Insertion and Deletion of Words: Determinism and Reversibility

Lila Kari *

Academy of Finland and Department of Mathematics
University of Turku
20500 Turku
Finland

Abstract. The paper addresses two problems belonging to the basic combinatorics of words. They are connected with the operations of sequential insertion and deletion, which are nondeterministic versions of catenation and right/ left quotient. Necessary and sufficient conditions, under which the result of sequential insertion or deletion of two words is a singleton set, are given. Also the situation when the insertion and deletion are inverse to each other, which is not generally the case, is studied.

1 Introduction

Operations on languages are intensively studied in formal language theory. One of the main goals of the theory is to represent a family of languages as the closure of some atomic languages with respect to some operations. The theory of abstract families of languages deals with operations, many operations appear in formal language theory applications, and so on. The operations of *sequential insertion* and *sequential deletion* (called in the sequel, shortly, *insertion* and *deletion*), defined and studied in [2] play an important role in understanding the mechanisms of generating languages. The insertion and deletion operations are generalizations of catenation respectively right/left quotient and right/left derivative. The purpose of this paper is to study two specific problems concerning these operations. The problems belong to the very basic combinatorics of words.

The result of insertion (deletion) of a word into (from) another is in general a set with cardinality greater than one. In Section 3, necessary and sufficient condition under which the result of insertion or deletion of two words is a singleton set, are given.

In Section 4 we obtain some necessary and sufficient conditions under which the original word w results, after first inserting a word u into w and then deleting u from the result. The cryptographic connotations are obvious: after encrypting the message with a key, the decryption has to provide the original message.

Indeed, apart from formal language theory and combinatorics of words, these issues have recently become important in certain cryptographic considerations

* The work reported here is a part of the project 11281 of the Academy of Finland

(see [1], [4]). We will not enter a discussion on the connections with cryptography in this paper. For a detailed presentation of applications of formal languages in cryptography, see [5].

2 Basic Definitions and Notions

An alphabet is a finite nonempty set; its elements are called *letters* or *symbols*. If $\Sigma = \{a_1, \ldots, a_n\}$ is an alphabet then any sequence $w = a_{i_1} \ldots a_{i_k}$, $k \geq 0$, $a_{i_j} \in \Sigma$, $1 \leq j \leq k$, is called a *string (word)* over Σ. The length of the word w is denoted with $\lg(w)$ and, by definition, equals k. The word "consisting" of zero letters is denoted by λ and is called the *empty word*. Obviously, $\lg(\lambda) = 0$. The word consisting of repeating the word w j times is abbreviated w^j, with the convention $w^0 = \lambda$. The set of all words over Σ is denoted Σ^* and the set of all nonempty words, $\Sigma^* - \{\lambda\}$, is denoted Σ^+.

The set Σ^* is a monoid under the operation of catenation defined by:

$$uv = a_{i_1} \ldots a_{i_r} a_{j_1} \ldots a_{j_s} \ ,$$

where $u = a_{i_1} \ldots a_{i_r}$, $v = a_{j_1} \ldots a_{j_s}$, $r, s \geq 0$, a_{i_q}, $a_{j_p} \in \Sigma$ for $1 \leq q \leq r$, $1 \leq p \leq s$. The fact that r, s can be also zero means that the words u and v can be empty. The catenation operation is associative and λ is the neutral element of the monoid.

The *left quotient* of a word u by a word v is defined by:

$$v \backslash u = w \text{ iff } u = vw \ .$$

The *right quotient* of a word u by a word v is defined by:

$$u / v = w' \text{ iff } u = w'v \ .$$

Definition 1. *Let u, v be words over an alphabet Σ, The insertion of v into u is defined as:*

$$u \leftarrow v = \{u_1 v u_2 |\ u = u_1 u_2\} \ .$$

Example 1. Let $u = cd$, $v = a$. The insertion of v into u is $u \leftarrow v = \{acd, cad, cda\}$. Notice that $uv = cda$ is an element of the set $u \leftarrow v$. □

The insertion is neither an associative nor a commutative operation.

Definition 2. *Let u, v be words over an alphabet Σ. The deletion of v from u is defined as:*

$$u \rightarrow v = \{w \in \Sigma^* |\ u = w_1 v w_2, w = w_1 w_2, w_1, w_2 \in \Sigma^*\} \ .$$

Example 2. Let $u = abababa$ and $v = aba$, The result of the deletion of v from u is $u \rightarrow v = \{baba, abab, abba\}$. □

The deletion operation is neither associative nor commutative.

