On Quasi Orders of Words and the Confluence Property

Tero HARJU

Mathematics Department, University of Turku
FIN-20014 Turku, Finland

Lucian ILIE

Turku Centre for Computer Science TUCS
FIN-20520 Turku, Finland

Turku Centre for Computer Science
TUCS Technical Report No 129
September 1997

ISBN 952-12-0064-2
ISSN 1239-1891

Abstract

We investigate the confluence property, that is, the property of a language to contain, for any two words of it, one which is bigger, with respect to a given quasi order on the respective free monoid, than each of the former two. This property is investigated mainly for regular and context-free languages. As a consequence of our study, we give an answer to an old open problem raised by Haines concerning the effective regularity of the sets of subwords. Namely, we prove that there are families with a decidable emptiness problem for which the regularity of the sets of subwords is not effective.

1 Introduction

The investigation of the possibilities to obtain languages of finite words from infinite words, or conversely, constitutes an important aspect of the investigation of the combinatorial structures of formal languages. There are several classical ways to associate a set of finite words to an infinite word: one can take the set of all finite prefixes or finite factors of it or the set of all finite words which are not prefixes of it, [AG], [AFG], [Ber], [MP]. Conversely, for a language of finite words, one can associate infinite words considering the notions of limit [Ei] or adherence [BoN].

As is easy to see, a language L obtained by taking all finite factors (or prefixes) of an infinite word has the following property: for any two words $u, v \in L$ there is a word $w \in L$ (w is not neccessarily different from u and v) such that both u and v are factors of w. From this, there naturally arises the problem of investigating this property for arbitrary languages and, moreover, for arbitrary quasi orders on the free monoid (instead of the factor partial order). We call this property *confluence* (from the name of a similar property for binary relations) and we investigate it mainly for regular and context-free languages. As a consequence of our study, we give an answer to an open problem raised by Haines in [Ha]. While the concept of well-quasi-ordering has been frequently discovered, see, for instance, [ER], [Hi], [Kr1], [Ha] ([Kr2] gives a complete account on this), Haines seems to have been the first to prove that both sets, of subwords and of superwords, of any language are regular. Since it may happen that the two sets are not effectively regular, there naturally arises the problem of investigating when it is possible to find regular expressions of them. Van Leeuwen [vL] solved the problem for superwords and gave a method to find the set of subwords of any context-free language. It remained open whether or not there are families with a decidable emptiness problem for which the regularity of the sets of subwords is not effective. We give here a positive answer to this problem.

The paper is organized as follows. After giving the basic definitions in the next section, we prove in section 3 that the confluence problem w.r.t. the factor partial order is decidable for regular languages.

In section 4 we prove that the confluence problem w.r.t. the prefix partial order is decidable for context-free languages. The same problem for the factor partial order is undecidable. Some other undecidability results about context-free languages as well as some general undecidability results concerning the confluence property are presented. We mention that [Bea1], [Bea2], and [BeaN] contain results which are related to our sections 3 and 4.

In section 5 we deal with terminating relations.

Section 6 contains the answer to Haines' question, that is the existence of families of languages having a decidable emptiness problem but for which the regularity of the sets of subwords is not effective. The proof of this result relies heavily on two facts: the decidability of the confluence problem w.r.t. the subword partial order of regular languages and a general confluence property of languages, as proved in section 3.

In the last section, we give some results concerning closure properties.

2 Definitions

For an alphabet Σ, we denote by Σ^* the free monoid generated by Σ and by λ, the empty word, its identity. The free semigroup generated by Σ is denoted by Σ^+.

A *right-infinite* word over Σ is a word which is unbounded from the right and it is viewed as a function $\alpha : \mathbb{Z}_+ \longrightarrow \Sigma$ from the set of positive integers into the free monoid generated by Σ. Analogously, a *left-infinite* word is a function $\alpha : \mathbb{Z}_- \longrightarrow \Sigma$. The set of all right-(left-)infinite words is denoted by Σ^ω (${}^\omega\Sigma$). For a finite word $w \in \Sigma^*$, $|w|$ denotes the length of w and we put $w^\omega = www\ldots \in \Sigma^\omega$ and ${}^\omega w = \ldots www \in {}^\omega\Sigma$.

A *bi-infinite* (two-sided) word is an infinite word without any end. We can define a bi-infinite word as a function $\alpha : \mathbb{Z} \longrightarrow \Sigma$ or, in fact, as an equivalence class of the set $\Sigma^\mathbb{Z}$ with respect to the equivalence relation defined for $\alpha, \beta \in \Sigma^\mathbb{Z}$ by $\alpha \sim \beta$ if and only if there is an integer k such that for any $n \in \mathbb{Z}, \alpha(n) = \beta(n+k)$. We denote by ${}^\omega\Sigma^\omega$ the set of all bi-infinite words over Σ.

For an infinite word $\alpha \in \Sigma^\omega \cup {}^\omega\Sigma \cup {}^\omega\Sigma^\omega$, we denote by $Fact(\alpha)$ the set of all finite factors of α, for $\alpha \in \Sigma^\omega$, $Pref(\alpha)$ denotes the set of all finite prefixes of α, and for $\alpha \in {}^\omega\Sigma$, $Suf(\alpha)$ denotes the set of all finite suffixes of α.

For other formal language theory notions and results we refer to [HU] and [Sa].

Given a quasi order $\leq$ on Σ^*, we say that a language $L \subseteq \Sigma^*$ is *confluent w.r.t.* $\leq$ if and only if for any $x, y \in L$ there is $z \in L$ such that $x \leq z$ and $y \leq z$. Notice that the empty language is confluent w.r.t. any quasi order on Σ^*.

Some of the most important partial orders on Σ^* in combinatorics on words, see [CK], are as follows, where we suppose that u, v are arbitrary finite words over Σ:

prefix:	$u \leq_p v$ iff there is $w \in \Sigma^*$ such that $v = uw$,
factor:	$u \leq_f v$ iff there are $w, z \in \Sigma^*$ such that $v = wuz$,
subword:	$u \leq_s v$ iff $u = a_1 a_2 \ldots a_n, n \geq 0, a_i \in \Sigma, 1 \leq i \leq n,$ $v = v_1 a_1 v_2 a_2 \ldots v_n a_n v_{n+1}, v_i \in \Sigma^*, \text{for } 1 \leq i \leq n+1.$

(Notice that what we call here a factor and a subword is called sometimes a subword and a scattered subword, respectively.)

The *down-set of L w.r.t.* $\leq$, denoted $\text{DOWN}_{\leq}(L)$, is the set

$$\text{DOWN}_{\leq}(L) = \{w \in \Sigma^* \mid w \leq u \text{ for some } u \in L\}.$$

For instance, $\text{DOWN}_{\leq_f}(L)$ is the set of all factors of words in L, $\text{DOWN}_{\leq_p}(L)$ is the set of all prefixes of words in L, etc. The down-operator $\text{DOWN}_{\leq}$ is monotone and idempotent.

Examples

1. The regular language

$$L_1 = a^* \cup b^*$$

is confluent w.r.t. none of the partial orders $\leq_p, \leq_f$, or $\leq_s$ since, for any nonempty $w_1 \in a^*, w_2 \in b^*$, there is no word in L_1 which contains both w_1 and w_2 as factors, subwords, or prefixes, respectively.

2. The context-free language

$$L_2 = \{a^n b^n \mid n \geq 0\}$$

is confluent w.r.t. $\leq_f$ because, for any $a^n b^n, a^m b^m \in L_2$, we have, for any $p \geq \max\{m, n\}$, $a^n b^n \leq_f a^p b^p, a^m b^m \leq_f a^p b^p$, and $a^p b^p \in L_2$. It follows that L_2 is confluent w.r.t. $\leq_s$ too. But, since there is no word in L_2 which has as prefixes both ab and $a^2 b^2$, L_2 is not confluent w.r.t. $\leq_p$. Notice that also Σ^* is not confluent w.r.t. $\leq_p$, for any alphabet $\Sigma, |\Sigma| \geq 2$.

3. Consider two morphisms $g, h : \Sigma^* \longrightarrow \Delta^*$ and the associated equality set

$$E(g, h) = \{w \in \Sigma^+ \mid g(w) = h(w)\}.$$

Because $E(g, h)$ is closed under concatenation, it follows that it is confluent w.r.t. both partial orders $\leq_f$ and $\leq_s$.

