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

Formale Sprachen: DNA Computing

Einseitige Sticker-Sprachen sind regulär

Ein Sticker-System $\gamma = (V, \rho, A, D)$ ist einseitig (*one-sided*), wenn in jedem Baustein $(u, v) \in D$ $u = \lambda$ oder $v = \lambda$, d.h. in jedem Berechnungsschritt wird nur links oder nur rechts etwas angebaut. Es soll gezeigt werden, dass die von einseitigen Sticker-Systemen erzeugten Sprachen regulär sind. Als erster Schritt wird eine korrekte Übersetzung einseitiger Sticker-Systeme in kontextfreie Grammatiken konstruiert. Die Idee dabei ist, die bei Berechnungen entstehenden *Sticky Ends* zu Nonterminals zu machen und in den Regeln auszudrücken, wie sich das aktuelle Molekül beim Berechnen an seinen Enden verändert. Eine Schwierigkeit dabei ist, dass potentiell unendlich viele *Sticky Ends* beim Berechnen vorkommen können. Es stellt sich heraus, dass man sich auf Berechnungen beschränken kann, deren *Sticky Ends* nie länger als die sind, die bereits in den Axiomen und Bausteinen vorkommen. Weitere Regeln modellieren die Anfänge von Berechnungen und die Abschlüsse von kompletten Berechnungen. Explizit ist die Übersetzung folgendermaßen definiert:

$$\xrightarrow{\gamma = (V, \rho, A, D)} \boxed{\text{OSL-2-CF}} \xrightarrow{G(\gamma) = (N(\gamma), \left[{V \atop V}\right]_\rho, P(\gamma), S)}$$

mit $N(\gamma) = \{S\} \cup \{\binom{r}{s}_l, \binom{v}{w}_r \mid r = \lambda \text{ oder } s = \lambda, v = \lambda \text{ oder } w = \lambda, |rs| \leq d(\gamma) \leq |vw|\}$.
Dabei ist $d(\gamma)$ die Länge des längsten Einzelstrangs, der in D direkt oder in A und D als *Sticky End* vorkommt, und $|w|$ bezeichnet die Länge eines Wortes w. $P(\gamma)$ ist so gewählt, dass sich komplette Berechnungen von γ als Ableitungen in $G(\gamma)$ simulieren lassen. Entsprechend enthält $P(\gamma)$ gerade alle Regeln, die in den folgenden Überlegungen verwendet werden.

Jede Berechnung beginnt mit $x_0 = \binom{r}{s}\left[{t \atop u}\right]\binom{v}{w} \in A$. Mit einer Regel $S ::= \binom{r}{s}_l \left[{t \atop u}\right] \binom{v}{w}_r$ lässt sich das durch die Ableitung $S \to \hat{x}_0$ simulieren, wobei $\hat{x}_0$ die rechte Regelseite ist.

Betrachte nun eine Berechnung $x_0 \overset{k}{\Longrightarrow} x_k$ der Länge k, wobei die *Sticky Ends* von x_k nicht länger als $d(\gamma)$ sind. Es sei vorausgesetzt, dass diese Berechnung durch eine Ableitung $S \overset{*}{\longrightarrow} \hat{x}_k$ in $G(\gamma)$ simuliert werden kann, wobei $\hat{x}_k$ aus x_k entsteht, indem man die *Sticky Ends* durch die entsprechenden Nonterminals ersetzt.

Sei ferner vorausgesetzt, dass das rechte *Sticky End* oben sitzt und an diesem Ende weitergebaut wird. Dann gibt es vier Fälle:

(1)

$$x_0 \overset{k}{\Longrightarrow} x_k = \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{matrix} v \\ \end{matrix} \Longrightarrow \ldots \left.\begin{matrix} x\ v\ t' \\ y\ w\ u' \end{matrix}\right] \begin{matrix} v' \\ w' \end{matrix} = x_{k+1}$$

mit $({}_w \left[\begin{smallmatrix} t' \\ u' \end{smallmatrix}\right] \begin{smallmatrix} v' \\ w' \end{smallmatrix}) \in D$ und $\left[\begin{smallmatrix} v \\ w \end{smallmatrix}\right] \in WK_\rho(V)$. Außerdem gilt $v' = \lambda$ oder $w' = \lambda$ und $|v'w'| \leq d(\gamma)$. Deshalb kann die Ableitung $S \xrightarrow{*} \hat{x}_k$ geeignet verlängert werden zu