Note. The definitions of insertion and deletion can be extended to languages in the natural fashion. Indeed, for languages L_1, L_2 over Σ,

$$L_1 \leftarrow L_2 = \bigcup_{u \in L_1, v \in L_2} (u \leftarrow v) \text{ and } L_1 \rightarrow L_2 = \bigcup_{u \in L_1, v \in L_2} (u \rightarrow v) \ .$$

3 Determinism: When Is the Result a Singleton

The catenation and the right and left quotient of words are deterministic operations in the sense that the result of the operation is, in all three cases, a single word. The insertion and deletion are nondeterministic versions of catenation respectively right and left quotient. The result of the insertion (deletion) of one word into (from) another is in general a set whose cardinality is greater than one. A natural problem that arises is under what circumstances the insertion or the deletion of two words is deterministic, that is, produces as result a singleton set.

The structural property of words which influences the answer to this problem is whether or not they are *bordered* (the terminology is due to [7]). Before this, the notion of a primitive word is introduced.

Definition 3. *A word $u \in \Sigma^+$ is called a primitive word if $u = g^i$, $g \in \Sigma^+$, $i \geq 1$, implies that $i = 1$.*

Every word in Σ^+ can be expressed uniquely as a power of a primitive word (see [3], [7], p.7).

Definition 4. *A word $u \in \Sigma^+$ is called bordered if $u = xy = yx'$ for some $x, y, x' \in \Sigma^+$.*

A word which is not bordered will be called *unbordered*. Clearly, an unbordered word is primitive. Thus the set of unbordered words is a proper subfamily of the set of primitive words.

Example 3. The following words over $\Sigma = \{a, b\}$ are bordered: *aba*, *ababab*, *ababa*. The words *aab*, *abb*, a^2b^2 are unbordered. □

The following lemmas (see [7], pp.6-11) will be needed in the sequel:

Lemma 1. *Let x, y be words in Σ^* such that $xy \neq \lambda$. If $xy = yx$ then there uniquely exist a primitive word $g \in \Sigma^+$ and naturals, $i, j \geq 0$, $i + j > 0$, with the property $x = g^i$, $y = g^j$.*

Lemma 2. *If $g \in \Sigma^+$ is a primitive word such that $g = xy = yx$ for some $x, y \in \Sigma^*$, then $x = \lambda$ or $y = \lambda$.*

For a bordered primitive word we have the following property (see [8]):

Lemma 3. *Let u be a bordered primitive word in Σ^+. Then u can be expressed as $u = xyx$ for some $x, y \in \Sigma^+$.*

The following two theorems give necessary and sufficient conditions under which the result of the deletion of a word from another is a singleton.

Note. Let u, w be words in Σ^*. If $u = \lambda$ then $w \rightarrow u$ is a singleton, namely $\{w\}$. If $w = \lambda$ then $w \rightarrow u$ is a singleton iff $u = \lambda$. Therefore we will deal in the following only with the case where u and w are nonempty words.

Theorem 1. *If w, u are words in Σ^+ and u is a power of an unbordered word $g \in \Sigma^+$, $u = g^i$, $i \geq 1$, then the statements (a) and (b) are equivalent:*

(a) The set $w \to u$ is a singleton;
(b) (1) The word w is of the form $w = \alpha g^j \beta$, $j \geq i$, $\alpha, \beta \in \Sigma^*$,
(2) The number j is maximal with this property, i.e., α does not contain g as a suffix and β does not contain g as a prefix,
(3) Neither α nor β contains u as a subword.

Proof. $(a) \Longrightarrow (b)$ Let us assume that $w, u \in \Sigma^+$ as in the theorem.If $w \to u$ is a singleton, for any two decompositions of w as $w = xuy = euf$ with $x, y, e, f \in \Sigma^*$, we have $xy = ef$. Let us choose x, y, e, f in such a way that the two occurrences of u are the rightmost and the leftmost one. Consider now all the possible relative positions of x, y and e, f.

• If $\lg(eu) \leq \lg(x)$ then:

$$w = eu\underbrace{x_2uy}_{f} = \underbrace{eux_2}_{x}uy,\ x_2 \in \Sigma^* \ .$$

The equality $xy = ef$ implies in this case that $eux_2y = ex_2uy$ that is, $ux_2 = x_2u$. According to Lemma 1, x_2 and u are powers of the same primitive word. As $u = g^i$, g primitive, we deduce that $x_2 = g^k$, $k \geq 0$. The word w can be then written as

$$w = eg^ig^kg^iy = eg^{k+2i}y \ .$$

Taking now $\alpha = e$ and $\beta = y$, $(b)(1)$ holds. Our choice of x, y, e, f guarantees that also $(b)(2)$ and $(b)(3)$ hold.

