On Sturmian and episturmian words, and related topics

AMY GLEN

Combinatorics on words plays a fundamental role in various fields of mathematics, not to mention its relevance in theoretical computer science and physics. Most renowned among its branches is the theory of infinite binary sequences called *Sturmian words*, which are fascinating in many respects, having been studied from combinatorial, algebraic, and geometric points of view. The most well-known example of a Sturmian word is the ubiquitous *Fibonacci word*, the importance of which lies in combinatorial pattern matching and the theory of words. Properties of the Fibonacci word and, more generally, Sturmian words have been extensively studied, not only because of their significance in discrete mathematics, but also due to their practical applications in computer imagery (digital straightness), theoretical physics (quasicrystal modelling) and molecular biology.

The history of Sturmian words dates back to the astronomer J. Bernoulli III (1772) and, as described in Venkov's book [38], there also exists some early work by Christoffel (1875) and Markoff (1882). The first detailed investigation of Sturmian words was carried out in 1940 by Morse and Hedlund [33], who studied such words under the framework of symbolic dynamics and, in fact, introduced the term "Sturmian", named after the mathematician Charles François Sturm (1803–1855). It would be superfluous here to attempt to review the extensive literature on Sturmian words; instead, we refer the reader to the comprehensive surveys by Berstel [7] and Berstel and Séébold [12]. Allouche and Shallit's recent book [4] also contains a chapter on the subject, as does the book [35]. Moreover, our bibliography lists some more recent papers, which are not referred to in these surveys, and many other works that are not necessarily cited in the text of the thesis.

Despite having been an extremely active subject of research over the last twenty five years or so, there still remain numerous unanswered questions and problems concerning Sturmian words and, in particular, a recent natural generalization of them – the so-called *episturmian words* (introduced in 2001 by Droubay, Justin, and Pirillo [22]). The aim of this thesis is to go about solving some of them. On the whole, our approach is purely combinatorial in nature, but number theory will also be prevalent throughout, particularly when our attention turns to a transcendence theorem.

OUTLINE OF THESIS

In Chapter 2, we fix some notation and terminology, and present the necessary background material required for the original work in later chapters.

Our initial aim is to completely describe where palindromes appear in a given Sturmian word, as suggested by Jean Berstel [6, 7, 8, 9, 10, 11, 12]. A palindrome is a finite word that reads the same backwards as forwards. Palindromes are important tools used in the study of factors of Sturmian words (e.g., [16, 17, 19, 21, 23]). Due to their connections with *complexity* and *recurrence*, the palindromic factors of an infinite word provide an insight into its inherent structure; as such, there is also much general interest in palindromes (for instance, see [14, 15] and the survey [3]). With the aforementioned aim in mind, we begin our journey in Chapter 3 by investigating certain palindromic factors of the Fibonacci word **f**. We first consider previous results on factorizations of the Fibonacci word into *singular words*, which are palindromes, originally defined by Wen and Wen [39]. In a similar vein, we define a particular family of palindromes called *circular words*. After proving some properties of circular words, we establish decompositions of the Fibonacci word with respect to such words, thus generalizing the 'singular' decompositions of **f**.

Chapter 4 extends recent results of Levé and Séébold [31] on *generalized singular words* of the Fibonacci word. Specifically, we shall be concerned with characteristic Sturmian words c_α and $c_{1-\alpha}$ where $\alpha \in (0,1)$ is an irrational number with continued fraction expansion $[0; 1 + d_1, \overline{d_2, \ldots, d_n}]$, $d_n \geqslant d_1 \geqslant 1$. For the case when $\alpha = [0; 2, r, r, r, \ldots]$ for some $r \geqslant 1$, we shall establish a decomposition of each *conjugate* (or *suffix*) of c_α (and hence $c_{1-\alpha}$) into *generalized adjoining singular words*, by considering powers of the morphism by which it is generated (see [24]).

In Chapter 5, we complete our study of palindromes in Sturmian words, by describing precisely where palindromes occur in characteristic Sturmian words. In particular, given any palindromic factor u of c_α, we establish a decomposition of c_α with respect to the occurrences of u. Such a decomposition shows exactly where u appears in c_α, and this is directly related to the continued fraction expansion of α (see [26]).

In Chapter 6, attempting to extend results of Chapter 4 to infinite words over alphabets with more than two letters, we investigate the *Tribonacci sequence* ζ; a natural generalization of the Fibonacci word to a three-letter alphabet. The Tribonacci sequence (or *Rauzy word* [36]) and, more generally, the *k-bonacci word*, are special examples of *episturmian words* (see [22, 27, 29, 30, 34] for example), which include the well-known *Arnoux-Rauzy sequences* [5]. After considering numerous properties of the factors of the Tribonacci sequence, we explicitly determine all integer powers of words occurring in ζ. This is followed, in Chapter 7, by a similar analysis of integer powers occurring in a class of standard episturmian words whose *directive words* resemble those of characteristic Sturmian words. Here, key tools are canonical decompositions and a generalization of

singular words, which we call *singular n-words of the r-th kind*. We also prove a division property of such episturmian words with respect to the *lexicographic order*, generalizing a result of de Luca [20]. The majority of Chapter 7 appears in the paper [25].

