CS–2008–06

Continuous Languages

Thomas Ang and Janusz Brzozowski

Technical Report 06

David R. Cheriton School of Computer Science
University of Waterloo

March 12, 2008
Revised April 21, 2008

Continuous Languages[*]

Thomas Ang and Janusz Brzozowski
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, ON, Canada N2L 3G1
`tang@student.cs.uwaterloo.ca, brzozo@uwaterloo.ca`

Abstract

A language is prefix-continuous if it satisfies the condition that, if a word w and its prefix u are in the language, then so is every prefix of w that has u as a prefix. Prefix-continuous languages include prefix-closed languages at one end of the spectrum, and prefix-free languages, which include prefix codes, at the other. In a similar way, we define suffix-, bifix-, factor-, and subword-continuous languages and their closed and free counterparts. We generalize these notions to arbitrary binary relations on Σ^*. This provides a common framework for diverse languages such as codes, factorial languages and ideals. We examine the relationships among these languages and their closure properties.

1 Introduction

Prefix-continuous languages were introduced in connection with trace-assertion specifications [5, 6], where a software module is modeled by an automaton in which the states are represented by words over the input alphabet. It was shown in [5], for deterministic automata, that the automaton is well-behaved if the set of words representing the states is prefix-continuous. This result was extended to nondeterministic automata in [4]. Applications of these methods to the specification of software modules were discussed in [6]. In this paper we consider some theoretical aspects of prefix-continuous and related languages.

Let Σ be an alphabet, and Σ^*, the free monoid generated by Σ, with ϵ as the empty word. A language over an alphabet Σ is any subset of Σ^*. If $L \subseteq \Sigma^*$, the complement of L with respect to Σ^* is denoted by $\overline{L}$. When convenient, we use

---

[*] This research was supported by the Natural Sciences and Engineering Research Council of Canada under grant no. OGP000871.

the customary notation for regular expressions, with + for union, juxtaposition for concatenation, and * for Kleene closure.

We generalize the concept of prefix-continuity to continuity with respect to an arbitrary binary relation. Suppose $\trianglelefteq$ is a binary relation on Σ^*; if $u \trianglelefteq v$ and $u \neq v$, we write $u \triangleleft v$. Let $\trianglerighteq$ be the converse binary relation, that is, let $u \trianglerighteq v$ if and only if $v \trianglelefteq u$.

Definition 1 *A language L is* $\trianglelefteq$-continuous[1] *if $u \trianglelefteq v$, $u \trianglelefteq w$, and $v \trianglelefteq w$ with $u, w \in L$ imply $v \in L$. It is* $\trianglelefteq$-free *if $v \triangleleft w$ and $w \in L$ imply $v \notin L$. It is* $\trianglelefteq$-closed *if $v \trianglelefteq w$ and $w \in L$ imply $v \in L$. It is* $\trianglerighteq$-closed *if $v \trianglerighteq w$ and $w \in L$ imply $v \in L$.*

If the binary relation is understood, we call a language simply *continuous, free, closed,* or *converse-closed*. Notice that $\trianglelefteq$-free and $\trianglelefteq$-closed languages are two extreme special cases of $\trianglelefteq$-continuous languages at the opposite ends of the continuous spectrum. Note also that a language is $\trianglerighteq$-continuous if and only if it is $\trianglelefteq$-continuous. Similarly, a language is $\trianglerighteq$-free if and only if it is $\trianglelefteq$-free. Hence we get nothing new by considering the converse relation in these two cases. In the third case, of $\trianglerighteq$-closed languages, we do get a new class.

There is an extensive literature on codes characterized as antichains with respect to binary relations in free monoids; see, for example, [7, 9, 13] and the references contained therein. It is not our purpose in this paper to deal with this topic in depth, but only to point out how various classes of these languages fit into the framework of continuous languages, and to study the closure properties of continuous languages.

We use the following terminology and notation. If $u, v, w \in \Sigma^*$ and $w = uv$, then u is a *prefix* of w and v is a *suffix* of w. If v is a prefix of w, we write $v \leq w$; if also $v \neq w$, then $v < w$. If v is a suffix of w, we write $v \preceq w$; if also $v \neq w$, then $v \prec w$. If $w = xvy$ for some $v, x, y \in \Sigma^*$, then v is a *factor* of w. Note that a prefix or suffix of w is also a factor of w. If v is a factor of w, we write $v \sqsubseteq w$; if also $v \neq w$, then $v \sqsubset w$. If $w = w_0 a_1 w_1 \cdots a_n w_n$, where $a_1, \ldots, a_n \in \Sigma$, and $w_0, \ldots, w_n \in \Sigma^*$, then $v = a_1 \cdots a_n$ is a *subword* of w; note that every factor of w is a subword of w.[2] If v is a subword of w, we write $v \models w$; if also $v \neq w$, then $v \vdash w$. The relations $\leq$, $\preceq$, $\sqsubseteq$, and $\models$ are partial orders on Σ^*.

We apply Definition 1 to four special cases:

$\trianglelefteq$ is $\leq$: If we use the relation 'is a prefix of', then we get prefix-continuous languages [5]. Prefix-free languages, except $\{\epsilon\}$, are prefix codes [3], prefix-closed

---

[1] Languages continuous with respect to a partial order have been called 'convex' in [13].

[2] The word 'subword' is often used to mean 'factor'; here by a 'subword' of w we mean a subsequence of w.

languages[3] are complements of right ideals, and converse-closed languages are the right ideals (have the form $L\Sigma^*$, $L \subseteq \Sigma^*$; see Proposition 8).

$\trianglelefteq$ **is** $\preceq$**:** If we use the relation 'is a suffix of', then we get the suffix-continuous languages. Suffix-free languages, except $\{\epsilon\}$, are suffix codes [3], suffix-closed languages are complements of left ideals, and converse-closed languages are the left ideals (Σ^*L; see Proposition 8).

$\trianglelefteq$ **is** $\sqsubseteq$**:** If we use the relation 'is a factor of',[4] we get factor-continuous languages. Factor-free languages, except $\{\epsilon\}$, are infix codes [9, 13], factor-closed languages are factorial languages [10], which are complements of two-sided ideals, and converse-closed languages are the ideals ($\Sigma^*L\Sigma^*$; see Proposition 7).

