Theoretical Computer Science 245 (2000) 115–133

Theoretical Computer Science

www.elsevier.com/locate/tcs

Shuffle and scattered deletion closure of languages ☆

Masami Ito [a,*], Lila Kari [b], Gabriel Thierrin [b]

[a] *Faculty of Science, Kyoto Sangyo University, Kamigamo-Motoyama, Kita-ku, Kyoto 603, Japan*
[b] *Department of Computer Science, University of Western Ontario, London, Ont., Canada N6A 5B7*

Abstract

We introduce and study the notion of *shuffle residual* of a language L: the set containing the words whose shuffle with L is completely included in L. Several properties and a characterization of the shuffle residual of a language are obtained. The *shuffle closure* of a language L (the smallest language that is shuffle closed and contains L) is investigated. Moreover, conditions for the existence of maximal languages whose shuffle residual equals a given language are obtained. The paper also considers an operation dual to shuffle, namely *scattered deletion*: the scattered deletion of a word w from u consists of the words obtained by sparsely deleting from u the letters of w, in the order in which they appear in w. The *scattered deletion residual* and *scattered deletion closure* of a language are defined and studied. Finally, relationships and interdependencies between shuffle, scattered deletion, and other insertion and deletion operations are obtained. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Shuffle operation; Shuffle closure; Shuffle residual; Maximal shuffle residual; RSS-type laguage; Insertion and deletion operations

1. Introduction

The shuffle operation, being in some sense a mathematical model of parallel computation, has been intensively studied in formal language theory. For example, some types of regular expressions of shuffle operators are dealt with in [1, 2, 13–17]. A related decidability problem for commutative regular languages is solved in [8]. A constrained form of shuffle product, namely the literal shuffle is defined in [3], while a special kind of literal shuffle product of a language is studied in [9]. A relation between shuffle closed languages and automata is studied in [7]. Shuffle operations on partial ordered

☆ Research supported by Grant OGP0007877 of the Natural Sciences and Engineering Research Council of Canada and Grant-in-Aid for Science Research 10440034 of the Ministry of Education, Science and Culture of Japan.

* Corresponding author.

E-mail addresses: ito@ksuvx0.kyoto-su.ac.jp (M. Ito), lila@csd.uwo.ca (L. Kari), gab@csd.uwo.ca (G. Thierrin).

PII: S0304-3975(99)00277-7

sets can be found in [4, 5]. A systematic study of insertion operations, which are related to the shuffle operation, is contained in [18], and a continuation of this line of research can be found in [11].

This paper introduces the notion of shuffle residual of a language L as consisting of the words whose shuffle with words in L is completely included in L. Properties and characterizations of the shuffle residual of a language are obtained. Moreover, the shuffle closure of a language L, which is the smallest shuffle closed language that contains L, is characterized. Finally, conditions for the existence of maximal languages whose shuffle residual equals a given language are obtained. In addition, the paper addresses similar issues related to a dual notion of shuffle, namely scattered deletion [18]. Relations between shuffle, scattered deletion, and other insertion and deletion operations like insertion, deletion and dipolar deletion are also obtained.

The paper is organized as follows. The end of this section contains some basic formal language definitions and notations. In Section 2 the notion of shuffle residual of a language is defined. Some properties of the shuffle residual of a language are obtained, as well as a characterization of the shuffle residual of a given language L. The second notion to be considered in the section is the shuffle closure of a language, introduced in [12]. A characterization of the shuffle closure of a given language is obtained. The shuffle closure of singleton sets is also considered.

Section 3 addresses the issue of the maximal language whose shuffle residual equals a given set L. Several conditions for the existence of such languages are obtained. Finally, a generalization of the notion of shuffle residual is introduced.

Section 4 investigates issues similar to those of Section 2, but this time for an operation that is, in some sense, "inverse" to the shuffle operation: the scattered deletion operation. (The scattered deletion of a word w from u consists of sparsely deleting from u the letters of w, in the order in which they appear in w.)

Finally, Section 5 studies relations and interdependencies between shuffle, scattered deletion and other insertion and deletion operations like insertion, deletion and dipolar deletion. A property of shuffle-base of a language is also given.

In the following, an alphabet X is a finite nonempty set. The cardinality of X, i.e. the number of letters in X, is denoted by $|X|$. Let X^* be the free monoid generated by X under the catenation operation, and let $X^+ = X^* \setminus \{1\}$, where 1 denotes the empty word of X^*. For the sake of simplicity, if $X = \{a\}$ then we write a^+ and a^* instead of $\{a\}^+$ and $\{a\}^*$. If $L \subseteq X^*$ then L^+ denotes the set of all possible catenations of words in L, and $L^* = L^+ \cup \{1\}$. In particular, if $L = \{w\}$, then we write w^+ and w^* instead of $\{w\}^+$ and $\{w\}^*$, respectively. If $u \in X^*$, then $|u|$ denotes the length of u, that is, the number of letters in u. Moreover, if $a \in X$, then the number of occurrences of the letter a in the word u is denoted by $|u|_a$. Let $L \subseteq X^*$. By $alph(L)$ we denote the alphabet of L, i.e. $a \in alph(L)$ if and only if a occurs in at least one word in L.

Let X be an alphabet and u, v be two words in X^*. The *shuffle product* of u and v is denoted by $u \diamond v$ and is defined by

$$u \diamond v = \{u_1 v_1 u_2 v_2 \ldots u_k v_k \mid u = u_1 u_2 \ldots u_k, v = v_1 v_2 \ldots v_k, k \geqslant 1, u_i, v_i \in X^*, 1 \leqslant i \leqslant k\}.$$

Obviously, $u \diamond v = v \diamond u$ and $(u \diamond v) \diamond w = u \diamond (v \diamond w)$ hold for all $u, v, w \in X^*$.

Now, let $A, B \subseteq X^*$. By the shuffle product of A and B we mean

$$A \diamond B = \bigcup_{u \in A,\, v \in B} (u \diamond v).$$

It is easy to see that $A \diamond B = B \diamond A$ and that $A \diamond (B \diamond C) = (A \diamond B) \diamond C$.

For further definitions and notations in formal language theory and theory of codes the reader is referred to [6, 19, 20], respectively.

2. Shuffle closure

Let $L \subseteq X^*$. To the language L we associate a set called the *shuffle-residual* of L, which consists of all words x with the following property: if $u \in L$, the result of the shuffle $u \diamond x$ is included in L. Formally, the shuffle-residual of L is denoted by $shRes(L)$ and is defined by

$$shRes(L) = \{x \in X^* \mid \forall u \in L, u \diamond x \subseteq L\}.$$

Example. Let $X = \{a, b\}$. Then,
- $shRes(X^*) = X^*$;
- $shRes(L_{ab}) = L_{ab}$, where $L_{ab} = \{x \in X^* \| x|_a = |x|_b\}$;
- if $L = \{a^n b^n \mid n \geqslant 0\}$ then $shRes(L) = \{1\}$;
- if $L_1 = (a^2)^*$, $L_2 = aL_1$ then $shRes(L_1) = L_1$ and $shRes(L_2) = L_1$;
- if $L = b^*ab^*$ then $shRes(L) = b^* = shRes^2(L)$;
- if $L = aX^*b$ then $shRes(L) = L$.

The following results give some basic properties of the shuffle residual of a language.

Proposition 2.1. *$shRes(M) \diamond shRes(N) \subseteq shRes(M \diamond N)$ and $shRes(M) \cap shRes(N) \subseteq shRes(M \cup N)$.*

Proof. Let $u \in shRes(M) \diamond shRes(N)$. This means that there exist $m \in shRes(M)$ and $n \in shRes(N)$ such that $u \in m \diamond n$. We have that

$$(M \diamond N) \diamond (m \diamond n) = (M \diamond m) \diamond (N \diamond n) \subseteq M \diamond N.$$

Note that the equality does not always hold. For example, let $X = \{a, b\}$, $M = ab^*$ and $N = ba^*$. Then, $shRes(M) = \{1\} = shRes(N)$. On the other hand, $M \diamond N = \{u \in X^* \mid |u|_a \geqslant 1, |u|_b \geqslant 1\}$, therefore $shRes(M \diamond N) = X^*$.

