COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA 16 (1) (1996) 19–40

NON-REPETITIVE WORDS: AGES AND ESSENCES

JAMES D. CURRIE[1]

Received March 7, 1990
Revised January 11, 1995

This paper introduces the notions of age and essence of an infinite word w. Using these notions, the author studies the set L of infinite non-repetitive words over $\{1,2,3\}$, and its proper subsets L_{121}, $L_{121,323}$, $L_{121,212}$, where words of L_{121} ($L_{121,323}$; $L_{121,212}$) do not contain 121 (121,323; 121,212) as subwords. Motivated by the question 'How many essentially different non-repetitive words over $\{1,2,3\}$ exist?' the author counts the equivalence classes of L, L_{121}, $L_{121,323}$, $L_{121,212}$ under agreement in a final segment, agreement in age, and in essence.

1. Introduction

A word w is *non-repetitive* if no two adjacent blocks in w are identical. For example, the word 'orange' is non-repetitive. On the other hand, the word 'banana' is repetitive, since 'an' occurs next to itself in 'banana'. Infinite non-repetitive words over $S=\{1,2,3\}$ have been used in mathematics to construct various pathological objects. Notably, non-repetitive words figure in Adjan and Novikov's solution of the Burnside problem for groups [24]. Other applications of non-repetitive words have been to logic [6], to ordered sets [30] and to symbolic dynamics [23].

The existence of infinite non-repetitive words over finite alphabets has been known since the turn of the century [29]. More recently, questions concerning the existence of infinite words avoiding other patterns have arisen [3, 14]. A non-repetitive word w is said to *avoid* p if there is no non-erasing substitution h such that $w=uh(p)v$. If an ω-word over Σ avoids p, we say that p is *avoidable* on Σ. The problem of finding an algorithm to determine when a given p is avoidable on a given Σ remains open [3, 2] and is an active area of research [7, 26].

At least two distinct approaches to showing the existence of infinite words avoiding patterns over finite alphabets have been taken. The standard way to show the existence of an infinite word avoiding some pattern has been to construct such a word by the iteration of a suitably chosen substitution. For example, in [1], the

Mathematics Subject Classification (1991): 68 Q, 03 C

[1] This work was supported by an NSERC operating grant.

substitution used is $f_0 : S^* \to S^*$, given by

$$f_0(1) = 123,$$
$$f_0(2) = 13,$$
$$f_0(3) = 2.$$

For all $n \in \mathbb{N}$, $f_0^n(1)$ is a non-repetitive word; moreover, since the first symbol in $f_0(1)$ is a 1, $f_0^n(1)$ is always a prefix of $f_0^{n+1}(1)$. It thus makes sense to define an infinite word $f_0^\omega(1)$ having $f_0^n(1)$ as an initial segment for each n. The word $f_0^\omega(1)$ is an infinite non-repetitive word.

A second approach to showing the existence of infinite non-repetitive words over $\{1,2,3\}$ is implicit in the work of Shelton and Soni [28] and has been generalized in [11]. In this approach, one studies the branching of the tree of finite words over alphabet Σ avoiding pattern p. If this tree branches relatively frequently, then one can give a non-constructive proof of the existence of infinite words over Σ avoiding p.

It may be that by combining these two approaches a general algorithm for determining when p is avoidable over Σ can be found. In the case where it is 'easy' to avoid p, so that the tree of finite words over Σ avoiding p branches frequently, the approach of Shelton applies. On the other hand, in the case where the tree of finite words avoiding p branches infrequently, experience shows that substitutions are easy to find. For example, in writing [8] the author found that by imposing additional conditions on non-repetitive words one produces extremal cases where ω-words can *only* be produced by iterating specific substitutions. In this paper we investigate this situation, trying to pin down the nature of these extremal cases.

Let L be the set of non-repetitive words of type ω over S. What is the size of L? Clearly if $w = uv$ is a non-repetitive word of type ω over S, then so is v whenever u is a finite prefix of w. Discarding the first letter of w, then the second, etc., gives a countable infinity of non-repetitive words; these words are all different because w is non-repetitive. In some sense however, these words are all the same. They all have a final segment in common.

In [3] it is shown that there are uncountably many non-repetitive words of type ω over S, 'no two of which have a final segment in common.' Here we may omit the last clause, and simply say that there are uncountably many non-repetitive words of type ω over S, since only countably many words of type ω can have any given final segment. In [28] it is shown that if u is a prefix of an infinite non-repetitive word over S, then u is a prefix of uncountably many non-repetitive words over S.

We see then that fixing the prefix of a non-repetitive word of type ω does not significantly restrict the choice of word. Let us consider substitution f_0 above. A study of f_0 shows that for $v \in S^*$, $f_0(v)$ only contains 23 in the context $123 = f_0(1)$, and $f_0(v)$ only contains 32 in the context 321. Thus if $v \in S^*$, $f_0(v)$ does not contain 121 or 323. If $\{a_1, a_2, \ldots, a_k\}$ is a set of words, denote by $L_{a_1,a_2,\ldots,a_k}$ the sublanguage of L consisting of words not containing any of $a_1, a_2, \ldots, a_k$ as a subword. Thus $f_0^\omega(1) \in L_{121,323}$. We show later that if $v \in L_{121,323}$, then a final segment of v is of the form $f_0(u)$, where $u \in L_{121,323}$. By induction, for any $n \in \mathbb{N}$,

v has a final segment of the form $f_0^n(u)$ for some $u \in L_{121,323}$. We see that in some sense there is 'only one' word in $L_{121,323}$, i.e. $f_0^\omega(1)$. Thus restricting subwords of non-repetitive words can produce extremal cases.

In this paper we explore in what sense $L_{121,323}$ is 'rigid', so that in considering $L_{121,323}$ we are forced to find f_0. Although $L_{121,323}$ contains uncountably many words, in some sense there is 'only one word' in $L_{121,323}$; there is only one word in *essence*, where the essence of a word v is the set of those finite subwords of v appearing in every final segment of v. We also show that the more obvious equivalence relation between infinite words, that of containing the same finite subwords, is too weak to characterize $L_{121,323}$.

Remark 1.1. Our desire is to illustrate a general phenomenon connected with words avoiding patterns, noticed during the writing of [8]. However, it turns out that our example $L_{121,323}$ is related to a much-studied particular case in combinatorics on words, that of irreducible words on 2 symbols. A word w is *irreducible* (or *overlap-free*) if we cannot write $w = aBBbc$ where b is the first letter of B. Irreducible words have been well-studied, and are characterized in [16, 27].

Irreducible words have been used in the study of dynamical systems. In such a setting, it is most natural to study two-way infinite sequences. Indeed, Thue spent most of his important 1912 paper considering two-way infinite sequences. While Thue also looks at one-way infinite sequences (our words of type ω), one must be careful in attempting to translate his results to words of type ω. For example, the following exercise, based on Thue's work, appears in [21, page 38]:

'With the notations of Theorem 2.3.1: $\mathbf{b}$ is a square-free word such that neither aba nor $acbca$ is a factor of $\mathbf{b}$, if and only if $\delta(\mathbf{b})$ has no overlapping factor. (See Thue 1912.)'

Here $\delta(a) = a$, $\delta(b) = ab$, $\delta(c) = abb$. Unfortunately, if $\mathbf{b} = cbc$, then $\delta(\mathbf{b}) =$ $abbababb$, which contains the overlap $babab$. In fact, one can find counter-examples to this exercise even where $\mathbf{b}$ is a one-way infinite word. However, the result of the exercise does hold when we look at two-way infinite words.

A similar correspondence, essentially the inverse of δ, arises in the 'folklore' of non-repetitive words: We construct a non-repetitive word over $\{0,1,2\}$ in two stages. First we construct an irreducible word over $\{0,1\}$. Let $t = \{t_i\}_{i \geq 0}$ be the ω-word over $\{0,1\}$ where t_i is the sum modulo 2 of the digits of the binary representation of i. Thus, for example, $t_5 = 0 \equiv 1+0+1 \pmod 2$. Then

$$t = 011010011001\ldots$$

which is the famous Thue-Morse sequence, and is irreducible. Now let $s = \{s_i\}_{i \geq 1}$ be the sequence where s_i counts the number of 1's between the i^{th} and $(i+1)^{st}$ 0's of t. Thus

$$s = 210201.$$

With this 'counting map' one *almost* gets a bijection from irreducible words over $\{0,1\}$ to non-repetitive words over $\{0,1,2\}$ not containing 010 or 212. However, the word

$$\hat{t} = 0t_3t_4t_5\ldots$$

is irreducible, but corresponds to

$$01020\ldots,$$

which contains 010.

