The Decidability of the Generalized Confluence Problem for Context-Free Languages

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

Turku Centre for Computer Science
TUCS Technical Report No 100
March 1997
ISBN 951-650-973-8
ISSN 1239-1891

Abstract

A language $L \subseteq \Sigma^*$ is confluent with respect to a given quasi order $\leq$ on Σ^* if, for any $x, y \in L$, there is $z \in L$ such that $x \leq z$ and $y \leq z$. L is confluent with respect to $\leq$ in generalized sense if it is a finite union of languages confluent with respect to $\leq$. We investigate in this paper the decidability of the generalized confluence problem with respect to the prefix partial order and the factor partial order for context-free languages, thus generalizing previuos results concerning the decidability of the ordinary confluence problem.

1 Introduction

Starting from the property of the language of all factors of an infinite word to contain, for any two words of it, one which has the former two as factors, one introduced in [HI] the notion of confluence. Generally, for any quasi order on a free monoid, the confluence property is the property of a language to contain, for any two words of it, one which is bigger, with respect to the given quasi order, than each of the former two. As shown in [HI], the decidability of the confluence problem is closely related to the effective regularity of down-sets, where the down-set of a language is the set of all words smaller than the words in the language.

The confluence property was generalized in [IS] by considering languages which are finite unions of confluent languages. It was shown there that the generalized confluence property is closely related to the property of a quasi order of being a well quasi order.

We study in this paper the decidability of the generalized confluence problem with respect to the prefix partial order and the factor partial order for context-free languages. Our results generalize the similar ones for the ordinary confluence problem in [HI].

The paper is organized as follows. After giving the basic definitions in the next section, we prove in section 3 that the generalized confluence problem with respect to the prefix partial order is decidable for context-free languages. It turns out that this decidability is closely connected with the decidability of the slenderness problem for context-free languages. In section 4 we prove that the same problem is undecidable for the factor partial order.

2 Definitions

For an alphabet Σ, we denote by Σ^* the free monoid generated by Σ and by λ, the empty word, its identity. For a word $w \in \Sigma^*$, $|w|$ denotes the length of w.

Given a quasi order $\leq$ on Σ^*, we say that a language $L \subseteq \Sigma^*$ is *confluent w.r.t.* $\leq$ if, for any $x, y \in L$, there is $z \in L$ such that $x \leq z$ and $y \leq z$. For a positive integer $k \geq 1$, $L \subseteq \Sigma^*$ is called *k-confluent w.r.t.* $\leq$ if

$$L = \bigcup_{i=1}^{k} L_i,$$

for some $L_i \subseteq \Sigma^*, L_i$ confluent w.r.t. $\leq$, for any $1 \leq i \leq k$. L is *confluent w.r.t.* $\leq$ *in generalized sense* if L is k-confluent w.r.t. $\leq$, for some $k \geq 1$.

Notice that this is a generalization of the ordinary confluence since 1-confluent is the same as confluent. Also remark the the empty language is confluent w.r.t. any quasi order and, for any two quasi orders $\leq_1$ and $\leq_2$ on Σ^*, if $\leq_1 \subseteq \leq_2$, then any language confluent w.r.t. $\leq_1$ is confluent also w.r.t. $\leq_2$.

The *down-set of L w.r.t.* $\leq$, denoted $\downarrow_{\leq} L$, is the set

$$\downarrow_{\leq} L = \{w \in \Sigma^* \mid w \leq u \text{ for some } u \in L\}.$$

For any quasi order $\leq$ on Σ^*, the down-operator $\downarrow_{\leq}$ is monotone and idempotent.

The two partial orders we deal with in this paper are as follows, where u, v are finite words over an alphabet Σ:

prefix: $u \leq_p v$ iff there is $w \in \Sigma^*$ such that $v = uw$,
factor: $u \leq_f v$ iff there is $w, z \in \Sigma^*$ such that $v = wuz$,

Examples

1. The language

$$L_1 = \{a^n b \mid n \geq 0\}$$

is obviously confluent w.r.t. $\leq_f$ but it is not confluent w.r.t. $\leq_p$. In fact, L_1 is not confluent w.r.t. $\leq_p$ even in generalized sense. Indeed, if R is a nonempty subset of L_1 such that R is confluent w.r.t. $\leq_p$, then R has exactly one element, since any two elements of L_1 are incomparable w.r.t. $\leq_p$. Thus, as L_1 is infinite, it cannot be written as a finite union of languages confluent w.r.t. $\leq_p$.

2. For any language $L \subseteq \Sigma^*$, the catenation closure of L, L^*, is confluent w.r.t. any quasi order $\leq$ on Σ^* which is an extension of the factor partial order $\leq_f$.

