The synchronized graphs trace the context-sensitive languages
Chloé Rispal

To cite this version:
Chloé Rispal. The synchronized graphs trace the context-sensitive languages. INFINITY 2002, 4th International Workshop on Verification of Infinite-State Systems, Aug 2003, Brno, Czech Republic. 10pp. hal-00620118

HAL Id: hal-00620118
https://hal-upec-upem.archives-ouvertes.fr/hal-00620118
Submitted on 3 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The synchronized graphs trace the context-sensitive languages

Rispal Chloé ¹

Irisa
Campus de Beaulieu
Rennes, France

Abstract

Morvan and Stirling have proved that the context-sensitive languages are exactly the traces of graphs defined by transducers with labelled final states. We prove that this result is still true if we restrict to the traces of graphs defined by synchronized transducers with labelled final states. From their construction, we deduce that the context-sensitive languages are the languages of path labels leading from and to rational vertex sets of letter-to-letter rational graphs.

1 Introduction

As for formal languages, an infinite graph hierarchy exists. First of all, Muller and Schupp [11] have defined the transition graphs of pushdown automata. Then, Courcelle has defined the family of equational graphs which are the graphs generated by deterministic graph grammars [5]. Caucal has extended these families to prefix recognizable graphs which are the prefix transitions of recognizable systems [3]. More recently, Morvan has introduced the rational graphs which are recognized by word transducers with labelled final states [9]. Finally, Caucal has presented the transition graphs of Turing machines [4].

A trace of a graph is the language of path labels leading from and to finite vertex sets. Traces of graphs are a link between infinite graph hierarchy and

¹ Email: rispal@univ-mlv.fr

©2002 Published by Elsevier Science B. V.
the Chomsky hierarchy of languages. The traces of finite graphs are the rational languages, the traces of prefix recognizable graphs are the context-free languages \cite{1}, the traces of rational graphs are the context-sensitive languages \cite{10} and finally, the traces of Turing graphs are the recursively enumerable languages \cite{4}.

A particular rational relation is the left-synchronized relation which is recognized by a letter-to-letter transducer followed by a recognizable relation for each final state \cite{6} and \cite{7}. These left-synchronized relations form a boolean algebra and are recognized by deterministic transducers. A graph is synchronized if it is isomorphic to some graph having words as vertices and such that each labelled transition is a left-synchronized relation. The synchronized graphs are the automatic graphs of Blumensath and Grädel \cite{1}. In this paper, we adapt the construction of Morvan and Stirling \cite{10} to prove that the context-sensitive languages are exactly the traces of synchronized graphs. We also characterize the context-sensitive languages as the languages of path labels leading from and to rational vertex sets of letter-to-letter rational graphs.

2 Rational synchronized graphs

Let $N$ be a finite alphabet. We denote by $N^*$ the set of words over letters of $N$, and we write $\epsilon$ for the empty word.

A transducer $T$ is defined by a finite subset of $Q \times N^* \times N^* \times Q$ of labelled edges where $Q$ is a finite set of states, by a set $I \subseteq Q$ of initial states, and by a set $F \subseteq Q$ of final states. So a transducer is a finite automaton labelled by pairs of words. Any transition $(p, u, v, q)$ of a transducer $T$ will be denoted by $p \xrightarrow{u/v} q$ or by $p \xrightarrow{u/v} q$ when $T$ is understood.

A path $p_0 \xrightarrow{u_1/v_1} p_1 \cdots p_{n-1} \xrightarrow{u_n/v_n} p_n$ with $u = u_1 \ldots u_n$ and $v = v_1 \ldots v_n$ is labelled $u/v$ and is denoted by $p_0 \xrightarrow{u/v} p_n$. A path is \textit{successful} if it leads from an initial state to a final one. A pair $(u, v) \in N^* \times N^*$ is recognized by a transducer if there exists a successful path labelled $u/v$. A relation is \textit{rational} if it is recognized by a transducer.

\textbf{Example 2.1} The following transducer:

\begin{center}
\begin{tikzpicture}
\node (p) at (0,0) [draw, circle] {$p$};
\node (q) at (1,0) [draw, circle] {$q$};
\path [->] (p) edge node [above] {B/AA} (q);
\path [->] (q) edge node [above] {B/AA} (p);
\end{tikzpicture}
\end{center}
with initial state $p$ and final state $q$ recognizes the rational relation
\[ \{ (A^n B^m, B^n A^{2m}) \mid n \geq 0, m > 0 \}. \]

From studies concerning rational relations, Elgot and Mezéi \cite{elgot1967} and then
Frougny and Sakarovitch \cite{frougny1994} have defined the subfamily of left-synchronized
relations.

