The synchronized graphs trace the context-sensitive languages

Chloé Rispal

Electronic Notes in Theoretical Computer Science 68 No. 6 (2002)
URL: http://www.elsevier.nl/locate/entcs/volume68.html 16 pages

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

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 [3], the traces of rational graphs are the context-sensitive languages [10] and finally, the traces of Turing graphs are the recursively enumerable languages [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 [6] and [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 [1]. In this paper, we adapt the construction of Morvan and Stirling [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.

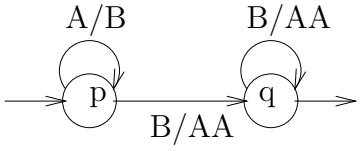
2 Rational synchronized graphs

Let N be a finite alphabet. We denote by N^* the set of words over letters of N, and we write ε 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 $p \xrightarrow[T]{u/v} q$ or by $p \xrightarrow{u/v} q$ when T is understood.

A path $p_0 \xrightarrow{u_1/v_1} p_1 \dots p_{n-1} \xrightarrow{u_n/v_n} p_n$ with $u = u_1...u_n$ and $v = v_1...v_n$ is labelled u/v and is denoted by $p_0 \xRightarrow[T]{u/v} p_n$. A path is *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 *rational* if it is recognized by a transducer.

Example 2.1 *The following transducer:*



with initial state p and final state q recognizes the rational relation

$$\{\ (A^nB^m\ ,\ B^nA^{2m})\ \mid\ n \geq 0, m > 0\ \}.$$

From studies concerning rational relations, Elgot and Mezei [6] and then Frougny and Sakarovitch [7] have defined the subfamily of left-synchronized relations.

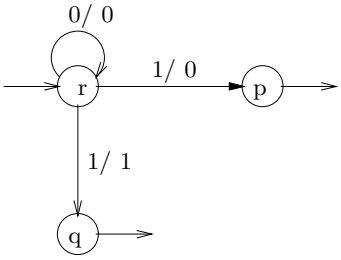
If a transducer has labels over $N \times N$ it is called a *letter-to-letter 2-automaton*: it is a transducer labelled by pairs of letters instead of pairs of words. Adding a rational terminal function completing one side of the recognized pairs, it recognizes a left-synchronized relation.

Definition 2.2 *A relation over $N^* \times N^*$ is left-synchronized if it is recognized by a letter-to-letter 2-automaton with terminal function taking values in*

$$Dif_{Rat} \quad = \quad (Rat(N^*) \times \{\varepsilon\})\ \cup\ (\{\varepsilon\} \times Rat(N^*))$$

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

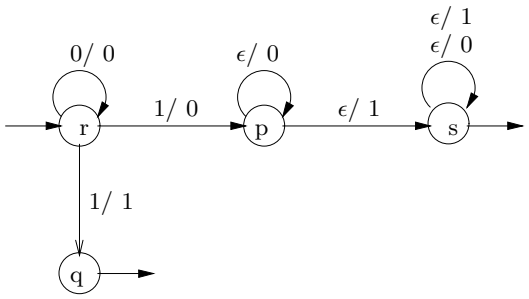
Example 2.3 *For all integer p, the relation $|_p$ defined by $x|_p y$ if x is a power of p dividing y, is left-synchronized. For instance, in base two with weak weigths on the left, $|_2$ is recognized by the following letter-to-letter 2-automaton:*



with the terminal function f defined by $f(p) = (\varepsilon, 0)^*(\varepsilon, 1)\{(\varepsilon, 0), (\varepsilon, 1)\}^*$ and $f(q) = (\varepsilon, \varepsilon)$

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
$\{p \xrightarrow{A/B} q | p, q \in Q \wedge A, B \in N\}$ while the second part contains edges of the form $\{p \xrightarrow{A/\epsilon} q | p, q \in Q \wedge A \in N\}$ exclusive or $\{p \xrightarrow{\epsilon/B} q | p, q \in Q \wedge B \in N\}$.