3. For any alphabet Σ such that card$(\Sigma) \geq 2$, the total language Σ^* is not confluent w.r.t. $\leq_p$, even in the generalized sense. To prove this, consider two different letters of Σ, say a and b. For any $k \geq 1$, consider the following $k+1$ words in Σ^*: $ab^{k+1}, a^2b^k, \ldots, a^{k+1}b$. Since, for any two words x, y of the previous $k+1$, there is no word $z \in \Sigma^*$ such that $x \leq_p z$ and $y \leq_p z$, it follows that Σ^* is not k-confluent w.r.t. $\leq_p$.

For all formal language theory results we refer to [HU] and [Sa] and for all combinatorics on words results we refer to [CK] and [Lo].

In what concerns the decidability of the generalized confluence problem w.r.t. a given quasi order $\leq$, there are two problems. We formulate them for context-free languages but they can be considered for any family of languages.

1. Given a fixed $k \geq 1$, is it decidable whether or not an arbitrary context-free language is k-confluent w.r.t. $\leq$?

2. Is it decidable whether or not an arbitrary context-free language is confluent w.r.t. $\leq$ in generalized sense, that is, is there some $k \geq 1$ such that the given language is k-confluent w.r.t. $\leq$?

We prove in the next sections that both problems are decidable for the prefix partial order $\leq_p$ and undecidable for the factor partial order $\leq_f$.

3 Prefixes of words

We prove in this section that both problems, the k-confluence problem w.r.t. $\leq_p$ and the generalized confluence problem w.r.t. $\leq_p$ are decidable for context-free languages.

For an infinite word α, we denote by $Pref(\alpha)$ the set of all finite prefixes of α. The family of languages which are prefixes of infinite words is denoted by $\mathcal{F}_{pref}$, that is,

$$\mathcal{F}_{pref} = \{L \mid L = Pref(\alpha) \text{ for some infinite word } \alpha\}.$$

Our first result is a connection between k-confluence and the family $\mathcal{F}_{pref}$. It generalizes the following result in [HI].

Lemma 3.1 ([HI]). *Any language $L \subseteq \Sigma^*$ is in the family $\mathcal{F}_{pref}$ if and only if the following conditions are fulfilled: (i) L is infinite, (ii) $L = \downarrow_{\leq_p} L$, (iii) L is confluent w.r.t. $\leq_p$.*

We need also the next result from [IS]. (It was proved first for $k = 1$ in [HI].)

Lemma 3.2 ([IS]). *For any quasi order $\leq$ on Σ^* and $k \geq 1$, a language $L \subseteq \Sigma^*$ is k-confluent w.r.t. $\leq$ if and only if $\downarrow_{\leq} L$ is k-confluent w.r.t. $\leq$.*

Lemma 3.3. *Let $k \geq 1$ and $L \subseteq \Sigma^*$ an infinite language such that $L = \downarrow_{\leq_p} L$. Then L is k-confluent w.r.t. $\leq_p$ if and only if*

$$L = \bigcup_{i=1}^{k} L_i$$

such that, for any $1 \leq i \leq k$, $L_i \subseteq \Sigma^*$ and L_i is either finite and confluent w.r.t. $\leq_p$ or else infinite and $L_i \in \mathcal{F}_{pref}$.

Proof. Take $k \geq 1$ and $L \subseteq \Sigma^*$ such that L is infinite and $L = \downarrow_{\leq_p} L$. Suppose first that L is k-confluent w.r.t. $\leq_p$ and put

$$L = \bigcup_{i=1}^{k} L_i,$$

for some $L_i \subseteq \Sigma^*$, L_i confluent w.r.t. $\leq_p$, for any $1 \leq i \leq k$. Since

$$\downarrow_{\leq_p} L = \downarrow_{\leq_p} (\bigcup_{i=1}^{k} L_i) = \bigcup_{i=1}^{k} \downarrow_{\leq_p} L_i,$$

we may suppose that $L_i = \downarrow_{\leq_p} L_i$, for any $1 \leq i \leq k$. Notice that we may indeed make this assumption since L_i is confluent w.r.t. $\leq_p$ and so, by Lemma 3.2, $\downarrow_{\leq_p} L_i$ is also confluent w.r.t. $\leq_p$.

If, for some $1 \leq i \leq k$, L_i is infinite, as $L_i = \downarrow_{\leq_p} L_i$ and L_i is confluent w.r.t. $\leq_p$, it follows by Lemma 3.1 that $L_i \in \mathcal{F}_{pref}$. One implication is proved.

For the converse one, it is enough to observe that any language in the family $\mathcal{F}_{pref}$ is confluent w.r.t. $\leq_p$. The proof is concluded.