If a transducer has labels over $\mathbb{N} \times \mathbb{N}$ it is called a **letter-to-letter 2-automaton**: it is a transducer labelled by pairs of letters instead of pairs of words. Adding
a rational terminal function completing one side of the recognized pairs, it recognizes a left-synchronized relation.

**Definition 2.2** A relation over $\mathbb{N}^\ast \times \mathbb{N}^\ast$ is left-synchronized if it is recognized
by a letter-to-letter 2-automaton with terminal function taking values in
\[ \text{Dif}_{\text{Rat}} = (\text{Rat}(\mathbb{N}^\ast) \times \{ \varepsilon \}) \cup (\{ \varepsilon \} \times \text{Rat}(\mathbb{N}^\ast)) \]

That is a left-synchronized relation is a finite union of elementary relations
of the form $R.S$ where $R \in \text{Rat}((\mathbb{N} \times \mathbb{N})^\ast)$ and $S \in \text{Dif}_{\text{Rat}}$

**Example 2.3** For all integer $p$, the relation $\mid_p$ defined by $x \mid_p y$ if $x$ is a power of
$p$ dividing $y$, is left-synchronized. For instance, in base two with weak weighs
on the left, $\mid_2$ is recognized by the following letter-to-letter 2-automaton:

\[
\begin{array}{c}
\text{with the terminal function } f \text{ defined by } f(p) = (\varepsilon, 0)^\ast(\varepsilon, 1)(\langle \varepsilon, 0 \rangle, (\varepsilon, 1))^\ast \\
\text{and } f(q) = (\varepsilon, \varepsilon)
\end{array}
\]

As the terminal function is rational, it can be introduced in the transducer. A **left-synchronized transducer** is a transducer such that each path leading from
an initial vertex to a final one can be divided in two parts: The first one only
contains edges of the form
\[ \{ p \xrightarrow{A/B} q \mid p, q \in Q \land A, B \in N \} \]
while the second part contains edges of the form
\[ \{ p \xrightarrow{A/\varepsilon} q \mid p, q \in Q \land A \in N \} \] exclusive or \[ \{ p \xrightarrow{\varepsilon/B} q \mid p, q \in Q \land B \in N \} \].

**Example 2.4** The following left-synchronized transducer recognizes the left-
synchronized relation of example 2.3.
A right-synchronized relation is defined symmetrically using a rational initial function. The left-synchronized relations form a subfamily of rational relations with useful closure properties.

**Theorem 2.5** [6] The synchronized relations form a boolean algebra.

We will use also particular left-synchronized relations. A binary relation $R$ is recognizable if it is a finite union of products $S \times T$ where $S, T \in \text{Rat}(N^*)$. A binary relation $R$ over words is of bounded length difference if there exists an integer $b$ such that $| |u| - |v|| \leq b$ for any $(u, v) \in R$.

**Proposition 2.6** [7] The family of synchronized relations contains the recognizable relations and the rational relations of bounded length difference.

Let $\mathcal{A}$ be a finite set of labels. A simple edge labelled graph is a subset of $V \times \mathcal{A} \times V$ where $V$ is an arbitrary set of vertices. For any label $a \in \mathcal{A}$, the $a$-transition of a graph $G$ is the relation $\xrightarrow{a} G := \{ (s, t) \mid (s, a, t) \in G \}$. A graph $G \subseteq N^* \times \mathcal{A} \times N^*$ is left-synchronized if for each $a \in \mathcal{A}$, the relation $\xrightarrow{a}$ is left-synchronized. An arbitrary graph is synchronized if it is isomorphic to some left-synchronized graph.

**Definition 2.7** A graph $G$ is synchronized (respectively rational, rational of bounded length difference) if it is isomorphic to some graph $G \subseteq N^* \times \mathcal{A} \times N^*$ such that for each $a \in \mathcal{A}$, the relation $\xrightarrow{a}$ is left-synchronized (respectively rational, rational and of bounded length difference).

Note that the synchronized graph family is also the closure by isomorphism of the right-synchronized graphs.

**Example 2.8** The following grid:
is synchronized because we can code its vertices by words to get the following left-synchronized graph \( G \) defined by \( \frac{a}{\varepsilon} = (A, A)^*(B, A)(B, B)^*(B, B) \) and \( \frac{b}{\varepsilon} = (A, A)^*(B, B)^*(B, B) \). Note that it is also a rational graph of bounded length difference.

Synchronized graphs are the automatic graphs of Blumensath and Grädel [1]. These graphs have a decidable first order theory. But the accessibility of these graphs is undecidable in general.

3 Traces of synchronized graphs

A trace of a graph \( G \) is the language \( L(G, I, F) \) of path labels leading from a set \( I \) of initial vertices to a set \( F \) of final vertices:

\[
L(G, I, F) = \{ u \mid \exists s \in I \exists t \in F, s \xrightarrow{u} t \}
\]

but with the condition that \( I \) and \( F \) are finite.

