A note on antichains of words

James D. Currie*
Department of Mathematics and Statistics
University of Winnipeg
Winnipeg, Manitoba
Canada R3B 2E9
currie@uwinnipeg.ca

Submitted: July 9, 1995; Accepted: October 14, 1995

Abstract

We can compress the word 'banana' as $xyyz$, where x = 'b', y = 'an',z = 'a'. We say that 'banana' *encounters* yy. Thus a 'coded' version of yy shows up in 'banana'. The relation 'u encounters w' is transitive, and thus generates an order on words. We study antichains under this order. In particular we show that in this order there is an infinite antichain of binary words avoiding overlaps.

AMS Subject Classification: 68R15, 06A99
Key Words: overlaps. antichains, words avoiding patterns

1 Introduction

The study of words avoiding patterns is an area of combinatorics on words reaching back at least to the turn of the century, when Thue proved [29] that one can find arbitrarily long words over a 3 letter alphabet in which no two adjacent subwords are identical. If w is such a word, then w cannot be written $w = xyyz$ with y a non-empty word. In modern parlance, we would say that w **avoids** yy. A word which can be written as $xyyz$ is said to **encounter** yy. Thue also showed that there are arbitrarily long words over a 2 letter alphabet avoiding yyy. One can quickly check that no word of length 4 or more over a 2 letter alphabet avoids yy. We say that yyy is **avoidable** on 2 letters, or **2-avoidable** whereas yy is **unavoidable** on 2 letters. On the other hand, yyy is certainly 3-avoidable, because any word avoiding yy must avoid yyy also.

Bean, Ehrenfeucht and McNulty [4], and independently Zimin [30], characterized words which are avoidable on some large enough finite alphabet. If p is a word over an n letter alphabet, then

*This work was supported by an NSERC operating grant.

p is avoidable on some finite alphabet if Z_n avoids p, where words Z_n are defined recursively by

$$\begin{aligned} Z_1 &= 1 \\ Z_{n+1} &= Z_n(n+1)Z_n, n \in \mathbb{N}. \end{aligned}$$

Thus the pattern *abcacb* is a pattern over a 3 letter alphabet, and is avoided by $Z_3 = 1213121$. It follows that *abcacb* is avoidable on some large enough finite alphabet. The size of the smallest alphabet on which *abcacb* is avoidable isn't known. No avoidable pattern is known which is not 4-avoidable. The following conjecture is given by Baker [2]:

Conjecture 1.1 Every avoidable pattern is 4-avoidable.

The following problem has been open since 1979 [4]:

Problem 1.2 Find an algorithm which given a word p determines the smallest k such that p is k-avoidable.

Cassaigne and Roth [8, 27] studied avoidable binary and ternary patterns p, giving when possible the smallest k for which p is k-avoidable. In such work, the most important patterns are the minimal ones; as discussed above, yyy is 3-avoidable because it contains yy. Similarly, if w is 3-avoidable, then so is w^R, the reverse of w. It follows from the work of Cassaigne and Roth that a binary pattern is 2-avoidable exactly when it encounters one of xxx, $xyxyx$, $xyxxy$, $xxyxyy$, $xyxyyx$ and $xxyyx$. Consideration of minimal k-avoidable patterns leads to problems such as the following, posed in [3]:

Problem 1.3 Write $u \geq w$ if u encounters w or the reversal of w. This relation is a quasi-order, and factoring out the resulting equivalence relation gives a partial order. Let $\mu(w)$ be the size of the smallest alphabet on which w is avoidable. For avoidable w, is there an infinite antichain on $\mu(w)$ letters such that each member of the antichain avoids w?

Perhaps the posing of this problem is too ambitious. An affirmative solution would imply that for any avoidable word w, it takes no more letters to avoid w and w^R simultaneously than it does to avoid w alone. Strong evidence to the contrary is provided by the example $w = abcacb$. It seems likely that *abcacb* is 2-avoidable; there are words of length 1000 over $\{0, 1\}$ avoiding *abcacb*. However, *abcacb* and *bcacba* are not simultaneously 2-avoidable.[1]

For this reason, it seems that a better question to ask is the following:

Problem 1.4 Write $u \geq w$ if u encounters w. For avoidable w, is there an infinite antichain on $\mu(w)$ letters such that each member of the antichain avoids w?