• If $\lg(e) < \lg(x) < \lg(eu)$ then:

$$w = e\underbrace{u_1u_2}_{u}\underbrace{u_3y}_{f} = \underbrace{eu_1}_{x}\underbrace{u_2u_3}_{u}y,\ u_1, u_2, u_3, \in \Sigma^+ \ .$$

The equality $xy = ef$ implies $eu_1y = eu_3y$, that is, $u_1 = u_3$. As $u = u_1u_2 = u_2u_1$, according to Lemma 1 we obtain that u_1 and u_2 are powers of the same primitive word, which is g. Therefore $u_1 = g^k$, $u_2 = g^{i-k}$, $k > 0$, which implies:

$$w = eu_1u_2u_1y = eg^{i+k}y, k > 0 \ .$$

Taking now $\alpha = e$ and $\beta = y$, $(b)(1)$ holds. Our choice of x, y, e, f implies also $(b)(2)$ and $(b)(3)$.

• If $\lg(e) = \lg(x)$ then there is only one occurrence of the word u in w and (b) obviously holds.

For $(b) \Longrightarrow (a)$ assume that $w, u \in \Sigma^+$ are as in the theorem and that (b) holds. As $j \geq i$ there exists a $k \geq 0$ such that $j = i + k$. Argue indirectly and assume that $w \to u$ is not a singleton, that is, there exists a word in $w \to u$ which is differs from $\alpha g^k \beta$.

Remark. Because u is a power of the unbordered word g, two occurrences of u can overlap only on powers of g.

We shall consider in the following all the possible cases $w = \alpha u g^k \beta = xuy$, $x, y \in \Sigma^*$, which can lead to the situation that $xy \neq \alpha g^k \beta$.

• If $\lg(\alpha u) \leq \lg(x)$ then

$$w = \underbrace{\alpha u x_1}_{x}\, u\, \underbrace{y_1 \beta}_{y} = \alpha u\, \underbrace{x_1 u y_1}_{g^k} \beta,\ x_1, y_1 \in \Sigma^* \ .$$

Note that u cannot overlap with α or β because g is unbordered and $(b)(2)$, $(b)(3)$ hold.

As we have assumed that $xy \neq \alpha g^k \beta$ we have that $\alpha u x_1 y_1 \beta \neq \alpha g^k \beta$ which implies that $g^i x_1 y_1 \neq g^k$. As $g^k = x_1 g^i y_1$, this is a contradiction. Our assumption was false, therefore this case cannot hold.

• If $\lg(\alpha) < \lg(x) < \lg(\alpha u)$ then:

$$w = \underbrace{\alpha x_1}_{x}\, \underbrace{u_1 u_2}_{u}\, \underbrace{y_1 \beta}_{y} = \alpha\, \underbrace{x_1 u_1}_{u}\, \underbrace{u_2 y_1}_{g^k} \beta,\ x_1, u_1 \in \Sigma^+, u_2, y_1 \in \Sigma^* \ .$$

As $u = x_1 u_1 = u_1 u_2 = g^i$ and g is an unbordered word we have that $u_1 = g^{i_1}$, $u_2 = g^{i_2} = x_1$, $i_1, i_2 > 0$. The fact that $xy \neq \alpha g^k \beta$ implies $\alpha x_1 g^{k-i_2} \beta \neq \alpha g^k \beta$ that is, $x_1 g^{k-i_2} \neq g^k$. As we have shown that $x_1 = g^{i_2}$, this is a contradiction. Our assumption was false, therefore this case cannot hold either.

As all the possible cases led to contradictions, our assumption that $w \to u$ is not a singleton is false. The proof of the second implication, and therefore of the theorem, is complete. □

The proof of the implication $(a) \Longrightarrow (b)$ did not use the fact that g is an unbordered word.

The reverse implication does not hold if g is not unbordered. For example, if $w = ababa$ and $u = aba$, taking $\alpha = ab$, $\beta = \lambda$, $g = aba$, the condition (b) is satisfied. However the set $w \to u = \{ab, ba\}$ is not singleton. A stronger condition than (b) is needed to assure that $w \to u$ is a singleton, if u is a power of a primitive bordered word.

Theorem 2. *Let w, u be words in Σ^+. If u is a power of a primitive bordered word $g \in \Sigma^+$, $u = g^i$, $i \geq 1$, then the statements (a) and (b) are equivalent:*

(a) The set $w \to u$ is a singleton.

(b) (1) The word w is of the form $w = \alpha g^j \beta$, $j \geq i$, $\alpha, \beta \in \Sigma^*$,

(2) The number j is maximal with this property, i.e., α does not contain g as a suffix and β does not contain g as a prefix,

(3) Neither α nor β contains u as a subword,

(4) For any decomposition of g, $g = xy = yx'$ where $x, y, x' \in \Sigma^+$ we have: $\alpha \neq \alpha' g^{i-1} x$, $\forall \alpha' \in \Sigma^*$ and $\beta \neq x' g^{i-1} \beta'$, $\forall \beta' \in \Sigma^*$.

