Acta Cybernetica **19** (2009) 445–464.

Languages Convex with Respect to Binary Relations, and Their Closure Properties*

Thomas Ang† and Janusz Brzozowski†

Abstract

A language is prefix-convex 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-convex 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-convex languages and their closed and free counterparts. This provides a common framework for diverse languages such as codes, factorial languages and ideals. We examine the relationships among these languages. We generalize these notions to arbitrary binary relations on the set of all words over a given alphabet, and study the closure properties of such languages.

Keywords: closed, closure, code, convex, factor, factorial, free, ideal, relation, language, prefix, subword, suffix

1 Introduction

This section introduces our basic terminology and notation, defines the scope of our work, and states some preliminary observations. Previous research is described in Section 2.

A note concerning the terminology is in order. We have used the term *continuous languages* in several publications [1, 5, 6, 7]. However, the term *convex languages* had been used for the same concept much earlier in [20]. Consequently we revert to the earlier terminology here.

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.

*This work was supported by the Natural Sciences and Engineering Research Council of Canada under grant no. OGP0000871.

†David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1, E-mail: tang@student.cs.uwaterloo.ca,brzozo@uwaterloo.ca

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 $v \trianglelefteq u$. The reflexive-and-transitive closure of $\trianglelefteq$ is denoted by $\trianglelefteq^*$.

Definition 1 and Proposition 1 below are generalizations of some results of Haines [10] and Thierrin [20]. See Section 2 for a further discussion.

Definition 1. *A language L is* $\trianglelefteq$-convex *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 $v \in L$.*

For an arbitrary relation $\trianglelefteq$ on Σ^*, let

$$_{\trianglelefteq}L = \{v \in \Sigma^* \mid v \trianglelefteq^* w \text{ for some } w \in L\}$$

and

$$L_{\trianglelefteq} = \{v \in \Sigma^* \mid v \trianglerighteq^* w \text{ for some } w \in L\}.$$

The following are easily verified:

Proposition 1. *Let $\trianglelefteq$ be an arbitrary relation on Σ^*. Then*

1. *A language is $\trianglelefteq$-convex if and only if it is $\trianglerighteq$-convex.*
2. *A language is $\trianglelefteq$-free if and only if it is $\trianglerighteq$-free.*
3. *Every $\trianglelefteq$-closed language and every $\trianglerighteq$-closed language is $\trianglelefteq$-convex.*
4. *A language is $\trianglelefteq$-closed if and only if its complement is $\trianglerighteq$-closed.*
5. *A language L is $\trianglelefteq$-closed ($\trianglerighteq$-closed) if and only if $L = {}_{\trianglelefteq}L$ ($L = L_{\trianglelefteq}$).*

Example 1. For $w \in \Sigma^*$, let $|w|_2$ be the length of w modulo 2. Let $\Sigma = \{a\}$ and let $\trianglelefteq$ be the binary relation $\leq_2$ defined by

$$u \leq_2 v \text{ if } |u|_2 \leq |v|_2.$$

The $\leq_2$-convex languages are $J = a^*$, $K = a(aa)^*$, $L = (aa)^*$, and $\emptyset$. The $\leq_2$-closed languages are J, L, and $\emptyset$. The $\leq_2$-free languages are $\emptyset$ and all the singleton languages $\{w\}$, for $w \in \Sigma^*$. Note that there are $\leq_2$-free languages that are not $\leq_2$-convex. For instance, $\{aa\}$ is $\leq_2$-free but not $\leq_2$-convex, because $aa \leq_2 \varepsilon$, $aa \leq_2 aa$, $\varepsilon \leq_2 aa$, but $\varepsilon \notin L$.

Proposition 2. *If $\trianglelefteq$ is antisymmetric, then every $\trianglelefteq$-free language is $\trianglelefteq$-convex. If $\trianglelefteq$ is reflexive and every $\trianglelefteq$-free language is $\trianglelefteq$-convex then $\trianglelefteq$ is antisymmetric.*

Proof. Suppose L is $\trianglelefteq$-free and $\trianglelefteq$ is antisymmetric. If L is not $\trianglelefteq$-convex, then there exist $u, w \in L$, $v \notin L$, such that $u \trianglelefteq v$, $u \trianglelefteq w$, and $v \trianglelefteq w$. Thus $u \trianglelefteq v$ and $v \trianglelefteq w$ and, since $\trianglelefteq$ is antisymmetric, we have $u \neq w$. Thus we have $u, w \in L$ and $u \triangleleft w$, contradicting that L is $\trianglelefteq$-free.

Conversely, suppose $\trianglelefteq$ is reflexive and every $\trianglelefteq$-free language is $\trianglelefteq$-convex, but $\trianglelefteq$ is not antisymmetric. Then there exist $v, w \in \Sigma^*$ such that $w \trianglelefteq v$, $v \trianglelefteq w$ and $v \neq w$. Since $\trianglelefteq$ is reflexive, we have $w \trianglelefteq w$. The language $\{w\}$ is $\trianglelefteq$-free but not $\trianglelefteq$-convex. Note that, if reflexivity is absent, v and w do not violate convexity. □

Usually in our applications we deal with partial order relations, so reflexivity and antisymmetry hold. If the binary relation is understood, we call a language *convex, free, closed,* or *converse closed*.

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[1] of w. If v is a subword of w, we write $v \Subset w$; if also $v \neq w$, then $v \subset w$. The relations $\leq$, $\preceq$, $\sqsubseteq$, and $\Subset$ are partial orders on Σ^*.

We apply Definition 1 to the following special cases:

$\trianglelefteq$ **is** $\leq$**:** If we use the relation "is a prefix of", then we get prefix-convex languages [6]. Prefix-free languages, except $\{\varepsilon\}$, are prefix codes [4], prefix-closed languages are complements of right ideals, and converse closed languages are the right ideals, that is, have the form $L\Sigma^*$, $L \subseteq \Sigma^*$. See Proposition 7.

