On One-Rule Grid Semi-Thue Systems

Michel Latteux, Yves Roos

Fundamenta Informaticae XX (2012) 1–16
IOS Press

On One-Rule Grid Semi-Thue Systems

Michel Latteux
Laboratoire d'Informatique Fondamentale de Lille
Université Lille 1
michel.latteux@lifl.fr

Yves Roos
Laboratoire d'Informatique Fondamentale de Lille
Université Lille 1
yves.roos@lifl.fr

Abstract. The family of one-rule grid semi-Thue systems, introduced by Alfons Geser, is the family of one-rule semi-Thue systems such that there exists a letter c that occurs as often in the left-hand side as the right-hand side of the rewriting rule. We prove that for any one-rule grid semi-Thue system S, the set $S(w)$ of all words obtainable from w using repeatedly the rewriting rule of S is a constructible context-free language. We also prove the regularity of the set $\mathrm{Loop}(S)$ of all words that start a loop in a one-rule grid semi-Thue systems S.

Keywords: one-rule semi-Thue system, termination, grid semi-Thue system.

1. Introduction

Semi-Thue systems, that are the non-symmetrical version of Thue systems introduced by Axel Thue in 1914, serve as a model for rewriting systems. Thus they are of primordial interest for computational problems. For several years they have been intensively studied and several deep results have been obtained [10, 3, 2, 21, 19, 22]. However some intriguing decidability problems remain open. With a semi-Thue system S, one associates the set S_∞ of words that start an infinite derivation in S. The recursiveness of S_∞ is called the termination problem for S and the emptiness of S_∞ is the uniform termination problem for S. The best result on the set S_∞ is that the termination problem and the uniform termination problem are undecidable for 3-rules semi-Thue systems ([14]). Clearly the termination problem is decidable for length-preserving semi-Thue systems, contrarily to the uniform termination problem that has been shown undecidable for length-preserving semi-Thue systems

([1]). This result remains true for 9-rules semi-Thue systems ([19]) and for length-two semi-Thue systems ([18]).

Another set naturally associated with semi-Thue systems is the set $\mathrm{Loop}(S)$ that brings together the words w that start a non-null derivation in S toward a word containing w as a factor. For length-preserving semi-Thue systems, the emptiness of $\mathrm{Loop}(S)$ is equivalent to the emptiness of S_∞ but this equivalence does not hold for arbitrary semi-Thue systems, as shown in [7] where a 2-rules semi-Thue system S is presented with $\mathrm{Loop}(S) = \emptyset$ and $S_\infty \neq \emptyset$. Moreover, for an arbritary semi-Thue system S, the set $\mathrm{Loop}(S)$ need not be recursive.

One-rule semi-Thue systems are the simplest rewriting systems. Indeed, they are defined by two words l, r, noted $S = \{l \to r\}$. For a word w, $S(w)$ is the set of words obtainable from w by replacing repeatedly l by r. It is clear that, except the particular case $l = r$, $w \in S_\infty$ if and only if the set $S(w)$ is infinite. However, despite numerous efforts for more than twenty years, these decidability problems remain open and have become challenging problems. We observe that these problems can be explained in a few minutes to non-scientific people and surely they point out a deep lack of understanding of the rewriting notion. To get new ideas in order to solve these difficult problems, it seems natural to examine closely particular one-rule semi-Thue systems. Several interesting partial results have been obtained in that direction([20, 23, 16, 5, 6, 17]). Here we continue the study of a natural subclass of one-rule semi-Thue systems, introduced by Alfons Geser in [4], and called *one-rule grid semi-Thue systems*.

It is obvious that the system $S = \{l \to r\}$ with $l \neq r$ is uniformly terminating if $|r| \leq |l|$ or if there is a letter x such that $|r|_x < |l|_x$. So, when studying the termination, we can assume that there is a letter x such that $|r|_x > |l|_x$ and there is no letter y such that $|r|_y < |l|_y$. Then a borderline case is when there is a unique letter x such that $|r|_x > |l|_x$ and $|r|_y = |l|_y$ for $y \neq x$. The family of one-rule grid semi-Thue system satisfies a slightly more general condition. This family, that we denote $\mathcal{S}_{\mathrm{grid}}$, is composed of all one-rule semi-Thue systems having a letter c such that $|r|_c = |l|_c = k > 0$[1]. In [4], Alfons Geser has given a nice decidable characterization of one-rule grid semi-Thue systems that are uniformly terminating by proving that such a system is non-uniformly terminating if and only if it has a loop of length 1 or 2. To know whether a one-rule semi-Thue system has a loop of length 1 or 2 was previously shown decidable by Winfried Kurth in [11].

First, we give a new formulation of this decidable characterization of non-uniformly terminating systems in $\mathcal{S}_{\mathrm{grid}}$: there exist words x, y such that ly is a left factor of xr and xl is a right factor of ry. We show that the words x and y are unique and give a simple way to compute these two words. This permits, when considering a nonterminating system $S = \{l \to r\} \in \mathcal{S}_{\mathrm{grid}}$, to give a very precise form of the two words l and r and to get as a consequence that if $|r|_c = |l|_c = k$ is odd l needs to be a factor of r.

Then we prove that, for any S in $\mathcal{S}_{\mathrm{grid}}$, the set $\mathrm{Loop}(S)$ is a constructible regular language by giving a simple rational expression. This property can not be generalized to arbitrary 1-rule semi-Thue system: for instance $\mathrm{Loop}(\{c \to caca\})$ is not regular. Concerning the link between $\mathrm{Loop}(S)$ and S_∞, it is proved in [4] that the family $\mathcal{S}_{\mathrm{grid}}$ satisfies the equality $S_\infty = S^{-1}(A^*\mathrm{Loop}(S)A^*)$, that is $w \in S_\infty$ if and only if there is a derivation from w to a word in the regular language $A^*\mathrm{Loop}(S)A^*$ where A is the alphabet of lr. Unfortunately, up to now, this relation does not imply that S_∞ is a regular language as shown in [20] when $l \in a^*b^*$. Then we prove that for any S in $\mathcal{S}_{\mathrm{grid}}$ and for any

[1] In the original definition of Alfons Geser, a one-rule grid semi-Thue system is a one rule semi-Thue system $S = \{l \to r\}$ having a letter c such that $|r|_c \leq |l|_c$. In this paper we do not consider the case $|r|_c < |l|_c$ for which S is trivially uniformly terminating.

word w, the set $S(w)$ is a constructible context-free language. Note that this result does not hold for arbitrary one-rule semi-Thue systems ([12]). As a matter of fact, $S(w)$ is a bounded context-free language. So we get both the decidability of the termination problem[2] in $\mathcal{S}_{\mathrm{grid}}$ and the decidability of the common descendant problem in $\mathcal{S}_{\mathrm{grid}}$. In the last section, we give an argument in favour of the regularity of S_∞ by proving that it is a regular language in the case $|r|_c = |l|_c = 2$.

2. Preliminaries and Notations

In the following, A will denote a finite alphabet, A^* the free monoid over A and ε the empty word in A^*. For a word $w \in A^*$ and a letter $a \in A$, $|w|$ denotes the length of the word w and $|w|_a$ denotes the number of occurrences of the letter a in the word w.