Proof. $(a) \Longrightarrow (b)$ Let w, u be as in the theorem.

If $w \to u$ is a singleton, using the proof of Theorem 1 and the remark following it, $(b)(1)$, $(b)(2)$ and $(b)(3)$ hold. Therefore w is of the form $w = \alpha g^j \beta$, $j \geq i$. As $j \geq i$ there exists $k \geq 0$ such that $j = i + k$.

Argue indirectly and assume that $(b)(4)$ does not hold. This means that one of the following cases holds:

• $\alpha = \alpha' g^{i-1} x$ where $\alpha' \in \Sigma^*$, $x \in \Sigma^+$ and there exists $y, x' \in \Sigma^+$ such that $g = xy = yx'$,

• $\beta = x' g^{i-1} \beta'$ where $\beta' \in \Sigma^*$, $x' \in \Sigma^+$ and there exist $y, x \in \Sigma^+$ such that $g = xy = yx'$.

We shall consider the first case, the other one being symmetric. The word w can be written as:

$$w = \alpha' g^{i-1} x g^{i+k} \beta = \alpha' g^{i-1} x (yx')^{i+k} \beta = \alpha' \underbrace{g^{i-1} xy}_{g^i = u} x' g^{i+k-1} \beta \ .$$

As $w \to u$ is a singleton the words $\alpha' x' g^{i+k-1} \beta$ and $\alpha' g^{i-1} x g^k \beta$ are equal. This equality leads to the following chain of implications:

$$x' g^{i+k-1} = g^{i-1} x g^k \Longrightarrow x' g^{i-1} = g^{i-1} x \Longrightarrow$$

$$x' \underbrace{yx' \ldots yx'}_{i-1} = \underbrace{xy \ldots xy}_{i-1} x \text{ and, as } \lg(x') = \lg(x), \Longrightarrow$$

$$x = x' \Longrightarrow g = xy = yx$$

According to Lemma 2, the last equality implies that either x or y equals λ. This contradicts our assumption that $x, y \in \Sigma^+$.

All the possible cases led to contradiction and therefore our assumption that $(b)(4)$ does not hold was false.

For the implication $(b) \Longrightarrow (a)$, let w, u be words in Σ^+, satisfying (b). Therefore u and w are nonempty words, $w = \alpha g^j \beta$, $u = g^i$, $j \geq i \geq 1$, ($g \in \Sigma^+$ primitive and bordered) such that (b) holds. As $j \geq i$ there exists $k \geq 0$ such that $j = i + k$.

From (b) it follows that an arbitrary occurrence of u in w overlaps with neither α nor β. Assume, for example, that u overlaps with α. Then we have:

$$w = \underbrace{\alpha_1 u_1}_{\alpha} \underbrace{u_2 u_3}_{u} g^k \beta, \alpha_1 \in \Sigma^*, u_1, u_2, u_3 \in \Sigma^+, u = u_1 u_2 = u_2 u_3 \ .$$

As $u = u_1 u_2 = u_2 u_3$, if any of u_i, $i = 1, 2, 3$ would be a power of g then u_1 would equal u_3. This, in turn, would imply that α contains g as a suffix – a contradiction with $(b)(2)$. Therefore we can assume that none of u_i, $i = 1, 2, 3$ is a power of g and we have:

$$u = \underbrace{g^q g_1}_{u_1} \underbrace{g_2 g^p}_{u_2} = \underbrace{g_2 g^p}_{u_2} u_3, q + p + 1 = i, g_1, g_2 \in \Sigma^+, g = g_1 g_2 \ .$$

If $p > 0$ and $q = 0$ then the preceding equality implies:

$$u = g_1 g_2 (g_1 g_2)^p = g_2 (g_1 g_2)^p u_3 \ ,$$

and as $\lg(g_1g_2)$ =$\lg(g_2g_1)$ we conclude that $g = g_1g_2 = g_2g_1$. According to Lemma 2 this implies $g_1 = \lambda$ or $g_2 = \lambda$– a contradiction with our assumption $g_1, g_2 \in \Sigma^+$.

If $p > 0$ and $q > 0$ then:

$$u = \underbrace{g_1g_2 \dots g_1g_2}_{q \text{ times}} g_1g_2(g_1g_2)^p = g_2(g_1g_2)^p u_3 \ ,$$

which implies that $g_1g_2 = g_2g_1$ and leads to the same contradiction.

If $p = 0$ then $\underbrace{g^q g_1}_{u_1}\underbrace{g_2}_{u_2} = g_2u_3$. As $q = i - 1$ we obtain that $\alpha = \alpha_1 g^{i-1} g_1$ where $g = g_1g_2 = g_2u_3$, and $g_1, g_2, u_3 \in \Sigma^+$, which contradicts $(b)(4)$.