For the second inclusion, let $u \in shRes(M) \cap shRes(N)$. The fact that $u \in shRes(M)$ implies that $M \diamond u \subseteq M$. The fact that $u \in shRes(N)$ implies that $N \diamond u \subseteq N$. Consequently, we have $(M \cup N) \diamond u \subseteq M \cup N$. Note that the equality does not always hold. For example, if $M = (X^2)^*$ and $N = X(X^2)^*$ then $shRes(M) = shRes(N) = M$, but $shRes(M \cup N) = X^*$. $\square$

A language L is *commutative* if for any $w \in L$, L contains all the words obtained from w by arbitrarily permuting its letters. For a word $u = a_1 a_2 \dots a_k \in X^*$, $k \geqslant 0$ we define

$$com(u) = \{a_{s(1)} a_{s(2)} \dots a_{s(k)} \mid s \text{ a permutation of } \{1, \dots, k\}\}.$$

that is, $com(u)$ contains all the words obtained by arbitrarily permuting the letters of u. If $L \subseteq X^*$ then

$$com(L) = \bigcup_{u \in L} com(u).$$

Proposition 2.2. *$shRes(L)$ is a submonoid of X^* that is moreover closed under shuffle. If L is a commutative language, then $shRes(L)$ is also a commutative language.*

Proof. Let $x, y \in shRes(L)$ and $u \in L$. Then $u \diamond x \subseteq L$ and consequently $(u \diamond x) \diamond y \subseteq L$. As shuffle is associative, we have that $u \diamond (x \diamond y) \subseteq L$, that is, $x \diamond y \subseteq shRes(L)$. This implies the closure of $shRes(L)$ under shuffle. In particular, $xy \in x \diamond y$ belongs to $shRes(L)$. Since $1 \in shRes(L)$, $shRes(L)$ is not empty. For the second claim, let $x \in shRes(L)$. We have that $u \diamond x \subseteq L$. As L is commutative, $com(u \diamond x) \subseteq L$. In particular, $u \diamond com(x) \subseteq com(u \diamond x) \subseteq L$, which implies $com(x) \subseteq shRes(L)$, i.e., $shRes(L)$ is commutative. $\square$

In the following, we give some properties and characterize $shRes(L)$ for a given language L. We begin by defining the *iterated shuffle* operation as

$$L_1 \diamond^* L_2 = \bigcup_{n=0}^{\infty} (L_1 \diamond^n L_2),$$

where $L_1 \diamond^0 L_2 = L_1$ and $L_1 \diamond^{n+1} L_2 = (L_1 \diamond^n L_2) \diamond L_2$.

Lemma 2.1. *Let $L \subseteq X^*$ and let $u, v \in shRes(L)$. Then $(v \diamond^* u) \subseteq shRes(L)$.*

Proof. Let $w \in (v \diamond^* u)$. There exists $k \geqslant 0$ such that $w \in (v \diamond^k u)$.

We will show, by induction on k, that $w \in shRes(L)$. If $k = 0$, then $w = v \in shRes(L)$. Assume the assertion holds true for k and take $w \in (v \diamond^{k+1} u)$ and $z \in L$. Then, $w \in \alpha \diamond u$ where $\alpha \in (v \diamond^k u)$. According to the induction hypothesis, $v \diamond {}^k u \subseteq shRes(L)$ therefore $\alpha \in shRes(L)$. As $\alpha, u \in shRes(L)$ and by Proposition 2.2, $shRes(L)$ is closed under the shuffle operation, $\alpha \diamond u \subseteq shRes(L)$. This implies $w \in shRes(L)$. $\square$

Proposition 2.3. *Let $L \subseteq X^*$. Then $shRes^2(L) = shRes(shRes(L)) = shRes(L)$.*

Proof. Assume $u \in shRes(shRes(L))$. As $1 \in shRes(L)$, we have $u = 1u \in shRes(L)$, i.e. $shRes(shRes(L)) \subseteq shRes(L)$. Assume now that $u \in shRes(L)$. Let $v \in shRes(L)$. Obviously, $v \diamond u \in (v \diamond^* u)$. By Lemma 2.1, $v \diamond^* u \subseteq shRes(L)$, hence $u \in shRes(shRes(L))$, i.e. $shRes(L) \subseteq shRes(shRes(L))$. $\square$

In order to characterize the shuffle residual of a language L we need to introduce an operation which is, in a sense, "inverse" to shuffle: the *scattered deletion*. Let L_1, L_2 be two languages over X. The *scattered deletion* of L_2 from L_1 is defined as (see [18]):

$$L_1 \mapsto L_2 = \{w \in X^* \mid u_1v_1u_2v_2 \dots u_kv_ku_{k+1} \in L_1,\\ v_1v_2 \dots v_k \in L_2, u_1u_2 \dots u_{k+1} = w, k \geqslant 1, u_i, v_i \in X^*\}.$$

The scattered deletion of a word v from u sparsely erases the letters of v from u, in the same order in which they occur in v, but irrespective of their position. A language L is called *scattered deletion closed*, or shortly, *sd-closed*, iff $u \in L$ and $v \in L$ imply $u \mapsto v \in L$.

We are now ready to construct the set *shRes*(L) for a given language L.

Proposition 2.4. *If L is a language in X^* then shRes$(L) = (L^c \mapsto L)^c$. Here K^c is meant the language $X^* - K$ for $K \subseteq X^*$.*

Proof. Take $x \in shRes(L)$. Assume, for the sake of contradiction, that $x \notin (L^c \mapsto L)^c$. Then, $x \in (L^c \mapsto L)$, that is, there exist $v \in L^c$, $u \in L$ such that $x \in v \mapsto u$. Note that $x \in v \mapsto u$ iff $v \in x \diamond u$. We arrived at a contradiction, as $x \in shRes(L)$ and $u \in L$ but the word v in $x \diamond u$ belongs to L^c.

Consider now a word $x \in (L^c \mapsto L)^c$. If $x \notin shRes(L)$, there exists $u \in L$, such that $(u \diamond x) \cap L^c \neq \emptyset$. Let $v \in (u \diamond x) \cap L^c$. This implies $x \in (v \mapsto u) \subseteq (L^c \mapsto L)$ – a contradiction with the initial assumptions about x. □

Corollary 2.1. *If a language L is regular, then shRes(L) is regular and can be effectively constructed.*

Proof. It follows as the family of regular languages is closed under scattered deletion (see [18]) and complementation. □

A language L such that $L \subseteq shRes(L)$ is called *shuffle-closed* or, shortly, *sh-closed*. A language L is sh-closed iff $u \in L$ and $v \in L$ imply $u \diamond v \subseteq L$. As a consequence, note that every sh-closed language is a submonoid of X^*. Note that a language L is sh-closed iff $L \diamond L \subseteq L$.

In general, submonoids of X^* are not sh-closed. For example, let $X = \{a, b, c\}$ and let $L = (a(bc)^*)^*$. Then L is a submonoid that is not sh-closed, because $a, abc \in L$, but $abac \notin L$.

Proposition 2.5. *shRes$(L) = L$ if and only if L is sh-closed and $1 \in L$.*

Proof. ($\Rightarrow$) Done.

($\Leftarrow$) If L is sh-closed then $L \subseteq shRes(L)$. For the other inclusion note that for every $u \in shRes(L)$, the fact that $1 \in L$ implies $1u \in L$, that further implies $shRes(L) \subseteq L$. □

If nonempty, the intersection of sh-closed languages is also an sh-closed language. Let L be a nonempty language and let I_L be the family of all the sh-closed languages containing L. This family is nonempty because $X^* \in I_L$. The intersection

$$sfc(L) = \bigcap_{L_i \in I_L} L_i$$

of the languages of the family I_L is clearly an sh-closed language containing L and it is called the *shuffle-closure* of L, or shortly, *sh-closure* of L. The sh-closure of a language L is the smallest sh-closed language containing L.

Proposition 2.6. *The shuffle closure of a language L is $sfc(L) = L \diamond^* L$.*

Proof. "$sfc(L) \subseteq L \diamond^* L$". Obvious, as $L \diamond^* L$ is sh-closed and L is included in $L \diamond^* L$.

"$L \diamond^* L \subseteq sfc(L)$". We show by induction on k that $L \diamond^k L \subseteq sfc(L)$. For $k = 0$ the assertion holds, as $L \subseteq sfc(L)$.