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 $S, T \in Rat(N^*)$. 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.*

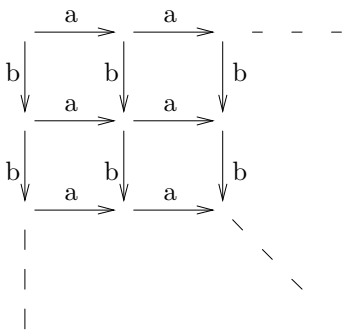
Let $\mathcal{A}$ be a finite set of labels. A simple edge labelled *graph* is a subset of $V \times \mathcal{A} \times V$ where V is an arbitrary set of *vertices*. For any label $a \in \mathcal{A}$, the *a-transition* of

a graph G is the relation $\xrightarrow[G]{a} := \{ (s,t) \mid (s,a,t) \in G \}$. A graph $G \subseteq N^* \times \mathcal{A} \times N^*$ is *left-synchronized* if for each $a \in \mathcal{A}$, the relation $\xrightarrow[G]{a}$ 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 $\xrightarrow[G]{a}$ 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 rigth-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 $\xrightarrow[G]{a} = (A,A)^*(B,A)(B,B)^*(\varepsilon,B)$ *and* $\xrightarrow[G]{b} = (A,A)^*(B,B)^*(\varepsilon,B)$. *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:

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

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 $G \subseteq N^*\times\mathcal{A}\times N^*$ and two rational sets $I, F \in Rat(N^*)$ such that $L = L(G,I,F)$.

Lemma 3.1 *Let $G \subseteq N^*\times\mathcal{A}\times N^*$ be a left-synchronized graph.*
Let $I, F \in Rat(N^*)$ and $i, f \notin N$.
There exists a left-synchronized graph $H \subseteq (N^* \cup \{i,f\})\times\mathcal{A}\times(N^* \cup \{i,f\})$ such that

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

Proof. i) For all $a \in \mathcal{A}$, the relation $\xrightarrow[G]{a}$ 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

$$\xrightarrow[G']{a} \ := \ \xrightarrow[G]{a} \ \cap \ \overline{N^* \times F}$$

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

$$F_a \ := \ Dom(\xrightarrow[G]{a} \ \cap \ N^* \times F)$$

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 $a \in \mathcal{A}$,

$$\xrightarrow[\overline{G}]{a} \ := \ \xrightarrow[G']{a} \ \cup \ F_a \times a \times \{f\}.$$

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

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

□

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 Γ_2 is of the form $AB \to AC$ with $B \neq C$ and every rule of Γ_1 is of the form $A \to a$ 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 $\{v \in \mathcal{A}^* \mid \exists\, u \in L_{Lin}\,,\ u \xrightarrow[\Gamma]{*} v\}$ for some 2-system Γ.*

Given a context-sensitive language L, we first look for a graph G_{Lin} such that $L = L(G_{Lin}, L_{Lin}, \{\varepsilon\})$. Let Γ be a 2-system. From Γ_2, we define the relation R_2 recognized by the following transducer T_2:

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

$$(A,B,C) \xrightarrow{B/D} (A,B,D) \quad \text{for all } A,B,C,D \in N \text{ such that } BC \underset{\Gamma_2}{\longrightarrow} BD \qquad \text{(type 2)}$$

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

$$(A,B,C) \xrightarrow{]A/]} F \quad \text{for all } A,B,C \in N \qquad \text{(type 4)}$$

This transducer starting at I and ending at F recognizes pairs of the form

$$([AA_1 \ldots A_m]B, [BB_1 \ldots B_m])$$

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 $A_iB_i \underset{\Gamma_2}{\longrightarrow} A_{i+1}B_{i+1}$. 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 $[AAA]B \ R_2 \ [BCD]$ because under the context A, letter B can be rewritten to C and then to D. The following derivation:*

$$ABAA \underset{\Gamma_2}{\longrightarrow} ACAA \underset{\Gamma_2}{\longrightarrow} ADAA \underset{\Gamma_2}{\longrightarrow} ADEA \underset{\Gamma_2}{\longrightarrow} ADEE$$

is represented as follows:

A	B	A	A
A	C	A	A
A	D	A	A
A	D	E	A
A	D	E	E

We have $[AAAAA]B \ R_2 \ [BCDDD]$ and $[BCDDD]A \ R_2 \ [AAAEE]$ and $[AAAEE]A \ R_2 \ [AAAAE]$.

Let us give an elementary property of transducer T_2.

