# Unbordered partial words

By: F. Blanchet-Sadri, C.D. Davis, Joel Dodge, Robert Mercaş, Robert, Margaret Moorefield

F. Blanchet-Sadri, C.D. Davis, J. Dodge, R. Mercas and M. Moorefield, "Unbordered Partial Words." *Discrete Applied Mathematics*, Vol. 157, 2009, pp 890-900. DOI: 10.1016/j.dam.2008.04.004

Made available courtesy of Elsevier: http://dx.doi.org/10.1016/j.dam.2008.04.004



## Abstract:
An unbordered word is a string over a finite alphabet such that none of its proper prefixes is one of its suffixes. In this paper, we extend the results on unbordered words to unbordered partial words. Partial words are strings that may have a number of "do not know" symbols. We extend a result of Ehrenfeucht and Silberger which states that if a word u can be written as a concatenation of nonempty prefixes of a word v, then u can be written as a unique concatenation of nonempty unbordered prefixes of v. We study the properties of the longest unbordered prefix of a partial word, investigate the relationship between the minimal weak period of a partial word and the maximal length of its unbordered factors, and also investigate some of the properties of an unbordered partial word and how they relate to its critical factorizations (if any).

## Keywords:
Words, Partial words, Unbordered words, Unbordered partial words

## Article:
### 1. Introduction
*Periodicity* and *borderedness* are two fundamental properties of words that play a role in several research areas including string searching algorithms [9–11,14], data compression [16], theory of codes [3], sequence assembly [15] and superstrings [7] in computational biology, and serial data communication systems [8]. It is well known that these two word properties do not exist independently from each other.

Let $A$ be a nonempty finite set, also called an alphabet. Consider a nonempty word $u = a_0a_1 \dots a_{n-1}$ with $a_i \in A$. Then a *period* of $u$ is a positive integer $p$ such that $a_i = a_{i+p}$ for $0 \leq i < n - p$. The word $u$ is called *bordered* if one of its proper prefixes is one of its suffixes. The length of the longest such prefix (also called longest border) is the length of u minus the length of its shortest period. The word u is called *unbordered* otherwise. In other words, it is *unbordered* if it has no proper period. For example, *abaabb* is unbordered while *abaab* is bordered. Unbordered words turn out to be primitive, that is, they cannot be written as a power of another word. Unborderedness has the following important property: Different occurrences of an unbordered factor u in a word v never overlap. A related property is that no primitive word $u$ can be an inside factor of $uu$. Fast algorithms for testing primitivity of words can be based on this property [10].

The study of unbordered partial words was initiated in [4]. Partial words are strings that may have a number of "do not know" symbols. In this paper, we pursue this study by extending some more results on unbordered words to unbordered partial words. We begin in Section 2 by reviewing basic concepts on words and partial words. In Section 3, we recall a result of Ehrenfeucht and Silberger [13] which states that if a word $u$ can be written as a concatenation of nonempty prefixes of a word $v$, then $u$ can be written as a unique concatenation of nonempty unbordered prefixes of $v$, and we extend this result to partial words. In Section 4, we give more results on concatenations of prefixes. In particular, we study the properties of the longest unbordered prefix of a partial word. We also investigate the relationship between the minimal *weak* period of a partial word and the maximal length of its unbordered factors. In Section 5, we investigate some of the properties of an unbordered partial word and how they relate to its critical factorizations (if any). Blanchet-Sadri and Wetzler extended the well-known critical factorization theorem to partial words and their result states that the minimal weak period of a *nonspecial* partial word can be locally determined in at least one position [6]. Finally, we prove that, with regard to Chomsky hierarchy, the set of all partial words over an arbitrary nonunary fixed finite alphabet having a critical factorization is a context sensitive language that is not context-free.

## 2. Preliminaries

Fixing an alphabet $A$, we first review the basic concepts on words and partial words over $A$.

### *2.1. Words*

A *string* or *word* $u$ over $A$ is a finite concatenation of symbols or letters from $A$. The number of symbols in $u$, or *length* of $u$, is denoted by $|u|$. For any word $u$, $u[i..j - 1]$ is the factor of u that starts at position $i$ and ends at position $j - 1$ (it is called *proper* if $0 \leq i < j \leq |u|$ and ($i > 0$ or $j < |u|$)). In particular, $u[0..j - 1]$ is the prefix of $u$ that ends at position $j - 1$ and $u[i..|u| - 1]$ is the suffix of $u$ that begins at position $i$. The factor $u[i..j - 1]$ is the empty word if $i \geq j$ (the empty word is denoted by $\varepsilon$). The set of all finite length words over $A$ (length greater than or equal to zero) is denoted by $A^*$. It is a monoid under the associative operation of concatenation or product of words where $\varepsilon$ serves as the identity, and it is referred to as the *free monoid* generated by A. Similarly, the set of all nonempty words over A is denoted by $A^+$. It is a semigroup under the operation of concatenation of words and is referred to as the *free semigroup* generated by $A$.

For a word $u$, the powers of u are defined inductively by $u^0 = \varepsilon$ and, for any $n \geq 1$, $u^n = uu^n - 1$. If $u$ is nonempty, then $v$ is a root of $u$ if $u = v^n$ for some positive integer $n$. The shortest root of $u$, denoted by $\sqrt{u}$, is called the *primitive root* of $u$, and $u$ is itself called *primitive* if $\sqrt{u} = u$. If $u = (\sqrt{u})^n$, then $\sqrt{u}$ is the unique primitive word $v$ and $n$ is the unique positive integer such that $u = v^n$. All positive powers of u have the same primitive root.

A word of length $n$ over $A$ can be defined by a total function $\underline{u} : \{0,..., n - 1\} \rightarrow A$ and is usually represented as $u = a_0a_1...a_{n-1}$ with $a_i \in A$. A positive integer $p$ is a period of $u$ if for all $0 \leq i < n - p$ we have $a_i = a_{i+p}$. This can be equivalently formulated, for $p \leq n$, by $u = xv = wx$ for some words $x$, $v$, $w$ satisfying $|v| = |w| = p$. For a word $u$, there exists a *minimal period* which is denoted by $p(u)$. A nonempty word $u$ is *unbordered* if $p(u) = |u|$. Otherwise, it is *bordered*. A nonempty word $x$ is a border of a word $u$ if $u = xv = wx$ for some nonempty words v and $w$. Unbordered words turn out to be primitive.

### *2.2. Partial words*

A partial word $u$ of length $n$ over $A$ is a partial function $u : \{0,..., n - 1\} \rightarrow A$. For $0 \leq i < n$, if $u(i)$ is defined, then we say that $i$ belongs to the *domain* of $u$, denoted by $i \in D(u)$, otherwise we say that $i$ belongs to the *set* of *holes* of $u$, denoted by $i \in H(u)$. A (*full*) word over $A$ is a partial word over $A$ with an empty set of holes.

For convenience, we will refer to a partial word over $A$ as a word over the enlarged alphabet $A_\Diamond = A \cup \{\Diamond\}$, where $\Diamond$ represents a "do not know" symbol. So a partial word $u$ of length $n$ over $A$ can be viewed as a total function $u : \{0,..., n - 1\} \rightarrow A \cup \{\Diamond\}$ where $u(i) = \Diamond$ whenever $i \in H(u)$. For example, $u = a \Diamond bbc \Diamond cb$ is a partial word of length 8 where $D(u) = \{0, 2, 3, 4, 6, 7\}$ and $H(u) = \{1, 5\}$. We can thus define for partial words concepts such as concatenation, powers, etc. in a trivial way.

The length of a partial word $u$ over $A$ is denoted by $|u|$, while the set of distinct letters in $A$ occurring in $u$ is denoted by $a(u)$. For the set of all partial words over $A$ with an arbitrary number of holes we write $A_\Diamond^*$. The set $A_\Diamond^*$ is a monoid under the operation of concatenation where $\varepsilon$ serves as the identity element. If $X \subset A$*o, then the *cardinality* of $X$ is denoted by $||X||$.

For partial words, we use the same notions of prefix, suffix and factor, as for full ones. The unique *maximal common prefix* of $u$ and $v$ will be denoted by pre($u$, $v$). Now, if $u \in A_\Diamond^*$ and $0 \leq i < j \leq |u|$, then $u[i..j - 1]$

denotes the factor $u(i) \dots u(j - 1)$. For a subset $X$ of $A_\Diamond^*$, we denote by $P(X)$ the set of prefixes of elements in $X$ and by $S(X)$ the set of suffixes of elements in $X$. If $X$ is the singleton $\{u\}$, then $P(X)$ (respectively, $S(X)$) will be abbreviated by $P(u)$ (respectively, $S(u)$).

