

## Wörter ohne Wiederholungen

von

Jean Berstel

*"Für die Entwicklung der logischen Wissenschaften wird es, ohne Rücksicht auf etwaige Anwendungen, von Bedeutung sein, ausgedehnte Felder für Spekulation über schwierige Probleme zu finden."*

*Axel Thue, 1912*

### 1. Einleitung

Mit diesem Satz beginnt Axel Thue's zweite Arbeit über wiederholungsfreie Folgen. Das dort mit Überzeugung ausgedrückte Bekenntnis zur Grundlagenforschung könnte wörtlich in einem Ansuchen um Forschungsmittel bei irgendeinem Gremium übernommen werden. Immerhin sind aber jetzt schon über 70 Jahre seit dem Erscheinen der grundlegenden Arbeiten von Thue vergangen, und der geduldigste Geldgeber darf fragen, wozu das Ganze nütze ist. Bevor also auf technische Einzelheiten eingegangen wird, sollen "etwaige Anwendungen" kurz skizziert werden.

Die erste Anwendung ist wohl von M. Morse [28], Er hat die von Thue 1906 beschriebene Folge wiederentdeckt und einige ihrer Eigenschaften studiert. Es ging ihm um die Existenz einer Folge, deren Teilwörter genügend oft vorkommen. Daraus konnte er Schlüsse für die symbolische Dynamik ziehen. Das zweite Zitat wiederholungsfreier Folgen steht historisch im Zusammenhang mit der Burnside'schen Vermutung. Im sehr langen Beweis der Existenz von Gruppen, die dieser Vermutung widersprechen, wird am Ende eine unendliche Folge ohne dritte Potenz gebraucht. (siehe Adjan [2], auch Hall [19]). In den letzten Jahren ist die Existenz wiederholungsfreier Folgen immer weiteren Kreisen bekannt geworden, und man findet die eine oder die andere dieser Folgen in verschiedenen Arbeiten, vor allem der theoretischen Informatik, zitiert. Sie werden meistens als Beispiele oder als Gegenbeispiele verwendet. Dies ist der Fall für Brzozowski et al. [7], Goldstine [18], Shyr [44] und andere.

Das Interesse für wiederholungsfreie Folgen ist in den letzten Jahren bedeutend angewachsen. Die Problematik hat jetzt wohl endgültig den Kreis der amüsanten Probleme für Sonntagsbeilagen verlassen. In der theoretischen Informatik, die ja unter anderem die verschiedenen Erzeugungs- bzw. Analysemethoden von Wörtern untersucht, haben insbesondere Algorithmen zur Erkennung von Wiederholungen (Apostolico, Preparata [3], Crochemore [11], [14], Main, Lorentz [26,27]) sowie Homomorphismen mit vorgeschriebenen Eigenschaften, wie weiter unten beschrieben, Anklang gefunden.

Die Theorie der wiederholungsfreien Folgen hat sich zu einem stattlichen Teilgebiet der Kombinatorik entwickelt, mit Verzweigungen zur Zahlentheorie (siehe z.B. Christol, Kamae, Mendes-France, Ranzy [10]) und zur Theorie der formalen Sprachen. In diesem Artikel sollen einige neuere Ergebnisse vorgestellt werden. Es gibt inzwischen mehrere Übersichtsarbeiten, und zwar von Salomaa [36, 37], und von Lothaire [25]. Eine kurze historische Beschreibung ist auch von Hedlund [21] gegeben worden. Siehe auch die entsprechenden Arbeiten im kürzlich erschienenen Tagungsband, der von Cummings [15] herausgegeben worden ist, und Berstel [5].

Diese Übersichtsarbeit ist wie folgt gegliedert: nach allgemeinen Definitionen in Teil 2 gibt der Teil 3 einige Ergebnisse für überlappungsfreie Wörter wieder. Teil 4 enthält die Beschreibung quadratfreier Wörter, und Teil 5 quadratfreier Morphismen. Einige Verallgemeinerungen werden im letzten Teil gebracht.

## 2. Bezeichnungen

Ein Alphabet  $A$  ist eine Menge, deren Elemente Buchstaben genannt werden. Ein Wort ist die Folge von Buchstaben

$$w = a_1 a_2 \dots a_n, \quad \text{mit } a_i \in A.$$

Die Länge  $n$  des Wortes  $w$  wird mit  $|w|$  bezeichnet.

Das freie Monoid aller Wörter über dem Alphabet  $A$  bezeichnet man mit  $A^*$ .

Ein (einseitig) unendliches Wort ist eine Abbildung

$$\underline{a} : \mathbb{N} \rightarrow A$$

geschrieben als

$$\underline{a} = a_0 a_1 a_2 \dots a_{11} \dots$$

Ein Faktor eines (endlichen oder unendlichen) Wortes ist ein (endliches) Wort, das in dem genannten Wort mindestens einmal vorkommt. Ein Quadrat ist ein Wort der Form

$$uu$$

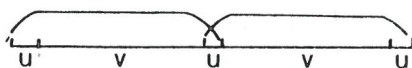
mit  $u$  verschieden vom leeren Wort.

Eine Überlappung ist ein Wort der Form

$$uvuvu$$

wobei das Wort  $u$  nicht leer ist. Der Begriff kommt daher, daß im Wort  $uvuvu$  der Faktor  $uvu$  zweimal auftritt, und daß sich die zwei Okkurrenzen überlappen (siehe Fig. 2.1).