Morvan and Stirling [10] have proved that the traces of rational graphs are the context-sensitive languages. So any trace of a synchronized graph is a context-sensitive language. It remains to show that any context-sensitive language \( L \) is also the trace of a synchronized graph. We get this result by adapting the construction of [10].

We only need to find a left-synchronized graph \( G \subseteq N^* \times A \times N^* \) and two rational sets \( I, F \in \text{Rat}(N^*) \) such that \( L = L(G, I, F) \).

**Lemma 3.1** Let \( G \subseteq N^* \times A \times N^* \) be a left-synchronized graph.

Let \( I, F \in \text{Rat}(N^*) \) and \( i, f \notin N \).

There exists a left-synchronized graph \( H \subseteq (N^* \cup \{i, f\}) \times A \times (N^* \cup \{i, f\}) \) such that

\[
L(G, I, F) = L(H, \{i\}, \{f\}).
\]
Proof. i) For all $a \in \mathcal{A}$, the relation $\frac{a}{c} \mapsto$ is left-synchronized.

We define the graph $G'$ by erasing all edges of $G$ leading to a terminal state of $F$. This graph $G'$ is still left-synchronized as for all $a \in \mathcal{A}$, the relation 

$$\frac{a}{c'} := \frac{a}{c} \cap N^* \times F$$

is a synchronized relation as the intersection of a synchronized relation with a recognizable relation (using Theorem 2.5 and Proposition 2.6). For all $a \in \mathcal{A}$, we denote 

$$F_a := \text{Dom}(\frac{a}{c} \cap N^* \times F)$$

the set of vertices which are source of an erased edge. This set is rational as a domain of a rational relation. Then we create new edges leading from those vertices to the vertex $f$. More precisely, we define the graph $\overline{G}$ such that for all $a \in \mathcal{A}$,

$$\frac{a}{c} := \frac{a}{c'} \cup F_a \times \{f\}.$$ 

This relation is left-synchronized as the union of a left-synchronized relation with a recognizable set. Moreover and by construction,

$$L(G, I, F) = L(\overline{G}, I, \{f\}).$$

ii) Denoting by $\overline{u}$ the mirror of $u \in \mathcal{A}^*$ and by $G^{-1}$ the graph such that $p \frac{a}{c} q$ if and only if $q \frac{a}{c^{-1}} p$, we apply i) in order to get a unique initial vertex:

$$L(\overline{G}, I, \{f\}) = L(\overline{G}^{-1}, \{f\}, I) \overset{(i)}{=} L(\overline{G}^{-1}, \{f\}, \{i\}) = L(\overline{G}^{-1}, \{i\}, \{f\}).$$

There are different ways to characterize a context-sensitive language $L$. As Morvan and Stirling [10], we choose the ‘left’ form due to Penttonen [12].

**Definition 3.2** A rewriting system $\Gamma = \Gamma_1 \cup \Gamma_2$ is a 2-system if every rule of $\Gamma_2$ is of the form $AB \rightarrow AC$ with $B \neq C$ and every rule of $\Gamma_1$ is of the form $A \rightarrow a$ where $A, B, C$ are letters of the non-terminal alphabet $N$ and $a \in \mathcal{A}$.

Context-sensitive languages are obtained by derivation of a 2-system from a linear language.
Theorem 3.3 \[12\] There exists a linear language $L_{Lin}$ such that every context-sensitive language is \{ $v \in A^* \mid \exists u \in L_{Lin}, u \xrightarrow{\gamma} v$ \} for some 2-system $\Gamma$.

Given a context-sensitive language $L$, we first look for a graph $G_{Lin}$ such that $L = L(G_{Lin}, L_{Lin}, \{ \varepsilon \})$. Let $\Gamma$ be a 2-system. From $\Gamma$, we define the relation $R_2$ recognized by the following transducer $T_2$:

$I \xrightarrow{[B/A]} (A, B, A) \text{ for all } A, B \in N$ \hspace{1cm} (type 1)

$(A, B, C) \xrightarrow{B/D} (A, B, D) \text{ for all } A, B, C, D \in N \text{ such that } BC \xrightarrow{r_2} BD$ \hspace{1cm} (type 2)

$(A, B, C) \xrightarrow{D/C} (A, D, C) \text{ for all } A, B, C, D \in N$ \hspace{1cm} (type 3)

$(A, B, C) \xrightarrow{[A]} F \text{ for all } A, B, C \in N$ \hspace{1cm} (type 4)

This transducer starting at $I$ and ending at $F$ recognizes pairs of the form $([A_{A_1} \ldots A_{m}][B], [B_{B_1} \ldots B_{m}])$ meaning that under the successive context $A, A_1, \ldots, A_m$ the letter $B$ can be rewritten successively $B, B_1, \ldots, B_m$. If the context does not change: $A_i = A_{i+1}$, and one can apply a rule $A_iB_i \xrightarrow{r_2} A_{i+1}B_{i+1}$. Note that it is possible even if $B_i = B_{i+1}$ as a rule of type 3 can be applied with $B = D$. If the context changes: $A_i \neq A_{i+1}$, we copy the letter $B_i = B_{i+1}$.

