SIMULATION OF SCATTERED CONTEXT GRAMMARS BY SYMBIOTIC E0L GRAMMARS

Ing. Tomáš KOPEČEK, Doctoral Degree Programme (2)
Dept. of Information Systems, FIT, BUT
E-mail: kopecek@fit.vutbr.cz

Supervised by: Dr. Alexander Meduna

ABSTRACT

This paper contains more examples to formerly introduced concept of formal language equivalency. That is, for two models, there is a substitution by which we change each string of every yield sequence in one model so that sequence ofs string resulting from this change represents a yield sequence in the other equvalent model, these two models closely simulates each other; otherwise they do not. In this paper are shown two cases of such simulations.

1 INTRODUCTION

In the [1] was introduced quite new method of compraing two grammatical systems. Before this paper there was almost vague comparations of grammars limited by similarity of generated languages. This new approach comes with comparing not only generated languages but also similarity of generating process.

Because we have many different transformations from one type of grammar to another in the theory of formal languages, we sometimes want to describe similarity of such converted grammars. On the second hand, we need to examine this similarity in the practice. For example we try to find some usable representation of some grammar for use in a compiling system. We can do some transformations but we still want to achieve same result in new grammar with almost same number of derivation steps and so on.

So, the concepts of *m-close simulation* and some others were introduced in [1].

In the section 2 are recalled some well-known notions of the formal language theory. Section 3 introduces new conversion from scattered context grammars to symbiotic E0L grammars. Next section deals with description of derivation simulation of previous section. Here are repeated some needed definitions of concepts of derivation similarity and showed theorem about previous conversion. Section 5 includes proved results as a whole.

2 PRELIMINARIES

We assume that the reader is familiar with the language theory (see [2], [4], [6]).

Let V be an alphabet. V^* denotes the free monoid generated by V under the operation of concatenation. Let ε be the unit of V^* and $V^+ = V^* - \{\varepsilon\}$. Given a word, $w \in V^*$, $|w|$ represents the length of w and $alph(w)$ denotes the set of all symbols occuring in w. Moreover, $sub(w)$ denotes the set of all subwords of w. Let R be a binary relation on a set W. Instead of $u \in R(v)$, where $u, v \in W$, we write vRu in this paper.

A *scattered context grammar* is an ordered quadruple $G = (V,T,P,S)$, where V, T, and S are the total alphabet of G, the set of terminals $T \subseteq V$, and the axiom $S \in V - T$, respectively. P is a finite set of productions of the form $(A_1,\ldots,A_n) \to (x_1,\ldots,x_n)$, for some $n \geq 1$, where $A_i \in V - T$ and $x_i \in V^*$ form $1 \leq i \leq n$. If $p \in P$ is of the form $(A_1,A_2,\ldots,A_n) \to (x_1,x_2,\ldots,x_n)$, $u = u_1A_1u_2A_2\ldots u_nA_nu_{n+1}$, $v = u_1x_1u_2x_2\ldots u_nx_nu_{n+1}$ and $u_i \in V^*$, for $i = 1,2,\ldots,n$, then u directly derives v in G according to p, denoted by $u \Rightarrow_G v[p]$ or, simply $u \Rightarrow v$. In a standard manner, we extend $\Rightarrow_G$ to $\Rightarrow_G^n$, where $n \geq 0$, and based on $\Rightarrow_G^n$, we define $\Rightarrow_G^*$, which is transitive and reflexive closure of $\Rightarrow$. Let $S \Rightarrow_G^* x$ is called a successful derivation. The language of G, $L(G)$, is defined as $L(G) = \left\{x : S \Rightarrow_G^* x, x \in T^*\right\}$. For any $p \in P$ of the form $(A_1,A_2,\ldots,A_n) \to (x_1,x_2,\ldots,x_n), left(p)$ means string $A_1A_2\ldots A_n$ and $right(p)$ string $x_1x_2\ldots x_n$.

A *symbiotic E0L grammar* (see [3]) is a quadruple $G = (W,T,P,S)$, where W, T, and S are the set of generators $W \subseteq (V \cup V^2)$, the set of terminals $T \subseteq V$, and the axiom $S \in V - T$, respectively. P is a finite set of productions of the form $A \to x$, $A \in V$, $x \in V^*$. The direct derivation relation is defined in the following way: let $x, y \in W^*$ such that $x = a_1a_2\ldots a_n$, $a_i \in V$, $y = y_1y_2\ldots y_n$, $y_i \in V^*$, and productions $a_i \to y_i \in P$ for all $i = 1,\ldots,n$. Then, x directly derives y, $x \Rightarrow_G y$. The language of G is $L(G) = \left\{w \in T^* : S \Rightarrow_G^* w\right\}$.

3 SIMULATION OF SCATTERED CONTEXT GRAMMARS

Construction 1.