Fig. 2.1 :



Eine Überlappung

BEISPIEL: Über dem Alphabet  $\{0,1\}$  ist

$$011001100$$

eine Überlappung.

Ein Wort ist quadratfrei, wenn keiner seiner Faktoren ein Quadrat ist. Analog heißt ein Wort überlappungsfrei, wenn keiner seiner Faktoren eine Überlappung ist. So ist z. B. das Wort

$$011010011010100$$

bestimmt nicht quadratfrei, wohl aber überlappungsfrei.

Der Begriff einer Potenz kann allgemeiner gefaßt werden. So nennt man Sesquipotenz ein Wort der Form

$$(uv)^n u .$$

So ein Wort ist eine rationale Potenz mit gebrochenem Exponenten

$$n + s$$

wobei

$$0 \leq s = \frac{|u|}{|uv|} < 1$$

Ein Morphismus ist eine Abbildung

$$\alpha : A^* \rightarrow B^*$$

so daß für alle

$$u = a_1 a_2 \dots a_n \quad a_i \in A$$

gilt

$$\alpha(u) = \alpha(a_1)\alpha(a_2)\dots\alpha(a_n)$$

Einen Morphismus

$$\alpha : A^* \rightarrow A^*$$

kann man iterieren, insbesondere, wenn für einen Buchstaben  $a \in A$  das Bild  $\alpha(a)$  auch mit  $a$  beginnt. Dann ist

$$\alpha(a) = au$$

für ein Wort  $u$ , und man sieht leicht ein, daß allgemein jedes Wort  $\alpha^{n+1}(a)$  mit  $\alpha^n(a)$  beginnt. Daraus ergibt sich, daß die Folge

$$(\alpha^n(a))_{n \geq 0}$$

zu einem (endlichen oder unendlichen) Wort konvergiert, das mit

$$\alpha^\omega(a)$$

bezeichnet und das durch Iteration von  $\alpha$  auf  $a$  erzeugte Wort genannt wird.

### 3. Überlappungsfreie Wörter

Es sei  $A = \{0,1\}$  und der Morphismus

$$\mu : A^* \rightarrow A^*$$

gegeben durch

$$\mu(0) = 01$$

$$\mu(1) = 10$$

Iteriert man den Morphismus  $\mu$ , so erhält man zwei Folgen von Wörtern

$$u_n = \mu^n(0),$$

$$v_n = \mu^n(1)$$

deren Werte verbunden sind durch die Formeln

$$u_{n+1} = u_n v_n, v_{n+1} = v_n u_n \quad (n \geq 0)$$

Die ersten Elemente der Folgen sind

$$u_0 = 0$$

$$v_0 = 1$$

$$u_1 = 01$$

$$v_1 = 10$$

$$u_2 = 0110$$

$$v_2 = 1001$$

$$u_3 = 01101001$$

$$v_3 = 10010110$$

$$u_4 = 0110100110010110 \quad v_4 = 1001011001011001$$

Im Sinne der obengenannten Konstruktion erzeugt ein (eigentlich zwei) unendliches Wort  $\mu^\omega(0)$ , das das Thue-Morse Wort genannt und mit  $m$  bezeichnet wird.

$$\underline{m} = \mu^\omega(0) = 011010011001011010010110\dots$$

Folgende Eigenschaft der Thue-Morse Folge erlaubt eine direkte Berechnung ihrer Elemente.

BEOBACHTUNG: (Morse [28]). Es sei  $m_k$  der k-te Buchstabe des Wortes m, wobei die Numerierung mit 0 anfängt. Dann ist  $m_k = 0$  oder  $m_k = 1$ , je nachdem ob die Anzahl der "1" in der Binärentwicklung von k gerade oder ungerade ist.

Für uns ist folgender Satz interessant, der die Existenz unendlicher überlappungsfreier Folgen zeigt

SATZ: (Thue [45], Morse [28]). Das Wort m ist überlappungsfrei.

Die Einfachheit der Konstruktion und die Vielzahl der Eigenschaften der Thue-Morse-Folge hat einerseits zu Verallgemeinerungen angeregt (siehe z.B. Cerny [9]), andererseits ein genaueres Studium spezieller Fragen ermöglicht. So sind z.B. überlappungsfreie Quadrate, wie die im Anfangsstück der Thue-Morse-Folge unterstrichenen Blöcke, jetzt wohlbekannt.

Fig. 3.1  $m = \underline{0110} \underline{1001} \underline{1001} \underline{0110} \underline{1001} \underline{0110} \dots$  *Quadrate in der Thue-Morse-Folge*

SATZ: Quadrate, die in der Thue-Morse-Folge vorkommen, haben die Länge  $2^n$  oder  $3 \cdot 2^n$ , für ein  $n \geq 0$ .

Die Beobachtung ist vor kurzem bedeutend verschärft worden. Einerseits wurde folgender Satz bewiesen:

SATZ: (Shelton, Soni) [43] Jedes überlappungsfreie Quadrat über zwei Buchstaben hat die Länge  $2^n$  oder  $3 \cdot 2^n$  für eine ganze Zahl  $n \geq 0$ .