$\trianglelefteq$ **is** $\models$**:** If the relation is 'is a subword of',[5] we get subword-continuous languages. Subword-free languages, except $\{\epsilon\}$, are hypercodes [9, 13], subword-closed languages are of the form $L = \overline{K} = \overline{\bigcup_{a_1\cdots a_i \in L} \Sigma^* a_1 \Sigma^* \cdots a_i \Sigma^*}$, and converse-closed languages are of the form K above (see Proposition 9).

$\leq$ **and** $\preceq$**:** If a language is both prefix- and suffix-continuous it is *bifix-continuous*. If it is both prefix- and suffix-free it is *bifix-free*; it is then a bifix code.[6] If it is both prefix- and suffix-closed, it is *bifix-closed*.

The remainder of the paper is structured as follows. In Section 2 we show the relations among the prefix-continuous and suffix-continuous classes of languages and their subclasses. In Section 3 we study the closure properties of the X-continuous, X-closed and X-free classes of languages, where X stands for prefix, suffix, bifix, factor or subword. All three of these types of classes are closed under intersection, and the X-closed languages are closed under union. The prefix (suffix) classes are closed under left (right) quotient, and the subword classes are closed under both types of quotients. All classes are closed under inverse homomorphism. The closure properties of X-converse-closed classes are the same as those of the X-closed classes, as is shown in Section 4. Closure under concatenation is studied in Section 5: all the X-free and X-closed classes are closed under concatenation.

---

[3] Languages closed under the taking of nonempty prefixes and suffixes have been called 'prefixial' and 'suffixial', respectively in [1].

[4] This is called the 'infix order' in [7, 9, 13].

[5] This order is called the 'embedding order' in [7, 9, 13].

[6] The word 'bifix' is sometimes used to describe a word that is both a prefix and a suffix. Here we follow [8, 13]. The term 'biprefix' is used in [3].

2 Continuous Languages

For convenience, we first consider $\trianglelefteq$-continuous, $\trianglelefteq$-free, and $\trianglelefteq$-closed languages, where $\trianglelefteq$ ranges over $\{\leq, \preceq, \sqsubseteq, \models\}$. If a nonempty language is prefix-continuous (respectively, suffix-, bifix-, factor-, or subword-continuous), then it is prefix-closed (respectively, suffix-, bifix-, factor-, or subword-closed) if and only if it contains ϵ. The empty language $\emptyset$ and the language $\{\epsilon\}$ vacuously satisfy the $\trianglelefteq$-continuous, $\trianglelefteq$-free, and $\trianglelefteq$-closed conditions. Also, since ϵ is a prefix, suffix, factor, and subword of every word, $\emptyset$ and $\{\epsilon\}$ are the only two languages that are both $\trianglelefteq$-free and $\trianglelefteq$-closed.

We use the term "factor-closed" to keep our terminology consistent. However, these languages are known as *factorial* languages. Factorial languages are defined as factor-closed languages, for example, in [1, 10], and as bifix-closed languages, for example, in [11]. This is justified in view of the following:

Remark 1 *A language is factor-closed if and only if it is bifix-closed.*

Proof: If L is factor-closed, then it is also bifix-closed, since every prefix and suffix is also a factor. Conversely, let L be a bifix-closed language and let $w \in L$. Suppose v is any factor of $w = xvy$; then $xv \in L$ since xv is a prefix of w, and $v \in L$ because v is a suffix of xv. Therefore L is factor-closed. $\square$

Factorial languages have received considerable attention. For example, their decompositions are studied in [1], their combinatorial properties in [10], and their complexity issues in [12]. We return to these languages later.

Figure 1 shows the various classes of languages partially ordered under set containment, where P, S, B, F, and W, stand for prefix, suffix, bifix, factor, and subword, respectively, PC, PF and PCL stand for prefix-continuous, prefix-free, and prefix-closed languages, *etc.* The classes in rectangular boxes are closed under concatenation; we discuss this later.

Proposition 1 *All containments shown in Fig. 1 are proper, and there are no other containments, except those implied by transitivity.*

Proof: First, we verify that the containments shown do indeed hold. Any class of the form BX, where $X \in \{C, CL, F\}$ is the intersection of PX and SX, by definition. Also, $BX \supseteq FX$, because every prefix and suffix is a factor, and $FX \supseteq WX$, because every factor is a subword. This explains the solid lines. Next, for $Y \in \{P, S, B, F, W\}$, classes YCL and YF are special cases of YC; this accounts for the dotted lines.

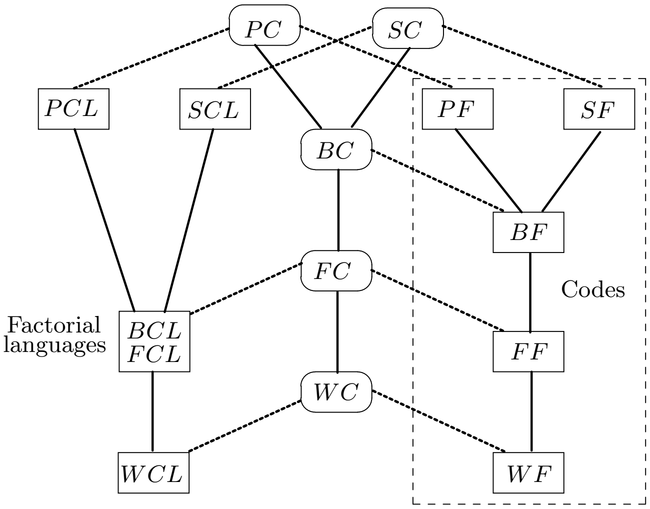

Figure 1: Classes of continuous languages.

Second, we show that no class contains any other class except as shown, or implied by transitivity of set containment. We consider each class in turn, starting with the maximal ones.

The prefix-continuous class PC**:** It suffices to show that PC contains neither SCL nor SF. We have $L_1 = \{\epsilon, a, ba\} \in SCL \setminus PC$, and $L_2 = \{a, abb\} \in SF \setminus PC$.

The suffix-continuous class SC**:** Use left-right symmetry with PC.

The prefix-closed class PCL**:** It suffices to show that PCL contains neither SCL nor WF. Since PC does not contain SCL, neither does its subclass PCL. Also, $L_3 = \{a, b\} \in WF \setminus PCL$.

The suffix-closed class SCL**:** Use left-right symmetry with PCL.

The prefix-free class PF**:** It suffices to show that PF contains neither SF nor WCL. Since PC does not contain SF, neither does PF. Also, $L_4 = \{\epsilon, a\} \in WCL \setminus PF$.