Note that $R_2$ is a bounded length difference relation.

Example 3.4 Let $\Gamma_2 = \{ (AB, AC), (AC, AD), (DA, DE), (EA, EE) \}$. We have $[AAA][B] \ xrightarrow{R_2} [BCD]$ because under the context $A$, letter $B$ can be rewritten to $C$ and then to $D$. The following derivation:

$ABAA \xrightarrow{r_2} ACAA \xrightarrow{r_2} ADAA \xrightarrow{r_2} ADEA \xrightarrow{r_2} ADEE$

is represented as follows:

\[
\begin{array}{cccc}
\text{A} & \text{B} & \text{A} & \text{A} \\
\text{A} & \text{C} & \text{A} & \text{A} \\
\text{A} & \text{D} & \text{A} & \text{A} \\
\text{A} & \text{D} & \text{E} & \text{A} \\
\text{A} & \text{D} & \text{E} & \text{E} \\
\end{array}
\]

We have $[AAAA][B] \ xrightarrow{R_2} [BCDDD]$ and $[BCDDD][A] \ xrightarrow{R_2} [AAAEE]$ and $[AAAEE][A] \ xrightarrow{R_2} [AAAAE]$.

Let us give an elementary property of transducer $T_2$. 

7
Lemma 3.5 If \( I \xrightarrow{[UA]/[BVC]} \tau_2 \xrightarrow{s} \) with \( A, B, C \in N \) and \( U, V \in N^* \) then \( s = (B, A, C) \).

**Proof.** By induction on the length of any non-empty derivation from \( I \). □

Consider a word \( X_1 \in L_{Ln} \) of size \( n \) and a derivation \( X_1 \xrightarrow{\tau_2} X_2 \xrightarrow{\tau_2} \ldots \xrightarrow{\tau_2} X_m \) to a word \( X_m \). Given the \( m \) successive letters at a position \( i \) according to the derivation, the transducer gives the \( m \) successive letters at position \( i + 1 \).

For any words \( X, Y \in N^* \) of the same length \( n \), we denote by \( X \triangle Y \) the cardinal of \( \{ 1 \leq i \leq n \mid X(i) \neq Y(i) \} \).

Lemma 3.6 The two following properties are equivalent:

a) \( X_1 \xrightarrow{\tau_2} X_2 \xrightarrow{\tau_2} \ldots \xrightarrow{\tau_2} X_m \)

b) \( [X_1(i - 1)X_2(i - 1)\ldots X_m(i - 1)]X_1(i) \xrightarrow{R_2} [X_1(i)\ldots X_m(i)] \)

for all \( 2 \leq i \leq |X_1|; \)

\( |X_{j - 1}| = |X_j| \) , \( X_{j - 1} \triangle X_j = 1 \) and \( X_{j - 1}(1) = X_j(1) \) for all \( 2 \leq j \leq m \).

The words \( X_1, \ldots, X_m \) of same length \( n \) are represented as follows.

| \( X_1 \) | \( X_1(1) \) | \( X_1(2) \) | \ldots | \( X_1(i - 1) \) | \( X_1(i) \) | \ldots | \( X_1(n) \) 
| \( X_2 \) | \( X_2(1) \) | \( X_2(2) \) | \ldots | \( X_2(i - 1) \) | \( X_2(i) \) | \ldots | \( X_2(n) \) 
| \( X_3 \) | \( X_3(1) \) | \( X_3(2) \) | \ldots | \( X_3(i - 1) \) | \( X_3(i) \) | \ldots | \( X_3(n) \) 
| \ldots | \ldots | \ldots | \ldots | \ldots | \ldots | \ldots | \ldots |
| \( X_j \) | \( X_j(1) \) | \( X_j(2) \) | \ldots | \( X_j(i - 1) \) | \( X_j(i) \) | \ldots | \( X_j(n) \) 
| \ldots | \ldots | \ldots | \ldots | \ldots | \ldots | \ldots | \ldots |
| \( X_{m - 1} \) | \( X_{m - 1}(1) \) | \( X_{m - 1}(2) \) | \ldots | \( X_{m - 1}(i - 1) \) | \( X_{m - 1}(i) \) | \ldots | \( X_{m - 1}(n) \) 
| \( X_m \) | \( X_m(1) \) | \( X_m(2) \) | \ldots | \( X_m(i - 1) \) | \( X_m(i) \) | \ldots | \( X_m(n) \) 

