FORAYS BEYOND DENDRICITY

HERMAN GOULET-OUELLET

In their landmark paper published in 1991 [1], Arnoux and Rauzy proposed a generalization of Sturmian words to larger alphabets now known as *Arnoux–Rauzy words*. This expanded on earlier work of Rauzy [10] in which he defined a numeration system based on the tribonacci word, analogous to the Zeckendorff numeration; more generally, Arnoux–Rauzy words are related to Ostrowski numeration systems (see [2]). Recall that an infinite word on n letters ($n \geq 2$) is Arnoux–Rauzy if it is recurrent and it admits, at every length k, a unique left special factor and a unique right special factor (not necessarily distinct) having exactly n left and right extensions respectively. Arnoux and Rauzy showed that, similar to the Sturmian case, these infinite words admit many different interpretations: they are, at the same time, abstract continued fraction algorithms defined via S-adic representations; infinite paths in the tree of *standard tuples*; or symbolic encodings of trajectories in partitions of certain dynamical systems.

A key element in Arnoux and Rauzy's original work was their description of the evolution of the *Rauzy graphs*, certain subgraphs of de Bruijn graphs determined by the factors of an infinite word. In an Arnoux–Rauzy word, the Rauzy graph at rank $n+1$ is obtained from the graph at rank n using either a "*fente*" or an "*éclatement*", illustrated in Figure 1.

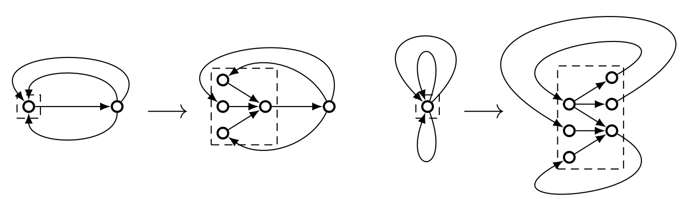

FIGURE 1. "Fentes" and "éclatements" described by Arnoux and Rauzy.

A generalization of Arnoux–Rauzy words was proposed in 2015 by Berthé et al. [3] called *dendric words*. These are infinite words defined by the *tree condition*, which involves the *extension graphs*: simple undirected graphs constructed using extensions of factors in the infinite word. An infinite word is said to be dendric when all of its extension graphs are trees. Take for instance the unique fixed point $\mathbf{d} = \mathtt{0210210100210}\cdots$ of the primitive substitution $\mathtt{0} \mapsto \mathtt{0210}, \mathtt{1} \mapsto \mathtt{10}, \mathtt{2} \mapsto \mathtt{210}$. It is a ternary dendric word which is not Arnoux–Rauzy; some of its extension graphs are depicted in Figure 2.

In addition to Arnoux–Rauzy words, dendric words include codings of regular interval exchanges [4] and many other examples in-between (like the dendric word $\mathbf{d}$ mentioned above). They agree with Sturmian words in the binary case, and in the

This work was supported by the ANR via the "Codys" project (ANR-18-CE40-0007).

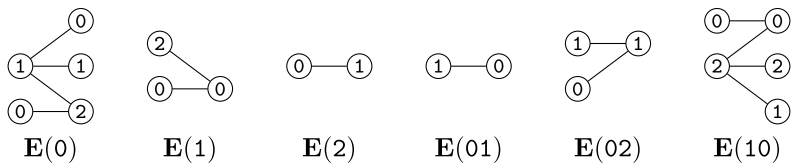

FIGURE 2. Some extension graphs of the dendric word $\mathbf{d} = 0210210100210 \cdots$.

ternary case, a full S-adic characterization was recently obtained by Gheeraert et al. [7]; the general case remains broadly misunderstood. Nonetheless, dendric words possess many remarkable combinatorial and dynamical properties: stability under taking *derived words* and some form of *rigidity* [5]; exact linear growth in factor complexity [3]; and stability under complete bifix decoding [6].

Dendric words also show remarkable regularity from an *algebraic* standpoint. This is embodied by the *Return Theorem* of Berthé et al. [3]. Recall that, given an infinite word $\mathbf{x} \in A^{\mathbb{N}}$ and a finite factor u of $\mathbf{x}$, a *return word* to u in $\mathbf{x}$ is a word r such that ru is a factor of $\mathbf{x}$ starting with u and having exactly two occurrences of u. By a *return set* of $\mathbf{x}$, we mean the set of all return words to a given finite factor u of $\mathbf{x}$. Let also F_n be the free group over n letters.