The suffix-free class SF**:** Use left-right symmetry with PF.

The bifix-continuous class BC**:** It suffices to show that BC does not contain any class from $\{PCL, SCL, PF, SF\}$. This follows because $L_5 = \{\epsilon, a, ab\} \in$

$PCL \setminus BC$, $L_1 = \{\epsilon, a, ba\} \in SCL \setminus BC$, $L_6 = \{b, aab\} \in PF \setminus BC$, and $L_2 = \{a, abb\} \in SF \setminus BC$.

The bifix-free class BF: It suffices to show that BF does not contain any class from $\{WCL, PF, SF\}$. We have $L_4 = \{\epsilon, a\} \in WCL \setminus BF$, $L_6 = \{b, aab\} \in PF \setminus BF$, and $L_2 = \{a, abb\} \in SF \setminus BF$.

The factor-continuous class FC: It suffices to show that FC does not contain any class from $\{PCL, SCL, BF\}$. Since BC does not contain PCL, or SCL, neither does FC. Also, $L_7 = \{b, aba\} \in BF \setminus FC$.

The bifix-closed class BCL: It suffices to show that BCL does not contain any class from $\{PCL, SCL, WF\}$. Since FC does not contain PCL or SCL, neither does BCL. Also, $L_8 = \{a\} \in WF \setminus BCL$.

The factor-free class FF: It suffices to show that FF does not contain any class from $\{WCL, BF\}$. Since BF does not contain WCL, neither does FF. Since FC does not contain BF, neither does FF.

The subword-continuous class WC: It suffices to show that WC contains neither BCL nor FF. We have $L_9 = \{\epsilon, a, b, ab, ba, aba\} \in BCL \setminus WC$, and $L_{10} = \{aa, abba\} \in FF \setminus WC$.

The subword-closed class WCL: It suffices to show that WCL contains neither BCL nor WF. Since WC does not contain BCL, neither does WCL. Also, $L_8 = \{a\} \in WF \setminus WCL$.

The subword-free class WF: It suffices to show that WF contains neither WCL nor FF. Since FF does not contain WCL, neither does WF. Also, $L_{10} = \{aa, abba\} \in FF \setminus WF$. □

Remark 2 $PC \cap SCL = PCL \cap SCL = BCL = SC \cap PCL$.

Proof: By definition, $BCL = PCL \cap SCL$. From Fig. 1, we have $PC \cap SCL \supseteq BCL$. Conversely, if L is suffix-closed, then it contains ϵ, which is also a prefix of every word; thus, if L is also prefix-continuous, then it is prefix-closed, and hence bifix-closed. The last equality follows by left-right symmetry. □

2.1 One-Letter Alphabets

The length of a word $w \in \Sigma^*$ is $|w|$, and w^R is the reverse of w. The reverse of L is $L^R = \{w^R \mid w \in L\}$.

Languages over one-letter alphabets have very special properties. Note that, if $L \subseteq \{a\}^*$, then $L = L^R$. Also, the statements 'u is a prefix of w', 'u is a suffix of w', 'u is a factor of w', and 'u is a subword of w' are all equivalent to each other and to '$|u| \leq |w|$'. Thus the following are easily verified:

Proposition 2 *If $\Sigma = \{a\}$, and $L \subseteq \Sigma^*$, then the following hold:*

1. *If X stands for 'prefix', 'suffix', 'bifix', 'factor', or 'subword', then all the statements of the form X-continuous are equivalent, all the statements of the form X-free are equivalent, and all the statements of the form X-closed are equivalent.*
2. *L is prefix-continuous if and only if it is empty, or has the form $\{a^i \mid m \leq i \leq m+n\}$, or $\{a^i \mid m \leq i\}$, for some $m \geq 0, n \geq 0$.*
3. *L is prefix-closed if and only if it is empty, or has the form $\{a^i \mid 0 \leq i \leq m\}$, for some $m \geq 0$, or $\{a^i \mid 0 \leq i\}$.*
4. *L is prefix-free if and only if it is empty, or contains only one word.*
5. *If $K, L \subseteq \Sigma^*$ are prefix-continuous, then so is KL.* $\square$

3 Closure in $\trianglelefteq$-Continuous Languages

We first consider the closure properties of continuous, free, and closed classes of languages. Converse-closed classes are studied in Section 4.

Proposition 3 *If $K, L \subseteq \Sigma^*$ are $\trianglelefteq$-continuous ($\trianglelefteq$-free, or $\trianglelefteq$-closed), then so is $M = K \cap L$.*

Proof: If M is not $\trianglelefteq$-continuous, there exist $u, w \in M$ and $v \notin M$ such that $u \lhd v$, $u \lhd w$, and $v \lhd w$. Since $u, w \in K$ and $u, w \in L$, and K and L are $\trianglelefteq$-continuous, we have $v \in K$ and $v \in L$, which contradicts that $v \notin M$.

If M is not $\trianglelefteq$-free, there exist $v, w \in M$ such that $v \lhd w$. Since $v, w \in K$, this contradicts that K is $\trianglelefteq$-free.

If M is not $\trianglelefteq$-closed, there exist $w \in M$, $v \notin M$ such that $v \lhd w$. Then either $v \notin K$ or $v \notin L$. In the first case, $w \in K$ and $v \notin K$ contradicts that K is $\trianglelefteq$-closed. In the second case, L cannot be $\trianglelefteq$-closed. $\square$

Corollary 1 *All the classes in Fig. 1 are closed under intersection.*

The following is easily verified:

Proposition 4 *If $K, L \subseteq \Sigma^*$ are $\unlhd$-closed, then so is $K \cup L$.*

Corollary 2 *All the closed classes, PCL, SCL, $BCL = FCL$, and WCL, are closed under union.*

The remaining classes in Fig. 1 are not closed under union. Let $K = \{\epsilon\}$, $L = \{aa\}$; both languages are X-continuous and X-free for all $X \in \{P, S, B, F, W\}$. However, $K \cup L$ is neither X-continuous nor X-free.

None of the classes is closed under complementation. The language $L = \{a\}$ is in XC for all $X \in \{P, S, B, F, W\}$, but its complement is not. Also, L is in XF, but $\overline{L}$ is not. The language $K = \{\epsilon\}$ is in XCL, but $\overline{K}$ is not.

If $x \in \Sigma^*$ and $L \subseteq \Sigma^*$, then the *left quotient* of L by x is $x^{-1}L = \{w \in \Sigma^* \mid xw \in L\}$. The *right quotient* of L by x is $Lx^{-1} = \{w \in \Sigma^* \mid wx \in L\}$.