The smoothing effect of infinity in Mathematics is well-known; adding a point at infinity to the complex plane simplifies the theory considerably. Two-way infinite irreducible words have no boundary. There exist uncountably many such words over $\{0,1\}$, but they are equivalent with respect to their ages. On the other hand, as regards ages there are uncountably many one-way infinite irreducible words over $\{0,1\}$. The finite case is even messier. It is only very recently that an efficient way of computing the number of irreducible words of a given length has been given [7].

Notwithstanding the foregoing, it is clear that $L_{121,323}$ is 'close' to the set of irreducible ω-words over $\{0,1\}$. However, we are not simply studying another interesting property of irreducible words. To corroborate this, we give in an appendix our analysis of $L_{121,212}$, parallel to our analysis of $L_{121,323}$. (The two-way infinite analog of $L_{121,212}$ was studied in Thue's 1912 paper [29].)

The author thanks the anonymous referees for their helpful comments.

2. Notation

Our notation follows the usual notation of automata theory. However, to make room for infinite words, we stretch definitions as necessary. Let Σ be a finite set. A *word over* Σ is a (finite or countably infinite) sequence of elements of Σ. In the case that w is a countably infinite sequence of letters of Σ, we refer to w as an *ω-word.* We refer to Σ as an *alphabet*, to its elements as *letters*. The set of all finite words over Σ is denoted by Σ^*. We denote the set of ω-words over Σ by Σ^ω. We take a naive view of words as strings of letters (with the infinite ones running off the right-hand side of the page); thus the concatenation of two words w and v, written wv, is simply the string consisting of the letters of w followed by the letters of v. (This makes sense when w is a finite word, or if w is infinite and v is empty.)

Say that v is a *subword* of w if we can write $w=uvz$; u, v, $z\in\Sigma^*$.

If $w=uv$, then we say that u is a *prefix* of w; v is a *suffix* of w. The *empty word*, denoted by ε, is the word with no letters in it. When w is finite, denote by $|w|$ the *length* of w, equal to the number of letters 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 $w\in\Sigma^*$, $w=a_1a_2\ldots a_m$; $a_i\in\Sigma$ for $i=1$ to m. Then $h(w)=h(a_1)h(a_2)\ldots h(a_m)$.

Let $h:\Sigma^*\to\Sigma^*$ be a substitution on Σ. We call h *2-onto* if whenever $\sigma\in\Sigma$, and $u\in\Sigma^*$ is a non-repetitive word, $|u|=2$, then u is a subword of $h^m(\sigma)$ for some $m\in\mathbb{N}$. We call h *increasing* if for some $m\in\mathbb{N}$ we have $|h^m(u)|>|h(u)|$ for all $u\in\Sigma$.

A word w over alphabet Σ is *non-repetitive* if we cannot write $w=xyyz$; x, y, $z\in\Sigma^*$, $y\neq\varepsilon$. That is, w is non-repetitive if no subword of w appears twice in a row in w. The term *square-free* is also used for such words in the literature.

We say that two ω-words with a common suffix *agree in a final segment*. Agreement in a final segment is an equivalence relation on words of type ω, and we can speak of equivalence classes, representatives etc.

Let $w_1, w_2, w_3, \ldots$ be a sequence of words, with w_i a proper prefix of w_{i+1} for each $i\in\mathbb{N}$. We then define $v=\lim_{n\to\infty} w_n$ to be that unique ω-word having each w_i as a prefix. Note that if $h:\Sigma^*\to\Sigma^*$ is a substitution, and for some $a\in\Sigma$, a is a prefix of $h(a)$, then $h^i(a)$ is a prefix of $h^{i+1}(a)$ for each $i\in\mathbb{N}$. In such a case, we define $h^\omega(a)=\lim_{n\to\infty} h^n(a)$.

Let $w\in\Sigma^\omega$ be a word of type ω. The *age* of w is the set of all finite subwords of w. We denote the age of w by $\operatorname{Age} w$. (The notion of the age of an infinite structure comes from mathematical logic. See [9] for example.) The *essence* of w is the set of all finite subwords of w that appear in w as subwords infinitely often; thus the essence of w is the intersection of the ages of final segments of w. We denote the essence of w by $\operatorname{Ess} w$.

We fix $S=\{1,2,3\}$. Let L be the set of non-repetitive subwords over S. If $\{a_1,a_2,\ldots,a_k\}$ is a set of words, denote by $L_{a_1,a_2,\ldots,a_k}$ the sublanguage of L consisting of words not containing any of $a_1,a_2,\ldots,a_k$ as a subword. Thus $f_0^\omega(1)\in L_{121,323}$.

3. Results

We have several structural notions for non-repetitive words: final segments, age, essence. Each of these notions gives us a different equivalence relation on words; two words may agree in a final segment, have the same age or essence. Which of these notions captures the sense in which L is 'bigger than' $L_{121,323}$? In this paper we count the number of equivalence classes of L, L_{121}, $L_{121,323}$, $L_{121,212}$ under agreement in a final segment, in age, and in essence. The results are tabulated as follows:

equivalence classes with respect to	L	L_{121}	$L_{121,323}$	$L_{121,212}$
final segments	2^ω	2^ω	2^ω	2^ω
age	2^ω	2^ω	2^ω	2^ω
essence	2^ω	2^ω	1	1

To establish the information in this table, it suffices (since $L_{121,323}$, $L_{121,212}\subseteq L_{121}\subseteq L$) to prove the following propositions:

Proposition 3.1. *There are uncountably many words in $L_{121,323}$.*

Proposition 3.2. *There are uncountably many words in* $L_{121,323}$, *no two of which have the same age.*

Proposition 3.3. *Let* v, $w \in L_{121,323}$. *Then* $\mathrm{Ess}\, v = \mathrm{Ess}\, w$.

Proposition 3.4. *There are uncountably many words with different ages in* L_{121}.

Proposition 3.5. *Let* v, $w \in L_{121,212}$. $\mathrm{Ess}\, v = \mathrm{Ess}\, w$.

Proposition 3.6. *There are uncountably many words with different ages in* $L_{121,212}$.

Proposition 3.1 follows from results in [16, 21] via the correspondence between irreducible words and $L_{121,13231}$ mentioned in the introduction. Proofs of Propositions 3.5, 3.6 are in the Appendix.

4. Proofs of Propositions 3.1–3.4

In this section, $f : S^* \to S^*$ will be the substitution given by

$$f_0(1) = 123$$
$$f_0(2) = 13$$
$$f_0(3) = 2.$$

Lemma 4.1. *Let* $v \in S^*$ *be a non-repetitive word not containing* 121 *or* 323 *as a subword. Then* $f_0(v)$ *is non-repetitive, and does not contain* 121 *or* 323 *as a subword. Thus* $f_0^\omega(1)$ *is non-repetitive, and does not contain* 121 *or* 323.

Proof. See [8, 9, 29]. ∎

Lemma 4.2. *Let* $v \in S^*$ *be a non-empty non-repetitive word not containing* 121 *or* 323. *Suppose that* $w = 323v$ *is a non-repetitive word. Then* $f_0(w)$ *is non-repetitive, and does not contain* 121 *or* 323.

We saw in the introduction that $f_0(w)$ can never contain 121 or 323. It remains to show that $f_0(w)$ is non-repetitive. An inspection of f_0 shows that if $f_0(323) = 2132$ is a subword of $f_0(u)$, then 323 is a subword of u.

Now w is non-repetitive by hypothesis, therefore v commences with a 1. (Otherwise w commences 3233 or 3232, and w is repetitive.) Thus $23v$ does not contain 121 or 323 (or else v would, contrary to assumption), so that $f_0(23v)$ is non-repetitive by the previous lemma. Now suppose that $f_0(w)$ is repetitive. We may thus write $f_0(w) = xxy$ for some x, y in $S^*, x \neq \varepsilon$. Both $f_0(323) = 2132$ and x are prefixes of $f_0(w)$. We have two mutually exclusive possibilities:

Case 1. The word 2132 is a prefix of x, and we can write $x = 2132x'$. Then $2132 f_0(v) = f_0(w) = xxy = 2132x'2132x'y$, and 2132 is a subword of $f_0(v)$. Thus v contains 323, which is impossible.