Two words u and v are *conjugate* if there exist words x and y such that $u = xy$ and $v = yx$. It is well known that two words u and v are conjugate if and only if there exists a word z such that $uz = zv$. A word u is a *factor* of a word v if there exist two words w_1 and w_2 such that $v = w_1uw_2$ and we denote by $\mathrm{F}(v)$ the set of the factors of the word v. We denote by $\mathrm{RF}(w)$ (respectively $\mathrm{LF}(w)$) the set of *right factors* (respectively *left factors*) of the word w, that is:

$$\mathrm{RF}(w) = \{v \in A^* \mid \exists u \in A^*, w = uv\},$$

$$\mathrm{LF}(w) = \{u \in A^* \mid \exists v \in A^*,\ w = uv\}.$$

A semi-Thue system over an alphabet A is a subset $S \subseteq A^* \times A^*$. Members of S are denoted $l \underset{S}{\rightarrow} r$ (or $l \rightarrow r$ if there is no ambiguity). One-step derivation, denoted $\underset{S}{\rightarrow}$ ($\rightarrow$ if no ambiguity), is the binary relation over words defined by : $\forall u, v \in A^*, u \rightarrow v$ iff there exists $l \rightarrow r \in S$ and $\alpha, \beta \in A^*$ such that $u = \alpha l \beta$ and $v = \alpha r \beta$. The relation $\overset{*}{\rightarrow}$ (resp. $\overset{+}{\rightarrow}$) is the reflexive and transitive closure (resp. transitive closure) of the relation $\rightarrow$ and, for any word $w \in A^*$, we shall denote $S(w)$ the set $S(w) = \{w' \in A^* \mid w \overset{*}{\underset{S}{\rightarrow}} w'\}$ and $S^{-1}(w) = \{w' \in A^* \mid w' \overset{*}{\underset{S}{\rightarrow}} w\}$. We extend these notations to languages: for any language $L \subseteq A^*$, $S(L) = \cup_{w \in L} S(w)$ and $S^{-1}(L) = \cup_{w \in L} S^{-1}(L)$. For a derivation $w = w_0 \rightarrow w_1 \cdots \rightarrow w_n = w'$, n is called the length of the derivation that will be denoted by $w \overset{n}{\rightarrow} w'$.

We note $w \overset{\infty}{\underset{S}{\longrightarrow}}$ iff there is an infinite derivation starting on w and we denote by S_∞ the set $S_\infty = \{w \in A^* \mid w \overset{\infty}{\underset{S}{\longrightarrow}}\}$.

For a semi-Thue system S, for any positive integer n, we denote $\mathrm{Loop}_n(S) = \{w \in A^* \mid \exists x, y \in A^*, w \overset{n}{\rightarrow} xwy\}$ and $\mathrm{Loop}(S) = \cup_{n>0} \mathrm{Loop}_n(S)$.

The *termination problem* for the alphabet A and the semi-Thue system $S \subseteq A^* \times A^*$ is the following:

instance: a word $w \in A^*$

[2]This result already appears implicitly inside proofs used by Alfons Geser in [4] to solve the uniform termination problem in $\mathcal{S}_{\mathrm{grid}}$.

question: Does every derivation (modulo S) starting on w have finite length? (that is does $w \notin S_\infty$?)

The *uniform termination problem* for a class $\mathcal{S}$ of semi-Thue systems is the following :

instance: an alphabet A and a finite semi-Thue system $S \subseteq A^* \times A^*$ which belongs to $\mathcal{S}$

question: Do all derivations (modulo S) starting from all words $w \in A^*$ have finite length? (that is does $S_\infty = \emptyset$?)

To get shorter, we say that a system S *is nonterminating* if the uniform termination problem has a negative answer for S. In the sequel, we focus on $\mathcal{S}_{\mathrm{grid}}$, the family of one-rule grid semi-Thue systems.

Definition 2.1. A one-rule grid semi-Thue system $S = \{u \to v\}$ is a one-rule semi-Thue system such that there exists a letter c with $|u|_c = |v|_c = k > 0$.

3. Uniform termination in $\mathcal{S}_{\mathrm{grid}}$

In this section, we state some consequences of the following Geser's result:

Proposition 3.1. ([4]) A one-rule grid semi-Thue system $S = \{u \to v\}$ is non-uniformly terminating iff one of the following properties is satisfied[3]:

1. u is a factor of v.
2. there exist words g, h, k such that $u = gh$, $v = hk$ and ggh is a factor of hkk.

Example 3.1. The simplest example of one-rule grid semi-Thue system satisfying the property 2 of proposition 3.1 but not the property 1 is $S = \{cac \to acca\}$with $g = c$, $h = ac$, $k = ca$ and $ggh = ccac$ is a factor of $hkk = accaca$.

Let $S = \{u \to v\} \in \mathcal{S}_{\mathrm{grid}}$. The words u and v belong to A^* where A is an alphabet that contains a letter c with $|u|_c = |v|_c = k > 0$. If we denote $B = A \setminus \{c\}$, then

- $u = lcu_1 \dots u_{k-1}cr$
- $v = l'cv_1 \dots v_{k-1}cr'$

for some words $l, r, l', r', u_1, \dots, u_{k-1}, v_1, \dots, v_{k-1} \in B^*$. We get, as a consequence of the proposition 3.1, the following corollary that is also directly proved in [13]:

Corollary 3.1. Let $S = \{u \to v\} \in \mathcal{S}_{\mathrm{grid}}$ with $u \neq v$. Then S is nonterminating iff one of the two following properties is satisfied:

(i) u is a factor of v

(ii) there exist $x, y, s, e \in A^*$ such that $xv = uye$ and $vy = sxu$.

[3] More precisely, Alfons Geser proved in [4] that a one-rule grid semi-Thue system $S = \{u \to v\}$ is non-uniformly terminating if and only if $\mathrm{Loop}_1(S) \cup \mathrm{Loop}_2(S) \neq \emptyset$ that is proved decidable by Winfried Kurth in [11].

Proof: Clearly, if (ii) is satisfied, S is nonterminating since $xu \to xv = uye \to vye = sxue$. Now, if $u \notin \mathrm{F}(v)$ and if the property 2 of the proposition 3.1 is satisfied, then $hkk = sgghe$ for some s, e. Since $|g|_c = |k|_c$, it follows $s, e \in B^*$. Suppose that $g, k \in B^*$ then, since $h \in l'cA^*$ and $hkk = sgghe$, we get $l' = sggl'$ that implies $g = \varepsilon$ and $u = h \in \mathrm{F}(v)$, a contradiction. Thus $|g|_c = |k|_c > 0$ and $k = ye$ for some $y \in A^*$. Taking $x = g$, we get $xv = ghk = uye$, $vye = vk = hkk = sgghe = sxue$, hence $vy = sxu$. □

The following lemma states that if $u \neq v$ the conditions (i) and (ii) of the corollary 3.1 are mutually exclusive. This lemma also clarifies the relationship between the words x, y, s and e when the property (ii) of the corollary 3.1 is satisfied.

Lemma 3.1. Let $S = \{u \to v\} \in \mathcal{S}_{\mathrm{grid}}$. If the property (ii) of the corollary 3.1 is satisfied and $u \neq v$, we have

(i) $|x|_c = |y|_c > 0$ and $s, e \in B^+$,

(ii) $u \notin \mathrm{F}(v)$,

(iii) $u = xu'' = u'y$ where u'' is the longest word in $\mathrm{RF}(u) \cap \mathrm{LF}(v)$ and u' is the longest word in $\mathrm{LF}(u) \cap \mathrm{RF}(v)$,

(iv) x and y are conjugate; s and e are conjugate.

Proof: If the property (ii) of the corollary 3.1 is satisfied, there are e and s in A^* such that $xv = uye$ (1) and $vy = sxu$ (2). From (1), we get $|x|_c = |ye|_c$ and from (2), $|y|_c = |sx|_c$. So $|x|_c = |y|_c$ and $|e|_c = |s|_c = 0$ that is $e, s \in B^*$. Moreover $|x|_c = |y|_c > 0$ otherwise x and y belong to B^* and, from (1), we obtain $xl' = l$ and from (2), $l' = sxl$. It follows $x = s = \varepsilon$ and we similarly obtain $y = e = \varepsilon$. This leads to the contradiction $u = v$. Hence $x \in lcA^*$ and $y \in A^*cr$ so v belongs to $slcA^* \cap A^*cre$ and $l' = sl$, $r' = re$.

Let us now suppose that u is a factor of v. We prove that this leads to a contradiction. Indeed, if $v = s'ue'$ for some $e', s' \in B^*$, then the equality (1) gives $xs'ue' = uye \in A^*cre' \cap A^*cre$ and it follows that $e' = e$. Similarly equality (2) implies $s' = s$ and we obtain $xs'u = uy$ and $ue'y = xu$. Therefore $e' = s' = \varepsilon$ and $u = v$; this contradiction proves (ii).

We can now finish the proof of (i): from (1) and (2), we get $xsxu = xvy = uyey$ that implies $u \in \mathrm{LF}((xsx)^*)$. Suppose $s = \varepsilon$ then $u \in \mathrm{LF}((x)^*)$ and $vy = xu \in \mathrm{LF}(x^*)$. It follows $v \in \mathrm{LF}(x^*)$ that implies $u \in \mathrm{F}(v)$, a contradiction. Similarly we can show $e \neq \varepsilon$.

