildung 5.5 auf die Seite 65):

$$S \longrightarrow \left| \begin{array}{l} \xrightarrow{1} aTbcA \xrightarrow{(m-1)*3} a^mTb^mcA \longrightarrow \left| \begin{array}{l} \xrightarrow{4} a^mb^mcA \xrightarrow{*6} a^mb^mc(b^+c)^+ \\ \xrightarrow{5} a^mAb^mcA \xrightarrow{*6} a^m(b^+c)^+b^mc(b^+c)^+ \end{array}\right. \\ \\ \\ \xrightarrow{2} aTbc \xrightarrow{(m-1)*3} a^mTb^mc \longrightarrow \left| \begin{array}{l} \xrightarrow{4} a^mb^mc \\ \xrightarrow{5} a^mAb^mc \xrightarrow{*6} a^m(b^+c)^+b^mc \end{array}\right. \end{array}\right.$$

Abbildung 5.5: Ableitungsschema der Grammatik G

Somit folgt, dass L=L(G) ist und dass jedes Wort $z :=$

$a^m b^{m_1} c b^{m_2} c \ldots b^{m_p} c$ aus L so viele Ableitungsbäume in G hat, wie die Anzahl der i's mit $m = m_i$, wobei $i \in \{1, 2, \ldots, p\}$.
Wegen $|\{i | m = m_i, i \in \{1, 2, \ldots, p\}\}| \leq p \leq n := |z|$,
gilt $da_G(n) = O(n)$.

q. e. d.

Bevor wir die Existenz von polynomiell mehrdeutigen Sprachen nachweisen, zeigen wir ein ähnliches Ergebnis zum Satz 4.3.1 (auf die Seite 48).

Satz 5.2.2 *Für jedes feste d aus $\mathbb{R}_+$ können wir eine Grammatik G_d für L entwerfen, so dass $da_{G_d}(n) \leq dn$ für genügend großes n ist.*

Beweis Wir teilen die Sprache L in zwei (kontextfreie) disjunkte Sprachen auf:
$L_1 := \{a^m b^{m_1} c b^{m_2} c \ldots b^{m_p} c \big| p \in \mathbb{N}; m, m_1, m_2, \ldots, m_p \in \mathbb{N};\ \mathbf{m} \leq \mathbf{k}\ \textit{und}\ \exists i \in \{1, 2, \ldots, p\}\ \textit{mit}\ m = m_i\}$
und $L_2 := \{a^m b^{m_1} c b^{m_2} c \ldots b^{m_p} c \big| p \in \mathbb{N}; m, m_1, m_2, \ldots, m_p \in \mathbb{N};\ \mathbf{m} > \mathbf{k}\ \textit{und}\ \exists i \in \{1, 2, \ldots, p\}\ \textit{mit}\ m = m_i\}$
wobei k eine Konstante ist, die wir später geeignet festlegen werden.

- Die Sprache L_1 ist als Vereinigung endlich vieler regulärer Sprachen regulär. Denn $L_1 = \bigcup_{m=1}^{k} a^m (b^{\neq m} c)^* b^m c (b^+ c)^*$ und $a^m (b^{\neq m} c)^* b^m c (b^+ c)^*$ ist für festes m regulär. Es gibt für L_1 also eine eindeutige (reguläre) Grammatik $G_1 = (N_1, \{a, b, c\}, P_1, S_1)$.

- Die Sprache L_2 ist kontextfrei. Es ist nicht schwer für L_2 eine Grammatik $G_2 = (N_2, \{a, b, c\}, P_2, S_2)$ zu entwerfen, so dass N_1 und N_2 disjunkt sind und für jedes Wort $w := a^m b^{m_1} c b^{m_2} c \ldots b^{m_p} c$ aus L_2 gilt: $da_{G_2}(w) = \Big| \{i | m_i = m\} \Big|$. Für G_2 bietet sich eine ähnliche Grammatik wie die Grammatik G (auf Seite 65) aus dem Beweis zu Satz 5.2.1 (auf die Seite 55) an. Dabei ist zu beachten, dass $m > k$ für festes k ist.

Aus $\left|\{i|m_i = m\}\right| \leq \frac{|w|}{k}$. folgt $da_{G_2}(w) \leq \frac{|w|}{k}$. Das heißt $da_{G_2}(w) \leq \frac{n}{k}$. Setzen wir $k = \left\lceil \frac{1}{d} \right\rceil$, so erhalten wir $da_{G_2}(n) \leq dn$.

Für die Grammatik $\widehat{G} := G_1 \dot{\cup} G_2 = (N_1 \dot{\cup} N_2 \dot{\cup} \{S\}, \{a,b,c\}, P_1 \dot{\cup} P_2 \dot{\cup} \{S \to S_1S_2\}, S)$ gilt : L=L($\widehat{G}$) und $da_{\widehat{G}}n) \leq max(da_{G_1}(n), da_{G_2}(n))$, denn $\widehat{G} := G_1 \dot{\cup} G_2$ mit G_1 und G_2 disjunkt. das heißt $da_{\widehat{G}}(n) \leq max(1, dn)$ damit ist $da_{\widehat{G}}(n) = dn$, für genügend großes n.

q. e. d.

5.3 polynomiell vom Grad k mehrdeutige Sprachen

Wir weisen nun die Existenz polynomiell vom Grad k mehrdeutiger Sprachen nach.