Assume that $L \diamond^k L \subseteq sfc(L)$ and consider a word $u \in L \diamond^{k+1} L = (L \diamond^k L) \diamond L$. Then $u \in w \diamond v$ where $w \in L \diamond^k L$ and $v \in L$. As both $L \diamond^k L$ and L are included in $sfc(L)$ and $sfc(L)$ is sh-closed, we deduce that $w \diamond v \subseteq sfc(L)$, i.e., $u \in sfc(L)$. The induction step, and therefore the requested equality are proved. □

Remark that, if L is a regular (context-free) language, then $sfc(L)$ is not in general a regular (context-free) language. Indeed, this follows because the families of regular and context-free languages are not closed under iterated shuffle. For example, let $L = \{a, b, c\}$. The iterated shuffle of L into L is $L_{abc} = \{w \in \{a, b, c\}^* \mid |w|_a = |w|_b = |w|_c\}$ which is not a context-free language.

Note that if L is sh-closed then $L \diamond^* L = L$. Indeed, as L is sh-closed, we have that $L = sfc(L)$. On the other hand, according to Proposition 2.6, $sfc(L) = L \diamond^* L$.

On X^* we can define an order relation, called the *embedding order* and denoted by $\leqslant_h$. For two words $u, v \in X^*$, we say that $u \leqslant_h v$ iff there exists a $w \in X^*$ such that $v \in u \diamond w$. A language $H \subseteq X^+$ is called a *hypercode* iff for all $u, v \in H$, $u \leqslant_h v$ implies $u = v$. A hyperdode is always finite.

The following result, proved in [12], relates the notions of shuffle and scattered deletion with the notion of hypercode.

Proposition 2.7. *Let M be a submonoid of X^* and $M \neq \emptyset$, $M \neq \{1\}$. Then M is sh-closed and sd-closed if and only if M is generated by a hypercode (generated refers to the shuffle operation).*

We conclude this section by considering the particular case where the language whose sh-closure we are studying is a singleton. The shuffle closure of a singleton word u, which generalizes the notion of monogenic closure of a word, [21], is denoted by $[u]$ and called the *monogenic shuffle closure* of u.

Proposition 2.8. *If $u, v \in X^*$ then $v \in [u]$ if and only if $[v] \subseteq [u]$.*

Proof. The implication $\Leftarrow$ is immediate as $v \in [v]$. For the reverse implication let $v \in [u]$. Then $v \in u \diamond^n u$ for some n. This implies $v \diamond^m v \subseteq u \diamond^{nm} u \subseteq [u]$, which shows that $[v] \subseteq [u]$. $\square$

Proposition 2.9. *Let $u \in X^+$. Then the following are equivalent.*
(a) *$[u]$ is regular,*
(b) *$u \in a^+$ for some $a \in X$,*
(c) *$[u]$ is closed under scattered deletion.*

Proof. (b) $\Rightarrow$ (a) or (c): obvious.

(a) $\Rightarrow$ (b) Assume $[u]$ regular. Suppose that $u = a^i bx$ $(i \geqslant 1)$, $a, b \in X$, $a \neq b$, $x \in X^*$. For all $v \in [u]$ we have that $|v|_b / |v|_a = \text{constant} = \alpha = |u|_b / |u|_a$. For all $n \geqslant 1$ $a^{in}(bx)^n \in [u]$. As $[u]$ is regular, according to the Pumping Lemma, if we take a large enough n, there exists a $p \geqslant 1$ with the property that $w = a^{in+p}(bx)^n \in [u]$. This implies $|w|_b / |w|_a < \alpha$ – a contradiction.

(c) $\Rightarrow$ (b) Suppose $u \notin a^+$ for any $a \in X$. Then $u = a^i bx$ for some $a, b \in X$, $a \neq b$, $i \geqslant 1$, and $x \in X^*$. We have $u^2 = a^i bxa^i bx = a^i(bxa^i)bx$. As $[u]$ is scattered deletion closed, $bxa^i \in [u]$. As $|a^i bx| = |bxa^i|$ and both words are in $[u]$, this implies $a^i bx = bxa^i$, which is impossible. Therefore, $u \in a^+$ for some $a \in X$. $\square$

3. Maximal shuffle residuals

This section will address conditions for the existence of maximal languages whose shuffle residual equals a given language, as well as a generalization of the notion of shuffle residual. Let $L \subseteq X^*$ be an sh-closed language with $1 \in L$. By $\mathscr{M}_X(L)$, we denote the set $\{M \subseteq X^* \mid shRes(M) = L$ and M is maximal in the sense of inclusion relation$\}$.

Recall that a language $L \subseteq X^*$ is sd-closed iff $L \mapsto L \subseteq L$. A language that is sh-closed and sd-closed has been called *ssh-closed* in [12]. For example, X^* and L_{ab} are sd-closed languages that are also sh-closed. Furthermore, they are both submonoids of X^*.

Definition 3.1. An sh-closed language $L \subseteq X^*$ is said to be an RSS-type language if it contains a regular ssh-closed language L_0 with $alph(L) = alph(L_0)$.

Proposition 3.1. *An sh-closed language $L \subseteq X^*$ is an RSS-type language if and only if, for any $a \in alph(L)$, $a^+ \cap L \neq \phi$.*

Proof. ($\Rightarrow$) Let $a \in alph(L)$. Then $a \in alph(L_0)$. Since L_0 is an ssh-closed language, by [12], $a^+ \cap L_0 \neq \phi$ and hence $a^+ \cap L \neq \phi$.

($\Leftarrow$) Let $alph(L) = \{a_1, a_2, \ldots, a_n\}$ and let p_i be a positive integer such that $a_i^{p_i} \in L$ for any i, $1 \leqslant i \leqslant n$. Moreover, let $L_0 = (a_1^{p_1})^* \diamond (a_2^{p_2})^* \diamond \cdots \diamond (a_n^{p_n})^*$. Then $L_0 \subseteq L$ and L_0 is a regular ssh-closed language. $\square$

In what follows, $L \subseteq X^*$ is assumed to be an RSS-type language which contains a regular ssh-closed language L_0 with $alph(L) = alph(L_0)$.

Lemma 3.1. *Let $M \subseteq X^*$ with $shRes(M) = L$. Then M can be represented as $M = \bigcup_{i \in I}(a_i \diamond L_0)$ where $a_i \in X^*$, $i \in I$.*

Proof. Obvious from the fact $L_0 \diamond M = M$. $\square$

Lemma 3.2. *Let $M \subseteq X^*$ with $shRes(M) = L$. If $alph(L) = X$, then there exists a positive integer p satisfying the following condition: For any $u \in X^*$, there exists $\beta \in X^*$ such that $|\beta| \leqslant p$ and $u \in \beta \diamond L_0$.*

Proof. Let $X = \{a_1, a_2, \ldots, a_n\}$. Since L_0 is ssh-closed, for any i, $1 \leqslant i \leqslant n$, there exists a positive integer p_i such that $(a_i^{p_i})^* \subseteq L_0$. Now let $u \in X^*$. Then $u \in u' \diamond (a_1^{p_1})^* \diamond (a_2^{p_2})^* \diamond \cdots \diamond (a_n^{p_n})^*$ where $0 \leqslant |u'|_{a_i} < p_i$ for any i, $1 \leqslant i \leqslant n$. Let $p = \sum_{i=1}^{n}(p_i - 1)$. Then $u \in u' \diamond L_0$ and $|u'| \leqslant p$. $\square$

Definition 3.2. By C_β, we denote the set $\beta \diamond L_0$.