$$S \xrightarrow{*} \hat{x}_k = \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{pmatrix} v \\ \lambda \end{pmatrix}_r \to \ldots \left.\begin{matrix} x\ v\ t' \\ y\ w\ u' \end{matrix}\right] \begin{pmatrix} v' \\ w' \end{pmatrix}_r = \hat{x}_{k+1}$$

falls $\left(\begin{smallmatrix} v \\ \lambda \end{smallmatrix}\right)_r ::= \left[\begin{smallmatrix} v & t' \\ w & u' \end{smallmatrix}\right] \left(\begin{smallmatrix} v' \\ w' \end{smallmatrix}\right)_r \in P(\gamma)$.

(2)

$$x_0 \overset{k}{\Longrightarrow} x_k = \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{matrix} v \\ \end{matrix} \Longrightarrow \ldots \left.\begin{matrix} x\ v \\ y\ w \end{matrix}\right]_{w'} = x_{k+1}$$

mit $(\lambda, \left(\begin{smallmatrix} \lambda \\ ww' \end{smallmatrix}\right)) \in D$ und $\left[\begin{smallmatrix} v \\ w \end{smallmatrix}\right] \in WK_\rho(V)$. Außerdem gilt $|w'| \leq |ww'| \leq d(\gamma)$. Deshalb kann die korrespondierende Ableitung wieder geeignet verlängert werden zu

$$S \xrightarrow{*} \hat{x}_k = \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{pmatrix} v \\ \lambda \end{pmatrix}_r \to \ldots \left.\begin{matrix} x\ v \\ y\ w \end{matrix}\right] \begin{pmatrix} \lambda \\ w' \end{pmatrix}_r = \hat{x}_{k+1},$$

falls $\left(\begin{smallmatrix} v \\ \lambda \end{smallmatrix}\right)_r ::= \left[\begin{smallmatrix} v \\ w \end{smallmatrix}\right] \left(\begin{smallmatrix} \lambda \\ w' \end{smallmatrix}\right)_r \in P(\gamma)$.

(3)

$$x_0 \overset{k}{\Longrightarrow} x_k = \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{matrix} v'\ v'' \\ \end{matrix} \Longrightarrow \ldots \left.\begin{matrix} x\ v' \\ y\ w' \end{matrix}\right] \begin{matrix} v'' \\ \end{matrix} = x_{k+1}$$

mit $(\lambda, \left(\begin{smallmatrix} \lambda \\ w' \end{smallmatrix}\right)) \in D$ und $\left[\begin{smallmatrix} v' \\ w' \end{smallmatrix}\right] \in WK_\rho(V)$ und $v = v'v''$. Außerdem gilt $|v''| \leq |v'v''| = |v| \leq d(\gamma)$. Deshalb kann die korrespondierende Ableitung erneut geeignet verlängert werden zu

$$S \xrightarrow{*} \hat{x}_k = \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{pmatrix} v'v'' \\ \lambda \end{pmatrix}_r \to \ldots \left.\begin{matrix} x\ v' \\ y\ w' \end{matrix}\right] \begin{pmatrix} v'' \\ \lambda \end{pmatrix}_r = \hat{x}_{k+1},$$

falls $\left(\begin{smallmatrix} v \\ \lambda \end{smallmatrix}\right)_r ::= \left[\begin{smallmatrix} v' \\ w' \end{smallmatrix}\right] \left(\begin{smallmatrix} v'' \\ \lambda \end{smallmatrix}\right)_r \in P(\gamma)$.

(4)

$$x_0 \overset{k}{\Longrightarrow} x_k = \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{matrix} v \\ \end{matrix} \Longrightarrow \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{matrix} v\overline{v} \\ \end{matrix} = x_{k+1}$$

mit $(\lambda, \left(\begin{smallmatrix} \overline{v} \\ \lambda \end{smallmatrix}\right)) \in D$. Die korrespondierende Ableitung kann wieder geeignet verlängert werden:

$$S \xrightarrow{*} \hat{x}_k = \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{pmatrix} v \\ \lambda \end{pmatrix}_r \to \ldots \left.\begin{matrix} x \\ y \end{matrix}\right] \begin{pmatrix} v\overline{v} \\ \lambda \end{pmatrix}_r = \hat{x}_{k+1},$$

falls $\binom{v}{\lambda}_r ::= \binom{v\overline{v}}{\lambda}_r \in P(\gamma)$. Aber das ist nach Definition der Nonterminals nur möglich, wenn $|v\overline{v}| \leq d(\gamma)$. Bei einer Berechnung ist diese Bedingung allerdings nicht immer erfüllt, so dass der Fall $|v\overline{v}| > d(\gamma)$ noch weiter betrachtet werden muss, was im nächsten Abschnitt passiert.