Input: A scattered context grammar, $G = (V,T,P,S)$

Output: A symbiotic E0L grammars, G'

Algorithm: At first, we introduce a new alphabet, $V' = V \cup \{@,\#,S'\} \cup V'' \cup \widetilde{T}, V'' = \{\langle i,j\rangle : 0 < i \leq Card(P), 0 \leq j \leq k\}, \widetilde{T} = \{\widetilde{a} : a \in T\}$. Let τ be a homomorphism from T to $\widetilde{T}$ such that $\tau(a) = \widetilde{a}$ for all $a \in T$. Define a language W, over V' as $W = V \cup \{@,\#,S'\} \cup \widetilde{T} \cup \{\langle i,j\rangle \langle i,j\rangle : 0 < i \leq Card(P), 0 \leq j \leq k\}$. Then, construct a symbiotic E0L grammar $G' = (W,T,P',S')$, where the set of productions is defined in the following way:

1. add $S' \to @S\#, @ \to \varepsilon$ and $\# \to \varepsilon$ to P';

2. for every production $n{:}\,(A_1,A_2,\ldots,A_k) \to (x_1,x_2,\ldots,x_k) \in P$ add these rules to P' (where n is a label, $0 <\leq Card(P)$:

$$\begin{array}{rcl} A_1 & \to & \langle n,0\rangle\,\tau(x_1)\,\langle n,1\rangle \\ A_2 & \to & \langle n,1\rangle\,\tau(x_2)\,\langle n,2\rangle \\ & \vdots & \\ A_k & \to & \langle n,k-1\rangle\,\tau(x_k)\,\langle n,k\rangle \end{array}$$

3. add $@ \rightarrow @\langle i,0\rangle, 0 < i \leq Card(P)$ to P';

4. add $\# \rightarrow \langle i,k\rangle \#$, to P' for each production i: $(A_1,A_2,\ldots,A_k) \rightarrow (x_1,x_2,\ldots,x_k) \in P$;

5. for each $A \in V \cup \widetilde{T}$ add productions of this form to P': $A \rightarrow \langle i,j\rangle A \langle i,j\rangle, 0 < i \leq Card(P), 0 \leq j \leq k$;

6. add these productions to P': $\langle i,j\rangle \rightarrow \varepsilon, 0 < i \leq Card(P), 0 \leq j \leq k$;

7. add production $\widetilde{a} \rightarrow a$ for each $a \in T$ to P'.

Theorem 1. *Let $G = (V,T,P,S)$ be a scattered context grammar. Let G' be a symbiotic E0L grammar constructed by Construction 1 with G as its input. Then, $L(G) = L(G')$.*

4 DERIVATION SIMULATIONS

4.1 DEFINITIONS

Now we have to repeat some needed definitions. Definitions as a whole were introduced in [1] and there can be found reasons of their existence and so on. Here we only repeat their readings because they will be used in the following subsections.

Definition 1. A *string-relation system* is a quadruple $\Psi = (W, \Rightarrow, W_0, W_F)$, where W is a language, $\Rightarrow$ is a binary relation on W, $W_0 \subseteq W$ is a set of *start strings*, and $W_F \subseteq W$ is a set of *final strings*.

Every string, $w \in W$, represents a 0-step string-relation sequence in Ψ. For every $n \geq 1$, a sequence $w_0, w_1, \ldots w_n$, $w_i \in W$, $0 \leq i \leq n$, is an *n-step string-relation sequence*, symbolically written as $w_0 \Rightarrow w_1 \Rightarrow \ldots \Rightarrow w_n$, if for each $0 \leq i \leq n-1$, $w_i \Rightarrow w_{i+1}$.

If there is a string-relation sequence $w_0 \Rightarrow w_1 \Rightarrow \ldots \Rightarrow w_n$, where $n \geq 0$, we write $w_0 \Rightarrow^n w_n$. Furthermore, $w_0 \Rightarrow^* w_n$ means that $w_0 \Rightarrow^n w_n$ for some $n \geq 0$, and $w_0 \Rightarrow^+ w_n$ means that $w_0 \Rightarrow^n w_n$ for some $n \geq 1$. Obviously, from the mathematical point of view, $\Rightarrow^+$ and $\Rightarrow^*$ are the transitive closure of $\Rightarrow$ and the transitive and reflexive closure of $\Rightarrow$, respectively.

Let $\Psi = (W, \Rightarrow, W_0, W_F)$ be a string-relation system. A string-relation sequence in Ψ, $u \Rightarrow^* v$, where $u, v \in W$, is called a *yield sequence*, if $u \in W_0$. If $u \Rightarrow^* v$ is a yield sequence and $v \in W_F$, $u \Rightarrow^* v$ is *successful*.