Lemma 3.3. *Let $u, v \in C_\beta$ and let $u \leqslant_h v$. Then $v \diamond L_0 \subseteq L_0$.*

Proof. First, $(v \mapsto u) \neq \phi$. The assumption $u, v \in C_\beta$ implies that $(v \mapsto u) \subseteq com(L_0 \mapsto L_0) = com(L_0) = L_0$. Hence $v \in u \diamond L_0$. Therefore, $v \diamond L_0 \subseteq (u \diamond L_0) \diamond L_0 = u \diamond L_0$. $\square$

Proposition 3.2. *Assume $alph(L) = X$. Let $M \subseteq X^*$ with $shRes(M) = L$. Then M is regular.*

Proof. Let $M = \bigcup_{i \in I}(\alpha_i \diamond L_0)$. Then, by Lemma 3.2, there exists a positive integer q and $\beta_j \in X^*, 1 \leqslant j \leqslant q$ such that $\{\alpha_i \mid i \in I\} = \bigcup_{j \in \{1,2,\ldots q\}}(\beta_j \diamond L_0)$. Let $D_j = C_{\beta_j} \cap \{\alpha_i \mid i \in I\}$ for any j, $1 \leqslant j \leqslant q$. Note that each D_j contains a maximal hypercode H_j in D_j. Let $H_j^{\downarrow} = \{u \in D_j \mid \exists v \in H_j, u_{\leqslant_h} v\}$ for any j, $1 \leqslant j \leqslant q$ and let $E = \bigcup_{j \in \{1,\ldots,q\}} H_j^{\downarrow}$. Remark that $E \subseteq \{\alpha_i \mid i \in I\}$ and E is finite. Let $\alpha \in \{\alpha_i \mid i \in I\}$. Then there exists j, $1 \leqslant j \leqslant q$ such that $\alpha \in D_j$. By the definition of $H_j^{\downarrow}$, $\alpha \in H_j$ or $\alpha_k \leqslant_h \alpha$ for some $\alpha_k \in H_j^{\downarrow}$. In the former case, $\alpha \diamond L_0 \subseteq E \diamond L_0$. In the latter case, $\alpha \in \alpha_k \diamond L_0$ and hence $\alpha \diamond L_0 \subseteq \alpha_k \diamond L_0 \subseteq E \diamond L_0$. Therefore, $E \diamond L_0 \subseteq \bigcup_{i \in I}(\alpha_i \diamond L_0) \subseteq E \diamond L_0$ and $M = E \diamond L_0$. Since E and L_0 are regular, M is regular. $\square$

Corollary 3.1. *An RSS-type language is regular.*

Proof. Since $L = shRes(M)$ and M is regular, L is regular. $\square$

Remark that, if $alph(L) \subset X$, then the statement in Proposition 3.2 does not hold true. For instance, let $M = L \cup (\bigcup_{n \geqslant 1}(b^{n!} \diamond L))$ where L is an RSS-type language and $b \in X \backslash alph(L)$. Then $shRes(M) = L$ but M is not regular.

Proposition 3.3. *Assume* $alph(L)=X$. *Let* $M\subseteq X^*$ *with* $shRes(M)=L$. *Then there exists* $N\supseteq M$ *such that* $N\in\mathscr{M}_X(L)$.

Proof. Let $M=M_0\subset M_1\subset M_2\subset\cdots$ be an ascending chain of languages such that $shRes(M_i)=L$ for any i, $i\geqslant 0$. Moreover, let $M_i=\bigcup_{j\in I_i}(\alpha_{ij}\diamond L_0)$ for any i, $i\geqslant 0$. From the same reason as in the proof of Proposition 3.2, it follows that $\{\alpha_{ij}\mid i\geqslant 0, j\in I_i\}\subseteq\bigcup_{j\in\{1,2,\dots,q\}}C_{\beta_j}$. Let $D_k=C_{\beta_k}\cap\{\alpha_{ij}\mid i\geqslant 0, j\in I_i\}$ for any k, $1\leqslant k\leqslant q$. Let H_k be a maximal hypercode in D_k, let $H_k^{\downarrow}=\{u\in D_k\mid\exists v\in H_k, u\leqslant_h v\}$ for any k, $1\leqslant k\leqslant q$ and let $E=\bigcup_{k\in\{1,2,\dots,q\}}H_k^{\downarrow}$. Then E is finite. Suppose $M=M_0\subset M_1\subset M_2\subset\cdots$ is an infinite ascending chain. Since E is finite, there exists a possitive integer r such that $E\diamond L_0\subseteq M_r$. Let $\alpha_{(r+1)t}\in M_{r+1}\backslash M_r$. In the same way as in the proof of Proposition 3.2, $\alpha_{(r+1)t}\diamond L_0\subseteq E\diamond L_0\subseteq M_r$, a contradiction. Hence $M=M_0\subset M_1\subset M_2\subset\cdots$ is always a finite ascending chain. Consequently, $N=M_r\in\mathscr{M}_X(L)$. □

The situation is completely different for the case $alph(L)\subset X$.

Let $alph(L)=Y\subset X$ and let $Z=X\backslash Y$. Now let $Z^*=\{z_0,z_1,z_2,\dots\}$ where $z_0=1$ and $|z_i|\leqslant|z_{i+1}|$ for any i, $i\geqslant 0$. Now suppose there exists $M\in\mathscr{M}_X(L)$ and $M=\bigcup_{i\geqslant 0}(z_i\diamond M_i)$ where $M_i\subseteq Y^*$ for any i, $i\geqslant 0$. Note that $shRes(M_i)\supseteq L$ for any i, $i\geqslant 0$.

Lemma 3.4. $M_i\neq\phi$ *for any* i, $i\geqslant 0$.

Proof. Suppose $M_0=\phi$. Consider $N=M\cup L$. It is obvious that $N\supset M$ and $shRes(N)\supseteq L$. Since $M\in\mathscr{M}_X(L)$, $shRes(N)\supset L$. Let $x\in shRes(N)\backslash L$. Then there exists $m_i\in z_i\diamond M_i$, $i\geqslant 1$ such that $(m_i\diamond x)\cap L\neq\phi$. However, this is impossible because, for any $u\in L$, $|u|_Z=0$ but, for any $v\in m_i\diamond x$, $|v|_Z\geqslant|m_i|_Z\geqslant 1$. Hence $M_0\neq\phi$. Now suppose $M_i=\phi$ for some i, $i\geqslant 1$. Consider $N=M\cup(z_i\diamond L)$. Obviously, $N\supset M$ and $shRes(N)\supseteq L$. By the maximality of M, $shRes(N)\supset L$. Hence there exists $x\in shRes(N)\backslash L$ such that $(x\diamond M)\cap(z_i\diamond L)\neq\phi$. This implies that $|x|_Z>0$. Now consider $x^{|z_i|+1}\in shRes(N)\backslash L$. Then there exists $m\in M$ such that $(x^{|z_i|+1}\diamond m)\cap(z_i\diamond L)\neq\phi$. This yields a contradiction because, for any $u\in z_i\diamond L, |u|_Z=|z_i|$ but, for any $v\in x^{|z_i|+1}\diamond m, |v|_Z\geqslant|z_i|+1$. Hence $M_i\neq\phi$ for any i, $i\geqslant 0$. □

Lemma 3.5. *Let* $\mathscr{N}=\{i\mid i\geqslant 0, M_i\neq Y^*\}$. *Then* $\mathscr{N}$ *is infinite.*

Proof. Suppose there exists a positive integer n_0 such that, for any $n\geqslant n_0$, $M_n=Y^*$. Consider $z_n^2\in Z^+$. Obviously, $z_n^2\diamond M\subseteq M$. Hence $z_n^2\in shRes(M)$, a contradiction. This completes the proof of the lemma. □

Now let $K_i=shRes(M_i)$ for any i, $i\geqslant 0$. Recall that $K_i\supseteq L$ for any i, $i\geqslant 0$.

Lemma 3.6. $L=\bigcup_{i\geqslant 0}K_i$.

Proof. Obviously, $L\subseteq\bigcup_{i\geqslant 0}K_i$. Let $x\in\bigcup_{i\geqslant 0}K_i$. Since $x\in K_0, x\diamond M_0\subseteq M_0$. Therefore, $|x|_Z=0$ and $(z_i\diamond M_i)\diamond x\subseteq(z_i\diamond M_i)$ for any i, $i\geqslant 1$. This implies that $x\diamond M\subseteq M$, i.e. $x\in L$. This completes the proof of the lemma. □

Lemma 3.7. *Let* $N = (\bigcup_{i\geqslant 0, i\neq k}(z_i \diamond M_i)) \cup (z_k \diamond Y^*)$. *Then* $shRes(N) = \bigcap_{i\geqslant 0, i\neq k} K_i$.