In the penultimate chapter (Chapter 8), we prove that an irrational number is transcendental if the sequence of partial quotients in its continued fraction expansion forms a *strict* standard episturmian word. We also prove that certain *generalized Thue-Morse* continued fractions are transcendental. The author's interest in such transcendence results was sparked by a transcendence theorem concerning Sturmian words; specifically, a *Sturmian continued fraction* (of which the partial quotients form a Sturmian word) is transcendental, proved by Allouche *et al.* [2]. Using a different approach, Adamczewski and Bugeaud [1] very recently proved a result similar to our Theorem 8.12 with the weaker condition $\gamma > 1$ and the condition on frequencies removed.

It is well-known that the invertible substitutions on a two-letter alphabet are exactly the *Sturmian morphisms*. What can be said about invertible substitutions on an alphabet of more than two letters? Initially, one might suspect that they are exactly the *episturmian morphisms*. However, there exist invertible substitutions on three or more letters that are not episturmian. The structure of invertible substitutions on an arbitrary finite alphabet is far from being understood; it is our aim in the last chapter (Chapter 9) to elucidate their structure or at least characterize them. We do not succeed in this, but make a modest beginning towards the proof of a characterization (Conjecture 9.1).

Evidently, decompositions are a common theme throughout this thesis. Indeed, in Chapters 3 to 5, our main results involve decompositions of Sturmian words with respect to the occurrences of a given factor. Canonical *decompositions* of infinite words in terms of their 'building blocks' are also used as a key tool in Chapters 6 and 7. Further, in Chapter 9, we observe that an invertible substitution can be expressed as a finite *composition* of *indecomposable* substitutions, and so our attention turns to the latter.

FUTURE DIRECTIONS

There has recently been a surge in the number of papers on powers of words occurring in Sturmian words; for instance see [8, 13, 18, 28, 32, 37]. Of particular interest are the notions of *index* (or *critical exponent*), fractional and integer powers, and the supremum limit of powers of increasingly longer factors in a given sequence. In Chapter 7, we explicitly determine all integer powers occurring in a restricted class of standard episturmian words whose directive words resemble those of characteristic Sturmian words. It is an open problem to do the same for *all* episturmian words. Furthermore, it would also be interesting to generalize other known results on powers in Sturmian words to episturmian words, by expanding on the work in Chapter 7, and perhaps searching for further uses of the generalized singular words that we define there.

Much of the earlier work in this thesis concerns palindromic factors of characteristic

Sturmian words. In view of the current wide-spread interest in palindromes, further investigation of palindromes in episturmian words is merited. In particular, can one describe precisely where palindromes occur in a given episturmian word? We have done this for Sturmian words in Chapter 5.

In another direction, it remains an open problem to elucidate the structure of invertible substitutions on an arbitrary finite alphabet. Proving Conjecture 9.1 in Chapter 9, which gives a characterization of indecomposable (invertible) substitutions, will shed some light on this difficult and important problem.

REFERENCES

[1]   B. Adamczewski and Y. Bugeaud, 'On the complexity of algebraic numbers, II. Continued fractions', *Acta Math.* **195** (2005), 1–20.

[2]   J.-P. Allouche, J.L. Davison, M. Queffélec and L.Q. Zamboni, 'Transcendence of Sturmian or morphic continued fractions', *J. Number Theory* **91** (2001), 39–66.

[3]   J.-P. Allouche, M. Baake, J. Cassaigne and D. Damanik, 'Palindrome complexity', *Theoret. Comput. Sci.* **292** (2003), 9–31.

[4]   J.-P. Allouche and J. Shallit, *Automatic sequences: theory, applications, generalizations* (Cambridge University Press, Cambridge U.K., 2003).

[5]   P. Arnoux and G. Rauzy, 'Représentation géométrique de suites de complexité $2n + 1$', *Bull. Soc. Math. France* **119** (1991), 199–215.

[6]   J. Berstel, 'Axel Thue's work on repetitions in words', in *Proceedings of the 4th International Conference on Formal Power Series and Algebraic Combinatorics, Montréal* **11**, Publications du LaCIM (Université du Québec à, Montréal, 1992), **pp. 65–80**.

[7]   J. Berstel, 'Recent results on Sturmian words', in *Developments in Language Theory, Magdeburg 1995*, (J. Dassow, A. Salomaa, Editors) (World Scientific, Singapore, 1996).

[8]   J. Berstel, 'On the index of Sturmian words', in *Jewels Are Forever* (Springer-Verlag, Berlin, 1999), **pp. 287–294**.

[9]   J. Berstel, 'Recent results on extensions Sturmian words', *Internat. J. Algebra Comput.* **12** (2002), 371–385.