As all cases led to contradictions, our assumption that an occurrence of u in w can overlap with α was false. Similarly we can prove that no occurrence of u in w overlaps with β.

As u can overlap with neither α nor β, an occurrence of u in w can appear only in the "g-part" of w. This means that an arbitrary occurrence of u in w can have only one of the following locations:

- $w = \alpha g^{k_1-1} g_1 \underbrace{g_2 g^{i-1} g_1}_{u} g_2 g^{k_2} \beta,\ g_1, g_2 \in \Sigma^+, g_1g_2 = g \ ,$

where $k > 0$ and $k_1 + k_2 = k$. We have assumed here that $k_1 > 0$ and $k_2 \geq 0$. The case when $k_2 > 0$ and $k_1 \geq 0$ is similar.

As $u = g^i = (g_1g_2)^i = g_2(g_1g_2)^{i-1}g_1$ we deduce that $g = g_1g_2 = g_2g_1$ which, together with Lemma 2, leads to a contradiction with our assumption that g_1, $g_2 \in \Sigma^+$. We conclude that such a situation cannot occur.

- $w = \alpha g^{k_1} \underbrace{g^i}_{u} g^{k_2}, k \geq 0, k_1 + k_2 = k \ .$

In this situation, the erasing of u from w produces the word $\alpha g^k \beta$, regardless of the values of k_1 and k_2, $k_1 + k_2 = k$.

We conclude that the only possible occurrence of u in w yields $w \rightarrow u = \{\alpha g^k \beta\}$. Therefore $w \rightarrow u$ is a singleton, that is, (a) holds. This completes the proof of the second implication, and therefore of the theorem. □

The following theorem gives a necessary and sufficient condition under which the result of the insertion between two words is a singleton set.

Theorem 3. *Let u, w be words in Σ^*. The set $w \leftarrow u$ is a singleton iff one of the following cases holds:*

(i) The words w, u have the forms $w = a^p$, $u = a^i$, $a \in \Sigma$, $p, i > 0$;
(ii) Either u or w (or both) is equal with λ.

Proof. The "if"-part is obvious. For the "only if"-part let u, w be in Σ^+ such that $w \leftarrow u$ is a singleton. We will show that in this case (i) holds. The fact that $w \leftarrow u$ is a singleton implies that for any decomposition of w as $w = xy$, $x, y \in \Sigma^*$ we have that $xuy = uxy = xyu$, all being elements of the set $w \leftarrow u$. From the equality $xuy = uxy$ and using Lemma 1, we deduce that x and u are

powers of the same primitive word, $x = g^j$, $u = g^i$, $g \in \Sigma^+$, $j \geq 0$, $i > 0$. Analogously, from $xuy = xyu$ we deduce that $y = g^k$, $k \geq 0$, being a power of the same primitive word as u. As x, y were arbitrary words with the property $xy = w$, taking for example x the first letter of w we conclude that u is of the form $u = a^i$, $a \in \Sigma$, $i > 0$ and w is of the form $w = xy = a^{j+k}$, $j \geq 0$, $k \geq 0$, $j + k > 0$. Taking $p = j + k$ the proof of the "only if"-part is complete. □

4 Conditions for Reversibility

The catenation and the right and left quotient of words possess the property that given the result of the operation and one of the operands, the other operand can be recovered. Indeed, if x, y, z are words in Σ^* then $xy = z$ iff $x = z/y$ iff $y = x\backslash z$. The insertion and deletion of words do not have this property. In general, if $x \leftarrow y = z$ then $\{x\} \subseteq z \rightarrow y$ and if $x \rightarrow y = z$ then $\{x\} \subseteq z \leftarrow y$, but the reverse inclusions do not hold. The following theorems will deal with circumstances under which, given the result of the insertion and the inserted word, the other operand can be obtained. The problem can be stated shortly : "When is $(w \leftarrow u) \rightarrow u$ equal with $\{w\}$?", where $u, w \in \Sigma^*$. Besides the fact that u is a power of a primitive bordered or unbordered word, the answer to this problem is influenced by whether or not u is a subword of w.

Note. If $u = \lambda$ or $w = \lambda$ then $(w \leftarrow u) \rightarrow u = \{w\}$. Therefore we will consider in the following only the case where u and w are nonempty words.

Theorem 4. *Let u, w be two words in Σ^+ such that u is not a subword of w. If u is a power of an unbordered word then $(w \leftarrow u) \rightarrow u = \{w\}$.*

Proof. Let u, w be as in the theorem, such that $u = g^i$, $g \in \Sigma^+$, $i \geq 1$, and g is an unbordered word. Let xuy be an arbitrary word in $(w \leftarrow u)$, where $x, y \in \Sigma^*$, $w = xy$.

If the only occurrence of u in xuy is the one inserted, then $xuy \rightarrow u = xy = \{w\}$.