**Proof.** i) Let us show that \( (a) \implies (b) \).

By definition of \( \Gamma_2 \), we have, for all \( 2 \leq j \leq m \),

\[ |X_{j - 1}| = |X_j| \text{ and } X_{j - 1} \triangle X_j = 1 \text{ and } X_{j - 1}(1) = X_j(1) \, . \]

Let us show that

\[ [X_1(i - 1)X_2(i - 1)\ldots X_m(i - 1)]X_1(i) \xrightarrow{R_2} [X_1(i)\ldots X_m(i)] \]
by induction on $m \geq 1$.

**Basis case**: $m = 1$. For all $2 \leq i \leq |X_1|$, we have

$$[X_1(i-1)]X_1(i) \ 	ext{R}_2 \ [X_1(i)]$$

considering the path

$$I \stackrel{[X_1(i-1)/X_1(i)]}{\xrightarrow{r_2}} (X_1(i), X_1(i-1), X_1(i)) \ 	ext{R}_2 \ [X_1(i)]$$

**Inductive case**: $m \Rightarrow m + 1$.

Suppose the implication for a derivation of length $m$ and let

$$X_1 \xrightarrow{r_2} \ldots \xrightarrow{r_2} X_m \xrightarrow{r_2} X_{m+1}.$$ 

There exists $2 \leq k \leq |X_1|$ such that $X_m(k) \neq X_{m+1}(k)$ and for all $i \neq k$, $X_m(i) = X_{m+1}(i)$.

Let $2 \leq i \leq |X_1|$. We want to show that

$$[X_1(i-1) \ldots X_m(i-1) X_{m+1}(i-1)]X_1(i) \ 	ext{R}_2 \ [X_1(i) \ldots X_{m+1}(i)].$$

By inductive hypothesis, we have

$$[X_1(i-1) \ldots X_m(i-1)]X_1(i) \ 	ext{R}_2 \ [X_1(i) \ldots X_m(i)].$$

Using Lemma 3.5, we have

$$I \stackrel{[X_1(i-1) \ldots X_m(i-1)/X_1(i) \ldots X_m(i)]}{\xrightarrow{r_2}} (X_1(i), X_m(i-1), X_m(i)).$$

We distinguish the two complementary cases below.

**Case 1**: $i \neq k$. We add an edge of type 3.

$$(X_1(i), X_m(i-1), X_m(i)) = (X_1(i), X_m(i-1), X_{m+1}(i))$$

$$X_{m+1}(i-1)/X_{m+1}(i) \xrightarrow{r_2} (X_1(i), X_{m+1}(i-1), X_{m+1}(i)).$$

**Case 2**: $i = k$. We have the rule $X_m(i-1) X_m(i) \ 	ext{R}_2 \ X_{m+1}(i-1) X_{m+1}(i)$.

To this rule is associated the following edge of type 2:

$$(X_1(i), X_m(i-1), X_m(i)) X_{m+1}(i-1)/X_{m+1}(i) \xrightarrow{r_2} (X_1(i), X_{m+1}(i-1), X_{m+1}(i)).$$

Finally, we add the edge leading to the final state:

$$(X_1(i), X_{m+1}(i-1), X_{m+1}(i)) \ 	ext{R}_2 \ [X_1(i)]$$

We get the result for $m + 1$ and the direct implication.
ii) Let us show that \( (b) \implies (a) \).
Suppose that \([X_1(i-1)\ldots X_m(i-1)]X_1(i)\ R_2\ [X_1(i)\ldots X_m(i)]\) for all \(2 \leq i \leq |X_1|\)
and \(|X_{j-1}| = |X_j|\) and \(X_{j-1} \triangle X_j = 1\) and \(X_1(j-1) = X_1(j)\) for all \(2 \leq j \leq m\).
Let \(2 \leq j \leq m\). Let us show that \(X_{j-1} \xrightarrow{r_2} X_j\).
As \(X_{j-1} \triangle X_j = 1\), there exists a unique \(2 \leq k \leq |X_1|\) such that \(X_{j-1}(k) \neq X_j(k)\).
Moreover \(X_{j-1}(1) = X_j(1)\) so \(k \neq 1\) and \(X_{j-1}(k-1) = X_j(k-1)\).
We have \([X_1(k-1)\ldots X_m(k-1)]X_1(k)\ R_2\ [X_1(k)\ldots X_m(k)]\).
Lemma 3.5 gives the existence of the following edge
\[(X_1(k), X_{j-1}(k-1), X_{j-1}(k)) \xrightarrow{r_2} X_j(k), X_j(k-1), X_j(k)).\]
This edge is of type 2 and gives the existence of the following rule of \(\Gamma_2\)
\[X_{j-1}(k-1)X_{j-1}(k) \xrightarrow{r_2} X_j(k-1)X_j(k)\].
\[
\]
Let \(L\) be a context-sensitive language obtained by derivation of a 2-system \(\Gamma\) from \(L_{Lin}\). Adding to \(T_2\) the set of edges \(\{F \xrightarrow{A/A} F \mid A \in N\}\), we get a transducer recognizing a rational graph \(G_{Lin}\) of bounded length difference with edges of the form \([U]\)AW \(\rightarrow\) [AV]W. If \(X_1 \in L_{Lin}\) with \(|X_1|=n\) and \(X_1 \xrightarrow{m-1} X_m\), the graph \(G_{Lin}\) contains the following path:
\[X_1(1)^m]X_1(2)\ldots X_1(n) \rightarrow [X_1(2)\ldots X_m(2)]X_1(3)\ldots X_1(n) \rightarrow [X_1(n)\ldots X_m(n)]\]
If we add edges of the form \([U] \rightarrow \varepsilon\) for any word \(U\) and if we label edges of \(G\) such that \([U]\)AW \(\xrightarrow{a}[AV]W\) if the last letter of \(U\) can be derived to \(a\) according to \(\Gamma_1\) then we get a left-synchronized graph \(G\) such that \(L = L(G, L_{Lin}, \{\varepsilon\})\). The problem is that \(L_{Lin}\) is not rational. In order to reduce \(L_{Lin}\) to a rational set, we complete \(T_2\) to a transducer generating words of \(L_{Lin}\) successively from left to right.
Let \(Gr\) be a grammar in Greibach normal form generating \(L_{Lin}\) from a non-terminal \(S\). Each rule of \(Gr\) is of the form \(Z \rightarrow AW\) where \(Z \in N_r\) is a non-terminal of \(Gr\), \(A \in N\) is a terminal (which is also a non-terminal of \(\Gamma\)) and \(W \in N^*_r\) is a non-terminal word of \(Gr\). Let the transducer
\[T_2^2 := T_2 \cup \{F \xrightarrow{Z|U} F' \mid (Z, U) \in Gr\} \cup \{F' \xrightarrow{Z|Z} F'' \mid Z \in N_r\}\]
10
where $F'$ is a new state of the transducer. We denote by $R'_2$ the relation recognized by $T'_2$ from $I$ to $F'$. This relation is still of bounded length difference. Let

$$L_{Rat} := \{ [A^m]B W | S \xrightarrow{2} AB W \land A, B \in N \land W \in N_r^* \land m \geq 1 \}.$$ 

Let us reformulate Lemma 3.6 for derivations starting from $L_{Lin}$.

**Lemma 3.7** Let $X_1, \ldots, X_m \in N^*$ and $n = |X_1|$. The two following properties are equivalent:

a) There exists $W_1, \ldots, W_{n-1} \in N_r^*$ such that

$$[X_1(1) \ldots X_m(1)]X_1(2)W_1 \in L_{Rat} \text{ and } W_{n-1} = \varepsilon$$

and $[X_1(n-1) \ldots X_m(n-1)]X_1(n) \xrightarrow{R_2} [X_1(n) \ldots X_m(n)]$

and $[X_1(i-1) \ldots X_m(i-1)]X_1(i)W_{i-1} \xrightarrow{R'_2} [X_1(i) \ldots X_m(i)]X_1(i+1)W_i$

for all $2 \leq i < n$

and $|X_j-i| = |X_j|$ and $X_{j-1} \triangle X_j = 1$ and $X_{j-1}(1) = X_j(1)$

for all $2 \leq j \leq m$.

**Proof.** i) We suppose (a) and show (b).

As $X_1 \in L_{Lin}$, we consider the derivation from $S$ to $X_1$ according to $Gr$: there exists non-terminal words $W_1, \ldots, W_{n-2}$ of $Gr$ such that

$$S \xrightarrow{2} X_1(1)X_1(2)W_1 \xrightarrow{2} \ldots \xrightarrow{2} X_1(1) \ldots X_1(n-1)W_{n-2} \xrightarrow{2} X_1(1) \ldots X_1(n)$$

By Lemma 3.6 we have for all $2 \leq i \leq |X_1|

$$[X_1(i-1) \ldots X_m(i-1)]X_1(i) \xrightarrow{R_2} [X_1(i) \ldots X_m(i)]$$

and $|X_{j-1}| = |X_j|$ and $X_{j-1} \triangle X_j = 1$ and $X_{j-1}(1) = X_j(1)$ for all $2 \leq j \leq m$.