Theorem 1 (Return Theorem)**.** *Let $\mathbf{x}$ be a uniformly recurrent dendric word over n letters. Every return set of $\mathbf{x}$ is a basis of F_n.*

The proof of the Return Theorem is interesting in its own right, and circles back to Arnoux and Rauzy's original approach: it relies on the idea that the Rauzy graphs evolve following a series of "fentes" and "éclatements", where each "fente" corresponds to a left special factor which is not right special and each "éclatement" corresponds to a bispecial factor. The dendric case is a *strict* generalization of the Arnoux–Rauzy case because (1) several operations can occur in parallel (i.e. at the same rank) and (2) the "éclatements" may vary depending on the shapes of the extension graphs; see Figure 3. Despite these new complications, the "éclatements" occurring in dendric words still enjoy a desirable property which translates to the Return Theorem: the converse operations—let us say "*écroulements*"—can be written as a sequence of elementary foldings in the sense of Stallings [11].

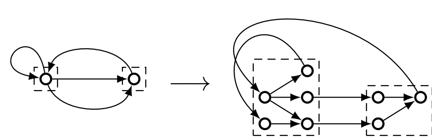

FIGURE 3. Example of parallel "fente" and "éclatement" in the dendric word $\mathbf{d} = 0210210100210 \cdots$.

The aim of this talk is to explore a generalization of dendricity defined in terms of generalized extension graphs, which we call *suffix-connectedness.* This condition still guarantees that the "éclatements" occurring in the Rauzy graphs are reasonably well-behaved, in the sense that the corresponding "écroulements" are, once again, a sequence of elementary foldings. Translated in an algebraic language, this gives the following generalization of the Return Theorem [8].

Theorem 2. *Let* $\mathbf{x}$ *be a uniformly recurrent suffix-connected word over* n *letters. The subgroups of* F_n *generated by the return sets of* $\mathbf{x}$ *all lie in the same conjugacy class and their rank is* $n - m + 1$*, where* m *is the number of connected components of the extension graph of the empty word.*

To illustrate the suffix-connectedness condition, we present a non-dendric suffix-connected example, the fixed point $\mathbf{s} = \mathtt{0302303012223}\cdots$ of the substitution on 4 letters $\mathtt{0} \mapsto \mathtt{030}, \mathtt{1} \mapsto \mathtt{230}, \mathtt{2} \mapsto \mathtt{122}, \mathtt{3} \mapsto \mathtt{23}$. This example features multiple parallel "éclatements" with shapes impossible to encounter in the dendric case, as depicted in Figure 4. The descriptions of the corresponding "écroulements" in terms of elementary foldings are accordingly more complicated.

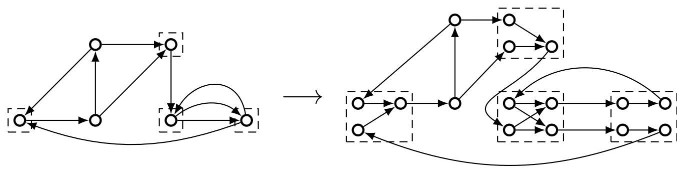

FIGURE 4. "Fentes" and "éclatements" with non-dendric shapes in the suffix-connected word $\mathbf{s} = \mathtt{0302303012223}\cdots$.

In this example, the return sets offer more variety: their cardinalities oscillate between different values, and they generate two different (but conjugate) subgroups of rank 3 in F_4. We use this example as an opportunity to present some of the tools and ideas used to establish suffix-connectedness for fixed points of primitive substitutions, including an algorithm for computing bispecial factors due to Klouda [9].

REFERENCES

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

[2] V. Berthé. Autour du système de numération d'Ostrowski. *Bull. Belg. Math. Soc. Simon Stevin*, 8(2):209–239, 2001.

[3] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176(4):521–550, 2015.

[4] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Bifix codes and interval exchanges. *J. Pure Appl. Algebra*, 219(7):2781–2798, 2015.

[5] V. Berthé, F. Dolce, F. Durand, J. Leroy, and D. Perrin. Rigidity and substitutive dendric words. *Int. J. Found. Comput.*, 29(05):705–720, 2018.

[6] F. Dolce and D. Perrin. Eventually dendric shift spaces. *Ergod. Theory Dyn. Syst.*, 41(7):2023–2048, 2020.

[7] F. Gheeraert, M. Lejeune, and J. Leroy. $\mathcal{S}$-adic characterization of minimal ternary dendric shifts. *Ergod. Theory Dyn. Syst.*, 42(11):3393–3432, 2022.

[8] H. Goulet-Ouellet. Suffix-connected languages. *Theoret. Comput. Sci.*, 923:126–143, 2022.

[9] K. Klouda. Bispecial factors in circular non-pushy D0L languages. *Theoret. Comput. Sci.*, 445:63–74, 2012.

[10] G. Rauzy. Nombres algébriques et substitutions. *Bull. Soc. Math. Fr.*, 79:147–178, 1982.

[11] J. R. Stallings. Topology of finite graphs. *Invent. Math.*, 71:551–565, 1983.

IRIF, UNIVERSITÉ PARIS CITÉ, 75013 PARIS, FRANCE
Email address: hgoulet@irif.fr