In fact, this a particular case of a more general result: if a language L satisfies $L = L^+$, then L is confluent w.r.t. any quasi order $\leq$ which is an extension of $\leq_f$.

3 Confluence of regular languages

For a quasi order $\leq$ on Σ^* we define the following families of languages:

$$\begin{aligned}
\mathcal{F}^r_{\leq} &= \{L \subseteq \Sigma^* | L = \text{DOWN}_{\leq}(Pref(\alpha)), \text{ for some } \alpha \in \Sigma^{\omega}\},\\
\mathcal{F}^l_{\leq} &= \{L \subseteq \Sigma^* | L = \text{DOWN}_{\leq}(Suf(\alpha)), \text{ for some } \alpha \in {}^{\omega}\Sigma\},\\
\mathcal{F}^{bi}_{\leq} &= \{L \subseteq \Sigma^* | L = \text{DOWN}_{\leq}(Fact(\alpha)), \text{ for some } \alpha \in {}^{\omega}\Sigma^{\omega}\},\\
\mathcal{F}_{\leq} &= \mathcal{F}^r_{\leq} \cup \mathcal{F}^l_{\leq} \cup \mathcal{F}^{bi}_{\leq}.
\end{aligned}$$

The following lemma establishes a connection between the property of a language of being confluent w.r.t. $\leq_f$ and of belonging to the family $\mathcal{F}_{\leq_f}$.

Lemma 3.1. *A language $L \subseteq \Sigma^*$ is in the family $\mathcal{F}_{\leq_f}$ if and only if the following conditions are fulfilled:*
(i) L is infinite,
(ii) $L = \text{DOWN}_{\leq_f}(L)$,
(iii) L is confluent w.r.t. $\leq_f$.

Proof. Suppose first that $L \in \mathcal{F}_{\leq_f}$. This is equivalent to the fact that there exists an infinite word $\alpha \in \Sigma^{\omega} \cup {}^{\omega}\Sigma \cup {}^{\omega}\Sigma^{\omega}$ such that $L = Fact(\alpha)$. The conditions (i) and (ii) are obviously fulfilled.

Let us prove now (iii). For, take $x, y \in L$. Then $x, y \in Fact(\alpha)$ and there is $w \in Fact(\alpha) = L$ such that $x \leq_f w, y \leq_f w$. Consequently, L is confluent w.r.t. $\leq_f$.

For the converse part, suppose that $L \subseteq \Sigma^*$ is a language satisfying (i), (ii), and (iii). We construct an infinite word α following the *algorithm* below:

1. Write L as a totally ordered infinite set (the order being arbitrary)
$$L = \{u_1, u_2, u_3, \ldots\}.$$
2. Define inductively the finite words v_n, for all $n \geq 1$, as follows:
   (a) $v_1 = u_1$,
   (b) For any $n \geq 2$, supposing that $v_1, v_2, \ldots, v_{n-1}$ are defined, take v_n as the word of L having the property that $u_{n-1} \leq_f v_n$ and $v_{n-1} \leq_f v_n$ (v_n exists since L is confluent w.r.t. $\leq_f$ by (iii)).
3. Since $v_n \leq_f v_{n+1}$, it follows that v_{n+1} is obtained from v_n by adding some (possibly empty) words at its ends. Thus, the limit $\alpha = \lim_{n\to\infty} v_n$ is well defined and infinite since L is infinite.

4. Depending on the direction(s) in which the sequence $(v_n)_{n\geq 1}$ extends unboundedly, α can be a left-, right-, or bi-infinite word. (As we will see in a moment, all the three cases are possible.)

Let us prove that $L = Fact(\alpha)$.

If $w \in L$, then $w = u_n$, for some $n \geq 1$ and so $w \leq_f v_n, v_n \in Fact(\alpha)$. Thus, $w \in Fact(\alpha)$.

Conversely, for $w \in Fact(\alpha)$, since we supposed that the sequence $(v_n)_{n\geq 1}$ grows unboundedly to the left (respectively, to the right, in both directions) if α is left- (respectively, right-, bi-) infinite, there must be an $n \geq 1$ such that v_n contains w as a factor. As $v_n \in L$ and $L = \text{DOWN}_{\leq_f}(L)$, w must be in L and our equality is proved. ■

As mentioned above, all the three cases are indeed possible. For instance, as it is easy to prove, the language

$$L_1 = a^* \cup ba^* = Fact(ba^\omega)$$

belongs to the family $\mathcal{F}^r_{\leq_f}$ but is not contained in any of the other two families, the language

$$L_2 = a^* \cup a^*b = Fact({}^\omega ab)$$

in only in the family $\mathcal{F}^l_{\leq_f}$, and

$$L_3 = a^*ba^* \cup a^* = Fact({}^\omega aba^\omega)$$

is only in $\mathcal{F}^{bi}_{\leq_f}$.

Let us further remark that the infinite word α in the construction above might not be unique. (See [HI] for details.)

We intend to prove now that the property of confluence w.r.t. $\leq_f$ is decidable for regular languages. For this, we recall some results from [HI].

Lemma 3.2 ([HI]). *It is decidable whether or not an arbitrary regular language is in the family $\mathcal{F}^r_{\leq_f}$.*

Since the family of regular languages is closed under mirror image, the same result holds for left-infinite words.

Corollary 3.3. *It is decidable whether or not an arbitrary regular language is in the family $\mathcal{F}^l_{\leq_f}$.*

Lemma 3.4 ([HI]). *It is decidable whether or not an arbitrary regular language is in the family $\mathcal{F}^{bi}_{\leq_f}$.*

The following lemma states that, when deciding the confluence property, one may work with the down-set of a language instead of the language itself.

Lemma 3.5. *For a quasi order $\leq$ on Σ^*, any language $L \subseteq \Sigma^*$ is confluent w.r.t. $\leq$ if and only if the down-set of L, $\text{DOWN}_{\leq}(L)$, is confluent w.r.t. $\leq$.*

Proof. Suppose that L is confluent w.r.t $\leq$ and take $x, y \in \text{DOWN}_{\leq}(L)$. It follows that there exist two words $u, v \in L$ such that $x \leq u$ and $y \leq v$. As L is confluent w.r.t $\leq$, there is a word $w \in L$ such that $u \leq w$ and $v \leq w$. Since $\leq$ is transitive, we get $x \leq w, y \leq w$. Now, because $\leq$ is reflexive, $L \subseteq \text{DOWN}_{\leq}(L)$, so $w \in \text{DOWN}_{\leq}(L)$ and therefore $\text{DOWN}_{\leq}(L)$ is confluent w.r.t. $\leq$.

Conversely, suppose that $\text{DOWN}_{\leq}(L)$ is confluent w.r.t $\leq$ and take $x, y \in L$. As $x, y \in \text{DOWN}_{\leq}(L)$, we can find $w \in \text{DOWN}_{\leq}(L)$ such that $x \leq w$ and $y \leq w$. By the definition of $\text{DOWN}_{\leq}(L)$, there must be a word $z \in L$ such that $w \leq z$. But now $x \leq z, y \leq z$, so L is confluent w.r.t. $\leq$ as claimed. ∎

Theorem 3.6. *It is decidable whether or not an arbitrary regular language is confluent w.r.t. $\leq_f$.*

Proof. Obviously, the problem of confluence w.r.t $\leq_f$ is decidable for finite languages. (In fact this is decidable for any quasi order $\leq$ for which it can be decided whether or not two arbitrary words are related by $\leq$.) Now, by Lemma 3.2, Corollary 3.3, and Lemma 3.4, it follows that it is decidable whether or not an arbitrary regular language L belongs to the family $\mathcal{F}_{\leq_f}$. But, by Lemma 3.1, an infinite language L with $L = \text{DOWN}_{\leq_f}(L)$ is in the family $\mathcal{F}_{\leq_f}$ if and only if it is confluent w.r.t. $\leq_f$. Since L and $\text{DOWN}_{\leq_f}(L)$ are simultaneously confluent w.r.t. $\leq_f$ (Lemma 3.5), using the fact that the finiteness problem is decidable for regular languages, our result follows. ∎

Notice that it is possible to prove in a similar way that it is decidable whether or not an arbitrary regular language is confluent w.r.t. $\leq_p$ but we will prove in the next section a more general result, namely that this problem is decidable for context-free languages.