Andererseits hat Harju eine vollständige Beschreibung zyklisch überlappungsfreier Wörter gegeben (siehe auch Thue [46]), woraus sofort folgt:

SATZ: (Harju [20]) Über zwei Buchstaben 0 und 1 sind genau die folgenden Wörter überlappungsfreie Quadrate:

$$u_n^2, v_n^2, (u_n v_n u_n)^2, (v_n u_n v_n)^2 \quad (n \geq 0)$$

un die dazu konjugierten Wörter.

Eine genaue Analyse überlappungsfreier Wörter über zwei Buchstaben ist Restivo, Salemi [43] gelungen. Sie haben beweisen können, daß ein überlappungsfreies Wort im wesentlichen durch ein paar Buchstaben schon festgelegt ist. Es soll hier nicht der genaue Wortlaut wiedergegeben werden, sondern der Sinn ihres Satzes erläutert werden. Man setze

$$a_n = u_{2n} \quad b_n = v_{2n} \quad n \geq 0$$

das heißt

$$\begin{array}{ll} a_0 = 0 & b_0 = 1 \\ a_1 = 0110 & b_1 = 1001 \\ a_2 = 0110100110010110 & \dots \\ \dots & \dots \end{array}$$

SATZ: (Restivo, Salemi) Jedes überlappungsfreie Wort  $w$  über  $\{0,1\}$  hat eine eindeutige Zerlegung der Form

$$w = g_0 g_1 \dots g_{k-1} u d_{k-1} \dots d_1 d_0$$

mit

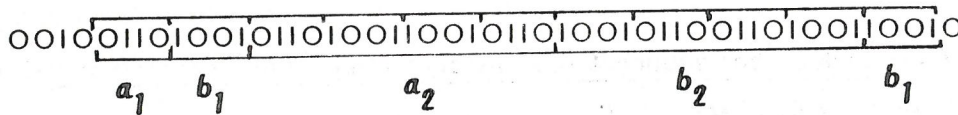
$$g_i \in G_i, d_i \in D_i \quad i = 0, \dots, k-1$$

und  $u$  ein Produkt von mindestens 2 und höchstens 11 Wörtern in  $\{a_k, b_k\}$ . Dabei sind  $G_i$  und  $D_i$  die Mengen von Wörtern die durch Knoten in den Bäumen der Figur 3.2 dargestellt sind.



Fig. 3.2: Die zwei Mengen  $G_i$  und  $D_i$

BEISPIEL: Beim folgenden überlappungsfreien Wort ist  $k = 2$ .



Das "Innere" eines überlappungsfreien Wortes hat also eine festgelegte Form. Es ist nämlich ein Faktor des Thue-Morse-Wortes. Aus ihrer Charakterisierungssatz haben Restivo, Salemi in der gleichen Arbeit folgendes Korollar erhalten.

SATZ: Es sei  $\gamma_n$  die Anzahl der überlappungsfreien Wörter der Länge  $n$  über 2 Buchstaben  
Dann gilt

$$\gamma_n \leq C n^{\log_2 15}$$

Wir erwähnen zum Abschluß dieses Teiles, noch einige Ergebnisse, die die Konstruktion überlappungsfreier Wörter betreffen. Die Thue-Morse-Folge wird durch einen einfachen Morphismus erzeugt. Man kann sich fragen, ob es einen anderen, einfacheren oder eventuell komplizierteren Morphismus gibt, der diese Folge im gleichen Sinne erzeugt. Die Antwort ist im wesentlichen: nein.

SATZ: (Pansiot [31]) Es sei  $h$  ein Morphismus, der durch Iteration die Thue-Morse-Folge erzeugt:

$$h^{\omega}(0) = \underline{m} = 01101001\dots = \mu^{\omega}(0)$$

Dann ist  $h$  eine Potenz von  $\mu$ .

Der Satz kann etwas anders formuliert werden, indem man die Menge  $M$  aller Morphismen  $h$  betrachtet, für die  $h(\underline{m}) = \underline{m}$  gilt. Natürlich ist  $M$  ein Monoid für die Komposition von Morphismen, und der Satz von Pansiot besagt, daß  $M$  zyklisch und von  $\mu$  erzeugt ist.

Eine etwas schärfere Formulierung des Satzes wurde von Harju [20] gegeben. Ein Morphismus  $h : A^* \rightarrow B^*$  heiße überlappungsfrei wenn er überlappungsfreie Wörter in überlappungsfreie Wörtern abbildet. Natürlich ist der Morphismus  $\mu$  überlappungsfrei, und wenn ein überlappungsfreier Morphismus durch Iteration ein unendliches Wort erzeugt, dann ist dieses überlappungsfrei. Die umgekehrte Behauptung gilt auch nach folgendem Satz.

SATZ: (Harju) Es sei  $h : \{0,1\}^* \rightarrow \{0,1\}^*$  ein überlappungsfreier Morphismus. Dann ist

$$h = \mu^n \text{ oder } h = \pi\mu^n$$

für ein  $n \geq 0$ , wobei  $\pi$  der Isomorphismus ist, der 0 und 1 vertauscht.

Beide Sätze sind Konsequenzen des folgenden Satzes

SATZ: (Séébold [38,39,40]) Es sei  $a$  ein unendliches überlappungsfreies Wort, das durch einen Morphismus  $h$  über zwei Buchstaben erzeugt ist:

$$\underline{a} = h^\omega(0)$$

Dann ist  $a = \underline{m}$  die Thue-Morse-Folge.

Der Satz besagt also, daß es nur ein unendliches Wort gibt, das sowohl überlappungsfrei als auch durch Iteration eines Morphismus erzeugt ist. Zur besseren Handhabung überlappungsfreier Wörter müssen also bessere Werkzeuge bereitgestellt werden.