A *factorization* of a partial word $u$ is any tuple $(u_0, u_1, \dots, u_{i-1})$ of partial words such that $u = u_0u_0 \dots u_{i-1}$. For a subset $X$ of $A_\Diamond^*$ and an integer $i > 0$, the set

$$\{u_0u_1 \dots u_{i-1} \mid u_0, \dots, u_{i-1} \in X\}$$

is denoted by $X^i$. The submonoid of $A_\Diamond^*$ generated by $X$ will be denoted by $X^*$ where $X^* = Si>0Xi$ and $X0 = \{E\}$. The subsemigroup of $A_\Diamond^*$ generated by $X$ is denoted by $X^+$ where $X^+ = \bigcup_{i\geq 0} X^i$. By definition, each partial word $u$ in $X^*$ admits at least one factorization $(u_0, u_1, \dots, u_{i-1})$ whose elements are all in $X$. Such a factorization is called an *X-factorization*.

*2.2.1. Containment and compatibility*

If $u$ and $v$ are two partial words of equal length, then $u$ is said to be *contained in* $v$, denoted by $u \subset v$, if all elements in $D(u)$ are in $D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$. If $u \subset v$ but $u \neq v$, then this will be denoted by $u \sqsubset v$. Partial words $u$ and $v$ are called *compatible* if there exists a partial word $w$ such that $u \subset w$ and $v \subset w$. This is denoted by $u \uparrow v$. The *least upper bound* of $u$ and $v$ is denoted by lub$(u, v)$. By this we mean $u \subset$ lub$(u, v)$ and $v \subset$ lub$(u, v)$ and $D(\text{lub}(u, v)) = D(u) \cup D(v)$. For example, $u = a \Diamond b \Diamond \Diamond c$ and $v = ab \Diamond c \Diamond c$ are compatible and lub$(u, v) = abbc \Diamond c$.

The following rules are used for computing with partial words.

**Lemma 1** *([2]). Let $u, v, w, x, y$ e $A_\Diamond^*$. The following hold*:

*Multiplication: If $u \uparrow v$ and $x \uparrow y$, then $ux \uparrow vy$.*
*Simplification: If $ux \uparrow vy$ and $|u| = |v|$, then $u \uparrow v$ and $x \uparrow y$.*
*Weakening: If $u \uparrow v$ and $w \subset u$, then $w \uparrow v$.*

The following result extends to partial words the *equidivisibility property* of words, or, *lemma of Lévi*.

**Lemma 2** *([2]). Let $u, v, x, y \in A_\Diamond^*$ be such that $ux \uparrow vy$.*
- *If $|u| > |v|$, then there exist $w, z$ e $A_\Diamond^*$ such that $u = wz$, $v \uparrow w$, and $y \uparrow zx$.*
- *If $|u| < |v|$, then there exist $w, z$ e $A_\Diamond^*$ such that $v = wz$, $u \uparrow w$, and $x \uparrow zy$.*

*2.2.2. Periodicity*

A *period* of a partial word $u$ over $A$ is a positive integer $p$ such that $u(i) = u(j)$ whenever $i, j \in D(u)$ and $i \equiv j$ mod $p$. In this case $u$ is called *p-periodic*. A *weak period* of $u$ is a positive integer $p$ such that $u(i) = u(i + p)$ whenever $i, i + p \in D(u)$. In this case $u$ is called *weakly p-periodic*. The partial word $u = baab \Diamond abca$ is weakly 3-periodic but is not 3-periodic. The latter shows a difference between partial words and full words since every weakly $p$-periodic full word is $p$-periodic. Also even if the length of a partial word $u$ is a multiple of a weak period of $u$, then $u$ is not necessarily a power of a shorter partial word. The minimal period and the minimal weak period of $u$ are denoted by $p(u)$ and $p'(u)$, respectively.

This notion of weak period can be equivalently formulated as follows.

**Lemma 3.** *For an integer $p$, the partial word $u \in A_\Diamond^*$ is weakly $p$-periodic if and only if the containments $u \subset xv$ and $u \subset wx$ hold for some partial words $x, v, w$ satisfying $|v| = |w| = p$.*

**Proof.** Write $u$ as $v_1v_2 \dots v_kr$ where $|v_1| = |v_2| = \cdots = |v_k| = p$ and $0 \leq |r| < p$, and $v_k$ as $st$ where $|s| = |r|$. Set $x_1 = v_1 \dots v_{k-1}s$ and $x_2 = v_2 \dots v_kr$.

If the containments $u \subset xv$ and $u \subset wx$ hold for some partial words $x$, $v$, $w$ satisfying $|v| = |w| = p$, then both $v_1...v_{k-1}s \subset x$ and $v_2...v_kr \subset x$ hold, and so $v_1...v_{k-1}s \uparrow v_2...v_kr$. By Simplification, $v_1 \uparrow v_2,...,v_{k-1} \uparrow v_k$ and $s \uparrow r$. Now, let $i, i + p \in D(u)$. Then $i = lp + j$ for some $0 \leq l < k$ and $0 \leq j < p$. If $l < k - 1$, then we get $u(i) = v_{l+1}(j) = v_{l+2}(j) = u(i + p)$ since $v_{l+1} \uparrow v_{l+2}$ and $j \in D(v_{l+1}) \cap D(v_{l+2})$, and if $l = k - 1$, then $u(i) = v_k(j) = s(j) = r(j) = u(i + p)$ since $s \uparrow r$ and $j \in D(s) \cap D(r)$. In either case, $u$ is weakly $p$-periodic. Conversely, if $p$ is a weak period of $u$, then $v_i \uparrow v_{i+1}$ for all $1 \leq i < k$ and $s \uparrow r$. Thus $x_1 \uparrow x_2$, and there exists $x$ such that $x_1 \subset x$ and $x_2 \subset x$. Setting $v = tr$ and $w = v1$, we get $u = x_1v \subset xv$ and $u = wx_2 \subset wx$ with $|v| = |w| = p$. □

A partial word $u$ is *primitive* if there exists no word $v$ such that $u \subset v^n$ with $n > 2$. Note that the empty word is not primitive, and that if $v$ is primitive and $v \subset u$, then $u$ is primitive as well. If $u$ is a nonempty partial word, then there exists a primitive word $v$ and a positive integer $n$ such that $u \subset v^n$. Uniqueness does not hold for partial words. For example, if $u = a\Diamond$, then $u \subset a^2$ and $u \subset ab$ for distinct letters $a$, $b$. For $u, v \in A_\Diamond^*$, if there exists a primitive word $x$ such that $uv \subset x^n$ for some positive integer $n$, then there exists a primitive word $y$ such that $vu \subset y^n$. Consequently, if $uv$ is primitive, then $vu$ is primitive [4].

A nonempty partial word $u$ is *bordered* if one of its proper prefixes is compatible with its suffix of the same length. Otherwise, no nonempty words $x$, $v$, $w$ exist such that $u \subset xv$ and $u \subset wx$ and $u$ is called *unbordered*. It is easy to see that if $u$ is unbordered and $u \subset u'$, then $u'$ is unbordered as well. In [4], an extension of a result on words to partial words allows us to conclude that unbordered partial words are primitive. This comes from the fact that if $u$ is a nonempty unbordered partial word, then $p(u) = |u|$. We call $x$ a border of $u$ if $u \subset xv$ and $u \subset wx$ for some $v$ and $w$ with $0 < |x| < |u|$. A border $x$ of $u$ is called *minimal* if $|x| > |y|$ implies that $y$ is not a border of $u$.

**3. Concatenations of prefixes**

For $u, v \in A_\Diamond^*$, we write $u \ll v$ if there exists a sequence $v_0,...,v_{n-1}$ of prefixes of $v$ such that $u = v_0...v_{n-1}$. Obviously, $\varepsilon \ll u$ and $u \ll u$. Also, if $u \ll v$ and $v \ll w$, then $u \ll w$.