4 The context-free case

We prove in this section that, rather unexpectedly, the confluence problem w.r.t. $\leq_p$ is decidable for context-free languages. The same problem for $\leq_f$ is undecidable.

We state first a result similar to Lemma 3.1 for the prefix order instead of the factor one.

Lemma 4.1. *Any language $L \subseteq \Sigma^*$ is in the family $\mathcal{F}^r_{\leq_p}$ if and only if the following conditions are fulfilled:*

(i) L is infinite,

(ii) $L = \text{DOWN}_{\leq_p}(L)$,

(iii) L is confluent w.r.t. $\leq_p$.

Proof. The proof is similar to the one of Lemma 3.1 with the difference that (using the same notations as in the algorithm there) in this case, for all $n \geq 1$, v_n is a prefix of v_{n+1}, and so the infinite word obtained using the algorithm is always a right-infinite one. Consequently, the family $\mathcal{F}_{\leq_f}$ is replaced by $\mathcal{F}^r_{\leq_p}$. ∎

The following result seems to be well-known but, for the sake of the completeness, we present a very short proof.

Lemma 4.2. *It is decidable whether or not an arbitrary context-free language is in the family $\mathcal{F}^r_{\leq_p}$.*

Proof. Take $L \subseteq \Sigma^*$ a context-free language and denote by c_L the usual constant for L from the Bar-Hillel pumping lemma. It is easy to see that $L \in \mathcal{F}^r_{\leq_p}$ if and only if $L = Pref(uv^\omega)$, for some $u, v \in \Sigma^*, |u| \leq c_L, |v| \leq c_L$. Now $Pref(uv^\omega)$ is a bounded context-free language and the equivalence between two context-free languages one of which is bounded is decidable (see, for instance, [Gi]). ∎

We may state now

Theorem 4.3. *It is decidable whether or not an arbitrary context-free language is confluent w.r.t. $\leq_p$.*

Proof. Since the property in discussion is decidable for finite languages, we may restrict the problem to infinite languages. Using Lemma 3.5, we may further restrict our family of languages to those which are closed under taking prefixes. (Notice that $\text{DOWN}_{\leq_p}(L)$ is effectively context-free for L context-free.) But now, by Lemma 4.1, for an infinite context-free language L with $L = \text{DOWN}_{\leq_p}(L)$, L belongs to the family $\mathcal{F}^r_{\leq_p}$ if and only if L is confluent w.r.t. $\leq_p$. Finally, the former problem is decidable by Lemma 4.2 and, using the fact that the finiteness problem is decidable for context-free languages, our decidability result follows. ∎

Remark. Notice that for any family of languages closed under mirror operation (and the families of regular and context-free languages have this property) all results proved for the prefix order hold true also for the suffix order by left-right duality.

Lemma 4.4. *Let $\mathcal{L}$ be a family of languages effectively closed under union, catenation with symbols, and λ-free catenation closure. If the confluence problem w.r.t. $\leq_f$ is decidable in $\mathcal{L}$, then so is the inclusion problem (and hence the equivalence problem).*

Proof. We observe first that the emptiness problem is decidable in $\mathcal{L}$. Indeed, if $L \in \mathcal{L}$, then

$$L' = \#L\#_1 \cup \#L\#_2 \in \mathcal{L},$$

where $\#, \#_1, \#_2$ are new symbols, and L' is confluent w.r.t. $\leq_f$ if and only if $L = \emptyset$. Indeed, if $L = \emptyset$, then L' is confluent w.r.t. $\leq_f$. Conversely, if $L \neq \emptyset$, then choose $w \in L$ and consider the words $\#w\#_1, \#w\#_2 \in L'$. Since there is no $z \in L'$ having both $\#w\#_1$ and $\#w\#_2$ as factors, a contradiction is obtained. Thus, instead of deciding whether or not $L = \emptyset$, it is enough to decide whether or not L' is confluent w.r.t. $\leq_f$ which is possible, by hypothesis, since L' is effectively in $\mathcal{L}$.

Let us prove that the inclusion problem is decidable for $\mathcal{L}$. Take two arbitrary languages $L_1, L_2 \in \mathcal{L}$. Suppose that $L_1 \subseteq \Sigma^*, L_2 \subseteq \Sigma^*$, and construct the language

$$L_3 = \#L_1\# \cup (\#L_2\#\#)^+$$

where $\# \notin \Sigma$ is a new symbol. By hypothesis, $L_3 \in \mathcal{L}$. We prove first the following claim.

Claim. L_3 is confluent w.r.t. $\leq_f$ if and only if either $L_1 \subseteq L_2$ or else $\mathrm{card}(L_1) = 1$ and $L_2 = \emptyset$.

Proof of Claim. Suppose first that $L_1 \subseteq L_2$. It follows that

$$\textsc{down}_{\leq_f}(L_3) = \textsc{down}_{\leq_f}(\#L_1\#) \cup \textsc{down}_{\leq_f}((\#L_2\#\#)^+) = \textsc{down}_{\leq_f}((\#L_2\#\#)^+)$$

since

$$\textsc{down}_{\leq_f}(\#L_1\#) \subseteq \textsc{down}_{\leq_f}(\#L_2\#) \subseteq \textsc{down}_{\leq_f}((\#L_2\#\#)^+).$$

Because $(\#L_2\#\#)^+$ is confluent w.r.t. $\leq_f$ (see the observation in example 3, section 2), it follows by Lemma 3.5 that $\textsc{down}_{\leq_f}((\#L_2\#\#)^+)$ is also confluent w.r.t. $\leq_f$. Now, $\textsc{down}_{\leq_f}(L_3)$ is confluent w.r.t. $\leq_f$ and, using again Lemma 3.5, we get that L_3 is confluent w.r.t. $\leq_f$.

Suppose now that $\mathrm{card}(L_1) = 1$ and $L_2 = \emptyset$. If we put $L_1 = \{w\}$, then

$$L_3 = \{\#w\#\}$$

and it is obvious that L_3 is confluent w.r.t. $\leq_f$ since any singleton is. One implication is proved.

For the other one, suppose that L_3 is confluent w.r.t. $\leq_f$ and $L_1 \not\subseteq L_2$. Take $u \in L_1 - L_2$. Suppose now that $L_2 \neq \emptyset$ and take $v \in L_2$. Since L_3 in confluent w.r.t. $\leq_f$, there is a $w \in L_3$ such that $\#u\# \leq_f w$ and $\#v\#\# \leq_f w$. Because of the three symbols $\#$ in the second word, we must have $w \in (\#L_2\#\#)^+$ and put

$$w = \#w_1\#\#\#w_2\#\#\ldots\#w_n\#\#, n \geq 1, w_i \in L_2, 1 \leq i \leq n.$$

From $\#u\# \leq_f w$, we get that $u = w_i$, for some $1 \leq i \leq n$. Thus $u \in L_2$, a contradiction. Consequently, $L_2 = \emptyset$ and $L_3 = \#L_1\#$.

Suppose that $\mathrm{card}(L_1) \geq 2$ and take $u, v \in L_1, u \neq v$. Using again the confluence of L_3 w.r.t. $\leq_f$, we find $w \in L_3$ with $\#u\# \leq_f w$ and $\#v\# \leq_f w$. If $w = \#x\#$, then $u = x = v$, a contradiction. Consequently, $\mathrm{card}(L_1) = 1$ ($L_1 \neq \emptyset$ because $L_1 \not\subseteq L_2$) and our claim is proved.

We give now an *algorithm* for deciding whether or not $L_1 \subseteq L_2$.

1 Decide whether or not $L_2 = \emptyset$ (the emptiness problem is decidable for $\mathcal{L}$).

  1.1 If *yes*, then go to step 2.

  1.2 If *no*, then go to step 3.

2 Decide whether or not $L_1 = \emptyset$ (again, possible).

  2.1 If *yes*, the output **yes** (obviously correct).

  2.2 If *no*, then output **no** (again correct).

3 Decide whether or not L_3 is confluent w.r.t. $\leq_f$ (the confluence problem w.r.t. $\leq_f$ is decidable for $\mathcal{L}$ by hypothesis and L_3 is effectively in $\mathcal{L}$).

  3.1 If *yes*, then output **yes** (correct by our Claim).

  3.2 If *no*, then output **no** (again, correct by our Claim).

The proof is concluded. ■

From Lemma 4.4, we get immediately the following result.

Theorem 4.5. *It is undecidable whether or not an arbitrary context-free language is confluent w.r.t.* $\leq_f$.