A binary relation is *left-invariant* (*right-invariant*) if $u \unlhd v$ implies $xu \unlhd xv$ ($ux \unlhd vx$).[7]

Proposition 5 *If $\unlhd$ is left-invariant, and L is $\unlhd$-continuous ($\unlhd$-free or $\unlhd$-closed), then $M = x^{-1}L$ is $\unlhd$-continuous ($\unlhd$-free or $\unlhd$-closed), for any $x \in \Sigma^*$. The same holds if 'left' is replaced by 'right' and '$x^{-1}L$' by 'Lx^{-1}'.*

Proof: Suppose L is $\unlhd$-continuous. If M is not $\unlhd$-continuous, then there exist $u, w \in M$ and $v \notin M$ such that $u \lhd v$, $u \lhd w$, and $v \lhd w$. If $\unlhd$ is left-invariant, then $xu \lhd xv$, $xu \lhd xw$, and $xv \lhd xw$, and xu and $xw \in L$, while $xv \notin L$. This contradicts that L is $\unlhd$-continuous.

Suppose L is $\unlhd$-free. If M is not $\unlhd$-free, there exist $v, w \in M$ such that $v \lhd w$; then $xv, xw \in L$. If $\unlhd$ is left-invariant, then $xv \lhd xw$, which contradicts that L is $\unlhd$-free.

Suppose L is $\unlhd$-closed. If M is not $\unlhd$-closed, there exist $w \in M$, $v \notin M$ such that $v \lhd w$; then $xw \in L$ and $xv \notin L$. If $\unlhd$ is left-invariant, then $xv \lhd xw$, which contradicts that L is $\unlhd$-closed.

The claim for the case where $\unlhd$ is right-invariant follows by duality. $\square$

Corollary 3 *The classes PC, PCL and PF are closed under left quotient, SC, SCL and SF are closed under right quotient, and WC, WCL and WF are closed under both quotients.*

Remark 3 *The classes BC, BF, FC, FCL and FF are not closed under either type of quotient. For let $L = \{\epsilon, a, b, ab, ba, aba\}$; then L is bifix-continuous, factor-continuous and factor-closed, but $a^{-1}L = \{\epsilon, b, ba\}$ and La^{-1} are not. Also, $L = \{bb, bab\}$ is bifix-free and factor-free, but $b^{-1}L = \{b, ab\}$ and Lb^{-1} are neither.*

---

[7] The terms 'left compatible' and 'right compatible' are used in [9, 13].

If S is a set, then 2^S is the set of all subsets of S. Let Σ and Δ be alphabets. A *homomorphism* is a map $h : \Sigma^* \to \Delta^*$ such that $h(uv) = h(u)h(v)$ for all $u, v \in \Sigma^*$. If $L \subseteq \Sigma$, then $h(L) = \bigcup_{w \in L}\{h(w)\}$. The *inverse homomorphism* of h is $h^{-1} : h(\Sigma^*) \to 2^{\Sigma^*}$ defined by $h^{-1}(x) = \{w \in \Sigma^* \mid h(w) = x\}$, for all $x \in h(\Sigma^*)$. If $L \subseteq h(\Sigma^*)$, then the inverse image of L under h is $h^{-1}(L) = \{w \in \Sigma^* \mid h(w) \in L\}$. A *substitution* is a map $s : \Sigma^* \to 2^{\Delta^*}$ such that $s(\epsilon) = \{\epsilon\}, s(uv) = s(u)s(v)$ for all $u, v \in \Sigma^*$, and $s(L) = \bigcup_{w \in L}\{s(w)\}$.

None of the classes is closed under homomorphism. If $\Sigma = \Delta = \{a\}$, $h(a) = aa$, $L = \{\epsilon, a\}$, then $h(L) = \{\epsilon, aa\}$, L is in XC and in XCL, for all $X \in \{P, S, B, F, W\}$, but $h(L)$ is not. Also, if $L = \{a, b\}$, $h(a) = \epsilon$, $h(b) = a$, then $h(L) = \{\epsilon, a\}$. Now L is in XF, but $h(L)$ is not. It follows that none of the classes is closed under substitution.

Let $\trianglelefteq$ be a binary relation on Σ^*, and $\trianglelefteq'$, a binary relation on Δ^*. Then h is a *relation homomorphism*[8] if $u \trianglelefteq v$ implies $h(u) \trianglelefteq' h(v)$.

Proposition 6 *Let $(\Sigma^*, \trianglelefteq)$ and $(\Delta^*, \trianglelefteq')$ be free monoids with binary relations, let $h : \Sigma^* \to \Delta^*$ be a relation homomorphism, and let $K \subseteq h(\Sigma^*)$. If K is $\trianglelefteq'$-continuous ($\trianglelefteq'$-free, or $\trianglelefteq'$-closed), then $L = h^{-1}(K)$ is $\trianglelefteq$-continuous ($\trianglelefteq$-free, or $\trianglelefteq$-closed).*

Proof: Suppose K is $\trianglelefteq'$-continuous, but L is not $\trianglelefteq$-continuous. Then there exist $u, w \in L$, $v \notin L$ such that $u \triangleleft v$, $u \triangleleft w$, and $v \triangleleft w$. Since h is a relation homomorphism, we also have $h(u), h(w) \in K$, $h(v) \notin K$, and $h(u) \triangleleft' h(v)$, $h(u) \triangleleft' h(w)$, and $h(v) \triangleleft' h(w)$, which contradicts that K is $\trianglelefteq'$-continuous.

Suppose K is $\trianglelefteq'$-free, but $L = h^{-1}(K)$ is not $\trianglelefteq$-free. Then there exist $v, w \in L$ such that $v \triangleleft w$. Since h is a relation homomorphism, we also have $h(v) \triangleleft' h(w)$, which contradicts that K is $\trianglelefteq'$-free.

Suppose K is $\trianglelefteq'$-closed, but $L = h^{-1}(K)$ is not $\trianglelefteq$-closed. Then there exist $w \in L$, $v \notin L$ such that $v \triangleleft w$. If h is a relation homomorphism, then $h(w) \in K$, $h(v) \notin K$, and $h(v) \triangleleft' h(w)$, which contradicts that K is $\triangleleft'$-closed. $\square$

Corollary 4 *All the classes in Fig. 1 are closed under inverse homomorphism.*

Proof: If u is a prefix (suffix, factor, or subword) of v and h is a homomorphism, then $h(u)$ is a prefix (suffix, factor, or subword) of $h(v)$. Thus, in all cases we have a relation homomorphism. $\square$

---

[8] In the terminology of [7], the relation $\trianglelefteq$ is *compatible* with h (in the case where $\trianglelefteq = \trianglelefteq'$).