For (iii), let us first observe that $|x| < |u|$ and $|y| < |u|$: otherwise, we may assume that $|y| \leq |x|$; then $|x| \geq |u|$ and $x = uz$ for some word z. It follows that $xu = uzu \in \mathrm{RF}(vy)$ with $|y| \leq |x| = |zu|$ and u is a factor of v which leads to a contradiction. Then we have $u = xu'' = u'y$ with u'' in $\mathrm{RF}(u) \cap \mathrm{LF}(v)$ and u' in $\mathrm{LF}(u) \cap \mathrm{RF}(v)$. We shall now prove that u'' is the longest word in $\mathrm{RF}(u) \cap \mathrm{LF}(v)$ and we could similarly show that u' is the longest word in $\mathrm{LF}(u) \cap \mathrm{RF}(v)$.

We have $x = lcu_1 \ldots u_{j-1}cu'_j$ with $u'_j \in \mathrm{LF}(u_j)$ and, if we denote p the first index such that $u_p \neq v_p$ (p exists since $u \notin \mathrm{F}(v)$), we have $j \leq p$. Assume now that $j < p$ and set $z = lcu_1 \ldots u_{p-j}c$, then we have $xsz \in \mathrm{LF}(u)$ and $sxz \in \mathrm{LF}(v)$. But this implies $s = \varepsilon$ and $u_p = u_{p-j} = v_p$, a contradiction. Thus $p = j$ and it follows $u_j = u'_j sl$ and $v_j = u'_j l$ with $s \neq \varepsilon$. Moreover, since $ye \in \mathrm{RF}(ue) \cap \mathrm{RF}(v)$ with $|x|_c = |y|_c$, we obtain $k \geq 2p$ and $xsx' \in \mathrm{LF}(u)$, $sxx' \in \mathrm{LF}(v)$, with $x' = lcu_1 \ldots u_{p-1}c$.

In order to show that u'' is the longest word in $\mathrm{RF}(u) \cap \mathrm{LF}(v)$, assume that there exists some z such that $|z| < |x|$ and $u \in \mathrm{LF}(zv)$. Then $|z|_c < |x|_c$, $zsx \in \mathrm{LF}(xsx')$ and $sxsx = szsxt$ with $t \neq \varepsilon$. Hence xs is not a primitive word and $xs = (x_r s)^{q+1}$ where $x_r s$ is the primitive root of xs. That implies $x = (x_r s)^q x_r$ and $x' = (x_r s)x'_r$ with $x'_r \in \mathrm{LF}(x_r)$, $|x'_r|_c = |x_r|_c > 0$, $z = (x_r s)^i x_r$ with $i < q$ and $zsx = xzs$. Since $u \in \mathrm{LF}(zv)$, we have $xsx' \in \mathrm{LF}(zsxx') = \mathrm{LF}(xszx')$ and $x' = (x_r s)^q x'_r \in \mathrm{LF}(zx') = \mathrm{LF}((x_r s)^i x_r (x_r s)^q x'_r)$. Thus $(x_r s)^{q-i} x'_r \in \mathrm{LF}(x_r (x_r s)^q x'_r)$ and $sx'_r \in \mathrm{LF}(x_r s)$. But that implies $s = \varepsilon$, a contradiction that finishes the proof of (iii). Observe that it implies the unicity of x and y. Moreover, if there are several c such that $|u|_c = |v|_c$, the value of x and y does not depend on the choice of a particular c.

Let us now prove (iv). Since $|x|_c = |y|_c$, we can write $x = x_0 c x_1 \dots c x_t$ and $y = y_0 c y_1 \dots c y_t$ with $\forall 0 \leq i \leq t, x_i, y_i \in B^*$. From (1) and (2), we can deduce $xvy = xsxu = uyey$ and it follows that yey and xsx are conjugate. Thus $cy_t e y_0 c \dots cy_t y_0 c \in \mathrm{F}((xsx)^*)$. Since $s, e \in B^+$, it is easy to verify that we must have $y_t e y_0 = x_t s x_0$ and $y_t y_0 = x_t x_0$. Suppose that $|y_t| \geq |x_t|$ (the case $|y_t| \leq |x_t|$ is symmetric) then $y_t = x_t z$ for some z and it follows $zy_0 = x_0$. Then we get $x_t z e y_0 = x_t s z y_0$ so s and e are conjugate. Now from the equality $y_t e y y_0 = x_t s x x_0$ we obtain $x_t z e y y_0 = x_t s x z y_0$ that implies $zey = szy = sxz$ therefore x and y are conjugate. □

Since the equalities $vy = sxu$ and $xv = uye$ imply $xvy = xsxu = uyey$, it follows that $u \in \mathrm{LF}((xsx)^*), u \in \mathrm{RF}((yey)^*), v \in \mathrm{LF}((sxx)^*)$ and $v \in \mathrm{RF}((yye)^*)$. Then we can state the precise form of u and v for a nonterminating one-rule grid semi-Thue system:

Proposition 3.2. $S \in \mathcal{S}_{\mathrm{grid}}$ is nonterminating iff one of the following three conditions is satisfied

1. $u \in \mathrm{F}(v)$
2. $u = (xsx)^n x'$ and $v = (sxx)^n sx'$ with $n > 0$ and $x' \in \mathrm{LF}(x) \cap \mathrm{LF}(sx)$
3. $u = (xsx)^n xsx'$ and $v = (sxx)^n sxx'e$ with $n \geq 0, x = x'x''$ and $x''s = ex''$.

Proof: Let us first observe that these three conditions are sufficient by verifying that for any system S satisfying 1, 2 or 3, $\mathrm{Loop}(S) \neq \emptyset$. It is clear for 1 since, in this case, $u \in \mathrm{Loop}_1(S)$. If S satisfies 2 then $x(xsx)^n x' \in \mathrm{Loop}_2(S)$: indeed, in this case we get $x = x'x''$ and $sx' = x'x'''$ for some words x'', x''' and $x(xsx)^n x' \to x(sxx)^n sx' = (xsx)^n x'x''sx' \to (sxx)^n sx'x''sx' = sx(xsx)^n x'x'''$. Finally, if S satisfies 3, $x(xsx)^n xsx' \to x(sxx)^n sxx'e = (xsx)^n xsx'x''x'e \to (sxx)^n sxx'ex''x'e = (sxx)^n sxx'x''sx'e = sx(xsx)^n xsx'e$, therefore $x(xsx)^n xsx' \in \mathrm{Loop}_2(S)$. It now remains to prove that the condition (1 or 2 or 3) is necessary.

Suppose that 1 is not satisfied, then we can consider three cases since $u \in \mathrm{LF}((xsx)^*)$:

(i) $u = (xsx)^n x'$ with $x = x'x''$.
First we can observe that $v = (sxx)^n sx'$ since $v \in \mathrm{LF}((sxx)^*)$ and $|v| = |u| + |s|$. Moreover $n > 0$ since $|x|_c < |u|_c$. It remains to prove that $x' \in \mathrm{LF}(sx)$. The property (iii) of the lemma 3.1 gives $y \in \mathrm{RF}(u)$, therefore $y = x''x'$ since $|x| = |y|$. From the equality $xv = uye$ we obtain $x(sxx)^n sx' = (xsx)^n x'x''x'e$ and $x' \in \mathrm{LF}(sx')$ since $|e| = |s|$.

(ii) $u = (xsx)^n xs'$ with $s' \in \mathrm{LF}(s)$.
Since $v \in \mathrm{LF}((sxx)^*)$ and $|v| = |u| + |s|$, it follows $v = (sxx)^n sxx'$ with $|x'| = |s'|$. Then $x' \in B^*$ since $|u|_c = |v|_c$ and $s \in B^*$. It follows that $x = x'x''$ for some x''. Let us now consider the factorization $x = zct$ with $t \in B^*$ (we know that $|x|_c > 0$) then we obtain

$u = u_1cts'$ and $v = v_1ctx'$ with $|ts'| = |tx'|$ and $ts', tx' \in B^*$, $u_1, v_1 \in A^*$. This is a contradiction since $u = u_1cr$ and $v = v_1cre$ with $|re| > |r|$. The case $u = (xsx)^n xs'$ with $s' \in \mathrm{LF}(s)$ is finally impossible.