As proved in [MP], it is undecidable whether or not an arbitrary context-free language is the set of finite factors of some right-infinite word. Using similar techniques, we prove this undecidability result for any quasi order containing $\leq_f$.

Theorem 4.6. *If $\leq$ is a quasi order on Σ^* which is an extension of $\leq_f$, then it is undecidable whether or not an arbitrary context-free language is in any of the families $\mathcal{F}^r_{\leq}, \mathcal{F}^l_{\leq}$, or $\mathcal{F}^{bi}_{\leq}$.*

Proof. Consider a two-letter alphabet $\{a, b\}$ and

$$\{(x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \mid x_i, y_i \in \{a, b\}^+, 1 \leq i \leq n\}$$

an arbitrary instance of the Post Correspondence Problem. Construct the following languages:

$$\begin{aligned}
&L(x) = \{ba^{i_k} \ldots ba^{i_1} c x_{i_1} \ldots x_{i_k} \mid k \geq 1, 1 \leq i_j \leq n, 1 \leq j \leq k\},\\
&L(y) = \{ba^{i_k} \ldots ba^{i_1} c y_{i_1} \ldots y_{i_k} \mid k \geq 1, 1 \leq i_j \leq n, 1 \leq j \leq k\},\\
&L_{x,y} = L(x) c\, mi(L(y)),\\
&L_{mi} = \{w_1 c w_2 c\, mi(w_2) c\, mi(w_1) \mid w_1, w_2 \in \{a, b\}^*\},\\
&L_{PCP} = L_{x,y} \cap L_{mi},
\end{aligned}$$

where $mi(w)$ denotes the mirror image of w and $c \notin \{a, b\}$ is a new letter. It is well-known (see, for instance, [Gi]) that the complement $\overline{L_{PCP}}$ of L_{PCP} is context-free.

We show that $\overline{L_{PCP}}$ belongs to any of the three families above if and only if the above instance of the Post Correspondence Problem has no solution.

Notice first that, since $\leq_f \subseteq \leq$, if $x \in \textsc{down}_{\leq_f}(\overline{L_{PCP}})$, then $x \leq_f y$, for some $y \in \overline{L_{PCP}}$, and so $x \leq y$. Thus

$$\overline{L_{PCP}} \subseteq \textsc{down}_{\leq_f}(\overline{L_{PCP}}) \subseteq \textsc{down}_{\leq}(\overline{L_{PCP}}).$$

Suppose now that $\overline{L_{PCP}} \in \mathcal{F}_{\leq}$. Then, as $\textsc{down}_{\leq}$ is idempotent, $\overline{L_{PCP}} = \textsc{down}_{\leq}(\overline{L_{PCP}})$ and, using the observation above,

$$\overline{L_{PCP}} = \textsc{down}_{\leq_f}(\overline{L_{PCP}}) = \{a, b, c\}^*$$

since we may suppose that any solution of the PCP is a factor of a non-solution. (Indeed, the instances of the PCP which violate this condition are exactly those having all words as solutions, so they are trivial and can be eliminated.) Thus, $L_{PCP} = \emptyset$.

Conversely, if $L_{PCP} = \emptyset$, then $\overline{L_{PCP}} = \{a, b, c\}^*$. Chose now infinite words $\alpha_1 \in \Sigma^\omega, \alpha_2 \in {}^\omega\Sigma, \alpha_3 \in {}^\omega\Sigma^\omega$ such that $Fact(\alpha_i) = \{a, b, c\}^*$, for any $1 \leq i \leq 3$. Now, as $\leq$ is reflexive, it follows that $\overline{L_{PCP}} = \textsc{down}_{\leq}(Fact(\alpha_3))$, so $\overline{L_{PCP}} \in \mathcal{F}^{bi}_{\leq}$. Using again the fact that $\leq_f \subseteq \leq$, we get

$$\{a, b, c\}^* = Fact(\alpha_1) = \textsc{down}_{\leq_f}(Pref(\alpha_1)) \subseteq \textsc{down}_{\leq}(Pref(\alpha_1))$$

so $\overline{L_{PCP}} = \textsc{down}_{\leq}(Pref(\alpha_1))$ and $\overline{L_{PCP}} \in \mathcal{F}^r_{\leq}$.

Analogously, $\overline{L_{PCP}} = \text{DOWN}_{\leq}(Suf(\alpha_2))$, thus $\overline{L_{PCP}} \in \mathcal{F}^l_{\leq}$. The proof is completed. ■

Using the tools in the proof of Theorem 4.6, we prove another undecidability result about context-free languages. It concerns the equality between a context-free language and its down-set w.r.t. any quasi order which is an extension of the prefix partial order $\leq_p$.

Corollary 4.7. *For any quasi order $\leq$ on Σ^* such that $\leq_p \subseteq \leq$, it is undecidable, for an arbitrary context-free language L, whether or not $L = \text{DOWN}_{\leq}(L)$.*

Proof. Using the notations in Theorem 4.6, we have that $\overline{L_{PCP}}$ is context-free and we claim that $\overline{L_{PCP}} = \text{DOWN}_{\leq}(\overline{L_{PCP}})$ if and only if the above instance of the Post Correspondence Problem has no solution.

Indeed, if $L_{PCP} = \emptyset$, then $\overline{L_{PCP}} = \{a, b, c\}^*$ and so $\overline{L_{PCP}} = \text{DOWN}_{\leq}(\overline{L_{PCP}})$. Conversely, suppose that $\overline{L_{PCP}} = \text{DOWN}_{\leq}(\overline{L_{PCP}})$. From $\leq_p \subseteq \leq$, we get

$$\overline{L_{PCP}} \subseteq \text{DOWN}_{\leq_p}(\overline{L_{PCP}}) \subseteq \text{DOWN}_{\leq}(\overline{L_{PCP}}),$$

so $\overline{L_{PCP}} = \text{DOWN}_{\leq_p}(\overline{L_{PCP}})$. Since we may assume, as in the proof of Theorem 4.6, that any solution is a prefix of a non-solution, we have $\text{DOWN}_{\leq_p}(\overline{L_{PCP}}) = \{a, b, c\}^*$, so $L_{PCP} = \emptyset$. The proof is concluded. ■

Remark. Notice that the undecidability of whether or not $L = \text{DOWN}_{\leq_p}(L)$ for an arbitrary context-free language L does not invalidate the proof of Theorem 4.3 since there we do not decide whether or not $L = \text{DOWN}_{\leq_p}(L)$, but replace L by $\text{DOWN}_{\leq_p}(L)$, which can be effectively done for L context-free.

We give now two general undecidability results concerning the confluence problem. The first concerns the factor partial order $\leq_f$.

Theorem 4.8. *Let $\leq$ be a quasi order on Σ^* such that $\leq \subseteq \leq_f$ and $\mathcal{L}$ a family of languages which is effectively closed under union and catenation with symbols such that the emptiness problem is undecidable for languages in $\mathcal{L}$. Then the confluence problem w.r.t. $\leq$ is undecidable for languages in $\mathcal{L}$.*

Proof. Consider a language $L \subseteq \Sigma^*, L \in \mathcal{L}$, and three new symbols $\#, \#_1, \#_2 \notin \Sigma$. Construct the language

$$L_1 = \#L\#_1 \cup \#L\#_2.$$

We have that L_1 is effectively in $\mathcal{L}$ and claim that L_1 is confluent w.r.t. $\leq$ if and only if $L = \emptyset$.

If $L = \emptyset$, then $L_1 = \emptyset$ and so L_1 is confluent w.r.t. $\leq$. Conversely, suppose that L_1 is confluent w.r.t. $\leq$. If L is not empty, then take $w \in L$. Now $\#w\#_1, \#w\#_2 \in L_1$ and, since L_1 is confluent w.r.t. $\leq$, there must be a $v \in L_1$ such that $\#w\#_1 \leq v$ and $\#w\#_2 \leq v$. Without loss of generality, we may suppose that $v \in \#L\#_1$ and let us put $v = \#u\#_1, u \in L$. From $\#w\#_1 \leq \#u\#_1$ we get $u = w$, since $\leq$ is contained in $\leq_f$. But now, $\#w\#_1 \leq v$ means $\#w\#_1 \leq \#w\#_2$, which is impossible. Thus $L = \emptyset$.