**Theorem 1** *([13]). Let $u \in A^+$, $v \in A^*$ be such that $u \ll v$. Then there exists a unique sequence $v_0,...,v_{n-1}$ of nonempty unbordered prefixes of $v$ such that $u = v_0...v_{n-1}$.*

Our main result in this section is to extend Theorem 1 to partial words (see Theorem 2). In order to do this, we introduce two types of bordered partial words: the *well bordered* and the *badly bordered* partial words.

**Definition 1.** Let $u \in A_\Diamond^+$ be bordered. Let $x$ be a minimal border of $u$, and set $u = x_1 v = wx_2$ where $x_1 \subset x$ and $x_2 \subset x$. We call $u$ *well bordered* if $x_1$ is unbordered. Otherwise, we call $u$ *badly bordered*.

Note that if a nonempty partial word $u$ is well bordered then $x_2$ can be either bordered or unbordered, and the same is true if $u$ is badly bordered. Also since $x_1$ is a prefix of $u$, Definition 1 is of special interest to the main topic of this section entitled "Concatenations of Prefixes".

For convenience, we will at times refer to a minimal border of a well-bordered partial word as a *good border* and of a badly bordered partial word as a *bad border*.

As a result of $x$ being a bad border, we have the following lemma.

**Lemma 4.** *Let $u \in A_\Diamond^+$ be badly bordered. Let $x$ be a minimal border of $u$, and set $u = x_1v = wx_2$ where $x_1 \subset x$ and $x_2 \subset x$. Then there exists $i$ such that $i \in H(x_1)$ and $i \in D(x_2)$.*

Proof. Since $x_1$ is bordered, $x_1 = r_1s_1 = s_2r_2$ for nonempty partial words $r_1$, $r_2$, $s_1$, $s_2$ where $s_1 \subset s$ and $s_2 \subset s$ for some $s$. If no $i$ exists such that $i \in H(x_1)$ and $i \in D(x_2)$, then $x_2$ must also be bordered. So $x2 = r'_1s'_1 = s'_2r'_2$ where

$r'_1 \subset r_1$, $r'_2 \subset r_2$, $s'_1 \subset s$ and $s'_2 \subset s$, thus $s_2 \uparrow s'_1$. This means that there exists a border of $u$ of length shorter that $|x|$ which contradicts the fact that $x$ is a minimal border of $u$. □

Our goal is to extend Theorem 1 to partial words or to construct, given any partial words $u$ and $v$ satisfying $u \ll v$, a unique sequence of nonempty unbordered prefixes of $v$, $v_0,...,v_{n-1}$, such that $u \uparrow v_0...v_{n-1}$. We will see that if during the construction of the sequence a badly bordered prefix is encountered, then the desired sequence may not exist. We first prove two propositions.

**Proposition 1.** *If $v \in A_\Diamond^*$, then there do not exist two distinct compatible sequences of nonempty unbordered prefixes of $v$.*

**Proof.** Suppose that $v_0...v_{n-1} \uparrow v'_0...v'_{m-1}$ where each $v_i$ and each $v'_i$ is a nonempty unbordered prefix of $v$. If there exists $i > 0$ such that $|v_0| = |v'_0|,...,|v_{i-1}| = |v'_{i-1}|$ and $|v_i| < |v'_i|$, then $v_0 = v'_0, ... , v_{i-1} = v'_{i-1}$ and $v_i$ is a prefix of $v'_i$. By simplification, $v_i...v_jx \uparrow v'_i$ where $i \leq j < n$-1 and $x$ is a nonempty prefix of $v_{j+1}$. The fact that $x$, $v'_i$ are prefixes of $v$ satisfying $|v'_i| > |x|$ implies that $x$ is a prefix of $v'_i$. In addition, $x$ is compatible with the suffix of length $|x|$ of $v'_i$, and consequently $v'_i$ is bordered. Similarly, there exists no $i \geq 0$ such that $|v_0| = |v'_0|,...,|v_{i-1}| = |v'_{i-1}|$ and $|v_i| > |v'_i|$. Clearly, $n = m$ and uniqueness follows. □

**Proposition 2.** *Let $u \in A_\Diamond^+$ be bordered. Let $x$ be a minimal border of $u$, and set $u = x_1v = wx_2$ where $x_1 \subset x$ and $x_2 \subset x$. Then the following hold:*

1. *The partial word $x$ is unbordered.*
2. *If $u$ is well bordered, then $u = x_1 u'x_2 \subset xu'x$ for some $u'$.*

**Proof.** For Statement 1, assume that $r$ is a border of $x$, that is, $x \subset rs$ and $x \subset s'r$ for some nonempty partial words $r, s, s'$. Since $u \subset xv$ and $x \subset rs$, we have $u \subset rsv$, and similarly, since $u \subset wx$ and $x \subset s'r$, we have $u \subset ws'r$. Then $r$ is a border of $u$. Since $x$ is a minimal border of $u$, we have $|x| \leq |r|$ contradicting the fact that $|r| < |x|$. This proves (1).

For Statement 2, if $|v| < |x|$, then $u = wtv$ for some $t$. Here $x_1 = wt = t'w'$ for some $t'$, $w'$ satisfying $|t| = |t'|$ and $|v| = |w| = |w'|$. Since $x_1 \uparrow x_2$, we have $t'w' \uparrow tv$ and by simplification, $t' \uparrow t$. The latter implies the existence of a partial word $t''$ such that $t' \subset t''$ and $t \subset t''$. So $x_1 = t'w' \subset t''w'$ and $x_1 = wt \subset wt''$. Then $t''$ is a border of $x_1$ and $x_1$ is bordered. According to the definition of $u$ being well bordered, $x_1$ is an unbordered partial word and this leads to a contradiction. Hence, we have $|v| \geq |x|$ and, for some $u'$, we have $v = u'x_2$ and $w = x_1u'$, and $u = wx_2 = x_1u'x_2 \subset xu'x$. This proves (2). □

Note that Proposition 2 implies that if $u \in A^+$ is bordered, then $u$ is well bordered. In this case, $u = xu'x$ where $x$ is the minimal border of $u$.

**Lemma 5.** *If $u, v \in A_\Diamond^+$ are such that $u = v_0...v_{n-1}$ where $v_0,...,v_{n-1}$ is a sequence of nonempty unbordered prefixes of $v$, then there exists a unique sequence $v'_0,...,v'_{m-1}$ of nonempty unbordered prefixes of $v$ such that $u \uparrow v'_0...v'_{m-1}$ (the desired sequence is just $v_0,...,v_{n-1}$).*

**Proof.** The statement follows immediately from Proposition 1. □

The badly bordered partial words are now split into the specially bordered and the nonspecially bordered partial words according to the following definition.

**Definition 2.** Let $u \in A_\Diamond^+$ be a partial word that is badly bordered. Let $x$ be a minimal border of $u$, and set $u = x_1v = wx_2$ where $x_1 \subset x$ and $x_2 \subset x$. If there exists a proper factor $x'$ of $u$ such that $x_1 \not\uparrow x'$ and $x' \uparrow x_2$, then we call $u$ specially bordered. Otherwise, we call $u$ nonspecially bordered.

**Lemma 6.** *Let $v \in A_\diamond^+$ be badly bordered. Let y be a minimal border of v, and set $v = y_1w' = wy_2$ where $y_1 \subset y$ and $y_2 \subset y$ (and thus $y_1 \uparrow y_2$). If there exists a sequence $v_0,...,v_{m-1}$ of nonempty unbordered prefixes of v such that $v \uparrow v_0...v_{m-1}$, then $|y1| < |v_{m-1}|$ and v is specially bordered.*

**Proof.** By Definition 1, $y_1$ is bordered. If $|y_1| = |v_{m-1}|$, then both $y_1$ and $v_{m-1}$ are prefixes of $v$, and thus $y_1 = v_{m-1}$. We get that $y_1$ is unbordered, a contradiction. If $|y_1| > |v_{m-1}|$, then set $y_2 = z_1v'$ where $|v'| = |v_{m-1}|$. Since both $y_1$ and $v_{m-1}$ are prefixes of $v$, we get that $v_{m-1}$ is a prefix of $y_1$. So $y_1 = v_{m-1}z_2$ for some $z_2$, and $v = v_{m-1}z_2w' = wz_1v'$ with $v_{m-1} \uparrow v'$. Thus $v$ has a border of length $|v_{m-1}| < |y_1| = |y|$ contradicting the fact that $y$ is a minimal border. And so $|y_1| < |v_{m-1}|$.

