

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 [6, 7], 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 [6], 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 [5]. Applications of these methods to the specification of software modules were discussed in [7]. 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 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. Let \preceq be a binary relation on Σ^* ; if $u \preceq v$ and $u \neq v$, we write $u \triangleleft v$. Let \succeq be the converse relation, that is, let $u \succeq v$ if and only if $v \triangleleft u$.

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

Definition 1 A language L is \triangleleft -continuous¹ if $u \triangleleft v$, $u \triangleleft w$, and $v \triangleleft w$ with $u, w \in L$ imply $v \in L$. It is \triangleleft -free if $v \triangleleft w$ and $w \in L$ imply $v \notin L$. It is \triangleleft -closed if $v \triangleleft w$ and $w \in L$ imply $v \in L$. It is \triangleright -closed if $v \triangleright w$ 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 \triangleleft -free and \triangleleft -closed languages are two extreme special cases of \triangleleft -continuous languages at the opposite ends of the continuous spectrum. Note also that a language is \triangleright -continuous if and only if it is \triangleleft -continuous. Similarly, a language is \triangleright -free if and only if it is \triangleleft -free. Hence we get nothing new by considering the converse relation in these two cases. In the third case, of \triangleright -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, [8, 10, 14] 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 $w = w_0a_1w_1 \cdots a_nw_n$, where $a_1, \dots, a_n \in \Sigma$, and $w_0, \dots, w_n \in \Sigma^*$, 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:

- \triangleleft is \leq : If we use the relation ‘is a prefix of’, then we get prefix-continuous languages [6]. Prefix-free languages, except $\{\epsilon\}$, are prefix codes [4], prefix-closed languages³ are complements of right ideals, and converse-closed languages are the right ideals (have the form $L\Sigma^*$, $L \subseteq \Sigma^*$; see Proposition 7).
- \triangleleft 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 [4], suffix-closed languages are complements of left ideals, and converse-closed languages are the left ideals (Σ^*L ; see Proposition 7).
- \triangleleft 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 [10, 14], factor-closed languages are factorial languages [11], which are complements of two-sided ideals, and converse-closed languages are the ideals ($\Sigma^*L\Sigma^*$; see Proposition 6).

¹Languages continuous with respect to a partial order have been called ‘convex’ in [14].

²The word ‘subword’ is often used to mean ‘factor’; here by a ‘subword’ of w we mean a subsequence of w .

³Languages closed under the taking of nonempty prefixes and suffixes have been called ‘prefixial’ and ‘suffixial’, respectively in [2].

⁴This is called the ‘infix order’ in [8, 10, 14].

\triangleleft **is** \models : If the relation is ‘is a subword of’,⁵ we get subword-continuous languages. Subword-free languages, except $\{\epsilon\}$, are hypercodes [10, 14], subword-closed languages are of the form $L = \overline{K} = \overline{\bigcup_{a_1 \dots a_i \in L} \Sigma^* a_1 \Sigma^* \dots a_i \Sigma^*}$, and converse-closed languages are of the form K above (see Proposition 8).

\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.

2 Continuous Languages

For convenience, we first consider \triangleleft -continuous, \triangleleft -free, and \triangleleft -closed languages, where \triangleleft 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 \triangleleft -continuous, \triangleleft -free, and \triangleleft -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 \triangleleft -free and \triangleleft -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 [2, 11], and as bifix-closed languages, for example, in [12]. 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

⁵This order is called the ‘embedding order’ in [8, 10, 14].

⁶The word ‘bifix’ is sometimes used to describe a word that is both a prefix and a suffix. Here we follow [9, 14]. The term ‘biprefix’ is used in [4].

Factorial languages have received considerable attention. For example, their decompositions are studied in [2], their combinatorial properties in [11], and their complexity issues in [13]. 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.

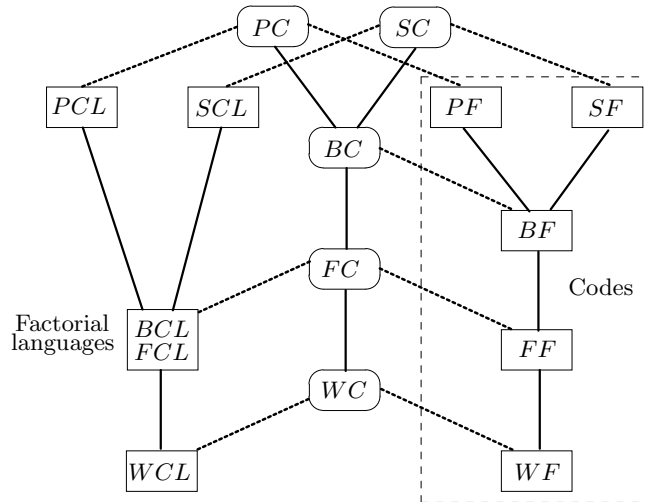


Figure 1: Classes of continuous languages.