$\trianglelefteq$ **is** $\preceq$**:** If we use the relation "is a suffix of", then we get the suffix-convex languages. Suffix-free languages, except $\{\varepsilon\}$, are suffix codes [4], suffix-closed languages are complements of left ideals, and converse closed languages are the left ideals, that is, have the form $\Sigma^* L$. See Proposition 7.

$\trianglelefteq$ **is** $\sqsubseteq$**:** If we use the relation "is a factor of"[2], we get factor-convex languages. Factor-free languages, except $\{\varepsilon\}$, are infix codes [19], factor-closed languages are factorial languages [15], which are complements of two-sided ideals, and converse closed languages are the ideals, that is, have the form $\Sigma^* L \Sigma^*$. See Proposition 6.

$\trianglelefteq$ **is** $\Subset$**:** If we use the relation "is a subword of"[3], we get subword-convex languages. Subword-free languages, except $\{\varepsilon\}$, are hypercodes [19], subword-closed languages are of the form $K = \overline{L} = \bigcup_{a_1 \cdots a_i \in L} \Sigma^* a_1 \Sigma^* \cdots a_i \Sigma^*$, and converse closed languages are of the form L above. See Section 2.

If a language is both prefix- and suffix-convex it is *bifix-convex*[4]. If it is both prefix- and suffix-free it is *bifix-free*; if it is not $\{\varepsilon\}$, it is then a bifix code [4]. If it is both prefix- and suffix-closed, it is *bifix-closed*. Note that bifix-closed and bifix-free languages can be defined as $(\leq \cup \preceq)$-closed and $(\leq \cup \preceq)$-free languages, respectively, but bifix-convex languages cannot be derived from a single relation.

The remainder of the paper is structured as follows. Previous work on convex languages is described in Section 2. In Section 3 we show the relations among the

[1] The word "subword" is often used to mean "factor"; here by a "subword" we mean a subsequence.

[2] This relation is called the "infix order" in [19].

[3] This relation is called the "embedding order" in [10].

[4] The word "bifix" is sometimes used to describe a word that is both a prefix and a suffix. Here we follow [12, 18]. The term "biprefix" is used in [4].

prefix- and suffix-convex classes of languages and their subclasses. In Section 4 we study the closure properties of the X-convex, X-closed and X-free classes of languages, where X stands for prefix, suffix, bifix, factor or subword. The converse X-closed classes are considered in Section 5. Special properties of closure under concatenation and star are studied in Section 6, and Section 7 concludes the paper.

2 Previous results and generalizations

For consistency, we use our notation and terminology when discussing previous work, but some of the key original terms are also mentioned.

In 1969 Haines proved the following results [10]:

Theorem 1 (Haines)**.** *Every subword-free language is finite.*

He called the subword relation *embedding*. He also defined (what we call) the *subword closure* of any language $L \subseteq \Sigma^*$ which is the set of all words that are subwords of words in L:

$$_{\Subset}L = \{u \in \Sigma^* \mid u \text{ is a subword of } v \text{ for some } v \in L\}.$$

Dually, he defined (what we call) the *converse subword closure* of any language $L \subseteq \Sigma^*$ which is the set of all words that contain a word of L as a subword:

$$L_{\Subset} = \{v \in \Sigma^* \mid u \text{ is a subword of } v \text{ for some } u \in L\}.$$

Theorem 2 (Haines)**.** *For any $L \subseteq \Sigma^*$, there exist finite languages F and G, such that*

$$_{\Subset}L = \overline{F_{\Subset}} = \overline{\bigcup_{a_1 \cdots a_i \in F} \Sigma^* a_1 \Sigma^* \cdots a_i \Sigma^*},$$

and

$$L_{\Subset} = G_{\Subset} = \bigcup_{a_1 \cdots a_i \in G} \Sigma^* a_1 \Sigma^* \cdots a_i \Sigma^*.$$

Theorem 3 (Haines)**.** *The languages $_{\Subset}L$ and $L_{\Subset}$ are regular, for every $L \subseteq \Sigma^*$.*

Haines noted that Theorem 1 is false for the factor relation, because $L = \{ab^n a \mid n \geq 1\}$ is an infinite factor-free language. It is also false for the prefix and suffix relations, since L is an infinite prefix- and suffix-free language.

For a discussion of earlier work related to the results of Haines see the paper by Kruskal [14].

In 1973 Thierrin introduced convex languages for the subword partial order [20]. He called a language *convex* if it is $\Subset$-convex, *left convex* if it is $\Subset$-closed, and *right-convex* if it is $\ni$-closed. He also defined a language to be *strongly convex* if it is closed under nonempty subwords, that is, if $v \neq \varepsilon$, $v \Subset w$, and $w \in L$ implies $v \in L$. This last concept is outside the scope of this work; we refer the reader to [20].

Proposition 3 (Thierrin)**.** *A language is $\Subset$-convex if and only if it is an intersection of a $\Subset$-closed and a $\Supset$-closed language. Equivalently, a language L is $\Subset$-convex if and only if there exist $\Subset$-closed languages M and N such that $L = M \setminus N$.*

Corollary 1 (Thierrin)**.** *Every $\Subset$-convex language is regular.*

Proposition 3 can be generalized with an added condition on $\trianglelefteq$.

Proposition 4. *If there exist $M, N \subseteq \Sigma^*$ such that M and N are $\trianglelefteq$-closed and $L = M \setminus N$, then L is $\trianglelefteq$-convex. If $\trianglelefteq$ is transitive and L is $\trianglelefteq$-convex, then there exist $M, N \subseteq \Sigma^*$ such that M and N are $\trianglelefteq$-closed and $L = M \setminus N$.*

Proof. Suppose $L = M \setminus N$, where M and N are $\trianglelefteq$-closed, and L is not $\trianglelefteq$-convex. Then there exists a triple $(u \in L, v \notin L, w \in L)$, such that $u \trianglelefteq v$, $u \trianglelefteq w$, and $v \trianglelefteq w$. We must have $u, w \in M$, and $u, w \notin N$, that is, $u, w \in \overline{N}$. If $v \in M$, then also $v \in N$ and $v \notin \overline{N}$. This means that $\overline{N}$ is not $\trianglelefteq$-convex. But, $\overline{N}$ is $\trianglerighteq$-closed by Proposition 1 (4), and every $\trianglerighteq$-closed language is $\trianglelefteq$-convex by Proposition 1 (3), which is a contradiction. Hence we must have $v \notin M$. But this now means that M is not $\trianglelefteq$-convex, and hence cannot be $\trianglelefteq$-closed—again a contradiction.