Since $v \uparrow v0 ... v_{m-1}$, we have $|v_{m-1}| \leq |v|$. Since $v_{m-1}$ is a prefix of $v$, and $v = y_1w'$ and $|v_{m-1}| > |y_1|$ there exists $z_1$ such that $y_1z_1 = v_{m-1}$. Since $v = wy_2$ and $v_{m-1}$ is compatible with a suffix of $v$, we have $v_{m-1} \uparrow z_2y_2$ for some $z_2$. Thus, we get that $v_{m-1} = y_1z_1 \uparrow z_2y_2$. Since $v_{m-1} \uparrow z_2y_2$, set $v_{m-1} = z_3y_3$ where $z_3 \uparrow z_2$ and $y_3 \uparrow y_2$. So $v_{m-1} = z_3y_3 = y_1z_1$. If $y_3 \uparrow y_1$, then $v_{m-1}$ is bordered, a contradiction with the fact that $v_{m-1}$ is unbordered. Thus $y_3 \not\uparrow y_1$, and since $v_{m-1}$ is a prefix of $v$, we have that $v$ is specially bordered. □

The following example illustrates Lemma 6.

**Example 1.** Consider the partial word

$v = aa \diamond aabbaaaaa \diamond b.$

Here, v is specially bordered (indeed, it has the factor *abb* such that $aa\diamond \not\uparrow abb$ and $a \diamond b \uparrow abb$) and is compatible with a sequence of some of its unbordered prefixes. Indeed, the compatibility

$aa \diamond aabbaaaaa \diamond b$ f $(aa \diamond aabb)$ $(aa \diamond aabb)$

holds. The shortest border of $v$ is *aab* which has length shorter than $aa \diamond aabb$.

**Lemma 7.** *Let $v \in A_\diamond^+$ be well bordered. Then there exists a longest sequence $v_0, v_1,...,v_{m-1}$ of nonempty prefixes of v such that $v \uparrow v_0v_1...v_{m-1}$, $v_j$ is unbordered for every $1 < j < m$, and $v_0$ is unbordered or badly bordered. Moreover, if $v_0$ is badly bordered, then no sequence of nonempty unbordered prefixes of v exists that is compatible with v.*

**Proof.** Let $y_0$ be a minimal border of $w_0 = v$, and set $w_0 = x_0w'_1 = w_1x'_0$ where $x_0 \subset y_0$ and $x'_0 \subset y_0$ (and thus $x_0 \uparrow x'_0$). By Definition 1, $x_0$ is unbordered, and

$v = w_1x'_0 \uparrow w_1x_0$ (1)

where both $w_1$ and $x_0$ are prefixes of $w_0$ (and hence of $v$). If $w_1$ is unbordered, then $v$ is compatible with a sequence of its nonempty unbordered prefixes.

If $w_1$ is badly bordered, then no sequence $v'_0,...,v'_{m'-1}$ of nonempty unbordered prefixes of $v$ exists that is compatible with $w_1$ unless $w_1$ is specially bordered and $|y_1| < |v'_{m'-1}|$ by Lemma 6 (here $y_1$ is a minimal border of $w_1$). If this is the case, then $w_1$ may be compatible with such a sequence of nonempty unbordered prefixes of $v$, and if so replace $w_1$ on the right-hand side of the compatibility in (1) by $v'_0 ... v'_{m'-1}$. If this is not the case, then no sequence of nonempty unbordered prefixes of $v$ exists that is compatible with $v$.

If $w_1$ is well bordered, then repeat the process. Let $y_1$ be a minimal border of $w_1$, and set $w_1 = x_1w'_2 = w_2x'_1$ where $x_1 \subset y_1$ and $x'_1 \subset y_1$ (and thus $x_1 \uparrow x'_1$). By Definition 1, $x_1$ is unbordered, and

$v = w_2x'_1x'_0 \uparrow w_2x_1x_0$ (2)

where both $w_2$ and $x_1$ are prefixes of $w_1$ (and hence of $v$, since $w_1$ is a prefix of $v$) and $x_0$ is a prefix of $v$.

Let $w_0, w_1,...,w_{j-1}$ be the longest sequence of nonempty well-bordered prefixes defined in this manner. For all $0 \leq k < j$, let $y_k$ be a minimal border of $w_k$, and set $w_k = x_kw'_{k+1} = w_{k+1}x'_k$ where $x_k \subset y_k$ and $x'_k \subset y_k$ (and thus $x_k \uparrow x'_k$). Again by Definition 1, $x_0,...,x_{j-1}$ are unbordered. We have $w_{j-1} = w_jx'_{j-1} \uparrow w_jx_{j-1}$ and thus by induction,

$v = w_jx'_{j-1}...x'_0 \uparrow w_jx_{j-1} ... x_0$ (3)

where $wj$, $x_{j-1}$, ... , $x0$ are prefixes of $w0$ (and hence of $v$). Now, if $w_j$ is unbordered, then $v$ is compatible with a sequence of some of its nonempty unbordered prefixes. If $w_j$ is badly bordered, then proceed as in the case above when $w_1$ is badly bordered.

We can thus equate $v$ with sequences of shorter and shorter factors that are some of its prefixes or compatible with some of its prefixes and the existence of the required sequence $v_0,...,v_{m-1}$ is established. □

**Theorem 2.** *Let $u, v \in A_\Diamond^+$ be such that $u \ll v$, and let $v_0,...,v_{m-1}$ be a longest sequence of nonempty prefixes of v satisfying $u \uparrow v_0...v_{m-1}$. Then, either all $v_i$'s are unbordered, or u is not compatible with the concatenation of any sequence of unbordered prefixes of v. In the latter case, some of the $v_i$'s are badly bordered while the others are unbordered.*

**Proof.** If $v_0,...,v_{m-1}$ are unbordered, then by Lemma 5 we get the unique sequence of nonempty unbordered prefixes of $v$ whose concatenation is compatible with $u$. If any of the prefixes are well or badly bordered, then proceed as in Lemma 6 or Lemma 7. □

**Example 2.** Consider the partial words

$u = aaaa \Diamond babbaaaaa \Diamond baa$ and $v = aa \Diamond babbaaaaa \Diamond b$.

We have a factorization of u in terms of nonempty prefixes of v. Here, the compatibility

$u \uparrow (a)\ (a)\ (aa \Diamond babbaaaaa \Diamond b)\ (a)\ (a)$

consists of unbordered and badly bordered prefixes of $v$ and is a longest such sequence ($aa \Diamond babbaaaaa \Diamond b$ is specially bordered and is not compatible with any sequence of nonempty unbordered prefixes of $v$). We can check that no sequence of nonempty unbordered prefixes of $v$ exists that is compatible with $u$.

## 4. More results on concatenations of prefixes

In this section, we give more results on concatenations of prefixes. In particular, we study the properties of the longest unbordered prefix of a partial word. We also investigate the relationship between the minimal weak period of a partial word and the maximal length of its unbordered factors. Our main results in this section (Theorems 3 and 4) extend a result of Ehrenfeucht and Silberger [ 13] which states that if $u = xv$ is a nonempty unbordered word where $x$ is the longest unbordered proper prefix of $u$, then $v$ is unbordered.

If $u \in A_\Diamond^+$, then unb(u) denotes the longest unbordered prefix of $u$. A result of Ehrenfeucht and Silberger shows that if $u, v \in A^*$ are such that $u = \text{unb}(u)v$, then $v \ll \text{unb}(u)$ [ 13]. This does not extend to partial words as $u = (ab)(\Diamond b) = \text{unb}(u)v$ provides a counterexample. However, the following lemma does hold.

**Lemma 8.** *Let $u \in A_\Diamond^+$, $v \in A_\Diamond^*$ be such that $u = \text{unb}(u)v$. Then $u \ll \text{unb}(u)$ if and only if $v \ll \text{unb}(u)$.*

**Proof.** If $v \ll \text{unb}(u)$, then obviously $u \ll \text{unb}(u)$. For the other direction, since $u \ll \text{unb}(u)$, we can write $u = u_0u_1...u_{n-1}$ where each $u_i$ is a nonempty prefix of unb($u$). We can suppose that $v \neq \varepsilon$. Then $\text{unb}(u) = u_0...u_ku'$ for some $k <$ n-1 and some prefix $u'$ of $u_{k+1}$. Since unb($u$) is unbordered, we have that $u' = \varepsilon$, that $k = 0$, and hence that $\text{unb}(u) = u_0$. It follows that $v = u_1...u_{n-1}$ and $v \ll \text{unb}(u)$. □

We get the following corollary.