The proof of the following result is given in [1]. The special case of languages over a one-letter alphabet is also discussed there.

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

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. \square

3 Closure in \triangleleft -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 2 *If $K, L \subseteq \Sigma^*$ are \triangleleft -continuous (\triangleleft -free, or \triangleleft -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 $u \triangleleft v$, $u \triangleleft w$, and $v \triangleleft w$. 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 $v \triangleleft w$. 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 $v \triangleleft w$. 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 3 *If $K, L \subseteq \Sigma^*$ are \trianglelefteq -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 \bar{L} is not. The language $K = \{\epsilon\}$ is in XCL , but \bar{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 $u \trianglelefteq v$ implies $xu \trianglelefteq xv$ ($ux \trianglelefteq vx$).⁷

Proposition 4 *If \trianglelefteq is left-invariant, and L is \trianglelefteq -continuous (\trianglelefteq -free or \trianglelefteq -closed), then $M = x^{-1}L$ is \trianglelefteq -continuous (\trianglelefteq -free or \trianglelefteq -closed), for any $x \in \Sigma^*$. If \trianglelefteq is right-invariant, and L is \trianglelefteq -continuous (\trianglelefteq -free or \trianglelefteq -closed), then $M = Lx^{-1}$ is \trianglelefteq -continuous (\trianglelefteq -free or \trianglelefteq -closed), for any $x \in \Sigma^*$.*

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

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

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

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

⁷The terms ‘left compatible’ and ‘right compatible’ are used in [10, 14].

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.*

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 $h : \Sigma^* \rightarrow \Delta^*$ 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 $h^{-1} : h(\Sigma^*) \rightarrow 2^{\Sigma^*}$ 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 $s : \Sigma^* \rightarrow 2^{\Delta^*}$ 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 \triangleleft be a binary relation on Σ^* , and \triangleleft' , a binary relation on Δ^* . Then h is a *relation homomorphism*⁸ if $u \triangleleft v$ implies $h(u) \triangleleft' h(v)$.

Proposition 5 *Let $(\Sigma^*, \triangleleft)$ and $(\Delta^*, \triangleleft')$ be free monoids with binary relations, let $h : \Sigma^* \rightarrow \Delta^*$ be a relation homomorphism, and let $K \subseteq h(\Sigma^*)$. If K is \triangleleft' -continuous (\triangleleft' -free, or \triangleleft' -closed), then $L = h^{-1}(K)$ is \triangleleft -continuous (\triangleleft -free, or \triangleleft -closed).*

Proof: Suppose K is \triangleleft' -continuous, but L is not \triangleleft -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 \triangleleft' -continuous.

Suppose K is \triangleleft' -free, but $L = h^{-1}(K)$ is not \triangleleft -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 \triangleleft' -free.

Suppose K is \triangleleft' -closed, but $L = h^{-1}(K)$ is not \triangleleft -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

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

4 Converse-Closed Languages

We now consider the remaining continuity property: converse-closure. The following result is proved in [11]:

Proposition 6 *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:

Proposition 7 *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 6 in [11]. 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. \square

Proposition 8 *A language L is subword-closed if and only if it is the complement of a language of the form $M = \bigcup_{a_1 \dots a_i \in K} \Sigma^* a_1 \Sigma^* \dots a_i \Sigma^*$, 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 6 in [11]. 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 $a_1, \dots, a_i \in \Sigma$, $a_1 \dots a_i \in K$ implies $w_0 a_1 w_1 \dots a_i w_i \in K$ for all $w_0, \dots, w_i \in \Sigma^*$, that is, $K = \bigcup_{a_1 \dots a_i \in K} \Sigma^* a_1 \Sigma^* \dots a_i \Sigma^* = M$, and $L = W(L) = \overline{K}$. Note that $W(L)$ is regular if L is regular. Conversely, suppose $L = \overline{M} = \overline{\bigcup_{a_1 \dots a_i \in K} \Sigma^* a_1 \Sigma^* \dots a_i \Sigma^*}$ for some K , $w = w_0 b_1 w_1 \dots b_n w_n \in L$, and $v = b_1 \dots b_n \notin L$, for $w_0, \dots, w_n \in \Sigma^*$ and $b_1, \dots, b_n \in \Sigma$. 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. \square

