ON SUBSTITUTIONS CLOSED UNDER DERIVATION: EXAMPLES

VÁCLAV KOŠÍK AND ŠTĚPÁN STAROSTA

Abstract. We study infinite words fixed by a morphism and their derived words. A derived word is a coding of return words to a factor. We exhibit two examples of sets of morphisms which are closed under derivation — any derived word with respect to any factor of the fixed point is again fixed by a morphism from this set. The first example involves standard episturmian morphisms, and the second concerns the period doubling morphism.

1. Introduction

In 1998 Fabien Durand characterized primitive substitutive sequences, i.e., morphic images of fixed points of primitive substitutions. A crucial role in his characterization is played by the notion “derived word”. Any primitive substitutive sequence $u$ is uniformly recurrent, i.e., for each factor $w$, the distances between consecutive occurrences of $w$ in $u$ are bounded. Or equivalently, there are only finitely many gaps between neighbouring occurrences of $w$. An infinite word coding ordering of these gaps (seen as finite words) is called the derived word to $w$ in $u$ and is denoted $d_u(w)$.

The mentioned main result of [2] says that a uniformly recurrent word is primitive substitutive if and only if the set of derived words to all prefixes of $u$ is finite. If moreover, $u$ is fixed by a primitive substitution, then the derived word to a prefix $w$ of $u$ is fixed by a primitive substitution as well. In other words, given any primitive substitution $\varphi$, there exists a finite list $L = \{\varphi_1, \varphi_2, \ldots, \varphi_k\}$ of primitive substitutions such that for each prefix $w$ of $u$, the fixed point of $\varphi$, the derived word $d_u(w)$ is fixed by a substitution $\varphi_i$ from $L$. An algorithm which to a given Sturmian substitution creates such list $L$ is described in [7].

On the other hand, if $w$ is a non-prefix factor of $u$, then it seems that $d_u(w)$ is fixed by a substitution only exceptionally. In [5], this phenomenon is studied for fixed points of Sturmian substitutions. For this purpose, the following new notion has been introduced:

Definition 1. A finite non-empty set $M$ of primitive substitutions is said to be closed under derivation if the derived word $d_u(w)$ to any factor $w$ of any fixed point $u$ of $\varphi \in M$ is fixed (after a suitable renaming of letters) by a substitution $\psi \in M$. 

(V. Košík) Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, 115 19, Prague 1, Czech Republic

(Š. Starosta) Department of Applied Mathematics, Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, 160 00, Prague 6, Czech Republic
A primitive substitution \(\xi\) is said to be closeable under derivation if it belongs to a set \(M\) closed under derivation.

Sturmian substitutions closeable under derivation are characterized in [5]. The aim of this contribution is to provide two new examples of sets \(M\) closed under derivation.

In our first example, in Section 4, the set \(M\) is a finite subset of the monoid of episturmian morphisms. In this case, all substitutions in \(M\) act on the same alphabet. In our second example, in Section 5, the substitutions in \(M\) act on alphabets with distinct cardinality. An inspiration for the second example comes from a recent result by Huang and Wen in [4], where a curious property of the period doubling substitution \(\psi(a) = ab\) and \(\psi(b) = aa\) was observed.

2. Preliminaries

Let \(\mathcal{A}\) denote an alphabet — a finite set of symbols. A word over \(\mathcal{A}\) is a finite sequence \(u = u_1u_2\ldots u_n\), where \(u_i \in \mathcal{A}\) for all \(i = 1, 2, \ldots, n\). The length of the word \(u\) is denoted \(|u|\) and is equal to \(n\). The set of all words over \(\mathcal{A}\) together with the operation concatenation forms a free monoid \(\mathcal{A}^*\), its neutral element is the empty word \(\varepsilon\). If \(u = pws \in \mathcal{A}^*\), then \(w\) is a factor of \(u\), \(p\) is a prefix of \(u\), and \(s\) is a suffix of \(u\). For \(w = uv\), we write \(u = uw^{-1}\) and \(v = u^{-1}w\).