Consequently, we have reduced the decidability of the confluence problem w.r.t. $\leq$ to the decidability of the emptiness problem which is, by hypothesis, undecidable for $\mathfrak{L}$. The proof is concluded. ∎

From the proof of Theorem 4.8 we obtain directly the following corollary.

Corollary 4.9. *For any quasi order $\leq$ on Σ^* such that $\leq \subseteq \leq_f$, it is undecidable whether or not the intersection of two arbitrary context-free languages is confluent w.r.t. $\leq$.*

Our second general undecidability result concerns the subword relation $\leq_s$.

Theorem 4.10. *Let $\leq$ be a quasi order on Σ^* such that $\leq \subseteq \leq_s$ and $\mathfrak{L}$ a family of languages which is effectively closed under union with singletons and catenation with symbols such that the emptiness problem is undecidable for languages in $\mathfrak{L}$. Then the confluence problem w.r.t. $\leq$ is undecidable for languages in $\mathfrak{L}$.*

Proof. Consider a language $L \subseteq \Sigma^*, L \in \mathfrak{L}$, and two new symbols $\#_1, \#_2 \notin \Sigma$. Construct the language

$$L_1 = L\#_1 \cup \{\#_2\}.$$

We have that L_1 is effectively in $\mathfrak{L}$ and we claim that L_1 is confluent w.r.t. $\leq$ if and only if $L = \emptyset$. One implication is clear. For the converse one, suppose that $L \neq \emptyset$ and take $w \in L$. Since L_1 is confluent w.r.t.$\leq$, it follows that there must be a word $v \in L_1$ such that $w\#_1 \leq v$ and $\#_2 \leq v$. If $v \in L\#_1$, then it is not possible that $\#_2 \leq v$. If $v = \#_2$, then $w\#_1 \leq v$ does not hold. In any case a contradiction is obtained. Because $\mathfrak{L}$ has an undecidable emptiness problem, our theorem is proved. ∎

Remark. Notice that the statement of Theorem 4.10 holds for an extention of the partial order $\leq_s$, namely for the so-called Parikh quasi order, defined as follows: $u \leq_{\mathcal{P}} v$ iff all letters of u appear, in some order, in v.

It follows from above that, in particular, the confluence problem w.r.t. any quasi order $\leq \subseteq \leq_{\mathcal{P}}$ is undecidable for context-sensitive languages.

5 Terminating relations

For a relation $\leq$ on Σ^*, its inverse is denoted $\geq$ and defined by $x \geq y$ iff $y \leq x$. If $\leq$ is a partial order, then so is $\geq$. A quasi order $\leq$ on Σ^* is called *well founded* if there is no infinite descending chain w.r.t. $\leq$, that is, for all infinite sequences of words $w_1 \geq w_2 \geq w_3 \geq \ldots$ there exists $n \geq 1$ such that $w_n \leq w_{n+1}$. A quasi order $\leq$ is *terminating* if its inverse $\geq$ is well founded.

In this section we investigate the confluence of terminating quasi orders.

Notice that, in what follows, the inverse relations of $\leq_p, \leq_f$, and $\leq_s$ are denoted by $\geq_p, \geq_f$, and $\geq_s$, respectively.

Lemma 5.1. *If $\leq$ is a terminating quasi order on Σ^*, then any language $L \subseteq \Sigma^*$ is confluent w.r.t. $\leq$ if and only if there exists a unique word $x_L \in L$ such that, for all $y \in L$, $y \leq x_L$.*

Proof. Suppose that $\leq$ is a terminating quasi order on Σ^* and $L \subseteq \Sigma^*$ is a language confluent w.r.t. $\leq$. We may suppose that L is non-empty since any empty language satisfies our property. Thus, there exists a word, say x_1, in L. If there is no $y \in L, y \neq x_1$, such that $x_1 \leq y$, then we choose $x_L = x_1$. Then, for any $y \in L$, since L is confluent, there is $w \in L$ such that $x_L \leq w, y \leq w$. But, we must have $x_L = w$ by the choice of x_L. So $y \leq x_L$ and we are done. If there is a $x_2 \in L - \{x_1\}$ such that $x_1 \leq x_2$ then we use the same reasoning with x_2 instead of x_1. Since $\leq$ is terminating, after a finite number $n \geq 1$ of steps we find $x_n \in L$ such that for no $y \in L, y \neq x_n$, $x_n \leq y$. It follows as above that $x_L = x_n$ is the word looked for. If there is another one, say x'_L, we have $x_L \leq x'_L$ and $x'_L \leq x_L$. As $\leq$ is antisymmetric, $x_L = x'_L$ so x_L is unique.

The converse implication is trivial. ∎

Lemma 5.2. *For any partial order $\rho \in \{\geq_p, \geq_f, \geq_s\}$ and any language $L \subseteq \Sigma^*$, L is confluent w.r.t. ρ if and only if there is a unique word $x_L \in L$ such that $|x_L| = \min_{x \in L} |x|$ and $y \rho x_L$ for all $y \in L$.*

Proof. Follows directly from Lemma 5.1 since all the three partial orders in discussion are terminating and $x \rho y$ implies $|x| \leq |y|$ with $|x| = |y|$ if and only if $x = y$. ∎

The problem of confluence w.r.t. a terminating partial order seems to be easier than the problem for a non-terminating one. Indeed, we show that for the inverses of all the three relations considered above, the confluence problem is decidable for a very large family of languages.

Theorem 5.3. *For any partial order $\rho \in \{\geq_p, \geq_f, \geq_s\}$ and any family $\mathfrak{L}$ of languages such that*

(i) $\mathfrak{L}$ is closed under gsm mappings,

(ii) for any language in $\mathfrak{L}$, the set of minimal length words is effectively computable,

it is decidable whether or not an arbitrary language in $\mathfrak{L}$ is confluent w.r.t. ρ.

Proof. Suppose that we have a a family of languages $\mathfrak{L}$ with the properties (i) and (ii) above and $\rho \in \{\geq_p, \geq_f, \geq_s\}$. For an arbitrary language $L \in \mathfrak{L}$, $L \subseteq \Sigma^*$, we use the following *algorithm* to decide whether or not L is confluent w.r.t. ρ.

1. Compute the set of minimal length words of L, say $L_{\min}$ (this is possible by (ii)).

2. If $\mathrm{card}(L_{\min}) \geq 2$, then answer **no** (the answer is correct by Lemma 5.2).

3. If $\mathrm{card}(L_{\min}) = 0$, then answer **yes**. (It follows that $L_{\min} = \emptyset$, so $L = \emptyset$ and the answer is correct.)

4. Now $\mathrm{card}(L_{\min}) = 1$ and put $L_{\min} = \{x_L\}$. By Lemma 5.2, we have to check only whether or not $y \rho x_L$ for all $y \in L$.

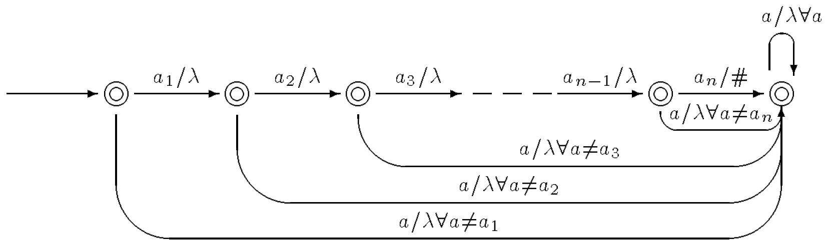
5. If $x_L = \lambda$, then answer **yes**.

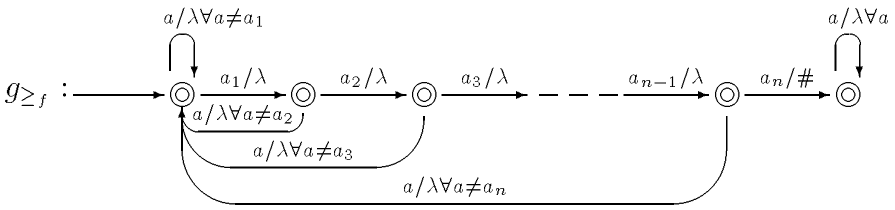
6. We may suppose now that all words in L are non-empty. Consider a new symbol $\#$ and construct a (deterministic) gsm g_ρ such that, for any $y \in L$,

$$g_\rho(y) = \begin{cases} \#, & \text{if } y \rho x_L, \\ \lambda, & \text{otherwise.} \end{cases}$$

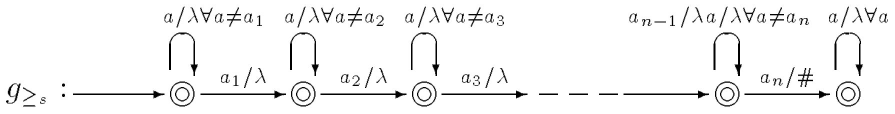
Suppose that $x_L = a_1 a_2 \dots a_n$, $a_i \in \Sigma$, $1 \leq i \leq n$. Then our gsm's are:

$g_{\geq_p}$:





Notice that all states are final for both $g_{\geq_p}$ and $g_{\geq_f}$ so each of them maps any word into a unique one. Moreover, $g_{\geq_p}$ ($g_{\geq_f}$) outputs at most one symbol which is $\#$ and this only in the case when the input contains $a_1a_2\dots a_n$ as a prefix (factor, respectively).