Satz 5.3.1 *Seien $k \in \mathbb{N}$ und L die Sprache aus dem Beweis für Satz 5.2.1. (auf Seite 55)Die Sprache L^k ist polynomiell vom Grad k mehrdeutig.*

Beweis Den Fall k=1 haben wir im Satz 5.2.1 (auf Seite 55) behandelt. Wir brauchen nur noch den Fall für $k \geq 2$ zu betrachten. Seien $k \geq 2$ und G=(N,Σ, P, S) eine CFG für L^k. Sei h die Konstante aus OGDEN's Lemma (Satz 2.4.1 auf Seite 16) für die Grammatik G. Für jedes $p \in \mathbb{N}$ betrachten wir die Wörter z aus L^k der Form $z = z_{i_1} z_{i_2} \ldots z_{i_k}$ mit $z_{i_j} := a^h (b^{h+h!}c)^{i_j - 1} b^h c (b^{h+h!}c)^{p - i_j}$, wobei j=1,...,k und $i_j = 1, \ldots, p$. Markieren wir genau alle a's in jedem z_{i_α} mit $\alpha \in \{1, 2, \ldots, k\}$ und gehen wir wie beim Beweis vom Satz 4.2.1(auf Seite 36) vor, indem wir nur die h markierten a's eines z_{i_α} berücksichtigen, so können wir mit Hilfe von OGDEN's Lemma zeigen, dass z eine Zerlegung $\hat{u_{i_\alpha}} v_{i_\alpha} w_{i_\alpha} x_{i_\alpha} \hat{y_{i_\alpha}}$ hat, so dass folgendes gilt:

1. $S \overset{+}{\Longrightarrow} \hat{u_{i_\alpha}} A_{i_\alpha} \hat{y_{i_\alpha}} \overset{+}{\Longrightarrow} \hat{u_{i_\alpha}} v_{i_\alpha} A_{i_\alpha} x_{i_\alpha} \hat{y_{i_\alpha}} \overset{+}{\Longrightarrow} \hat{u_{i_\alpha}} v_{i_\alpha} w_{i_\alpha} x_{i_\alpha} \hat{y_{i_\alpha}} = z$

2. In w_{i_α} tritt mindestens eine der h markierten Positionen von z_{i_α} auf.

3. **Entweder** treten sowohl in $\hat{u_{i_\alpha}}$ als auch in v_{i_α} markierte Positionen von z_{i_α} auf **oder** es treten sowohl in x_{i_α} als auch in $\hat{y_{i_\alpha}}$ markierte Positionen von z_{i_α} auf.

4. In $v_{i_\alpha} w_{i_\alpha} x_{i_\alpha}$ treten höchstens h der markierten Positionen von z_{i_α} auf,

5.
$$\begin{aligned} S &\overset{+}{\Longrightarrow} \hat{u_{i_\alpha}} A_{i_\alpha} \hat{y_{i_\alpha}} \\ &\overset{+}{\Longrightarrow} \hat{u_{i_\alpha}} v_{i_\alpha} A_{i_\alpha} x_{i_\alpha} \hat{y_{i_\alpha}} \\ &\overset{+}{\Longrightarrow} \dots \\ &\overset{+}{\Longrightarrow} \hat{u_{i_\alpha}} v^q_{i_\alpha} A_{i_\alpha} x^q_{i_\alpha} \hat{y_{i_\alpha}} \\ &\overset{+}{\Longrightarrow} \hat{u_{i_\alpha}} v^q_{i_\alpha} w_{i_\alpha} x^q_{i_\alpha} \hat{y_{i_\alpha}} \in L^k \textit{ wobei } q \in \mathbb{N}_0. \end{aligned}$$

   (insbesondere für q=0 gilt $S \overset{+}{\Longrightarrow} \hat{u_{i_\alpha}} A_{i_\alpha} \hat{y_{i_\alpha}} \overset{+}{\Longrightarrow} \hat{u_{i_\alpha}} w_{i_\alpha} \hat{y_{i_\alpha}}$)

Als nächstens zeigen wir das

Lemma 5.3.1 *Für die nach* OGDEN*'s Lemma existierende Zerlegung* $z = \hat{u_{i_\alpha}} v_{i_\alpha} w_{i_\alpha} x_{i_\alpha} \hat{y_{i_\alpha}}$ *gilt folgendes:*

$$\begin{aligned} &\hat{u_{i_\alpha}} = z_{i_1} \dots z_{i_{\alpha-1}} u_{i_\alpha} && u_{i_\alpha} = a^{r_\alpha} \textit{ und } 1 \le r_\alpha \le h-2, \\ &v_{i_\alpha} = a^{s_\alpha} && 1 \le s_\alpha \le h-2, \\ &w_{i_\alpha} = a^{h-r_\alpha-s_\alpha} (b^{h+h!}c)^{i_\alpha -1} b^{t_\alpha} && 0 \le t_\alpha \le h-1, \\ &x_{i_\alpha} = b^{s_\alpha}, && \\ &\hat{y_{i_\alpha}} = y_{i_\alpha} z_{i_{\alpha+1}} \dots z_{i_k} && y_{i_\alpha} = b^{h-t_\alpha-s_\alpha} c (b^{h+h!}c)^{p-i_\alpha}. \end{aligned}$$

Beweis (Lemma 5.3.1) Aus OGDEN's Lemma folgt: **Entweder** treten sowohl in $\hat{u_{i_\alpha}}$ als auch in v_{i_α} markierte Positionen von z_{i_α} auf **oder** es treten sowohl in x_{i_α} als auch in $\hat{y_{i_\alpha}}$ markierte Positionen von z_{i_α} auf.

Wenn sowohl x_{i_α} als auch $\hat{y_{i_\alpha}}$ markierte Positionen von z_{i_α} haben, dann gilt $x_{i_\alpha} \in a^+$ und $x_{i_\alpha} \in z_{i_\alpha}$, denn w_{i_α} hat gemäß OGDEN's Lemma mindestens eine markierte Position von z_{i_α} (also ein a) und x_{i_α} kommt zwischen w_{i_α} und y_{i_α} in z vor.
Sei $x_{i_\alpha} = a^s \in z_{i_\alpha}$ mit $1 \leq s \leq h-2$. Daraus folgt: w_{i_α} endet mit einem a aus z_{i_α} und $\hat{y_{i_\alpha}} = a^t(b^{h+h!}c)^{i_\alpha-1}b^h c(b^{h+h!}c)^{p-i_\alpha} z_{i_{\alpha+1}} \ldots z_{i_k}$, wobei $1 \leq t \leq h-2$.