□

We need one more definition from [PS]. For a language $L \subseteq \Sigma^*$, denote

$$l_n(L) = \mathrm{card}\{w \in L \mid |w| = n\},$$

for any $n \geq 0$. For any $k \geq 0$, the language L is called *k-slender* if and only if $l_n(L) \leq k$, for any $n \geq 0$. L is called *slender* if it is k-slender, for some $k \geq 1$.

Remind the following two results from [Il1] and [Il2].

Lemma 3.4 ([Il1]). *A context-free language $L \subseteq \Sigma^*$ is slender if and only if it is a finite union of sets of the form*

$$\{uv^n wx^n y \mid n \geq 0\},$$

where $u, v, w, x, y \in \Sigma^*$.

Lemma 3.5 ([Il2]). *It is decidable whether or not an arbitrary context-free language $L \subseteq \Sigma^*$ is slender. In the affirmative case, the smallest constant k such that L is k-slender is computable. Also, a representation of L as in Lemma 3.4 can be effectively found.*

Lemma 3.6. *Let $k \geq 1$ and $L \subseteq \Sigma^*$ be an infinite context-free language. Then L is k-confluent w.r.t. $\leq_p$ if and only if the following two conditions are fulfilled:*

(i) $\downarrow_{\leq_p} L$ is k-slender,

(ii) $\downarrow_{\leq_p} L$ is a finite union

$$\downarrow_{\leq_p} L = \bigcup_{i=1}^{k} L_i,$$

for some $L_i \subseteq \Sigma^*$, such that for any $1 \leq i \leq k, L_i$ is either finite and confluent w.r.t. $\leq_p$ or else infinite and of the form

$$L_i = \downarrow_{\leq_p} \{z_{i1} z_{i2}^n \mid n \geq 0\},$$

for some $z_{i1}, z_{i2} \in \Sigma^*, z_{i2} \neq \lambda$.

Proof. Take $k \geq 1$ and $L \subseteq \Sigma^*$, such that L is k-confluent w.r.t. $\leq_p$. By Lemma 3.2, the down-set $\downarrow_{\leq_p} L$ is confluent w.r.t. $\leq_p$ and so, by Lemma 3.3, we have

$$\downarrow_{\leq_p} L = \bigcup_{i=1}^{k} L_i,$$

for some $L_i \subseteq \Sigma^*, L_i$ either finite and confluent w.r.t. $\leq_p$ or else infinite and $L_i \in \mathfrak{F}_{pref}$.

Clearly, if $L_i \in \mathfrak{F}_{pref}$, then $l_n(L_i) = 1$, for any $n \geq 0$. On the other hand, if L_i is finite, then $l_n(L_i) \leq 1$, for any $n \geq 0$. Indeed, if $u, v \in L_i$ such that $|u| = |v|$, then, as L_i is k-confluent w.r.t. $\leq_p$, there is $w \in L_i$ such that $u \leq_p w$ and $v \leq_p w$. Thus, either $u \leq_p v$ or $v \leq_p u$. In any case, it follows that $u = v$.

Therefore, $l_n(\downarrow_{\leq_p} L) \leq k$, for any $n \geq 0$, and (i) is proved.

The following claim will be useful for the proof of (ii).

***Claim* 1.** $\downarrow_{\leq_p} L$ contains no infinite antichain of $\leq_p$.

Proof of Claim 1. Let $A \subseteq \downarrow_{\leq_p} L$ be an antichain of $\leq_p$. We have

$$A = \bigcup_{i=1}^{k} (L_i \cap A).$$

If $u, v \in L_i \cap A$, then, as L_i is confluent w.r.t. $\leq_p$, there is $w \in L_i$ such that $u \leq_p w$ and $v \leq_p w$. Thus either $u \leq_p v$ or $v \leq_p u$. In both cases, since A is an antichain of $\leq_p$, it follows that $u = v$, so $\mathrm{card}(L_i \cap A) \leq 1$. Therefore, $\mathrm{card}(A) \leq k$ and so A is finite. The claim is proved.