[10] J. Berstel and A. de Luca, 'Sturmian words, Lyndon words and trees', *Theoret. Comput. Sci.* **178** (1997), 171–203.

[11] J. Berstel and P. Séébold, 'A characterization of Sturmian morphisms', in *Mathematical Foundations of Computer Science 1993*, (A.M. Borzyszkowski, S. Sokolowski, Editors), Lecture Notes in Computer Science **711** (Springer-Verlag, Berlin, 1993), **pp. 281–290**.

[12] J. Berstel and P. Séébold, 'Sturmian words', in *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications **90** (Cambridge University Press, Cambridge U.K., 2002), **pp. 45–110**.

[13] V. Berthé, C. Holton and L.Q. Zamboni, 'Initial powers of Sturmian words', *Acta Arith.* (to appear).

[14] J.-P. Borel and C. Reutenauer, 'Palindromic factors of billiard words', *Theoret. Comput. Sci.* **340** (2005), 334–348.

[15] S. Brlek, S. Hamel, M. Nivat and C. Reutenauer, 'On the palindromic complexity of infinite words', *Internat. J. Found. Comput. Sci.* **15** (2004), 293–306.

[16] T.C. Brown, 'Descriptions of the characteristic sequence of an irrational', *Canad. Math. Bull.* **36** (1993), 15–21.

[17] W.-F. Chuan, 'Symmetric Fibonacci words', *Fibonacci Quart.* **31** (1993), 251–255.

[18] D. Damanik and D. Lenz, 'The index of Sturmian sequences', *European J. Combin.* **23** (2002), 23–29.

[19] A. de Luca, 'A combinatorial property of the Fibonacci words', *Inform. Process. Lett.* **12** (1981), 193–195.

[20] A. de Luca, 'A division property of the Fibonacci word', *Inform. Process. Lett.* **54** (1995), 307–312.

[21] A. de Luca and F. Mignosi, 'Some combinatorial properties of Sturmian words', *Theoret. Comput. Sci.* **136** (1994), 361–385.

[22] X. Droubay, J. Justin and G. Pirillo, 'Episturmian words and some constructions of de Luca and Rauzy', *Theoret. Comput. Sci.* **255** (2001), 539–553.

[23] X. Droubay and G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.* **223** (1999), 73–85.

[24] A. Glen, 'Conjugates of characteristic Sturmian words generated by morphisms', *European J. Combin.* **25** (2004), 1025–1037.

[25] A. Glen, 'Powers in a class of $\mathcal{A}$-strict standard episturmian words', in *Words* 2005, 5th *International Conference on Word*, Publications du LaCIM **36** (Université du Québec à Montréal, 2005), pp. 249–263.

[26] A. Glen, 'Occurrences of palindromes in characteristic Sturmian words', *Theoret. Comput. Sci.* **352** (2006), 31–46.

[27] J. Justin, 'Episturmian morphisms and a Galois theorem on continued fractions', *Theor. Inform. Appl.* **39** (2005), 207–215.

[28] J. Justin, G. Pirillo, 'Fractional powers in Sturmian words', *Theoret. Comput. Sci.* **255** (2001), 363–376.

[29] J. Justin and G. Pirillo, 'Episturmian words and episturmian morphisms', *Theoret. Comput. Sci.* **276** (2002), 281–313.

[30] J. Justin and G. Pirillo, 'Episturmian words: shifts, morphisms and numeration systems', *Internat. J. Found. Comput. Sci.* **15** (2004), 329–348.

[31] F. Levé and P. Séébold, 'Conjugation of standard morphisms and a generalization of singular words', *Bull. Belg. Math. Soc. Simon Stevin* **10** (2003), 737–747.

[32] F. Mignosi and G. Pirillo, 'Repetitions in the Fibonacci infinite word', *Theor. Inform. Appl.* **26** (1992), 199–204.

[33] M. Morse and G.A. Hedlund, 'Symbolic dynamics II: Sturmian trajectories', *Amer. J. Math.* **62** (1940), 1–42.

[34] G. Pirillo, 'Inequalities characterizing standard Sturmian and episturmian words', *Theoret. Comput. Sci.* **341** (2005), 276–292.

[35] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lecture Notes in Mathematics **1794** (Springer-Verlag, Berlin, 2002).

[36] G. Rauzy, 'Nombres algébriques et substitutions', *Bull. Soc. Math. France* **110** (1982), 147–178.

[37] D. Vandeth, 'Sturmian words and words with a critical exponent', *Theoret. Comput. Sci.* **242** (2000), 283–300.

[38]   B.A. Venkov, *Elementary number theory* (Wolters-Noordhoff, Netherlands, 1970).

[39]   Z.-X. Wen and Z.-Y. Wen, 'Some properties of the singular words of the Fibonacci word', *European J. Combin.* **15** (1994), 587–598.

Discipline of Pure Mathematics
School of Mathematical Sciences
The University of Adelaide
South Australia, 5005
Australia
e-mail: amy.glen@gmail.com