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 9 *A language L is \triangleright -closed if and only if it is the complement of a \triangleleft -closed language.*

Proof: Suppose L is \triangleright -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 \triangleleft -closed. Similarly, if \overline{L} is \triangleleft -closed, then L is \triangleright -closed. \square

Note that the languages \emptyset and Σ^* are both \triangleleft -closed and \triangleright -closed.

Propositions 6–8 provide characterizations of \triangleleft -closed languages for the cases where \triangleleft 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 \triangleright -closed, then it is \triangleleft -continuous.*

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

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

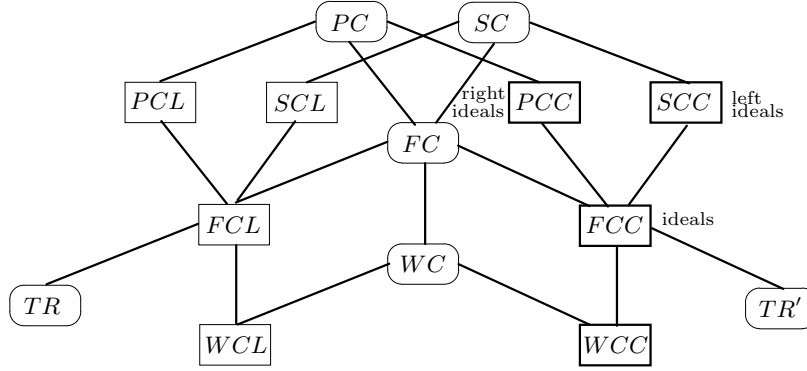


Figure 2: Classes of converse-closed languages.

The classes XCC in Fig. 2 are the converse-closed classes. (We explain TR and TR' later.) Each converse-closed class $XCC = \{\bar{L} \mid L \in XCL\}$ 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 10 *If $K, L \subseteq \Sigma^*$ are \triangleright -closed, then so are $K \cap L$ and $K \cup L$. If \triangleright is left-invariant, and L is \triangleright -closed, then $x^{-1}L$ is \triangleright -closed, for any $x \in \Sigma^*$. The same holds if ‘left’ is replaced by ‘right’ and ‘ $x^{-1}L$ ’ by ‘ Lx^{-1} ’. Let $(\Sigma^*, \triangleleft)$ and $(\Delta^*, \triangleleft')$ be free monoids with binary relations, let $h : \Sigma^* \rightarrow \Delta^*$ be a relation homomorphism, and let $K \subseteq h(\Sigma^*)$. If K is \triangleright' -closed, then $h^{-1}(K)$ is \triangleright -closed.*

Proof: This follows by Propositions 2–5. \square

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 \sqsubseteq -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 \sqsubseteq -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 an interesting subclass that deserves to be mentioned. For details we refer the reader to the literature [3, 4, 12]. 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 [3]. A language L is *transitive* if for all $u, w \in L$, there exists $x \in \Sigma^*$ such that $v = uwx \in L$. 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 [3]:

Proposition 11 *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 $L = (\epsilon + b^*c)(b + ab^*c)^*(\epsilon + ab^*) = \Sigma^*(ab^*a + cb^*c)\Sigma^*$. Here the language $G = ab^*c + b$ is a circular code [4] and is a minimal generating set of M . The monoid $M = G^*$ is very pure, and $L = F(M)$ is transitive.*

Proposition 12 *Let $h : \Sigma^* \rightarrow \Delta^*$ 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 is a $y \in \Delta^*$ such that $xyz \in K$. Since K is factorial, also $y \in K$. So there exists $v \in L$ such that $h(v) = y$. Since $h(uvw) = h(u)h(v)h(w) = xyz \in K$, $uvw \in L$, 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. \square

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 $a \triangleleft b$, $aa \triangleleft ab \triangleleft ba \triangleleft bb$, $aaa \triangleleft aab \triangleleft aba \triangleleft \dots \triangleleft bbb$, 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. \square

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

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

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

Proposition 13 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. \square

Lemma 1 The binary relations \leq , \preceq , \sqsubset and \vdash 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 $v \neq \epsilon$.