This question was answered in the affirmative in the case where w is 2-avoidable in [18] Note that for the sake of studying antichains it is unnecessary to move from the quasi-order to a partial order.

In a related paper [9] it was shown that for any $\epsilon > 0$ there is an infinite antichain of such ternary words avoiding x^k for $7/4 < k < 7/4 + \epsilon$. Note that 7/4 is the **threshold** of repetition

[1] Thanks to Kirby Baker for the use of his software which allowed the author to make these discoveries concerning *abcacb*.

for words over a 3 letter alphabet; if $r < 7/4$, we can find at most finitely many words over a 3 letter alphabet avoiding x^r.

The threshold of repetition for a 2 letter alphabet is 2. A word containing no subwords of the form x^k for $k > 2$ is called **overlap-free**. A word is overlap-free exactly when it avoids both the patterns xxx and $xyxyx$.

In this note we prove that any binary word which avoids overlaps is an element of an infinite antichain of binary words avoiding overlaps.

2 Preliminaries

An **alphabet** Σ is a set whose elements are called **letters**. A **word** w over Σ is a finite string of letters from Σ. The **length** of word w is the number of letters in w, denoted by $|w|$. Thus $|banana| = 6$, for example. The language consisting of all words over Σ is denoted by Σ^*. If $x, y \in \Sigma^*$, the **concatenation** of x and y, written xy, is simply the string consisting of x followed by y. The word with no letters is called the **empty word** and is denoted by ϵ. Suppose $w \in \Sigma^*$. We call word x a **prefix** of w if $w = xy$, some $y \in \Sigma^*$. Similarly word y is a **suffix** of w if we can write $w = xy$, some $x \in \Sigma^*$. We call y a **subword** of w if we can write $w = xyz$, some $x, z \in \Sigma^*$. In the case that x and z are non-empty, y is an **internal subword** of w.

Let Σ, T be alphabets. A **substitution** $h : \Sigma^* \to T^*$ is a function generated by its values on Σ. That is, suppose $w \in \Sigma^*$, $w = a_1a_2 \dots a_m$; $a_i \in \Sigma$ for $i = 1$ to m. Then $h(w) = h(a_1)h(a_2) \dots h(a_m)$. A substitution is **non-erasing** if for every $a \in \Sigma$, $|h(a)| \neq \epsilon$.

Let w, v be finite words over some alphabet Σ. We say that w **encounters** v if $w = xh(v)y$ for some non-erasing substitution $h : \Sigma^* \to \Sigma^*$. Otherwise we say that w **avoids** v. If x is a word we denote by x^n the word consisting of x repeated n times in a row. Thus $x^2 = xx, x^3 = xxx$ and so on. We call a word w a k**-power** if $w = x^k$, some $x \neq \epsilon$. A 2-power is also called a **square**. An **overlap** is a word of form xxx or $xyxyx$ for some words x and y. A word w is k**-power free** if we cannot write $w = xyz$, where y is a k-power. Thus w is k-power free if w avoids x^k. Similarly one speaks of **square-free** or **overlap-free** words.

An ω**-word** over alphabet Σ is an infinite sequence of letters of Σ. If $w = \{w_i\}_{i \in \mathbb{N}}$ is an ω-word over Σ, then each finite initial segment $w_1, w_2, \dots, w_n$ of w will correspond to some word $w_1w_2 \dots w_n$ of Σ^*. In this case we say that $w_1w_2 \dots w_n$ is a **prefix of** ω**-word** w**.** If u is an ω-word over Σ we say that u encounters w if some finite prefix of u encounters w. Otherwise, we say that u avoids w.

We say that w is k**-avoidable** if the set of words over Σ avoiding w is infinite, for some, hence for any, alphabet Σ of size k. Equivalently, w is k-avoidable if there is an ω-word over an alphabet of size k which avoids w. If w is k-avoidable for some $k \in \mathbb{N}$ we say that w is **avoidable**. Otherwise, w is **unavoidable**. Let S be a set of words. We say that v avoids S if v avoids each $w \in S$.