Für v_{i_α} kommen folgende Möglichkeiten in Frage: ($v_{i_\alpha} \in a^*$ und $v_{i_\alpha} \in z_{i_\alpha}$) oder (v_{i_α} beginnt mit einem Zeichen aus $z_{i_1} \ldots z_{i_{\alpha-1}}$)

Ist $v_{i_\alpha} \in a^*$ und $v_{i_\alpha} \in z_{i_\alpha}$, dann ist

$\bar{z} = \hat{u_{i_\alpha}} v_{i_\alpha}^2 w_{i_\alpha} x_{i_\alpha}^2 \hat{y_{i_\alpha}} = z_{i_1} \ldots z_{i_{\alpha-1}} a^{h+s+|v_{i_\alpha}|}(b^{h+h!}c)^{i_\alpha-1}b^h c(b^{h+h!}c)^{p-i_\alpha} z_{i_{\alpha+1}} \ldots z_{i_k} \notin L^k$, weil $a^{h+s+|v_{i_\alpha}|}(b^{h+h!}c)^{i_\alpha-1}b^h c(b^{h+h!}c)^{p-i_\alpha} \notin L$, denn $1 < s + |v_{i_\alpha}| < h-2$.
Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Beginnt v_{i_α} mit einem Zeichen aus $z_{i_1} \ldots z_{i_{\alpha-1}}$, so ist $v_{i_\alpha} w_{i_\alpha} = vca^{h-s-t}$, wobei das Terminal c das letzte Zeichen c in $z_{i_{\alpha-1}}$ ist, dann gilt:

$$\begin{aligned}
\bar{z} &= \hat{u_{i_\alpha}} v_{i_\alpha}^2 w_{i_\alpha} x_{i_\alpha}^2 \hat{y_{i_\alpha}} \\
&= \hat{u_{i_\alpha}} v_{i_\alpha} v_{i_\alpha} w_{i_\alpha} x_{i_\alpha} x_{i_\alpha} \hat{y_{i_\alpha}} \\
&= \hat{u_{i_\alpha}} v_{i_\alpha} vca^{h-s-t} a^s a^s a^t (b^{h+h!}c)^{i_\alpha-1} b^h c (b^{h+h!}c)^{p-i_\alpha} z_{i_{\alpha+1}} \ldots z_{i_k} \\
&= \hat{u_{i_\alpha}} v_{i_\alpha} vca^{h+s} (b^{h+h!}c)^{i_\alpha-1} b^h c (b^{h+h!}c)^{p-i_\alpha} z_{i_{\alpha+1}} \ldots z_{i_k} \notin L^k,
\end{aligned}$$

weil $a^{h+s}(b^{h+h!}c)^{i_\alpha-1}b^h c(b^{h+h!}c)^{p-i_\alpha} \notin L$.

Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Damit folgt: x_{i_α} und $\hat{y_{i_\alpha}}$ haben nicht beide markierte Positionen von z_{i_α}. D. h. sowohl $\hat{u_{i_\alpha}}$ als auch v_{i_α} müssen markierte Positionen von z_{i_α} haben. Da w_{i_α} mindestens eine markierte Position von z_{i_α} hat und x_{i_α} zwischen $\hat{u_{i_\alpha}}$ und

w_{i_α} in z vorkommt, folgt $v_{i_\alpha} \in a^+$ und $v_{i_\alpha} \in z_{i_\alpha}$.
Damit ist $v_{i_\alpha} = a^s \in z_{i_\alpha}$, wobei $1 \leq s \leq h-2$.
Daraus folgt: w_{i_α} beginnt mit einem a aus z_{i_α} und $\hat{u_{i_\alpha}} = z_{i_1} \dots z_{i_{\alpha-1}} a^r$, wobei $1 \leq r \leq h-2$.

Für x_{i_α} kommen zwei Möglichkeiten in Frage: (x_{i_α} endet mit einem Zeichen aus $z_{i_{\alpha+1}} \dots z_{i_k}$) oder ($x_{i_\alpha}$ endet mit einem Zeichen aus z_{i_α}, das heißt $x_{i_\alpha} \in z_{i_\alpha}$).

Endet x_{i_α} mit einem Zeichen aus $z_{i_{\alpha+1}} \dots z_{i_k}$, so ist $w_{i_\alpha} x_{i_\alpha} = a^{h-r-s}(b^{h+h!}c)^{i_\alpha - 1} b^h c (b^{h+h!}c)^{p-i_\alpha} ax$, wobei das a das erste a in $z_{i_{\alpha+1}}$ ist. Dann gilt aber

$$
\begin{aligned}
\bar{z} &= \hat{u_{i_\alpha}} v_{i_\alpha}^2 w_{i_\alpha} x_{i_\alpha}^2 \hat{y_{i_\alpha}} \\
&= \hat{u_{i_\alpha}} v_{i_\alpha} v_{i_\alpha} w_{i_\alpha} x_{i_\alpha} x_{i_\alpha} \hat{y_{i_\alpha}} \\
&= z_{i_1} \dots z_{i_{\alpha-1}} a^r a^s a^s a^{h-r-s} (b^{h+h!}c)^{i_\alpha-1} b^h c (b^{h+h!}c)^{p-i_\alpha} a x x_{i_\alpha} \hat{y_{i_\alpha}} \\
&= z_{i_1} \dots z_{i_{\alpha-1}} a^{h+s} (b^{h+h!}c)^{i_\alpha-1} b^h c (b^{h+h!}c)^{p-i_\alpha} a x x_{i_\alpha} \hat{y_{i_\alpha}} \notin L^k,
\end{aligned}
$$

weil $a^{h+s}(b^{h+h!}c)^{i_\alpha-1} b^h c (b^{h+h!}c)^{p-i_\alpha} \notin L$.

Dies führt zu einem Widerspruch zu der Aussage von OGDEN's Lemma.