Proof. That $\bigcap_{i\geqslant 0, i\neq k} K_i \subseteq shRes(N)$ is obvious. Assume $x \in shRes(N)$. If $|x|_Z > 0$, then $x^{|z_k|+1} \in shRes(N)$. Since $x^{|z_k|+1} \diamond N \subseteq \bigcup_{i\geqslant k+1}(z_i \diamond M_i)$ and $z_k \diamond M_k \subseteq z_k \diamond Y^*, x^{|z_k|+1} \diamond M \subseteq M$, i.e. $x^{|z_k|+1} \in shRes(M) = L$, a contradiction. Hence $|x|_Z = 0$. Since $|x|_Z = 0, x \diamond M_i \subseteq M_i$ for any $i, k \neq i \geqslant 0$, i.e. $x \in \bigcap_{i\geqslant 0, i\neq k} K_i$. This completes the proof of the lemma. □

Let $\mathscr{A} = \{w \in X^* \mid \exists i \geqslant 0, w \notin K_i, \forall j, j \neq i, w \in K_j\}$. Note that $\mathscr{A} = \{w \in X^* \mid \exists i \geqslant 0, w \notin K_i, \forall j, \ j \neq i, w \diamond L_0 \subseteq K_j\}$. As the proof of Proposition 3.2, there exists a finite set $\mathscr{B}$ with $\mathscr{B} \subseteq \mathscr{A}$ such that, for any $w \in \mathscr{A}$ there exists $w' \in \mathscr{B}$ with $w \in w' \diamond L_0$. Now let $w \notin K_i$ and let $w' \notin K_j$. Suppose $i \neq j$. Then $w' \in K_i$ and hence $w \in w' \diamond L_0 \subseteq K_i$, a contradiction. Therefore, $i = j$. Let $\mathscr{B} = \{w_1, w_2, \ldots, w_r\}$. Moreover, for any i, $1 \leqslant i \leqslant r$, we choose some integer $f(i)$ such that $w_i \notin K_{f(i)}$. Then the following is now obvious.

Lemma 3.8. *Let* $w \in \mathscr{A}$. *Then* $w \notin \bigcap_{1\leqslant i\leqslant r} K_{f(i)}$.

Proposition 3.4. *Let* $alph(L) = Y \subset X$. *Then* $\mathscr{M}_X(L) = \phi$.

Proof. By Lemma 3.5, there exists a positive integer t such that $M_t \neq Y^*$ and $t \notin \{f(1), f(2), \ldots, f(r)\}$. Let $u \notin L$. If $u \in \mathscr{A}$, then $u \notin \bigcap_{1\leqslant i\leqslant r} K_{f(i)}$ and hence $u \notin \bigcap_{i\geqslant 0, i\neq t} K_i$. If $u \notin \mathscr{A}$, then there exist at least two distinct integers i and j such that $u \notin K_i \cup K_j$. Therefore, $u \notin \bigcap_{i\geqslant 0, i\neq t} K_i$. Now let $N = (\bigcup H_{i\geqslant 0, i\neq t}(z_i \diamond M_i)) \cup (z_t \diamond Y^*)$. Obviously, $N \supset M$. By Lemma 3.7, $shRes(N) = \bigcap_{i\geqslant 0, i\neq t} K_i = L$. This contradicts the maximality of M and hence $\mathscr{M}_X(L) = \phi$. □

We consider now similar questions for a generalization of the notion of shuffle residual. The shuffle residual of a language consists of the words x whose shuffle $L \diamond x$ is completely included in L. We can relax this condition by only requiring that at least one word from $L \diamond x$ belongs to L. The notion obtained in this way generalizes the notion of shuffle residual. More precisely, the *generalized shuffle residual* of a language L, denoted by g-$shRes(L)$ is defined as follows.

Definition 3.3. Let $M \subseteq X^*$. Then $g\text{-}shRes(M) = \{x \in X^* \mid \exists y \in M, (x \diamond y) \cap M \neq \emptyset\}$.

The following results give some properties of the generalized shuffle residual of a language.

Proposition 3.5. *If* M *is a semigroup, then* g-$shRes(M)$ *is a monoid.*

Proof. (i) $1 \in M$: obvious.

(ii) $x, y \in g\text{-}shRes(M)$ implies that there exist $z_1, z_2 \in M$ such that $(x \diamond z_1) \cap M \neq \emptyset$ and $(y \diamond z_2) \cap M \neq \emptyset$. This implies that $(xy \diamond z_1z_2) \cap M \neq \emptyset$, that is, $xy \in g\text{-}shRes(M)$. □

Proposition 3.6. *If M is sh-closed, then g-shRes(M) is sh-closed.*

Proof. If $x, y \in g\text{-}shRes(M)$ this means that there exist $z_1, z_2 \in M$ such that $(x \diamond z_1) \cap M \neq \emptyset$ and $(y \diamond z_2) \cap M \neq \emptyset$.

Let z be a word in $x \diamond y$. As $z_1 \diamond z_2 \subseteq M$ we have that

$$z \diamond (z_1 \diamond z_2) \subseteq (x \diamond y) \diamond (z_1 \diamond z_2) = (x \diamond z_1) \diamond (y \diamond z_2),$$

which implies that $[z \diamond (z_1 \diamond z_2)] \cap M \neq \emptyset$. $\square$

Proposition 3.7. *If $M \subseteq X^*$ is regular, then g-shRes(M) is regular.*

Proof. Let M be a regular language accepted by the finite automaton $A = (S, X, \delta, s_0, F)$, where S is the set of states, X is the alphabet, δ is the transition function, s_0 the initial state, and F the set of final states of A. Denote by $\bar{X} = \{\bar{a} \mid a \in X\}$ and $\tilde{X} = X \cup \bar{X}$.

Consider the function $\delta_1 : S \times \tilde{X} \to S$ defined as $\delta_1(s, a) = \delta(s, a)$ if $a \in X$ and $\delta_1(s, \bar{a}) = s$ if $\bar{a} \in \bar{X}$.

Consider another function $\delta_2 : S \times \bar{X} \to S$ defined by $\delta_2(s, a) = \delta(s, a)$ if $a \in X$ and $\delta_2(s, \bar{a}) = \delta(s, a)$ if $\bar{a} \in \bar{X}$.

Define now the automaton

$$\tilde{A} = (S, \tilde{X}, \tilde{\delta}, (s_0, s_0), \{(s, t) \mid s \in F,\ t \in F\}),$$

where the transition function $\tilde{\delta} : (S \times S) \times \tilde{X} \to S \times S$ is $\tilde{\delta}((s, t), b) = (\delta_1(s, b), \delta_2(t, b))$ for $b \in \tilde{X}$.

It is not difficult to see that

$$\begin{aligned} L(\tilde{A}) = \ & \{z \in \tilde{X}^* \mid \exists x_i \in \{1\} \cup X, \exists y_i \in \{1\} \cup X,\ z = x_1 \bar{y}_1 x_2 \bar{y_2} \ldots x_n \bar{y}_n, x_1 x_2 \ldots x_n \in M, \\ & x_1 y_1 x_2 y_2 \ldots x_n y_n \in M,\ \text{where } \bar{1} = 1\}. \end{aligned}$$

Let ρ be the morphism of $\tilde{X}^*$ into X^* defined by $\rho(a) = 1$ if $a \in X$ and $\rho(\bar{a}) = a$ if $\bar{a} \in \bar{X}$.

Then it is easy to see that $g\text{-}shRes(M) = \rho(L(\tilde{A}))$, hence $g\text{-}shRes(M)$ is regular. $\square$

Denote now by $\mathscr{G} = \{L \subseteq X^* \mid \exists M \subseteq X^* \text{ such that } g\text{-}shRes(M) = L\}$.

Remark 3.1. The following statement is not always true: "For any $L \subseteq \mathscr{G}$ there exists a maximal $M \subseteq X^*$ such that $g\text{-}shRes(M) = L$.

Proof. $\{1\} \in \mathscr{G}$. Suppose that there exists a maximal $M \subseteq X^*$ such that $g\text{-}shRes(M) = \{1\}$. Let $\alpha \notin M$ and let $\tilde{M} = M \cup \{\alpha\}$. It is obvious that $\{1\} \in g\text{-}shRes(\tilde{M})$. Suppose that $1 \neq x \notin g\text{-}shRes(\tilde{M})$. Since $g\text{-}shRes(\tilde{M})$ is a monoid, $x^n \in g\text{-}shRes(\tilde{M})$ for any $n, n \geqslant 1$. Let $n \geqslant 1$ such that $n|x| > |\alpha|$. Note that $(M \cup \{\alpha\}) \diamond x^n \subseteq M \cup \alpha$. However, since $x^n \notin g\text{-}shRes(M)$, there exists $m \in M$ such that $x^n \diamond m \nsubseteq M$. Hence $\alpha \in x^n \diamond m$, but this contradicts the assumption $n|x| > |\alpha|$. Therefore $g\text{-}shRes(\tilde{M}) = \{1\}$. This means that M is not maximal. $\square$

4. Scattered deletion closure

This section parallels Section 2 by considering a notion analogous to the shuffle residual of a language, but this time in relation to the scattered deletion operation.