(iii) $u = (xsx)^n xsx'$ with $x' \in \mathrm{LF}(x)$.
Since $v \in \mathrm{LF}((sxx)^*)$ and, from the property (iv) of the lemma 3.1, $|v| = |u| + |e|$, we get $v = (sxx)^n sxx'x'''$ with $|x'''| = |e|$. Moreover $v \in \mathrm{RF}((yye)^*)$ therefore $x''' = e$ and $v = (sxx)^n sxx'e$. Now, from the equality $xv = uye$, we can deduce that $y = x''x'$ since $|x| = |y|$. Let us consider now the equality $vy = sxu$: we get $(sxx)^n sxx'ey = sx(xsx)^n xsx'$ and it follows $x'ey = xsx'$. Since $x = x'x''$ and $y = x''x'$, we finally obtain $ex'' = x''s$.

□

Example 3.2. Let $S = \{a^2ca^5c \to a^4ca^3ca^2\}$. In this case, $s = e = a^2$ and from the property (iii) of the lemma 3.1, $x = a^2ca$. Then we are in the case 3 of the proposition 3.2 with $n = 0$, $x' = a^2c$ and $x'' = a$.

The condition 2 of the proposition 3.2 implies $x' \in \mathrm{LF}(sx')$ so $x' \in B^*$. It follows that $|u|_c = 2n|x|_c$ is even. Similarly, the condition 3 of the proposition 3.2 implies that $x'' \in B^*$. Then $|x'|_c = |x|_c$ and $|u|_c = 2n|x|_c$ is even. We obtain:

Corollary 3.2. Let $S = \{u \to v\}$ such that $|u|_c = |v|_c$ is odd. Then S is nonterminating iff $u \in \mathrm{F}(v)$.

4. The construction of $\mathrm{Loop}(S)$ with S in $\mathcal{S}_{\mathrm{grid}}$

The aim of this section is to prove that for any system $S = \{u \to v\}$ in $\mathcal{S}_{\mathrm{grid}}$, the set $\mathrm{Loop}(S)$ is a constructible regular language. Clearly, if $u = v$ then $\mathrm{Loop}(S) = A^*uA^*$ and in the rest of this section, we suppose that $u \neq v$.

We shall use the following lemma:

Lemma 4.1. Let $S = \{u \to v\}$ in $\mathcal{S}_{\mathrm{grid}}$ that is nonterminating with $u \notin \mathrm{F}(v)$. If $z_0cw_0cz'_0 \xrightarrow{4} z_4cw_4cz'_4$ with $z_0, z'_0, z_4, z'_4 \in B^*$ and $|w_0| = |w_4|$ then $z_0cw_0cz'_0 \in B^*(xu + uy)B^*$.

Proof: We consider the derivation $z_0cw_0cz'_0 \to z_1cw_1cz'_1 \to z_2cw_2cz'_2 \to z_3cw_3cz'_3 \to z_4cw_4cz'_4$ with $z_i, z'_i \in B^*$. Observe first that $|cw_0c|_c > |u|_c$ otherwise we can not apply a derivation step on $z_1cw_1cz'_1$ since $u \notin \mathrm{F}(v)$. That implies that the first and the last occurrences of the letter c can not be both involved in a same step of the derivation. From the proposition 3.2, $|w_{i+1}| = |w_i|$ if the derivation involves the first or the last occurrence of the letter c, else $|w_{i+1}| > |w_i|$. From the hypothesis, we are in the first case and $\forall i, |w_i| = |w_0|$. We can also suppose that $z_0 = l$ and $z'_0 = r$. Then $u \in \mathrm{LF}(lcw_0cr) \cup \mathrm{RF}(lcw_0cr)$ and we consider the case $lcw_0cr = uy_0 \to vy_0$. Since $su \notin \mathrm{LF}(v)$, we necessarily have $u \in \mathrm{RF}(vy_0)$. It follows that $vy_0 = sx_1u \to sx_1v = suy_1e \to svy_1e = s^2x_2ue$. We can verify that $|y_0| = |y_1|$ and $y_0, y_1 \in \mathrm{RF}(u)$. Finally $y_0 = y_1$ and we get $vy_0 = sx_1u$ and $x_1v = uy_0e$. Now, from the property (ii) of the corollary 3.1 and from the property (iii) of the lemma 3.1, we get $y_0 = y$, $x_1 = x$ that imply $lcw_0cr = uy$. Observe that if $z_0 \neq l$ or $z'_0 \neq r$ we still have $z_0cw_0cz'_0 \in B^*uyB^*$. If $u \in \mathrm{RF}(lcw_0cr)$, then $z_0cw_0cz'_0 \in B^*xuB^*$ that proves the lemma. □

Let us denote, for any word $w \in A^*cA^*$, $\mathrm{int}(w) = (B^*)^{-1}w(B^*)^{-1} \cap cA^* \cap A^*c$.

Lemma 4.2. Let S be nonterminating with $u \neq v$.

1. If $w \xrightarrow{*} w'$ with $u \in \mathrm{F}(v)$ or $|w|_c > |u|_c$, then $|\mathrm{int}(w')| \geq |\mathrm{int}(w)|$,
2. if $w \xrightarrow{+} w' \xrightarrow{*} zwz'$ then $|\mathrm{int}(w)| = |\mathrm{int}(w')|$

Proof: For the property 1, it is sufficient to prove that $w \to w'$ implies $|\mathrm{int}(w')| \geq |\mathrm{int}(w)|$. That is clear if $u \in \mathrm{F}(v)$ since $|v| > |u|$ and $\mathrm{int}(v) = \mathrm{int}(u)$. If $u \notin \mathrm{F}(v)$ then the property (ii) of the corollary 3.1 is satisfied. Since $|w|_c > |u|_c$ we have to consider three cases:

1. $w = zuw''$ and $w' = zvw''$ with $z \in B^*$, $w'' \in (A^* \setminus B^*)$,
2. $w = w''uz$ and $w' = w''vz$ with $z \in B^*$, $w'' \in (A^* \setminus B^*)$,
3. $w = w''uw'''$ with $w'', w''' \in (A^* \setminus B^*)$.

Let us consider the case 1. From the property (iv) of the lemma 3.1, $|\mathrm{int}(u)| = |\mathrm{int}(v)| + |e|$, moreover $u \in A^*cr$ and $v \in A^*cre$ so we get $|\mathrm{int}(w')| = |\mathrm{int}(w)|$. The second case can be proved similarly and for the third case, we clearly obtain $|\mathrm{int}(w')| > |\mathrm{int}(w)|$ since $|u| < |v|$. For the property 2, we have $\mathrm{int}(zwz') = \mathrm{int}(w)$ and $w \in S_\infty$. So, if $u \notin \mathrm{F}(v)$, $|w|_c > |u|_c$ from the lemma 4.1. Then property 1 yields: $|\mathrm{int}(w)| \leq |\mathrm{int}(w')| \leq |\mathrm{int}(zwz')| = |\mathrm{int}(w)|$. □

Lemma 4.3. If the property (ii) of the corollary 3.1 is satisfied and $u \neq v$ then

1. $v \notin \mathrm{F}(xuB^*)$,
2. $v \notin \mathrm{F}(B^*uy)$,
3. $xu \notin A^*uA^+$,
4. $uy \notin A^+uA^*$.

Proof: From the property (iii) of the lemma 3.1, $u = gy$ where g is the longest word in $\mathrm{LF}(u) \cap \mathrm{RF}(v)$. Assume first that $v \in \mathrm{F}(xu)$. Since $v \in slcA^*$ and $xu \in lcA^*$, $v \notin \mathrm{LF}(xu)$. Then $u = g'y'$ with $g' \in \mathrm{LF}(u) \cap \mathrm{RF}(v)$ and $|y'| < |x| = |y|$. Thus $|g| < |g'|$, a contradiction. Assume now that $v \in \mathrm{F}(xuB^*)$. Since $v \in A^*cre$ and $u \in A^*cr$, we get $v \in \mathrm{RF}(xue)$ and $v = ue$, a contradiction. Assume now that $xu \in A^*uA^+$. Consider the equality $vy = sxu$. If $u \in \mathrm{LF}(xu)$ then $v = su$, a contradiction. Otherwise, $u = g'y'$ with $g' \in \mathrm{LF}(u) \cap \mathrm{RF}(v)$ and $|y'| < |y|$. Thus $|g| < |g'|$, a contradiction. Properties 2 and 4 can be proved similarly. □

We obtain as a consequence:

Corollary 4.1. If the property (ii) of the corollary 3.1 is satisfied and $u \neq v$ then $S^{-1}(B^*(xu + uy)B^*) = B^*(xu + uy)B^*$.