A morphism \(\varphi\) is a mapping \(\varphi : \mathcal{A}^* \mapsto \mathcal{B}^*\) such that \(\varphi(uv) = \varphi(u)\varphi(v)\) for all \(u, v \in \mathcal{A}^*\). A morphism \(\varphi : \mathcal{A}^* \mapsto \mathcal{A}^*\) is called primitive if there exists an iteration \(k \in \mathbb{N}\) such that for any pair \(a, b\) of letters from \(\mathcal{A}\), the letter \(a\) occurs in \(\varphi^k(b)\). In accordance with Durand’s terminology, a morphism \(\varphi\) is a substitution if there exist \(a \in \mathcal{A}\) and \(w \in \mathcal{A}^*, w \neq \varepsilon\) such that \(\varphi(a) = aw\) and \(|\varphi^n(a)|\) tends to infinity with growing \(n\).

An infinite word over \(\mathcal{A}\) is an infinite sequence \(u = u_0u_1u_2\ldots\) from \(\mathcal{A}^\mathbb{N}\). A finite word \(w\) of length \(n\) is a factor of \(u\) if there exists an index \(i \in \mathbb{N}\), such that \(w = u_iu_{i+1}\ldots u_{i+n-1}\). The index \(i\) is called an occurrence of \(w\) in \(u\). The set of all factors of \(u\) is denoted by \(\mathcal{L}(u)\). If each factor \(w\) of \(u\) has infinitely many occurrences, then \(u\) is recurrent. A return word to \(w\) in \(u\) is a factor \(r = u_iu_{i+1}\ldots u_{j-1}\), where \(i < j\) are two consecutive occurrences of \(w\) in \(u\). The word \(rw\) is called a complete return word to \(w\) in \(u\) and obviously, \(rw\) is a factor of \(u\). The set of all return words to \(w\) in \(u\) is denoted by \(\mathcal{R}_u(w)\). If the set \(\mathcal{R}_u(w)\) is finite, say \(\mathcal{R}_u(w) = \{r_0, r_1, \ldots, r_{k-1}\}\), then \(u\) can be written as a concatenation \(u = pr_{i_0}r_{i_1}r_{i_2}\ldots\), where \(p\) is the prefix of \(u\) such that the factor \(w\) occurs in \(pw\) exactly once. The infinite word \(i_0i_1i_2\ldots\) over the alphabet \(\{0, 1, 2, \ldots, k - 1\}\) is the derived word to \(w\) in \(u\) and is denoted \(\mathcal{D}_u(w)\). A recurrent infinite word \(u\) is uniformly recurrent if the set \(\mathcal{R}_u(w)\) is finite for all \(w \in \mathcal{L}(u)\).

The domain of a morphism \(\varphi : \mathcal{A}^* \mapsto \mathcal{B}^*\) is naturally extended to \(\mathcal{A}^\mathbb{N}\) by putting \(\varphi(u) = \varphi(u_0u_1u_2\ldots) = \varphi(u_0)\varphi(u_1)\varphi(u_2)\ldots\). A word \(u\) is purely substitutive if there exists a substitution \(\varphi\) over \(\mathcal{A}\) such that \(u = \varphi(u)\), i.e. \(u\) is a fixed point of \(\varphi\). A word \(v\) over \(\mathcal{B}\) is substitutive if \(v = \psi(u)\), where \(\psi : \mathcal{A}^* \mapsto \mathcal{B}^*\) is a morphism and \(u\) is a purely substitutive word. If \(u\) is fixed by a primitive substitution, then \(v\) is primitive substitutive. A well known fact is that a primitive substitutive word is uniformly recurrent (c.f. [2]).
3. The set of derived words to factors of an infinite word

In this section we list several simple properties of the set

$$\text{Der}_f(u) = \{d_u(w) : w \in L(u)\}.$$  

First, we show that only some special factors need to be examined to describe \(\text{Der}_f(u)\). A letter \(a \in A\) is a right extension of \(w \in L(u)\) if \(wa \in L(u)\). Note that any factor of \(u\) has at least one right extension. A factor \(w \in L(u)\) is right special if it has at least two distinct right extensions. Analogously, we define left special.

A factor which is simultaneously right and left special is bispecial.

**Proposition 2.** Let \(u\) be an infinite recurrent word over \(A\) and \(w \in L(u)\).

1. If \(w\) is not left special, then \(R_u(aw) = aR_u(w)a^{-1}\), where \(a \in A\) is the unique left extension of \(w\). Moreover, if \(w\) is not a prefix of \(u\), then \(d_u(aw) = d_u(w)\).
2. If \(w\) is not right special, then \(R_u(wa) = R_u(w)\) and \(d_u(wa) = d_u(w)\), where \(a \in A\) is the unique right extension of \(w\).