Conversely, suppose that $\trianglelefteq$ is transitive, and L is $\trianglelefteq$-convex. Let $M = {}_{\trianglelefteq}L$; then M is $\trianglelefteq$-closed by definition. Let

$$N = {}_{\trianglelefteq}L \setminus L = \{v \in \Sigma^* \mid v \notin L \text{ and } v \trianglelefteq^* w \text{ for some } w \in L\}.$$

Since $\trianglelefteq$ is transitive, we also have

$$N = \{v \in \Sigma^* \mid v \notin L \text{ and } v \trianglelefteq w \text{ for some } w \in L\}.$$

We claim that N is also $\trianglelefteq$-closed. For suppose $u \trianglelefteq v$ for some $v \in N$ such that $v \trianglelefteq w$, for some $w \in L$. Then $u \trianglelefteq w$ by transitivity. If $u \in L$, then L cannot be $\trianglelefteq$-convex because of the triple $(u \in L, v \notin L, w \in L)$. Hence we must have $u \notin L$, and N is $\trianglelefteq$-closed. Now $M \setminus N = M \cap \overline{N} = {}_{\trianglelefteq}L \cap (\overline{{}_{\trianglelefteq}L} \cup L) = {}_{\trianglelefteq}L \cap L = L$. □

Example 2. Let $\Sigma = \{a\}$, and let $\trianglelefteq = \{(\varepsilon, a), (a, aa)\}$; then $\trianglelefteq$ is not transitive since (ε, aa) is not in the relation. If $L = \{\varepsilon, aa\}$, then L is $\trianglelefteq$-convex, but not $\trianglelefteq$-closed, since $a \trianglelefteq aa$, $aa \in L$ and $a \notin L$. Suppose L can be expressed as $L = M \setminus N$, where both M and N are $\trianglelefteq$-closed. Then M must contain L and be closed; hence $a \in M$. Now N must contain a; otherwise $a \in M \setminus N$, and $M \setminus N \neq L$. However, since N must be closed, it must contain ε, since $\varepsilon \trianglelefteq a$. But then $M \setminus N$ does not contain ε, which is a contradiction. Therefore, Proposition 3 does not hold here.

On the other hand, lack of transitivity does not prevent *all* languages from satisfying Proposition 3. For example, the language $K = \{a\}$ is $\trianglelefteq$-convex and not $\trianglelefteq$-closed, but can be expressed as $K = \{\varepsilon, a\} \setminus \{\varepsilon\}$, which is a difference between two $\trianglelefteq$-closed languages.

Proposition 5 (Thierrin)**.** *If L is a $\Subset$-closed or a $\Supset$-closed language over Σ, then the syntactic monoid M_L is finite and contains a disjunctive zero z such that $ab = z$, $a, b \in M_L$, implies $axb = z$ for every $x \in M_L$.*

A language L is *noncounting* [20] if there exists an integer $k \geq 0$ such that, for arbitrary $u, v, w \in \Sigma^*$, $uv^k w \in L$ if and only if $uv^{k+1}w \in L$.

Corollary 2 (Thierrin)**.** *Every $\Subset$-convex, $\Subset$-closed, and $\Supset$-closed language is a noncounting regular language, and hence, a star-free language.*

Corollary 2 does not hold for the prefix, suffix, and factor relations. For example, $K = (aa)^*b\Sigma^*$, $L = \Sigma^*b(aa)^*$ and $M = \Sigma^*b(aa)^*b\Sigma^*$ are converse prefix-closed, converse suffix-closed and converse factor-closed, respectively, but they are not noncounting. However, convex languages with respect to these three relations are noncounting in the case of the one-letter alphabet.

Properties of $\Subset$-free languages were studied by Shyr and Thierrin [19] under the name of *hypercodes*. There is an extensive literature on codes characterized as antichains with respect to binary relations in free monoids. For example, languages that are both factor-free codes and $\Subset$-convex are studied in [9]. See also [11, 13, 18] and the references contained therein for further examples. 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 convex languages, and to study the closure properties of convex languages.

In 1990, de Luca and Varricchio characterized factor-closed languages, which they called *factorial:*

Proposition 6 (De Luca & Varricchio)**.** *A language L is factorial (that is, $\sqsubseteq$-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.*

In Proposition 6, 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 [15]. Let $\leq L$ be the set of all prefixes of words in L; thus, if L is prefix-closed, then $L = \leq L$. Now let $K = \overline{\leq L}$. One verifies that $u \in K$ implies $uv \in K$ for all $v \in \Sigma^*$, that is, $K = K\Sigma^*$, and $L = \leq L = \overline{K} = \overline{K\Sigma^*}$. Note that K 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. □

Prefix-convex languages were studied in connection with trace-assertion specifications [6, 7] (under the name of prefix-continuous languages). Here 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-convex. This result was extended to nondeterministic automata in [5]. Applications of these methods to the specification of software modules were discussed in [7].

Closure properties studied by Thierrin [20] are discussed in later sections.

