Binomial complexities and Parikh-collinear morphisms

Joint work Michel Rigo and Markus A. Whiteland (ULiège)

Manon Stipulanti (ULiège) FNRS Postdoc Researcher

DLT 2022
University of South Florida (Tampa, USA)
May 9th, 2022

Notation

- For a finite alphabet A, A^* = the set of finite words over A.
- Infinite words are written in **bold**.
- $|w|$ = the length of the word w = # of letters in w.
- $|w|_a$ = # of occurrences of the letter a in w.
- In a word,
  A factor is a subsequence made of consecutive letters;
  A (scattered) subword is simply a subsequence made of letters.

Example: $|\texttt{reappear}| = 8$, $|\texttt{reappear}|_{\texttt{a}} = 2 = |\texttt{reappear}|_{\texttt{e}}$

reappear	factor	subword
	re**appear**	re**ap**p**e**ar
	re**app**ear	**r**e**ap**pear
	reapp**ear**	**re**appe**ar**

- $\mathrm{Fac}_n(\mathbf{x})$ = the set of length-n factors of an infinite word $\mathbf{x}$, $n \geq 0$.

In CoW, the combinatorial structure of infinite words is sometimes studied by means of their set of factors.

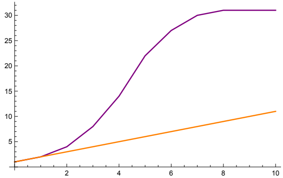
For instance, the factor complexity $\mathsf{p}_{\mathbf{x}} \colon \mathbb{N} \to \mathbb{N}, n \mapsto \#\mathrm{Fac}_n(\mathbf{x})$ gives structural information on $\mathbf{x}$.

Theorem (Morse, Hedlund, 1938)

$\mathbf{x}$ is ultimately periodic iff $\mathsf{p}_{\mathbf{x}}$ is bounded.

$\mathbf{x}$ is Sturmian iff $\mathsf{p}_{\mathbf{x}}(n) = n + 1 \ \forall n \geq 0$.

(Sturmian=binary, aperiodic, minimal factor complexity)



Variation: Count not all but "different enough" factors, measured by equivalence relations.

Main ingredients: binomial coefficients

Definition (Lothaire, 1997)

The binomial coefficient $\binom{u}{v}$ of $u, v \in A^*$ is the number of times v occurs as a (scattered) subword of u.

Example: $\binom{101001}{101} = 6$

101001 101001 101001
101001 101001 101001

Remark: Generalization of usual binomial coefficients of integers as $\binom{a^n}{a^k} = \binom{n}{k}$.

One "famous" motivation in CoW:

The reconstruction problem

Given an integer n, what is the smallest integer k such that each length-n word is uniquely determined by the set of all its length-k subwords (with multiplicities)?

Still open but bounds have been obtained, and several variations considered.

Main ingredients: binomial equivalence relations

Let $k \geq 1$ be an integer.
Two words $u, v \in A^*$ are k-binomially equivalent if $\binom{u}{x} = \binom{v}{x}$ for all $x \in A^*$ with $|x| \leq k$.
We write $u \sim_k v$.

Example: $0110 \sim_2 1001$ but $0110 \not\sim_3 1001$

x	0	1	00	01	10	11	000	001
$\binom{0110}{x}$	2	2	1	2	2	1	0	0
$\binom{1001}{x}$	2	2	1	2	2	1	0	1

Remark:
$\sim_k$ is a generalization* of abelian equivalence since $u \sim_{ab} v$ iff $u \sim_1 v$.
($u, v \in A^*$ are abelian equivalent if they are permutations of each other.)

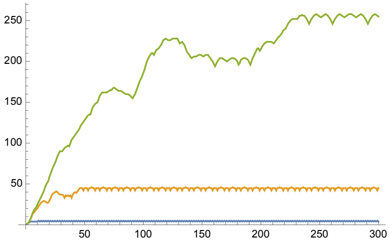
*Note also that other generalizations of abelian equivalence exist in the literature.

Definition (Rigo, Salimov, 2015)

Let $k \geq 1$. The k-binomial complexity function of $\mathbf{x}$ is the map $\mathsf{b}_{\mathbf{x}}^{(k)} : \mathbb{N} \to \mathbb{N},\ n \mapsto \#(\mathrm{Fac}_n(\mathbf{x})/\!\sim_k)$.