Damit $G(\gamma)$ dieselbe Sprache wie γ erzeugt, müssen nur komplette Berechnungen simuliert werden. Bei denen aber wird jedes zwischendurch bestehende *Sticky End* später aufgefüllt. Wie kann die Berechnung in Punkt 4 weitergehen? Da die Arbeiten am linken und rechten Ende völlig unabhängig voneinander sind, kann angenommen werden, dass erst alle Schritte rechts ausgeführt werden, bevor links gearbeitet wird. Demnach könnten als nächstes weitere Schritte kommen, die das vorhandene *Sticky End* verlängern. Das ändert die Situation nicht grundsätzlich. Auf jeden Fall muss nämlich nach der vorläufig letzten Verlängerung des *Sticky End* eine Auffüllung erfolgen, die mit einem Einzelstrang arbeitet, weil echte *Sticky Ends* in D zu kurz sind.

(5)

$$x_0 \overset{k}{\Longrightarrow} x_k = \ldots \begin{matrix} x \\ y \end{matrix}\Big] \, v \overset{*}{\Longrightarrow} \ldots \begin{matrix} x \\ y \end{matrix}\Big] \, v\overline{v} \Longrightarrow \ldots \begin{matrix} x\, v' \\ y\, w \end{matrix}\Big] \, v''$$

mit $|v\overline{v}| > d(\gamma), (\lambda, \binom{\lambda}{w}) \in D, v\overline{\overline{v}} = v'v''$ und $\begin{bmatrix} v' \\ w \end{bmatrix} \in WK_\rho(V)$.

Dieser Schritt lässt sich aber mit den Verlängerungen vertauschen, was man wie folgt einsehen kann:

1. Fall: $v = u'u''$ mit $\begin{bmatrix} u' \\ w \end{bmatrix} \in WK(V) : \ldots \begin{matrix} x \\ y \end{matrix}\Big] \, v = \ldots \begin{matrix} x \\ y \end{matrix}\Big] \, u'u'' \Longrightarrow \ldots \begin{matrix} x\, u' \\ y\, w \end{matrix}\Big] \, u'' \overset{*}{\Longrightarrow} \ldots \begin{matrix} x\, u' \\ y\, w \end{matrix}\Big] \, u''\overline{\overline{v}}$.
Wegen $u'u''\overline{\overline{v}} = v\overline{\overline{v}} = v'v''$ und $|u'| = |w| = |v'|$ folgt $u' = v'$ und $u''\overline{\overline{v}} = v''$, so dass beide Berechnungen dasselbe liefern.

2. Fall: $w = w'w''$ mit $\begin{bmatrix} v \\ w' \end{bmatrix} \in WK_\rho(V)$ und $\overline{\overline{v}} = tt'$ mit $\begin{bmatrix} t \\ w'' \end{bmatrix} \in WK_\rho(V)$ und $vt = v'$:
$\ldots \begin{matrix} x \\ y \end{matrix}\Big] \, v \Longrightarrow \ldots \begin{matrix} x\; v \\ y\; w' \end{matrix}\Big] \, w'' \overset{*}{\Longrightarrow} \ldots \begin{matrix} x\; v\; t \\ y\; w'\; w'' \end{matrix}\Big] \, \overline{t}k$.
Wegen $v\overline{\overline{v}} = vt\overline{t} = v'\overline{t} = v'v''$ folgt $\overline{t} = v''$, so dass wieder dasselbe Molekül herauskommt.

In beiden Fällen kann man also einen $(k+1)$-ten Berechnungsschritt finden, der nach Punkt 2 oder 3 simuliert werden kann. Wiederholt man diese Konstruktion oft genug, verschwindet das Problem, weil am Ende kein *Sticky End* bleibt, so dass man eine komplette Berechnung $x_0 \overset{*}{\Longrightarrow} \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$ immer simulieren kann durch eine Ableitung $S \overset{*}{\longrightarrow} \binom{\lambda}{\lambda}_l \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix} \binom{\lambda}{\lambda}_r$. Dabei ist ausgenutzt, dass man mit *Sticky End*s unten so umgehen kann wie oben und links wie rechts. Um zu terminieren, braucht man noch zwei weitere Regeln: $\binom{\lambda}{\lambda}_l ::= \lambda$ und $\binom{\lambda}{\lambda}_r ::= \lambda$. Die Regeln sind außerdem so gewählt, dass jede Ableitung vom Startsymbol aus einer Berechnung entspricht, wenn sie nicht gerade die Länge 0 hat. Das alles zusammengenommen ergibt die Korrektheit der Übersetzung auf der Ebene der Molekülsprachen: $LM(\gamma) = L(G(\gamma))$.