3 Examples of convex languages

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

Factorial languages are defined as factor-closed languages, for example, in [2, 15], and as bifix-closed languages, for example, in [16]. 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. □

Factorial languages have received considerable attention. For example, their decompositions are studied in [2], their combinatorial properties in [15], and their complexity issues in [17]. 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-convex, prefix-free, and prefix-closed languages, *etc.* The classes in small rectangular boxes are closed under concatenation; we discuss this later. The classes in the large rectangle correspond to codes. The only difference between the solid and dashed lines is that the dashed lines indicate free and closed languages as special cases of convex languages, while solid lines show classes defined by changing the underlying binary relation.

Proposition 8. *All containments shown in Figure 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 $Y \in$

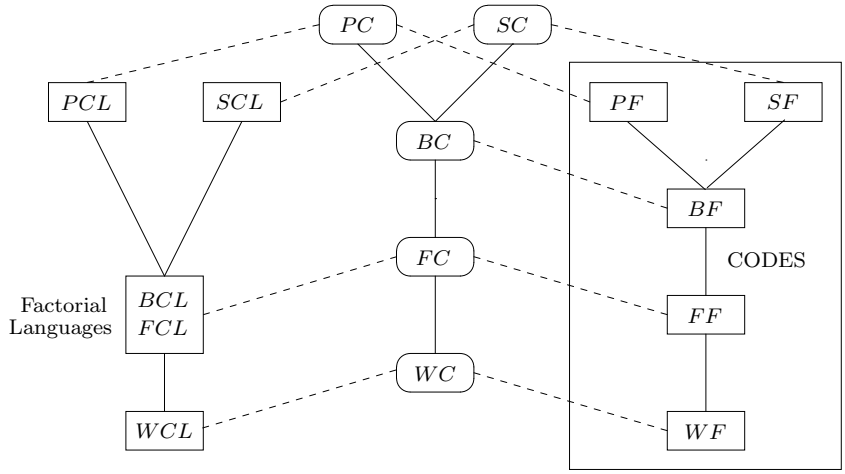

Figure 1: Classes of convex languages

$\{P, S, B, F, W\}$, classes YCL and YF are special cases of YC by Propositions 1(3) and 2; this accounts for the dashed lines.

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

The suffix-convex 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, $L_4 = \{\varepsilon, a\} \in WCL \setminus PF$.

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

The bifix-convex class BC: It suffices to show that BC does not contain any class from $\{PCL, SCL, PF, SF\}$. This follows because $L_5 = \{\varepsilon, a, ab\} \in PCL \setminus BC$, $L_1 = \{\varepsilon, a, ba\} \in SCL \setminus BC$, $L_6 = \{b, aab\} \in PF \setminus BC$, 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 $L_4 = \{\varepsilon, a\} \in WCL \setminus BF$, $L_6 = \{b, aab\} \in PF \setminus BF$, and $L_2 = \{a, abb\} \in SF \setminus BF$.

The factor-convex 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-convex class WC: It suffices to show that WC contains neither BCL nor FF. We have $L_9 = \{\varepsilon, a, b, ab, ba, aba\} \in BCL \setminus WC$, 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 is enough to show that WF contains neither WCL nor FF. Since FF does not contain WCL, neither does WF. Also, $L_{10} = \{aa, abba\} \in FF \setminus WF$.

This completes the proof. □

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

Proof. By definition, $BCL = PCL \cap SCL$. From Figure 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-convex, then it is prefix-closed, and hence bifix-closed. The last equality follows by left-right symmetry. □

3.1 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 9. *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-convex 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-convex if and only if it is empty, or has the form $\{a^i \mid m \leq i \leq m+n\}$, or $\{a^i \mid m \leq i\} = a^m a^*$, 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 $\{a^i \mid 0 \leq i\} = a^*$.*
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-convex, then so is KL.*

4 Closure in $\trianglelefteq$-convex languages

Thierrin [20] proved the closure results of Table 1 for subword-convex languages.

Table 1: Thierrin's closure results for the subword relation.

	convex	closed	converse closed
intersection	yes	yes	yes
union	no	yes	yes
complement	no	no	no
concatenation	no	yes	yes
star	no	yes	no

We generalize and extend these results. We first consider the closure properties of convex, free, and closed classes of languages. Converse closed classes are studied in Section 5.

4.1 Intersection, union and complement

Proposition 10. *If $K, L \subseteq \Sigma^*$ are $\trianglelefteq$-convex ($\trianglelefteq$-free, or $\trianglelefteq$-closed), then so is $M = K \cap L$.*

Proof. If M is not $\trianglelefteq$-convex, there exist $u, w \in M$ and $v \notin M$ such that $u \trianglelefteq v$, $u \trianglelefteq w$, and $v \trianglelefteq w$. Since $u, w \in K$ and $u, w \in L$, and K and L are $\trianglelefteq$-convex, 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 $v \trianglelefteq w$. 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. □

Corollary 3. *All the classes in Figure 1 are closed under intersection.*

The following is easily verified:

Proposition 11. *If $K, L \subseteq \Sigma^*$ are $\trianglelefteq$-closed, then so is $K \cup L$. If L^n is $\trianglelefteq$-closed for $n \geq 1$, then so is $\bigcup_{n=1}^{\infty} L_n$.*

Corollary 4. *All the closed classes, PCL, SCL, $BCL = FCL$, and WCL, are closed under union.*

Remark 3. The remaining classes in Figure 1 are not closed under union. Let $K = \{\varepsilon\}$, $L = \{aa\}$; both languages are X-convex and X-free for all $X \in \{P, S, B, F, W\}$. However, $K \cup L$ is neither X-convex nor X-free.

Remark 4. 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 $K = \{\varepsilon\}$ is in XCL, but $\overline{K}$ is not.

4.2 Concatenation and star

In general, convex languages are not closed under concatenation and star, even if the relation $\trianglelefteq$ is one of prefix, suffix, factor or subword relations.

Remark 5. If $L \subseteq \Sigma^*$ is prefix-(suffix-, bifix-, factor-, or subword-)convex, then L^2 is not necessarily prefix-(suffix-, bifix-, factor-, or subword-)convex. Hence these languages are not closed under concatenation.