Case 2. The word x is a proper prefix of 2132. This means that the first letter of x must occur twice in 2132, since $f_0(w)=2132f_0(v)=xxy$. Therefore we must have $x=213$. Then $f_0(v)$ commences with 13, and v commences with a 2, contradicting our earlier observation that v must commence with a 1.

Since both these cases are impossible, we conclude that $f_0(w)$ is non-repetitive, as required. ∎

Lemma 4.3. *If $w = u323v$ where $u,\ v \in S^*$, $u,\ v \neq \varepsilon$, then $f_0(w)$ is repetitive. (In such a case, we say that w contains a 323 internally.)*

Proof. If w is repetitive, then of course $f_0(w)$ is repetitive. Suppose that w is non-repetitive. Then 323 must be preceded and followed in w by 1's, so that w contains 13231. But then $f_0(w)$ contains $f_0(13231)=12\ 321\ 321\ 23$, which repeats 321. ∎

Proof of Proposition 3.1. Consider the substitutions $f_0 : S^* \to S^*$ and $g_0 : S^* \to S^*$ given by

$$f_0(1) = 123, \qquad g_0(1) = 2,$$

$$f_0(2) = 13, \qquad g_0(2) = 31,$$

$$f_0(3) = 2, \qquad g_0(3) = 321.$$

One sees that g_0 is obtained from f_0 by interchanging the roles of 1 and 3. Note that if $v \in S^*$, then $f_0^2(v)$ commences with a 1; that is, we can write $f_0^2(v)=1w$ for some w.

Define substitutions $\mu_0 : S^* \to S^*$ and $\mu_1 : S^* \to S^*$ by $\mu_0 = f_0^3$ and $\mu_1 = f_0^2 g_0$. The substitutions f_0 and g_0 are clearly 1-1. It follows that if $v \in S^*$, then $\mu_0(v) \neq \mu_1(v)$. Let $T=\{0,1\}$. Define $\phi : T^* \to S^*$ recursively by

$$\begin{aligned} \phi(\varepsilon) &= 1, \\ \phi(0v) &= \mu_0(\phi(v)), \\ \phi(1v) &= \mu_1(\phi(v)). \end{aligned}$$

Thus $\phi(01)=\mu_0(\phi(1))=\mu_0(\mu_1(\phi(\varepsilon)))=\mu_0(\mu_1(1))$, for example. By Lemma 4.1, (which will also hold true with g_0 in place of f_0) it follows that if $v \in T^*$, then $\phi(v)$ is a non-repetitive word in S^*, not containing 121 or 323. We see that if $v \in T^*$, then $\phi(v)$ commences with a 1. Write $\phi(v)=1x$. Let $u=u_1u_2\dots u_m, u_i \in T$ for each i. Suppose $w=uv$. Then $\phi(w)=\mu_{u_1}(\mu_{u_2}(\dots\mu_{u_m}(\phi(v))\dots)=\mu_{u_1}(\mu_{u_2}(\dots\mu_{u_m}(1x))\dots)$ which has $\mu_{u_1}(\mu_{u_2}(\dots\mu_{u_m}(1)\dots))=\phi(u)$ as a prefix. We see then that if $u,v \in T^*$, and u is a prefix of v, then $\phi(u)$ is a prefix of $\phi(v)$.

Thus if $v=v_1v_2\dots v_n\dots$ is an ω-word, with each $v_i \in T$, then it makes sense to define $\phi(v)=lim_{n\to\infty}\phi(v_1v_2\dots v_n)$. Also, since $\mu_0(w)\neq\mu_1(w)$ if w is a non-empty word in S^*, and μ_0, μ_1 are 1-1, it follows by induction that if $u,v \in T^*$, and $u \neq v$, then $\phi(u)\neq\phi(v)$.

Let W be the set of ω-words over T. It follows that $V=\{\phi(w): w \in W\}$ is an uncountable set of distinct non-repetitive words over S, not containing 121 or 323. This establishes Proposition 3.1. ∎

Proof of Proposition 3.2. If w is a non-repetitive word over S^* beginning with 123, write $w = 123w'$. For such a w, let $g_1(w) = 323w'$.

Define maps $h_0 : S^* \to S^*$ and $h_1 : S^* \to S^*$ by $h_0 = f_0^5$ and $h_1 = f_0^2 g_1 f_0^3$. If $v \in S^*$, then $h_0(v) \neq h_1(v)$, since $h_0(v)$ commences with $f_0^2(1)$, while $h_1(v)$ starts with $f_0^2(3)$, and f is 1-1.

Again let $T = \{0,1\}$ and define $\phi : T^* \to S^*$ recursively by

$$\phi(\varepsilon) = 1,$$
$$\phi(0v) = h_0(\phi(v)),$$
$$\phi(1v) = h_1(\phi(v)).$$

By Lemmas 4.1, 4.2, we see that $\phi(v) \in L_{121,323}$ for all $v \in T^*$. Again, if u, $v \in T^*$, and u is a prefix of v, then $\phi(u)$ is a prefix of $\phi(v)$, and ϕ can be extended to ω-words. Also, since $h_0(w) \neq h_1(w)$ if w is a non-empty word in S^*, it follows that if u, $v \in T^*$, and $u \neq v$, then $\phi(u) \neq \phi(v)$.

One other fact about ϕ is to be noted; earlier we remarked that if $f_0(v)$ contains $f_0(323)$, then v contains 323. Suppose that $\psi = h_{u_1} h_{u_2} \ldots h_{u_k} f_0^2$ where $u_i \in T$ for each i. If $\psi(v)$ contains $\psi(323)$, then v contains 323. This can be established by induction on k.

Let W be the set of ω-words over T. Again we will have that $V = \{\phi(w) : w \in W\}$ is an uncountable set of distinct words in $L_{121,323}$. Moreover, each word in V has a different age.

Let u, v be ω-words over T, $u \neq v$. We shall show that one of $\phi(u)$ and $\phi(v)$ contains a subword not found in the other. Since $u \neq v$, suppose they first differ in the $k+1^{st}$ symbol. Say without loss of generality that $u = u_1 u_2 \ldots u_k 0 u'$, $v = u_1 u_2 \ldots u_k 1 v'$, where the $u_i, v_i \in T, u', v' \in T^*$. Let $\psi = h_{u_1} h_{u_2} \ldots h_{u_k}$. Then $\phi(u) = \psi(\phi(0u'))$ and $\phi(v) = \psi(\phi(1v'))$. We claim that $\psi(f_0^2(323))$ is a subword of $\phi(v)$, but not of $\phi(u)$.

Now $\phi(u) = \psi(f_0^2(f_0^3(\phi(u'))))$. Since $\phi(u)$ is non-repetitive, as noted earlier, this means that $f_0^2(f_0^3(\phi(u')))$ is non-repetitive. From Lemma 2, it follows that $f_0^3(\phi(u'))$ does not contain a 323 internally. Of course, $f_0^3(\phi(u'))$ begins with a 1, and thus does not contain 323 as a subword at all. It follows that $\psi(f_0^2(323))$ is not a subword of $\phi(u)$. On the other hand, $\phi(v) = \psi(f_0^2 g_1(f_0^3(\phi(v'))))$ has $\psi(f_0^2(323))$ as a prefix. We have thus demonstrated Proposition 3.2. ∎

Proof of Proposition 3.3. Suppose $v \in L_{121,323}$. A final segment v' of v will commence with a 1, and we can write $v = 1b_1 1b_2 1b_3 \ldots 1b_n 1 \ldots$ so that none of the b_n contains a 1. What do the pieces $1b_n$ look like? The non-repetitive words starting with 1 and containing only a single 1 are 1, 12, 13, 123, 132, 1232, 1323. However, v cannot contain 11, 121 or 1323, so that $1b_n$ must be either $A = 13$, $B = 123$, $C = 132$ or $D = 1232$. Since BD and $BA1$ are repetitive, B only appears in v in the context BC. Also AC and $DC1$ are repetitive, so C can occur in v' only at the beginning, or in the context BC. Thus a final segment of v is concatenated from the pieces A, BC and D. However, $A=f_0^2(3)$, $BC=f_0^2(1)$ and $D=f_0^2(2)$. We conclude that a final segment of v is of the form $f_0^2(w)=f_0(f_0(w))$ for some $w\in S^*$. As noted earlier in the introduction, $f_0(w)$ cannot contain 121 or 323. Also $f_0(w)$ is non-repetitive, since v is. In summary, a final segment of v is of the form $f_0(v')$ where $v'\in L_{121,323}$. Induction gives the following result:

Lemma 4.4. *Let $n\in\mathbb{N}$ be given. If $v\in L_{121,323}$, then a final segment of v is of the form $f_0^n(w)$, where $w\in L_{121,323}$.*

Lemma 4.5. *Let v be an ω-word over Σ. Let h be an increasing, 2-onto substitution on Σ. Let $g:\Sigma^*\to T$ be an increasing substitution. Suppose that for each $n\in\mathbb{N}$ there is an ω-word w such that $g(h^n(w))$ is a final segment of v. Then $\operatorname{Ess} v=\operatorname{Ess} g(h^\omega(1))$.*

Proof. Suppose that the conditions of the lemma hold and that $u\in\operatorname{Ess} g(h^\omega(1))$. Pick $m\in\mathbb{N}$. We show that u occurs as a subword of v at least m times. Since $u\in\operatorname{Ess} g(h^\omega(1))$, pick $n\in\mathbb{N}$ such that u appears as a subword of $g(h^n(1))$ at least m times. Since h is 2-onto, pick $k\in\mathbb{N}$ so large that for any $\sigma\in\Sigma$, 1 is a subword of $h^k(\sigma)$. Thus u appears at least m times in $h^{n+k}(w)$ for any non-empty word w over Σ. By the conditions of the lemma, for some w, word $g(h^{n+k}(w)$ will be a suffix of v. Thus u occurs as a subword of v at least m times. Since m was arbitrary, $\operatorname{Ess} g(h^\omega(1))\subseteq\operatorname{Ess} v$.

Suppose that $u\in\operatorname{Ess} v$. Pick $n\in\mathbb{N}$ such that $|g(h^n(\sigma))|>|u|$ whenever $\sigma\in\Sigma$. This is possible since g,h are increasing. Since $u\in\operatorname{Ess} v$, we can find w such that $g(h^n(w))$ is a final segment of v and u is a subword of $g(h^n(w))$. In fact, by our choice of n, u will be a subword of $g(h^n(y))$ for some subword y of w with $|y|=2$. Since h is 2-onto, u will be a subword of $g(h^r(1))$, some $r\in\mathbb{N}$. It follows that $u\in\operatorname{Ess} g(h^\omega(1))$. ∎

Corollary 4.6. *If $v\in L_{121,323}$ then $\operatorname{Ess} f_0^\omega(1)=\operatorname{Ess} v$.*

Proof. We apply the previous lemma with $g=h=f_0$. ∎

This establishes Proposition 3.3. ∎

Proof of Proposition 3.4. Consider the word

$$\begin{aligned} f_0^4(1) &= f_0^3(123)\\ &= f_0^2(123132)\\ &= f_0(123132123213) = 1231\underline{321232131}23132131232. \end{aligned}$$

This word contains $\alpha=321232131$ as a subword as indicated. Thus $f_0^\omega(1)$ contains α as a subword infinitely often. Suppose w is a word arising from $f_0^\omega(1)$ by replacing some occurrences of α by the word $\beta=32\,12321\,323\,12321\,31$.

Remark 4.7. No prefix of α is a proper suffix of α. Thus if $\alpha x = y\alpha$ for non-empty words x, y then $|x| = |y| \geq |\alpha|$. A similar remark holds true for β.

Remark 4.8. Neither $f_0^\omega(1)$ nor β contains any of 22, 321321, 123123, 3131, 1313 or 121 as a subword. One sees that w cannot contain any of these subwords either.

Thus if $w = x32\,12321\,32y$, some x, y then the first letter of y must be a 3. We can thus write $w = x32\,12321\,323y'$. Since 323 is not a subword of $f_0^\omega(1)$, this means that $w = x\beta y''$, for some y''. Put briefly, if β' is a prefix of β with $|\beta'| \geq 9$ then if $w = x\beta' y$ for some x, y, then $w = x\beta y'$, some y'. A long prefix of β can only appear in w as part of β.

Remark 4.9. If β' is a prefix of β with $|\beta'| \geq 9$, then if $w = x\beta' y$ for some x, y, then $w = x\beta y'$, some y'.

Suppose $w = x_0\,31\,2321\,31y$, some x_0, y. Since β and $f_0^\omega(1)$ do not contain 33, neither does w. Thus the last letter of x_0 is either a 2 or a 1.

Case I. Write $x_0 = x_1 2$. Then $w = x_1 23\,12321\,31y$. By Remark 4.8, the last letter of x_1 must be a 3. Then $w = x'\beta y$, some x' with $x'\beta = x_o 3\,12321\,31$.

Case II. Write $x_0 = x_1 1$, and $w = x_1 131\,232\,131y$. By Remark 4.8, we can write $x_1 = x_2 2$, and $y = 2y_1$. Then $w = x_2\,213\,123\,213\,12y_1$. Again by Remark 4.8, we can write $y_1 = 3y_2$, so that $w = x_2\,213123\,213123\,y_2$, so that w contains a repetition of 213123.

Changing all β's in w back to α's removes this repetition, so an occurrence of β in w must intersect somehow an occurrence of $z = 213123213123$. An inspection of β shows this is impossible; β and z do not contain each other, and no prefix of β is a suffix of z or vice versa. Thus the present case is impossible.

Remark 4.10. If β' is a suffix of β with $|\beta'| \geq 8$, then if $w = x\beta' y$ for some x, y, then $w = x'\beta y$, where $x\beta' = x'\beta$.

Lemma 4.11. *Word w is non-repetitive.*

Proof. Suppose that w is repetitive. This means that a finite prefix of w is repetitive, so that replacing finitely many α's in $f_0^\omega(1)$ by β's can cause a repetition. Let w be chosen such that w contains the least possible occurrences of β. Suppose that $w = xuuy$, $u \neq \varepsilon$. Since replacing β's by α's in w removes the repetition uu, uu must intersect some occurrence of β. There are four possibilities:

Case 1: Word uu is a subword of β.
Case 2: Word β overlaps word uu on the left.
Case 3: Word β overlaps uu on the right.
Case 4: Word β is a subword of uu.

We attack these cases separately.

Case 1. Word uu is a subword of β. This is impossible, since β is non-repetitive.

Case 2. Word β overlaps uu on the left. Write $\beta = \beta'\beta''$ such that $\beta'' \neq \varepsilon$, $uu = \beta'' v$ and $w = x\beta'' vy$. If $|\beta''| \leq |u|$, then we can write $u = u'\beta''$ for some u'. In any case,

$|\beta''| < |uu|$, since uu is not a subword of β. If $|u| < |\beta''| < |uu|$, then let $\beta'' = uc$. Then c is a proper suffix of β which is a prefix of u.

In all cases, a proper suffix of β is a prefix of u. Renaming if necessary, write $\beta = \beta'\beta''$, $u = \beta''u''$, $\beta'' \neq \varepsilon$.

Case 2(a). $|\beta''| \leq 7$.

In this case, β'' is a suffix of α. Therefore, $uu = \beta''u''u$ is a suffix of $\alpha u''u$, and $w^* = x\alpha u''uy$ contains the repetition uu. This is impossible, as w^* contains one fewer β than w.

Case 2(b). $|\beta''| \geq 8$.

Note that $w = x\beta''u''\beta''u''y$. By Remark 4.10, we can write $w = x'\beta u''\beta''u''y$, and also $w = x''\beta u''y$, where $x'\beta u'' = xu$ and $x''\beta u'' = xuu$

If $|x''| < |x'\beta|$, then a proper prefix of β will be a suffix of β. This is impossible by Remark 4.7. We may thus assume that $|x''| \geq |x'\beta|$, so that we can write $w = x'\beta u^*\beta u''y$, where $u'' = u^*u^{**}$. However, this means that $w^* = x'\alpha u^*\alpha u''y = x'\alpha u^*\alpha u^*u^{**}y$ is repetitive, but contains two fewer β's than w, which is impossible.