Lemma 3.5 *If* $I \overset{[UA/[BVC}{\underset{T_2}{\Longrightarrow}} s$ *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 $X_1 \in L_{Lin}$ of size n and a derivation $X_1 \underset{\Gamma_2}{\longrightarrow} X_2 \underset{\Gamma_2}{\longrightarrow} \ldots \underset{\Gamma_2}{\longrightarrow} X_m$ 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) $X_1 \underset{\Gamma_2}{\longrightarrow} X_2 \underset{\Gamma_2}{\longrightarrow} \ldots \underset{\Gamma_2}{\longrightarrow} X_m$

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

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

Proof. i) Let us show that (a) $\Longrightarrow$ (b).

By definition of Γ_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

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

by induction on $m \geq 1$.

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

$$[X_1(i-1)]X_1(i) \ R_2 \ [X_1(i)]$$

considering the path

$$I \ \xrightarrow[T_2]{[X_1(i-1)/[X_1(i)} \ (X_1(i), X_1(i-1), X_1(i)) \ \xrightarrow[T_2]{]X_1(i)/]} \ F.$$

Inductive case : $m \Longrightarrow m+1$.

Suppose the implication for a derivation of length m and let $X_1 \underset{\Gamma_2}{\longrightarrow} \ldots \underset{\Gamma_2}{\longrightarrow} X_m \underset{\Gamma_2}{\longrightarrow} X_{m+1}$.

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

$$[X_1(i-1)\ldots X_m(i-1)X_{m+1}(i-1)]X_1(i) \ \ R_2 \ \ [X_1(i)\ldots X_{m+1}(i)].$$

By inductive hypothesis, we have

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

Using Lemma 3.5, we have

$$I \ \xrightarrow[T_2]{[X_1(i-1)\ldots X_m(i-1)/[X_1(i)\ldots X_m(i)} \ (X_1(i), X_m(i-1), X_m(i)).$$

We distinguish the two complementary cases below.

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

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

$$\xrightarrow[T_2]{X_{m+1}(i-1)/X_{m+1}(i)} (X_1(i), X_{m+1}(i-1), X_{m+1}(i)).$$

Case 2 : $i = k$. We have the rule $X_m(i-1)X_m(i) \ \ \Gamma_2 \ \ X_{m+1}(i-1)X_{m+1}(i)$.

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

$$(X_1(i), X_m(i-1), X_m(i)) \ \xrightarrow[T_2]{X_{m+1}(i-1)/X_{m+1}(i)} \ (X_1(i), X_{m+1}(i-1), X_{m+1}(i)).$$

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

$$(X_1(i), X_{m+1}(i-1), X_{m+1}(i)) \ \xrightarrow[T_2]{]X_1(i)/]} \ F.$$

We get the result for $m+1$ and the direct implication.

ii) Let us show that (b) $\Longrightarrow$ (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 $X_{j-1} \underset{\Gamma_2}{\longrightarrow} X_j$.
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

$$(X_1(k), X_{j-1}(k-1), X_{j-1}(k)) \overset{X_j(k-1)/X_j(k)}{\underset{T_2}{\longrightarrow}} (X_1(k), X_j(k-1), X_j(k)).$$

This edge is of type 2 and gives the existence of the following rule of Γ_2

$$X_{j-1}(k-1)X_{j-1}(k) \longrightarrow X_j(k-1)X_j(k).$$

□

Let L be a context-sensitive language obtained by derivation of a 2-system Γ 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 $[U]AW \to [AV]W$. If $X_1 \in L_{Lin}$ with $|X_1| = n$ and $X_1 \underset{\Gamma_2}{\overset{m-1}{\longrightarrow}} X_m$, the graph G_{Lin} contains the following path:

$$[X_1(1)^m]X_1(2)\ldots X_1(n) \to [X_1(2)\ldots X_m(2)]X_1(3)\ldots X_1(n) \ldots \to [X_1(n)\ldots X_m(n)]$$

If we add edges of the form $[U] \to \varepsilon$ 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 Γ_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 $Z \to AW$ where $Z \in N_r$ is a non-terminal of Gr, $A \in N$ is a terminal (which is also a non-terminal of Γ) and $W \in N_r^*$ is a non-terminal word of Gr. Let the transducer

$$T_2' := T_2 \cup \{F \xrightarrow{Z/U} F' \mid (Z, U) \in Gr\} \cup \{F' \xrightarrow{Z/Z} F' \mid Z \in N_r\}$$