Since $\downarrow_{\leq_p} L$ is slender, it follows by Lemma 3.4 that it is a finite union of the form

$$\downarrow_{\leq_p} L = \bigcup_{i=1}^{m} \{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\},$$

for some $m \geq 1$ and $u_i, v_i, w_i, x_i, y_i \in \Sigma^*, 1 \leq i \leq m$. Thus

$$\downarrow_{\leq_p} L = \bigcup_{i=1}^{m} \downarrow_{\leq_p} \{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\},$$

where we suppose that no set in the union in the right-hand side is redundant, that is, for no $i_0, 1 \leq i_0 \leq m$,

$$\downarrow_{\leq_p} L = \bigcup_{\substack{i=1 \\ i \neq i_0}}^{m} \downarrow_{\leq_p} \{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\},$$

***Claim* 2.** If, for some $1 \leq i \leq m$, $v_i x_i \neq \lambda$, then

$$\downarrow_{\leq_p} \{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\} = \downarrow_{\leq_p} \{z_{i1} z_{i2}^n \mid n \geq 0\},$$

for some $z_{i1}, z_{i2} \in \Sigma^*, z_{i2} \neq \lambda$.

Proof of Claim 2. Because $v_i x_i \neq \lambda$, it follows that the set $\{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\}$ is infinite. As, by Claim 1, there is no infinite antichain of $\leq_p$ in $\downarrow_{\leq_p} L$, we can find a strictly increasing infinite sequence of positive integers $(i_n)_{n \geq 1}$ such that

$$u_i v_i^{i_n} w_i x_i^{i_n} y_i \leq_p u_i v_i^{i_{n+1}} w_i x_i^{i_{n+1}} y_i,$$

for any $n \geq 1$. Thus, for any $n, p \geq 1$,

$$w_i x_i^{i_n} y_i \leq_p v_i^{i_{n+p} - i_n} w_i x_i^{i_{n+p}} y_i.$$

Suppose that $v_i \neq \lambda$ and $x_i \neq \lambda$. Then, as the length of $v_i^{i_{n+p} - i_n}$ tends to infinity with p, for a large enough i_n, $x_i^{i_n}$ and $v_i^{i_{n+p} - i_n}$ will overlap each other on a part longer than $|x_i| + |v_i|$. By Fine and Wilf's theorem, it follows that x_i and v_i are powers of conjugates of the same word. Thus, there are $z', z'' \in \Sigma^*$ such that $x_i = (z'z'')^{r_1}, v_i = (z''z')^{r_2}$, for some $r_1 > 0, r_2 > 0$. Then $w_i = (z''z')^{r_3} z''$, for some $r_3 \geq 0$ and $y_i = (z'z'')^{r_4} t'$, for some $r_4 \geq 0$ and $t' \in \Sigma^*, t' \leq_p z'z''$. Let us put $z'z'' = t't'', t'' \in \Sigma^*$. Therefore

$$\begin{aligned}
\{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\} &= \{u_i (z''z')^{nr_2} (z''z')^{r_3} z'' (z'z'')^{nr_1} (z'z'')^{r_4} t' \mid n \geq 0\} \\
&= \{u_i z'' (z'z'')^{r_3 + r_4 + n(r_1 + r_2)} t' \mid n \geq 0\} \\
&= \{u_i z'' (t't'')^{r_3 + r_4 + n(r_1 + r_2)} t' \mid n \geq 0\} \\
&= \{z_{i1} z_{i2}^n \mid n \geq 0\},
\end{aligned}$$

for

$$z_{i1} = u_i z'' t' (t'' t')^{r_3 + r_4},$$
$$z_{i2} = (t'' t')^{r_1 + r_2}.$$

Notice that $z_{i2} \neq \lambda$.

The cases $v_i \neq \lambda, x_i = \lambda$ and $v_i = \lambda, x_i \neq \lambda$ are treated similarly. The claim is proved.

***Claim* 3.** *$m \leq k$.*

Proof of Claim 3. By Lemma 3.3, $\downarrow_{\leq_p} L$ is a finite union of k sets of prefixes of some words, finite or infinite, say

$$L = \bigcup_{i=1}^{k} R_i.$$

So far, we proved that $\downarrow_{\leq_p} L$ is finite union of m sets of prefixes of some words which are either finite or infinite and ultimately periodic. Since we have supposed that none of the words in the second representation is redundant, it follows that

$$\mathrm{card}\{i \mid R_i \text{ is infinite}\} \geq \mathrm{card}\{i \mid z_{i2} \neq \lambda\}.$$

Indeed, if $\downarrow_{\leq_p} \{z_{i1} z_{i2}^n \mid n \geq 0\} \cap R_j$ is infinite, then $\downarrow_{\leq_p} \{z_{i1} z_{i2}^n \mid n \geq 0\} \subseteq R_j$, since $R_j \in \mathfrak{F}_{pref}$. It is also clear that

$$\mathrm{card}\{i \mid R_i \text{ is finite}\} \geq \mathrm{card}\{i \mid z_{i2} = \lambda\}.$$

Therefore $m \leq k$ and the claim is proved.

By Claim 2 and Claim 3, one implication in the statement is proved.