Damit gilt: $x_{i_\alpha} \in z_{i_\alpha}$. Daraus folgt: $\hat{u_{i_\alpha}} v_{i_\alpha} w_{i_\alpha} x_{i_\alpha} \hat{y_{i_\alpha}} = z_{i_1} \dots z_{i_{\alpha-1}} u_{i_\alpha} v_{i_\alpha} w_{i_\alpha} x_{i_\alpha} y_{i_\alpha} z_{i_{\alpha+1}} \dots z_{i_k}$, wobei $u_{i_\alpha} v_{i_\alpha} w_{i_\alpha} x_{i_\alpha} y_{i_\alpha} = z_{i_\alpha}$. Da k fest ist, ist $\hat{u_{i_\alpha}} v_{i_\alpha}^q w_{i_\alpha} x_{i_\alpha}^q \hat{y_{i_\alpha}} \in L^k$ äquivalent zu $u_{i_\alpha} v_{i_\alpha}^q w_{i_\alpha} x_{i_\alpha}^q y_{i_\alpha} \in L$ für alle q aus $\mathbb{N}_0$. Damit lässt sich die Zerlegung von z auf eine Zerlegung von z_{i_α}. zurückführen, die die Bedingungen vom Lemma 5.2.1 auf Seite 55 erfüllt. Wir können also das Lemma 5.2.1 auf die Zerlegung $u_{i_\alpha} v_{i_\alpha} w_{i_\alpha} x_{i_\alpha} y_{i_\alpha} = z_{i_\alpha}$ anwenden, dann erhalten wir:

$$\begin{array}{ll}
u_{i_\alpha} = a^{r_\alpha} & 1 \le r_\alpha \le h-2, \\
v_{i_\alpha} = a^{s_\alpha} & 1 \le s_\alpha \le h-2, \\
w_{i_\alpha} = a^{h-r_\alpha-s_\alpha}(b^{h+h!}c)^{i_\alpha-1}b^{t_\alpha} & 0 \le t_\alpha \le h-1, \\
x_{i_\alpha} = b^{s_\alpha}, & \\
y_{i_\alpha} = b^{h-t_\alpha-s_\alpha}c(b^{h+h!}c)^{p-i_\alpha} & .
\end{array}$$

Und schließlich:

$$\begin{array}{ll}
\hat{u_{i_\alpha}} = z_{i_1}\dots z_{i_{\alpha-1}}u_{i_\alpha} & u_{i_\alpha} = a^{r_\alpha} \; und \; 1 \le r_\alpha \le h-2, \\
v_{i_\alpha} = a^{s_\alpha} & 1 \le s_\alpha \le h-2, \\
w_{i_\alpha} = a^{h-r_\alpha-s_\alpha}(b^{h+h!}c)^{i_\alpha-1}b^{t_\alpha} & 0 \le t_\alpha \le h-1, \\
x_{i_\alpha} = b^{s_\alpha}, & \\
\hat{y_{i_\alpha}} = y_{i_\alpha}z_{i_{\alpha+1}}\dots z_{i_k} & y_{i_\alpha} = b^{h-t_\alpha-s_\alpha}c(b^{h+h!}c)^{p-i_\alpha}.
\end{array}$$

q. e. d.

Wir setzen nun den Beweis von Satz 5.3.1 (auf Seite 67) fort.

Wir bemerken, dass v_{i_α} und x_{i_α} vollständig in z_{i_α} enthalten sind. Damit hat der Ableitungsbaum einen A_{i_α}–Pumpbaum, der die Anzahl der Symbole von v_{i_α} und x_{i_α} gleichmäßig erhöht. Dies gilt für jedes α aus $\{1,\dots,k\}$. Das heißt, jeder Ableitungsbaum B(z) in G mit der Front $z = z_{i_1}z_{i_2}\dots z_{i_k}$ hat k nebeneinander stehende A_{i_1}–, A_{i_2}–,… A_{i_k}–Pumpbäume und sieht wie folgt aus (siehe Abbildung 5.6 auf die Seite 72):

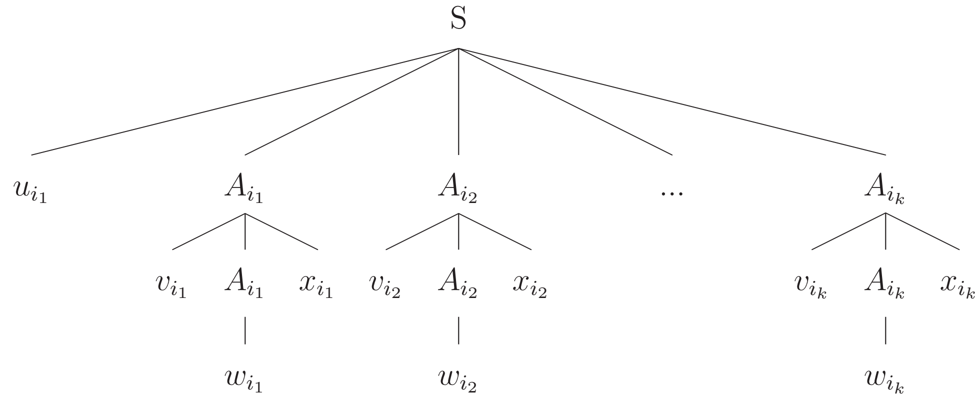

Abbildung 5.6: Ein Ableitungsbaum B(z) für ein $z = z_{i_1} z_{i_2} \dots z_{i_k}$

Aus dem Ableitungsbaum B(z) für z lässt sich eine Ableitung konstruieren, die folgendermaßen aussieht:

$$S \stackrel{+}{\Longrightarrow} u_{i_1} A_{i_1} y_{i_1} \dots z_{i_{\alpha-1}} u_{i_\alpha} A_{i_\alpha} y_{i_\alpha} \dots u_{i_k} A_{i_k} y_{i_k} \tag{5.3.1}$$

wobei $A_{i_\alpha} \stackrel{+}{\Longrightarrow} v_{i_\alpha} A_{i_\alpha} x_{i_\alpha}$, und $A_{i_\alpha} \stackrel{+}{\Longrightarrow} w_{i_\alpha}$

Pumpen wir nun jeden A_{i_j}–Pumpbaum des Baumes B(z) $q_{i_j} = \frac{h!}{s_{i_j}} + 1$ mal, so erhalten wir einen Ableitungsbaum T(z) für das Wort $(a^{h+h!}(b^{h+h!}c)^p)^k$. Der Ableitungsbaum T(z) ist in der Abbildung 5.7 auf die Seite 73 dargestellt.