Observation: For all $k \geq 1$, $u \sim_{k+1} v \implies u \sim_k v$, so

$$\mathsf{b}_{\mathbf{x}}^{(1)}(n) \leq \mathsf{b}_{\mathbf{x}}^{(2)}(n) \leq \cdots \leq \mathsf{b}_{\mathbf{x}}^{(k)}(n) \leq \mathsf{b}_{\mathbf{x}}^{(k+1)}(n) \leq \cdots \leq \mathsf{p}_{\mathbf{x}}(n) \quad \forall n \in \mathbb{N}.$$



Motivation and goal

	abelian complexity	k-binomial complexity
theory (general behavior, properties, etc.)	rich	not much is known

What is desirable:

- Compute the k-binomial complexity of a given infinite word (but it is sometimes challenging).
- Compare, in some ways, the k- and $(k+1)$-binomial complexities of a given word.
- Understand the k-binomial complexity of large classes of words (automatic words, morphic words, Sturmian words, episturmian words, etc.).

In this talk

- $\nearrow$ theory on binomial complexities and,
- 1 question (or 2) related to them.

A little bit of history

Since their introduction in 2015 by Rigo and Salimov, several results on binomial complexities have been obtained:

(type of) word	property	reference
Sturmian word $\mathbf{s}$	$\mathsf{b}_{\mathbf{s}}^{(k)} = \mathsf{p}_{\mathbf{s}} \ \forall\, k \geq 2$	Rigo, Salimov, 2015
Tribonacci word $\mathbf{z}$ (fixed point of $0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 0$)	$\mathsf{b}_{\mathbf{z}}^{(k)} = \mathsf{p}_{\mathbf{z}} \ \forall\, k \geq 2$	Lejeune, Rosenfeld, Rigo, 2020 Lejeune's PhD thesis, 2021
Thue–Morse word $\mathbf{t}$ (fixed point of $0 \mapsto 01, 1 \mapsto 10$) + wider class of words (described later on)	$\mathsf{b}_{\mathbf{t}}^{(k)}$ bounded by a constant depending on k	Rigo, Salimov, 2015
Thue–Morse word $\mathbf{t}$	precise values of $\mathsf{b}_{\mathbf{t}}^{(k)} \ \forall\, k \geq 1$	Lejeune, Leroy, Rigo, 2020 Lejeune's PhD thesis, 2021
Generalized Thue–Morse words	precise values of $\mathsf{b}^{(2)}$	Lü, Chen, Wen, Wu, 2021

Developing the theory a little further

The binomial complexities of the Thue–Morse word $\mathbf{t}$ are bounded.

Theorem (Lejeune, Leroy, Rigo, 2020)

$$\text{For all } k \geq 1, \quad \mathsf{b}_{\mathbf{t}}^{(k)}(n) = \begin{cases} \mathsf{p}_{\mathbf{t}}(n), & \text{if } n \leq 2^k - 1; \\ 3 \cdot 2^k - 3, & \text{if } n \equiv 0 \pmod{2^k} \text{ and } n \geq 2^k; \\ 3 \cdot 2^k - 4, & \text{otherwise.} \end{cases}$$

$\mathbf{t}$ is the fixed point of $\varphi\colon 0 \mapsto 01, 1 \mapsto 10$ for which

$$\begin{pmatrix} |\varphi(0)|_0 \\ |\varphi(0)|_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} |\varphi(1)|_0 \\ |\varphi(1)|_1 \end{pmatrix}.$$

We say that φ is Parikh-constant.

Definition: Let $A = \{a_1, \ldots, a_k\}$ with the order $a_1 < a_2 < \cdots < a_k$. The Parikh vector of a word $w \in A^*$ is

$$\Psi(w) = \begin{pmatrix} |w|_{a_1} & |w|_{a_2} & \cdots & |w|_{a_k} \end{pmatrix}^\top.$$

A morphism $f\colon A^* \to B^*$ is Parikh-constant if $\Psi(f(a)) = \Psi(f(b))$ for all letters $a, b \in A$.

Theorem (Rigo, Salimov, 2015)

A fixed point of a Parikh-constant morphism has bounded k-binomial complexity for all $k \geq 1$.

Question: Is there a larger class of words having bounded binomial complexities?

A morphism $f\colon A^* \to B^*$ is Parikh-collinear if, for all letters $a, b \in A$, $\Psi(f(b)) = r_{a,b}\Psi(f(a))$ for some rational number $r_{a,b}$.