The converse implication is clear since L and $\downarrow_{\leq_p} L$ are k-confluent w.r.t. $\leq_p$ at the same time. The proof is concluded.

□

Remark. It is clear that condition (i) in Lemma 3.6 does not imply that L is k-confluent w.r.t. $\leq_p$. We notice that, even with the hypothesis that L is confluent w.r.t. $\leq_p$ in generalized sense, it still does not follow that L is k-confluent w.r.t. $\leq_p$. For instance, the language

$$L = Pref(ab^\omega) \cup Pref(a^2 b^\omega) \cup \{b\},$$

where b^ω is the infinite word $bbb\ldots$, is 2-slender but it is not 2-confluent w.r.t. $\leq_p$. L is 3-confluent w.r.t. $\leq_p$.

We can prove now the decidability of the k-confluence problem w.r.t. $\leq_p$ for context-free languages.

Theorem 3.7. *For any $k \geq 1$, it is decidable whether or not an arbitrary context-free language $L \subseteq \Sigma^*$ is k-confluent w.r.t. $\leq_p$.*

Proof. Take $k \geq 1$ and an arbitrary context-free language $L \subseteq \Sigma^*$. We give the following algorithm to decide whether or not L is k-confluent w.r.t. $\leq_p$.

Algorithm 3.8.

1. Decide whether or not L is finite [possible for context-free languages]. If yes, then decide directly whether or not L is k-confluent w.r.t. $\leq_p$ [obviously possible], else go to step 2.

2. Construct $\downarrow_{\leq_p} L$ [this is effectively constructable for L context-free].

3. Decide whether or not $\downarrow_{\leq_p} L$ is slender [possible by Lemma 3.5]. If yes, then go to step 4, else answer **no** [correct by Lemma 3.6].

4. Compute the minimal k' such that $\downarrow_{\leq_p} L$ is k'-slender [possible by Lemma 3.5].

5. If $k' \leq k$, then go to step 6, else answer **no** [correct by Lemma 3.6].

6. Find a representation of $\downarrow_{\leq_p} L$ as in Lemma 3.4 [possible by Lemma 3.5].

7. Check whether this decomposition verifies the conditions in the statement of Lemma 3.6 [this is clearly possible]. If yes, then answer **yes**, else answer **no** [correct by Lemma 3.6].

□

We move now to the generalized confluence problem w.r.t. $\leq_p$. We intend to prove that also this problem is decidable for context-free languages. We need the following result which is a corollary of Lemma 3.6.

Corollary 3.9. *Let $L \subseteq \Sigma^*$ be a context-free language. Then L is confluent w.r.t. $\leq_p$ in generalized sense if and only if $\downarrow_{\leq_p} L$ is slender and there is a deconposition of L*

$$\downarrow_{\leq_p} L = \bigcup_{i=1}^{k} \downarrow_{\leq_p} \{u_i v_i^n \mid n \geq 0\},$$

for some $k \geq 1$ and $u_i, v_i \in \Sigma^*, 1 \leq i \leq k$.

Proof. Take $L \subseteq \Sigma^*$ a context-free language which is confluent w.r.t. $\leq_p$ in generalized sense. Then, for some $k \geq 1$, L is k-confluent w.r.t. $\leq_p$ and the claim follows by Lemma 3.6.

The converse implication is clear since the set of prefixes of a finite or infinite word is confluent w.r.t. $\leq_p$.

□

As a consequence of Corrolary 3.9, we have the next result.

Corollary 3.10. *If $L \subseteq \Sigma^*$ is a context-free language which is confluent w.r.t. $\leq_p$ in generalized sense, then L is regular.*

We can give now the decidability result.

Theorem 3.11. *It is decidable whether or not an arbitrary context-free language is confluent w.r.t. $\leq_p$ in generalized sense.*

Proof. Let $L \subseteq \Sigma^*$ be a an arbitrary context-free language. The algorithm to decide whether or not L is confluent w.r.t. $\leq_p$ in generalized sense is clear now from Corollary 3.9.

Algorithm 3.12.

1. Decide whether or not L is finite [possible for context-free languages]. If yes, then answer **yes** [obviously correct], else go to step 2.
2. Construct $\downarrow_{\leq_p} L$ [this is effectively constructable for L context-free].
3. Decide whether or not $\downarrow_{\leq_p} L$ is slender [possible by Lemma 3.5]. If yes, then go to step 4, else answer **no** [correct by Corollary 3.9].
4. Find a representation of $\downarrow_{\leq_p} L$ as in Lemma 3.4 [possible by Lemma 3.5].
5. Check whether this decomposition verifies the conditions in the statement of Corollary 3.9 [this is clearly possible]. If yes, then answer **yes**, else answer **no** [correct by Corollary 3.9].