Proof: Let us prove that $S^{-1}(B^*xuB^*) \subseteq B^*(xu + uy)B^*$. It is clearly sufficient to consider a single step of derivation. Let $w_1uw_2 \to w_1vw_2 = w_1'xuw_2'$ with $w_1, w_2 \in A^*$ and $w_1', w_2' \in B^*$. From the lemma 4.3, $v \notin \mathrm{F}(xuB^*)$ therefore there exist $t \in B^*$ and $z \in A^*$ such that $w_1' = w_1t$, $w_2 = zw_2'$ and $txu = vz$. Since $v \in slcA^*$ and $x \in lcA^*$ it follows $t = s$ that implies $z = y$ and $w_1uw_2 = w_1uyw_2' \in B^*uyB^*$. We can prove similarly that $S^{-1}(B^*uyB^*) \subseteq B^*(xu + uy)B^*$. □

Proposition 4.1.

1. If $v = su$ with $s \in B^+$ then $\mathrm{Loop}(S) = \mathrm{RF}(s^*)uA^*$,
2. If $v = ue$ with $e \in B^+$ then $\mathrm{Loop}(S) = A^*u\mathrm{LF}(e^*)$,
3. If $v = sue$ with $s, e \in B^+$ then $\mathrm{Loop}(S) = \mathrm{RF}(s^*)u\mathrm{LF}(e^*)$.

Proof: Let $S = \{u \to su\}$ with $s \in B^+$ and $w \in \mathrm{Loop}(S)$. Then $w = \alpha u\beta \to \alpha su\beta \xrightarrow{*} z\alpha u\beta z'$ with $z, z' \in B^*$. From the lemma 4.2, we have $|\mathrm{int}(\alpha su\beta)| = |\mathrm{int}(\alpha u\beta)|$ and, since $s \in B^+$, we get $\alpha \in B^*$. Moreover we obtain by induction on the length $n > 0$ of the derivation that $z\alpha u\beta z' = \alpha s^n u\beta$. It follows $\alpha s^n = z\alpha$ and $\alpha \in \mathrm{RF}(\alpha s^{i\times n})$ for any integer $i > 0$. Therefore $\alpha \in \mathrm{RF}(s^*)$ and $\mathrm{Loop}(S) \subseteq \mathrm{RF}(s^*)uA^*$. Since the converse inclusion is immediate we finally get $\mathrm{Loop}(S) = \mathrm{RF}(s^*)uA^*$. The other cases can be proved similarly. □

Lemma 4.4. If the property (ii) of the corollary 3.1 is satisfied and $u \neq v$ then for any word $\alpha, \beta \in B^*$

1. $S(\alpha xu\beta) = \{\alpha s^n xue^n\beta \mid n \geq 0\} \cup \{\alpha s^{n-1}uye^n\beta \mid n > 0\}$,
2. $S(\alpha uy\beta) = \{\alpha s^n uye^n\beta \mid n \geq 0\} \cup \{\alpha s^{n-1}xue^n\beta \mid n > 0\}$.

Proof: We prove that $S(\alpha xu\beta) = \{\alpha s^n xue^n\beta \mid n \geq 0\} \cup \{\alpha s^{n-1}uye^n\beta \mid n > 0\}$: from the lemma 4.3, $xu \notin A^*uA^+$ and $uy \notin A^+uA^*$ therefore the first steps of a derivation from w are $w = \alpha xu\beta \to \alpha xv\beta = \alpha uye\beta \to \alpha vye\beta = \alpha sxue\beta$ and the property is proved by induction on the length of the derivation. The second property of this lemma can be proved similarly. □

Proposition 4.2. If the property (ii) of the corollary 3.1 is satisfied and $u \neq v$ then

$$\mathrm{Loop}(S) = \mathrm{RF}(s^*)(xu + uy)\mathrm{LF}(e^*)$$

Proof: Let $w \in \mathrm{Loop}(S)$ then $w \xrightarrow{n} zwz'$ with $z, z' \in B^*$ and $n > 0$. We can suppose $n \geq 4$ and there exists a derivation $w \to w_1 \to w_2 \to w_3 \to w_4 \xrightarrow{*} zwz'$ with $z, z' \in B^*$. From the lemma 4.2, we get $|\mathrm{int}(w_4)| = |\mathrm{int}(w)|$ and, from the lemma 4.1, we get $w = \alpha xu\beta$ or $w = \alpha uy\beta$ with $\alpha, \beta \in B^*$. Let us suppose that $w = \alpha xu\beta$ then $S(w) = \{\alpha s^n xue^n\beta \mid n \geq 0\} \cup \{\alpha s^{n-1}uye^n\beta \mid n > 0\}$ from the lemma 4.4. Moreover, from the lemma 4.3 we deduce that $xu \neq uy$ since $u \notin \mathrm{LF}(xu)$. On the other hand, from the property (iii) of the lemma 3.1, we have $xu \in lcA^*cr$ and $uy \in lcA^*cr$ then it follows that $\mathrm{int}(uy) \neq \mathrm{int}(xu) = \mathrm{int}(w)$. That implies that if $w = \alpha xu\beta \xrightarrow{+} zwz' = z\alpha xu\beta z'$ we must have $z\alpha xu\beta z' = \alpha s^n xue^n\beta$ for some $n > 0$ therefore $\alpha \in \mathrm{RF}(s^*)$ and $\beta \in \mathrm{LF}(e^*)$. The case $w = \alpha uy\beta$ is symmetric and we finally get $\mathrm{Loop}(S) \subseteq \mathrm{RF}(s^*)(xu+uy)\mathrm{LF}(e^*)$. We can easily verify the converse inclusion since $xu \xrightarrow{*} sxue$ and $uy \xrightarrow{*} suye$. □

We directly obtain from the two previous propositions:

Corollary 4.2. If $S \in \mathcal{S}_{\text{grid}}$ then $\text{Loop}(S)$ is a constructible regular language.

The condition $S \in \mathcal{S}_{\text{grid}}$ in the corollary 4.2 is necessary as shown by the following example:

Example 4.1. Consider $S = \{c \to caca\}$. We shall show that $\text{Loop}(S) \cap c(ca)^+a^* = L = c\{(ca)^n a^p \mid n > p \geq 0\}$. Since $c \xrightarrow{*} (ca)^{n-1}ca^{n-1}$, it follows that for $n > p \geq 0$, $c(ca)^n a^p \xrightarrow{*} c(ca)^n a^p a^{n-1-p}(ca)^{n-1}a^p$ and $L \subseteq \text{Loop}(S)$. For the reverse inclusion, remark that $(ca+a)^*$ is S-closed. So, if $cw = c(ca)^n a^p \xrightarrow{*} \alpha cw\beta$ for some $\alpha, \beta \in A^*$, then $\alpha = \varepsilon$ and $w \xrightarrow{+} w\beta$ with $|\beta|_c > 0$. Thus $\beta = \beta' a^p$, $(ca)^n \xrightarrow{*} (ca)^n a^p \beta' \in \text{LF}(S(c))$ and $p < n$ since $\forall z \in \text{LF}(S(c)), 2|z|_c > |z|_a$.

5. Testing membership in S_∞ with S in $\mathcal{S}_{\text{grid}}$

The aim of this section is to show that the termination problem of a system $S \in \mathcal{S}_{\text{grid}}$ is decidable. Recall that this problem is, given $S \in \mathcal{S}_{\text{grid}}$ and a word $w \in A^*$ to know whether $w \in S_\infty$. Clearly, for a system $S = \{u \to v\}$ with $u \in \text{F}(v)$, $S_\infty = A^*uA^*$ and we conjecture that for any system $S \in \mathcal{S}_{\text{grid}}$ the set S_∞ is a constructible regular language. This is true when $|u|_c = |v|_c = 2$ as it is proved in the next section but here we must find another way in order to prove that S_∞ is a recursive language for any $S \in \mathcal{S}_{\text{grid}}$. More precisely, we show that for any $S \in \mathcal{S}_{\text{grid}}$ and any word w, the set $S(w)$ is a constructible bounded context-free language.