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

$$L_{Rat} \ := \ \{ \ [A^m]BW \mid S \xrightarrow[Gr]{2} ABW \ \wedge \ A, B \in N \ \wedge \ W \in N_r^* \ \wedge \ m \geq 1 \ \}.$$

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) $X_1 \underset{\Gamma_2}{\longrightarrow} X_2 \underset{\Gamma_2}{\longrightarrow} \ldots \underset{\Gamma_2}{\longrightarrow} X_m$ *and* $X_1 \in L_{Lin}$

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

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

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

and $[X_1(i-1)\ldots X_m(i-1)]X_1(i)W_{i-1} \ R'_2 \ [X_1(i)\ldots X_m(i)]X_1(i+1)W_i$
for all $2 \leq i < n$

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

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

$$S \xrightarrow[Gr]{2} X_1(1)X_1(2)W_1 \underset{Gr}{\longrightarrow} \ldots \underset{Gr}{\longrightarrow} X_1(1)\ldots X_1(n-1)W_{n-2} \underset{Gr}{\longrightarrow} X_1(1)\ldots X_1(n)$$

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

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

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 $W_{i-1}(1)$:

$$W_{i-1} \ = \ ZV \underset{Gr}{\longrightarrow} UV \ = \ X_2(i+1)W_i.$$

We complete the preceeding path leading to F with the edge $F \xrightarrow{Z/U} F'$ and then with edges $F' \xrightarrow[T'_2]{Z/Z} F'$ for V. Thus, we have

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

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 $X_1 \underset{\Gamma_2}{\longrightarrow} X_2 \underset{\Gamma_2}{\longrightarrow} \ldots \underset{\Gamma_2}{\longrightarrow} X_m$.
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 $S \underset{Gr}{\overset{2}{\longrightarrow}} X_1(1)X_1(2)W_1$. Thus $S \underset{Gr}{\overset{*}{\longrightarrow}} X_1(1)\ldots X_1(n) = X_1$ hence $X_1 \in L_{Lin}$. □

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 Γ.

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 Γ such that

$$L \;\; = \;\; \{ \; v \; \in \mathcal{A}^* \mid \exists \; u \in L_{Lin} \, , \; u \underset{\Gamma}{\overset{*}{\longrightarrow}} v \; \}.$$

For all letter $a \in \mathcal{A}$, we denote by

$$N_a \;\; := \;\; \{ \; A \in N \mid A \underset{\Gamma_1}{\longrightarrow} a \; \}$$

the set of non-terminals generating the terminal a in Γ.
We define the graph G_0 such that for any $a \in \mathcal{A}$,

$$\underset{G_0}{\overset{a}{\longrightarrow}} \;\; := \;\; R'_2 \; \cap \; [N^*N_a]NN_r^* \times ([N^+]NN_r^* \cup [N^+]) \; .$$

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

$$G \;\; := \;\; G_0 \; \cup \; \bigcup_{a \in \mathcal{A}} \{ \; [UA] \overset{a}{\longrightarrow} [UA]\$ \mid U \in N^* \; \wedge \; A \in N_a \; \}$$

is also of bounded length difference.
We recall that

$$L_{Rat} \;\; := \;\; \{ \; [A^m]BW \mid S \underset{Gr}{\overset{2}{\longrightarrow}} ABW \; \wedge \; m \geq 1 \; \}$$

where S is the axiom of Gr and let

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

We have

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

$$\Longleftrightarrow$$

$$\text{there exists } \ X_1, \ldots, X_m \ \in N^* \ \text{ of length } n \text{ such that}$$
$$X_1 \ \in L_{Lin} \ \text{ and } \ X_1 \underset{\Gamma_2}{\longrightarrow} X_2 \underset{\Gamma_2}{\longrightarrow} \ldots \underset{\Gamma_2}{\longrightarrow} X_m \ \text{ and } \ X_m(i) \underset{\Gamma_1}{\longrightarrow} u(i) \ \text{ for all } 1 \leq i \leq n$$

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

$$\text{there exists non-terminal words } \ W_1, \ldots, W_{n-1} \ \text{ of } \ Gr \ \text{ such that}$$
$$[X_1(1)\ldots X_m(1)]X_1(2)W_1 \ \in I_{Rat} \ , W_{n-1} = \varepsilon$$
$$[X_1(1)\ldots X_m(1)]X_1(2)W_1 \underset{G_0}{\overset{u(1)}{\longrightarrow}} [X_1(2)\ldots X_m(2)]X_1(3)W_2 \underset{G_0}{\overset{u(2)}{\longrightarrow}} \ldots \underset{G_0}{\overset{u(n-1)}{\longrightarrow}} [X_1(n)\ldots X_m(n)]$$
$$\text{and } \ X_m(n) \in N_{u(n)}$$

$$\Longleftrightarrow$$

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

Thus

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

□

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 $L(G, L_{Rat}, F_{Rat})$ 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 3.8 gives the converse.