□

The next corollary is a direct consequence of Theorem 3.7, Theorem 3.11, and Lemma 3.6.

Corollary 3.13. *For a context-free language L which is confluent w.r.t. $\leq_p$ in generalized sense, the minimal k such that L is k-confluent w.r.t. $\leq_p$ is computable. Also*

$$\min\{k' \mid \downarrow_{\leq_p} L \textit{ is } k'\textit{-slender}\} \leq \min\{k'' \mid L \textit{ is } k''\textit{-confluent w.r.t. } \leq_p\}.$$

4 Factors of words

We deal in this section with the factor partial order $\leq_f$ and prove that the k-confluence problem w.r.t. $\leq_f$ and the generalized confluence problem w.r.t. $\leq_f$ are undecidable for context-free languages.

We prove first a general undecidability result concerning the decidability of the k-confluence problem w.r.t. $\leq_f$.

Lemma 4.1. *Let $k \geq 1$ and $\mathfrak{L}$ be a family of languages effectively closed under union, catenation with letters, and λ-free catenation closure such that the inclusion problem is undecidable in $\mathfrak{L}$. Then also the k-confluence problem is undecidable in $\mathfrak{L}$.*

Proof. We argue by contradiction. Suppose that the k-confluence problem w.r.t. $\leq_f$ is decidable in $\mathfrak{L}$.

***Claim* 1.** The emptiness problem is decidable in $\mathfrak{L}$.

Proof of Claim 1. Take an arbitrary language $L \in \mathfrak{L}, L \subseteq \Sigma^*$, and construct the language

$$L_1 = \bigcup_{i=0}^{k} \#L\#_i,$$

where $\#, \#_0, \#_1, \ldots, \#_k \notin \Sigma$ are new letters. It follows by hypothesis that $L_1 \in \mathfrak{L}$ and we claim that L is empty if and only if L_1 is k-confluent w.r.t. $\leq_f$.

If $L = \emptyset$, then $L_1 = \emptyset$, so L_1 is k-confluent w.r.t. $\leq_f$.

Conversely, suppose that L_1 is k-confluent w.r.t. $\leq_f$ but $L \neq \emptyset$ and take $w \in L$. We have by definition

$$L_1 = \bigcup_{i=1}^{k} R_i,$$

for some $R_i \subseteq \Sigma^*, R_i$ confluent w.r.t. $\leq_f$. Consider the words $\#w\#_i, 0 \leq i \leq k$. They belong all to L_1. But no R_i can contain two of them. Indeed, if $\#w\#_{i_1}, \#w\#_{i_2} \in R_i$, for some $0 \leq i_1, i_2 \leq k, i_1 \neq i_2$, then, as R_i is confluent w.r.t. $\leq_f$, there is $z \in R_i$ such that $\#w\#_{i_1} \leq_f z$ and $\#w\#_{i_2} \leq_f z$. Then, from $\#w\#_{i_1} \leq_f z$, we get $z = \#w\#_{i_1}$ so $\#w\#_{i_2} \not\leq_f z$, a contradiction. Therefore L is empty.

Consequently, to decide whether or not L is empty, we decide whether or not L_1 as constructed above is k-confluent w.r.t. $\leq_f$, which is possible by our assumption, since L_1 is effectively in $\mathfrak{L}$. The claim is proved.

We show that also the inclusion problem is decidable in $\mathfrak{L}$, thus obtaining a contradiction. For, take $L_1, L_2 \in \mathfrak{L}, L_1 \subseteq \Sigma^*, L_2 \subseteq \Sigma^*$, and construct the language

$$L = \#L_1\#_0 \cup \bigcup_{i=1}^{k} (\#L_2\#_0\#_i)^+,$$

where $\#, \#_0, \#_1, \ldots, \#_k \notin \Sigma$ are new letters. Then, by hypothesis, $L \in \mathfrak{L}$.

***Claim* 2.** L is k-confluent w.r.t. $\leq_f$ if and only if either $L_1 \subseteq L_2$ or else $1 \leq \mathrm{card}(L_1) \leq k$ and $L_2 = \emptyset$.