Case 3. We can write $uu = v\beta'$ where $\beta = \beta'\beta''$ and $w = xv\beta'y$. Reasoning as in the previous case we can assume without loss of generality that $u = u'\beta'$ and $w = xuu'\beta y$.

Case 3(a). $|\beta'| \leq 8$. In this case $w^* = xuu'\alpha y$ contains the subword uu, and one less β than w. The reasoning is as in Case 2(a) above.

Case 3(b). $|\beta'| \geq 9$. Here we use Remark 4.9 to conclude that $w = xu'\beta u''u'\beta u''y'$, some y' where $u = u'\beta u''$. Then $w^* = xu^*u^{**}\alpha u^{**}\alpha y$, is repetitive, an impossibility. The reasoning is as in Case 2(b) above.

Case 4. $uu = c\beta d$, some c, d.

Case 4(a). $|c|$, $|d| \leq |u|$. Then $u = c\beta' = \beta''d$ and $\beta = \beta'\beta''$. But $|\beta| = 17$, so either $|\beta'| \geq 9$, which is dealt with as in Case 2(b), or $|\beta''| \geq 9$, which is dealt with as in Case 3(b).

Case 4(b). $|c| > |u|$ or $|d| > u$. Then $u = v_1\beta v_2$, for some v_1, v_2, and $w = xv_1\beta v_2 v_1\beta v_2 y$. But then $w^* = xv_1\alpha v_2 v_1\alpha v_2 y$ is repetitive, with fewer β's than w.

This completes our case by case analysis. In all cases, the assumption that w is repetitive leads to a contradiction. Thus w is non-repetitive, as desired. ∎

Combining this lemma with Remark 4.8, we see that whenever w is obtained from $f_0^\omega(1)$ by replacing α's with β's we have $w \in L_{121}$. Here is a 1-1 correspondence between ω-words w obtained in this way and the set of all ω-words over $T = \{0,1\}$: Given w, let $u = \{u_i\}_{i=1}^{\infty}$ be the word where $u_i = 1$ exactly when the i^{th} occurrence of α in $f_0^\omega(1)$ is replaced by β in w. Thus u records the pattern by which α's alternate with β's in w. Thanks to this correspondence, or rather, its inverse, we can show there are uncountably many ω-words in L_{121} with different ages by showing that there are uncountably many words in T^ω with different ages.

Let $v \in T^\omega, v = \{v_i\}_{i=1}^\infty$. Consider the map $\nu : T^\omega \to T^\omega$ given by $\nu(v) = \{u_j\}_{i=1}^\infty$ where $u_j = 1$ exactly when for some $k, v_k = 1$ and $j \equiv 2^{k-1} \pmod{2^k}$. For example, if $v_1 = 1$, then $u_j = 1, j$ odd. One verifies that if $v \neq v'$, then $\nu(v), \nu(v')$ have different ages. There are thus uncountably many words in T^ω with different ages. It follows that there are uncountably many words in L_{121} with different ages.

This establishes Proposition 3.4. ∎

5. Appendix: Proofs of Propositions 3.5–3.6

Let $R = \{a, b, c, d, e\}$. Let $f_1 : R^* \to S^*, f_2 : R^* \to R^*$ be given by

$$
\begin{aligned}
f_1(a) &= 123, & f_2(a) &= adcbebc,\\
f_1(b) &= 1232, & f_2(b) &= adcbedc,\\
f_1(c) &= 13, & f_2(c) &= aebc,\\
f_1(d) &= 132, & f_2(d) &= aebedc,\\
f_1(e) &= 1323, & f_2(e) &= aebedcbebc.
\end{aligned}
$$

Lemma 5.1. *Let $w \in L_{121,212}$. Then for every $n \in \mathbb{N}$ a final segment of w is of the form $f_1(f_2^n(v_n))$, some ω-word $v_n \in R^*$.*

Proof. Replacing w by a final segment if necessary, write

$$w = 1w_1 1w_2 1w_3 \ldots$$

where each $w_i \in \{2,3\}^*$. Since the maximal non-repetitive words over $\{2,3\}^*$ are 232 and 323, we see that for each i, we must have $1w_i \in \{12, 123, 1232, 13, 132, 1323\}$. In fact, we can exclude the possibility that $1w_i = 12$, since in such a case $1w_i1 = 121$ would be a subword of w. Thus $w = f_1(u)$, some $u \in R^*$.

Various observations concerning possible subwords of u may be made. For example, $ab \notin \operatorname{Age} u$, or else $f_1(ab) = 1231232$ is a subword of w, which is impossible, since w is non-repetitive. Similarly, ea is not a subword of u. Also, ba, $da \notin \operatorname{Age} u$, or 212 appears in w. By arguments of this sort one shows that u can be walked on digraph D of Figure 1.

We note that five additional restrictions on subwords of u hold:

1. No word of the form $dzbza, z \in R^*$ can be a subword of u. Otherwise $f_1(dzbza) = 1\ 32f_1(z)12\ 32f_1(z)12\ 3$ is a subword of w. This is impossible, since $32f_1(z)12$ is repeated.
2. No word of the form $azezd, z \in R^*$ can be a subword of u. Otherwise $f_1(azezd) = 12\ 3f_1(z)132\ 3f_1(z)132$ is a subword of w. This is impossible, since $3f_1(z)132$ is repeated.

3. No word of the form $ezcz, z \in R^*$ can be a subword of u. Otherwise $f_1(ezcz)1 = 132\ 3f_1(z)1\ 3f_1(z)1$ is a subword of w. This is impossible, since $3f_1(z)1$ is repeated.
4. No word of the form $zazb, z \in R^*$ can be a subword of u. Otherwise $f_1(zazb) = f_1(z)12\ f_1(z)12\ 3$ is a subword of w. This is impossible, since $f_1(z)12$ is repeated.
5. No word of the form $zdze, z \in R^*$ can be a subword of u. Otherwise $f_1(zdze) = f_1(z)132\ f_1(z)132\ 3$ is a subword of w. This is impossible, since $f_1(z)132$ is repeated.

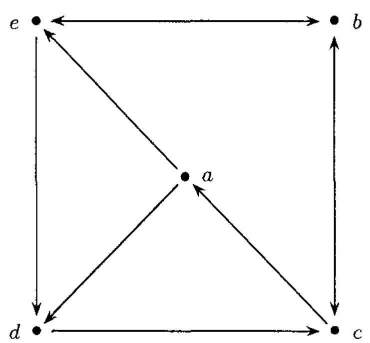

Fig. 1. Transition digraph D

Replacing u by a final segment if necessary, suppose that u begins with an a. Then $u = au_1au_2au_3\ldots$ where $u_i \in \{b,c,d,e\}^*$. Keeping in mind that au_i must be non-repetitive for each i, and also our five other restrictions on subwords of u, we restrict the possible values of au_i to those enumerated in Figure 2.

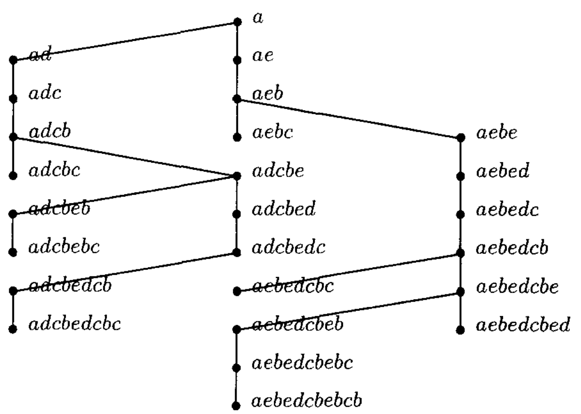

Fig. 2. Candidate values for the au_i

Reference to Figure 1 reminds us that au_i must always have a c as its last letter. The possible values of au_i are thus among the following:

adc^{**}, $adcbc^*$, $adcbebc$, $adcbedc$, $adcbedcbc^*$, $aebc$, $aebedc$, $aebedcbc^*$, $aebedcbebc$.

Here the singly-starred words are ruled out, since if az_i has one of these values, then u contains subword az_ia, which has $dcbca$ as a suffix. This is impossible since $dcbca$ is of the form $dzbza$. The doubly-starred word adc cannot appear twice in u, since otherwise if $az_i = adc$, then $caz_iad = cad\,cad$ or $caz_iae = cad\,cae$ will be a subword of u. Clearly $cadcad$ is repetitive, while $cadcae$ is outlawed by restriction 5 on subwords of u.