Proof. $L = \{a, b, ab, ad, ca\}$ is prefix-, suffix-, factor-, and subword-convex, but L^2 is not, since $ab, abca \in L^2$ but $abc \notin L^2$, and $ab, adab \in L^2$, but $dab \notin L^2$. □

For concatenation of converse closed languages, see Section 5. For closed and free languages see Section 6.

Remark 6. If $L \subseteq \Sigma^*$ is prefix-(suffix-, bifix-, factor-, or subword-)convex, then L^* is not necessarily prefix-(suffix-, bifix-, factor-, or subword-)convex. The same holds if we replace "convex" by "free" or "converse closed".

Proof. If $\Sigma = \{a\}$, $L = \{aa\}$ is prefix-, suffix-, bifix-, factor-, and subword-convex, and -free, but $(aa)^*$ is not. Also, $L = aaa^*$ is converse X-closed for all X, but $L^* = L \cup \{\varepsilon\}$ is not. □

For the star of closed languages see Section 6.2.

4.3 Quotients

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 $xu \trianglelefteq xv$ ($ux \trianglelefteq vx$).[5]

[5]The terms 'left compatible' and 'right compatible' are used in [13, 18].

Proposition 12. *If $\trianglelefteq$ is left-invariant, and L is $\trianglelefteq$-convex ($\trianglelefteq$-free or $\trianglelefteq$-closed), then $M = x^{-1}L$ is $\trianglelefteq$-convex ($\trianglelefteq$-free or $\trianglelefteq$-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 $\trianglelefteq$-convex. If M is not $\trianglelefteq$-convex, then there exist $u, w \in M$ and $v \notin M$ such that $u \trianglelefteq v$, $u \trianglelefteq w$, and $v \trianglelefteq w$. If $\trianglelefteq$ is left-invariant, then $xu \trianglelefteq xv$, $xu \trianglelefteq xw$, and $xv \trianglelefteq xw$, and xu and $xw \in L$, while $xv \notin L$. This contradicts that L is $\trianglelefteq$-convex.

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 $v \trianglelefteq w$; then $xw \in L$ and $xv \notin L$. If $\trianglelefteq$ is left-invariant, then $xv \trianglelefteq xw$, which contradicts that L is $\trianglelefteq$-closed.

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

Example 1 shows that the invariance conditions of Proposition 12 are not necessary. The relation $\leq_2$ is not left-invariant, since $aa \leq_2 a$ but $a(aa) \not\leq_2 a(a)$. The left (and right) quotient of J by any word is J. The left quotient of K by w is K (respectively L), if w has even (respectively odd) length. Similarly, the left quotient of L by w is L (respectively K), if w has even (respectively odd) length. The left quotient of $\emptyset$ is $\emptyset$. Thus the left quotient of every $\leq_2$-convex language is $\leq_2$-convex. Similarly, the left quotient of every $\leq_2$-free language is $\leq_2$-free. On the other hand, the left quotient of L by a is K, which is not $\leq_2$-closed.

Corollary 5. *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 7. The classes BC, BF, FC, FCL and FF are not closed under either type of quotient. For let $L = \{\varepsilon, a, b, ab, ba, aba\}$; then L is bifix-convex, factor-convex and factor-closed, but $a^{-1}L = \{\varepsilon, b, ba\}$ and $La^{-1} = \{\varepsilon, b, ab\}$ are not. Also, $L = \{bb, bab\}$ is bifix-free and factor-free, but $b^{-1}L = \{b, ab\}$ and $Lb^{-1} = \{b, ba\}$ are neither.

4.4 Homomorphism and inverse homomorphism

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 $s(\varepsilon) = \{\varepsilon\}, s(uv) = s(u)s(v)$ for all $u, v \in \Sigma^*$, and $s(L) = \bigcup_{w \in L}\{s(w)\}$.

Remark 8. None of the classes from Figure 1 is closed under homomorphism. If $\Sigma = \Delta = \{a\}$, $h(a) = aa$, $L = \{\varepsilon, a\}$, then $h(L) = \{\varepsilon, aa\}$, L is in XC and in XCL,

for all $X \in \{P, S, B, F, W\}$, but $h(L)$ is not. Also, if $L = \{a, b\}$, $h(a) = \varepsilon$, $h(b) = a$, then $h(L) = \{\varepsilon, a\}$. Now L is in XF, but $h(L)$ is not. It follows that none of the classes from Figure 1 is closed under substitution.

Let $\trianglelefteq$ be a binary relation on Σ^*, and $\trianglelefteq'$, a binary relation on Δ^*. Then h is a $(\trianglelefteq, \trianglelefteq')$-*homomorphism*[6] if $u \trianglelefteq v$ implies $h(u) \trianglelefteq' h(v)$.

Proposition 13. *Let $(\Sigma^*, \trianglelefteq)$ and $(\Delta^*, \trianglelefteq')$ be free monoids with binary relations, let $h : \Sigma^* \to \Delta^*$ be a $(\trianglelefteq, \trianglelefteq')$-homomorphism, and let $K \subseteq h(\Sigma^*)$. If K is $\trianglelefteq'$-convex ($\trianglelefteq'$-free, or $\trianglelefteq'$-closed), then $L = h^{-1}(K)$ is $\trianglelefteq$-convex ($\trianglelefteq$-free, or $\trianglelefteq$-closed).*

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

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 $(\trianglelefteq, \trianglelefteq')$-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 $v \trianglelefteq w$. If h is a $(\trianglelefteq, \trianglelefteq')$-homomorphism, then $h(w) \in K$, $h(v) \notin K$, and $h(v) \trianglelefteq' h(w)$, which contradicts that K is $\trianglelefteq'$-closed. □

Corollary 6. *All the classes in Figure 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, we have a $(\trianglelefteq, \trianglelefteq')$-homomorphism for all $\trianglelefteq \in \{\leq, \preceq, \sqsubseteq, \Subset\}$. □

5 Converse closed languages

For $X \in \{P, S, F, W\}$, let XCC be the class of converse closed languages corresponding to the prefix, suffix, factor, and subword relations, respectively. By Proposition 1 (3), all these languages are convex. Similarly, let XC represent the convex classes and XCL, the closed classes.