**Proof.** Item (1): First assume that \(w\) is not left special and \(w\) is not a prefix of \(u\). The integer \(i\) is an occurrence of \(w\) in \(u\) if and only if \(i - 1\) is an occurrence of \(aw\) in \(u\). Consequently, \(r \in R_u(w)\) if and only if \(ara^{-1} \in R_u(w)\) and the ordering of the return words to \(w\) in \(u\) and the ordering the return words to \(aw\) in \(u\) coincide.

Let \(0\) be an occurrence of \(w\), i.e., \(w\) is a prefix of \(u\). Then a return word \(r\) to \(w\) and \(rw\) have an occurrence \(0\). We have to show that even for such \(r\) the word \(ara^{-1}\) belongs to \(R_u(aw)\). Indeed, the word \(u\) is recurrent and thus \(rw\) has an occurrence \(j > 0\). As \(w\) always precedes the letter \(a\) and \(a\) is a suffix of \(r\) we can conclude that \(ara^{-1}\) is a return word to \(aw\) in \(u\).

Item (2): The proof is analogous. \(\square\)

We formulate a straightforward corollary of Proposition 2.

**Proposition 3.** Let \(u\) be an infinite recurrent word over \(A\). We have

$$\text{Der}_f(u) = \{d_u(w) : w \text{ is a right special prefix of } u\} \cup \{d_u(w) : w \text{ is a bispecial factor of } u\}.$$

The following claim is taken from Durand’s article. His proof is constructive and provides an algorithm for finding a suitable morphism.

**Proposition 4** ([2]). Let \(u \in A^n\) be a fixed point of a primitive morphism \(\phi\) and \(w\) be a prefix of \(u\). The word \(d_u(w)\) is fixed by a primitive morphism as well.

**Sketch of the proof.** We do not repeat the whole proof, we only describe the construction of a primitive morphism fixing \(d_u(w)\).

Let \(r_0, r_1, \ldots, r_{k-1}\) be the return words to \(w\). Since \(u\) is fixed by \(\phi\), the image \(\phi(w)\) has a prefix \(w\) and thus \(\phi(r_iw)\) has a prefix \(\phi(r_i)w\). As \(w\) is a prefix and a suffix of \(\phi(r_i)w\), the factor \(\phi(r_i)\) is concatenation of several return words to \(w\), i.e. we can find unique indices \(s_1, s_2, \ldots, s_{\ell_i} \in \{0, 1, \ldots, k - 1\}\) such that \(\phi(r_i) = r_{s_1}r_{s_2}\cdots r_{s_{\ell_i}}\).

It is easy to check that the morphism given by

$$\delta : i \mapsto s_1s_2\cdots s_{\ell_i}, \text{ for each } i \in \{0, 1, \ldots, k - 1\}$$

is primitive and fixes \(d_u(w)\). All details can be found in [2]. \(\square\)
Proposition 5. Let $u \in \mathcal{A}^\mathbb{N}$ be a fixed point of a primitive morphism $\varphi$ and $w \in \mathcal{L}(u)$. The word $d_u(w)$ is primitive substitutive.

Proof. Let $pw$ be the shortest prefix of $u$ containing the factor $w$. Denote by $r_0, r_1, \ldots, r_{k-1}$ the return words to $pw$ and by $\tilde{r_0}, \tilde{r_1}, \ldots, \tilde{r}_{j-1}$ the return words to $w$. As $w$ is a prefix and a suffix of the factor $p^{-1}r_ipw$, the word $p^{-1}r_ipw$ can be written as concatenation of the return words to $w$, i.e. $p^{-1}r_ipw = \tilde{r}_{s_1} \tilde{r}_{s_2} \cdots \tilde{r}_{s_{k-1}}$ for some indices $s_1, s_2, \ldots, s_{k-1} \in \{0, 1, \ldots, j-1\}$. Define a morphism $\psi : \{0, 1, \ldots, k-1\}^* \rightarrow \{0, 1, \ldots, j-1\}^*$ by

\[ \psi : i \mapsto s_is_{i+1} \cdots s_{k-1} \quad \text{for each } i \in \{0, 1, \ldots, k-1\}. \]

It follows that $d_u(w) = \psi(d_u(pw))$. By Proposition 4, $d_u(pw)$ is fixed by a primitive substitution. \qed

We finish this section by an example.