Else, the second occurrence of u must overlap the first, as we have assumed that u is not a subword of w. Moreover, because u is a power of an unbordered word g, they must overlap on powers of g. Under these circumstances, the erasing of the second ocurrence of u from xuy produces also w.

In all the possible cases the erasing of an occurrence of u from an arbitrary word of $(w \leftarrow u)$ produced w, and therefore we can conclude that $(w \leftarrow u) \rightarrow u = \{w\}$. □

The reverse implication does not hold. For example, taking $w = cd$, $u = aba$, we have that u is not a subword of w and that $(w \leftarrow u) \rightarrow u = \{w\}$ but u is not a power of an unbordered word.

Theorem 5. *Let u, w be words in Σ^+ such that u is not a subword of w. If u is a power of a primitive bordered word $g \in \Sigma^+$, $u = g^i$, $i \geq 1$ then the following statements are equivalent:*

(i) *The set* $(w \leftarrow u) \rightarrow u$ *is a singleton, namely* $\{w\}$.
(ii) *For any decomposition of* g, $g = xy = yx'$, $x, y, x' \in \Sigma^+$, *the word* w *contains neither* $g^{i-1}x$ *nor* $x'g^{i-1}$ *as a subword.*

Proof. We will prove first that $\neg(ii) \Longrightarrow \neg(i)$. Let u, w be as in the theorem such that (ii) does not hold. There exists a decomposition of g, $g = xy = yx'$ where $x, y, x' \in \Sigma^+$ such that $w = \alpha g^{i-1} x \beta$, $\alpha, \beta \in \Sigma^*$. The case where w is of the form $w = \alpha x' g^{i-1} \beta$ is symmetric. The word

$$\alpha g^{i-1} x g^i \beta = \alpha \underbrace{g^{i-1} xy}_{u} x' (yx')^{i-1} \beta$$

belongs to $w \leftarrow u$ and therefore both words $\alpha g^{i-1} x \beta$ and $\alpha x' g^{i-1} \beta$ are in the set $(w \leftarrow u) \rightarrow u$.

If we assume that $(w \leftarrow u) \rightarrow u$ is a singleton, we obtain $g^{i-1}x = x'g^{i-1}$ which implies

$$\underbrace{xyxy \ldots xy}_{(i-1) \text{ times}} x = x' \underbrace{yx' \ldots yx'}_{(i-1) \text{ times}} \ .$$

The last equality shows that $x = x'$, which implies $g = xy = yx$. According to Lemma 2 either x or y equals λ, which contradicts our assumption $x, y \in \Sigma^+$. Consequently, we conclude that $(w \leftarrow u) \rightarrow u$ is not a singleton.

For $(ii) \Longrightarrow (i)$, let w, u be words as in the theorem such that (ii) holds. Assume that $(w \leftarrow u) \rightarrow u$ is not a singleton. This means that there exist a word $\alpha \in (w \leftarrow u)$ and two occurrences of u in α, which produce different words after being erased. As u is not a subword of w, these two occurrences of u must overlap:

$$\alpha = xuy = euf = \underbrace{eu_1}_{x} \underbrace{u_2 u_3}_{u} y = e \underbrace{u_1 u_2}_{u} \underbrace{u_3 y}_{f} \in (w \leftarrow u) \ ,$$

where $x, y, e, f \in \Sigma^*$, $u_1, u_2, u_3 \in \Sigma^+$ and $w = xy$, $xy \neq ef$. The words $xy = eu_1 y$, $ef = eu_3 y$ are not equal and therefore $u_1 \neq u_3$.

If any of u_i, $i = 1, 2, 3$, is a power of g, as $u = u_1 u_2 = u_2 u_3$, we obtain $u_1 = u_3$ – a contradiction. We will assume therefore that none of u_i, $i = 1, 2, 3$, is a power of g. This implies:

$$\underbrace{g^k g_1}_{u_1} \underbrace{g_2 g^p}_{u_2} = \underbrace{g_2 g^p}_{u_2} u_3, g_1, g_2 \in \Sigma^+, p + k + 1 = i \ .$$

If $p > 0$ we obtain that $g_1 g_2 = g_2 g_1$ which, together with Lemma 2, implies $g_1 = \lambda$ or $g_2 = \lambda$. This contradicts the fact that $g_1, g_2 \in \Sigma^+$.

If $p = 0$ then:

$$\underbrace{g^k g_1}_{u_1} g_2 = g_2 u_3 = u \ ,$$

and, as $w = xy = eu_1 y = eg^k g_1 y$ and $k = i - 1$, this implies that w contains a subword of the form $g^{i-1} g_1$ with $g = g_1 g_2 = g_2 u_4$, $g_1, g_2, u_4 \in \Sigma^+$. We have

arrived at a contradiction with (ii). All cases led to contradictions and therefore our assumption that $(w \leftarrow u) \rightarrow u$ is not a singleton was false.