Fix an alphabet Σ. The relation 'w encounters v' is a quasi-order on Σ^* which we will abbreviate by $w \geq v$. We will be interested in the quasi-ordered set $\langle \Sigma^*, \geq \rangle$.

Lemma 2.1 *Suppose that $\mathcal{A} \subseteq \Sigma^*$ is an infinite antichain. Then there is an ω-word over Σ avoiding $\mathcal{A}$.*

Proof: Let $\mathcal{A} = \{w_i\}_{i=1}^{\infty}$. If w is a non-empty word, denote by w' the word obtained from w by deleting the last letter.

We claim that for each $i \in \mathbb{N}, w'_i$ avoids $\mathcal{A}$: If $v \leq w'_i$ then $v \leq w_i$ by transitivity. Thus if $j \neq i$, then w'_i avoids w_j, because w_j and w_i are incomparable. On the other hand, since w'_i is shorter than w_i, certainly w'_i avoids w_i.

Since $\mathcal{A}$ is an infinite set of words over a finite alphabet, $\mathcal{A}$ contains arbitrarily long words. Thus the set $\mathcal{A}' = \{w'_i\}_{i=1}^{\infty}$ contains arbitrarily long words of Σ^* avoiding $\mathcal{A}$. It follows by König's Infinity Lemma that there is an ω-word over Σ avoiding $\mathcal{A}$.□

Lemma 2.2 *Let S be a finite set of avoidable words. Then there is an ω-word over a finite alphabet avoiding S.*

Proof: This is proved in [4, 30]. Let $S = \{s_i : 1 \leq i \leq m\}$. For each i pick $n_i \in \mathbb{N}$ and an ω-word $w_i = \{w_{ij}\}_{j=1}^{\infty}$ over an alphabet Σ_i of size n_i avoiding s_i. Then the word $w = \{(w_{1j}, w_{2j}, \ldots, w_{mj})\}_{j=1}^{\infty}$ over the alphabet $\prod_{i=1}^{m} \Sigma_i$ avoids s.□

To avoid S it suffices to avoid a maximal antichain of minimal elements of S. Thus we could get by with a version of this lemma in which S was restricted to be an antichain.

Corollary 2.3 *Suppose that $\mathcal{A} \subseteq \Sigma^*$ is a finite antichain. Then there is an ω-word over some finite alphabet avoiding $\mathcal{A}$.*

Remark 2.4 It is striking that infinite antichains over finite sets are easier to avoid than finite antichains! That is, to avoid a finite antichain over S it may be necessary to move to a larger alphabet, whereas this is not the case with infinite antichains. To give a concrete example, let S be the set of all words of length 7 over $\Sigma = \{a, b\}$. Each word of S is 2-avoidable, but any binary word of length 7 or more encounters an element of S.

An image of xxx or $xyxyx$ under a non-erasing substitution is called an **overlap**. Note that a prefix of an overlap will be a square.

Theorem 2.5 *There is an infinite antichain of binary words avoiding overlaps.*

Proof: Following Thue [29], Define the map $h : \{a, b\}^* \rightarrow \{a, b\}^*$ by $h(a) = ab, h(b) = ba$. Let l = aabaab. Thus $l^R = baabaa$. We see that the prefixes of l which are suffixes of l^R are exactly a, aa, $aabaa$.

Lemma 2.6 *Word l is avoided by $h^{\omega}(a)$.*

Proof: This was proved by Cassaigne [8, Section 2.6, Théorème 2.2]. □

Corollary 2.7 *The word l^R, the reverse of l, is avoided by $h^{\omega}(a)$.*

Let $n \in \mathbb{N}$. Then we can write $h^{2n+2}(a) = abbabaabu_nbaababba$ for some word u_n. Let $m_n = aabaabu_nbaabaa$.

Remark 2.8 Word l is a prefix and suffix of each m_n, but every internal subword of m_n is a subword of $h^{\omega}(a)$. It follows that l doesn't appear in m_n internally. We note that for each n, m_n is a palindrome.