4 Converse-Closed Languages

We now consider the remaining continuity property: converse-closure. The following result is proved in [10]:

Proposition 7 *A language L is factorial (that is, factor-closed) if and only if it is the complement of a two-sided ideal, that is, if and only if $L = \overline{\Sigma^* K \Sigma^*}$, for some language K. Moreover, K can be taken to be regular if L is regular.*

We have analogous results for prefix-closed and suffix-closed languages; we include the proof for completeness.

Proposition 8 *A language L is prefix-closed (suffix-closed) if and only if it is the complement of a right (left) ideal, that is, if and only if $L = \overline{K\Sigma^*}$, ($L = \overline{\Sigma^* K}$) for some language K. Moreover, K can be taken to be regular if L is regular.*

Proof: The proof parallels the proof of Proposition 7 in [10]. Let $P(L)$ be the set of all prefixes of words in L; thus, if L is prefix-closed, then $L = P(L)$. Now let $K = \overline{P(L)}$. One verifies that $u \in K$ implies $uv \in K$ for all $v \in \Sigma^*$, that is, $K = K\Sigma^*$, and $L = P(L) = \overline{K} = \overline{K\Sigma^*}$. Note that $\overline{P(L)}$ is regular if L is regular. Conversely, suppose $L = \overline{K\Sigma^*}$ for some K, $w = uv \in L$, and $u \notin L$. Then $u \in K\Sigma^*$, $u = u'u''$, for some $u' \in K$, $u'' \in \Sigma^*$, and $w = u'u''v$ must also be in $K\Sigma^*$, which is a contradiction. Thus L is prefix-closed.

A dual argument proves the result for suffix-closed languages. □

Proposition 9 *A language L is subword-closed if and only if it is the complement of a language of the form $M = \bigcup_{a_1 \cdots a_i \in K} \Sigma^* a_1 \Sigma^* \cdots a_i \Sigma^*$, for some language K. Moreover, K can be taken to be regular if L is regular.*

Proof: The proof also parallels the proof of Proposition 7 in [10]. Let $W(L)$ be the set of all subwords of words in L; thus, if L is subword-closed, then $L = W(L)$. Now let $K = \overline{W(L)}$. For $a_1, \ldots, a_i \in \Sigma$, $a_1 \cdots a_i \in K$ implies $w_0 a_1 w_1 \cdots a_i w_i \in K$ for all $w_0, \ldots, w_i \in \Sigma^*$, that is, $K = \bigcup_{a_1 \cdots a_i \in K} \Sigma^* a_1 \Sigma^* \cdots a_i \Sigma^* = M$, and $L = W(L) = \overline{K}$. Note that $W(L)$ is regular if L is regular. Conversely, suppose $L = \overline{M} = \overline{\bigcup_{a_1 \cdots a_i \in K} \Sigma^* a_1 \Sigma^* \cdots a_i \Sigma^*}$ for some K, $w = w_0 b_1 w_1 \cdots b_n w_n \in L$, and $v = b_1 \cdots b_n \notin L$, for $w_0, \ldots, w_n \in \Sigma^*$ and $b_1, \ldots, b_n \in \Sigma$. Then $v \in M$ and v has a subword, say $u \in K$. Hence w also has u as a subword, and $w \in M$, which is a contradiction. □

For example, let $K = \{aa\}$, and $M = \Sigma^* a \Sigma^* a \Sigma^*$. Then $\overline{M} = \epsilon + a + b^* + b^* a b^*$ is subword-closed.

Proposition 10 *A language L is $\trianglerighteq$-closed if and only if it is the complement of a $\trianglelefteq$-closed language.*

Proof: Suppose L is $\trianglerighteq$-closed; then $v \triangleright w$ and $w \in L$ implies $v \in L$. Thus $v \triangleright w$ and $v \notin L$ implies $w \notin L$. Equivalently, $w \triangleleft v$ and $v \in \overline{L}$ implies $w \in \overline{L}$, that is, $\overline{L}$ is $\trianglelefteq$-closed. Similarly, if $\overline{L}$ is $\trianglelefteq$-closed, then L is $\trianglerighteq$-closed. $\square$

Note that the languages $\emptyset$ and Σ^* are both $\trianglelefteq$-closed and $\trianglerighteq$-closed.

Propositions 7–9 provide characterizations of $\trianglelefteq$-closed languages for the cases where $\trianglelefteq$ is $\leq$, $\preceq$, $\sqsubseteq$, and $\models$.

For $X \in \{P, S, F, W\}$, let XCC be the the class of converse-closed languages corresponding to the prefix, suffix, factor, and subword relations, respectively. Similarly, let XC represent the continuous classes and XCL, the closed classes.

Remark 4 *If $L \subseteq \Sigma^*$ is $\trianglerighteq$-closed, then it is $\trianglelefteq$-continuous.*

Proof: This follows, because $\trianglerighteq$-closure is a special case of $\trianglerighteq$-continuity which coincides with $\trianglelefteq$-continuity. $\square$

Corollary 5 *We have $XCC \subseteq XC$ for all $X \in \{P, S, F, W\}$.*

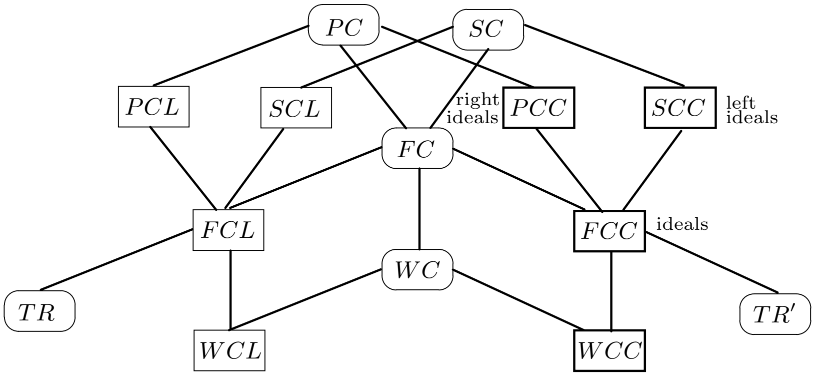

Figure 2: Classes of converse-closed languages.