The classes XCC in Figure 2 are the converse closed classes, which are shown in double rectangles. (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 $X \in \{P, S, F, W\}$.

Applying Propositions 10, 11, 12, and 13 for intersection, union, quotient, and inverse homomorphism, respectively, to the relation $\trianglerighteq$, we obtain:

Corollary 7. *All the classes of the form XCC are closed under intersection, union, and inverse homomorphism. Moreover, PCC is closed under left quotient, SCC, under right quotient, and WCC, under both.*

[6] In the terminology of [11], the relation $\trianglelefteq$ is *compatible* with h (in the case where $\trianglelefteq = \trianglelefteq'$).

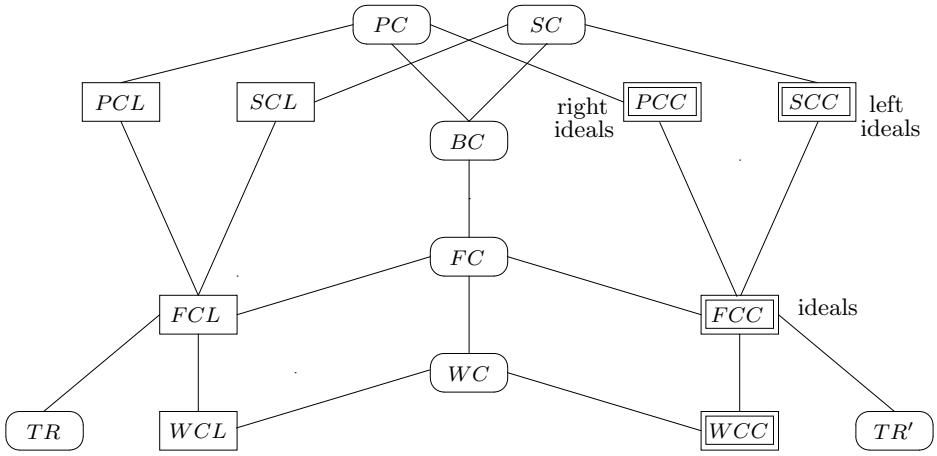

Figure 2: Classes of converse closed languages

Remark 9. The class 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.

Remark 10. No class XCC is closed under homomorphism. For let $\Sigma = \Delta = \{a, b\}$, $h(a) = h(b) = b$, and $L = \{\varepsilon, a\}$. 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 $K = \overline{h(\overline{L})} = \varepsilon + \Sigma^* a\Sigma^*$ is not in XCL, since $b \notin K$.

Remark 11. All the classes of the form XCC are closed under concatenation, because we have $(L\Sigma^*)(K\Sigma^*) = (L\Sigma^* K)\Sigma^*$, *etc.* Also, all the classes are closed under positive closure, which is defined as $L^+ = LL^*$, because $L\Sigma^* \supseteq L\Sigma^* L\Sigma^*$, *etc.* However, converse closed classes are not closed under star if $\epsilon \notin L$, because $\{\varepsilon\} \cup L\Sigma^*$ is not a right ideal, *etc.*

5.1 Transitive sofic languages

Factorial languages contain an interesting subclass which we discuss next; for more details we refer the reader to the literature [3, 4, 16]. A language $M \subseteq \Sigma^*$ is a *monoid* if it contains ε and is closed under concatenation. A monoid L 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 Figure 2, and TR' is the class of their complements. The following characterization is given in [3]:

Proposition 14 (Béal & Perrin)**.** *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 3. Let $\Sigma^* = \{a, b, c\}$, let $M = (ab^*c + b)^*$, and let $L = F(M)$. To find $F(M)$ we construct a nondeterministic finite automaton $\mathcal{N}$ from the minimal deterministic finite automaton $\mathcal{D}$ for M as follows. All the states of M, except the rejecting "dead" state, are made both accepting and initial. This guarantees that $\mathcal{N}$ accepts precisely all the factors of words of M. We then determinize $\mathcal{N}$ using the standard subset construction to obtain the minimal deterministic finite automaton $\mathcal{A}$ for $L = F(M)$. We leave the details to the reader; the automaton $\mathcal{D}$ has only three states, and $\mathcal{A}$ has only four. From $\mathcal{A}$ we can find the following regular expressions for L:

$$L = b^* + b^*c(b + ab^*c)^*(\varepsilon + ab^*) + b^*a(b + cb^*a)^*(\varepsilon + cb^*) = \overline{\Sigma^*(ab^*a + cb^*c)\Sigma^*}.$$

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 15. *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. □

Remark 12. Transitive sofic languages are not closed under left and right quotients, intersection, union, complement and concatenation. Let $\Sigma = \{a, b, c, d, e\}$, let L be the transitive sofic language L of Example 3, and let K be a similar language,

$$K = e^* + e^*c(e + de^*c)^*(\varepsilon + de^*) + b^*d(e + ce^*d)^*(\varepsilon + ce^*) = \overline{\Sigma^*(de^*d + ce^*c)\Sigma^*}.$$

Then $L \cap K = \varepsilon + c$, which is not transitive, because, for instance, $cxc \notin L \cap K$ for any $x \in \Sigma^*$. Also, for the language L of Example 3, $cac \in a^{-1}L$, but $a \notin a^{-1}L$; hence $a^{-1}L$ is not factorial. Similarly, $cac \in La^{-1}$, but $a \notin La^{-1}$; hence La^{-1} 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 $\varepsilon \notin L$.

5.2 Containments involving converse closed languages

Proposition 16. *All containments shown in Figures 1 and 2 are proper, and there are no other containments, except those implied by transitivity.*

Proof. We consider the classes starting with the largest ones.