We may thus write $u = f_2(v_1)$, some $v_1 \in R^*$. Clearly v_1 must be non-repetitive. We show that a final segment of v_1 can be walked on digraph D of Figure 1, and that restrictions 1–5 on subwords hold for a final segment of v_1. Our result will follow by the obvious induction.

Let us verify that a final segment of v_1 can be walked on digraph D of Figure 1. First, $cd \notin \mathrm{Age}\,v_1$, since otherwise $cf_2(c)f_2(d) = caeb\,caeb \in \mathrm{Age}\,u$, which is impossible. Similarly, ce, $de \notin \mathrm{Age}\,v_1$. Also, if $ea \in \mathrm{Age}\,v_1$, then $f_2(ea)a = aebe\ dcbebca\ dcbebca \in \mathrm{Age}\,u$, which is impossible. Similarly, $ac, ec \notin \mathrm{Age}\,v_1$. If $da \in \mathrm{Age}\,v_1$, then $f_2(da) = aebeb\,dc\,a\,dc\,b\,ebc \in \mathrm{Age}\,u$, which contradicts restriction 1 on subwords of u. Similarly, ba, $db \notin \mathrm{Age}\,v_1$. It now follows that $bd \notin \mathrm{Age}\,v_1$. Otherwise, since da, $db \notin \mathrm{Age}\,v_1$, we have $f_2(bd)ae = adc\ bedcae\ bedcae \in \mathrm{Age}\,u$, which is impossible. Finally, since ba, da, $ca \notin \mathrm{Age}\,v_1$, if ab is a subword of v_1 more than once, then so is cab. Then $f_2(cab) = ae\ bcadcbe\ bcadcbe\ dc \in \mathrm{Age}\,u$. This is impossible. Since none of ab, ac, ba, bd, cd, ce, da, db, de, ea, ec can appear in v_1 more than once, a final segment of v_1 can be walked on digraph D of Figure 1.

To finish our proof we verify that our other 5 restrictions on subwords hold for v_1:

1. No word of the form $dzbza, z \in R^*$ can be a subword of v_1. Otherwise $f_2(dzbza) = aebe\,dcf_2(z)adcbe\,dcf_2(z)adcbe\,bc$ is a subword of u.
2. No word of the form $ezcz, z \in R^*$ can be a subword of v_1. Otherwise $f_2(ezcz)a = aebedcb\,ebcf_2(z)a\,ebcf_2(z)a$ is a subword of u.
3. No word of the form $zazb$, $z \in R^*$ can be a subword of v_1. Otherwise $bcf_2(zazb) = bcf_2(z)adcbe\,bcf_2(z)adcbe\,dc$ is a subword of u.
4. No word of the form $zdze$, $z \in R^*$ can be a subword of v_1. Otherwise $f_2(zdze) = f_2(z)aebedc\,f_2(z)aebedcbebc$ is a subword of u.
5. No word of the form $zdze$, $z \in R^*$ can be a subword of u. Otherwise $f_2(zdze) = f_2(z)aebedc\ f_2(z)aebedc\ bebc$ is a subword of w. This is impossible, since $f_2(z)aebedc$ is repeated. ∎

Proof of Proposition 3.5. Combine the previous lemma with the proof of Lemma 4.5, with $g = f_1$, $h = f_2$, and appropriate changes of alphabet. ∎

Let $f : A^* \to B^*$ be a substitution. Suppose that in any solution of

$$x''f(b_1)f(b_2)\dots f(b_m)y' = y''f(c_1)f(c_2)\dots f(c_r)z',$$

$$f(x) = x'x'',\ f(y) = y'y'',\ f(z) = z'z'',\ x'' \neq \varepsilon,\ y'' \neq \varepsilon,\ z'' \neq \varepsilon$$
$$b_1,\ b_2,\ \ldots,\ b_m,\ c_1,\ c_2,\ \ldots,\ c_r,\ x,\ y,\ z \in A$$

we have $x'' = y'', y' = z', m = r$ and $b_i = c_i$ for $i = 1$ to m. In this case we say that f satisfies the *line-up condition*. Both f_1 and f_2 satisfy the line-up condition.

Lemma 5.2. *Let $f : A^* \to B^*$ be a substitution which satisfies the line-up condition. Let $u \in A^*$. Suppose that $f(u)$ is repetitive. Then either*

1. *$f(v)$ is repetitive for some subword v of u with $|v| \leq 2$ or*
2. *a subword of u has the form $xwywz$, some $x, y, z \in A$, some $w \in A^*$ where $f(xyz)$ is repetitive.*

Proof. Write $f(u) = abbc$, $b \neq \varepsilon$. If bb is a subword of $f(v)$, some subword v of $u, |v| \leq 2$, then we are done. Otherwise we can write

$$b = x''f(b_1)f(b_2)\ldots f(b_m)y' = y''f(c_1)f(c_2)\ldots f(c_r)z'$$

$$f(x) = x'x'',\ f(y) = y'y'',\ f(z) = z'z'';\ x'' \neq \varepsilon,\ y'' \neq \varepsilon,\ z'' \neq \varepsilon$$

$$\text{some } b_1,\ b_2,\ \ldots,\ b_m,\ c_1,\ c_2,\ \ldots,\ c_r,\ x,\ y,\ z \in A$$

where $xb_1 \ldots b_m yc_1 \ldots c_n z$ is a subword of u. Let $w = b_1b_2 \ldots b_m$. By the line-up condition, $xb_1 \ldots b_m yc_1 \ldots c_n z = xwywz$. Also, $x'' = y'', y' = z'$, so that $f(xyz) = x'x''y'y''z'z'' = x'\ x''y'\ x''y'\ z''$ and is repetitive. ∎

Corollary 5.3. *Let $f : A^* \to B^*$ be a substitution which satisfies the line-up condition. Let $u \in A^*$. Suppose that*

1. *$f(v)$ is non-repetitive whenever v is a two letter subword of u and*
2. *if xyz is a three letter word over A such that $f(xyz)$ is repetitive, then no subword of the form $xwywz$, some $w \in A^*$ appears in u.*

   Then $f(u)$ is non-repetitive.

Remark 5.4. If xyz is a repetitive three letter word then $x = y$ or $y = x$. If $x = y$ then $xwywz = xw\,xw\,z$ is repetitive. Similarly, if $y = z$, then $xwywz$ is repetitive. We thus have the following corollary.

Corollary 5.5. *Let $f : A^* \to B^*$ be a substitution which satisfies the line-up condition. Let $u \in A^*$ be non-repetitive. Suppose that*

1. *$f(v)$ is non-repetitive whenever v is a two letter subword of u and*
2. *if xyz is a non-repetitive three letter word over A such that $f(xyz)$ is repetitive, then no subword of the form $xwywz$, some $w \in A^*$ appears in u.*

   Then $f(u)$ is non-repetitive.

Proof. If xyz is a repetitive three letter word then by the above remark $xwywz$ cannot be a subword of u, which is non-repetitive. ∎

f_1	f_2
cab, *ea-*	*cab*, *ea-*
dba	*dba*
ac-, *ec-*, *-cd*, *-ce*	*ac-*, *ec-*, *-cd*, *-ce*
bd-, *-de*	*bda*, *bde*
aec, *aed*, *ced*	*aed*

Table. 3. Non-repetitive triples xyz such that $f_1(xyz), f_2(xyz)$ is repetitive

We plan to apply Corollary 5.5 with $f = f_1$ (with $f = f_2$). As a preliminary, we list the non-repetitive triples xyz over $\{a,b,c,d,e\}$ where $f = f_1(xyz)$ (where $f_2(xyz)$) is repetitive.

Here the inclusion of $ae-$ for f_1 signifies that $f_1(aez)$ is repetitive for any $z \in R$. One sees that the triples listed for f_2 form a subset of those listed for f_1.

Remark 5.6. For $f = f_1$ or f_2, the condition that u can be walked on digraph D of Figure 1 is stronger than condition 1 of Corollary 5.5.