Example: $f\colon 0 \mapsto 000111, 1 \mapsto 0110$ is Parikh-collinear

$$\Psi(f(0)) = \begin{pmatrix} |f(0)|_0 \\ |f(0)|_1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \qquad \Psi(f(1)) = \frac{2}{3}\Psi(f(0)).$$
$$\Psi(f(1)) = \begin{pmatrix} |f(1)|_0 \\ |f(1)|_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

Remark: A morphism is Parikh-collinear iff its adjacency matrix has rank 1.

Characterization in terms of abelian complexity:

Theorem (Cassaigne, Richomme, Saari, Zamboni, 2011)

A morphism is Parikh-collinear iff it maps all infinite words to words with bounded abelian complexity.

New characterization in terms of binomial complexities:

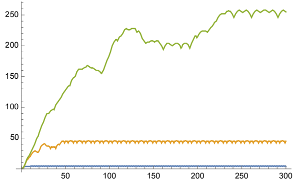
Theorem

A morphism is Parikh-collinear iff it maps, for all $k \geq 0$, all words with bounded k-bin. complexity to words with bounded $(k+1)$-bin. complexity.

Generalization of Rigo and Salimov:

Corollary

A fixed point of a Parikh-collinear morphism has bounded k-binomial complexity for all $k \geq 1$.



$f : 0 \mapsto 000111, 1 \mapsto 0110$

Question B

Binomial complexities are increasingly nested:

$$\mathsf{b}_{\mathbf{x}}^{(1)}(n) \le \mathsf{b}_{\mathbf{x}}^{(2)}(n) \le \cdots \le \mathsf{b}_{\mathbf{x}}^{(k)}(n) \le \mathsf{b}_{\mathbf{x}}^{(k+1)}(n) \le \cdots \le \mathsf{p}_{\mathbf{x}}(n) \quad \forall n \in \mathbb{N}\,.$$

So is it possible that the factor complexity coincides with some binomial complexity?

<u>Notation</u>: For two functions $\mathsf{f}, \mathsf{g} \colon \mathbb{N} \to \mathbb{N}$, we write $\mathsf{f} \prec \mathsf{g}$ when

- $\mathsf{f}(n) \le \mathsf{g}(n)$ for all $n \in \mathbb{N}$ and
- $\mathsf{f}(n) < \mathsf{g}(n)$ for infinitely many $n \in \mathbb{N}$.

Question B (Stabilization)

For each $k \ge 1$, does there exist a word $\mathbf{w}_k$ such that

$$\mathsf{b}_{\mathbf{w}_k}^{(1)} \prec \mathsf{b}_{\mathbf{w}_k}^{(2)} \prec \cdots \prec \mathsf{b}_{\mathbf{w}_k}^{(k-1)} \prec \mathsf{b}_{\mathbf{w}_k}^{(k)} = \mathsf{p}_{\mathbf{w}_k}?$$

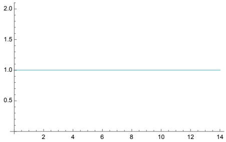
This question was inspired by Lejeune's PhD thesis.

First (naive) answer

Periodic words answer Question B.

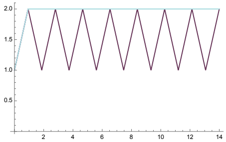
Examples:

$\mathbf{w} = 0\,0\,0 \cdots$



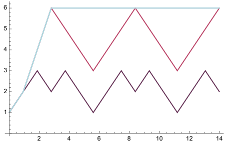
$\mathsf{b}_{\mathbf{w}}^{(1)} = \mathsf{p}_{\mathbf{w}}$

$\mathbf{w} = 01\,01\,01 \cdots$



$\mathsf{b}_{\mathbf{w}}^{(1)} \prec \mathsf{b}_{\mathbf{w}}^{(2)} = \mathsf{p}_{\mathbf{w}}$

$\mathbf{w} = 011001\,011001 \cdots$



$\mathsf{b}_{\mathbf{w}}^{(1)} \prec \mathsf{b}_{\mathbf{w}}^{(2)} \prec \mathsf{b}_{\mathbf{w}}^{(3)} = \mathsf{p}_{\mathbf{w}}$

But these words have

- bounded complexities and
- a rather simple structure.

Let's propose more "interesting" words*.

*No offense, I personally love periodic words.