1. $PC \not\supseteq SCC$: $L_{11} = (a+b)^*(aa+aaba) \in SCC$, because L_{11} is a left ideal, but L_{11} is not prefix-convex because $aa, aaba \in L_{11}$, but $aab \notin L_{11}$.
2. $SC \not\supseteq PCC$: $L_{12} = (aa+abaa)(a+b)^* \in PCC \setminus SC$, by a similar argument.
3. $PCL \not\supseteq TR'$: $L_{13} = (a+b+c)^*(ab^*a+cb^*c)(a+b+c)^* \in TR' \setminus PCL$, because $aa \in L_{13}$, but $a \notin L_{13}$.
4. $PCL \not\supseteq WCC$: $L_{14} = (a+b)^*a(a+b)^*a(a+b)^* \in WCC \setminus PCL$, because $aa \in L_{14}$, but $a \notin L_{14}$.
5. $SCL \not\supseteq TR'$: $L_{13} \in TR' \setminus SCL$.
6. $SCL \not\supseteq WCC$: $L_{14} \in WCC \setminus SCL$.
7. $PF \not\supseteq TR'$: $L_{13} \in TR' \setminus PF$, because $aaa \in L_{13}$ and $aa \in L_{13}$.
8. $PF \not\supseteq WCC$: $L_{14} \in WCC \setminus PF$, because $aaa \in L_{14}$ and $aa \in L_{14}$.
9. $SF \not\supseteq TR'$: $L_{13} \in TR' \setminus SF$.
10. $SF \not\supseteq WCC$: $L_{14} \in WCC \setminus SF$.
11. $PCC \not\supseteq WCL$: $L_4 = \{\varepsilon, a\} \in WCL \setminus PCC$.
12. $PCC \not\supseteq TR$: $L_{15} \in TR \setminus PCC$, where L_{15} is L from Example 3, because $a \in L_{15}$, but $aa \notin L_{15}$.
13. $PCC \not\supseteq WF$: $L_8 = \{a\} \in WF \setminus PCC$.
14. $SCC \not\supseteq WCL$: $L_4 = \{\varepsilon, a\} \in WCL \setminus SCC$.
15. $SCC \not\supseteq TR$: $L_{15} \in TR \setminus SCC$.
16. $SCC \not\supseteq WF$: $L_8 = \{a\} \in WF \setminus SCC$.
17. $FCC \not\supseteq PCC$: $L_{16} = a(a+b)^* \in PCC \setminus FCC$.
18. $FCC \not\supseteq SCC$: $L_{17} = (a+b)^*a \in SCC \setminus FCC$.
19. $WCC \not\supseteq FCC$: $L_{18} = (a+b)^*aa(a+b)^* \in FCC \setminus WCC$, since $aa \in L_{18}$, but $aba \notin L_{18}$.
20. $TR' \not\supseteq WCC$: $L_{14} \in WCC \setminus TR'$, because $\overline{L_{14}} = b^* + b^*ab^*$ is not transitive, since it has no word axa for any $x \in \Sigma^*$.
21. $WCC \not\supseteq TR'$: $L_{13} \in TR' \setminus WCC$, because $aa \in L_{13}$, but $aca \notin L_{13}$.

Hence all the classes shown in the two figures are distinct and there are no other containments. □

6 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 4. 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, the classes of $\trianglelefteq$-closed and $\trianglelefteq$-free languages are not closed under concatenation.

6.1 Free languages and 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 17. *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 $x_1x_2 \lhd y_1y_2$. 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. □

Lemma 1. *The binary relations $\leq$, $\preceq$, $\sqsubseteq$ and $\Subset$ 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 $uv \neq \varepsilon$. 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 $u \neq \varepsilon$, then $x_1 \sqsubset y_1$. If $ux_1 = y_1$ and $u = \varepsilon$, then $x_1 = y_1$, and $x_2 \sqsubset y_2$, since $v \neq \varepsilon$.

Now suppose that $x_1x_2 \Subset y_1y_2$; 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 $v_1 \cdots v_n \neq \varepsilon$. If $j < i$, then $x_1 \Subset y_1$. If $j > i$, then $x_2 \Subset y_2$. If $j = i$, and $v_0v_1 \cdots v_i' \neq \varepsilon$, then $x_1 \Subset y_1$. If $j = i$, and $v_0v_1 \cdots v_i' = \varepsilon$, then $x_2 \Subset y_2$. □

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

6.2 Closed languages and concatenation and star

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 18. *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. □

Lemma 2. *The binary relations $\leq$, $\preceq$, $\sqsubseteq$ and $\Subset$ 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 $\varepsilon \leq y_2$, we have $x_1 = x$, and $x_2 = \varepsilon$. 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 $x_2 = \varepsilon \sqsubseteq y_2$. 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 $x_1 = \varepsilon \sqsubseteq y_1$ and $x_2 = x \sqsubseteq y_2$.

Now suppose that $x \Subset y_1y_2 = v$; then $x = a_1 \cdots a_n$ and $v = v_0a_1v_1 \cdots a_nv_n$, where $v_0, \ldots, v_n \in \Sigma^*$, $a_1, \ldots, a_n \in \Sigma$, and, for some i we have $y_1 = v_0a_1v_1 \cdots a_iv_i'$ and $y_2 = v_i''a_{i+1}v_{i+1} \cdots a_nv_n$, where $v_i = v_i'v_i''$. If $i = n$, then $x_1 = x \Subset y_1$ and $x_2 = \varepsilon \Subset y_2$. If $i < n$, then $x = x_1x_2$, where $x_1 = a_1 \cdots a_i \Subset y_1$ and $x_2 = a_{i+1} \cdots a_n \Subset y_2$. □

Corollary 9. *The prefix-, suffix-, bifix- (= factor-), and subword-closed classes are closed under concatenation.*

A binary relation $\trianglelefteq$ is *ε-full* if $\varepsilon \trianglelefteq w$ for all $w \in \Sigma^*$. Note that all our example relations are ε-full.