Let $2 \leq i \leq n - 1$. We know that $W_i$ is obtained from $W_{i-1}$ by the rewriting of the non-terminal $W_{i-1}(1)$:

$$W_{i-1} = ZV \xrightarrow{gr} UV = X_2(i+1)W_i.$$ 

We complete the preceding path leading to $F$ with the edge $F \xrightarrow{Z/V} F'$ and then with edges $F' \xrightarrow{Z'/V'} F''$ for $V$. Thus, we have

$$[X_1(i-1) \ldots X_m(i-1)]X_1(i)W_{i-1} \xrightarrow{R'_2} [X_1(i) \ldots X_m(i)]X_1(i+1)W_i.$$ 

11
ii) We suppose (b) and show (a).

We cut the paths

\[ [X_1(i - 1) \ldots X_m(i - 1)] X_1(i) W_{i-1} R'_2 [X_1(i) \ldots X_m(i)] X_1(i + 1) W_i \]

which become

\[ [X_1(i - 1) X_2(i - 1) \ldots X_m(i - 1)] X_1(i) R_2 [X_1(i) \ldots X_m(i)] . \]

By Lemma 3.6, we have

\[ X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_m. \]

By hypothesis \([X_1(1) \ldots X_m(1)] X_1(2) W_1 \in L_{Rat} \) and \( X_1(1) = \ldots = X_m(1) \).

So \( S \xrightarrow{2_{Gr}} X_1(1) X_1(2) W_1 \). Thus \( S \xrightarrow{*_{Gr}} X_1(1) \ldots X_1(n) = X_1 \) hence \( X_1 \in L_{Lin} \). \( \square \)

The transducer \( T'_2 \) successively generates letters of \( X_1 \). Let us construct a graph of bounded length difference such that the language of path labels leading from the rational vertex set \( L_{Rat} \) to a rational vertex set \( F_{Rat} \) is the context-sensitive language defined by \( \Gamma \).

**Proposition 3.8** Any context-sensitive language is the language \( L(G, L_{Rat}, F_{Rat}) \) of path labels leading from a rational set of vertices \( L_{Rat} \) to another \( F_{Rat} \) and where \( G \) is a graph of bounded length difference.

**Proof.** Let \( L \) be a context-sensitive language. There exists a 2-system \( \Gamma \) such that

\[ L = \{ v \in A^* \mid \exists u \in L_{Lin}, \ u \xrightarrow{r} v \}. \]

For all letter \( a \in A \), we denote by

\[ N_a := \{ A \in N \mid A \xrightarrow{r_a} a \} \]

the set of non-terminals generating the terminal \( a \) in \( \Gamma \).

We define the graph \( G_0 \) such that for any \( a \in A \),

\[ \frac{a}{r_a} := R'_2 \cap [N^* N_a] N N^*_r \times ([N^+] N N^*_r \cup [N^+]) . \]

As \( R'_2 \) is a bounded length difference relation, so \( G_0 \) is and the following graph:

\[ G := G_0 \cup \bigcup_{a \in A} \{ [UA] \xrightarrow{a} [UA] \} \quad \text{for all} \ a \in N_a \]

is also of bounded length difference.

We recall that

\[ L_{Rat} := \{ [A^m] BW \mid S \xrightarrow{2_{Gr}} ABW \quad \text{and} \quad m \geq 1 \} \]

12
where $S$ is the axiom of $Gr$ and let

$$F_{\text{Rat}} := \lfloor [N^*] \rfloor.$$

We have

$$u \in L \text{ with } |u| = n > 1$$

$$\iff$$

there exists $X_1, \ldots, X_m \in N^*$ of length $n$ such that

$q \in L_{Lin}$ and $X_1 \xrightarrow{r_2} X_2 \xrightarrow{r_2} \ldots \xrightarrow{r_2} X_m$ and $X_m(i) \xrightarrow{r_1} u(i)$ for all $1 \leq i \leq n$

$$\iff \text{(by Lemma 3.7)}$$

there exists non-terminal words $W_1, \ldots, W_{n-1}$ of $Gr$ such that

$$[X_1(1) \ldots X_m(1)]X_1(2)W_1 \in I_{\text{Rat}}, W_{n-1} = \varepsilon$$

$$[X_1(1) \ldots X_m(1)]X_1(2)W_1 \xrightarrow{u(1)} [X_1(2) \ldots X_m(2)]X_1(3)W_2 \xrightarrow{u(2)} [X_1(3) \ldots X_m(3)] \xrightarrow{u(n-1)} [X_1(n) \ldots X_m(n)]$$

and $X_m(n) \in N_{u(n)}$

$$\iff$$

$$u \in L(G, L_{\text{Rat}}, F_{\text{Rat}})$$

Thus

$$L = L(G, L_{\text{Rat}}, F_{\text{Rat}}) \cup \{ u \in L \mid |u| \leq 1 \}$$

$\square$

It remains to apply Lemma 3.1 to get the following proposition:

**Proposition 3.9** Any context-sensitive language is trace of a synchronized graph.

Any synchronized graph is a rational graph, hence any trace of a synchronized graph is a context-sensitive language [10]. Proposition 3.9 gives the converse.