Towards a second answer

For the Thue–Morse word $\mathbf{t}$, fixed point of $\varphi : 0 \mapsto 01, 1 \mapsto 10$:

Theorem (Lejeune, Leroy, Rigo, 2020)

$$\text{For all } k \geq 1, \quad \mathsf{b}_{\mathbf{t}}^{(k)}(n) = \begin{cases} \mathsf{p}_{\mathbf{t}}(n), & \text{if } n \leq 2^k - 1; \\ 3 \cdot 2^k - 3, & \text{if } n \equiv 0 \pmod{2^k} \text{ and } n \geq 2^k; \\ 3 \cdot 2^k - 4, & \text{otherwise.} \end{cases}$$

Behavior not specific to $\mathbf{t}$, but appearing for a large class of words:

Theorem

Let j, k be integers with $1 \leq j \leq k$.
Let $\mathbf{y}$ be an aperiodic binary word and let $\mathbf{x} = \varphi^k(\mathbf{y})$.
Then

$$\mathsf{b}_{\mathbf{x}}^{(j)}(n) = \begin{cases} \mathsf{p}_{\mathbf{x}}(n), & \text{if } n \leq 2^j - 1; \\ \mathsf{b}_{\mathbf{t}}^{(j)}(n), & \text{otherwise.} \end{cases}$$

The proof is quite technical and requires concepts from Lejeune, Leroy and Rigo's paper.

Corollary

Let k be an integer.
Let $\mathbf{y}$ be an aperiodic binary word and let $\mathbf{x} = \varphi^k(\mathbf{y})$.
We have $\mathsf{b}_{\mathbf{x}}^{(1)} \prec \mathsf{b}_{\mathbf{x}}^{(2)} \prec \ldots \prec \mathsf{b}_{\mathbf{x}}^{(k)} \prec \mathsf{b}_{\mathbf{x}}^{(k+1)}$.

Proof: By the previous theorem, $\mathsf{b}_{\mathbf{x}}^{(1)} \prec \mathsf{b}_{\mathbf{x}}^{(2)} \prec \ldots \prec \mathsf{b}_{\mathbf{x}}^{(k)}$.
Showing $\mathsf{b}_{\mathbf{x}}^{(k)} \prec \mathsf{b}_{\mathbf{x}}^{(k+1)}$ requires more work (and involves concepts and techniques from Lejeune, Leroy and Rigo's paper).

Answer to Question B:

Theorem

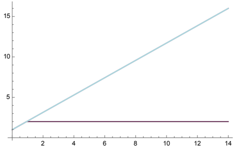
Let k be an integer.
Let $\mathbf{s}$ be a Sturmian word and let $\mathbf{s}_k = \varphi^k(\mathbf{s})$.
Then $\mathsf{b}_{\mathbf{s}_k}^{(1)} \prec \mathsf{b}_{\mathbf{s}_k}^{(2)} \prec \cdots \prec \mathsf{b}_{\mathbf{s}_k}^{(k+1)} \prec \mathsf{b}_{\mathbf{s}_k}^{(k+2)} = \mathsf{p}_{\mathbf{s}_k}$.

Proof: By the previous corollary, $\mathsf{b}_{\mathbf{s}_k}^{(1)} \prec \mathsf{b}_{\mathbf{s}_k}^{(2)} \prec \cdots \prec \mathsf{b}_{\mathbf{s}_k}^{(k+1)}$.
Showing $\mathsf{b}_{\mathbf{s}_k}^{(k+2)} = \mathsf{p}_{\mathbf{s}_k}$ requires more work.

Example: Let $\mathbf{s} = 010010100100\cdots$ be the Fibonacci word

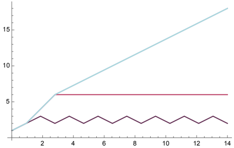
(fixed point of $0 \mapsto 01, 1 \mapsto 0$).

$\mathbf{s} = \varphi^0(\mathbf{s})$



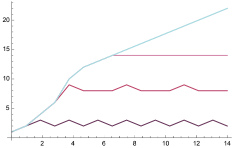
$\mathsf{b}^{(2)} = \mathsf{p}$

$\varphi^1(\mathbf{s})$



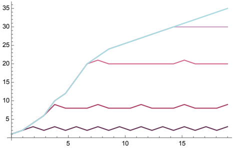
$\mathsf{b}^{(3)} = \mathsf{p}$

$\varphi^2(\mathbf{s})$



$\mathsf{b}^{(4)} = \mathsf{p}$

$\varphi^3(\mathbf{s})$



$\mathsf{b}^{(5)} = \mathsf{p}$

True for all Sturmian words.