Notice that if $a_1a_2\dots a_n$ appears as a subword of $y \in L$, then we may suppose that a_1 appears as the first occurrence of a_1 in y, a_2 appears as the first occurrence of a_2 after one of a_1, and so on and so forth. The converse is obvious and it follows that $g_{\geq_s}$ works as required.

7. Now, for any $L \in \mathfrak{L}$, $g_\rho(L) \in \{\{\lambda\}, \{\lambda, \#\}, \{\#\}\}$ and L is confluent w.r.t. ρ if and only if $g_\rho(L) = \{\#\}$.

8. Compute $g_\rho(L)_{\min}$. (This is possible by (ii), since $g_\rho(L) \in \mathfrak{L}$.)

9. If $g_\rho(L)_{\min} = \{\#\}$, then answer **yes**, otherwise answer **no**. (Notice that in both cases the answer is correct.)

Our proof is now complete. ■

Corollary 5.4. *For any partial order $\rho \in \{\geq_p, \geq_f, \geq_s\}$ and any full trio $\mathfrak{L}$ such that $\mathfrak{L}$ contains only recursive languages and the emptiness problem is decidable for $\mathfrak{L}$, it is decidable whether or not an arbitrary language in $\mathfrak{L}$ is confluent w.r.t. ρ.*

Proof. Both conditions in Theorem 5.3 are fulfilled because: (i) any full trio is closed under gsm mappings and (ii) the set of minimal length words $L_{\min}$ can be effectively computed for any language $L \in \mathfrak{L}$ by deciding first whether or not $L = \emptyset$ and in the case $L = \emptyset$ set $L_{\min} = \emptyset$ while for $L \neq \emptyset$ compute effectively $L_{\min}$ (possible because L is recursive). The result is proved. ■

As a direct consequence of Corollary 5.4 we get

Corollary 5.5. *For any partial order $\rho \in \{\geq_p, \geq_f, \geq_s\}$, it is decidable whether or not an arbitrary context-free language is confluent w.r.t. ρ.*

6 The regularity of down-sets

Haines [Ha] proved the following theorem concerning the down-sets w.r.t. $\leq_s$ and $\geq_s$.

Theorem 6.1 (Haines, [Ha]). *For any language L, both down-sets $\mathrm{DOWN}_{\leq_s}(L)$ and $\mathrm{DOWN}_{\geq_s}(L)$ are regular.*

In what follows, we give a generalization of this result.

We neeed some definitions. A quasi order $\leq$ on Σ^* is *monotone* if for any $x, y, u, v \in \Sigma^*$, $x \leq y$ implies $uxv \leq uyv$; $\leq$ is *well quasi order* if it is well founded and each set of pairwise incomparable elements in Σ^* is finite. L is *$\leq$-closed* if $x \in L$ and $x \leq y$ imply $y \in L$.

We will use the following generalization of Myhill-Nerode theorem from [EHR].

Theorem 6.2 (Ehrenfeucht, Haussler, Rozenberg, [EHR]). *Any language $L \subseteq \Sigma^*$ is regular if and only if L is $\leq$-closed under some monotone well quasi order $\leq$ on Σ^*.*

We can prove now the theorem concerning the regularity of down-sets.

Theorem 6.3. *For any monotone well quasi order $\leq$ on Σ^* and any language $L \subseteq \Sigma^*$, both sets $\mathrm{DOWN}_{\leq}(L)$ and $\mathrm{DOWN}_{\geq}(L)$ are regular.*

Proof. Since $\mathrm{DOWN}_{\geq}(L)$ is $\leq$-closed, the regularity of the set $\mathrm{DOWN}_{\geq}(L)$ follows from Theorem 6.2.

In order to show that $\mathrm{DOWN}_{\leq}(L)$ is also regular, it is enough to prove that its complement $\overline{\mathrm{DOWN}_{\leq}(L)}$ is regular. We claim that

$$\overline{\mathrm{DOWN}_{\leq}(L)} = \mathrm{DOWN}_{\geq}(\overline{\mathrm{DOWN}_{\leq}(L)}).$$

If this equality were proved, then, in virtue of the first part of our theorem, $\overline{\mathrm{DOWN}_{\leq}(L)}$ would be proved to be regular.

The inclusion of the left-hand member into the right-hand one is by the definition of $\mathrm{DOWN}_{\geq}$ and the reflexivity of $\geq$. For the converse inclusion, suppose that there is a word

$$w \in \mathrm{DOWN}_{\geq}(\overline{\mathrm{DOWN}_{\leq}(L)}) - \overline{\mathrm{DOWN}_{\leq}(L)} = \mathrm{DOWN}_{\geq}(\overline{\mathrm{DOWN}_{\leq}(L)}) \cap \mathrm{DOWN}_{\leq}(L).$$

It follows that there must be two words $u \in \overline{\mathrm{DOWN}_{\leq}(L)}$ and $v \in L$ such that $w \geq u$ and $w \leq v$. Then $u \leq w \leq v$ and so, $\leq$ being transitive, $u \leq v$. But now $u \in \mathrm{DOWN}_{\leq}(L) \cap \overline{\mathrm{DOWN}_{\leq}(L)} = \emptyset$, which is absurd. Our proof is complete. ∎

Notice that the result of Haines concerning the regularity of the sets of subwords and superwords follows from Theorem 6.3 and the following restricted form of Higman's theorem. For a simple proof of Theorem 6.4, see [Lo].

Theorem 6.4 (Higman, [Hi]). *If L is a language such that any two words in L are incomparable w.r.t. the subword partial order $\leq_s$, then L is finite.*

Haines raised also the following problem: when is it possible to find effectively the regular sets $\mathrm{DOWN}_{\leq_s}(L)$ and $\mathrm{DOWN}_{\geq_s}(L)$? It is easy to see that if it is possible to find effectively the two sets for all languages in a family $\mathcal{L}$, then $\mathcal{L}$ has a decidable emptiness problem. Van Leeuwen [vL] proved that for the superword partial order $\geq_s$ the converse statement practically holds.

Theorem 6.5 (van Leeuwen, [vL]). *For any family $\mathcal{L}$ effectively closed under intersection with regular sets, one can effectively determine $\mathrm{DOWN}_{\geq_s}(L)$ for each $L \in \mathcal{L}$ if and only if $\mathcal{L}$ has a decidable emptiness problem.*

For the subword partial order $\leq_s$, he gave an algorithm to find for any context-free language L the set $\mathrm{DOWN}_{\leq_s}(L)$; actually, he proved a more general result, namely that it is possible to find effectively the sets $\mathrm{DOWN}_{\leq_s}(L)$ and $\mathrm{DOWN}_{\geq_s}(L)$ for all languages in a family $\mathcal{L}$ if and only if this is possible for the algebraic closure of $\mathcal{L}$.

Theorem 6.6 (van Leeuwen, [vL]). *For any context-free language L, the down-set $\mathrm{DOWN}_{\leq_s}(L)$ is effectively regular.*

It remained open whether or not there are families of languages for which the emptiness problem is decidable but the regularity of the sets of subwords is not effective. In what follows, we give a positive answer to this problem and construct a wide range of families having these properties.

The following theorem is very important for our purpose.

Theorem 6.7. *It is decidable whether or not an arbitrary regular language is confluent w.r.t.* $\leq_s$.

Proof. Consider a regular language R and construct the language

$$L = \{u\#mi(v) \mid u, v \in R \text{ such that there is } w \in R \text{ with } u \leq_s w, v \leq_s w\}.$$

It is easy to see that R is confluent w.r.t. $\leq_s$ if and only if $L = R\#mi(R)$.