Definition 5.1. ([9]) A language $L \subseteq A^*$ is said to be *bounded* if there exist words $w_1, \ldots, w_n \in A^*$ such that $L \subseteq w_1^* \ldots w_n^*$.

As a matter of fact, it would be sufficient to prove that for any $S \in \mathcal{S}_{\text{grid}}$ and any word w, the set $S(w)$ is a constructible context-free language in order to decide whether $S(w)$ is finite. Nevertheless bounded languages have nice structural properties that will also permit to solve in this section the common descendant problem for any $S \in \mathcal{S}_{\text{grid}}$ that is to decide, given two words w and w', whether $S(w) \cap S(w')$ is not empty. We also observe that the image $S(w)$ of a word w by an arbitrary one-rule semi-Thue system S need not be context-free([12]).

Following Alfons Geser in [4], we first establish that if a derivation is applied to a word long enough, this word can then be factorized so that the derivation applies independently on each factor of the word. Let us denote $N' = \max\{|cu_i| \mid i < k\}$ and $N = \max\{N', |l|, |r|\}$.

Lemma 5.1. Let $w = z'cm'm''cz''$ with $m', m'' \in B^*$ and $z', z'' \in A^*$. If $|m'| \geq N$ and $|m''| \geq N$ then $S(w) = S(z'cm')S(m''cz'')$.

Proof: Clearly $S(z'cm')S(m''cz'') \subseteq S(w)$. Conversely, let $w' \in S(w)$, we prove by induction on the length n of the derivation from w to w' that $w' \in S(z'cm')S(m''cz'')$. The base case $w = w'$ is immediate. Otherwise $w \to w'' \xrightarrow{n-1} w'$. Since $|m'm''| > N$, the first step of the derivation applies on an occurrence of u that can only appear in $z'cm'$ or in $m''cz''$. Suppose it is in $z'cm'$, then $z'cm' \to z_1'cm_1'$ with $m_1' \in B^*$, $|m_1'| \geq |m'|$ and $w'' = z_1'cm_1'm''cz''$. Now, we can apply the induction hypothesis on $w'' \xrightarrow{n-1} w'$: $w' \in S(z_1'cm_1')S(m''cz'') \subseteq S(z'cm')S(m''cz'')$. The other case can be proved similarly. $\square$

We directly deduce:

Lemma 5.2. Let $zcwcz'$ with $z, z' \in B^*$ and $w \in A^*$. If $|cwc| \geq 2 \times (|w|_c + 2) \times N$ then there exist two constructible words $w', w'' \in (A^* \setminus B^*)$ such that $zcwcz' = w'w''$ and $S(zcwcz') = S(w')S(w'')$.

Then we can state:

Proposition 5.1. Let $S \in \mathcal{S}_{\text{grid}}$ satisfying the property (ii) of the corollary 3.1.

1. For any word $w \in A^*$, the set $S(w)$ is a bounded context-free language.
2. $S_\infty = S^{-1}(A^*(xu + uy)A^*) = S^{-1}(A^*\text{Loop}(S)A^*)$.

Proof: The proof is based on an induction over $|w|_c$. For the property 2, we have only to prove $S_\infty \subseteq S^{-1}(A^*(xu+uy)A^*)$ since the inclusions $S^{-1}(A^*(xu+uy)A^*) \subseteq S^{-1}(A^*\text{Loop}(S)A^*) \subseteq S_\infty$ are clear.

- If $|w|_c < |xu|_c$ then $S(w)$ is finite (therefore bounded context-free) from the lemma 4.1,
- If $|w|_c = |xu|_c$ then
  - if $w \notin B^*(xu + uy)B^*$ then $w \notin S^{-1}(B^*(xu + uy)B^*)$ from the corollary 4.1 and it follows that $S(w)$ is finite from the lemma 4.1,
  - if $w \in B^*(xu + uy)B^*$ then, from the lemma 4.4, $S(w)$ is a (constructible) bounded context-free language and w is in $S^{-1}(A^*(xu + uy)A^*)$
- If $|w|_c > |xu|_c$, we make a new induction on $K_w = 2 \times (|w|_c) \times N - |\text{int}(w)|$. If $K_w \leq 0$ then it follows from the lemma 5.2 that there exist two constructible words $w', w'' \in (A^* \setminus B^*)$ such that $w = w'w''$ and $S(w) = S(w')S(w'')$. As $|w'|_c < |w|_c$ and $|w''|_c < |w|_c$, we can apply the induction hypothesis and it follows that $S(w)$ is a constructible bounded context-free language. If $K_w > 0$, let us denote $S_1 = \{w' \in A^* \mid w \xrightarrow{<4} w'\}$ and $S_2 = \{w' \in A^* \mid w \xrightarrow{4} w'\}$. Clearly $S(w) = S_1 \cup (\bigcup_{w' \in S_2} S(w'))$. The family of bounded languages is closed by finite union, consequently it remains to prove that for any word w' in S_2, $S(w')$ is a bounded context-free language. Since $w \notin B^*(xu + uy)B^*$ it follows from the lemma 4.1 and the lemma 4.2 that, for any word $w' \in S_2$, $|\text{int}(w')| > |\text{int}(w)|$. Then $K_{w'} < K_w$ and, by the induction hypothesis, $S(w')$ is a bounded context-free language, moreover if $w' \in S_\infty$ then $w' \in S^{-1}(A^*(xu + uy)A^*)$ and it follows that $w \in S^{-1}(A^*(xu + uy)A^*)$ which concludes the proof of the proposition.

□

When $u \in \text{F}(v)$, a similar proof can be used to show that the property 1 is also satisfied in this case. Then we get as a corollary:

Corollary 5.1. The termination problem is decidable for any system $S \in \mathcal{S}_{\text{grid}}$.

Since it is proved in [8] that the non-emptiness of the intersection of two bounded context-free languages is decidable, we also obtain:

Corollary 5.2. The common descendant problem is decidable for any system $S \in \mathcal{S}_{\text{grid}}$.

6. The special case $|u|_c = |v|_c = 2$

It has been shown in the section 4 that $\mathrm{Loop}(S)$ is regular for S in $\mathcal{S}_{\mathrm{grid}}$ and that this result does not hold for arbitrary one-rule semi-Thue system. The status of S_∞ is different. Géraud Sénizergues has proved in [20] that S_∞ is regular when $u \in a^*b^*$, but the regularity of S_∞ is an open problem for arbitrary one-rule semi-Thue system. Note that the regularity of S_∞ does not hold for two-rules grid semi-Thue system as shown by the example $S = \{aca \to c, cc \to cca\}$ since it is easy to see that $S_\infty = A^*cA^*cA^*cA^* \cup \{a^i ca^p ca^q \mid p \le i + q\}$. We conjecture that S_∞ is a constructible regular language for any S in $\mathcal{S}_{\mathrm{grid}}$. As we are unable to prove this fact, we give an argument in favour by proving it in the simpler case $|u|_c = |v|_c = 2$. The following proof shows also that the structure of S_∞ can be involved and that the proof in the general case S in $\mathcal{S}_{\mathrm{grid}}$ could be tricky.

Clearly the question arises only when S satisfies the property (ii) of the corollary 3.1 since, when $u \in \mathrm{F}(v)$, $S_\infty = A^*uA^*$. Then we consider in the following a nonterminating system $S = \{u \to v\}$ with $u = lcu_1cr$, $v = slcv_1cre$, $xv = uye$ and $vy = sxu$ for $u_1, v_1, s, e \in B^*$ and $x, y \in A^*$. Then there exist words u_1' and u_1'' such that $u_1 = u_1'sl = reu_1''$, $v_1 = u_1'l = ru_1''$, $x = lcu_1'$ and $y = u_1''cr$

Let us denote $S_\infty^{\min} = S_\infty \setminus (A\, S_\infty \cup S_\infty\, A)$ and, for any word $w \in A^*$, $L_w = S^{-1}(A^*w) \setminus (S_\infty \cup A\, S^{-1}(A^*w))$ and $R_w = S^{-1}(wA^*) \setminus (S_\infty \cup S^{-1}(wA^*)A)$.