(Rigo, Salimov, 2015)

But... the binomial complexities $\mathsf{b}_{\mathbf{s}_k}^{(1)}, \ldots, \mathsf{b}_{\mathbf{s}_k}^{(k+1)}$ are bounded.

Question C

For each $k \geq 1$, does there exist a word $\mathbf{w}_k$ such that $\mathsf{b}^{(1)}_{\mathbf{w}_k}$ is unbounded and

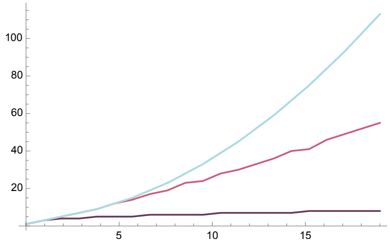
$$\mathsf{b}^{(1)}_{\mathbf{w}_k} \prec \mathsf{b}^{(2)}_{\mathbf{w}_k} \prec \cdots \prec \mathsf{b}^{(k-1)}_{\mathbf{w}_k} \prec \mathsf{b}^{(k)}_{\mathbf{w}_k} = \mathsf{p}_{\mathbf{w}_k}?$$

Answer for $k = 3$:

$\mathbf{h} = 011212212221222212222\cdots$

(fixed point of $0 \mapsto 01, 1 \mapsto 12, 2 \mapsto 2$)

- has $\mathsf{b}^{(1)}_{\mathbf{h}}$ unbounded
- has $\mathsf{b}^{(1)}_{\mathbf{h}} \prec \mathsf{b}^{(2)}_{\mathbf{h}} \prec \mathsf{b}^{(3)}_{\mathbf{h}} = \mathsf{p}_{\mathbf{h}}$.



What about larger values of k? The question remains open...

If time permits: Question A

Binomial complexities are increasingly nested:

$$\mathsf{b}_{\mathbf{x}}^{(1)}(n) \leq \mathsf{b}_{\mathbf{x}}^{(2)}(n) \leq \cdots \leq \mathsf{b}_{\mathbf{x}}^{(k)}(n) \leq \mathsf{b}_{\mathbf{x}}^{(k+1)}(n) \leq \cdots \leq \mathsf{p}_{\mathbf{x}}(n) \quad \forall n \in \mathbb{N}\,.$$

So between two consecutive functions, could the non-equality happen infinitely many times?

Question A

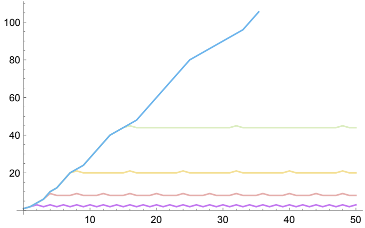
Does there exist an infinite word $\mathbf{w}$ such that $\mathsf{b}_{\mathbf{w}}^{(k)} \prec \mathsf{b}_{\mathbf{w}}^{(k+1)}$ for all $k \geq 1$?

This question was inspired by Lejeune's PhD thesis.

The Thue–Morse word $\mathbf{t}$

- answers Question A
- has $\mathsf{b}_{\mathbf{t}}^{(k)}$ bounded

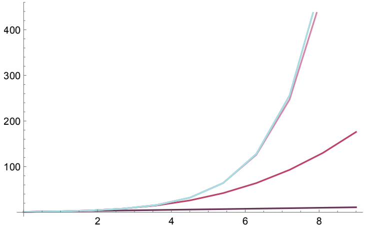
(Lejeune, Leroy, Rigo, 2015)



The binary Champernowne word $\mathbf{c} = 0\,1\,10\,11\,100\,101\,110\,111\cdots$

(the concatenation of all binary representations)

- answers Question A
- has $\mathsf{b}_{\mathbf{c}}^{(1)}$ unbounded
- is not morphic nor uniformly recurrent



A festival of answers: more "structured" words

The word $\mathbf{v} = \tau(g^\omega(a))$, where

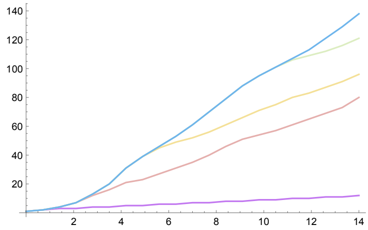
$$g \colon a \mapsto a0\alpha, 0 \mapsto 01, 1 \mapsto 10, \alpha \mapsto \alpha^2$$

$$g^\omega(a) := \lim_{n\to\infty} g^n(a)$$

$$\tau \colon a \mapsto \varepsilon, 0 \mapsto 0, 1 \mapsto 1, \alpha \mapsto 1,$$

- answers Question A
- has $\mathsf{b}_{\mathbf{v}}^{(1)}$ unbounded
- is binary and morphic

(morphic: the image, under a coding, of a fixed point of a morphism)



Grillenberger's construction gives a word $\mathbf{w} = 0100010101100111\cdots$ which

- answers Question A
- has $\mathsf{b}_{\mathbf{w}}^{(1)}$ unbounded
- is binary and uniformly recurrent

(uniformly recurrent: every factor occurs infinitely many times within bounded gaps).