Lemma 2.9 *Let $n \in \mathbb{N}$. Then the word m_n is overlap-free.*

Proof: Every internal subword of m_n is a subword of $h^\omega(a)$, and is overlap-free. It remains to show that no prefix or suffix of m_n is an overlap. As m_n is a palindrome, we need only show that no prefix of m_n is an overlap.

First note that no prefix of m_n is a square of length 12 or greater; otherwise the prefix of m_n of length 6 reappears internally, i.e. m_n contains an l internally, which is impossible. It follows that the shortest overlap which is a prefix of m_n has length at most 11, and is a prefix of m_1. Inspection shows that no prefix of m_1 of length 11 or less is an overlap.□

Theorem 2.10 *The set $\{m_n\}_{n\in\mathbb{N}}$ is an antichain.*

Proof: Let $i,j \in \mathbb{N}, i<j$. Clearly m_i doesn't encounter m_j since m_j is longer than m_i. Suppose, for the sake of a contradiction, that m_j encounters m_i. Say that $m_j = \alpha f(m_i)\beta$ where f is a non-erasing substitution.

If $\alpha \neq \epsilon$ then an internal subword of m_j encounters the prefix l of m_i. Thus a subword of $h^\omega(a)$ encounters l. This is impossible by Lemma 2.6. Thus $\alpha = \epsilon$. Symmetrically, $\beta = \epsilon$.

Since a is a prefix of m_i, $f(a)$ is a prefix of m_j. Thus l and $f(a)$ are both prefixes of m_j, and one must be a prefix of the other. Since a is an internal subword of m_i, $f(a)$ is an internal subword of m_j. Thus l is not a prefix of $f(a)$. We conclude that $f(a)$ is a proper prefix of l. Symmetrically, $f(a)$ must be a proper suffix of l^R, the reverse of l. Thus $f(a) = a$, aa or $aabaa$. However, aa is a subword of m_i, so that $f(aa)$ is a subword of the overlap-free word m_j. We conclude that $f(a) = a$.

We have so far determined $f(a)$. Now consider $f(b)$. Since $f(m_i) = m_j$ is longer than m_i, the length of $f(b)$ is greater than 1. Since aab and baa are subwords of m_i, both $aaf(b)$ and $f(b)aa$ are subwords of m_j. It follows that $f(b)$ has b as a prefix and a suffix, since otherwise aaa appears in m_j, a contradiction since m_j is overlap-free.

Since aab is a prefix of m_i, $f(aab) = aaf(b)$ is a prefix of m_j. However, another prefix of m_j is $aabaab$. Thus one of $f(b)$ and $baab$ is a prefix of the other. Since $f(b)$ begins and ends with a b, we conclude that $baab$ is a prefix of $f(b)$. As aab appears internally in m_i, we conclude that $f(aab)$ occurs internally in m_j. A prefix of $f(aab)$ is $aabaab = l$. It follows that l appears internally in m_j, and hence that l is a subword of $h^\omega(a)$. This contradicts Lemma 2.6.

The supposition that m_j encounters m_i leads to a contradiction. We conclude that m_i and m_j are incomparable, and that in fact $\{m_n\}_{n\in\mathbb{N}}$ is an antichain. □.

References

[1] S. Aršon, Démonstration de l'existence des suites asymétriques infinies, Mat. Sb., (N.S.) **2**, 769–779; Zbl. **18**, 115.

[2] Kirby A. Baker, Some problems on avoidability. Exposé au LITP, 1992.

[3] Kirby A. Baker, George. F. McNulty & Walter Taylor, Growth problems for avoidable words, Theoret. Comput. Sci. **69** (1989), no. 3, 319–345; MR **91f**:68109.

[4] Dwight R. Bean, Andrzej Ehrenfeucht & George McNulty, Avoidable Patterns in Strings of Symbols, Pacific J. Math. **85** (1979), 261–294; MR **81i**:20075.

[5] J. Brinkhuis, Non-repetitive sequences on three symbols, Quart. J. Math. Oxford Ser.(2) **34** (1983), 145–149; MR **84e**:05008.