**Corollary 1.** *Let $u \in A_\Diamond^*$, $v \in A_\Diamond^+$. Then the following hold:*

1. *If $u \ll \text{unb}(v)$, then $u \ll v$.*
2. *If $w \in A_\Diamond^*$ is such that $v = \text{unb}(v)w$ and $w \ll \text{unb}(v)$, then $u \ll v$ if and only if $u \ll \text{unb}(v)$.*

**Proof.** Statement 1 holds trivially. For Statement 2, by Lemma 8, $w \ll \text{unb}(v)$ if and only if $v \ll \text{unb}(v)$. Now, if $u \ll v$, then since $v \ll \text{unb}(v)$, by transitivity we get $u << \text{unb}(v)$. □

Statement 2 of Corollary 1 is not true in general. Indeed, u = ababac o aab and v = abac o aba provide a counterexample. To see this, v = (abac) (oaba) = unb(v)w and we have u << v since u = (ab) (abac o a) (ab) where ab and abac o a are prefixes of v. However u <<~ unb(v) (here w <<~ unb(v)). However, for u, v e A∗ , u << v if and only if u << unb(v) [ 13].

For u, v e A∗ o, when both u << v and v << u we write u ti v. The relation ti is an equivalence relation. A result on words states that for u, v e A∗ , u ti v if and only if unb(u) = unb(v) [ 13]. For partial words, the following holds.

**Proposition 3.** *For u, v* e $A_\Diamond^*$, *if u* t $\approx$ *v, then* unb($u$) = unb($v$).

**Proof.** Suppose that $u \approx v$. Set $v = \text{unb}(v)w$ for some partial word $w$. Since $u \ll v$, we can write $u = v_0...v_{n-1}$ where each $v_i$ is a nonempty prefix of $v$. Since $v \ll u$, there exists a sequence of nonempty prefixes of $u$, say $u_0,...,u_{m-1}$, such that $v = u_0u_1...u_{m-1}$. Since unb($v$) is a prefix of $v$, we have $\text{unb}(v) = u_0...u_ku'$ where $u'$ is a prefix of $u_{k+1}$ and $k < m$-1. Since unb($v$) is unbordered, we have $u' = \varepsilon$, $k = 0$, and $\text{unb}(v) = u_0$. Therefore, unb($v$) is an unbordered prefix of $u$. Hence, it is a prefix of unb($u$). Similarly, unb($u$) is a prefix of unb($v$). □

The converse of Proposition 3 does not necessarily hold for partial words as is seen by considering $u = aba\Diamond$ and $v = ab\Diamond b$. We have $\text{unb}(u) = ab = \text{unb}(v)$ but $u \not\approx v$.

If $v$ is an unbordered word and $w$ is a proper prefix of $v$ for which $u \ll w$, then $uv$ and $wv$ are unbordered [13]. For partial words, we can prove the following.

**Lemma 9.** *Let $u \in A_\Diamond^*$ be unbordered. Then the following hold:*

1. *If $v \in P(u)$ and $v \neq u$, then $vu$ is unbordered.*
2. *If $v \in S(u)$ and $v \neq u$, then $uv$ is unbordered.*

**Proof.** Let us prove Statement 1 (the proof of Statement 2 is similar). Set u = $vx$ for some $x$. If $vu = vvx$ is bordered, then there exist nonempty partial words $r$, $s$, $s'$ such that $vvx \subset rs$ and $vvx \subset s'r$. If $|r| \leq |v|$, then $u = vx$ is bordered by $r$. And if $|r| > |v|$, then $r = v'y$ where $|v'| = |v|$ and this implies that $u = vx$ is bordered by $y$. In either case, we get a contradiction with the assumption that $u$ is unbordered. □

**Lemma 10.** *If $v \in A_\Diamond^*$ is unbordered and $u \ll v$ and $u \neq v$, then $uv$ is unbordered.*

**Proof.** Since $u \ll v$, we can write $u = v_0v_1...v_{n-1}$ where each $v_i$ is a prefix of $v$. Therefore, any prefix of $u$ is a concatenation of prefixes of $v$. Assume that $uv$ is bordered by $y$. If $|y| > |u|$, then set $y = u'y'$ with $u \subset u'$. We get $y'$ a border of $v$ contradicting the fact that $v$ is unbordered. If $|y| < |u|$, then we have the following two cases:

*Case* 1. $y$ contains a prefix of $v_0$.

Here y contains a prefix of $v$ and also a suffix of $v$ and therefore, $y$ is a border of the unbordered word $v$.

*Case* 2. $v_0...v_k v' \subset y$ where $v'$ is a prefix of $v_{k+1}$.

If $v' = \varepsilon$, then $v_0...v_k \subset y$ where $v_k$ is a prefix of $v$. This results in a suffix of $y$ containing both a prefix and a suffix of $v$. Similarly, if $v' \neq \varepsilon$, then factor $y$ as $y = y_1 y_2$ where $v' \subset y_2$. Because $v'$ is a prefix of $v$, we can write $v = v'z \subset y_2 z$. But because $|y_2| < |v|$ and we have assumed that $uv$ is bordered by $y = y_1 y_2$, we must have that $v = z'v''$ with $v'' \subset y_2$. Therefore $y_2$ is a border for $v$. In either case, we get a contradiction with the fact that $v$ is unbordered. □

A result of Ehrenfeucht and Silberger [13] states that if $u$ = punb($u$)$v$ is a nonempty unbordered word where punb($u$) the longest proper unbordered prefix of $u$, then $v$ is unbordered. The partial word $u = ab \lozenge ac$ where punb($u$) = $ab$ and $v = \lozenge ac$ and the partial word u = $abaca \lozenge c$ where punb($u$) = $abac$ and $v = a \lozenge c$ provide counterexamples for partial words. However, when $v$ is full, the following theorem does hold.

**Theorem 3.** *Let* $u \in A_\lozenge^*$ *be unbordered. Then the following hold:*
1. *Let x be the longest proper unbordered prefix of u and let v be such that u = xv. If* $v \in A^*$, *then v is unbordered.*
2. *Let y be the longest proper unbordered suffix of u and let w be such that u = wy. If* $w \in A^*$, *then w is unbordered.*

**Proof.** We prove Statement 1 (Statement 2 can be proved similarly). Assume that $v$ is bordered. Since $v$ is full, there exist nonempty words $z$, $v'$ such that $v = zv'z$ where $z$ is the minimal border of $v$. Then $u$ = punb($u$)$zv'z$, so that punb($u$)$z$ is a proper prefix of $u$ such that |punb($u$)$z$| > |punb($u$)|. It follows that punb($u$)$z$ is bordered, and there exist nonempty partial words $r$, $r_1$, $r_2$, $s_1$, $s_2$ such that punb($u$)$z = r_1 s_1 = s_2 r_2$, $r_1$ c $r$ and $r_2 \subset r$ (here $r$ is a minimal border). Let us consider the following two cases:

Case 1. $|r| > |z|$.
In this case, $r2 = x'z$ where $x'$ is a nonempty suffix of punb($u$). Since $r1$ T $r2$, there exist partial words $x''$, $z'$ such that $r1 = x''z'$ where $x''$ T $x'$ and $z'$ T $z$. But then, $x''z's1 = r1s1$ = punb($u$)$z$ = $s2r2 = s2x'z$. It follows that $x''$ is a prefix of punb($u$) and $x'$ is a suffix of punb($u$) that are compatible. As a result, punb($u$) is bordered.

Case 2. $|r| \leq |z|$.
In this case, $r_2$ is a suffix of $z$ and set $z = sr_2$ for some $s$. We get $u$ = punb($u$)$zv'z = r_1 s_1 v' s r_2$ c $r s_1 v' s r$, whence $r$ is a border of the unbordered partial word $u$. □

A closer look at the proof of Theorem 3 allows us to show the following.

**Theorem 4.** *Let* $u \in A_\lozenge^+$. *Then the following hold:*
1. *Let x be the longest proper unbordered prefix of u and let v be such that u = xv. If v is bordered, then set* $v = z_1 v_1 = v_2 z_2$ *where* $z_1 \subset z$, $z_2 \subset z$ *and where z is a minimal border of v. Then* $xz_1$ *has a minimal border r such that* $|r| \leq |z|$. *Moreover, if v is well bordered, then* $|x| \geq |r|$.