Claim 1. L is context-free.

Proof of Claim 1. A push-down automaton $\mathcal{A}$ for L works as follows.

– any input without the marker $\#$ or with at least two markers is rejected.

– for an input of the form $u\#v$, $\mathcal{A}$ does:

– when reading u, it pushes nondeterministically into the stack a word $w \in R$ such that $u \leq_s w$;

– after the marker $\#$, $\mathcal{A}$ checks whether or not $v \leq_s mi(w)$ (which is equivalent to $mi(v) \leq_s w$).

A formal construction for $\mathcal{A}$ follows. Suppose that $\mathcal{B} = (\Sigma, Q, \delta_{\mathcal{B}}, q_0, F)$ is a complete finite automaton recognizing R. Then we construct

$$\mathcal{A} = (\Sigma, Q \times Q \cup Q, \Gamma = \Sigma \cup \{Z\}, \delta_{\mathcal{A}}, (q_0, q_0), Z, \emptyset)$$

with the transition mapping $\delta_{\mathcal{A}}$ given by

$$\begin{aligned}
&\delta_{\mathcal{A}}((q_1, q_2), a, Y) = \{(\delta_{\mathcal{B}}(q_1, a), \delta_{\mathcal{B}}(q_2, a)), aY\}, q_1, q_2 \in Q, a \in \Sigma, Y \in \Gamma,\\
&\delta_{\mathcal{A}}((q_1, q_2), \lambda, Y) = \{((q_1, \delta_{\mathcal{B}}(q_2, a)), aY) \mid a \in \Sigma\}, q_1, q_2 \in Q, Y \in \Gamma,\\
&\delta_{\mathcal{A}}((q_1^F, q_2^F), \#, Y) = \{(q_0, Y)\}, q_1^F, q_2^F \in F, Y \in \Gamma,\\
&\delta_{\mathcal{A}}(q_1, a, a) = \{(\delta_{\mathcal{B}}(q_1, a), \lambda)\}, q_1 \in Q, a \in \Sigma,\\
&\delta_{\mathcal{A}}(q_1, \lambda, a) = \{(q_1, \lambda)\}, q_1 \in Q, a \in \Sigma,\\
&\delta_{\mathcal{A}}(q_1^F, \lambda, Z) = \{(q_1^F, \lambda)\}, q_1^F \in F.
\end{aligned}$$

Clearly, for any $w \in \Sigma^*$, we have

$$((q_0, q_0), w, Z) \overset{*}{\vdash} (q_1, \lambda, \lambda) \text{ iff } w \in L.$$

The claim is proved.

Claim 2. $L = R\#mi(R)$ if and only if $\text{DOWN}_{\leq_s}(L) = \text{DOWN}_{\leq_s}(R\#mi(R))$.

Proof of Claim 2. In one direction the implication is obvious. Suppose that $\text{DOWN}_{\leq_s}(L) = \text{DOWN}_{\leq_s}(R\#mi(R))$. We have anyway $L \subseteq R\#mi(R)$. Take $u\#mi(v) \in R\#mi(R)$. We have that $R\#mi(R) \subseteq \text{DOWN}_{\leq_s}(R\#mi(R)) = \text{DOWN}_{\leq_s}(L)$, so $u\#mi(v) \in \text{DOWN}_{\leq_s}(L)$. Thus, for some $w \in L, u\#mi(v) \leq_s w$. Let us put $w = x\#mi(y)$. It follows that $u \leq_s x$ and $v \leq_s y$. But now, since $x\#mi(y) \in L$, there must be a word $z \in R$ with $x \leq_s z, y \leq_s z$, by the definition of L. We get $u \leq_s z, v \leq_s z$ and so $u\#mi(v) \in L$, which proves the claim.

Claim 3. It is decidable whether or not $L = R\#mi(R)$.

Proof of Claim 3. Since the languages $R\#mi(R)$ and L are context-free, we obtain by Theorem 6.6 that the down-sets $\mathrm{DOWN}_{\leq_s}(R\#mi(R))$ and $\mathrm{DOWN}_{\leq_s}(L)$ are effectively regular. As the equivalence problem is decidable for regular languages, using Claim 2, we are done.

Consequently, it is decidable whether or not R is confluent w.r.t. $\leq_s$ and the proof is concluded. ∎

The result in Theorem 6.7 can be extended to the context-free case.

Corollary 6.8. *It is decidable whether or not an arbitrary context-free language is confluent w.r.t.* $\leq_s$.

Proof. Take an arbitrary context-free language L. By Lemma 3.5, L is confluent w.r.t. $\leq_s$ if and only if $\mathrm{DOWN}_{\leq_s}(L)$ is confluent w.r.t. $\leq_s$. By Theorem 6.6, the down-set $\mathrm{DOWN}_{\leq_s}(L)$ is effectively regular and thus its confluence w.r.t. $\leq_s$ can be decided by Theorem 6.7. ∎

Remark. Since, by Theorem 4.3 and Corollary 6.8, the confluence problem w.r.t. any of the partial orders $\leq_p$ and $\leq_s$ is decidable for context-free languages, it follows that the result in Lemma 4.4 cannot be extended for any partial order $\leq$ such that either $\leq \subseteq \leq_f$ or $\leq_f \subseteq \leq$. In this sense, Lemma 4.4 characterizes the factor partial order $\leq_f$.

We can prove now our result concerning the effective regularity of the sets of subwords. We denote, for a language L, by $\mathrm{alph}(L)$ the set of all letters which appear in words of L.

Theorem 6.9. *Le $\mathcal{L}$ be a family of languages such that the emptiness problem is undecidable for languages in $\mathcal{L}$. Construct the family $\mathcal{L}'$ of languages by*

$$\mathcal{L}' = \{L\Sigma_L^* \cup \Delta_L^* \mid L \in \mathcal{L}\}$$

where

$$(\Sigma_L, \Delta_L) = \begin{cases} (\{a_1, \ldots, a_{p-1}\}, \{a_1, \ldots, a_{p-1}, a_{p+1}\}), & \textit{if } \mathrm{alph}(L) = \{a_1, \ldots, a_p\}, p \geq 2, \\ (\{a_2\}, \{a_3\}), & \textit{if } \mathrm{alph}(L) \subseteq \{a_1\}, \end{cases}$$

and two letters with different subscripts are considered to be different. Then the emptiness problem is decidable for languages in $\mathcal{L}'$ but there is no algorithm to compute the sets of subwords of the languages in $\mathcal{L}'$.

Proof. The emptiness problem is trivially decidable for languages in $\mathcal{L}'$. For the second part of the statement, we prove first the next claim.

Claim. For any $L \in \mathcal{L}$, the language $L\Sigma_L^* \cup \Delta_L^*$ is confluent w.r.t. $\leq_s$ if and only if L is empty.

Proof of Claim. If L is empty, then $L\Sigma_L^* \cup \Delta_L^* = \Delta_L^*$ which is confluent w.r.t. $\leq_s$.

Conversely, suppose that $L\Sigma_L^* \cup \Delta_L^*$ is confluent w.r.t. $\leq_s$ but L is not empty. If $\mathrm{alph}(L) = \{a_1, a_2, \ldots, a_p\}, p \geq 2$, take $w \in L$ such that $a_p \in \mathrm{alph}(w)$. Consider the words $w, a_{p+1} \in L\Sigma_L^* \cup \Delta_L^*$. Since $L\Sigma_L^* \cup \Delta_L^*$ is confluent w.r.t. $\leq_s$, there must be an $x \in L\Sigma_L^* \cup \Delta_L^*$ such that $w \leq_s x$ and $a_{p+1} \leq_s x$. Because $a_{p+1} \notin \mathrm{alph}(L\Sigma_L^*)$, we must have $x \in \Delta_L^*$. But now $a_p \notin \mathrm{alph}(x)$ thus $w \not\leq_s x$, a contradiction.

In the case when $\mathrm{alph}(L) \subseteq \{a_1\}$, as L is not empty, we have two cases: either $L = \{\lambda\}$ or there is a word, say w, in L such that $a_1 \in \mathrm{alph}(w)$. In the former case, we get that $L\Sigma_L^* \cup \Delta_L^* = a_2^* \cup a_3^*$ is confluent w.r.t. $\leq_s$, which is false. For the later one, consider $w, a_3 \in La_2^* \cup a_3^*$. As above, there is no word $x \in La_2^* \cup a_3^*$ with $w \leq_s x, a_3 \leq_s x$, again a contradiction. The claim is proved.