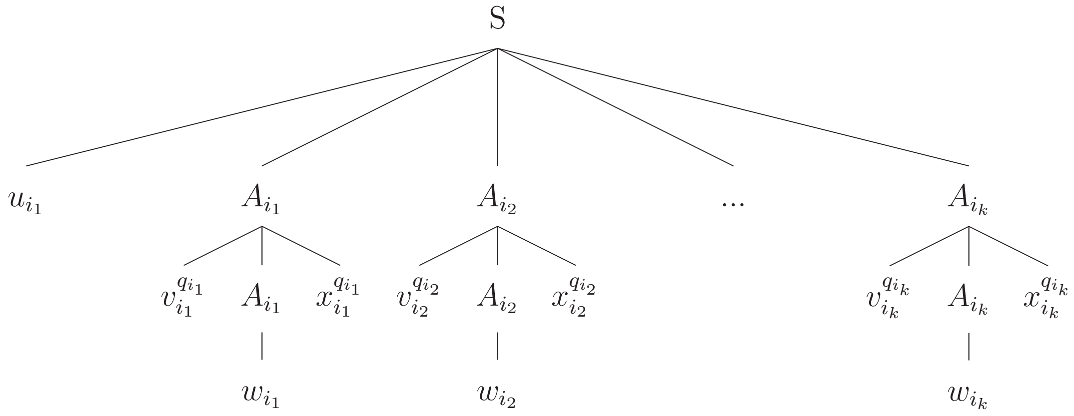

Abbildung 5.7: Jeden A_{i_j}–Pumpbaum von B(z) pumpen ergibt einen Ableitungsbaum T(z) für das Wort $(a^{h+h!}(b^{h+h!}c)^p)^k$

Von den Wörtern der Form $z = z_{i_1}z_{i_2}\ldots z_{i_k}$ mit $z_{i_j} := a^h(b^{h+h!}c)^{i_j-1}b^hc(b^{h+h!}c)^{p-i_j}$, j=1, …,k und $i_j = 1,\ldots,p$ gibt es p^k verschiedene Wörter. Damit gibt es p^k Ableitungsbäume der Form T(z). Wir zeigen nun, dass diese p^k Ableitungsbäume T(z) paarweise verschieden sind.

Seien

$z = z_{i_1}z_{i_2}\ldots z_{i_k}$ *mit* $z_{i_j} := a^h(b^{h+h!}c)^{i_j-1}b^hc(b^{h+h!}c)^{p-i_j}$

und

$\tilde{z} = z_{\tilde{i}_1}z_{\tilde{i}_2}\ldots z_{\tilde{i}_k}$ *mit* $z_{\tilde{i}_j} := a^h(b^{h+h!}c)^{\tilde{i}_j-1}b^hc(b^{h+h!}c)^{p-\tilde{i}_j}$

verschieden. Das heißt, es gibt ein j mit $i_j \neq \tilde{i}_j$. Es genügt zu zeigen, dass die Ableitungsbäume T(z) und T($\tilde{z}$) verschieden sind. Der Ableitungsbaum T(z) hat gemäß (5.3.1) eine Ableitung, die folgendermaßen aussieht

$$\begin{aligned} S &\stackrel{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{i_j-1}u_{i_j}v_{i_j}^{q_{i_j}}A_{i_j}x_{i_j}^{q_{i_j}}y_{i_j}(a^{h+h!}(b^{h+h!}c)^p)^{k-i_j} \\ &\stackrel{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{i_j-1}u_{i_j}v_{i_j}^{q_{i_j}}w_{i_j}x_{i_j}^{q_{i_j}}y_{i_j}(a^{h+h!}(b^{h+h!}c)^p)^{k-i_j} \\ &= (a^{h+h!}(b^{h+h!}c)^p)^k \in L^k \end{aligned}$$

Analog hat der Ableitungsbaum T($\tilde{z}$) eine Ableitung, die folgendermaßen aussieht:

$$\begin{array}{rcl} S & \stackrel{+}{\Longrightarrow} & (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1} u_{\tilde{i}_j} v_{\tilde{i}_j}^{q_{\tilde{i}_j}} A_{\tilde{i}_j} x_{\tilde{i}_j}^{q_{\tilde{i}_j}} y_{\tilde{i}_j} (a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j} \\ & \stackrel{+}{\Longrightarrow} & (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1} u_{\tilde{i}_j} v_{\tilde{i}_j}^{q_{\tilde{i}_j}} w_{\tilde{i}_j} x_{\tilde{i}_j}^{q_{\tilde{i}_j}} y_{\tilde{i}_j} (a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j} \\ & = & (a^{h+h!}(b^{h+h!}c)^p)^k \in L^k \end{array}$$

Nehmen wir an, dass die Ableitungsbäume T(z) und $T(z, \tilde{z})$ für das Wort $a^{h+h!}(b^{h+h!}c)^p$ gleich wären, also gleich einem Ableitungsbaum $T(z, \tilde{z})$. Dieser Ableitungsbaum $T(z, \tilde{z})$ mit der Front $(a^{h+h!}(b^{h+h!}c)^p)^k$ müßte sowohl einen A_{i_j}– als auch einen $A_{\tilde{i}_j}$–Pumpbaum haben. Damit können wir folgende Fälle unterscheiden:

1. Der Knoten A_{i_j} kommt links vom Knoten $A_{\tilde{i}_j}$ im $T(z, \tilde{z})$ vor (siehe Abbildung 5.8 auf die Seite 75)

2. Der Knoten A_{i_j} kommt rechts vom Knoten $A_{\tilde{i}_j}$ im Baum $T(z, \tilde{z})$ vor.

3. Der Knoten A_{i_j} ist ein Nachfolger des Knotens $A_{\tilde{i}_j}$ (siehe Abbildung 5.9 auf die Seite 76).

4. Der Knoten $A_{\tilde{i}_j}$ ist ein Nachfolger des Knotens A_{i_j}. Der 1. und 2. Fall (bzw. 3. und 4.) sind ähnlich, deshalb behandeln wir nur den 1. sowie den 3. Fall.