Let $L \subseteq X^*$ and define the set of sparse subwords of L by:

$$sps(L) = \{u \in X^* \mid u = a_1 \dots a_k, \text{ and } \exists v_1 a_1 v_2 a_2 \dots v_k a_k v_{k+1} \in L,\ a_i, v_i \in X^*\}.$$

To the language L one can associate the set consisting of all words x with the following property: x is sparse subword of at least one word of L, and the scattered deletion of x from any word of L containing x as sparse subword yields words belonging to L. The set defined in this way, denoted by $sdRes(L)$ and called the *scattered deletion residual* of L, is formally defined by

$$sdRes(L) = \{x \in sps(L) \mid \forall u \in L, u \mapsto x \subseteq L\}.$$

The condition that $x \in sps(L)$ has been added because otherwise $sdRes(L)$ would contain irrelevant elements: words which are not sparse subwords of any word of L and thus yield $\emptyset$ as a result of the scattered deletion from L.

Example. Let $X = \{a, b\}$. Then,
- $sdRes(X^*) = X^*$;
- $sdRes(L_{ab}) = L_{ab}$;
- if $L = \{a^n b^n \mid n \geqslant 0\}$ then $sdRes(L) = L$;
- if $L = b^* a b^*$ then $sdRes(L) = b^*$.

The following proposition gives some basic properties of the scattered deletion residual of a language.

Proposition 4.1. *Let* $L \subseteq X^*$.
(i) *If* $x, y \in sdRes(L)$ *and* $xy \in sps(L)$, *then* $xy \in sdRes(L)$.
(ii) *If* $sps(L)$ *is a submonoid of* X^*, *then* $sdRes(L)$ *is a submonoid of* X^*.
(iii) *If* L *is a commutative language, then* $sdRes(L)$ *is also commutative.*

Proof. (i) Let $x, y \in sdRes(L)$ with $xy \in sps(L)$. If $u \in L$,

$$u = u_1 x_1 u_2 x_2 \dots u_k x_k u_{k+1} y_1 u_{k+2} y_2 \dots u_{n+k} y_n u_{n+k+1}$$

then, as $x \in sdRes(L)$, $u_1 \dots u_{k+1} y_1 \dots y_n u_{n+k+1} \in L$ and, as $y \in sdRes(L)$ we can conclude that $u_1 \dots u_k \dots u_{n+k+1} \in L$. As the initial decomposition of u was arbitrary, we deduce that $u \mapsto xy \subseteq L$, which implies $xy \in sdRes(L)$.

(ii) Immediate from (i).

(iii) Let $x = x_1 x_2 \dots x_k \in sdRes(L)$, $x_i \in X^*$ for $1 \leqslant i \leqslant k$. As $x \in sps(L)$ and L is commutative, $com(x) \subseteq sps(L)$. Let $y = y_1 y_2 \dots y_k$, $y_i \in X$, be a word in $com(x)$ and let $u = u_1 y_1 \dots u_k y_k u_{k+1} \in L$, $u_i \in X^*$, $y_i \in X$. As L is commutative, the word

$u_1x_1 \ldots u_kx_ku_{k+1}$ is in L and, as $x \in sdRes(L)$, we have that $u_1 \ldots u_ku_{k+1} \in L$. This implies $u \mapsto y \subseteq L$, which means $y \in sdRes(L)$. □

In the following, we show how, for a given language L, the set $sdRes(L)$ can be constructed. The construction is similar to the one for $shRes(L)$.

Proposition 4.2. *If L is a language in X^* then $sdRes(L) = (L \mapsto L^c)^c \cap sps(L)$.*

Proof. Let $x \in sdRes(L)$. From the definition of $sdRes(L)$ it follows that $x \in sps(L)$. Assume that $x \notin (L \mapsto L^c)^c$. This means there exists $w \in L$, $v \in L^c$ such that $x \in (w \mapsto v)$. This further implies $v \in (w \mapsto x)$. We arrived at a contradiction as $x \in sdRes(L)$ but there exists a word $w \in L$ with $(w \mapsto x) \cap L^c = v \neq \emptyset$.

For the other inclusion, let $x \in (L \mapsto L^c)^c \cap sps(L)$. As $x \in sps(L)$, if $x \notin sdRes(L)$ then there exists $w \in L$ such that $v \in (w \mapsto x) \cap L^c \neq \emptyset$. This implies $x \in (w \mapsto v) \subseteq (L \mapsto L^c)$ – a contradiction with the initial assumption about x. □

The following, result connects the notions of shuffle and scattered deletion.

Proposition 4.3. *Let $L \subseteq X^*$ be an sh-closed language. Then L is sd-closed if and only if $L = (L \mapsto L)$.*

Proof. If L is sd-closed, $L \mapsto L \subseteq L$. Now let $u \in L$. Since L is sh-closed, $uu \in L$. Therefore $u \in (L \mapsto L)$, i.e. $L \subseteq (L \mapsto L)$. We can conclude that $L = (L \mapsto L)$. The other implication is obvious. □

If L is a nonempty language and if D_L is the family of all the sd-closed languages L_i containing L, then the intersection

$$\bigcap_{L_i \in D_L} L_i$$

of all the sd-closed languages containing L is an sd-closed language called the *scattered deletion closure* of L, or shortly, *sd-closure* of L. The sd-closure of L is the smallest sd-closed language containing L.

We will now define a sequences of languages whose union is the sd-closure of a given language L. Let

$$sdc_0(L) = L,$$

$$sdc_1(L) = sdc_0(L) \mapsto (sdc_0(L) \cup \{1\}),$$

$$sdc_2(L) = sdc_1(L) \mapsto (sdc_1(L) \cup \{1\}),$$

$$\ldots$$

$$sdc_{k+1}(L) = sdc_k(L) \mapsto (sdc_k(L) \cup \{1\}).$$

$$\ldots$$

Clearly $sdc_k(L) \subseteq sdc_{k+1}(L)$. Let

$$sdc(L) = \bigcup_{k \geqslant 0} sdc_k(L).$$

Proposition 4.4. *$sdc(L)$ is the sd-closure of the language L.*

Proof. Clearly $L \subseteq sdc(L)$. Let now $v \in sdc(L)$ and $u \in sdc(L)$. Then $v \in sdc_i(L)$ and $u \in sdc_j(L)$ for some integers $i, j \geqslant 0$. If $k = \max\{i, j\}$, then $v \in sdc_k(L)$ and $u \in sdc_k(L)$. This implies $(u \mapsto v) \subseteq sdc_{k+1}(L) \subseteq sdc(L)$. Therefore $sdc(L)$ is an sd-closed language containing L.

Let T be an sd-closed language such that $L = sdc_0(L) \subseteq T$. Since T is sd-closed, if $sdc_k(L) \subseteq T$ then $sdc_{k+1}(L) \subseteq T$. By an induction argument, it follows that $sdc(L) \subseteq T$. $\square$

Since, by [18], the family of regular languages is closed under scattered deletion, it follows that if L is regular, then the languages $sdc_k(L)$, $k \geqslant 0$, are also regular. However, it is an open question whether $sdc(L)$ is regular for any regular language $L \subseteq X^*$.

Recall that, for a language L, the *principal congruence P_L* is defined by

$$u \equiv v(P_L) \text{ iff } \forall x,\ y \in X^* \text{ we have } xuy \in L \Leftrightarrow xvy \in L.$$

When the principal congruence of L has a finite index (finite number of classes) the language L is regular.

If L is commutative, we have the following result.

Proposition 4.5. *Let $L \subseteq X^*$ be a regular language. If L is commutative, then its scattered deletion closure $sdc(L)$ is commutative and regular.*

Proof. Let us prove first that $sdc(L)$ is commutative. To this end, it is sufficient to show that $sdc_{k+1}(L)$ is commutative if $sdc_k(L)$ is commutative. Let $xuvy \in sdc_{k+1}(L)$. If $xuvy \in sdc_k(L)$, then we are done. Otherwise, by the definition of $sdc_{k+1}(L)$, there exist $w, z \in sdc_k(L)$ such that $w \in (xuvy \diamond z)$. Since $sdc_k(L)$ is commutative, $xuvyz \in sdc_k(L)$ and $xvuyz \in sdc_k(L)$. From the fact that $z, xvuyz \in sdc_k(L)$ and the definition of $sdc_{k+1}(L)$, it follows that $xvuy \in sdc_{k+1}(L)$, i.e. $sdc_{k+1}(L)$ is commutative.