Suppose that there is an algorithm to compute the sets of subwords for languages in $\mathcal{L}'$. Now, for any $L \in \mathcal{L}$, by the Claim above, L is empty if and only if $L\Sigma_L^* \cup \Delta_L^*$ is confluent w.r.t. $\leq_s$, which is, in turn, equivalent, by Lemma 3.5, to the fact that the set $\mathrm{DOWN}_{\leq_s}(L\Sigma_L^* \cup \Delta_L^*)$ is confluent w.r.t. $\leq_s$, which is decidable by Theorem 5.3 since $\mathrm{DOWN}_{\leq_s}(L\Sigma_L^* \cup \Delta_L^*)$ is effectively regular by our assumption. Thus, the emptiness problem is decidable for languages in $\mathcal{L}$, a contradiction. The theorem is proved. ■

7 Some closure properties for down-operators

We investigate in this section the closure under the down-operator $\mathrm{DOWN}_{\leq}$.

A general result concerning the closure under down-operators is the following.

Theorem 7.1. *Let $\leq \subseteq \Sigma^* \times \Sigma^*$ be a quasi order for which there is a gsm $g_{\leq}$ such that for any $w \in \Sigma^*, \mathrm{DOWN}_{\leq}(\{w\}) = g_{\leq}(w)$. Then any full trio is closed under the down-operator $\mathrm{DOWN}_{\leq}$.*

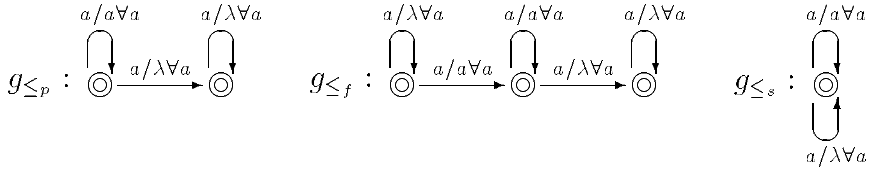
Proof. Consider a full trio $\mathcal{L}$ and a language $L \in \mathcal{L}$. We have then

$$\mathrm{DOWN}_{\leq}(L) = \bigcup_{w \in L} \mathrm{DOWN}_{\leq}(\{w\}) = \bigcup_{w \in L} g_{\leq}(w) = g_{\leq}(L).$$

Since any full trio is closed under gsm's, it follows that the family $\mathcal{L}$ is closed under the down-operator $\text{DOWN}_{\leq}$. ∎

Corollary 7.2. *Any full trio is closed under the down-operators* $\text{DOWN}_{\leq_p}$, $\text{DOWN}_{\leq_f}$, *and* $\text{DOWN}_{\leq_s}$.

Proof. We need, according to Theorem 7.1, only to show that, for any of the down-operators in the statement, there is a gsm as required. Below are the three constructions



It is easy to see from the pictures that the three gsm's work as required. ∎

Using again Theorem 6.3, we obtain a closure result which covers also the part concerning $\leq_s$ in Corollary 7.2.

Theorem 7.3. *Any full trio is closed under the down-operators* $\text{DOWN}_{\leq}$ *and* $\text{DOWN}_{\geq}$, *for any monotone quasi order* $\leq$.

Proof. Any full trio contains all regular languages. Consequently, the result follows from Theorem 6.3. ∎

8 Conclusion

We have investigated the confluence property. As proved in Lemma 3.5, for any quasi order $\leq$, a language L is confluent w.r.t. $\leq$ if and only if its down-set $\text{DOWN}_{\leq}(L)$ is confluent w.r.t. $\leq$. On the other hand, for any monotone well quasi order $\leq$, the down-set $\text{DOWN}_{\leq}(L)$ of any language L is regular, as proved in Theorem 6.3. It follows that, for any monotone well quasi order $\leq$ such that the confluence problem w.r.t. $\leq$ is decidable for regular languages, the effectiveness of this regularity for a certain family of languages $\mathcal{L}$ implies the decidability of the confluence problem w.r.t. $\leq$ in $\mathcal{L}$. Therefore, the confluence property is suitable for finding families for which the regularity of down-sets w.r.t. $\leq$ is not effective (as in Theorem 6.9).

There are two general problems which should be investigated (or their restrictions to some particular quasi orders):

Problem 1. When does the decidability of the confluence problem entail the effectiveness of the regularity of down-sets?

Problem 2. Find general conditions which are equivalent to the effectiveness of the regularity of down-sets (as, for instance, the one by van Leeuwen for $\geq_s$ in Theorem 6.5).

References

[AG] J.-M. Autebert, J. Gabarro, Iterated GSM's and co-CFL, *Acta Informatica* **26** (1989) 749–769.

[AFG] J.-M. Autebert, P. Flajolet, J. Gabarro, Prefixes of infinite words and ambiguous context-free languages, *Inform. Proc. Letters* **25** (1987) 211–216.

[Bea1] D. Beauquier, Thin homogeneous sets of factors, *Proc. of the 6th Conference of Foundations of Software Technology and Theoret. Comput. Sci.*, Lecture Notes in Computer Science 241 (1986) 239–251.

[Bea2] D. Beauquier, Minimal automaton of a rational cover, *Proc. ICALP'87*, Lecture Notes in Computer Science 267 (1987) 174–189.

[BeaN] D. Beauquier, M. Nivat, About rational sets of factors of a bi-infinite word, *Proc. ICALP'85*, Lecture Notes in Computer Science 194 (1985) 33–42.

[Ber] J. Berstel, Every iterated morphism yields a co-CFL, *Inform. Proc. Letters* **22** (1986) 7–9.

[BoN] L. Boasson, M. Nivat, Adherence of languages, *J. Comput. System Sci.* **20** (1980) 285–309.

[CK] C. Choffrut, J. Karhumäki, Combinatorics of Words, in G. Rozenberg, A. Salomaa, eds., *Handbook of Formal Languages, Vol 1* (Springer-Verlag, Berlin, Heidelberg, 1997) 329–438.

[dLV] A. de Luca, S. Varricchio, Unavoidable Regularity and Finiteness Conditions, in G. Rozenberg, A. Salomaa, eds., *Handbook of Formal Languages, Vol 1* (Springer-Verlag, Berlin, Heidelberg, 1997) 747–810.

[EHR] A. Ehrenfeucht, D. Haussler, G. Rozenberg, On regularity of context-free languages, *Theoret. Comput. Sci.* **27** (1983) 311–332.

[Ei] S. Eilenberg, *Automata, Languages and Machines*, Vol A (Academic Press, New York, 1974).

[ER] P. Erdös, R. Rado, Sets having divisor property, Solution to problem 4358, *Amer. Math. Monthly* **59** (1952) 255–257.

[Gi] S. Ginsburg, *The Mathematical Theory of Context-Free Languages* (McGraw-Hill Book Comp., 1966).

[Ha] L. H. Haines, On free monoids partially ordered by embedding, *J. Comb. Theory* **6** (1969) 94–98.

[HI] T. Harju, L. Ilie, Languages obtained from infinite words, submitted, 1995.

[Hi] G. Higman, Ordering by divisibility in abstract algebras, *Proc. London Math. Soc.* (3) **2** (1952) 326–336.

[HU] J. E. Hopcroft, J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, Reading, Mass., 1979).

[Kr1] J. B. Kruskal, The theory of well-partially-ordered sets, Doctoral Thesis, Princeton, 1954.

[Kr2] J. B. Kruskal, The theory of well-quasi-ordering; a frequently discovered concept, *J. Comb. Theory* (A) **13** (1972) 297–305.

[Lo] M. Lothaire, *Combinatorics on Words* (Addison-Wesley, Reading, Mass., 1983).

[MP] S. Marcus, Gh. Păun, Infinite words and their associated formal languages, in: C. Calude, M. J. J. Lennon, H. Maurer (eds.), *Salodays in Auckland* (Auckland Univ. Press, 1994) 95–99.

[Sa] A. Salomaa, *Formal Languages* (Academic Press, New York, 1973).

[vL] J. van Leeuwen, Effective constructions in well-partially-ordered free monoids, *Discrete Mathematics* **21** (1978) 237–252.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.abo.fi

University of Turku
- **Department of Mathematical Sciences**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**