#### 4. Quadratfreie Wörter

Unsere Kenntnis von quadratfreien Wörtern ist bei weitem nicht so entwickelt wie für überlappungsfreie Wörter. Wenn man quadratfreie Wörter über zwei Buchstaben, z.B. 0 und 1, sucht, dann ist man schnell fertig. Es gibt nur 6 solche Wörter, nämlich

$$0, 01, 010 \text{ und } 1, 10, 101$$

Die Lage ist ganz anders, sobald man einen dritten Buchstaben hinzunimmt – 1912 hat Thue folgenden Satz gezeigt:

SATZ: (Thue [46]) Es gibt unendliche quadratfreie Wörter über drei Buchstaben.

Von der von A. Thue eingeführten quadratfreien Folge sind mir vier verschiedene Konstruktionen bekannt. Sie scheinen mir erwähnenswert, weil sie die Vielzahl der Beschreibungsmöglichkeiten gut illustrieren.

A. Thue's Konstruktion beginnt mit dem Wort

$$\underline{m} = 0110100110010110 \dots$$

in dem, weil es überlappungsfrei ist, niemals ein Faktor 111 auftritt. Also kann  $\underline{m}$  (eindeutig) zerlegt werden in Blöcke, von denen jeder O, OI oder OII ist. Betrachtet man den Morphismus

$$h: \begin{array}{l} a \rightarrow OII \\ b \rightarrow OI \\ c \rightarrow O \end{array}$$

so erhält man das unendliche Wort

$$\underline{t} = h^{-1}(\underline{m}) = abcacbabcbac \dots$$

von dem man leicht beweisen kann, daß es quadratfrei ist.

Morse und Hedlund [29] sind anders vorgegangen. Auch sie gehen von dem Wort  $\underline{m}$  aus, interpretieren aber je zwei aufeinanderfolgende Buchstaben von  $\underline{m}$  als Binärdarstellung einer der Zahlen 0,1,2,3. Sie erhalten also aus

$$\underline{m} = 0110100110010110 \dots$$

das Wort

$$\tau(\underline{m}) = \underline{m}' = 132120132012132 \dots$$

wobei  $\tau$  durch

$$\tau: \begin{array}{l} OO \rightarrow 0 \\ OI \rightarrow 1 \\ IO \rightarrow 2 \\ II \rightarrow 3 \end{array}$$

gegeben ist. In einem zweiten Schritt werden jetzt 0 und 3 identifiziert. Genauer, sei

$$\beta: \begin{array}{l} 0 \rightarrow b \\ 1 \rightarrow a \\ 2 \rightarrow c \\ 3 \rightarrow b \end{array}$$

Dann ist die endgültige Folge das Wort

$$\beta(\underline{m}') = \beta(\tau(\underline{m})) = abcacbabcbac \dots = \underline{t}$$

die trotz der völlig verschiedenen Konstruktion, die gleiche Folge wie die Thue's ist!

Man kann abschließend noch bemerken, daß  $\underline{m}'$  auch durch Iteration eines Morphismus' erhältlich ist:

$$\underline{m}' = \alpha^{\omega}(1)$$



wobei

$$\alpha : \begin{array}{l} 0 \rightarrow 12 \\ 1 \rightarrow 13 \\ 2 \rightarrow 20 \\ 3 \rightarrow 21 \end{array}$$

Die dritte Konstruktion, die bei M. Hall [19] zu finden ist, und von Istreil [22] genau beschrieben wurde, ist wohl die einfachste. Es sei  $\sigma$  der durch

$$\sigma : \begin{array}{l} a \rightarrow abc \\ b \rightarrow ac \\ c \rightarrow b \end{array}$$

definierte Morphismus. Dann gilt

$$\sigma^\omega(a) = \underline{t}$$

Die letzte Beschreibung ist eine unveröffentlichte Konstruktion von G. Cousineau. Man betrachte folgende Wortfolge

$$\begin{array}{l} w_0 = \epsilon \\ w_{n+1} = w_n a \tilde{w}_n b \tilde{w}_n c w_n \end{array} \quad n \geq 0$$

( $\tilde{w}$  bezeichne das zu  $w$  gespiegelte Wort).

Dann gilt

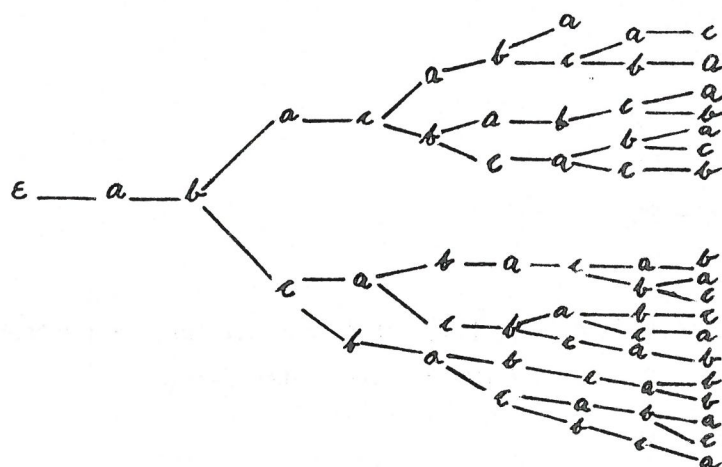
$$\underline{t} = \lim_{n \rightarrow \infty} w_n.$$

Die ersten Wörter dieser Folge sind

$$\begin{array}{l} w_0 = \epsilon \\ w_1 = abc \\ w_2 = abcacbabcabacabc \end{array}$$

Aus dieser Konstruktion ergibt sich sofort eine interessante Eigenschaft der Folge  $\underline{t}$ : wenn  $w$  ein Faktor von  $\underline{t}$  ist, dann ist das Spiegelbild von  $w$  auch Faktor von  $\underline{t}$ .

