Prof. Dr. Hans-Jörg Kreowski **Mai 2005**

Formale Sprachen: DNA Computing

Sticker-Systeme: ein Beispiel

Betrachte das folgende Sticker-System:

$$\gamma_1 = (\{a,b,c\}, \{(a,a),(b,b),(c,c)\}), \{\begin{bmatrix} a \\ a \end{bmatrix}\}, \{(\binom{b}{\lambda}, \binom{b}{\lambda}), (\binom{c}{\lambda}, \binom{\lambda}{\lambda}), (\binom{\lambda}{b}, \binom{\lambda}{\lambda}), (\binom{\lambda}{c}, \binom{\lambda}{b})\})$$

Mit den Bausteinen 1 und 2 lassen sich folgende Berechnungen bilden:

(1) $\begin{bmatrix} a \\ a \end{bmatrix} \overset{*}{\underset{1,2}{\Longrightarrow}} {}^{u}\begin{bmatrix} a \\ a \end{bmatrix}^{b^k}$ mit $count(b,u) = k$ & $u \in \{b,c\}^*$

und mit den Bausteinen 3 und 4 kann man Moleküle der Form ${}^{vw}\begin{bmatrix} a \\ a \end{bmatrix}^{b^l b^m}$ mit $w \in \{b,c\}^*$ und $count(c,w) = l$ bzgl. w und b^l auffüllen:

(2) ${}^{vw}\begin{bmatrix} a \\ a \end{bmatrix}^{b^l b^m} \overset{*}{\underset{3,4}{\Longrightarrow}} {}^{v}\begin{bmatrix} w\,a\,b^l \\ w\,a\,b^l \end{bmatrix}^{b^m}$

Damit gilt insbesondere:

(3) $\begin{bmatrix} a \\ a \end{bmatrix} \overset{*}{\underset{1,2}{\Longrightarrow}} {}^{u}\begin{bmatrix} a \\ a \end{bmatrix}^{b^k} \overset{*}{\underset{3,4}{\Longrightarrow}} \begin{bmatrix} u\,a\,b^k \\ u\,a\,b^k \end{bmatrix}$ für $u \in \{b,c\}^*$ und $count(b,u) = count(c,u) = k$

(4) $\{uab^k \mid u \in \{b,c\}^*, count(b,u) = count(c,u) = k\} \subseteq L(\gamma_1)$

Punkt 1 lässt sich durch Induktion über u zeigen.
IA: $\begin{bmatrix} a \\ a \end{bmatrix} \overset{0}{\underset{1,2}{\Longrightarrow}} \begin{bmatrix} a \\ a \end{bmatrix} = {}^{\lambda}\begin{bmatrix} a \\ a \end{bmatrix}^{b^0}$

IS: Nach IV gilt: $\begin{bmatrix} a \\ a \end{bmatrix} \overset{*}{\underset{1,2}{\Longrightarrow}} {}^{u}\begin{bmatrix} a \\ a \end{bmatrix}^{b^k}$. Diese Berechnung lässt sich durch die Bausteine 1 und 2 verlängern: ${}^{u}\begin{bmatrix} a \\ a \end{bmatrix}^{b^k} \underset{1}{\Longrightarrow} {}^{bu}\begin{bmatrix} a \\ a \end{bmatrix}^{b^{k+1}}$ und ${}^{u}\begin{bmatrix} a \\ a \end{bmatrix}^{b^k} \underset{2}{\Longrightarrow} {}^{cu}\begin{bmatrix} a \\ a \end{bmatrix}^{b^k}$. Das ergibt insgesamt $\begin{bmatrix} a \\ a \end{bmatrix} \overset{*}{\underset{1,2}{\Longrightarrow}} {}^{xu}\begin{bmatrix} a \\ a \end{bmatrix}^{b^l}$ mit $x \in \{b,c\}$ und $l = count(b,xu)$.

Punkt 2 ergibt sich analog mit Induktion über w.

Punkt 3 und 4 folgen unmittelbar aus Punkt 1 und 2.

Es gilt auch die Umkehrung von Punkt 4, d.h.

(5) $L(\gamma_1) \subseteq \{uab^k \mid u \in \{b,c\}^*, count(b,u) = count(c,u) = k\}$.

Denn $w \in L(\gamma_1)$ bedeutet nach Definition $\begin{bmatrix} a \\ a \end{bmatrix} \xRightarrow[1,2,3,4]{*} \begin{bmatrix} w \\ w \end{bmatrix}$. Diese Berechnung hat die Form von Punkt 3, falls alle Bausteine 1 und 2 vor allen Bausteinen 3 und 4 angewendet werden. Dann hat w die gewünschte Form.