Theorem 3.11 *The context-sensitive languages are the languages $L(G, L_{Rat}, F_{Rat})$ 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 $\diamond \notin N$. We can transform R into a letter-to-letter relation R_l such that*
$\forall (U, V) \in N^* \times N^*, \forall n \geq 0,$

$$(U \xrightarrow[R]{n} V) \iff (\exists k \geq 0, \exists k^{'} \geq 0 \text{ such that } U\diamond^k \xrightarrow[R_l]{n} V\diamond^{k^{'}})$$

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

Proposition 3.13 *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 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 $\diamond \notin N \cup N_r$. 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

$$\xrightarrow[G_0]{a} \;:=\; R_l \;\cap\; [N^*N_a]NN_r^*\diamond^* \times ([N^+]NN_r^*\diamond^* \cup [N^+]\diamond^*)$$

$$G \;:=\; G_0 \;\cup\; \bigcup_{a \in \mathcal{A}} \{\; [UA]\diamond^k \xrightarrow{a} \$^{|[UA]|+k} \mid U \in N^* \;\wedge\; A \in N_a \;\}$$

$$L_{Rat} \;:=\; \{\; [A^m]BW\diamond^k \mid S \xrightarrow[Gr]{2} ABW \;\wedge\; m \geq 1 \;\wedge\; k \geq 0\}$$

and

$$F_{Rat} \;:=\; \$^+$$

□

The converse is given by Theorem 3.11.

Theorem 3.14 *The context-sensitive languages are the languages* $L(G, L_{Rat}, F_{Rat})$ *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 langages 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 machin instead of using Pentonnen form.

Acknowledgement

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

References

[1] A. Blumensath and E. Grädel, *Automatic structures*, LICS **15**, (2000), 51–62.

[2] A. Carayol, *A characterization of context-sensitive langages*, report, (2001).

[3] D. Caucal, *On infinite transition graphs having a decidable monadic theory*, ICALP **23**, LNCS **1099**, F. Meyer aufder, Heide, B. Monien, (1996), 194–205 [a full version will appear in Theoretical Computer Science].

[4] D. Caucal *On transition graphs of Turing machines*, MCU **3**, LNCS **2055**, M. Margenstern, Y. Rogozhin, (2201), 177–189.

[5] B. Courcelle *Graph rewriting: an algebraic and logic approach*, Handbook of Theoretical Computer Science, (1990) B, 193–242.

[6] C. Elgot and J. Mezei *On relations defined by generalized finite automata*, IBM J. Res. Dev. **9**,(1965)) 47–68.

[7] C. Frougny and J. Sakarovitch *Synchronized rational relations of finite and infinite words*, Theoretical Computer Science **108**, (1993), 45–82.

[8] S.-Y. Kuroda *Classes of languages and linear-bounded automata*, Information and Control 7,(1964), 207–223 .

[9] C. Morvan *On rational graphs*, FOSSACS **3**, Berlin, J. Tiuryn, LNCS 1784, (2000), 252–266.

[10] C. Morvan and C. Stirling *Rational graphs trace context-sensitive langages*, MFCS **26**, Mariánské, Lázně, A. Pultr and J. Sgall, LNCS, (2001).

[11] D. Muller and P. Schupp *The theory of ends, pushdown automata, and second-order logic*, TCS **37**, (1985), 51–75.

[12] Penttonen *One-sided and two-sided context in formal grammars*, Information and Control **25**, (1974), 371–392.

[13] M.O. Rabin *Decidability of second order-theories and automata on finite trees*, Trans. Amer. Math. Soc **141**, (1969), 1–35.

[14] M. Rabin and D. Scott *Finite automata and their decision problems*, IBM J. Res. Dev. **3**, (1959), 125–144; republished: E. Moore, *Sequential machines*, Addison-Wesley (1965).