Für den 1. Fall sieht der Baum $T(z, \tilde{z})$ wie folgt aus:

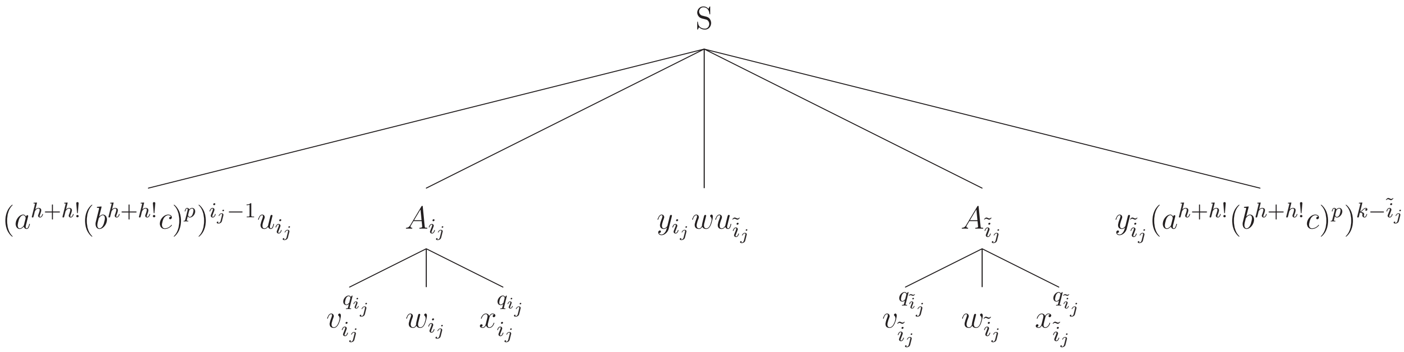

Abbildung 5.8: Der Knoten A_{i_j} kommt links vom Knoten $A_{\tilde{i}_j}$ inm $T(z, \tilde{z})$ vor

Hier hätte die Front des Baumes $T(z, \tilde{z})$ mindestens (k+1) Teilwörter der Form $a^{h+h!}(b^{h+h!}c)^p$, im Widerspruch zur Tatsache, dass die Front das Wort $(a^{h+h!}(b^{h+h!}c)^p)^k$ ist.

Für den 3. Fall sieht der Baum $T(z, \tilde{z})$ wie folgt aus:

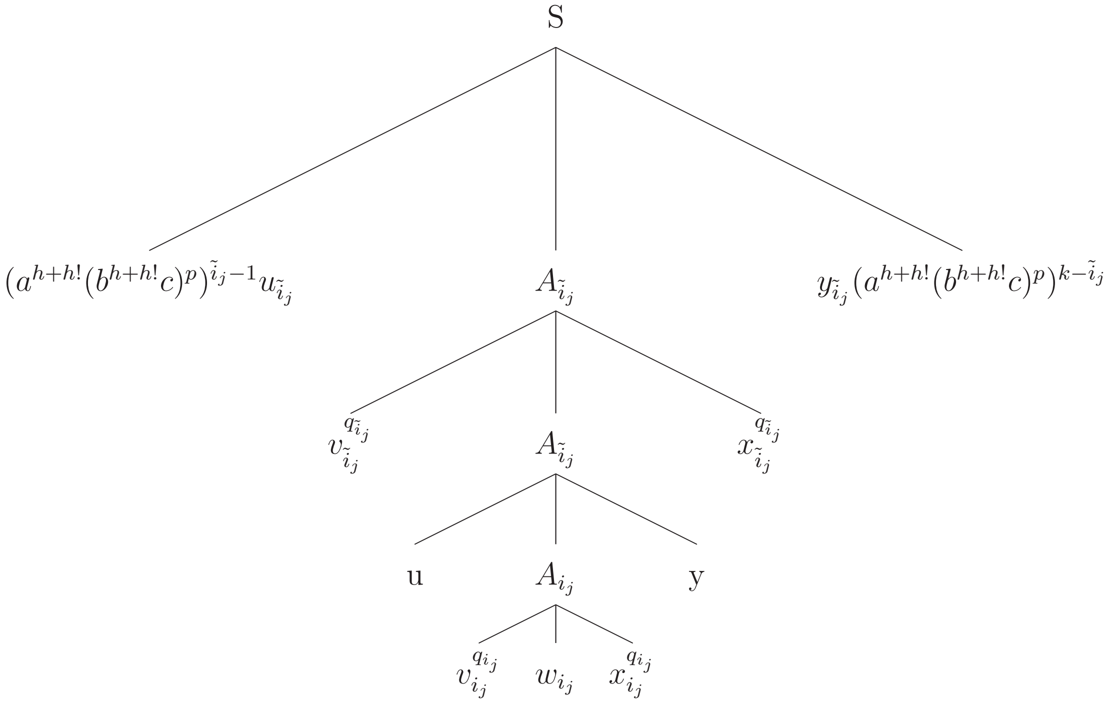

Abbildung 5.9: Der Knoten A_{i_j} ist ein Nachfolger des Knotens $A_{\tilde{i}_j}$

Hier erhalten wir:

$$
\begin{aligned}
S &\overset{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}u_{\tilde{i}_j}v_{\tilde{i}_j}^{q_{\tilde{i}_j}}A_{\tilde{i}_j}x_{\tilde{i}_j}^{q_{\tilde{i}_j}}y_{\tilde{i}_j}(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j} \\
&\overset{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}u_{\tilde{i}_j}v_{\tilde{i}_j}^{q_{\tilde{i}_j}}uA_{i_j}yx_{\tilde{i}_j}^{q_{\tilde{i}_j}}y_{\tilde{i}_j}(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j} \\
&\overset{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}u_{\tilde{i}_j}v_{\tilde{i}_j}^{q_{\tilde{i}_j}}uv_{i_j}^{q_{i_j}}A_{i_j}x_{i_j}^{q_{i_j}}yx_{\tilde{i}_j}^{q_{\tilde{i}_j}}y_{\tilde{i}_j}(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j} \\
&\overset{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}\underbrace{u_{\tilde{i}_j}v_{\tilde{i}_j}^{q_{\tilde{i}_j}}uv_{i_j}^{q_{i_j}}w_{i_j}x_{i_j}^{q_{i_j}}yx_{\tilde{i}_j}^{q_{\tilde{i}_j}}y_{\tilde{i}_j}}_{t_1}(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j} \\
&= (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}t_1(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j} \in L^k
\end{aligned}
$$

Da die Front des Baumes $T(z,\tilde{z})$ das Wort $(a^{h+h!}(b^{h+h!}c)^p)^k$ ist, folgt $t_1 = a^{h+h!}(b^{h+h!}c)^p$.