[6] T. C. Brown, Is there a sequence on four symbols in which no two adjacent segments are permutations of each other?, Amer. Math. Monthly **78** (1971), 886–888.

[7] Stanley Burris & Evelyn Nelson, Embedding the dual of in the lattice of equational classes of semigroups, Algebra Universalis, **1** (1971/72), 248–253; MR **45** #5257.

[8] Julien Cassaigne, Motifs évitables et régularité dans les mots, Thèse de Doctorat, Université Paris VI, Juillet 1994.

[9] Max Crochemore & Pavel Goralcik, Mutually avoiding ternary words of small exponent, Internat. J. Algebra Comput. **1** (1991) 407–410.

[10] James D. Currie, Non-repetitive walks in graphs and digraphs, PhD thesis, University of Calgary (1987).

[11] James D. Currie, Which graphs allow infinite non-repetitive walks? Discrete Math. **87** (1991), 249–260; MR **92a**:05124.

[12] James. D. Currie, Open problems in pattern avoidance, Amer. Math. Monthly **100** (1993), 790–793.

[13] James. D. Currie, Non-repetitive words: ages and essences, Combinatorica, to appear.

[14] F.M.Dekking,Strongly non-repetitive sequences and progression-free sets, J. Combin Theory Ser. A, **27** (1979), 181–185; MR **81b**:05027.

[15] Françoise Dejean, Sur un théorème de Thue, J. Combin. Theory Ser. A **13** (1972), 90–99.

[16] Roger C. Entringer, Douglas E. Jackson & J.A. Schatz, On non-repetitive sequences, J. Combin. Theory Ser. A **16** (1974), 159–164; MR **48** #10860.

[17] Earl D. Fife, Binary sequences which contain no BBb, Trans. Amer. Math. Soc. **261** (1980), 115–136; MR **82a**:05034

[18] Pavel Goralcik & Tomas Vanicek, Binary Patterns in Binary Words, Internat. J. Algebra Comput. **1** (1991) 387–391.

[19] Andres. del Junco, A transformation with simple spectrum which is not rank one, Canad. J. Math. **29** (1977), 655–663; MR **57** #6367.

[20] Jacques Justin, Characterization of the repetitive commutative semigroups, J. Algebra **21** (1972), 87–90; MR **46**#277.

[21] Juhani Karhumäki, On cube-free ω-words generated by binary morphisms, Discrete Appl. Math. **5** (1983), 279–297; MR **84j**:03081.

[22] Veikko Keränen, Abelian squares are avoidable on 4 letters, Automata, Languages and Programming: Lecture notes in Computer Sciences **623** (1992) Springer-Verlag 4152.

[23] Filippo Mignosi, Infinite words with linear subword complexity, Theoret. Comput. Sci. **65** (1989), 221–242; MR **91b**:68093.

[24] Marston Morse & Gustav A. Hedlund, Symbolic dynamics I, II, Amer. J. Math. **60** (1938), 815–866; **62** (1940) 142; MR **1**, 123d.

[25] P. S. Novikov & S. I. Adjan, Infinite periodic groups I, II, III, Izv. Akad. Nauk. SSSR Ser. Mat. **32** (1968), 212–244;251–524;709–731;MR **39** #1532a–c.

[26] P. A. B. Pleasants, Non-repetitive sequences, Proc. Cambridge Philos. Soc. **68** (1970), 267–274; MR **42** #85.

[27] Peter Roth, Every binary pattern of length six is avoidable on the two-letter alphabet, Interner Bericht 6/89, Fachbereich Informatik, Universität Frankfurt.

[28] Robert. O. Shelton & Raj P. Soni, Aperiodic words on three symbols I, II, III, J. reine agnew. Math. **321;327;330** (1981), 195–209;1–11;44–52; MR **82m**:05004a–c.

[29] Axel Thue, Über unendliche Zeichenreihen, Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana (1912), 1–67.

[30] A. Zimin, Blocking sets of terms, Mat. Sb. (N.S.) **119 (161)** (1982); Math. USSR Sbornik **47** (1984), 353–364.