Ansonsten gibt es eine Anwendung von 3 oder 4 direkt vor einer von 1 oder 2. Folgende vier Fälle können auftreten:

(i) $\begin{bmatrix} a \\ a \end{bmatrix} \xRightarrow{*} vb \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[3]{} v \begin{bmatrix} b\,x \\ b\,x \end{bmatrix} b^m \xRightarrow[1]{} bv \begin{bmatrix} b\,x \\ b\,x \end{bmatrix} b^{m+1} \xRightarrow{*} \begin{bmatrix} w \\ w \end{bmatrix}$

(ii) $\begin{bmatrix} a \\ a \end{bmatrix} \xRightarrow{*} vb \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[3]{} v \begin{bmatrix} b\,x \\ b\,x \end{bmatrix} b^m \xRightarrow[2]{} cv \begin{bmatrix} b\,x \\ b\,x \end{bmatrix} b^m \xRightarrow{*} \begin{bmatrix} w \\ w \end{bmatrix}$

(iii) $\begin{bmatrix} a \\ a \end{bmatrix} \xRightarrow{*} vc \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[4]{} v \begin{bmatrix} c\,x\,b \\ c\,x\,b \end{bmatrix} b^{m-1} \xRightarrow[1]{} bv \begin{bmatrix} c\,x\,b \\ c\,x\,b \end{bmatrix} b^m \xRightarrow{*} \begin{bmatrix} w \\ w \end{bmatrix}$

(iv) $\begin{bmatrix} a \\ a \end{bmatrix} \xRightarrow{*} vc \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[4]{} v \begin{bmatrix} c\,x\,b \\ c\,x\,b \end{bmatrix} b^{m-1} \xRightarrow[2]{} cv \begin{bmatrix} c\,x\,b \\ c\,x\,b \end{bmatrix} b^{m-1} \xRightarrow{*} \begin{bmatrix} w \\ w \end{bmatrix}$

In allen vier Fällen können die beiden Verlängerungsschritte in der Mitte vertauscht werden:

(i) $vb \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[1]{} bvb \begin{bmatrix} x \\ x \end{bmatrix} b^{m+1} \xRightarrow[3]{} bv \begin{bmatrix} b\,x \\ b\,x \end{bmatrix} b^{m+1}$

(ii) $vb \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[2]{} cvb \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[3]{} cv \begin{bmatrix} b\,x \\ b\,x \end{bmatrix} b^m$

(iii) $vc \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[1]{} bvc \begin{bmatrix} x \\ x \end{bmatrix} b^{m+1} \xRightarrow[4]{} bv \begin{bmatrix} c\,x\,b \\ c\,x\,b \end{bmatrix} b^m$

(iv) $vc \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[2]{} cvc \begin{bmatrix} x \\ x \end{bmatrix} b^m \xRightarrow[4]{} cv \begin{bmatrix} c\,x\,b \\ c\,x\,b \end{bmatrix} b^{m-1}$

Wenn man nach und nach alle Schrittpaare in falscher Reihenfolge umdreht, entsteht aus jeder kompletten Berechnung eine der Form von Punkt 3, so dass w die gewünschte Form hat. Denn beim Vertauschen wird die Zahl der Paare von Schritten, die falsch sortiert sind, kleiner, so dass nach spätestens quadratisch vielen Vertauschungen keine falsch sortierten Paare mehr existieren.

Mit Punkt 4 und 5 ergibt sich zusammenfassend:

(6) $L(\gamma_1) = \{uab^k \mid u \in \{b,c\}^*, count(b,m) = count(c,u) = k\}$.

Das ist ein signifikantes Resultat, weil sich so die Sprache $L(\gamma_1)$ als nicht kontextfrei erweist, was sich folgendermaßen einsehen lässt:

Angenommen, $L(\gamma_1)$ sei kontextfrei. Dann existiert nach dem Pumpinglemma für kontextfreie Sprachen eine Konstante p, derart dass für alle $z \in L((\gamma_1)$ mit $length(z) > p$ eine Zerlegung $z = uvwxy$ existiert mit $1 \leq length(vx) \leq length(vwx) \leq p$ und $uv^iwx^iy \in L(\gamma_1)$ für alle $i \in \mathbb{N}$. Betrachte insbesondere $b^kc^kab^k \in L(\gamma_1)$ mit $k \geq p$. Dann misslingt jeder Pumpversuch. Denn a muss in w liegen, weil sich die Zahl der a durch Pumpen veränderte, wenn a in v oder x läge, und die Zahl der b oder c vor a durch Pumpen verschieden würde zur Zahl der b hinter a, wenn a in u oder y läge. Damit aber beim Pumpen dann die Zahlenverhältnisse stimmen, muss v so viele b und c enthalten wie x b. Dann gilt aber $length(vwx) > k \geq p$, so dass wieder keine geeigneten Teilwörter gepumpt werden können. Das ist ein Widerspruch zur Annahme, so dass $L((\gamma_1)$ nicht kontextfrei sein kann.

Die Sprache ist allerdings kontextsensitiv, wie die folgende erzeugende Grammatik zeigt:

$$G_1 = (\{S\}, \{a, b, c\}, \{S ::= bcSb, S ::= a, bc ::= cb, cb ::= bc\}, S)$$

Mit den ersten beiden Regeln lässt sich für jedes k die Ableitung $S \overset{*}{\Rightarrow} (bc)^kab^k$ bilden. Mit den beiden letzten Regeln kann man die k b und c vor dem a in jede beliebige Reihenfolge bringen. An diesem Ergebnis ändert sich auch nichts, wenn man die b und c schon anfängt umzusortieren, bevor alle Anwendungen der Regeln 1 und 2 erfolgt sind. Wer sich daran stört, dass zwei linke Regelseiten terminal sind, möge in allen obigen Regeln c durch ein neues Nonterminal C ersetzen und die Regel $C ::= c$ hinzunehmen.

Das Beispiel zeigt, dass die Erzeugungsmächtigkeit von Sticker-Systemen mindestens bis in die Kontextsensitivität hineinreicht. Das ist aber auch bereits die Grenze des Möglichen, denn alle von Sticker-Systemen erzeugten Sprachen erweisen sich als kontextsensitiv.