Im Baum T(z,) können wir aber auch den A_{i_j}–Pumpbaum q_{i_j}+1 mal und den $A_{\tilde{i}_j}$–Pumpbaum $q_{\tilde{i}_j}+1$ mal pumpen. Dann gilt aber:

$$\begin{aligned}
S &\overset{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}u_{\tilde{i}_j}v_{\tilde{i}_j}^{q_{\tilde{i}_j}+1}A_{\tilde{i}_j}x_{\tilde{i}_j}^{q_{\tilde{i}_j}+1}y_{\tilde{i}_j}(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j}\\
&\overset{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}u_{\tilde{i}_j}v_{\tilde{i}_j}^{q_{\tilde{i}_j}+1}uA_{i_j}yx_{\tilde{i}_j}^{q_{\tilde{i}_j}+1}y_{\tilde{i}_j}(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j}\\
&\overset{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}u_{\tilde{i}_j}v_{\tilde{i}_j}^{q_{\tilde{i}_j}+1}uv_{i_j}^{q_{i_j}+1}A_{i_j}x_{i_j}^{q_{i_j}+1}yx_{\tilde{i}_j}^{q_{\tilde{i}_j}+1}y_{\tilde{i}_j}(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j}\\
&\overset{+}{\Longrightarrow} (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}\underbrace{u_{\tilde{i}_j}v_{\tilde{i}_j}^{q_{\tilde{i}_j}+1}uv_{i_j}^{q_{i_j}+1}w_{i_j}x_{i_j}^{q_{i_j}+1}yx_{\tilde{i}_j}^{q_{\tilde{i}_j}+1}y_{\tilde{i}_j}}_{t_2}(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j}\\
&= (a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}t_2(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j} \in L^k
\end{aligned}$$

$$\#_a(t_2) = \#_a(t_1) + |v_{\tilde{i}_j}| + |v_{i_j}| = h + h! + |v_{\tilde{i}_j}| + |v_{i_j}|$$

In jedem b-Block von t_2 ist die Anzahl der b's entweder h+h! oder $h + h! + |x_{i_j}|$ oder $h + h! + |x_{\tilde{i}_j}|$, also ungleich der Anzahl der a's in t_2, denn $|v_{i_j}| = |x_{i_j}|$ und $|v_{\tilde{i}_j}| = |x_{\tilde{i}_j}|$. Damit gilt, $t_2 \notin L$. Dies steht im Widerspruch zu $(a^{h+h!}(b^{h+h!}c)^p)^{\tilde{i}_j-1}t_2(a^{h+h!}(b^{h+h!}c)^p)^{k-\tilde{i}_j} \in L^k$.

Insgesamt können wir nun schließen, dass T(z) und T($\tilde{z}$) verschieden sind. Da es

insgesamt p^k verschiedene Wörter der Form $z = z_{i_1}z_{i_2}\ldots z_{i_k}$ mit $z_{i_j} := a^h(b^{h+h!}c)^{i_j-1}b^hc(b^{h+h!}c)^{p-i_j}$ j=1, ...,k und $i_j = 1,\ldots,p$ gibt, hat das Wort $a^{h+h!}(b^{h+h!}c)^p)^k$ mindestens p^k paarweise verschiedene Ableitungsbäume in G.

Aus $n := |a^{h+h!}(b^{h+h!}c)^p)^k| = k(p(h+h!+1)+h+h!)$ folgt $da_G(n) = \Omega(n^k)$.

Wir entwerfen nun eine Grammatik $\widehat{G}$ für L^k, so dass $da_{\widehat{G}}(n) = O(n^k)$ ist. Dazu bietet sich folgende Idee an. Wir nehmen die Grammatik G für L (aus dem Beweis von Satz 5.2.1 auf die Seite 65) und fügen ein neues Startsymbol $\widetilde{S}$ sowie eine neue Regel $\widetilde{S} \to S^k$ hinzu. Somit erhalten wir die Grammatik $G := (\{\widetilde{S}\}\dot\cup\{S,T,A\},\{a,b,c\},P\dot\cup\{\widetilde{S}\to S^k\},\widetilde{S})$, wobei P die Produktionsmenge aus der Grammatik G (aus dem Beweis von Satz 5.2.1 auf die Seite 65) ist. Wir bemerken, dass man aus $\widetilde{S}$ nur die Wörter der Form

$z = z_1 z_2 \ldots z_k$ ableiten kann, so dass

$\widetilde{S} \to S^k$,

$S \overset{+}{\Longrightarrow} z_i$, wobei $1 \leq i \leq k$.

Daraus folgt, dass

- $L(\widehat{G}) = L^k$.
- $da_{\widehat{G}}(z) = \prod_{i=1}^{n} (\textit{Anzahl der verschiedenen Ableitungen } (S \overset{+}{\Longrightarrow} z_i))$

Aus Satz 5.2.1 auf die Seite 55 folgt aber, dass die Anzahl der verschiedenen Ableitungen der Form $(S \overset{+}{\Longrightarrow} z_i) = O(|z_i|)$ ist. Somit folgt, $da_G(z) = O(|z|^k)$.

Damit ist $da_{\widehat{G}}(n) = O(n^k)$.

q. e. d.