Lemma 6.1. $S_\infty^{\min} \subseteq L_{xl}cu_1cR_r \cup L_lcu_1cR_{ry}$.

Proof: Let us consider $w_0 \in S_\infty^{\min}$ and let n be the length of the shortest derivation from w_0 to any word in $A^*(xu + uy)A^*$. Let us suppose that there exists a derivation from w_0 to $w_n \in A^*(xu)A^*$. Then $w_0 \xrightarrow{*} w_1 \xrightarrow{*} w_i \xrightarrow{*} w_n$ and $w_n = p_ncq_ncr_n$ with $p_n \in A^*xl$, $q_n = u_1$ and $r_n \in rA^*$. For any $i \in [0, n]$, let us consider the factorization $w_i = p_icq_icr_i$ with $|p_i|_c = |p_n|_c$ and $q_i \in B^*$. We claim that $q_0 = u_1$, else let j be the greatest index such that $q_j \ne u_1$. Then $w_{j+1} = p_{j+1}cu_1cr_{j+1}$ and $p_{j+1} \xrightarrow{*} p_n, r_{j+1} \xrightarrow{*} r_n$. This will lead to the following contradiction: there exists a derivation from w_0 to a word in $A^*(xu+uy)A^*$ whose length is strictly less than n. It follows that for any i, $q_i = u_1$ and $p_0 \xrightarrow{*} p_n$, $r_0 \xrightarrow{*} r_n$ therefore $p_0 \in S^{-1}(A^*xl), r_0 \in S^{-1}(rA^*)$. Since $w_0 \in S_\infty^{\min}$, it follows $p_0 \in L_{xl}, r_0 \in R_r$ and $w_0 \in L_{xl}cu_1cR_r$. Let us distinguish two cases for the step $w_j \to w_{j+1}$:

1. $q_j = u_1'l, r_j = u_1crz$ and $r_{j+1} = v_1crez$.
   Then $p_ncu_1' \in A^*lcu_1' = A^*x$ and $w' = p_ncq_jcr_j = p_ncu_1'lcu_1crz \in A^*xuA^*$. Clearly, the length of the derivation $w_0 \xrightarrow{*} w_j = p_jcq_jcr_j \xrightarrow{*} w' = p_ncq_jcr_j$ is strictly smaller than n.
2. $q_j = ru_1'', p_j = zlcu_1$ and $p_{j+1} = zslcv_1$.
   Then $u_1''cr_n \in u_1''crA^* = yA^*$ and $w' = p_jcq_jcr_n = zlcu_1cru_1''cr_n \in A^*uyA^*$. Clearly, the length of the derivation $w_0 \xrightarrow{*} w_j = p_jcq_jcr_j \xrightarrow{*} w' = p_jcq_jcr_n$ is strictly smaller than n.

□

Lemma 6.2. $S^{-1}(A^*u) \setminus A^*cu_1cr \subseteq S_\infty$.

Proof: Let $w \in S^{-1}(A^*u) \setminus A^*cu_1cr$. Then there exists a derivation: $w_0 \to w_1 \xrightarrow{*} w_i \xrightarrow{*} w_n \in A^*u$. Let j be the greatest index such that $w_j \notin A^*cu_1cr$. Then $w_j = p_juu_1''cr \in A^*uy$, and it follows $w \in S^{-1}(A^*uy) \subseteq S_\infty$. □

Let us denote $F = \{cv_1\} \cup \{cu_1crz \mid v_1 = rez\}$. Remark that F is finite and $S^{-1}(A^*l)F \subseteq S^{-1}(A^*xl)$.

Lemma 6.3. $L_{xl} \subseteq L_l F$.

Proof: Let $w \in L_{xl}$. Then $w \xrightarrow{*} zxl = zlcv_1$. Let us consider two cases:

1. If $w = w'cv_1$, we can deduce that $w' \xrightarrow{*} zl$, so $w' \in S^{-1}(A^*l)$. Since $w = w'cv_1 \notin S_\infty$, it follows $w' \notin S_\infty$ and since $w \notin AS^{-1}(A^*xl)$, it follows $w' \notin AS^{-1}(A^*l)$. Then $w' \in L_l$ and $w = w'cv_1 \in L_l F$.
2. Else $w = w'crz$ with $z \in B^*$ and $rz \neq v_1$. Then $w \xrightarrow{*} puz \to pvz \xrightarrow{*} p'cv_1$. We show that $rez = v_1$. Indeed, if it is not the case, $pslcv_1cr \in S^{-1}(A^*u) \setminus A^*cu_1cr \subseteq S_\infty$ which implies $pvz \in S_\infty$ and $w \in S_\infty$. Since $w = w'crz \xrightarrow{*} puz$, it follows $w'cr \in S^{-1}(A^*u) \setminus S_\infty \subseteq A^*cu_1cr$, so $w'crz = w''cu_1crz \xrightarrow{*} pvz \in A^*xl$. Finally $w'' \in S^{-1}(A^*l), w'' \notin S_\infty$ and $w'' \notin AS^{-1}(A^*l)$ which implies $w'' \in L_l$ and $w \in L_l F$.

□

Let us denote $H = \{z \mid l \in \mathrm{RF}(rez) \setminus \mathrm{RF}(rz)\}$. Remark that H is finite and $S^{-1}(A^*u)H \subseteq S^{-1}(A^*l)$.

Lemma 6.4. $L_l \subseteq l + L_u H$.

Proof: Let $w \in L_l$. If $w \in A^*l$ then $w = l$ otherwise $w = w'crz$ with $l \notin \mathrm{RF}(rz)$. Then there exists a derivation $w \xrightarrow{*} puz \to pvz$ with $pvz \in A^*crez \cap S^{-1}(A^*l)$. Suppose that $l \notin \mathrm{RF}(rez)$, then $pslcv_1cr \in S^{-1}(A^*u) \setminus A^*cu_1cr$ and it follows from the lemma 6.2 that pvz, and thus also w, is in S_∞. This contradiction implies $l \in \mathrm{RF}(rez)$ therefore $z \in H$ and $w'cr \in S^{-1}(A^*u)$. Moreover $w \notin S_\infty$ and $w \notin AS^{-1}(A^*l)$ imply that $w'cr \notin S_\infty$ and $w'cr \notin AS^{-1}(A^*u)$. Finally $w'cr \in L_u$ and $w \in L_u H$. □

Lemma 6.5. $L_u \subseteq L_l cu_1cr$.

Proof: Let $w \in L_u \subseteq S^{-1}(A^*u) \setminus S_\infty \subseteq A^*cu_1cr$. Then $w = w'cu_1cr \xrightarrow{*} w''cu_1cr$ with $w'' \in A^*l$. It follows that $w' \in S^{-1}(A^*l)$. Moreover $w' \notin S_\infty$ and $w' \notin AS^{-1}(A^*l)$ that imply $w' \in L_l$ and $w \in L_l cu_1cr$. □

Since, clearly, $S^{-1}(A^*l)cu_1crH \subseteq S^{-1}(A^*l)$, it follows from the lemmata 6.4 and 6.5 that $L_l \subseteq l(cu_1crH)^* \subseteq S^{-1}(A^*l)$. Then, from the lemma 6.3, we get $L_{xl} \subseteq l(cu_1crH)^*F \subseteq S^{-1}(A^*l)F \subseteq S^{-1}(A^*xl)$:

Lemma 6.6. $L_l \subseteq l(cu_1crH)^* \subseteq S^{-1}(A^*l)$ and $L_{xl} \subseteq l(cu_1crH)^*F \subseteq S^{-1}(A^*xl)$.

Symmetrically, denoting $F' = \{v_1c\} \cup \{zlcu_1c \mid v_1 = zsl\}$ and $H' = \{z \mid r \in \mathrm{LF}(zsl) \setminus \mathrm{LF}(zl)\}$, we get:

Lemma 6.7. $R_r \subseteq (H'lcu_1c)^*r \subseteq S^{-1}(rA^*)$ and $R_{ry} \subseteq F'(H'lcu_1c)^*r \subseteq S^{-1}(ryA^*)$.