Proof of Claim 2. Suppose first that $L_1 \subseteq L_2$. Then

$$\downarrow_{\leq_f} L = \downarrow_{\leq_f} (\#L_1\#_0) \cup \bigcup_{i=1}^{k} \downarrow_{\leq_f} (\#L_2\#_0\#_i)^+ = \bigcup_{i=1}^{k} \downarrow_{\leq_f} (\#L_2\#_0\#_i)^+$$

since

$$\downarrow_{\leq_f} (\#L_1\#_0) \subseteq \downarrow_{\leq_f} (\#L_2\#_0\#_i) \subseteq \downarrow_{\leq_f} (\#L_2\#_0\#_i)^+,$$

for any $1 \leq i \leq k$. As, clearly, $(\#L_2\#_0\#_i)^+$ is confluent w.r.t. $\leq_f$, by Lemma 3.2, $\downarrow_{\leq_f} (\#L_2\#_0\#_i)^+$ is confluent w.r.t. $\leq_f$. Thus $\downarrow_{\leq_f} L$ if k-confluent w.r.t. $\leq_f$ and so, again by Lemma 3.2, L is k-confluent w.r.t. $\leq_f$.

Suppose now that $1 \leq \mathrm{card}(L_1) \leq k$ and $L_2 = \emptyset$. Then, if $L_1 = \{w_1, w_2, \ldots w_p\}$, $1 \leq p \leq k$, then

$$L = \bigcup_{i=1}^{p} \{\#w_i\#_0\}.$$

As $\{\#w_i\#_0\}$ is confluent w.r.t. $\leq_f$, it follows that L is p-confluent w.r.t. $\leq_f$ hence also k-confluent w.r.t. $\leq_f$.

For the converse part, suppose that L is k-confluent w.r.t. $\leq_f$. Suppose also that $L_1 \not\subseteq L_2$ and take $w_1 \in L_1 \backslash L_2$. If $L_2 \neq \emptyset$, then take $w_2 \in L_2$. It follows that

$$\{\#w_1\#_0, \#w_2\#_0\#_1, \#w_2\#_0\#_2, \ldots, \#w_2\#_0\#_k\} \subseteq L$$

and a contradiction with the k-confluence w.r.t. $\leq_f$ of L is obtained as in the proof of Claim 1. Thus $L_2 = \emptyset$ and so

$$L = \#L_1\#_0.$$

If $\mathrm{card}(L_1) \geq k+1$, then take $z_1, z_2, \ldots, z_{k+1} \in L_1, z_i \neq z_j$, for any $i \neq j$. We have that

$$\{\#z_1\#_0, \#z_2\#_0, \ldots \#z_{k+1}\#_0\} \subseteq L$$

and again a contradiction with the k-confluence w.r.t. $\leq_f$ of L is obtained. Thus $\mathrm{card}(L_1) \leq k$. As $L_1 \not\subseteq L_2 = \emptyset$, we get also $\mathrm{card}(L_1) \geq 1$. The claim is proved.

***Claim* 3.** The inclusion problem is decidable in $\mathfrak{L}$.

Proof of Claim 3. Take two arbitrary language $L_1, L_2 \in \mathfrak{L}$ and construct the language L as above. We give the following algorithm to decide whether or not $L_1 \subseteq L_2$ and prove the claim.

Algorithm 4.2.

1. Decide whether or not $L_2 = \emptyset$ [the emptiness problem is decidable in $\mathfrak{L}$ by Claim 1]. If yes, then go to step 2, else go to step 3.

2. Decide whether or not $L_1 = \emptyset$ [possible by Claim 1]. If yes, then answer **yes** [obviously correct], else answer **no** [again correct].

3. Decide whether or not L is k-confluent w.r.t. $\leq_f$ [the k-confluence problem is decidable in $\mathfrak{L}$ by our assumption and L is effectively in $\mathfrak{L}$]. If yes, then answer **yes** [correct by Claim 2], else answer **no** [correct also by Claim 2].

Consequently, we have proved that the decidability of the k-confluence problem w.r.t. $\leq_f$ for languages in $\mathfrak{L}$ would entail the decidability of the inclusion problem for languages in $\mathfrak{L}$. Since the latter is undecidable in $\mathfrak{L}$, it follows that also the former is undecidable in $\mathfrak{L}$ and the result is proved.

□

As a direct corollary of Lemma 4.1, we obtain the undecidability of the k-confluence problem w.r.t. $\leq_f$ for context-free languages.

Theorem 4.3. *For any $k \geq 1$, it is undecidable whether or not an arbitrary context-free language is k-confluent w.r.t. $\leq_f$.*

We consider now the generalized confluence problem w.r.t. $\leq_f$ and give the following general undecidability result.