Example 6. Recall the period doubling substitution

\[ \psi(a) = ab \quad \text{and} \quad \psi(b) = aa, \]

and its fixed point

\[ z = abaaababaaababaaabaa \ldots. \]

- Any occurrence of the letter $b$ is preceded and followed by the letter $a$, therefore $b$ is neither right nor left special. By Proposition 2,

\[ d_z(b) = d_z(ab) = d_z(aba). \]

- There are two return words to $a$ in $z$, namely $r_0 = ab$ and $r_1 = a$. We can write

\[ z = r_0r_1r_0r_0r_0r_1r_1r_1r_1r_0r_1r_0r_1 \ldots \quad \text{and thus} \quad d_z(a) = 0110001101101 \ldots. \]

The word $d_z(a)$ is fixed by a substitution. To find it, we compute

\[ \psi(r_0) = \psi(ab) = abaa = r_0r_1r_1 \quad \text{and} \quad \psi(r_1) = \psi(a) = ab = r_0. \]

It follows from the proof of Proposition 4 that $d_z(a)$ is fixed by the substitution $\xi$ determined by

\[ \xi(0) = 011 \quad \text{and} \quad \xi(1) = 0. \]

4. Example 1: Standard episturmian morphisms

Let us recall the definition of standard Arnoux–Rauzy words and known results on morphisms fixing these words. All mentioned facts and further results can be found in the survey [3].

Definition 7. An infinite word $u \in \mathcal{A}^\mathbb{N}$ is Arnoux–Rauzy if

1. $u$ has exactly one right special factor of each length;
2. $wa \in \mathcal{L}(u)$ for every right special factor $w$ of $u$ and every letter $a \in \mathcal{A}$;
3. $\mathcal{L}(u)$ is closed under reversal, i.e. $v_1v_2 \cdots v_n \in \mathcal{L}(u)$ implies $v_nv_{n-1} \cdots v_1 \in \mathcal{L}(u)$.

An Arnoux–Rauzy word $u$ is standard if each of its prefixes is a left special factor of $u$. 


The Arnoux–Rauzy words represent a generalization of Sturmian words to multi-literal alphabets and share many properties with Sturmian words. A property which is important for a description of their derived words is that Arnoux–Rauzy words are aperiodic and by \( M \) they are also uniformly recurrent. Let \( M_A \) denote the monoid generated by standard episturmian morphisms \( L_a \) defined for every \( a \in A \) as follows:
\[
L_a : \begin{cases} 
    a \rightarrow a \\
    b \rightarrow ab \text{ for all } b \neq a.
\end{cases}
\]

To abbreviate the notation of elements of the monoid \( M_A \), we put \( L_z = L_{z_1} \circ L_{z_2} \circ \cdots \circ L_{z_n} \) for \( z = z_1z_2 \cdots z_n \in A^* \).

A morphism \( L_z \in M_A \) is primitive if and only if each letter from \( A \) occurs in \( z \).

Any primitive morphism in \( M_A \) has only one fixed point and this fixed point is a standard Arnoux–Rauzy word. On the other hand, if a standard Arnoux–Rauzy word is fixed by a primitive substitution, then it is fixed by a primitive morphism from the monoid \( M_A \).

Example 8. Let us consider the Tribonacci word \( u_\tau = abacabaabacababacabaa \cdots \) — the fixed point of the morphism \( \tau : a \rightarrow ab, b \rightarrow ac, c \rightarrow a \). The word \( u_\tau \) is a standard Arnoux-Rauzy word over \( \{a,b,c\} \) and it is fixed also by the morphism \( \tau^3 \). It is easy to check that \( \tau^3 = L_{abc} \) and thus the Tribonacci word is fixed by a substitution from \( M_A \).