Let $D(\Psi)$ and $SD(\Psi)$ denote the set of all yield sequences and all successful yield sequences in Ψ, respectively.

Definition 2. Let $\Psi = (W, \Rightarrow_\Psi, W_0, W_F)$ and $\Omega = (W', \Rightarrow_\Omega, W'_0, W'_F)$ be two s-r systems, and let σ be a substitution from W' to W. Furthermore, let d be a yield sequence in Ψ of the form $w_0 \Rightarrow_\Psi w_1 \Rightarrow_\Psi \ldots \Rightarrow_\Psi w_{n-1} \Rightarrow_\Psi w_n$, where $w_i \in W$, $0 \leq i \leq n$, for some $n \geq 0$. A yield sequence, h, in Ω *simulates d with respect to* σ, symbolically written as $h \rhd_\sigma d$, if h is of the form $y_0 \Rightarrow_\Omega^{m_1} y_1 \Rightarrow_\Omega^{m_2} \ldots \Rightarrow_\Omega^{m_{n-1}} y_{n-1} \Rightarrow_\Omega^{m_n} y_n$, where $y_j \in W'$, $0 \leq j \leq n$, $m_k \geq 1$, $1 \leq k \leq n$, and $w_i \in \sigma(y_i)$ for all $0 \leq i \leq n$. If, in addition, there exists $m \geq 1$ such that $m_k \leq m$ for each $1 \leq k \leq n$, then h *m-closely simulates d with respect to* σ, symbolically written as $h \rhd_\sigma^m d$.

Definition 3. Let $\Psi = (W, \Rightarrow_\Psi, W_0, W_F)$ and $\Omega = (W', \Rightarrow_\Omega, W'_0, W'_F)$ be two s-r systems, and let σ be a substitution from W' to W. Let $X \subseteq D(\Psi)$ and $Y \subseteq D(\Omega)$. *Y simulates X with respect to* σ, written as $Y \rhd_\sigma X$, if the following two conditions hold: for every $d \in X$, there is $h \in Y$ such that $h \rhd_\sigma d$ and for every $h \in Y$, there is $d \in X$ such that $h \rhd_\sigma d$.

Let m be a positive integer. *Y m-closely simulates X with respect to* σ, $Y \rhd_\sigma^m X$, provided that: for every $d \in X$, there is $h \in Y$ such that $h \rhd_\sigma^m d$ and for every $h \in Y$, there is $d \in X$ such that $h \rhd_\sigma^m d$.

Definition 4. Let $\Psi = (W, \Rightarrow_\Psi, W_0, W_F)$ and $\Omega = (W', \Rightarrow_\Omega, W'_0, W'_F)$ be two s-r systems. If there exists a substitution σ from W' to W such that $D(\Omega) \rhd_\sigma D(\Psi)$ and $SD(\Omega) \rhd_\sigma SD(\Psi)$, then Ω is said to be Ψ's *derivation simulator* and *successful-derivation simulator*, respectively. Furthermore, if there is an integer, $m \geq 1$, such that $D(\Omega) \rhd_\sigma^m D(\Psi)$ and $SD(\Omega) \rhd_\sigma^m SD(\Psi)$, Ω is called an *m-close derivation simulator* and *m-close successful-derivation simulator* of Ψ, respectively. If there exists a homomorphism ρ from W' to W such that $D(\Omega) \rhd_\rho D(\Psi)$, $SD(\Omega) \rhd_\rho SD(\Psi)$, $D(\Omega) \rhd_\rho^m D(\Psi)$, and $SD(\Omega) \rhd_\rho^m SD(\Psi)$, then Ω is Ψ's *homomorphic derivation simulator*, *homomorphic successful-derivation simulator*, *m-close homomorphic derivation simulator* and *m-close homomorphic successful-derivation simulator*, respectively.

4.2 DERIVATION SIMULATION OF SCATTERED CONTEXT GRAMMARS

Definition 5. Let $G = (V, T, P, S)$ be a scattered context grammar. Let $\Rightarrow_G$ be the direct derivation relation in G. For $\Rightarrow_G$ and every $l \geq 0$, set $\Delta(\Rightarrow_G, l) = \{x \Rightarrow_G y : x \Rightarrow_G y \Rightarrow_G^i w, x, y \in V^*, w \in T^*, i+1 = l, i \geq 0\}$.