2. *Let y be the longest proper unbordered suffix of u and let w be such that u = wy. If w is bordered, then set* $w = z1v1 = v2z2$ *where* $z1$ c $z$, $z2$ c $z$ *and where z is a minimal border of w. Then* $z2y$ *has a minimal border r such that* $|r| \leq |z|$. *Moreover, if w is well bordered, then* $|y| \geq |r|$.

**Proof.** We prove Statement 1 (Statement 2 can be proved similarly). Then $u$ = punb($u$)$z_1 v_1$, so that punb($u$)$z_1$ is a proper prefix of $u$ longer than punb($u$). It follows that punb($u$)$z_1$ is bordered, and there exist nonempty partial words $r$, $r_1$, $r_2$, $s_1$, $s_2$ such that punb($u$)$z_1 = r_1s_1 = s_2r_2$, $r_1 \subset r$ and $r_2 \subset r$ with $r$ a minimal border. If $|r| > |z|$, then $r_2 = x'z_1$ where $x'$ is a nonempty suffix of punb($u$). Since $r_1 \uparrow r_2$, there exist partial words $x''$, $z'$ such that $r_1 = x'' z'$ where $x'' \uparrow x'$ and $z' \uparrow z_1$. But then, $x'' z's_1 = r_1s_1 =$ punb($u$)$z_1 = s_2r_2 = s_2x'z_1$. It follows that $x''$ is a prefix of punb($u$) and $x'$ is a suffix of punb($u$) that are compatible. As a result, punb($u$) is bordered, which contradicts that punb($u$) is the longest unbordered proper prefix of $u$. And so $|r| < |z|$ and $r2$ is a suffix of $z1$. Set $z1 = sr2$ for some suffix $s$ of $s2$ ($s2$ = punb($u$)$s$). If we further assume that $v$ is well bordered, then we claim that |punb($u$)| > $|r|$. To see this, if |punb($u$)| < $|r|$, then set $r_1$ = punb($u$)$t$ and $z_1 = ts_1$ for some $t$. Since $r_1 \uparrow r_2$, there exist $x'$, $t'$ such that $r_2 = x't'$ and punb($u$) $\uparrow x'$ and $t \uparrow t'$. Since $r_2$ is a suffix of $z_1$, we have that $t'$ is a suffix of $z_1$. Consequently, $t$ is a prefix of $z_1$ and $t'$ is a suffix of $z_1$ that are compatible. So $z_1$ is bordered and we get a contradiction with $v$'s well borderedness, establishing our claim. □

The maximum length of the unbordered factors of a partial word $u$ is denoted by $\mu(u)$. Recall that $p(u)$ denotes the minimal period of a (full) word $u$. Ehrenfeucht and Silberger studied the relationship between $p(u)$ and $\mu(u)$ in [ 13]. Clearly, $\mu(u) \leq p(u)$. Here, we investigate the relationship between the minimal weak periods of a partial word $u$, $p'(u)$, and $\mu(u)$.

**Proposition 4.** *For all* $u \in A_\Diamond^*$, $\mu(u) < p'(u) < p(u)$.

**Proof.** Let $w$ be a factor of $u$ such that $|w| > p'(u)$. Factor $w$ as $w = xw_1 = w_2y$ where $|w_1| = |w_2| = p'(u)$. We have $x(i) = w(i)$ and $y(i) = w(i + p'(u))$. This means that whenever $x(i) \neq y(i)$, $i \in H(x)$ or $i \in H(y)$. Therefore $x \uparrow y$ and $w$ is bordered. So we must have that $\mu(u) \leq p'(u)$. □

For any partial word $u$, Proposition 4 gives an upper bound for the maximum length of the unbordered factors of $u$: $\mu(u) \leq p'(u)$. This relationship cannot be replaced by $\mu(u) < p'(u)$ as is seen by considering $u = aba\Diamond$ with $\mu(u) = p'(u) = 2$.

For any $v$, $w \in A_\Diamond^*$, if there exists a partial word $u$ such that $u \ll w$ and $u \subset v$, then we say that $v$ contains a concatenation of prefixes of $w$. Otherwise, we say that $v$ contains no concatenation of prefixes of $w$. Similarly, if $u \in P(w)$ and $u \subset v$, then we say that $v$ contains a prefix of $w$.

The following result extends to partial words a result on words which states that if $u$, $v$ are words such that $u$ = unb($u$)$v$unb($u$) and unb($u$) is not a factor of $v$, then $v$unb($u$) is unbordered (Corollary 2.5 in [12]).

**Proposition 5.** *Let* $u$, $v \in A_\Diamond^*$ *be such that* $u = hvh$ *where* $h$ *abbreviates* unb($u$). *If* $h$ *is not compatible with any factor of* $v$, *then* $vh$ *is unbordered if one of the following holds:*
1. *$v$ is full,*
2. *$v$ contains a prefix of $h$ or a concatenation of prefixes of $h$.*

**Proof.** For Statement 1, suppose that $v$ is full and there exist nonempty $x$, $w_1$, $w_2$ such that $vh \subset xw_1$ and $vh \subset w_2x$. We must have that $|x| \leq |v|$ or else $h$, which is unbordered, would be bordered by a factor of $x$. If $|h| < |x|$, then there exists $x' \in S(x)$ such that $h \subset x'$ and because $|x| \leq |v|$, there exists $v'$ a factor of $v$ with $v'$ c $x'$ and this says that $v' \uparrow h$, contradicting our assumption. Now, if $|h| \geq |x|$, then set $v = rv'$ and $h = h's$ where $|r| = |s| = |x|$. In this case, $r \subset x$ and $s \subset x$, and there exist nonempty $r \in P(v)$ and $s \in S(h)$ such that $r \uparrow s$. But $r$ is full and so $r \uparrow s$ implies that $s \subset r$. But then, by Lemma 9, we have that $hs$ is unbordered, and so $hr$ is an unbordered prefix of $u$ with length greater than $|h|$. This contradicts the assumption that $h$ = unb($u$), hence $vh$ must be unbordered.

For Statement 2, first assume that $v$ contains a prefix of $h$. Let $v' \in P(h)$ be such that $v' \subset v$. By Lemma 9, since $h$ is unbordered, we have that $v'h$ is unbordered. Now, assume that $v$ contains a concatenation of prefixes of $h$. Let $v'$ be such that $v' << h$ and $v' \subset v$. By Lemma 10, since $h$ is unbordered and $v' \ll h$, we have that $v'h$ is unbordered. In either case, since $v' \subset v$, $vh$ is unbordered as well. □

**5. Critical factorizations**

In this section, we first discuss so-called critical factorizations of a partial word w, then study some of their properties when *w* is unbordered (Proposition 6, and Corollaries 2 and 3), and finally investigate the position in the Chomsky hierarchy of the set of all partial words having a critical factorization (Theorems 5 and 6).

If w is a *nonspecial* partial word of length at least two, then there exists a factorization $(u, v)$ of $w$ with $u, v \neq \varepsilon$ such that the minimal local period of $w$ at position $|u|$ - 1 (as defined below) equals the minimal weak period of $w$ [5,6]. Such a factorization $(u, v)$ of $w$ is called *critical* and the position $|u|$ - 1 is called a *critical point* of *w*.

**Definition 3** (*[5]*). Let $w \in A_\Diamond^+$. A positive integer $p$ is called a local period of $w$ at position $i$ if there exist $u, v, x, y \in A_\Diamond^+$ such that $w = uv$, $|u| = i + 1$, $|x| = p$, $x \uparrow y$, and such that one of the following conditions holds for some partial words $r$, $s$:

1. $u = rx$ and $v = ys$ (internal square),
2. $x = ru$ and $v = ys$ (left-external square if $r \neq \varepsilon$),
3. $u = rx$ and $y = vs$ (right-external square if $s \neq \varepsilon$),
4. $x = ru$ and $y = vs$ (left- and right-external square if $r, s \neq \varepsilon$).

The minimal local period of $w$ at position $i$ is denoted by $p(w, i)$. Clearly, $1 < p(w, i) < p'(w) < |w|$.

There exist unbordered partial words that have no critical factorizations, like $w = a \Diamond bc$.