Proposition 19. *If $\trianglelefteq$ is antisymmetric, factoring, and ε-full, and L is $\trianglelefteq$-closed, then so is L^*.*

Proof. We adapt Thierrin's proof [20] given for the case where $\trianglelefteq$ is the subword relation. Suppose L is $\trianglelefteq$-closed. If $L = \emptyset$, then $L^* = \{\varepsilon\}$. If there is no $w \in \Sigma^*$, $w \neq \varepsilon$, such that $w \trianglelefteq \varepsilon$, then L^* is $\trianglelefteq$-closed. If there is such a w, then $\varepsilon \trianglelefteq w$, since $\trianglelefteq$ is ε-full. However, this contradicts the antisymmetry of $\trianglelefteq$. Thus, if $L = \emptyset$, then L^* is $\trianglelefteq$-closed. Therefore assume that $L \neq \emptyset$. Since $\trianglelefteq$ is ε-full, we must have $\varepsilon \in L$. We argue that L^* is $\trianglelefteq$-closed if L^n is $\trianglelefteq$-closed for each $n \geq 0$. As we have shown above, $L^0 = \{\varepsilon\}$ is $\trianglelefteq$-closed. If $n = 1$, then $L^n = L$, and L is $\trianglelefteq$-closed by assumption. For $n > 1$, since $\trianglelefteq$ is factoring, L^n is $\trianglelefteq$-closed by Proposition 18. Since the union of $\trianglelefteq$-closed languages is $\trianglelefteq$-closed by Proposition 11, we have our result. □

Corollary 10. *The prefix-, suffix-, bifix- (= factor-), and subword-closed classes are closed under star.*

7 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^*. Table 2 summarizes our closure results. In case closure holds only for some of the relations, we specify these relations in the table.

Table 2: Closure results for prefix, suffix, factor, and subword relations.

	convex	*closed*	*converse closed*	*free*
intersection	yes	yes	yes	yes
union	no	yes	yes	no
complement	no	no	no	no
concatenation	no	yes	yes	yes
Kleene star	no	yes	no	no
positive closure	no	yes	yes	no
left quotient	prefix	prefix	prefix	prefix
	subword	subword	subword	subword
right quotient	suffix	suffix	suffix	suffix
	subword	subword	subword	subword
homomorphism	no	no	no	no
inverse homomorphism	yes	yes	yes	yes

The problems of deciding whether a language specified by a deterministic or nondeterministic finite automaton is prefix-, suffix-, factor-, or subword-convex, -free, or -closed have been recently studied in [8].

Acknowledgment: We thank Larry Cummings, Helmut Jürgensen, Jacques Sakarovitch and Jeff Shallit for useful comments and pointers to references. We are grateful to two anonymous referees for suggesting numerous improvements.

References

[1] Ang, T. and Brzozowski, J. A. Continuous Languages. In Csuhaj-Varjú, E. and Ésik, Z., editors, *Proc. 12th Int. Conference on Automata and Formal Languages,* Computer and Automation Research Institute, Hungarian Academy of Sciences, 74–85, 2008.

[2] Avgustinovich, S. V. and Frid, A. E. A unique decomposition theorem for factorial languages. *Internat. J. Algebra Comput.,* (15): 149–160, 2005.

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

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

[5] Brzozowski, J. A. Representation of a class of nondeterministic semiautomata by canonical words. *Theoret. Comput. Sci.,* (356): 46–57, 2006.

[6] Brzozowski, J. A. and Jürgensen, H. Representation of semiautomata by canonical words and equivalences. *Internat. J. Found. Comput. Sci.,* (16): 831–850, 2005.

[7] Brzozowski, J. A. and Jürgensen, H. Representation of semiautomata by canonical words and equivalences, Part II: Applications to the specification of software modules. *Internat. J. Found. Comput. Sci.,* (18): 1065–1087, 2007.

[8] Brzozowski, J. A., Shallit, J. O. and Xu, Z. Decision problems for convex languages. In Dediu, A. H., Ionescu, A. M. and Martin-Vide, C., editors, *Proc. 3rd Int. Conference on Languages and Automata Theory and Applications,* LNCS (5457): 247–258. Springer, 2009.

[9] Guo, Y. Q., Shyr, H. J. and Thierrin, G. E-convex infix codes. *Order,* (3): 55–59,1986.

[10] Haines, L. H. On free monoids partially ordered by embedding. *J. Combin. Theory,* (6): 94–98,1969.

[11] Jürgensen, H., Kari, L. and Thierrin, G. Morphisms preserving densities. *Internat. J. Comput. Math.,* (78): 165–189, 2001.

[12] Jürgensen, H. and Konstantinidis, S. Codes. In Rozenberg, G. and Salomaa, A., editors: *Handbook of Formal Languages,* (1): 511–607. Springer, 1997.

[13] Jürgensen, H. and Yu, S. S. Relations on free monoids, their independent sets, and codes. *Internat. J. Comput. Math.,* (40): 17–46, 1991.

[14] Kruskal, J. B. The theory of well-quasi-ordering: a frequently discovered concept. *J. Combin. Theory,* (A) (13): 297–305,1972.

[15] De Luca, A. and Varricchio, S. Some combinatorial properties of factorial languages. In Capocelli, R., editor, *Sequences,* 258–266. Springer, 1990.

[16] Restivo, A. Finitely generated sofic systems. *Theoret. Comput. Sci.,* (65): 265–270, 1989.

[17] Shur, A. M. Factorial languages of low combinatorial complexity. In Ibarra, O. H. and Dang, Z., editors, *Proc. Developments in Language Theory,* LNCS (4036): 397–407. Springer, 2006.

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

[19] Shyr, H. J. and Thierrin, G. Hypercodes. *Inform. and Control,* (24): 45–54, 1974.

[20] Thierrin, G. Convex languages. In Nivat, M., editor, *Automata, Languages and Programming,* 481–492. North-Holland, 1973.

Received 25th September 2008