If $x_1x_2 \vdash y_1y_2$, 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 $v_i = v'_i v''_i$, $v_0, \dots, v_n \in \Sigma^*$, $a_1, \dots, a_n \in \Sigma$, 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 $x_1 \vdash y_1$; otherwise $x_2 \vdash y_2$. \square

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 $x \trianglelefteq y_1 y_2$ implies that $x = x_1 x_2$ for some $x_1, x_2 \in \Sigma^*$ such that $x_1 \trianglelefteq y_1, x_2 \trianglelefteq y_2$.

Proposition 14 *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 $x \trianglelefteq y_1 y_2$. Since \trianglelefteq is factoring, $x = x_1 x_2$, 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. \square

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

Proof: Suppose $x \leq y_1 y_2$; then $xv = y_1 y_2$ 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 $x = x_1 x_2$, where $x_1 = y_1$ and $x_2 v = y_2$. Then $x_1 \leq y_1$, and $x_2 \leq y_2$. A symmetric argument works for \preceq .

Suppose $x \sqsubseteq y_1 y_2$; then $uxv = y_1 y_2$, 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 $x = x_1 x_2$, where $ux_1 = y_1$ and $x_2 v = y_2$. 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 $x_2 = x \sqsubseteq y_2$.

If $x \models y_1 y_2 = v$, then $x = a_1 \cdots a_n$ and $v = v_0 a_1 v_1 \cdots a_n v_n$, where $v_0, \dots, v_n \in \Sigma^*$, $a_1, \dots, a_n \in \Sigma$, and we have $y_1 = v_0 a_1 v_1 \cdots a_i v'_i$ and $y_2 = v''_i a_{i+1} v_{i+1} \cdots a_n v_n$, for some i , 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 $x_2 = a_{i+1} \cdots a_n \models y_2$. \square

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 $aba, bab \notin KL$. \square

A result similar to Proposition 4.1 in [4] also holds for prefix-closed languages [1]:

Proposition 15 *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 .*

6 Conclusions

We have provided a common framework for several classes of languages, and have shown that their closure properties 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] T. Ang and J. Brzozowski, “Continuous Languages”, Tech. Rep. CS-2008-06, David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada. <http://www.cs.uwaterloo.ca/research/tr/2008/>
- [2] S. V. Avgustinovich and A. E. Frid, “A Unique Decomposition Theorem for Factorial Languages”, *Int. J. Algebra and Comput.* **15**, 149–160, 2005.
- [3] M. P. Béal and D. Perrin, “Une caractérisation des ensembles sofiques”, *C. R. Acad. Sci., Paris* **303**, 255–257, 1986.
- [4] J. Berstel and D. Perrin, *Theory of Codes*, Academic Press, 1985.
- [5] J. A. Brzozowski, “Representation of a Class of Nondeterministic Semiautomata by Canonical Words”, *Theoretical Comput. Sci.* **356**, 46–57, 2006.
- [6] J. A. Brzozowski and H. Jürgensen, “Representation of Semiautomata by Canonical Words and Equivalences”, *Int. J. Foundations Comp. Sci.* **16**, 831–850, 2005.
- [7] 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 Comp. Sci.* **18**, 1065–1087, 2007.
- [8] H. Jürgensen, L. Kari, and G. Thierrin, “Morphisms Preserving Densities.” *Int. J. Computer Math.* **78**, 165–189, 2001.
- [9] H. Jürgensen and S. Konstantinidis, “Codes.” In: *Handbook of Formal Languages*, **1**, G. Rozenberg and A. Salomaa, eds., 511–607, 1997.
- [10] H. Jürgensen and S. S. Yu, “Relations on Free Monoids, Their Independent Sets, and Codes.” *Int. J. Computer Math.* **40**, 17–46, 1991.
- [11] A. de Luca and S. Varricchio, “Some Combinatorial Properties of Factorial Languages”, in *Sequences*, R. Capocelli, ed., 258–266, Springer, 1990.
- [12] A. Restivo, “Finitely Generated Sofic Systems”, *Theoretical Computer Science* **65**, 265–270, 1989.
- [13] A. M. Shur, “Factorial Languages of Low Combinatorial Complexity”, *Developments in Language Theory*, LNCS **4036**, 397–407, Springer, Berlin, 2006.
- [14] H. J. Shyr, *Free Monoids and Languages*, Hon Min Book Co., Taichung, Taiwan, 2001.