Nachdem jetzt ein unendliches quadratfreies Wort über drei Buchstaben konstruiert ist, fragt man nach der allgemeinen Form endlicher oder unendlicher quadratfreier Wörter. Dazu eignet sich die Darstellung dieser Wörter durch einen Baum.



1 1 2 3 5 7 10 13 18

Fig 4.1: Der Baum der quadratfreien Wörter mit Anfang ab .

Die Anzahl der quadratfreien Wörter wächst anscheinend nicht sehr rasch. Wenn man mit  $c(n)$  die Anzahl der quadratfreien Wörter der Länge  $n$  über drei Buchstaben bezeichnet, dann sind Werte in Fig. 4.2 angegeben (aus Brandenburg [6]).

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	... 24
$c(n)$	3	6	12	18	30	42	60	78	108	144	204	264	342	456	7032

Fig. 4.2: Anzahl quadratfreier Wörter über 3 Buchstaben

SATZ: (Brandenburg [6]) Die Anzahl  $c(n)$  quadratfreier Wörter der Länge  $n$  über drei Buchstaben wächst exponentiell an. Es gilt

$$6 \cdot c_1^n \leq c(n) \leq 6 \cdot c_2^n$$

mit  $c_1 \geq 1.032$  und  $c_2 \leq 1.38$  .

Die Lage ist hier also verschieden von dem Wachstum überlappungsfreier Wörter, wie im Satz von Restivo und Salemi gegeben. Eine Begründung dieses Unterschiedes fehlt noch.

Genaueres über die Struktur des Baumes quadratfreier Wörter haben Shelton [41] und Shelton, Soni [42] gefunden. Dazu braucht man zwei Bezeichnungen. Ein Ast (Fig.4.3) ist ein endlicher Teilbaum des Baumes quadratfreier Wörter, und ein Kern ist ein Weg in diesem Baum von dem nur Äste abgehen (Fig. 4.4)

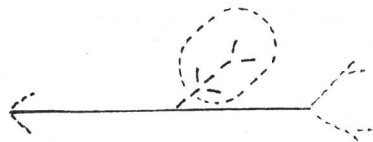


Fig. 4.3: Ein Ast

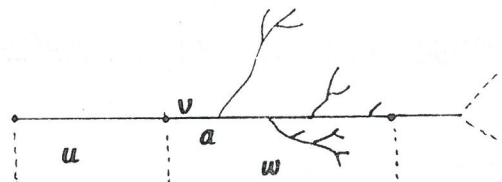


Fig. 4.4: Ein Kern

Genauer gesagt, ist ein Kern ein Paar  $c = (u, w)$  von Wörtern so daß Folgendes gilt

1.  $uw$  ist ein quadratfreies Wort
2. für jeden Linksfaktor  $va$  von  $w$ , mit  $a$  einem Buchstaben, ist die Menge der quadratfreien Wörter, die mit  $uvb$  ( $b \neq a$ ) beginnen, endlich.

Man bezeichnet die Länge von  $u$  als den Index des Kerns und die Länge von  $w$  als die Länge des Kerns.

SATZ: Es gibt eine Konstante  $A$  so daß für jeden Kern  $c=(u,w)$  die Ungleichung

$$|u| \geq A \cdot |w|^{3/2}$$

gilt.

Der Satz hat zwei sehr interessante Korollare.

SATZ: Die Menge quadratfreier Wörter über drei Buchstaben ist perfekt: wenn ein unendliches quadratfreies Wort mit dem (endlichen) Wort  $u$  beginnt, dann gibt es unendlich viele unendliche quadratfreie Wörter, die mit  $u$  beginnen.

SATZ: Es gibt eine Konstante  $K$  so daß Folgendes gilt: wenn ein quadratfreies Wort  $u$  verlängert werden kann in ein quadratfreies Wort  $uv$  der Länge  $|u| + K \cdot |u|^{3/2}$ , dann gibt es ein unendliches quadratfreies Wort, das mit  $u$  beginnt.

Als formale Sprache ist die Menge der quadratfreien Wörter natürlich weder rational (regulär), noch kontextfrei. Das zeigt man leicht mit den sog. Iterations- oder Pumping-lemmas. Eine schöne Herausforderung an die Spitzfindigkeit der Spezialisten der formalen Sprachen war die Vermutung, daß das Komplement dieser Sprache, also die Sprache aller Wörter, die ein Quadrat enthalten, auch nicht kontextfrei ist. Diese von Autebert, Beauquier, Boasson, Nivat [1] aufgestellte Vermutung ist jetzt von Ehrenfeucht, Rozenberg [17] und von Ross, Winkelmann [35] bewiesen. Ein schöner Beweis findet sich in der Arbeit von Ogden, Ross, Winkelmann [30].

SATZ: Die Menge der Wörter, die ein Quadrat enthalten, ist nicht kontextfrei.