Lemma 4.4. *Let $\mathfrak{L}$ be a family of languages effectively closed under union, catenation, and λ-free catenation closure such that all regular languages are in $\mathfrak{L}$. If it is undecidable for an arbitrary language $L \in \mathfrak{L}, L \subseteq \Sigma^*$, whether or not $L = \Sigma^*$, then the generalized confluence problem w.r.t. $\leq_f$ is undecidable for languages in $\mathfrak{L}$.*

Proof. We argue by contradiction. Suppose that the generalized confluence problem w.r.t. $\leq_f$ is decidable in $\mathfrak{L}$. Then we prove that it is decidable for an arbitrary language $L \in \mathcal{L}, L \subseteq \Sigma^*$, whether or not $L = \Sigma^*$. For, take an arbitrary such language L and construct the language

$$L_1 = \#\Sigma^*\#_1\#_2^*\#_1 \cup (\#L\#_1\#_2^*\#_1)^+,$$

where $\#, \#_1, \#_2 \notin \Sigma$ are new letters. Then, by hypothesis, $L_1 \in \mathcal{L}$.

Claim. L_1 is confluent w.r.t. $\leq_f$ in generalized sense if and only if $L = \Sigma^*$.

Proof of Claim. Suppose first that $L = \Sigma^*$. Then

$$\downarrow_{\leq_f} L_1 = \downarrow_{\leq_f} (\#\Sigma^*\#_1\#_2^*\#_1) \cup \downarrow_{\leq_f} (\#\Sigma^*\#_1\#_2^*\#_1)^+ = \downarrow_{\leq_f} (\#\Sigma^*\#_1\#_2^*\#_1)^+.$$

Then, as $(\#\Sigma^*\#_1\#_2^*\#_1)^+$ is confluent w.r.t. $\leq_f$, it follows, by Lemma 3.2, that $\downarrow_{\leq_f} (\#\Sigma^*\#_1\#_2^*\#_1)^+$ is confluent w.r.t. $\leq_f$, so $\downarrow_{\leq_f} L_1$ is confluent w.r.t. $\leq_f$ and so, again by Lemma 3.2, L_1 is confluent w.r.t. $\leq_f$. It follows also that L_1 is confluent w.r.t. $\leq_f$ in generalized sense.

Conversely, suppose that L_1 is k-confluent w.r.t. $\leq_f$, for some $k \geq 1$, but $L \subset \Sigma^*$ and take $w \in \Sigma^* \backslash L$. Put also

$$L_1 = \bigcup_{i=1}^{k} R_i,$$

for some $R_i \subseteq \Sigma^*, R_i$ confluent w.r.t. $\leq_f$. We have then

$$\bigcup_{i=0}^{k} \{\#w\#_1\#_2^i\#_1\} \subseteq L_1.$$

and we can find an $i, 1 \leq i \leq k$, and $i_1 \neq i_2, 0 \leq i_1, i_2 \leq k$, such that $\#w\#_1\#_2^{i_1}\#_1$, $\#w\#_1\#_2^{i_2}\#_1 \in R_i$. As R_i is confluent w.r.t. $\leq_f$, there is $z \in R_i$ such that $\#w\#_1\#_2^{i_1}\#_1 \leq_f z$ and $\#w\#_1\#_2^{i_2}\#_1 \leq_f z$. Because $w \notin L$, it follows that $z \notin (\#L\#_1\#_2^*\#_1)^+$, and so $z = \#x\#_1\#_2^j\#_1$, for some $x \in \Sigma^*, j \geq 0$. From $\#w\#_1\#_2^{i_1}\#_1 \leq_f \#x\#_1\#_2^j\#_1$, we obtain $w = x$ and $i_1 = j$. But, since $i_1 \neq i_2$, $\#w\#_1\#_2^{i_2}\#_1 \not\leq_f z$, a contradiction. It follows that $L = \Sigma^*$. As k above has been chosen arbitrarily, the claim is proved.

Hence, to decide whether or not $L = \Sigma^*$, we can decide whether or not L_1 is confluent w.r.t. $\leq_f$ in generalized sense, which is possible by our assumption, since L_1 is effectively in $\mathcal{L}$. But this contradicts the hypothesis. Therefore the generalized confluence problem w.r.t. $\leq_f$ is undecidable in $\mathcal{L}$ and the result is proved.

□

We get imediately the following result.

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

References

[CK] C. Choffrut, J. Karhumäki, *Combinatorics of Words*, in *Handbook of Formal Languages*, (G. Rozenberg, A Salomaa, eds.), to appear.

[HI] T. Harju, L. Ilie, On quasi orders of words and the confluence property, submitted.

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

[Il1] L. Ilie, On a conjecture about slender context-free languages, *Theoretical Computer Science* **132** (1994) 427–434.

[Il2] L. Ilie, On slender context-free languages, manuscript.

[IS] L. Ilie, A. Salomaa, On well quasi orders of free monoids, submitted.

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

[PS] Gh. Păun, A. Salomaa, Thin and slender languages, *Discrete Appl. Math.* **61** (1995) 257–270.

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

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**