The classes XCC in Fig. 2 are the converse-closed classes. (We explain TR and TR' later.) Each converse-closed class $XCC = \{\overline{L} \mid L \in XCL\}$ is in 1-1 correspondence with the corresponding closed class. Note that each class XC contains languages that are not in $XCL \cup XCC \cup XF$. For example, $\{a, aa\}$ is in XC but it is not in $XCL \cup XCC \cup XF$, for all $X \in \{P, S, F, W\}$.

Proposition 11 *If $K, L \subseteq \Sigma^*$ are $\trianglerighteq$-closed, then so are $K \cap L$ and $K \cup L$. If $\trianglerighteq$ is left-invariant, and L is $\trianglerighteq$-closed, then $x^{-1}L$ is $\trianglerighteq$-closed, for any $x \in \Sigma^*$. The same holds if 'left' is replaced by 'right' and '$x^{-1}L$' by 'Lx^{-1}'. Let $(\Sigma^*, \trianglelefteq)$ and $(\Delta^*, \trianglelefteq')$ be free monoids with binary relations, let $h : \Sigma^* \to \Delta^*$ be a relation homomorphism, and let $K \subseteq h(\Sigma^*)$. If K is $\trianglerighteq$-closed, then $h^{-1}(K)$ is $\trianglerighteq$-closed.*

Proof: This follows by Propositions 3–6. □

Corollary 6 *All the classes of the form XCC are closed under intersection, union, and inverse homomorphism. Moreover, PCC is closed under left quotient, SCC is closed under right quotient, and WCC, under both.*

Note that FCC is not closed under either quotient. Let $\trianglelefteq$ be $\sqsubseteq$, let $\Sigma = \{a, b\}$, and let $L = \Sigma^*aba\Sigma^*$. Then L is $\sqsupseteq$-closed, but $K = a^{-1}L = \Sigma^*aba\Sigma^* + ba\Sigma^*$ is not, because $ba \in K$, but $bba \notin K$. Symmetrically, La^{-1} is not $\sqsupseteq$-closed.

No class XCC is closed under homomorphism. For let $\Sigma = \Delta = \{a, b\}$, $h(a) = h(b) = b$, and $L = \{\epsilon, a\}$. Then $L \in XCL$ and $\overline{L} = (b + aa + ab)\Sigma^* = \Sigma^*(b + aa + ab) = \Sigma^*(b + aa + ab)\Sigma^* = \Sigma^*b\Sigma^* + \Sigma^*a\Sigma^*a\Sigma^* + \Sigma^*a\Sigma^*b\Sigma^* \in XCC$, for all $X \in \{P, S, F, W\}$. However, $h(\overline{L}) = bb^*$, and $K = \overline{h(\overline{L})} = \epsilon + \Sigma^*a\Sigma^*$ is not in XCL, since $b \notin K$.

Remark 5 *All the classes of the form XCC are closed under concatenation, because we have $(L\Sigma^*)(K\Sigma^*) = (L\Sigma^*K)\Sigma^*$, etc.*

4.1 Transitive Sofic Languages

Factorial languages contain a very interesting subclass that deserves to be mentioned. For more details we refer the reader to the literature [2, 3, 11]. A language $M \subseteq \Sigma^*$ is a *monoid* if it contains ϵ and is closed under concatenation. A monoid is *very pure* if $uv, vu \in L$ implies $u, v \in L$. A factorial language is called *sofic* if it is regular. A language L is *transitive* if for all $u, w \in L$, there exists $x \in \Sigma^*$ such that $v = uxw \in L$. Let $F(L)$ be the set of all factors of words in L.

Transitive sofic languages constitute the class TR in Fig. 2, and TR' is the class of their complements. The following characterization is given in [2]:

Proposition 12 *A language L is sofic and transitive if and only if there exists a very pure regular language M which is a monoid such that $L = F(M)$.*

Example 1 *Let $\Sigma^* = \{a, b, c\}$, let $M = (ab^*c + b)^*$, and let $L = F(M)$. One verifies that $L = (\epsilon + b^*c)(b + ab^*c)^*(\epsilon + ab^*) = \overline{\Sigma^*(ab^*a + cb^*c)\Sigma^*}$. Here the language $G = ab^*c + b$ is a circular code [3] and is a minimal generating set of M. The monoid $M = G^*$ is very pure, and $L = F(M)$ is transitive.*

Proposition 13 *Let $h : \Sigma^* \to \Delta^*$ be a homomorphism, let $K \subseteq h(\Sigma^*)$ and let $L = h^{-1}(K)$. If K is a transitive sofic language then so is L.*

Proof: Since K is regular, so is L, since regular languages are closed under inverse homomorphism. Suppose that u and w are in L, and let $h(u) = x$, $h(w) = z$. Since K is transitive, for every $x, z \in K$ there exists $y \in \Delta^*$ such that xyz is in K. Since K is factorial, we also have $y \in K$. Hence there exists $v \in L$ such that $h(v) = y$. Since $h(uvw) = h(u)h(v)h(w) = xyz \in K$, we also have $uvw \in L$, and we have shown that L is transitive. Finally, if $uvw \in L$ and $v \notin L$, then $h(uvw) \in K$ and $h(v) \notin K$, contradicting that K is factorial. Hence L is also factorial. Altogether, L is transitive sofic. □

Transitive sofic languages are not closed under either quotient, intersection, union, complement and concatenation. Let $\Sigma = \{a, b, c, d, e\}$, let L be the transitive sofic language L of Example 1, and let K be a similar language, $K = (\epsilon + e^*c)(e + de^*c)^*(\epsilon + de^*)$. Then $L \cap K = \epsilon + c$, which is not transitive. Also, for the language L of Example 1, $cac \in a^{-1}L$, but $a \notin a^{-1}L$; hence $a^{-1}L$ is not factorial. Moreover, let $\Sigma = \{a, b\}$, $K = a^*$, and $L = b^*$. Then K and L are transitive, but $K \cup L$ and KL are not. We have $a, b \in K \cup L$, but there is no $x \in \Sigma^*$ such that $axb \in K \cup L$. Also, $ab \in KL$, but there is no $x \in \Sigma^*$ such that $abxab \in L$. The complement of L is not factorial, since $\epsilon \notin L$.

5 Concatenation in Free and Closed Languages

The next example illustrates that, in general, $\trianglelefteq$-closed and $\trianglelefteq$-free languages are not closed under concatenation.