Next, let $G_1 = (V_1, T_1, P_1, S_1)$ and $G_2 = (V_2, T_2, P_2, S_2)$ be scattered context grammars. Let $\Rightarrow_{G_1}$ and $\Rightarrow_{G_2}$ be the derivation relations of G_1 and G_2, respectively. Let σ be a substitution from V_2 to V_1. *G_2 simulates G_1 with respect to* σ, $D(G_2) \rhd_D (G_1)$ in symbols, if there exists two natural numbers $k, l \geq 0$ so that the following conditions hold:

1. $\Psi_1 = (V_1^*, \Rightarrow_{G_1}, \{S_1\}, T_1^*)$ and $\Psi_2 = (V_2^*, \Rightarrow_{\Psi_2}, W_0, W_F)$ are string-relation systems corresponding to G_1 and G_2, respectively, where $W_0 = \{x \in V_2^* : S_2 \Rightarrow_{G_2}^k x\}$ and $W_F = \{x \in V_2^* : x \Rightarrow_{G_2}^l w, w \in T_2^*, \sigma(w) \subseteq T_1^*\}$;

2. relation $\Rightarrow_{\Psi_2}$ coincides with $\Rightarrow_{G_2} - \Delta(\Rightarrow_{G_2}, l)$;

3. $D(\Psi_2) \rhd_\sigma D(\Psi_1)$.

In case that $SD(\Psi_2) \rhd_\sigma SD(\Psi_1)$, *$G_2$ simulates successful derivations of G_1 with respect to* σ; in symbols, $SD(G_2) \rhd_\sigma SD(G_1)$.

Definition 6. Let G_1 and G_2 be scattered context grammars with total alphabets V_1 and V_2, terminal alphabets T_1 and T_2, and axioms S_1 and S_2, respectively. Let σ be a substitution from V_2 to V_1. *G_2 m-closely simulates G_1 with respect to* σ if $D(G_2) \rhd_\sigma D(G_1)$ and there exists $m \geq 1$ such that the corresponding string-relation systems Ψ_1 and Ψ_2 satisfy $D(\Psi_2) \rhd_\sigma^m D(\Psi_1)$. In symbols, $D(G_2) \rhd_\sigma^m D(G_1)$.

Analogously, *G_2 m-closely simulates successful derivations of G_1 with respect to* σ, denoted by $SD(G_2) \rhd_\sigma^m SD(G_1)$, if $SD(\Psi_2) \rhd_\sigma^m SD(\Psi_1)$ and there exists $m \geq 1$ such that $SD(G_2) \rhd_\sigma^m SD(G_1)$.

Definition 7. Let G_1 and G_2 be two scattered context grammars. If there exists a substitution σ such that $D(G_2) \rhd_\sigma D(G_1)$, then G_2 is said to be G_1's *derivation simulator*.

By analogy with Definition 7, the reader can also define *homomorphic*, *m-close*, and *successful-derivation simulators* of scattered context grammars.

Theorem 2. *Let $G = (V,T,P,S)$ be a scattered context grammar and $G' = (W,T,P',S')$ be a symbiotic E0L grammar constructed by using Construction 1 with G as its input. Then, there exists a homomorphism $\widetilde{\omega}$ such that $D(G') \rhd^1_{\widetilde{\omega}} D(G)$ and $SD(G') \rhd^1_{\widetilde{\omega}} SD(G)$.*

5 CONCLUSION

In this paper we have gained following results:

Theorems 1 and 2 show that for every scattered context grammar $G = (V,T,P,S)$, there exists a symbiotic E0L grammar $G' = (W',T,P',S')$ such that

1. $L(G) = L(G')$;
2. G' is a 1-close homomorphic derivation simulator of G';
3. G' is a 1-close homomorphic successful-derivation simulator of G;
4. To simulate G, G' uses one initial derivation step ($S' \Rightarrow_{G'} @S\#$) and one derivation step which removes auxiliary symbols ($\langle i,0\rangle\,\langle i,0\rangle\,\tau(t_1)\,\langle i,0\rangle \ldots \langle i,k\rangle\,\tau(t_n)\,\langle i,k\rangle\,\langle i,k\rangle \Rightarrow_{G'} t_1t_2\ldots t_n : 0 < i \leq Card(P), t_j \in T^*, 1 \leq j \leq n, n \geq 0$).

All theorems are given without proofs because of limited space. These proofs can be obtained from author. With almost same aparatus we have confirmed similar results with phrase-structured grammars.

REFERENCES

[1] Meduna, A., Švec M.: A Formalization of Derivation Similarity in the Formal Language Theory and its Illustration in Terms of Lindenmayer systems, Fundamenta Informaticae 20 (2004)

[2] Hamburger H., Richards D.: Logic and Language Models for Computer Science, Prentice Hall, London, 2002.

[3] Meduna A.: Symbiotic E0L Systems, Acta Cybernetica 10 (1992), 165–172.

[4] Meduna A.: Automata and Languages: Theory and Applications, Springer, London, 2000.

[5] Meduna A., Švec M.: Forbidding ET0L grammars, Theoretical Computer Science 306 (2003), 449–469.

[6] Parkes, A.: Introduction to Languages, Machines, and Logic: Computable Languages, Abstract Machines, and Formal Logic, Springer, London, 2002.