Da $G(\gamma)$ eine kontextfreie Grammatik ist, bleibt noch die Regularität zu zeigen. Außerdem soll noch nachgewiesen werden, dass sich daran nichts ändert, wenn man zu Wörtern übergeht.

Die Grammatik $G(\gamma)$ ist sehr speziell gebaut. In der rechten Seite von Regeln erscheinen zwei Nonterminals, wenn S die linke Seite ist. Alle anderen Regeln sind links- oder rechtslinear. Auerdem beginnt jede Ableitung, die in S startet und einen terminalen Wert liefert, mit der Anwendung einer Regel, die am linken und am rechten Ende je ein Nonterminal hat, wobei das linke Nonterminal nur durch linkslineare Regeln und das rechte nur durch rechslineare Regeln abgeleitet wird. Außerdem sind die linken und rechten Ableitungsschritte voneinander völlig unabhängig. Betrachtet man nun die Grammatiken $G\binom{r}{s}_l$, die alle l-Nonterminals als Nonterminals und alle Regeln mit l-Nonterminals als linker Seite enthalten und $\binom{r}{s}_l$ als Startsymbol besitzen, sowie analog die Grammatiken $G\binom{v}{w}_r$, so stellt sich heraus, dass jedes $\begin{bmatrix} w \\ w' \end{bmatrix} \in L(G(\gamma))$ die Form $\begin{bmatrix} w \\ w' \end{bmatrix} = \begin{bmatrix} x \\ x' \end{bmatrix} \begin{bmatrix} t \\ t' \end{bmatrix} \begin{bmatrix} y \\ y' \end{bmatrix}$ hat mit $\begin{bmatrix} x \\ x' \end{bmatrix} \in L(G\binom{r}{s}_l)$, $\begin{bmatrix} y \\ y' \end{bmatrix} \in L(G\binom{u}{v}_r)$ und $\binom{r}{s} \begin{bmatrix} t \\ t' \end{bmatrix} \binom{u}{v} \in A$. Da $G\binom{r}{s}_l$ links- und $G\binom{u}{v}_r$rechtslinear ist, sind die erzeugten Sprachen regulär. Da einelementige Sprachen, Konkatenationen und Vereinigungen regulärer Sprachen regulär sind, erweist sich $L(G(\gamma))$ als regulär.

$L(G(\gamma))$ wird von einem endlichen Automaten A erkannt. Ersetzt man bei dem das Eingabealphabet $\begin{bmatrix} V \\ V \end{bmatrix}_\rho$ durch V und in jedem Zustandsübergang das zugehörige Eingabesymbol $\begin{bmatrix} a \\ b \end{bmatrix}$ durch a, erkennt der resultierende Automat gerade die oberen Stränge der von A erkannten Moleküle und damit $L(\gamma)$.

Bezeichnet man die Klasse der durch einseitige Sticker-Systeme erzeugten Sprachen als OSL und die der regulären Sprachen als REG, so gilt also nach den vorausgegangenen Konstruktionen und Überlegungen: $OSL \subseteq REG$. Da sich auch $REG \subseteq RSL$ zeigen lässt, wobei RSL alle von regulären Sticker-Systemen erzeugten Sprachen enthält, und offensichtlich $RSL \subseteq OSL$ gilt (nur rechts Anbauen ist besonders einseitig), stellt sich insgesamt heraus:

$$REG = RSL = OSL.$$

Die Erzeugungskraft von Sticker-Sytemen über das Kontextfreie hinaus bis ins Kontextsensitive hinein kommt also vom gleichzeitigen Bauen an beiden Enden.

Darüber hinaus sind die *Sticky Ends* wichtig. Denn ein Verzicht auf echte *Sticky Ends*, wie bei glatten Sticker-Systemen, deren Atome und Bausteine nur vollständige Moleküle enthalten, führt zu Sticker-Sprachen, die linear sind.