Example 2 *Suppose $u \trianglelefteq v$ if and only if either $u = v$ or $|u| = |v|$ and u precedes v in the lexicographic order. Thus, for $\Sigma = \{a, b\}$, we have $a \lhd b$, $aa \lhd ab \lhd ba \lhd bb$, $aaa \lhd aab \lhd aba \lhd \cdots \lhd bbb$, etc. Let $K = \{a, bb\}$; then K is $\trianglelefteq$-free. However, $KK = \{aa, abb, bba, bbbb\}$ is not. Also, if $L = \{aa, ab\}$, then L is $\trianglelefteq$-closed. However, $LL = \{aaaa, aaab, abaa, abab\}$ is not. Hence, for this binary relation, $\trianglelefteq$-closed and $\trianglelefteq$-free languages are not closed under concatenation.* □

A binary relation $\trianglelefteq$ is *propagating* if $x_1x_2 \lhd y_1y_2$ implies that

$$(x_1 \lhd y_1) \vee (y_1 \lhd x_1) \vee (x_2 \lhd y_2) \vee (y_2 \lhd x_2),$$

for all $x_1, x_2, y_1, y_2 \in \Sigma^*$, where $\vee$ denotes disjunction.

Proposition 14 *If $\trianglelefteq$ is propagating, and K and L are $\trianglelefteq$-free, then so is KL.*

Proof: Suppose K and L are $\trianglelefteq$-free, but $M = KL$ is not. Then there are $x_1, y_1 \in K$, $x_2, y_2 \in L$ such that $x_1x_2 \triangleleft y_1y_2$. Since $\trianglelefteq$ is propagating, either x_1 and y_1 are unequal and comparable under $\trianglelefteq$, or x_2 and y_2 are. Thus either K or L is not $\trianglelefteq$-free, which is a contradiction. $\Box$

Lemma 1 *The binary relations $\leq$, $\preceq$, $\sqsubseteq$ and $\models$ are propagating.*

Proof: Suppose $x_1x_2 < y_1y_2$; then $x_1x_2v = y_1y_2$, where $v \in \Sigma^*$ is nonempty. If $x_1 < y_1$ or $x_1 > y_1$, the condition of the lemma is satisfied. If $x_1 = y_1$, then $x_2 < y_2$, and the lemma holds. A symmetric argument works for $\preceq$.

Suppose $x_1x_2 \sqsubset y_1y_2$; then $ux_1x_2v = y_1y_2$, for some $u, v \in \Sigma^*$, where $uv \neq \epsilon$. If $ux_1 < y_1$, then $x_1 \sqsubset y_1$. If $ux_1 > y_1$, then $x_2 \sqsubset y_2$. If $ux_1 = y_1$ and $u \neq \epsilon$, then $x_1 \sqsubset y_1$. If $ux_1 = y_1$ and $u = \epsilon$, then $x_1 = y_1$, and $x_2 \sqsubset y_2$, since $v \neq \epsilon$.

Now suppose that $x_1x_2 \vdash y_1y_2$; then $x_1 = a_1 \cdots a_j$, $x_2 = a_{j+1} \cdots a_n$, for some j, and $y_1 = v_0a_1v_1 \cdots a_iv_i'$ and $y_2 = v_i''a_{i+1}v_{i+1} \cdots a_nv_n$, for some i, where $v_i = v_i'v_i''$, $v_0, \ldots, v_n \in \Sigma^*$, $a_1, \ldots, a_n \in \Sigma$, and $v_1 \cdots v_n \neq \epsilon$. If $j < i$, then $x_1 \vdash y_1$. If $j > i$, then $x_2 \vdash y_2$. If $j = i$, and $v_0v_1 \cdots v_i' \neq \epsilon$, then $x_1 \vdash y_1$. If $j = i$, and $v_0v_1 \cdots v_i' = \epsilon$, then $x_2 \vdash y_2$. $\Box$

Corollary 7 *The prefix-, suffix-, bifix-, factor-, and subword-free classes are closed under concatenation.*

We now consider $\trianglelefteq$-closed languages. A binary relation $\trianglelefteq$ is *factoring* if $x \trianglelefteq y_1y_2$ implies that $x = x_1x_2$ for some $x_1, x_2 \in \Sigma^*$ such that $x_1 \trianglelefteq y_1$, $x_2 \trianglelefteq y_2$.

Proposition 15 *If $\trianglelefteq$ is factoring, and K and L are $\trianglelefteq$-closed, then so is KL.*

Proof: Suppose K and L are $\trianglelefteq$-closed, but $M = KL$ is not. Then there exist $x \notin M$, $y_1 \in K$, $y_2 \in L$ such that $x \triangleleft y_1y_2$. Since $\trianglelefteq$ is factoring, $x = x_1x_2$, where $x_1 \trianglelefteq y_1$ and $x_2 \trianglelefteq y_2$. If K and L are $\trianglelefteq$-closed, then $x_1 \in K$, $x_2 \in L$, and $x \in M$—a contradiction. $\Box$

Lemma 2 *The binary relations $\leq$, $\preceq$, $\sqsubseteq$ and $\models$ are factoring.*

Proof: Suppose $x \leq y_1y_2$; then $xv = y_1y_2$ for some $v \in \Sigma^*$. For $x \leq y_1$, since $\epsilon \leq y_2$, we have $x_1 = x$, and $x_2 = \epsilon$. If $x > y_1$, then $x = x_1x_2$, where $x_1 = y_1$ and $x_2v = y_2$. Then $x_1 \leq y_1$, and $x_2 \leq y_2$. A symmetric argument works for $\preceq$.

Suppose $x \sqsubseteq y_1y_2$; then $uxv = y_1y_2$, for some $u, v \in \Sigma^*$. If $ux \leq y_1$, then $x_1 = x \sqsubseteq y_1$ and $x_2 = \epsilon \sqsubseteq y_2$. If $ux > y_1$ and $u < y_1$, then $x = x_1x_2$, where $ux_1 = y_1$ and $x_2v = y_2$. Then $x_1 \sqsubseteq y_1$, and $x_2 \sqsubseteq y_2$. If $ux > y_1$ and $u \geq y_1$, then $x_1 = \epsilon \sqsubseteq y_1$ and $x_2 = x \sqsubseteq y_2$.