Lemma 5.7. *If n is a non-negative integer then $f_1(f_2^n(a))$ is non-repetitive, and does not contain* 121 *or* 212 *as a subword.*

Proof. We use induction. The result is true for $n=0$ since $f_1(f_2^0(a)) = f(a) = 123$. One verifies that $f_1(f_2(v))$ will never contain 121 or 212 as a subword for any $v \in R^*$. It is also clear that $f_2^n(a)$ can be walked on digraph D of Figure 1 for each non-negative n. Our result will thus follow by application of Corollary 5.5 if we can verify that $f_2^n(a)$ never contains a word of form $xwywz$ for any of the triples xyz listed for f_1 in Table 3. Suppose then that n is least such that $f_2^n(a)$ contains a word of form $xwywz$ for some triple xyz listed for f_1 in Table 3. It follows from Corollary 5.5 that $f_2^{n-1}(a)$ is non-repetitive. For $u \in R$ define $N(u) = \{v : uv \text{ is a directed edge of } D\}, N^-(u) = \{v : vu \text{ is a directed edge of } D\}$. We now rule out the problem triples one by one:

Triples *cab*, *ac-*

Suppose that $cuaub$ is a subword of $f_2^n(a)$. Since $b \notin N(a), u \neq \varepsilon$. The first letter of u must be in $N(c) \cap N(a) = \phi$, which is impossible. Thus $f_2^n(a)$ has no subword of form $cuaub$. Similarly, suppose that $aucu-$ is a subword of $f_2^n(a)$. Since $c \notin N(a), u \neq \varepsilon$. The first letter of u must again be in $N(c) \cap N(a)$.

Triple *-cd*

Suppose that $ucud$ is a subword of $f_2^n(a)$. Since $d \notin N(c), u \neq \varepsilon$. The last letter of u must be in $N^-(c) \cap N^-(d) = \phi$, which is impossible.

Triple *dba*

Suppose that $dubua$ is a subword of $f_2^n(a)$. Since $b \notin N(d), u \neq \varepsilon$. The first letter of u must be in $N(d) \cap N(b) = \{c\}$. The last letter of u must be in

$N^-(b) \cap N^-(a) = \{c\}$. This means $dubua = dcu'cbcu'ca$. However, then $f_2^n(a)$ contains a subword cbc. Perusal of the definition of f_2 shows this to be impossible.

Triple *ec-*

Suppose that $eucu$ is a subword of $f_2^n(a)$. Since $c \notin N(r), u \neq \varepsilon$. The first letter of u must be in $N(e) \cap N(c) = \{b\}$. Since cbc cannot be a subword of $f_2^n(a)$, the second letter of u is in $N(b) - \{c\} = \{e\}$. Thus $eucu = ebeu'cbeu'$. Examination of the definition of f_2 shows that ebe only appears in $f_2^n(a)$ in the context $aebedc$. Therefore $eucu = ebedcu''cbedcu''$. However, $cbedc$ only appears in $f_2^n(a)$ in the context $adcbedca$. Thus $eucu = ebedcau'''adcbedcau'''ad$. Finally, $ebedca$ only appears in $f_2^n(a)$ in the context $f_2(d)a$. Thus our assumption that $eucu$ is in $f_2^n(a)$ implies that $aeucu = aebedcau'''adcbedcau'''ad = f_2(d)f(u^*)f_2(b)f_2(u^*)ad$ appears in $f_2^n(a)$.

Since ad only appears in $f_2^n(a)$ as a prefix of $f_2(a)$ or $f_2(b)$, we see that $f_2^{n-1}(a)$ contains a subword du^*bu^*a or du^*bu^*b. The first possibility is ruled out by the minimality of n, the second because $f_2^{n-1}(a)$ is non-repetitive.

The foregoing argument can be recorded in an abbreviated form if we use the notation

$$x \dots y \dots z \Rightarrow xx^* \dots z^*yx^* \dots z^*z$$

to abbreviate 'The existence of a subword of the form $xuyuz$ in $f_2^n(a)$ implies the existence of a subword of the form $xx^*u^*z^*yx^*u^*z^*z$ in $f_2^n(a)$'. The argument is then

$$\begin{aligned} e \dots c \dots &\Rightarrow ebe \dots cbe \dots \\ &\Rightarrow ebedc \dots cbedc \dots \\ &\Rightarrow ebedca \dots adcbedca \dots ad \\ &\Rightarrow aebedca \dots adcbedca \dots ad \\ &\Rightarrow f_2(d) \dots f_2(b) \dots f_2(a) \text{ or } f_2(d) \dots f_2(b) \dots f_2(b). \end{aligned}$$

For the sake of brevity the rest of our case arguments will be outlined in this way.

Triple *aed*, *aec*

Here

$$\begin{aligned} a \dots e \dots &\Rightarrow adc \dots edc \dots \\ &\Rightarrow adcbe \dots edcbe \dots \\ &\Rightarrow adcbebc \dots aebedcbebc \dots aeb \\ &\Rightarrow f_2(a) \dots f_2(e) \dots f_2(c) \text{ or } f_2(a) \dots f_2(e) \dots f_2(d) \\ &\quad \text{or } f_2(a) \dots f_2(e) \dots f_2(e). \end{aligned}$$

Triple *ced*

Here

$$\begin{aligned} c\ldots e\ldots d &\Rightarrow c\ldots ae\ldots ad \\ &\Rightarrow cb\ldots aeb\ldots ad \\ &\Rightarrow cbe\ldots aebe\ldots ad \\ &\Rightarrow cbed\ldots aebed\ldots ad \\ &\Rightarrow cbedca\ldots aebedca\ldots ad \\ &\Rightarrow adcbedca\ldots aebedca\ldots ad \\ &\Rightarrow f_2(b)\ldots f_2(d)\ldots. \end{aligned}$$

Triple *-ce*

Here

$$\begin{aligned} \ldots c\ldots e &\Rightarrow \ldots bc\ldots be \\ &\Rightarrow \ldots ebc\ldots ebe \\ &\Rightarrow \ldots aebc\ldots aebe \\ &\Rightarrow \ldots f_2(c)\ldots f_2(d) \text{ or } \ldots f_2(c)\ldots f_2(e). \end{aligned}$$

Triple *ea-*

Here

$$\begin{aligned} e\ldots a\ldots &\Rightarrow ed\ldots ad\ldots \\ &\Rightarrow edc\ldots adc\ldots \\ &\Rightarrow edcb\ldots adcb\ldots \\ &\Rightarrow edcbebc\ldots adcbebc\ldots \\ &\Rightarrow f_2(e)\ldots f_2(a)\ldots \\ &\Rightarrow f_2(e)\ldots f_2(c)f_2(a)\ldots f_2(c) \\ & \text{since } N^-(a)=\{c\}. \end{aligned}$$

Thus eu^*au^* appears in $f_2^{n-1}(a)$, which is impossible.

Triple *-de*

What could be the identity of the letter '$-$'? Triple *ade* is impossible, since $N(a)\cap N(d)=\phi$. Similarly we can rule out *cde*, *ede*. On the other hand, *dde* is repetitive. Thus we need only eliminate the possibility of *bde*:

$$\begin{aligned} b\ldots d\ldots e &\Rightarrow b\ldots ad\ldots ae \\ &\Rightarrow bc\ldots adc\ldots ae \\ &\Rightarrow ebc\ldots adc\ldots ae \\ &\Rightarrow ebcbe\ldots adcbe\ldots ae. \end{aligned}$$

However, then *ecbc* appears in $f_2^n(a)$, which is impossible.