Finally we obtain the following regular expression for S_∞:

Proposition 6.1. $S_\infty = A^*l(cu_1crH)^*(Fcu_1c + cu_1cF')(H'lcu_1c)^*rA^*$

Proof: From the lemmata 6.1, 6.6 and 6.7, we obtain the following inclusions:

$$S_\infty = A^*S_\infty^{\min}A^* \subseteq A^*(L_{xl}cu_1cR_r \cup L_lcu_1cR_{ry})A^*$$
$$\subseteq A^*(l(cu_1crH)^*Fcu_1c(H'lcu_1c)^*r + l(cu_1crH)^*cu_1cF'(H'lcu_1c)^*r)A^*$$
$$\subseteq A^*(S^{-1}(A^*xl)cu_1cS^{-1}(rA^*) \cup S^{-1}(A^*l)cu_1cS^{-1}(ryA^*))A^*$$
$$\subseteq A^*(S^{-1}(A^*(xu+uy)A^*))A^* = S_\infty$$

□

Example 6.1. Let $S = \{a^2ca^5c \to a^4ca^3ca^2\}$ that was used in the example 3.2. Here $s = e = l = a^2$ and $r = \varepsilon$. Recall that $x = a^2ca$ and $y = a^3c$. We deduce $F = ca^3 + ca^5ca$, $H = \varepsilon + a$, $F' = a^3c$ and $H' = \emptyset$. Then

$$S_\infty = (a+c)^*a^2(ca^5c + ca^5ca)^*[(ca^3 + ca^5ca)ca^5c + ca^5ca^3c](a+c)^*$$

7. Conclusion and open questions

This paper deals with the family $\mathcal{S}_{\mathrm{grid}}$ of one-rule semi-Thue systems that are a borderline case for termination. Some of the results obtained here are shown to be false for arbitrary one-rule semi-Thue systems. For instance, it is proved in the section 4 that, for S in $\mathcal{S}_{\mathrm{grid}}$, $\mathrm{Loop}(S)$ is a constructible regular language whereas for $S = \{c \to caca\}$, $\mathrm{Loop}(S)$ is not regular. It is also proved that, for S in $\mathcal{S}_{\mathrm{grid}}$, $S(w)$ is always a context-free language whereas for $S = \{ba \to a^2b^2\}$, $S(b^2a^2)$ is not context-free [12]. At the contrary, some other results obtained here have been already shown to be true for other particular classes of one-rule semi-Thue systems. So we give new arguments in favour of some conjectures concerning arbitrary one-rule semi-Thue systems. For instance, it is proved that, for S in $\mathcal{S}_{\mathrm{grid}}$, $S_\infty = S^{-1}(A^*\mathrm{Loop}(S)A^*)$. This equality appears already in [15] and [23] for other subclasses of one-rule semi-Thue systems. So we can think :

Conjecture 1. $S_\infty = S^{-1}(A^*\mathrm{Loop}(S)A^*)$ for any one-rule semi-Thue system S.

Note that the validity of this conjecture would imply that if S is a nonterminating one-rule semi-Thue system, then $\mathrm{Loop}(S)$ is not empty. Moreover, it would give a way to decide whether a word initiates an infinite derivation or not, as mentioned in [7]. For S in $\mathcal{S}_{\mathrm{grid}}$, it is proved that $A^*\mathrm{Loop}(S)A^*$, the *ideal* generated by $\mathrm{Loop}(S)$, is finitely generated. Does this property hold for arbitrary one-rule Thue systems? It is also proved here that, for $S = \{u \to v\}$ with $|u|_c = |v|_c = 2$, S_∞ is a constructible regular language. This result was previously proved in [20] by Géraud Sénizergues for one-rule semi-Thue system with a left hand-side in 0^*1^*. So we can hope that :

Conjecture 2. If S is a one-rule semi-Thue system, then S_∞ is a constructible regular language.

This property would imply the decidability of the (uniform) termination problem. However, remark that these two conjectures do not give a way to decide whether or not a word is in $\mathrm{Loop}(S)$.

Another possible extension of our results deals with one-rule grid Thue systems. For S in $\mathcal{S}_{\mathrm{grid}}$, one can consider the Thue system $\hat{S} = S \cup S^{-1}$ and ask: Is it true that $\forall w$, $\hat{S}(w)$ is a context-free language? Is it true that if S and S^{-1} are terminating then $\forall w$, $\hat{S}(w)$ is a finite language? Remark that this last question can be also raised when S is an arbitrary one-rule Thue system.

References

[1] Anne-Cécile Caron. Linear bounded automata and rewrite systems: Influence of initial configurations on decision properties. In *TAPSOFT, Vol.1*, pages 74–89, 1991.

[2] Max Dauchet. Termination of rewriting is undecidable in the one-rule case. In *MFCS*, pages 262–270, 1988.

[3] Nachum Dershowitz. Termination of rewriting. *J. Symb. Comput.*, 3(1/2):69–116, 1987.

[4] Alfons Geser. Decidability of termination of grid string rewriting rules. *SIAM J. Comput.*, 31(4):1156–1168, 2002.

[5] Alfons Geser. Loops of superexponential lengths in one-rule string rewriting. In *RTA*, pages 267–280, 2002.

[6] Alfons Geser. Termination of string rewriting rules that have one pair of overlaps. In *RTA*, pages 410–423, 2003.

[7] Alfons Geser and Hans Zantema. Non-looping string rewriting. *ITA*, 33(3):279–302, 1999.

[8] Seymour Ginsburg. *The Mathematical Theory of Context-Free Languages*. McGraw-Hill, Inc., New York, NY, USA, 1966.

[9] Seymour Ginsburg and Edwin H. Spanier. Bounded algol-like languages. *Transactions of the American Mathematical Society*, 113(2):333–368, 1964.

[10] John V. Guttag, Deepak Kapur, and David R. Musser. On proving uniform termination and restricted termination of rewriting systems. *SIAM J. Comput.*, 12(1):189–214, 1983.

[11] Winfried Kurth. One-rule semi-thue systems with loops of length one, two or three. *ITA*, 30(5):415–429, 1996.

[12] Michel Latteux and Yves Roos. The image of a word by a one-rule semi-Thue system is not always context-free. `http://www.lifl.fr/~yroos/al/one-rule-context-free.pdf`, 2011.

[13] Michel Latteux and Yves Roos. A new proof for the uniform termination problem of one-rule grid semi-Thue systems. `http://www.lifl.fr/~yroos/al/one-rule-utermination.pdf`, 2011.

[14] Yuri Matiyasevich and Géraud Sénizergues. Decision problems for semi-thue systems with a few rules. In *LICS*, pages 523–531, 1996.

[15] Robert Mcnaughton. Well behaved derivations in one-rule semi-thue systems. Technical Report 95-15, Dept. of Computer Science, Rennsselaer Polytechnic Institute, 1995.

[16] Robert Mcnaughton. Semi-thue systems with an inhibitor. *J. Autom. Reason.*, 26(4):409–431, 2001.

[17] Wojciech Moczydlowski and Alfons Geser. Termination of single-threaded one-rule semi-thue systems. In *RTA*, pages 338–352, 2005.

[18] Masahiko Sakai and Yi Wang. Undecidable properties on length-two string rewriting systems. *Electronic Notes in Theoretical Computer Science*, 204:53 – 69, 2008. Proceedings of the 7th International Workshop on Reduction Strategies in Rewriting and Programming (WRS 2007).

[19] Géraud Sénizergues. Some undecidable termination problems for semi-thue systems. *Theor. Comput. Sci.*, 142(2):257–276, 1995.

[20] Géraud Sénizergues. On the termination problem for one-rule semi-thue system. In *RTA*, pages 302–316, 1996.

[21] Celia Wrathall. Confluence of one-rule thue systems. In *IWWERT '90: Proceedings of the First International Workshop on Word Equations and Related Topics*, pages 237–246, London, UK, 1992. Springer-Verlag.

[22] Wang Yi, Sakai Masahiko, and Nishida Naoki. Confluence of length preserving string rewriting systems is undecidable(theory of computer science and its applications). *RIMS Kokyuroku*, 1554:171–177, 2007.

[23] Hans Zantema and Alfons Geser. A complete characterization of termination of 0^p 1^q-> 1^r 0^s. *Appl. Algebra Eng. Commun. Comput.*, 11(1):1–25, 2000.