We will show next that $sdc(L)$ is regular. To this aim, we show that if $u \equiv v(P_{sdc_k(L)})$ then $u \equiv v(P_{sdc_{k+1}(L)})$. Let $u \equiv v(P_{sdc_k(L)})$ and let $xuy \in sdc_{k+1}(L)$. By the definition of $sdc_{k+1}(L)$, there exists $w, z \in sdc_k(L)$ such that $w \in (xuy \diamond z)$. Since $sdc_k(L)$ is commutative, $xuyz \in sdc_k(L)$. Hence $xvyz \in sdc_k(L)$. From the fact that $z \in sdc_k(L)$ and by the definition of $sdc_{k+1}(L)$, it follows that $xvy \in sdc_{k+1}(L)$. In the same way, $xvy \in sdc_{k+1}(L)$ implies $xuy \in sdc_{k+1}(L)$. Consequently, $u \equiv v(P_{sdc_{k+1}(L)})$ holds. This means that the number of congruence classes of $P_{sdc_{k+1}(L)}$ is smaller or equal to that of $P_{sdc_k(L)}$. Remark that

$$sdc_0(L) \subseteq sdc_1(L) \subseteq \cdots \subseteq sdc_n(L) \subseteq sdc_{n+1}(L) \cdots .$$

It can be shown, [10], that $sdc_t(L) = sdc_{t+1}(L)$ for some $t, t \geqslant 1$. Thus, $sdc(L) = sdc_t(L)$ which implies that $sdc(L)$ is regular. $\square$

5. Combining the operations

Besides examining the notion of a shuffle-base of a language, this section studies relations and interdependencies between shuffle, scattered deletion and various other insertion and deletion operations.

If L is a shuffle closed language then its shuffle base is defined as

$$J(L) = \{u \in L \mid u \neq 1, u \notin (L \backslash \{1\}) \diamond (L \backslash \{1\})\} = L \backslash [(L \backslash \{1\}) \diamond^{+} (L \backslash \{1\})],$$

i.e. $J(L)$ consists of the words of L that are not the result of shuffle of any nonempty words of L. Then $J(L)$ is uniquely determined and $L \backslash \{1\} = J(L) \diamond^{*} J(L)$. Properties of the shuffle base of a language have been investigated in [12].

The following result shows that if L is regular, its shuffle base is also regular. The proof is based on the fact that one can construct a generalized sequential machine (shortly, gsm; see [19] for a definition) g such that $g(L)$ is the set of words in L that can be obtained as a result of shuffle.

Proposition 5.1. *If L is a regular shuffle closed language then its shuffle base $J(L)$ is a regular language.*

Proof. Let L be a regular sh-closed language. We can assume, without loss of generality, that L is 1-free. Let $A = (X, S, s_0, F, P)$ be a finite deterministic automaton accepting L, where X is the alphabet, S is the set of states, s_0 is the initial state, F is the set of final states, and the rules of P are of the form $s_i a \rightarrow s_j, s_i, s_j \in S, a \in X$.

We will show that there exists a generalized sequential machine g, such that $g(L) = L \backslash J(L)$. As the family of regular languages is closed under gsm mappings and set difference, it will follow that $J(L)$ is regular.

Note first that, as L is shuffle closed, $L \backslash J(L) = \{u \in L \mid u \in L \diamond L\}$.

Consider now the gsm $g = (X, X, S \times S, (s_0, s_0), F \times F, P')$ where

$$\begin{aligned} P' = \ & \{(s_i, s_j)a \rightarrow a(s_i, s'_j) | s_j a \rightarrow s'_j \in P\} \\ & \cup \{(s_i, s_j)a \rightarrow a(s'_i, s_j) | s_i a \rightarrow s'_i \in P\} \end{aligned}$$

The idea of the construction is the following. We have constructed two copies of the set of states. Given a word $u_1 v_1 u_2 v_2 \ldots u_k v_k \in L$ as an input, the gsm g works as follows. The first component of a state makes sure that the word $u_1 u_2 \ldots u_k$ belongs to L while the second component makes sure that $v_1 v_2 \ldots v_k$ is in L. While scanning the letters of the input, the derivation affects either the first component or the second, but not both. That is, according to the choice made, the letter will either be considered to belong to $u_1 \ldots u_k$ or to $v_1 \ldots v_k$. A final state will be reached only if a final state is reached in both components, i.e. if both $u_1 \ldots u_k \in L$ and $v_1 \ldots v_k \in L$.

From the above explanations it follows that g reaches a final state iff the input word is of the form $u_1v_1u_2v_2\dots u_kv_k$ with $u_1\dots u_k\in L$ and $v_1\dots v_k\in L$, that is, iff the input is the result of the shuffle of two words in L. Consequently, $g(L)=\{u \mid u\in L\diamond L\}$. $\square$

A language $L\subseteq X^*$ is called *e-convex*, [20], iff $u\leqslant_h x\leqslant_h v$, and $u,v\in L$ imply $x\in L$. In particular, any hypercode is an e-convex language. A language L is called a μ-ideal of X^*, [22], iff $u_1u_2\in L$, $x\in X^*$ imply $u_1xu_2\in L$.

Proposition 5.2. *L is shuffle closed and e-convex if and only if L is a μ-ideal of Y^*, $Y\subseteq X$.*

Proof. ($\Rightarrow$) Let $Y=alph(L)$. Let $u\in L$ and $a\in Y$. Then there exists $v\in L$ with $v=xay$. Hence $u\diamond v\subseteq L$ and in particular, $xu_1au_2y\in L$ for any decomposition $u=u_1u_2$ of u. Clearly,

$$u\leqslant_h u_1au_2\leqslant_h xu_1au_2y, \text{ with } u=u_1u_2.$$

Therefore, $u_1au_2\in L$ for all decompositions $u=u_1u_2$ of u. If $x\in Y^*$, $x=x_1\dots x_k, x_i\in X$ then $u_1x_1u_2\in L$, which implies $u_1x_1x_2u_2\in L$ and so on. Finally we conclude that $u_1xu_2\in L$, i.e., L is a μ-ideal of Y^*.

($\Leftarrow$) Let $u,v\in L$. As L is a μ-ideal, by iteratedly inserting the letters of v into u we obtain words belonging to L, i.e., $u\diamond v\subseteq L$. Analogously, if $u\leqslant_h x\leqslant_h v$ and $u,v\in L$, then x can be obtained from u by inserting some letters, therefore $x\in L$. $\square$

In the remainder of this section we consider relations between various insertion and deletion operations.

Example.

- The language $L=\{a^nb^n \mid n\geqslant 0\}$ is sd-closed but is not shuffle closed.
- The language $L=aX^*b$, where $X=\{a,b\}$ is shuffle closed but is not sd-closed.
- Any shuffle closed language that is 1-free is an example of a language that is shuffle closed but not sd-closed.

The following results connect shuffle and scattered deletion with ordinary insertion and deletion operations. Recall that a language L is *insertion closed* or shortly *ins-closed*, iff for all $u,x,v\in X^*$, $x\in L$ and $uv\in L$ imply $uxv\in L$ (see [11]). Analogously, a language L is called *deletion closed*, or shortly *del-closed*, iff for all $u,x,v\in X^*$, $uxv\in L$ and $x\in L$ imply $uv\in L$ (see [11]).

Proposition 5.3. *Any language $L\subseteq X^*$ that is ins-closed and sd-closed is shuffle closed.*

Proof. It is enough to prove that the language L is commutative.

Let $xuvy\in L$. As L is ins-closed, the word $x(xuvy)uvy$ belongs to L. L is sd-closed and therefore $xvyu\in L$. The fact that L is ins-closed implies that $xv(xuvy)yu\in L$.

By deleting $xvyu$ and because L is sd-closed, we obtain $xvuy \in L$. This implies that L is commutative and, together with the fact that L is ins-closed, it implies that L is shuffle closed. □

Proposition 5.4. *An sh-closed language that is del-closed is not always an ssh-closed language.*

Proof. Let $X = \{a, b\}$ and let f be the following mapping of X^* into the set of integers:
(1) $f(1) = 0$
(2) $f(a) = 1, f(b) = -1$
(3) $f(a_1 a_2 \dots a_n) = \sum_{i=1}^{n} f(a_i)$ where $a_i \in X, 1 \leqslant i \leqslant n$.