We now investigate some of the properties of an unbordered partial word of length at least two and how they relate to its critical factorizations (if any).

**Definition 4.** Let $u, v \in A_\Diamond^+$. We say that $u$ and $v$ overlap if there exist partial words $r$, $s$ satisfying one of the following conditions:

1. $r \uparrow s$ with $u = ru'$ and $v = v's$,
2. $r \uparrow s$ with $u = u'r$ and $v = sv'$,
3. $u = ru's$ with $u' \uparrow v$,
4. $v = rv's$ with $v' \uparrow u$.

Otherwise we say that $u$ and $v$ do not overlap.

**Proposition 6.** Let $u, v \in A_\Diamond^+$. If $w = uv$ is unbordered, then $|u|$ - 1 is a critical point of $w$ if and only if $u$ and $v$ do not overlap.

**Proof.** Let us first consider the first implication and let us suppose that $u$ and $v$ overlap. If we have Type 1 overlap, then $w = ru'v's$ and $r \uparrow s$ for some partial words $r, s, u', v'$. This contradicts the fact that $w$ is unbordered. If we have Type 2 overlap, then $w = u'rsv'$ and there is an internal square at position $|u|$ - 1 of length $k = |r| = |s|$, so $p(w, |u| - 1) \leq k$. But because $w$ is unbordered, $p'(w) = |w|$. Of course we have that $k < |w|$ (otherwise we have Type 1 overlap), so this contradicts that $|u|$ - 1 is a critical point of $w$. If we have Type 3 overlap, then $w = ru'sv$ and there is a right-external square of length $|u's|$ at position $|u| - 1$. Because $v \neq \varepsilon$, $|u's| < |w| = p'(w)$ and we have that $|u|$ - 1 cannot be a critical point of $w$, a contradiction. The case for Type 4 overlap is very similar to Type 3.

For the other direction we have that $u$ and $v$ do not overlap and let us suppose that $|u|$ - 1 is not a critical point of $w$.

Since $|u|$ - 1 is not a critical point, there exist $x$ and $y$ defined as in Definition 3, with the length of $x$ strictly smaller than the minimal weak period of $w$. Let us now look at all the four conditions of the definition. If we have an internal square, then according to Definition 4 we have a Type 2 overlap of $u$ and $v$, which is a contradiction with our assumption. For a left-external, respectively right-external, square we get that either $u$ is compatible with a factor of $v$, or $v$ is compatible with a factor of $u$. Both cases contradict with the fact that $u$ and $v$ do not overlap, giving us a Type 4, respectively Type 3, overlap.

In the case we have a left- and right-external square we get that $x = ru$ and $y = vs$, where $x \uparrow y$ and $r, s \neq \varepsilon$. If $|r| < |v|$, then there exists $v'$ with $|v'| > 0$, such that $v = rv'$. Hence, since $ru \uparrow rv's$ we get a Type 2 overlap, $u \uparrow v's$, which is a contradiction with our initial assumption. If $|r| \geq |v|$, then there exists $r'$ such that $r = vr'$. This implies that $|w| = |uv| < |vr'u| = |ru| = |x| < p'(w) < |w|$, which is a contradiction. □

**Corollary 2.** *Let $u, v \in A_\diamond^+$. If $w = uv$ is unbordered and $|u| - 1$ is a critical point of $w$, then $w' = vu$ is unbordered as well.*

**Proof.** This is immediately implied by Proposition 6 and the fact that if $w' = vu$ is bordered, then $u$ and $v$ must overlap. □

**Corollary 3.** *Let $u, v \in A_\diamond^+$. If $w = uv$ is unbordered and $|u| - 1$ is a critical point of $w$, then $|v| - 1$ is a critical point of $w' = vu$.*

**Proof.** By Proposition 6, u and v do not overlap. By Corollary 2, w' is unbordered. Then by Proposition 6, the point JvJ — 1 is critical for w'. □

We end this section by considering the language

$CrFa = \{w \mid w$ is a partial word over $A$ that has a critical factorization$\}$

where $A$ denotes an arbitrary nonunary fixed finite alphabet (we will assume that $a$ and $b$ are two distinct letters of $A$). What is the position of *CrFa* in the Chomsky hierarchy? We prove that *CrFa* is a context sensitive language that is not context-free.

Let us first recall a version of the pumping lemma that is due to Bader and Moura [1], and is a generalization of the well-known Ogden's Lemma.

**Lemma 11** (*[1]*). *For any context-free language $L$, there exists $n \in \mathbb{N}$, the set of nonnegative integers, such that for all $z \in L$, if $d$ positions in $z$ are "distinguished" and $e$ positions are "excluded," with $d > n(e + 1)$, then there exist $u, v, w, x, y$ such that $z = uvwxy$ and*

1. *vx contains at least one distinguished position and no excluded positions,*
2. *if $r$ is the number of distinguished positions and $s$ is the number of excluded positions in $vwx$, then $r < n(s + 1)$,*
3. *for all $i \in \mathbb{N}$, $uv^iwx^iy \in L$.*

The above lemma says that for any context-free language $L$, there exists a natural number $n$, such that in any word $z \in L$, by marking any $d$ positions as "distinguished" and $e$ positions as "excluded" with $d > n(e + 1)$, we can decompose $z$ into five contiguous factors that satisfy the three statements. It is easy to observe that the only restrictions imposed by $d$ and $e$ are on the three inner factors $v$, $w$ and $x$.

**Theorem 5.** *The language CrFa is not context-free.*

**Proof.** Let us assume that the language CrFa is context-free. This implies that the previously defined pumping lemma holds. Let us take the word

$z = ba^{3n^3}ba^{n^3}\ \lozenge^{n^3}\ a^{n^3}ba^{3n^3}b$

where $n$ is the natural number from the lemma, and mark all symbols except the first and the last one as distinguished and these two as excluded. It is easy to check that $p'(z) = 3n^3 + 1$, $z$ has a critical factorization $(b, a^{3n^3}ba^{n^3}\ \lozenge^{n^3}\ a^{n^3}ba^{3n^3}b)$ and the number of distinguished positions is greater than $n^{(2+1)}$. From Lemma 11(1) we get that the first and the last occurrences of $b$ will never be part of either $v$ or $x$.

Let us first consider the case when $u = \varepsilon$. This implies, by Lemma 11(1), that $v = \varepsilon$. Hence, $w$ contains exactly one excluded position, implying $x = a^k$, where $0 < k < n^2$ by Lemma 11(2). In this case, for $i = 0$, we obtain the word

$ba^{3n^3-\mathrm{k}}ba^{n^3}\ \lozenge^{n^3}\ a^{n^3}ba^{3n^3}b$

which is not in *CrFa*, contradicting Lemma 11(3). To see that this word does not have a critical factorization, note that it has minimal weak period greater than $3n^3 + 1$. However, the minimal local periods at the positions defined by the factorization (b, $a^{3n^3-k}, b, a^{n^3}\ \lozenge^{n^3}\ a^{n^3}, b, a^{3n^3}b$) are $3n^3 - k + 1$, $3n^3 - k + 1$, $n^3 + 1$, $n^3 + 1$, $3n^3 + 1$ and $3n^3 + 1$ respectively, while the minimal local period at any other position is 1. Similarly we easily prove that it is impossible to have $y = \varepsilon$.

From now on, let us consider the cases where both u and y are nonempty. Then each of $u$ and $y$ contains an excluded position and so $vwx$ will all be distinguished. And therefore the length of $vwx$ is at most $n$ by Lemma 11(2).

When $vwx$ matches $a^*$ and $vwx$ is part of the 1st group of $a$'s, then $vwx = ak$ for some $0 < k < n$, and $v = a^{k_2}$ and $x = a^{k_2}$ with $k_1 > 0$ or $k_2 > 0$. In this case take $i = 0$. The 1st group of $a$'s is then reduced to $3n^3 - k_1 - k_2$, giving us the word

$ba^{3n^3-\ k_1k_2}ba^{n^3}\ \lozenge^{n^3}\ a^{n^3}ba^{3n^3}b$

that does not have a critical factorization (again, the minimal weak period is greater than $3n^3 + 1$ while the minimal local periods are smaller than or equal to $3n^3 + 1$). A similar argument works for the 2nd, 3rd and 4th groups of $a$'s. We are left with the cases when $vwx$ matches $a^*ba^*$, or $a^*\lozenge^*$ or $\lozenge^*a^*$.