## 5. Quadratfreie Morphismen

Ein Morphismus

$$h : A^* \rightarrow B^*$$

heißt quadratfrei wenn das Bild eines quadratfreien Wortes wieder quadratfrei ist. Zur Konstruktion quadratfreier Wörter sind solche Morphismen sehr willkommen.

BEISPIELE: 1. A. Thue [46] hat gezeigt, daß der durch

$$\begin{aligned} a &\rightarrow abcab \\ b &\rightarrow acabcb \\ c &\rightarrow acbcacb \end{aligned}$$

gegebene Morphismus quadratfrei ist.

2. Hingegen ist der Morphismus

$$\begin{aligned} a &\rightarrow abc \\ h : b &\rightarrow ac \\ c &\rightarrow b \end{aligned}$$

nicht quadratfrei, obwohl das unendliche Wort

$$\underline{t} = h^\omega(a)$$

quadratfrei ist. Es gilt nämlich

$$h(aba) = ab\underline{c}a\underline{c}abc.$$

Die zwei Tatsachen stehen nicht im Widerspruch, weil  $aba$  im Wort  $\underline{t}$  nicht vorkommt.

Ein quadratfreier Morphismus kann nicht einfach sein. Bezeichnet man mit der Größe eines Morphismus  $h$  die Zahl

$$\sum_{a \in A} |h(a)|$$

dann gilt nämlich folgender Satz:

SATZ: (Carpi [8]) Jeder quadratfreie Morphismus über drei Buchstaben, der keine Permutation des Alphabets ist, hat mindestens Größe 18.

Der von A. Thue gegebene Morphismus hat Größe 18, also war Thue auch da schon optimal.... Der 2. Morphismus hat hingegen nur Größe 6. Viele Autoren, mit Thue angefangen, haben hinreichende Bedingungen für quadratfreie Morphismen angegeben. Die genauesten hinreichenden und notwendigen Bedingungen stammen von M. Crochemore [12,13] der die folgenden Sätze zeigt.

SATZ: Ein Morphismus  $h : A^* \rightarrow B^*$  ist genau dann quadratfrei, wenn

- (i)  $h(x)$  für alle quadratfreien Wörter  $x$  der Länge 3 quadratfrei ist, und
- (ii) keines der Wörter  $h(a)$ ,  $a \in A$ , ein internes Präquadrat enthält.

Ein Wort  $u$  heißt dabei ein internes Präquadrat von  $h(a)$  wenn  $u$  ein Faktor von  $h(a)$  ist und es ein Wort  $x$  gibt so daß  $xa$  quadratfrei ist, jedoch  $h(xa)$  das Quadrat  $uu$  enthält

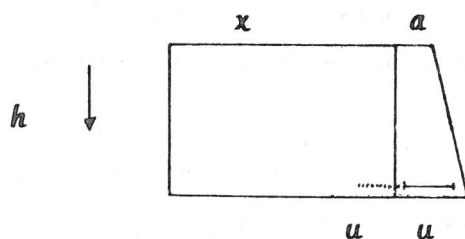


Fig. 5.1: Internes Präquadrat

Die Bedingungen des vorigen Satzes sind entscheidbar. Genauer gesagt, setze man

$$m(h) = \min \{ |h(a)| : a \in A \}$$

$$M(h) = \max \{ |h(a)| : a \in A \}$$

Dann gilt:

SATZ: Es sei  $h : A^* \rightarrow B^*$  ein Morphismus mit  $m(h) \neq 0$ .  $h$  ist genau dann quadratfrei, wenn  $h(x)$  quadratfrei ist für jedes quadratfreie Wort  $x$  der Länge

$$k = \max \left\{ 3, 1 + \left\lceil \frac{M(h)-3}{m(h)} \right\rceil \right\}$$

Für den interessanten Spezialfall von 3 Buchstaben kann dieser Satz noch verfeinert werden. Man erhält

SATZ: Es sei  $h : A^* \rightarrow B^*$  ein Morphismus mit  $\text{Card}(A) = 3$ .  $h$  ist genau dann quadratfrei, wenn  $h(x)$  quadratfrei ist für alle quadratfreien Wörter  $x$  der Länge 5.

## 6. Verallgemeinerungen

Von den möglichen Verallgemeinerungen wollen wir drei kurz und eine vierte etwas länger diskutieren. Zuerst kann man das Suchen nach quadratfreien Wörtern durch suchen nach Wörtern ohne kommutative Quadrate ersetzen. Ein kommutatives Quadrat ist ein Wort  $uv$ , wobei  $v$  eine Permutation von  $u$  ist. Über 3 Buchstaben haben alle Wörter der Länge 9 ein kommutatives Quadrat; nach Pleasants [33] weiß man, daß es unendliche Wörter ohne kommutative Quadrate über 5 Buchstaben gibt. Der Fall von 4 Buchstaben ist offen. Man kann auch die Kommutativität durch eine partielle Kommutativität zu ersetzen versuchen. Eine andere Art der Verallgemeinerung, die wir aber hier nicht diskutieren wollen, stammt von Bean, Ehrenfeucht und McNulty [4]. Es handelt sich um vermeidbare bzw. unvermeidbare Muster, die durch Wörter beschrieben sind.