K. Medková in [8] studies derived words of Arnoux–Rauzy words. She considers all Arnoux–Rauzy (not only standard) words, but she describes derived words only to prefixes of infinite words. To quote a consequence of one of her results we need to recall the cyclic shift operation on \( A^* \):
\[
\text{cyc}(z_1z_2 \cdots z_n) = z_nz_1 \cdots z_{n-1}.
\]

Proposition 9 (Theorem 24 in [8]). Let \( L_z \in M_A, z \in A^* \), be a primitive morphism and \( u \) be its fixed point. If \( w \) is a prefix of \( u \), then there exists \( k \in \{1,2,\ldots,|z|\} \) such that \( d_u(w) \) is fixed (up to a permutation of letters) by \( L_{\text{cyc}^k(z)} \). In particular, the word \( d_u(w) \) is a standard Arnoux–Rauzy word.

Theorem 10. Let \( z \) be a word in \( A^* \) such that each letter \( a \in A \) occurs in \( z \) at least once. The set
\[
M = \{ L_{\text{cyc}^k(z)} : k \in \{1,2,\ldots,|z|\} \}
\]
is closed under derivation.

Proof. Let \( v \) be a fixed point of \( L_v \) with \( v = \text{cyc}^k(z) \) for some \( k \in \{1,2,\ldots,|z|\} \). Since \( z \) contains each letter from \( A \), the word \( v \) contains all letters from \( A \) as well and thus \( L_v \) is primitive.

As \( v \) is a standard Arnoux–Rauzy word, each of its bispecial factors is a prefix of \( v \). By Proposition 3, only derived words to prefixes have to be considered. By Proposition 9, each such derived word is fixed (up to a permutation of letters) by a morphism \( L_{\text{cyc}^i(v)} \) for some \( i \in \{1,2,\ldots,|v|\} \). Obviously, this morphism belongs to \( M \).

Example 11. If we apply the previous theorem to the ternary word \( abc \), we obtain that the set \( M = \{ L_{abc}, L_{bca}, L_{cab} \} \) is closed under derivation. Nevertheless, all the 3 morphisms in \( M \) fix (up to a permutation of letters) the same word, namely
the Tribonacci word. This word is fixed by the substitution \( \tau \) given in Example 8. Therefore, the set \( \{ \tau \} \) is closed under derivation as well.

5. Example 2: The period doubling morphism

The aim of this section is to show that the period doubling substitution \( \psi \) determined by \( \psi(a) = ab \) and \( \psi(b) = aa \) is closeable under derivation. For this purpose, we first define the two following substitutions:

\[
\nu : \begin{cases} 
0 \mapsto 01, \\
1 \mapsto 02020101, \\
2 \mapsto 0202,
\end{cases} \quad \xi : \begin{cases} 
0 \mapsto 011, \\
1 \mapsto 0.
\end{cases}
\]

Next, we deduce several auxiliary statements which help us to prove the following main theorem.

**Theorem 12.** The sets \( \{ \psi, \xi, \nu \} \) and \( \{ \xi, \nu \} \) are closed under derivation.

First, we focus on the derived words of the fixed point \( z = abaaababaaabaaabaaabaa \cdots \) of the substitution \( \psi \). The following properties are immediate:

- \( bb \notin L(z) \). If \( a^i \in L(z) \), then \( i \leq 3 \).
- \( a \) and \( aa \) are bispecial factors of \( z \).
- Any bispecial factor of length more than 2 has a prefix \( ab \) and a suffix \( ba \).
- The longest common prefix of \( \psi(a) \) and \( \psi(b) \) is the letter \( a \); the longest common suffix of \( \psi(a) \) and \( \psi(b) \) is the empty word. It implies that \( \Phi(v) := \psi(v)a \) is bispecial whenever \( v \) is bispecial.

The converse of the very last property also holds (if \( \Phi(v) \) is not too short):

**Proposition 13.** Let \( w \) be a non-empty bispecial factor of \( z \) such that \( w \neq a \) and \( w \neq aa \). There exists a bispecial factor \( v \) such that \( \Phi(v) = w \).

**Proof.** As mentioned before, the bispecial factor \( w \) has a suffix \( ba \) and a prefix \( ab \). Hence, there exists a factor \( v \) such that \( \Phi(v) = \psi(v)a = w \) and \( a \) is both a prefix and a suffix of \( v \). It remains to show that \( v \) is bispecial. If it is not right special, then \( v \) is followed only by \( a \) or \( b \). But then \( w \) is followed only by \( b \) or \( a \), respectively, since \( \psi(va) = \psi(v)ab \) and \( \psi(vb) = \psi(v)aa \). Thus, \( w \) is right special. Similarly, \( v \) is left special, and therefore bispecial. \( \square \)

As the fixed point \( z \) has a bispecial factor \( aa \) which is not a prefix of \( z \), the description of derived words to non-prefix factors is more complicated than in the case of a fixed point of a standard episturmian morphism. The following notion will be very useful for this purpose.