**Theorem 3.10** The context-sensitive languages are the traces of synchronized graphs.

Moreover, using Lemma 3.1, we get that any language $L(G, L_{\text{Rat}}, F_{\text{Rat}})$ of path labels leading from and to a rational vertex set of a graph $G$ of bounded length difference is a context-sensitive language as the trace of a synchronized (thus rational) graph. Proposition 3.9 gives the converse.
Theorem 3.11 The context-sensitive languages are the languages \( L(G, L_{\text{Rat}}, F_{\text{Rat}}) \) of path labels leading from and to a rational vertex set of a graph \( G \) of bounded length difference.

The synchronized relation of bounded length difference \( R_2' \) we used in the proof of Proposition 3.8 can be completed into a letter-to-letter relation.

Lemma 3.12 Let \( R \subseteq N^* \times N^* \) be a left-synchronized relation and let \( \diamond \) be a symbol such that \( \diamond \not\in N \). We can transform \( R \) into a letter-to-letter relation \( R_l \) such that

\[
\forall (U, V) \in N^* \times N^*, \forall n \geq 0, \quad (U \frac{n_V}{R} V) \iff (\exists k \geq 0, \exists k' \geq 0 \text{ such that } U \diamond k \frac{n_V}{R} V \diamond k')
\]

Let \( T \) be a left-synchronized transducer recognizing \( R \). We construct the transducer \( T_l \) from \( T \) replacing each edge of the form \( p \rightarrow^A q \) (respectively \( p \overset{A}{\rightarrow} q \)) with \( A \in N \) by the edge \( p \overset{A}{\rightarrow}^\diamond q \) (respectively \( p \overset{A}{\rightarrow}^\diamond q \)). Then for each final vertex \( f \) of \( T \), create a new final state \( f' \) of \( T_l \) and add the edges \( f \overset{\diamond}{\rightarrow} f' \) and \( f' \overset{\diamond}{\rightarrow} f' \).

\( \square \)

Proposition 3.13 Any context-sensitive language is the language \( L(G, L_{\text{Rat}}, F_{\text{Rat}}) \) of path labels leading from a rational set of vertices \( L_{\text{Rat}} \) to another \( F_{\text{Rat}} \) and where \( G \) is a letter-to-letter rational graph.

Using Proposition 2.6 we get that \( R_2' \) is a left-synchronized relation. Let \( \diamond \) be a symbol such that \( \diamond \not\in N \cup N_r \). Using Lemma 3.12 we complete \( R_2' \) into a letter-to-letter relation \( R_l \). We get the result adapting the proof of Proposition 3.8 with

\[
\text{D} \frac{a}{c_0} := R_l \cap [N^*N_a]NN^*_a \diamond ^*([N^+]N[N^*a^* \diamond ^* \cup [N^+] \diamond ^*)
\]

\[
G := G_0 \cup \bigcup_{a \in A} \{ [UA]\diamond k \frac{a}{c_0} S \|^{|UA|+k} | U \in N^* \land A \in N_a \}
\]

\[
L_{\text{Rat}} := \{ [A^m]BW \diamond k | S \frac{2}{c_0} ABW \land m \geq 1 \land k \geq 0 \}
\]

and

\[
F_{\text{Rat}} := S^+
\]

\( \square \)
The converse is given by Theorem 3.11.

**Theorem 3.14** The context-sensitive languages are the languages \( L(G, L_{\text{Rat}}, F_{\text{Rat}}) \) of path labels leading from and to a rational vertex set of a letter-to-letter rational graph \( G \).

### 4 Conclusion

Since synchronized binary relations form a boolean algebra and are recognized by deterministic 2-automata, the consideration of context-sensitive languages as traces of synchronized graphs could help for the conjecture of determinism of context-sensitive languages [8]: does any context-sensitive language can be recognized by a deterministic linear bounded Turing machine? The characterization of context-sensitive languages using rational letter-to-letter graphs could also be useful to solve this problem as every connex component of a rational letter-to-letter graph is a finite graph. In [2] Arnaud Carayol considers globally deterministic sets of transducers (i.e. in a case of non-determinism, only one output produced is accepted). He shows that the traces of those graphs with rational initial vertex sets are deterministic context-sensitive languages. His proof suggests that we could have worked directly on LBA Turing machine instead of using Pentonnen form.

### Acknowledgement

Many thanks to Didier Caucal for his support and his help during the redaction of this paper.
References