Höhere Potenzen sind die einfachste Art der Verallgemeinerung. Einige hinreichende Bedingungen für Morphismen, die Wörter ohne  $k$ -te Potenz in Wörter gleicher Art überführen, stehen in der Arbeit von Bean et. al. Hier soll folgendes schöne Ergebnis von Karhumäki [23] zitiert sein:

SATZ: Es sei  $A = \{a,b\}$ , und  $h : A^* \rightarrow A^*$ ; ferner beginne  $h(a)$  mit dem Buchstaben  $a$ . Das unendliche Wort  $h(a)$  ist genau dann ohne dritte Potenz, wenn  $h^{10}(a)$  ohne dritte Potenz ist.

Ein Morphismus  $h : A^* \rightarrow B^*$  heiße potenzfrei, wenn für alle  $n \geq 2$  jedes Wort  $h(w)$  keine  $n$ -te Potenz enthält wenn  $w$  keine  $n$ -te Potenz enthält. Dann gilt

SATZ: (Leconte [24]) . Ein Morphismus  $h : A^* \rightarrow B^*$  ist genau dann potenzfrei, wenn  $h$  quadratfrei und wenn  $h(aa)$  kubusfrei für jeden Buchstaben  $a \in A$  ist.

Zum Schluß soll noch auf "kleinere" Potenzen eingegangen werden. Wie schon am Anfang dieser Arbeit erklärt, kann man ja für rationale Zahlen  $r \in \mathbb{Q}_+$  den Begriff der rationalen Potenz gewisser Wörter definieren, indem man die rationale Zahl  $r$  schreibt als

$$r = n + \frac{p}{q}, \text{ für } n \in \mathbb{Z}, 0 \leq p \leq q$$

Dann ist eine  $r$ -te Potenz ein Wort der Form

$$w = (uv)^n u \text{ mit } \frac{|u|}{|uv|} = \frac{p}{q}$$

Wir sagen, daß  $r$ -te Potenzen über  $q$  Buchstaben vermeidbar (bzw. unvermeidbar) sind, wenn es ein unendliches Wort ohne  $r$ -te Potenzen gibt (bzw. wenn jedes Wort, das lang

genug ist, eine  $r$ -te Potenz enthält). Wenn  $r$ -te Potenzen über  $q$  Buchstaben vermeidbar sind, dann sind auch  $r'$ -te Potenzen vermeidbar für  $r' > r$ . Das Gleiche gilt, mutatis mutandis, für unvermeidbare Potenzen. Folglich kann man setzen:

$$s(q) = \min \{r \in \mathbb{Q}_+ \mid r\text{-te Potenzen sind über } q \text{ Buchstaben vermeidbar}\}$$

$$S(q) = \max \{r \mid r\text{-te Potenzen sind über } q \text{ Buchstaben vermeidbar}\}$$

Natürlich gilt

$$s(q) = S(q)$$

und natürlich wird, für jedes  $q$ , nur einer der zwei Grenzwerte erreicht.

Die Berechnung der Zahlen  $s(q)$  ist einfach für  $q = 2$ , weil wegen der Unvermeidbarkeit der Quadrate  $S(2) \geq 2$ , wegen der Vermeidbarkeit jeder höheren Potenz, die durch die Existenz der Thue-Morse-Folge erwiesen ist, aber  $s(2) < r$  für alle  $r > 2$  gilt. Also gilt  $s(2) = 2$ .

Der Wert  $s(3)$  wurde von F. Dejean [16] bestimmt. Sie hat berechnet, daß jedes Wort der Länge 39 über 3 Buchstaben eine  $(\frac{7}{4})$ -te Potenz enthält, also einen Faktor der Form  $uvu$ , mit  $|u| = 3|v|$ . Das heißt, daß  $\frac{7}{4}$ -te Potenzen unvermeidbar und deshalb  $S(3) \geq \frac{7}{4}$  ist. Andererseits hat sie folgenden Satz bewiesen:

SATZ: Es gibt ein unendliches Wort über 3 Buchstaben, das keine  $r$ -te Potenz für  $r < \frac{7}{4}$  enthält.

Das heißt also, daß  $s(3) > r$  für alle  $r < \frac{7}{4}$  ist, und deshalb gilt  $s(3) = \frac{7}{4}$ .

Durch Einführung neuer Methoden ist es Pansiot [32] gelungen, auch den Wert für  $q = 4$  zu bestimmen, und zwar gilt, wie schon von Dejean vermutet,  $s(4) = \frac{7}{5}$ . Für höhere Werte von  $q$  ist die Bestimmung von  $s(q)$  offen.

References:

=====

- [1] Autebert, J./ Beauquier, J./Boasson, L./Nivat, M.: Quelques problèmes ouverts en théorie des langages algébriques, RAIRO Informatique théorique 13 (1979), 363-379
- [2] Adjan, S.: The Burnside problem and identities in groups, Erg. Math. Grenzber. vol. 95, Springer 1979
- [3] Apostolico, A./Preparata, F.: Optimal off-line detection of repetitions in a string, Theor. Comp. Sci 22 (1983), 297-315
- [4] Bean, D./Ehrenfeucht, A./McNulty, G.: Avoidable patterns in strings of symbols Pacific J. Math. 85(1979) , 261-294
- [5] Berstel, J.: Some recent results on squarefree words, in: Fontet, Mehlhorn (eds), "STACS 84", Lecture Notes in Comp. Sc. 166, Springer 1984, 14 -25 .
- [6] Brandenburg, F.: Uniformly growing k-th power-free homomorphisms, Theor. Comput. Sci. 23 (1983), 69-82
- [7] Brzozowski, J./Culik II, K./Gabrielian, A.: Classification of noncounting events, J. Comp. Syst. Sci. 5 (1971), 41-53
- [8] Carpi, A.: On the size of a square-free morphism on a three letter alphabet, Inf. Proc. Letters 16 (1983), 231-236
- [9] Cerny, A.: On generalized words of Thue-Morse, Tech. Report, Universite Paris VI, L.I.T.P. 83-44.
- [10] Christol, C./Kmae, T./Mendès-France, M./Rauzy, G.: Suites algébriques, automates et substitutions, Bull. Soc. Math. France 108 (1980), 401-419
- [11] Crochemore, M: An optimal algorithm for computing the repetitions in a word, Inf. Proc. Letters, 12 (1981), 244-250
- [12] —,— Sharp characterizations of squarefree morphisms, Theor. Comp. Sci. 18, (1982), 221-226
- [13] —,— Mots et morphismes sans carré, Annals of Discr. Math. 17 (1983), 235-245
- [14] —,— Recherche linéaire d'un carré dans un mot, C.R. Acad. Sci. Paris, 296, (1983), 781-784
- [15] Cummings, L. (ed.): Combinatorics on words, Progress and Perspectives, Academic Press 1983
- [16] Dejean, F.: Sur un théorème de Thue, J. Combinatorial Theory 13, (1972), 90-99
- [17] Ehrenfeucht, A./Rozenberg, G.: On the separating power of EOL systems, RAIRO Informatique 17, (1983), 13-22
- [18] Goldstine, J.: Bounded AFLs, J. Comput. Syst. Sci. 12,(1976), 399-419
- [19] Hall, M.: Generators and relations in groups - the Burnside problem, in: T.L. Saaty (ed.) "Lectures in Modern Mathematics", vol. II, John Wiley 1964
- [20] Harju, T.: Morphisms that avoid overlapping, Univ. of Turku, May 1983
- [21] Hedlund, G.: Remarks on the work of A. Thue on sentences, Nord. Mat. Tidskr. 16, (1967), 148-1
- [22] Istrail, S.: On irreducible languages and non rational numbers, Bull. Mat. Soc. Sci. Mat. R.s. Roumanie 21, (1977), 301-308
- [23] Karhumäki, J.: On cube-free  $\omega$ -words generated by binary morphisms, Discr. Appl. Math. 5, (1983), 279-297



- [24] Leconte, M.: A fine characterization of power-free morphisms, erscheint in T.C.S.
- [25] Lothaire, M.: Combinatorics on Words, Addison-Wesley 1983
- [26] Main, M./Lorentz, R.: An  $O(n \log n)$  algorithm for recognizing repetition, Washington State Univ. Tech. Report CS-79-056
- [27] —, — An  $O(n \log n)$  algorithm for finding all repetitions in a string, J. Algorithms, 1983, to appear
- [28] Morse, M.: Recurrent geodesics on a surface of negative curvature, Trans Am. Math. Soc. 22, (1921), 84-100
- [29] —, — with Hedlund, G.: Unending chess, symbolic dynamics and a problem in semigroups, Duke Math. J. 11, (1944), 1-7
- [30] Ogden, W./Ross, R./Winkelmann, K.: An "Interchange lemma" for context-free languages, Washington State University, Tech. Rep. CS-82-080
- [31] Pansiot, J.: The Morse sequence and iterated morphisms, Inf. Proc. Letters 12, (1981), 68-70
- [32] —, — A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, Discr. Appl. Math.
- [33] Pleasants, P.: Nonrepetitive sequences, Proc. Cambridge Phil. Soc. 68, (1970), 267-274
- [34] Restivo, A./Salemi, S.: On weakly square free words, Bull EATCS 21, (1983), 49-56
- [35] Ross, R./Winkelmann, K.: Repetitive strings are not context-free, RAIRO Inform. théor. 16, (1982), 191-199
- [36] Salomaa, A.: Morphisms on free monoids and language theory, in: R. Book (ed.): "Formal language theory: perspectives and open problems", Academic Press 1980
- [37] —, — Jewels of Formal Language Theory, Pitman 1981
- [38] Seebold, P.: Morphisms itérés, mot de Morse et mot de Fibonacci, C.R. Acad. Sci. Paris, 295, (1982), 439-441
- [39] —, — Sur les morphismes qui engendrent des mots infinis ayant des facteurs prescrits, 6. GI Tagung Theor. Informatik, Lect. Not. Comp. Sci. 145, (1983), 301-311
- [40] —, — Sequences generated by infinitely iterated morphisms, Tech. Report Univ. de Paris VII, L.I.T.P 83-13
- [41] Shelton, R.: Aperiodic words on three symbols, I, II, J. reine angew. Math. 321, (1981), 195-209, 327, (1981), 1-11
- [42] —, — with Soni, R.: Aperiodic words on three symbols III, J. reine angew. Math. 330, (1982), 44-52
- [43] —, — Chains and fixing blocks in irreducible binary sequences, University of Tennessee 1982
- [44] Shyr, J.: A strongly primitive word of arbitrary length and its applications, Int. J. Comp. Math. A-6 (1977), 165-170
- [45] Thue, A.: Über unendliche Zeichenreihen, Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania (1906), 1-22
- [45] —, — Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania (1912), 1-67