If $x$ matches $a^*ba^*$, then v is a string of $a$'s of length at most $n - 1$ with the $a$'s either from the 1st group or the 3rd group. In both cases, taking $i = 0$, we get a contradiction with the fact that the words

$ba^{4n^3-k_1}\ \ \lozenge^{n^3}\ a^{n^3}ba^{3n^3}b$

and

$ba^{3n^3}ba^{n^3}\ \lozenge^{n^3}\ a^{4n^3-k_2}b$

are in *CrFa* for some $0 < k_1, k_2 < n$. To see that the first word does not have a critical factorization, note that it has minimal weak period greater than $4n^3 + 1$. However, the minimal local periods at the positions defined by the factorization ($b$, $a^{4n^3-k_1}\ \lozenge^{n^3}\ a^{n^3}$, b, $a^{3n^3}$, $b$) are $4n^3 - k_1 + 1$, $n^3 + 1$, $3n^3 + 1$ and $3n^3 + 1$ respectively, while the minimal local period at any other position is 1. The case where $v$ matches $a^*ba^*$ is solved analogously to the previous one and hence we will omit its proof. By taking $i = 2$, a contradiction is reached in the cases where $v = a^{k_1}$ and $x = a^{k_2}$ for some $k_1$, $k_2$ with the $a$'s in $v$ from the 1st group of $a$'s, and the ones in $x$ from the 2nd group of $a$'s (respectively, with the $a$'s in $v$ from the 3rd group of $a$'s, and the ones in $x$ from the 4th group of $a$'s).

If $v = a^{k_1} \Diamond^{k_2}$ or $x = \Diamond^{k_1} a^{k_2}$ for some $k_1$, $k_2$, then we get that $x = \Diamond^{k_3}$, respectively $v = \Diamond^{k_3}$, with $0 < k_1 + k_2 + k_3 < n$. In both cases, taking $i = 2$, we obtain a word that does not have a critical factorization. When $v = \Diamond^{k_1} a^{k_2}$ or $x = a^{k_1} \Diamond^{k_2}$, we proceed similarly. The case $vx = a^k$ where $0 < k < n$, with the $a$'s from the 2nd or the 3rd group, is solved similarly.

Since all cases lead to contradictions we conclude that our assumption is false, hence the language *CrFa* is not context-free. □

**Theorem 6.** *The language CrFa is context sensitive.*

**Proof.** To prove this we will give an LBA (linear bounded automaton) that recognizes all partial words having a critical factorization. We recall that the factorization $(u, v)$ of input partial word $w$ is critical if the minimal local period of $w$ at position $|u|$ - 1 is equal to the minimal weak period of $w$, $p'(w)$.

Our LBA will have an input tape of size $3|w|$ and five auxiliary tapes of size at most $|w| + 1$, that we are going to describe next. We will denote the word on the input tape as inp.

The input tape will contain, starting from position $|w|$, the input word while all other positions will be filled in with ◊'s. Position $|w|$ (respectively, $2|w| - 1$) on the input tape can be easily recognized by using an auxiliary symbol \$ (respectively, #).

The first auxiliary tape, let us call it $P$, will have size 1w1 and will be used for the identification of the minimal weak period of our input word w. This can be easily done by using an unary numbering system that adds 1's until the minimal weak period is discovered. Since the minimal weak period of a word is greater than or equal to one, we start with a 1 symbol on the tape.

The second tape, $Z$, will be used for remembering the current position in the word. Hence, for position $i < |w|$, the head will be positioned on the input tape on the $(|w| + i)$th cell, and Tape $Z$ will contain $i$ ones. The tape is initialized with one 1 and has size $|w| + 1$.

The following tape, $X$, will have size $p'(w)$ and will be used for checking the size of the current minimal local period.

The last two tapes, called $Y_1$ and $Y_2$, will have sizes $p'(w)$. They will be used to save the words of length at most $p'(w)$, positioned to the left and right of the current position. More exactly these tapes will contain $x$ and $y$ from the definition of critical factorization.

We now describe how the LBA works, using the notation 1T1 for denoting the number of symbols present on Tape $T$:

1. Starting at position $|w|$ on the input tape, the head marks the current position and then moves to the right $|P|$ positions and checks if the symbols are compatible. This step is repeated until the condition is violated. If this happens, then a 1 is added to Tape $P$ and all symbols are unmarked. If the end of the word is reached, then the head moves left to the position $|w|$ and repeats the step for the first unmarked symbol. The step is repeated until all symbols are marked or $|P| = |w|$. This will give us the minimal weak period of the word.
2. Increment the value of $X$.
3. Starting at position $i$, where $i$ represents the sum between 1$w$1 and the number of 1's on Tape $Z$, the LBA copies the suffix of length $|X|$ (recall that the number of symbols present on Tape $X$, or $|X|$, is bounded by $p'(w)$) of the word inp[0..$i$) on Tape $Y_1$ and the prefix of length $|X|$ of the word inp[$i$..3$|w|$) on Tape $Y_2$.
4. Next the LBA checks if the word on Tape $Y_1$ is compatible with the word on Tape $Y_2$. This can easily be done just by comparing one symbol at a time while going in parallel on the two tapes. If the words are

compatible and the sum of 1's in $X$ is equal to $p'(w)$, then the automaton stops and outputs the position where a critical factorization is present (the LBA will accept the word). If the words are compatible and the sum of 1's in $X$ is not equal to $p'(w)$, then the automaton fills the $X$ tape with 1's and goes to the next step.
5. If $X$ is full, then the tape is brought to the initial configuration and the LBA adds a 1 on $Z$. If $Z$ is full, then the automaton stops and concludes that a critical factorization does not exist, hence, the LBA will reject the word. Otherwise, the LBA goes to Step 2.

It is easy to check that the algorithm will always stop. Since the construction of a linear bounded automaton that recognizes all partial words over $\{a, b\}$ having a critical factorization was possible, we conclude that *CrFa* is a context sensitive language. □

**Acknowledgments**
The first author gratefully acknowledges a research assignment from the University of North Carolina at Greensboro. A part of this assignment was carried out at the LIAFA: Laboratoire d'Informatique Algorithmique: Fondements et Applications of Université Paris 7, Paris, France.

## References

1. C. Bader, A. Moura, A generalization of Ogden's lemma, J. ACM 29 (1982) 404–407.
2. J. Berstel, L. Boasson, Partial words and a theorem of Fine and Wilf, Theoret. Comput. Sci. 218 (1999)135–141.
3. J. Berstel, D. Perrin, Theory of Codes, Academic Press, Orlando, FL, 1985.
4. F. Blanchet-Sadri, Primitive partial words, Discrete Appl. Math. 148 (2005)195–213.
5. F. Blanchet-Sadri, S. Duncan, Partial words and the critical factorization theorem,J. Combin. Theory Ser. A 109 (2005) 221–245. www.uncg.edu/mat/cft.
6. F. Blanchet-Sadri, N.D. Wetzler, Partial words and the critical factorization theorem revisited, Theoret. Comput. Sci. 385 (2007) 179–192. http://www.uncg.edu/mat/research/cft2.
7. D. Breslauer, T. Jiang, Z. Jiang, Rotations of periodic strings and short superstrings, J. Algorithms 24 (1997) 340–353.
8. P. Bylanski, D.G.W. Ingram, Digital Transmission Systems, IEE, 1980.
9. M. Crochemore, D. Perrin, Two-way string matching, J. Assoc. Comput. Mach. 38 (1991) 651–675.
10. M. Crochemore, W. Rytter, Text Algorithms, Oxford University Press, 1994.
11. M. Crochemore, W. Rytter, Jewels of Stringology, World Scientific, 2003.
12. J.P. Duval, Relationship between the period of a finite word and the length of its unbordered segments, Discrete Math. 40 (1982) 31–44.
13. Ehrenfeucht, D.M. Silberger, Periodicity and unbordered segments of words, Discrete Math. 26 (1979)101–109.
14. Z. Galil, J. Seiferas, Time-space optimal string matching, J. Comput. System Sci. 26 (1983) 280–294.
15. D. Margaritis, S. Skiena, Reconstructing strings from substrings in rounds, in: Proceedings of the 36th Annual Symposium on Foundations of Computer Science, 1995, pp. 613–620.
16. J.A. Storer, Data Compression: Methods and Theory, Computer Science Press, Rockville, MD, 1988.