Bemerkung 5.3.1 *Für die Menge der Palindrome* $PAL := \{w | w \in \{a, b\}^* \textit{ und } w = w^R\}$ *ist zu vermuten, dass* $(PAL)^k$ $\Theta(n^{k-1})$*–deutig ist, wobei* $k \in \mathbb{N}$ *ist. Insbesondere ist PAL eindeutig. Für k=2 erhalten wir die sogenannte* CRESTIN*–Sprache* $C := \{uv | u, v \in \{a, b\}^*; u = u^R \textit{ und } v = v^R\}$ *(vgl. [1]).*

Ausblick

In dieser Arbeit haben wir uns hauptsächlich mit kontextfreien Grammatiken und Sprachen beschäftigt, bei denen der Mehrdeutigkeitsgrad für jedes Wort von dessen Länge abhängt. Wir konnten konstant, exponentiell sowie polynomiell mehrdeutige Grammatiken und Sprachen angeben. Aus [10] haben wir die Tatsache zitiert, dass es keine Grammatiken (und somit keine Sprachen) gibt, deren Mehrdeutigkeit stärker als polynomiell, aber schwächer als exponentiell wächst (wie z. B. $\Theta(2^{\sqrt{n}})$)). Außerdem haben wir bewiesen, dass es weder Sprachen noch zyklenfreie Grammatiken gibt, deren Mehrdeutigkeitsgrad stärker als $2^{\Theta(n)}$ wächst (wie z B. $\Theta(n^n)$). Ferner haben wir gezeigt, dass die Mehrdeutigkeit einer Grammatik genau dann für einzelne Wörter unendlich ist, wenn die Grammatik (nützliche) Zyklen enthält.

Unbeantwortet blieb in dieser Arbeit die Frage, ob es $\Theta(n^r)$–deutige Grammatiken (bzw. Sprachen) gibt, wobei r eine nicht natürliche Zahl ist.

In [5, Seite 100] steht "...removing unit productions, cannot introduce ambiguities..." Das bedeutet, dass die Beseitigung von Kettenregeln keine Mehrdeutigkeiten einführen kann. Präziser können wir sagen, dass die Mehrdeutigkeit wohl vermindert wird bzw. gleichbleibt, falls man die Kettenregeln beseitigt. Es stellt sich aber die Frage, inwieweit die Beseitigung von Kettenregeln die Mehrdeutigkeit einer Grammatik ändert. Beseitigt man z. B. die Kettenregeln der für einzelne Wörter unendlich mehrdeutigen Grammatik mit den Regeln $S \rightarrow S|a$, erhält man die eindeutige Grammatik mit der Regel $S \rightarrow a$. Man könnte die Wirkung der Beseitigung von Kettenregeln auf die Mehrdeutigkeit einer Grammatik untersuchen. Wir halten es für sinnvoll, auch die Wirkung der Elimination von ε–Regeln auf die Mehrdeutigkeit einer Grammatik zu studieren. Die Mehrdeutigkeit einer Grammatik ändert sich aber nicht, wenn man nutzlose Produktionen und Symbole beseitigt.

Im Kapitel 4 ist uns das offene Problem begegnet, ob es überhaupt eine exponentiell mehrdeutige Sprache gibt, die für irgendein festes c aus $\mathbb{R}_+$ 2^{cn}–deutig ist.

Es ist zwar nicht entscheidbar[1] , ob eine gegebene Grammatik mehrdeutig ist. Im Laufe der Zeit hat man aber einige Kriterien für die Mehrdeutigkeit entwickelt, die im Einzelfall hilfreich sein können. In der Literatur konnten wir einige Kriterien[2]. für die Mehrdeutigkeit im Allgemeinen sowie für die exponentielle Mehrdeutigkeit finden. Es ist allerdings ein offenes Problem, ob es Kriterien für die polynomielle Mehrdeutigkeit gibt.

[1]siehe [5, Seite 200, Theorem 8.9].
[2]siehe [6, 10, 1]; [5, Seite 100; Lemma 4.6]

Literaturverzeichnis

[1] P. Flajolet, *Analytic models and ambiguity of CFLs*, Theoretical Computer Science (1987), no. 49, 283–309.

[2] M. A. Harrison, *Introduction to Formal language Theory*, Addison–Wesley, 1978.

[3] S. Heilbrunner, *Parsing automata approach to LR theory*, Theoret. Comput. Sci. (1981), no. 15, 117–157.

[4] C. Herzog, *Kellerautomaten mit beschränktem Nichtdeterminismus oder beschränkter Mehrdeutigkeit*, Master's thesis, Fachbereich Informatik, J.-W.–Goethe–Universität Frankfurt am Main, 1994.

[5] J. D. Ullman J. E. Hopcroft, *Introduction to Automata Theory, Languages and Computation*, Addison–Wesley, 1979.

[6] R. Kemp, *Mehrdeutigkeit kontextfreier Sprachen*, LNCS **14** (1974), 534–546.

[7] ______, *Skript zur Vorlesung Informatik IV*, Fachbereich Informatik, J.–W.–Goethe–Universität Frankfurt am Main, SS1997.

[8] H. Maurer, *The existence of context–free languages which are inherently ambiguous of any degree*, Department of Mathematics Research Series; University of Calgary, (1968).

[9] A. V. Aho & J. D. Ullman, *The Theory of Parsing, Translation and Compiling*, Prentice Hall, 1972.

[10] K. Wich, *Kriterien für die Mehrdeutigkeit kontextfreier Grammatiken*, Master's thesis, Fachbereich Informatik; Universität Frankfurt am Main, 1997.

[11] T. Ottmann; P. Widmayer, *Algorithmen und Datenstrukturen*, 3. auflage ed., Spektrum Akademischer Verlag, 1996.

[12] D. Wotschke, *Skript zur Vorlesung Informatik IV*, Fachbereich Informatik, J.–W.–Goethe–Universität Frankfurt am Main, SS1996.

Stichwortverzeichnis

Special characters

A

C

D

G

H

I

K

L

M