The proof of the second implication and consequently, of the theorem, is now complete. □

Theorem 6. *Let u, w be words in Σ^+, u a proper subword of w. Then $(w \leftarrow u) \rightarrow u = \{w\}$ iff $w = a^p$, $u = a^i$, $a \in \Sigma$, $p > i > 0$.*

Proof. The implication " $\Longleftarrow$ " is obvious. In order to show the reverse implication, let u, w be words in Σ^+ where u a is subword of w (not necessarily proper) and $(w \leftarrow u) \rightarrow u = \{w\}$. The word w can be expressed as $w = xuy$, $x, y \in \Sigma^*$, $u \in \Sigma^+$. This implies that both words $u(xuy)$ and $(xuy)u$ belong to $w \leftarrow u$ and therefore:

$$xuy, uxy, xyu \ \in (w \leftarrow u) \rightarrow u = \{w\} \ .$$

From the equality $xuy = uxy$ we deduce $xu = ux$. According to Lemma 1, x and u are powers of the same primitive word, $u = g^i$, $x = g^k$, $g \in \Sigma^+$, $k \geq 0$, $i \geq 1$.

From the equality $xuy = xyu$ we deduce $uy = yu$. According to Lemma 1, y and u are powers of the same primitive word g that is, $y = g^j$, $j \geq 0$.

The primitive word g is unbordered. Indeed, assume that g is bordered. Then, according to Lemma 3, g can be written as $g = \gamma v \gamma$, $\gamma, v \in \Sigma^+$. As $u = (\gamma v \gamma)^i$ and $w = (\gamma v \gamma)^{k+i+j}$ we deduce that both words:

$$(\gamma v \gamma)^{k+2i+j}, \text{ and } \gamma(\gamma v \gamma)^i v \gamma (\gamma v \gamma)^{k+i+j-1} = \gamma(\gamma v \gamma)^{i-1} \gamma v (\gamma v \gamma)^i (\gamma v \gamma)^{k+j} \ ,$$

are in the set $w \leftarrow u$ (the first word was obtained by catenating w and u and the second by inserting u after the first occurrence of γ.) This implies that both words:

$$(\gamma v \gamma)^{k+i+j} \text{ and } \gamma(\gamma v \gamma)^{i-1} \gamma v (\gamma v \gamma)^{k+j} \ ,$$

belong to $(w \leftarrow u) \rightarrow u$, which is a singleton. The equality of the above mentioned words implies the equality of their prefixes $\gamma v \gamma = \gamma \gamma v$ which further implies $v\gamma = \gamma v$. According to Lemma 1, γ and v are powers of the same primitive word, $\gamma = \delta^r$, $v = \delta^{r'}$, $\delta \in \Sigma^+$, $r, r' > 0$. We can rewrite g now as $g = \gamma v \gamma = \delta^{2r+r'}$, $2r+r' > 2$, which contradicts the fact that g is primitive. Our assumption was false, therefore g is an unbordered word.

Taking $p = i + j + k$ we have therefore proved that if u is a subword of w (proper or not) and $(w \leftarrow u) \rightarrow u = \{w\}$ then $u = g^i$, $w = g^p$, $g \in \Sigma^+$, $p \geq i > 0$, where g is an unbordered word.

Assume now that $u \neq \lambda$ is a proper subword of w and denote $k' = j + k$, $k' > 0$. Argue indirectly and assume that g contains at least two different letters, $g = a\alpha b\beta$, $a, b \in \Sigma$, $a \neq b$,$\alpha, \beta \in \Sigma^*$. Then both words:

$$(a\alpha b\beta)^{2i+k'} \text{ and } a\alpha(a\alpha b\beta)^i b\beta(a\alpha b\beta)^{i+k'-1} \ ,$$

are in the set $w \leftarrow u$ which implies that both:

$$(a\alpha b\beta)^{i+k'} \text{ and } a\alpha(a\alpha b\beta)^i b\beta(a\alpha b\beta)^{k'-1}$$

belong to $(w \leftarrow u) \rightarrow u$. Indeed, as $k' \geq 1$ we have another occurrence of u in $w \leftarrow u$ than the one inserted, namely the prefix of length $\lg(u)$ of $(a\alpha b\beta)^{i+k'-1}$. As $(w \leftarrow u) \rightarrow u$ is a singleton, the two words belonging to it are equal that is,

$$a\alpha b\beta(a\alpha b\beta)^{i+k'-1} = a\alpha(a\alpha b\beta)^i b\beta(a\alpha b\beta)^{k'-1} \ .$$

We arrive at a contradiction as, after erasing the prefix $a\alpha$, the above equality implies $a = b$ and we assumed that the letters a and b are distinct. Our assumption that g contains at least two different letters was false. As g is also primitive and unbordered we deduce that g is of the form $g = a$, $a \in \Sigma$ and consequently, $w = a^{i+k'}$, $i \geq 1$, $k' > 0$.