Triple *bd-*

What could be the identity of the letter '$-$'? By the previous case we cannot have *bde*. Triple *bda* is impossible, since $N^-(d)\cap N^-(a)=\phi$. Similarly we can rule out bdc. On the other hand, bdd is repetitive. Thus we need only eliminate the possibility of bdb:

$$\begin{aligned}
b\dots d\dots b &\Rightarrow bc\dots dc\dots b\\
&\Rightarrow bc\dots edc\dots eb\\
&\Rightarrow bc\dots bedc\dots beb\\
&\Rightarrow bc\dots dcbedc\dots dcbeb\\
&\Rightarrow bc\dots adcbedc\dots adcbeb\\
&\Rightarrow bca\dots adcbedca\dots adcbeb\\
&\Rightarrow f_2(z)\dots f_2(b)\dots f_2(a) \text{ or } f_2(z)\dots f_2(b)\dots f_2(b)\\
&\ \text{where } z\in\{a,c,e\}\\
&\Rightarrow f_2(a)\dots f_2(b)\dots f_2(a)\\
&\ \text{since } N(c)\cap N(b) = N(e)\cap N(b) = \phi.
\end{aligned}$$

It follows that a word $aubua$ is found in $f_2^{n-1}(a)$. However,

$$\begin{aligned}
a\dots a\dots a &\Rightarrow ae\dots cbe\dots ca\\
&\Rightarrow ae\dots edcbe\dots edca\\
&\Rightarrow aebc\dots aebedcbebc\dots aebedca\\
&\Rightarrow f_2(c)\dots f_2(e)\dots f_2(d)\\
&\ \text{which is impossible.}
\end{aligned}$$

∎

Let $\tilde{R}=\{f_2(a)v : v\in R^*\}$. Define $g_2 : \tilde{R}\to R^*$ by

$$g_2(adcbebcv) = abcv.$$

Lemma 5.8. *Let $n\in\mathbb{N}$. Then $g_2(f_2^n(a))$ is non-repetitive and contains no $xuyuz$ such that $f_1(xyz)$ is repetitive.*

Proof. Write $f_2^n(a) = adcbebcv$. Then $g_2(f_2^n(a)) = abcv$. Since bcv is a subword of $f_2^n(a)$, word bcv is non-repetitive and contains no $xuyuz$ such that $f_1(xyz)$ is repetitive. Suppose $abcv=\alpha\beta\beta\gamma, \beta\neq\varepsilon$. Then $\alpha=\varepsilon$, since bcv is non-repetitive. If $|\beta|=1$, then $b=c$, which is absurd. Suppose that $|\beta|>1$. Then ab is a subword of β. Then ab appears in $abcv$ twice, hence in bcv. This is a contradiction, since bcv is a subword of $f_2^n(a)$, which contains no ab.

Suppose that $abcv$ contains a subword $xuyuz$ such that $f_1(xyz)$ is repetitive. In this case we can in fact write $abcv=xuyuz\gamma$, some γ, since $xuyuz$ cannot be a subword of bcv. Thus $x=a$ or '$-$' (in the notation of Table 3). Referring to Table 3, we see that the triple xyz is one of $ac-, aec$ or aed. We eliminate these possibilities in two steps:

Triple *ac-*

Here

$$\begin{aligned}
a\dots c\dots &\Rightarrow ab\dots cb\dots ca\\
&\ \text{since our word commences } abc\\
&\Rightarrow abc\dots cbc\dots
\end{aligned}$$

This is impossible, since cbc is never a subword of $f_2^n(a)$.

Triple *aec*, *aed*

Here

$$\begin{array}{rcl} a\ldots e\ldots & \Rightarrow & abc\ldots ebc\ldots \\ & & \text{since our word commences } abc \\ & \Rightarrow & abc\ldots ebc\ldots \\ & \Rightarrow & abcf_2(d)\ldots ebcf_2(d)ebc\ldots \\ & & \text{since } f_2^n(a) \text{ must commence } f_2(ad). \\ & \Rightarrow & abcf_2(d)\ldots f_2(a)f_2(d)\ldots \text{ or } abcf_2(d)\ldots f_2(e)f_2(d)\ldots \\ & & \text{since } N^-(d)=\{a,e\}. \end{array}$$

Replacing abc by $f_2(a)$, it follows that in $f_2^n(a)$, we have either $f(a)f(u)f(a)f(u)$, some u, or $f(a)f(u)f(e)f(u)$. The first of these implies a repetition in $f_2^n(a)$, which is impossible, while the second case was shown impossible in the section of the proof of Lemma 4.11, dealing with triples *aed*, *aec*. ∎

Proof of Proposition 3.6. Let $\hat{L}_{121,212}$ be the set of non-repetitive words over R, walkable on D, and never containing $xuyuz$ for any triple xyz where $f_1(xyz)$ is repetitive. Using the construction of the proof of Proposition 3.2 with g_2 for g_1, f_2 for f_0, one shows that there are uncountably many words in $\hat{L}_{121,212}$ with different ages. However, if $u,v \in \hat{L}_{121,212}$ have different ages, then so do $f_1(u), f_1(v) \in L_{121,212}$. ∎

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, George. F. McNulty and Walter Taylor: Growth problems for avoidable words, *Theoret. Comput. Sci.*, **69** (1989), No. 3, 319–345.

[3] Dwight R. Bean, Andrzej Ehrenfeucht and George McNulty: Avoidable Patterns in Strings of Symbols, *Pacific J. Math.*, **85** (1979), 261–294.

[4] J. Brinkhuis: Non-repetitive sequences on three symbols, *Quart. J. Math. Oxford*, Ser.(2) **34** (1983), 145–149.

[5] 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.

[6] Stanley Burris and Evelyn Nelson: Embedding the dual of in the lattice of equational classes of semigroups, *Algebra Universalis*, **1** (1971/72), 248–253.

[7] Julien Cassaigne: Mots évitables et régularité dans les mots, Ph.D. thesis, L.I.T.P. (1994).

[8] JAMES D. CURRIE: Non-repetitive walks in graphs and digraphs, Ph.D. thesis, University of Calgary (1987).

[9] JAMES D. CURRIE: Which graphs allow infinite non-repetitive walks?, *Discrete Math.*, **87** (1991), 249–260.

[10] JAMES. D. CURRIE: Open problems in pattern avoidance, Amer. Math. Monthly **100** (1993), 790–793.

[11] JAMES. D. CURRIE: On the structure and extendibility of k-power free words, *Eur. J. Comb.*, to appear.

[12] JAMES. D. CURRIE: Subwords of non-repetitive words, *J. Combin. Theory*, Ser. A., to appear.

[13] F. M. DEKKING: Strongly non-repetitive sequences and progression-free sets, *J. Combin. Theory, Ser. A*, **27** (1979), 181–185.

[14] FRANÇOISE DEJEAN: Sur un théorème de Thue, *J. Combin. Theory, Ser. A*, **13** (1972), 90–99.

[15] ROGER C. ENTRINGER, DOUGLAS E. JACKSON and J. A. SCHATZ: On non-repetitive sequences, *J. Combin. Theory, Ser. A*, **16** (1974), 159–164.

[16] EARL D. FIFE: Binary sequences which contain no BBb, *Trans. Amer. Math. Soc.*, **261** (1980), 115–136.

[17] ANDRES DEL JUNCO: A transformation with simple spectrum which is not rank one, *Canad. J. Math.*, **29** (1977), 655–663.

[18] JACQUES JUSTIN: Characterization of the repetitive commutative semigroups, *J. Algebra*, **21** (1972), 87–90.

[19] JUHANI KARHUMÄKI: On cube-free ω-words generated by binary morphisms, *Discrete Appl. Math.*, **5** (1983), 279–297.

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

[21] M. LOTHAIRE: *Combinatorics on Words, Encyclopedia of Mathematics and its Applications* Vol. 17, Addison-Wesley, Reading Mass., 1983.

[22] FILIPPO MIGNOSI: Infinite words with linear subword complexity, *Theoret. Comput. Sci.*, **65** (1989), 221–242.

[23] MARSTON MORSE and GUSTAV A. HEDLUND: Symbolic dynamics I, II, *Amer. J. Math.*, **60** (1938), 815–866; **62** (1940) 142.

[24] P. S. NOVIKOV and S. I. ADJAN: Infinite periodic groups I, II, III, *Izv. Akad. Nauk. SSSR*, Ser. Mat. **32** (1968), 212–244; 251–524; 709–731.

[25] P. A. B. PLEASANTS: Non-repetitive sequences, *Proc. Cambridge Philos. Soc.*, **68** (1970), 267–274.

[26] PETER ROTH: Every binary pattern of length six is avoidable on the two-letter alphabet, *Acta Inf.*, **29** (1992), 95–106.

[27] ROBERT O. SHELTON and RAJ P. SONI: Chains and fixing blocks in irreducible binary sequences, *Discrete Math.*, **54** (1985), 93–99.

[28] ROBERT O. SHELTON and 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;

[29] AXEL THUE: Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana*, (1912), 1–67.

[30] WILLIAM T. TROTTER and PETER WINKLER: Arithmetic Progressions in Partially Ordered Sets, *Order*, **4** (1987), 37–42.

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

James D. Currie

Department of Mathematics
and Statistics
University of Winnipeg
Winnipeg, Manitoba
Canada R3B 2E9
currie@io.uwinnipeg.ca