Now suppose that $x \models y_1 y_2 = v$; then $x = a_1 \cdots a_n$ and $v = v_0 a_1 v_1 \cdots a_n v_n$, where $v_0, \ldots, v_n \in \Sigma^*$, $a_1, \ldots, a_n \in \Sigma$, and, for some i we have $y_1 = v_0 a_1 v_1 \cdots a_i v_i'$ and $y_2 = v_i'' a_{i+1} v_{i+1} \cdots a_n v_n$, where $v_i = v_i' v_i''$. If $i = n$, then $x_1 = x \models y_1$ and $x_2 = \epsilon \models y_2$. If $i < n$, then $x = x_1 x_2$, where $x_1 = a_1 \cdots a_i \models y_1$ and $x_2 = a_{i+1} \cdots a_n \models y_2$. □

Corollary 8 *The prefix-, suffix-, bifix- (= factor-), and subword-closed classes are closed under concatenation.*

Remark 6 *If $K, L \subseteq \Sigma^*$ are prefix- (suffix-, bifix-, factor-, or subword-) continuous, then KL, is not necessarily prefix- (suffix-, bifix-, factor-, or subword-) continuous.*

Proof: $K = \{a, ab\}$ and $L = \{b, ab\}$ are prefix-, suffix-, bifix-, factor-, and subword-continuous, but $KL = \{ab, aab, abb, abab\}$ is not, for $aba, bab \notin KL$. □

Before stating our next results, we quote (in our terminology) part of a proposition and a corollary from the theory of codes [3], p. 103.

Proposition 16 *Let Σ be an alphabet, let $K, (L_i)_{i \in I}$ be nonempty subsets of Σ^*, and let $(K_i)_{i \in I}$ be a partition of K. Set $M = \bigcup_{i \in I} K_i L_i$. Then the following are true: (a) If K and the L_i's are prefix-free, then M is prefix-free. (b) If M is prefix-free, then all L_i's are prefix-free.*

Corollary 9 *Let $K \subseteq \Sigma^+$, and $n \geq 1$. Then K is prefix-free if and only if K^n is prefix-free.*

A result similar to Proposition 16 holds for prefix-closed languages:

Proposition 17 *Let Σ be an alphabet, let $K, (L_i)_{i \in I}$ be nonempty subsets of Σ^*, and let $(K_i)_{i \in I}$ be a collection of subsets of K such that $K = \bigcup_{i \in I} K_i$. Let $M = \bigcup_{i \in I} K_i L_i$. If K and the L_i's are prefix-closed, then so is M.*

Proof: If M is not prefix-closed, then there is a word $w \in M$ such that $w = uv$ for some $u, v \in \Sigma^*$, and $u \notin M$. Since $w \in M$, we have $w = xy$, for some $x \in K_i$, $y \in L_i$. First, if $u \leq x$, then $u \in K_j$ for some j, since K is prefix-closed. Since L_j is prefix-closed, $\epsilon \in L_j$, and it follows that $u \in K_j L_j \subseteq M$, which is a contradiction. Second, if $x < u$, then $u = xy'$, $x \in K_i$, $y' \in \Sigma^*$, $y = y'y''$, and $v = y''$. Since $y \in L_i$, $y' \leq y$, and L_i is prefix-closed, we have $y' \in L_i$. Since also $x \in K_i$, we have $y \in M$—a contradiction. □

The analog of Part (b) of Proposition 16 does not hold. Let $\Sigma = \{a\}$, and $K = \Sigma^*$; then K is trivially a partition of itself. Let $L = \{\epsilon, aa\}$. Then $M = KL = \Sigma^*$ is prefix-closed, but L is not.

Of course, if K is prefix-closed, then so is K^n. In general, however, if K^n is prefix-closed, then K need not be. For example, if $K = \epsilon + a(aa)^*$, then $K^n = \Sigma^*$ for all $n \geq 2$, K^n is prefix-closed, and K is not. A finite example is $K = \{\epsilon, a, a^3, a^4\}$; here K^n is prefix-closed for all $n \geq 2$.

6 Conclusions

We have provided a common framework for several classes of languages, and we have shown that closure properties of these classes can be studied using binary relations on Σ^*.

Acknowledgment: We thank Larry Cummings, Helmut Jürgensen and Jeff Shallit for useful comments and pointers to references.

References

[1] S. V. Avgustinovich and A. E. Frid, "A Unique Decomposition Theorem for Factorial Languages", *Int. J. Algebra and Comput.* **15**, 149–160, 2005.

[2] M. P. Béal and D. Perrin, "Une caractérisation des ensembles sofiques", *C. R. Acad. Sci., Paris* **303**, 255–257, 1986.

[3] J. Berstel and D. Perrin, *Theory of Codes,* Academic Press, 1985.

[4] J. A. Brzozowski, "Representation of a Class of Nondeterministic Semiautomata by Canonical Words", *Theoretical Comput. Sci.* **356**, 46–57, 2006.

[5] J. A. Brzozowski and H. Jürgensen, "Representation of Semiautomata by Canonical Words and Equivalences", *Int. J. Foundations of Computer Sci.* **16**, 831–850, 2005.

[6] J. A. Brzozowski and H. Jürgensen, "Representation of Semiautomata by Canonical Words and Equivalences, Part II: Applications to the Specification of Software Modules", *Int. J. Foundations of Computer Sci.* **18**, 1065–1087, 2007.

[7] H. Jürgensen, L. Kari, and G. Thierrin, "Morphisms Preserving Densities." *Int. J. Computer Math.* **78**, 165–189, 2001.

[8] H. Jürgensen and S. Konstantinidis, "Codes." In: *Handbook of Formal Languages,* **1**, G. Rozenberg and A. Salomaa, eds., 511–607, 1997.

[9] H. Jürgensen and S. S. Yu, "Relations on Free Monoids, Their Independent Sets, and Codes." *Int. J. Computer Math.* **40**, 17–46, 1991.

[10] A. de Luca and S. Varricchio, "Some Combinatorial Properties of Factorial Languages", in *Sequences,* R. Capocelli, ed., 258–266, Springer, New York, 1990.

[11] A. Restivo, "Finitely Generated Sofic Systems", *Theoretical Computer Science* **65**, 265–270, 1989.

[12] A. M. Shur, "Factorial Languages of Low Combinatorial Complexity", *Developments in Language Theory,* LNCS **4036**, 397–407, Springer, Berlin, 2006.

[13] H. J. Shyr, *Free Monoids and Languages,* Hon Min Book Co., Taichung, Taiwan, 2001.