**Definition 14.** Let \( w \) be a non-empty factor of a fixed point \( x \) of a substitution \( \varphi \). Suppose there exist words \( y, y' \) and \( u = u_1u_2 \cdots u_n \) such that \( ywy' = \varphi(u) \), \( |y| < |\varphi(u_1)| \), \( |y'| < |\varphi(u_n)| \), and \( u \in L(x) \). If there is exactly one occurrence of \( w \) in \( \varphi(u) \), then we call \( u \) an ancestor of \( w \). The set of all ancestors of \( w \) is denoted by \( A(w) \). If there are more occurrences of \( w \) in \( \varphi(u) \), then we say \( w \) allows an ambiguous ancestor.

**Example 15.** Given the fixed point \( z = abaaababaaabaaabaaabaa \cdots \) of the period doubling substitution \( \psi \), the set of all ancestors of the factor \( aa \) is \( A(aa) = \{ b \} \) because \( \psi(b) = aa \) and \( y = \varepsilon, y' = \varepsilon \). Since \( \psi(ba) = aaab \), \( y = a, y' = b \) and there
are two occurrences of \(aa\) in \(\psi(ba)\), the factor \(aa\) allows an ambiguous ancestor. The prefix \(aba\) has two ancestors \(aa\) and \(ab\) and it does not allow an ambiguous ancestor.

**Proposition 16.** Let \(x\) be a fixed point of an injective substitution \(\varphi\) and \(w\) be a factor of \(x\) with a unique ancestor \(u\). Assume \(w\) does not allow an ambiguous ancestor. We have \(d_x(w) = d_x(u)\).

**Proof.** The infinite word \(x\) can be written as \(x = zr_{i_0}r_{i_1}r_{i_2}\cdots\), where \(r_{i_j} \in \mathcal{R}_x(u)\) for all \(j \in \mathbb{N}_0\). If \(u\) is a prefix, then \(z = \varepsilon\). By the definition of a return word, \(u\) is a prefix of the word \(r_{i_k}u\cdots\) for all \(k \in \mathbb{N}_0\). Since \(u\) is a unique ancestor of \(w\) and \(w\) does not allow an ambiguous ancestor, there are exactly two occurrences of \(w\) in \(\varphi(r_{i_k})\varphi(u)\). Let \(\varphi(u) = ywy'\).

\[
\varphi(r_{i_k}u) = yw \quad | \quad y'w
\]

**Figure 1.** An illustration of \(r_{i_k}u\) and \(\varphi(r_{i_k}u)\) in the proof of Proposition 16.

If we define \(r_{i_k}' := y^{-1}\varphi(r_{i_k})y\) as in Figure 1, then \(r_{i_k}' \in \mathcal{R}_x(w)\) for all \(k \in \mathbb{N}_0\) and we have

\[
x = \varphi(x) = \varphi(z)\varphi(r_{i_0})\varphi(r_{i_1})\varphi(r_{i_2})\cdots = \varphi(z)y'y^{-1}\varphi(r_{i_0})y'y^{-1}\varphi(r_{i_1})y'y^{-1}\varphi(r_{i_2})y'y^{-1}\cdots = z'r_{i_0}'r_{i_1}'r_{i_2}'\cdots.
\]

The derived words of \(u\) and \(w\) are both \(i_0i_1i_2\cdots\). \(\square\)

**Lemma 17.** Let \(v\) be a non-empty bispecial factor of the fixed point \(z\) of the period doubling substitution \(\psi\). We have \(d_z(\Phi(v)) = d_z(v)\).

**Proof.** Since \(v\) is bispecial, the word \(a\) is a suffix of \(v\) and thus \(\psi(v)\) has a suffix \(b\). It implies that \(\psi(v)\) is not right special. Therefore \(d_z(\psi(v)) = d_z(\psi(v)a) = d_z(w)\) with \(w = \Phi(v) = \psi(v)a\).