Consider the language $L = \{u \in X^* \mid \ u = vw \Rightarrow f(v) \geqslant 0$ and $f(u) = 0\}$.

Then it is not difficult to verify that L is an sh-closed language which is del-closed.

Indeed, let $u, v \in L$, $u = u_1 u_2 \dots u_n$ and $v = v_1 v_2 \dots v_n$. For a prefix of a word in the shuffle, $u_1 v_1 \dots u_k v_k, 1 \leqslant k \leqslant n$ we have that $f(u_1 v_1 \dots u_k v_k) = f(u_1 \dots u_k) + f(v_1 \dots v_k) \geqslant 0$. Moreover, $f(u_1 v_1 \dots u_n v_n) = f(u_1 \dots u_n) + f(v_1 \dots v_n) = 0$. This implies that the language L is sh-closed. Let now $uvw \in L$ and $v \in L$. We have that

$$f(uw) = f(u) + 0 + f(w) = f(u) + f(v) + f(w) = f(uvw) = 0.$$

If $u = u_1 u_2$ then $f(u_1) \geqslant 0$ since $u_1 u_2 vw \in L$. On the other hand, if $w = w_1 w_2$ then

$$f(uw_1) = f(u) + 0 + f(w_1) = f(u) + f(v) + f(w_1) = f(uvw_1) \geqslant 0$$

since $uvw_1 w_1 \in L$. This means L is del-closed.

Note that $L \subseteq aX^* \cup \{1\}$. Moreover, $|u|_b \neq 0$ for any $u \in L \backslash \{1\}$. Suppose L is ssh-closed. Then, according to the proof of Proposition 5.3, L is commutative, hence $L \cap bX^* \neq \emptyset$, a contradiction. Therefore, L is not ssh-closed. □

A language L is *dipolar deletion closed*, [11], or shortly *dipdel-closed* iff for all $u, x, v \in X^*, uxv \in L$ and $uv \in L$ imply $x \in L$.

Proposition 5.5. *A language L that is sh-closed and dipdel-closed is ssh-closed.*

Proof. It is enough to show that the language L is commutative. Let $xuvy \in L$. As L is sh-closed, the word $xu\mathbf{x}v\mathbf{u}y\mathbf{vy}$ is in L. As L is dipdel-closed, this implies that $xvuy \in L$. □

Lemma 5.1. *Let X be an alphabet and let G be a finite group. Moreover, let f be a morphism of X^* into G. Then $L = \{u \in X^* \mid \ f(u) = e\}$ is ins-closed and dipdel-closed.*

Proof. Let $v \in L$ and $uw \in L$. Then $f(uvw) = f(u)f(v)f(w) = f(u)f(w) = f(uw) = e$, i.e., $uvw \in L$, which means that L is ins-closed. Now, let $uw, uvw \in L$. Then $e = f(uw) = f(u)f(w)$ and $f(w) = f(u)^{-1}$. Since $e = f(uvw) = f(u)f(v)f(w) = f(u)f(v)$

$f(u)^{-1}$, we have $f(u)^{-1}ef(u)=f(u)^{-1}f(u)f(v)f(u)^{-1}f(u)=f(v)$, which implies $f(v)=e$, i.e., $v\in L$. This means L is dipdel-closed. $\square$

Denote by $\mathscr{S}_n$ the group of permutations of $\{1,\dots n\}$.

Proposition 5.6. *Let $|X|\geqslant 2$. Then there exists an ins-closed and dipdel-closed language $L\subseteq X^*$ which is not ssh-closed.*

Proof. Let $X=\{a,b,\dots\}$ and let $f(a)=(1\quad 2)\in\mathscr{S}_3, f(b)=(1\quad 3)\in\mathscr{S}_3$ and $f(c)=e\in\mathscr{S}_3$ for any $c\in X\setminus\{a,b\}$. Let $L=\{u\in X^*\mid f(u)=e\}$. By the preceding lemma it follows that L is ins-closed and dipdel-closed. Since $f(a^2b^2)=e$, $a^2b^2\in L$. On the other hand, since $f(abab)(1)=3$, $f(abab)\neq e$ and $abab\notin L$. Therefore L is not commutative and, according to Proposition 5.3, L is not ssh-closed. $\square$

A language $L\subseteq X^*$ is called *reflective* iff $uv\in L, u,v\in X^*$ imply $vu\in L$.

Proposition 5.7. *Let $L\subseteq X^*$ be an ins-closed language that is dipdel-closed. Then L is reflective.*

Proof. Let $uv\in L$. Then $uvuv\in L$. Since $uvuv=u(vu)v$ and $uv\in L, vu\in L$. $\square$

Note that if $L\subseteq X^*$ is an sh-closed language, $L\neq\{1\}$ and $L_n=L\setminus X^{[n]}$ where $X^{[n]}=\bigcup_{i=1}^{n}X^i$ then L_n is sh-closed for any n. This implies that there does not exist a minimal sh-closed language.

References

[1] T. Araki, N. Tokura, Decision problems for regular expressions with shuffle and shuffle closure operators, Trans. Inst. Electron. Commun. Eng. Japan J64-D, (1981) 1069–1073.

[2] T. Araki, N. Tokura, Flow languages equal recursively enumerable languages, Acta Inform. 15 (1981) 209–217.

[3] B. Berard, Literal shuffle, Theoret. Comput. Sci. 51 (1987) 281–299.

[4] S.L. Bloom, Z. Ésik, Nonfinite axiomatizability of shuffle inequalities, in: Proc. TAPSOFT'95, LNCS 915, Springer, Berlin, 1995, pp. 318–333.

[5] Z. Ésik, M. Bertol, Nonfinite axiomatizability of the equational theory of shuffle, in: Proc. ICALP'95, LNCS 944, Springer, Berlin, 1995, pp. 27–38.

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

[7] B. Imreh, M. Ito, A note on regular strongly shuffle-closed languages, Acta Cybernet. 11 (1994) 257–260.

[8] B. Imreh, M. Ito, M. Katsura, On shuffle closures of commutative regular languages, in: D.S. Bridges et al. (Eds.), Combinatorics, Complexity and Logic, Springer, Berlin, 1996, 276–288.

[9] M. Ito, G. Tanaka, Dense property of initial literal shuffles, Internat. J. Comput. Math. 34 (1990) 161–170.

[10] M. Ito, C. M. Reis, G. Thierrin, Adherence in finitely generated free monoids, Congr. Numer. 95 (1988) 73–83.

[11] M. Ito, L. Kari, G. Thierrin, Insertion and deletion closure of languages, Theoret. Comput. Sci. 183 (1997) 3–19.

[12] M. Ito, G. Thierrin, S.S. Yu, Shuffle-closed Languages, Publ. Math. Debrecen 46/3-4 (1995) 1–21.
[13] K. Iwama, Universe problem for unrestricted flow languages, Acta Inform. 19 (1983) 85–96.
[14] M. Jantzen, Extending regular operations with iterated shuffle, Theoret. Comput. Sci. 38 (1985) 223–247.
[15] J. Jedrzejowicz, Nesting of shuffle closure is important, Inform. Process. Lett. 25 (1987) 363–367.
[16] J. Jedrzejowicz, Infinite hierarchy of shuffle expressions over a finite alphabet, Inform. Process. Lett. 36 (1990) 13–17.
[17] J. Jedrzejowicz, Undecidability results for shuffle languages, Automata, Languages Combin. 1 (1996) 147–159.
[18] L. Kari, On insertion and deletion in formal languages, Ph.D. Thesis, University of Turku, Finland, 1991.
[19] A. Salomaa, Formal Languages, Academic Press, London, 1973.
[20] H.J. Shyr, Free Monoids and Languages, Hon Min Book, Taichung, 1991.
[21] H.J. Shyr, S.S. Yu, G. Thierrin, Monogenic e-closed languages and dipolar words, Discrete Math. 126 (1994) 339–348.
[22] G. Thierrin, Decomposition of some classes of subsets in a semigroup, Proc. of Conf. on Semigroups and Applications, Oberwolfach, 1978, Springer, Berlin, 1981.