Taking $p = i + k'$, the proof of the second implication is complete. □

Theorem 7. *If u is a word in Σ^+ then $(u \leftarrow u) \rightarrow u = \{u\}$ iff u is a power of an unbordered word.*

Proof. It has been shown in the proof of Theorem 6 that, if $u, w \in \Sigma^+$ and u is a subword of w (proper or not) then $(w \leftarrow u) \rightarrow u = \{w\}$ implies $u = g^i$, $w = g^p$, $p \geq i > 0$, where $g \in \Sigma^+$ is an unbordered word. Taking $u = w$, this proves the implication "$\Longrightarrow$" of the theorem.

For the reverse implication let $u \in \Sigma^+$ be a power of an unbordered word $g \in \Sigma^+$, $u = g^i$, $i \geq 1$.

Assume that there exists $w \in (u \leftarrow u) \rightarrow u$, $w \neq u$. Applying the operations in the reverse order, we deduce that $u \in (w \leftarrow u) \rightarrow u$. As $w \neq u$ but $\lg(w) = \lg(u)$, u is not a subword of w. According to Theorem 4 we have $(w \leftarrow u) \rightarrow u = \{w\}$, which implies $w = u$. This contradicts our assumption that $w \neq u$. Consequently, we can conclude that the set $(u \leftarrow u) \rightarrow u = \{u\}$, and therefore the proof for the second implication is complete. □

The last theorem of this section gives a necessary and sufficient condition under which the set $(w \rightarrow u) \leftarrow u$ is a singleton.

Theorem 8. *If u, w are words in Σ^* then $(w \rightarrow u) \leftarrow u = \{w\}$ iff one of the next cases holds:*

(i) The word w is equal with u;
(ii) The word u equals λ;
(iii) The words w, u are of the form $w = a^p$, $u = a^i$, $a \in \Sigma$, $p > i \geq 1$.

Proof. The implication " $\Longleftarrow$ " is obvious. For the reverse implication, assume that $w, u \in \Sigma^*$, such that $(w \rightarrow u) \leftarrow u = \{w\}$ and $w \neq u$, $u \neq \lambda$. We will show that in this case (iii) holds.

As u is a proper subword of w we can assume, without loss of generality, that it is a suffix of w, that is, $w = a\alpha u$, $a \in \Sigma$, $\alpha \in \Sigma^*$. The word $a\alpha$ is in the set $w \rightarrow u$ therefore both $au\alpha$ and $ua\alpha$ belong to $(w \rightarrow u) \leftarrow u = \{w\}$. The equality $au\alpha = ua\alpha$ implies that $u = a^i$, $i > 0$. The equality $a\alpha u = au\alpha$ implies that $w = a^p$, $p > 1$. As u is a proper subword of w, $p > i \geq 1$, and the proof of the second implication is complete. □

5 Open Problems

As problems for further research we could mention Theorem 4 which should be replaced with an "if and only if" condition. From the practical point of view, it would be also preferable to have more compact and easily testable necessary and sufficient conditions for the problems investigated in Sections 3 and 4. Interesting and cryptographically motivated seems to be the study of analogous problems with more sophisticated types of insertion and deletion. For example, the parallel, controlled, permuted and permuted scattered variants of insertion and deletion defined in [2], [6], could be of interest.

References

1. Andraşiu, M., Dassow, J., Păun, Gh., Salomaa, A.: Language-theoretical problems arising from Richelieu cryptosystems (to appear)
2. Kari, L.: On insertion and deletion in formal languages. Ph.D. Thesis University of Turku (1991)
3. Lyndon, R.C., Schutzenberger, M.P.: The equation $a^M = b^N c^P$ in a free group. Michigan Math.J. **9**(1962) 289-298
4. Păun, Gh., Salomaa, A.: Semi-commutativity sets – a cryptographically grounded topic. Bull.Math.Soc.Sci.Math.Roumanie (to appear)
5. Salomaa, A.: Public-key Cryptography. Springer Verlag Berlin (1990)
6. Sântean, L.: Six arithmetic-like operations on languages. Revue Roumaine de Linguistique Tome XXXIII (1988), Cahiers de linguistique theorique et applique Tome XXV (1988) **1** Janvier-Juin 65-73
7. Shyr, H.J.: Free Monoids and Languages. Lecture Notes, Institute of applied mathematics, National Chung-Hsing University Taichung Taiwan (1991)
8. Shyr, H.J., Thierrin, G., Yu, S.S.: Monogenic e-closed languages and dipolar words (to appear)