The word \(v\) is surely an ancestor of \(\psi(v)\). We show that it is the only ancestor. Suppose there is another ancestor \(t\) with \(t \neq v\). Since \(\psi\) is injective, there exist \(y, y' \neq \varepsilon\) such that \(y\psi(v)y' = \psi(t)\). It follows that \(y\) and \(y'\) are both letters. Thus, the last letter of \(\psi(v)\) is the first letter of \(\psi(a)\) or \(\psi(b)\) which is in both cases the letter \(a\) — a contradiction. Therefore \(A(\psi(v)) = \{v\}\) and it is not difficult to verify that \(\psi(v)\) does not allow an ambiguous ancestor when it contains at least one letter \(b\). By Proposition 16 we have \(d_z(v) = d_z(\psi(v)) = d_z(w)\). \(\square\)

**Proposition 18.** If \(w\) is a non-empty factor of \(z\), then \(d_z(w) = d_z(a)\) or \(d_z(w) = d_z(aa)\). If \(w\) is a non-empty prefix of \(z\), then \(d_z(w) = d_z(a)\).
Proof. By Proposition 2 we have to describe the derived words to right special prefixes and to bispecial factors only. First assume that $w$ is a bispecial factor of $z$. By Proposition 13, the factor $w$ can be obtained by iteration of the mapping $\Phi(v) = \psi(v)a$ starting from the two initial bispecial factors $a$ and $aa$ (in fact, this a special case of a general construction of bispecial factors from [6]). By Lemma 17, $d_z(w)$ equals to $d_z(a)$ or to $d_z(aa)$.

Now assume that $w$ is a right special prefix of $z$. As the initial bispecial factor $a$ is a prefix of $z$, the bispecial factor $\Phi^k(a)$ is a prefix of $z$ for each $k \in \mathbb{N}$. Therefore, any right special prefix $w$ of $z$ is left special as well. More specifically, any right special prefix of $z$ equals to $\Phi^k(a)$ for some $k \in \mathbb{N}$ and by Lemma 17, $d_z(w) = d_z(\Phi^k(a)) = d_z(a)$.

Now we show that both derived words to a factor of $z$ are fixed by primitive substitutions. We exploit the following simple tool.

**Observation 19.** Let $v$ be a fixed point of a morphism $\gamma$ and let $u = \alpha(v)$ where $\alpha$ is a morphism. If there exists a morphism $\beta$ such that $\alpha\gamma = \beta\alpha$, then $u$ is fixed by $\beta$.

**Proof.** $\beta(u) = \beta\alpha(v) = \alpha\gamma(v) = \alpha(v) = u$. \hfill $\Box$

**Proposition 20.** The derived word $d_z(a)$ is fixed by $\xi$ and the derived word $d_z(aa)$ is fixed by $\nu$ (where $\xi$ and $\nu$ are defined in (1)).

**Proof.** In Example 6 above, we show that the derived word $d_z(a)$ is fixed by the substitution $\xi$.

It remains to consider $d_z(aa)$. As $abaa$ is the shortest prefix of $z$ containing the bispecial factor $aa$, we use the construction from the proof of Proposition 5 to find a morphism $\alpha$ such that $d_z(aa) = \alpha(d_z(abaa))$. In our case $p = ab$ and $w = aa$. According to Proposition 18, the derived word $d_z(abaa)$ is fixed by $\xi$ since $d_z(a)$ is fixed by $\xi$. Thus, $d_z(abaa)$ is over a binary alphabet, and so the prefix $abaa$ has exactly two return words, say $r_0$ and $r_1$. These two return words can be found in the prefix of $z$ of length 16. They are

$$r_0 = abaaabab \quad \text{and} \quad r_1 = abaa.$$ 

It follows from the proof of Proposition 5 that $(ab)^{-1}r_0ab$ and $(ab)^{-1}r_1ab$ can be written as a concatenation of return words to $aa$. Specifically, $r_0' = a, r_1' = aababab, r_2' = aab$ are return words of $aa$ and $(ab)^{-1}r_0ab = r_0'r_1'$ and $(ab)^{-1}r_1ab = r_0'r_2'$. Hence, according to this claim we have

$$\alpha(0) = 01,$$
$$\alpha(1) = 02.$$

Note that since $d_z(abaa)$ is fixed by $\xi$, it is also fixed by $\xi^2$. By Observation 19, if the substitution $\nu$ satisfies $\alpha\xi^2 = \nu\alpha$, the proof is finished. This is very easy to verify:

$$\alpha\xi^2(0) = \alpha(01100) = 0102020101$$
$$\nu\alpha(0) = \nu(01) = 0102020101$$
$$\alpha\xi^2(1) = \alpha(011) = 010202$$
$$\nu\alpha(1) = \nu(02) = 010202.$$ \hfill $\Box$
Remark 21. The derived word \( d_z(aa) \) is also fixed by the morphism

\[
\begin{align*}
\eta(0) &= \varepsilon \\
\eta(1) &= 010202 \\
\eta(2) &= 01.
\end{align*}
\]

A proof is the same as the proof of Proposition 20, but at the end we have to verify the equality \( \alpha \xi = \eta \alpha \). The reason why we prefer \( \nu \) to \( \eta \) is that \( \eta \) is an erasing non-primitive morphism.

Corollary 22. If \( w \) is a non-empty factor of \( z \), then \( d_z(w) \) is fixed by \( \xi \) or \( \nu \).

Proof. The corollary follows from Propositions 18 and 20.

We conclude this section by the proof of our main result. For this purpose we need one more ingredient. It is a modification of Proposition 6, Item 5 from [2]. Its proof is almost identical with the proof of the original statement and thus we omit it.

Lemma 23. Let \( u \) be a uniformly recurrent word and let \( w \) be its factor. Set \( v = d_u(w) \). For a factor \( x \) of \( v \), there exists a factor \( y \) of \( u \) such that \( d_v(x) = d_u(y) \).

Proof of Theorem 12. Let \( v \) be a fixed point of the primitive substitution \( \xi \) and \( x \) be a factor of \( v \). By Proposition 20, we have \( v = d_z(a) \). By Lemma 23, there exists a factor \( y \) in \( z \) such that \( d_v(x) = d_z(y) \). Proposition 18 implies that \( d_v(x) \) equals \( d_z(a) \) or \( d_z(aa) \). Therefore, \( d_v(x) \) is fixed by \( \xi \) or \( \nu \).

The same reasoning gives that the derived word to any factor of the fixed point of \( \nu \) is fixed by \( \xi \) or by \( \nu \). By Definition 1, the set \( \{ \nu, \xi \} \) is closed under derivation.

As \( d_z(\varepsilon) = z \) and the derived word to any non-empty factor of \( z \) is fixed by \( \xi \) or by \( \nu \), the set \( \{ \nu, \xi, \psi \} \) is also closed under derivation.

Acknowledgments

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic, project no. CZ.02.1.01/0.0/0.0/16_019/0000778. We also acknowledge financial support of the Grant Agency of the Czech Technical University in Prague, grant No. SGS17/193/OHK4/3T/14. We thank to Michel Dekking for attracting our attention to the article [4].

References

[1] Droubay, X., Justin, J., Pirillo, G.: Episturmian words and some constructions of de Luca and Rauzy. Theoret. Comput. Sci. 255(1-2), 539–553 (2001). https://doi.org/10.1016/S0304-3975(99)00320-5
[2] Durand, F.: A characterization of substitutive sequences using return words. Discrete Math. 179(1–3), 89–101 (1998)
[3] Glen, A., Justin, J.: Episturmian words: a survey. Theor. Inf. Appl. 43(3), 403–442 (2009)
[4] Huang, Y.K., Wen, Z.Y.: Envelope words and the reflexivity of the return word sequences in the period-doubling sequence. Preprint available at https://arxiv.org/abs/1703.07157 (2017)
[5] Klouda, K., Pelantová, E., Starosta, Š.: Sturmian substitutions closed under derivation. In preparation (2019)
[6] Klouda, K.: Bispecial factors in circular non-pushy D0L languages. Theoret. Comput. Sci. 445(0), 63–74 (2012). https://doi.org/10.1016/j.tcs.2012.05.007
[7] Klouda, K., Medková, K., Pelantová, E., Starosta, Š.: Fixed points of Sturmian morphisms and their derivated words. Theoret. Comput. Sci. 743, 23–37 (2018). https://doi.org/10.1016/j.tcs.2018.06.037
[8] Medková, K.: derived sequences of Arnoux–Rauzy sequences. Submitted to WORDS 2019 (2019)