Construction (Grillenberger, 1973)

Start with $D_0 = \{0, 1\}$.
If D_n is constructed ($n \geq 0$), let w_n be the concatenation of words in D_n in lexicographic order, assuming $0 < 1$.
Then set $D_{n+1} := w_n D_n^2$.
The word $\mathbf{w} = \lim_{n\to\infty} w_n$ is uniformly recurrent.

References

- Allouche, J.P., Shallit, J.: Automatic sequences: Theory, applications, generalizations. Cambridge University Press, Cambridge (2003).
- Berthé, V., Karhumäki, J., Nowotka, D., Shallit, J.: Mini-workshop: combinatorics on words. Oberwolfach Rep. 7, 2195-2244 (2010).
- Cassaigne, J., Fici, G., Sciortino, M., Zamboni, L.Q.: Cyclic complexity of words. J. Comb. Theory, Ser. A **145**, 36–56 (2017).
- Cassaigne, J., Richomme, G., Saari, K., Zamboni, L.Q.: Avoiding Abelian powers in binary words with bounded Abelian complexity. Int. J. Found. Comput. S. **22**(4), 905–920 (2011).
- Cheraghchi, M., Gabrys, R., Milenkovic, O., Ribeiro, J.: Coded trace reconstruction. IEEE Transactions on Information Theory **66**(10), 6084–6103 (2020).
- Dudík, M., Schulman, L.J.: Reconstruction from subsequences. J. Comb. Theory, Ser. A **103**(2), 337–348 (2003).
- Frid, A.: Applying a uniform marked morphism to a word. Discrete Math. Theor. Comput. Sci. **3**(3), 125–139 (1999).
- Grillenberger, C.: Constructions of strictly ergodic systems. I. Given entropy. Z. Wahrscheinlichkeit. **25**, 323–334 (1973).
- Karhumäki, J., Saarela, A., Zamboni, L.Q.: On a generalization of abelian equivalence and complexity of infinite words. J. Comb. Theory, Ser. A **120**(8), 2189–2206 (2013).
- Lejeune, M.: On the k-binomial equivalence of finite words and k-binomial complexity of infinite words. Ph.D. thesis, University of Liège (2021), `http://hdl.handle.net/2268/259266`.

References

- Lejeune, M., Leroy, J., Rigo, M.: Computing the k-binomial complexity of the Thue–Morse word. J. Comb. Theory, Ser. A **176**, 44 (2020).
- Lejeune, M., Rigo, M., Rosenfeld, M.: Templates for the k-binomial complexity of the Tribonacci word. Adv. Appl. Math. **112**, 26 (2020).
- Lothaire, M.: Combinatorics on Words. Cambridge Mathematical Library. Cambridge University Press (1997).
- Lü, X.T., Chen, J., Wen, Z.X., Wu, W.: On the 2-binomial complexity of the generalized Thue-Morse words (2021), preprint `https://arxiv.org/abs/2112.05347`.
- Ochsenschläger, P.: Binomialkoeffizienten und shuffle-zahlen. Technischer bericht, Fachbereicht Informatik, T.H. Darmstadt (1981).
- Peltomäki, J.: Introducing privileged words: Privileged complexity of Sturmian words. Theor. Comput. Sci. **500**, 57–67 (2013).
- Rigo, M.: Relations on words. Indag. Math., New Ser. **28**(1), 183–204 (2017).
- Rigo, M., Salimov, P.: Another generalization of abelian equivalence: binomial complexity of infinite words. Theor. Comput. Sci. **601**, 47–57 (2015).
- ⋆ Rigo, M., Stipulanti, M., Whiteland, M.A.: Binomial complexities and Parikh-collinear morphisms (2022), preprint `https://arxiv.org/abs/2201.04603`.
- Whiteland, M.A.: Equations over the k-binomial monoids. In: WORDS 2021, LNCS, vol. 12847, pp. 185–197. Springer, Cham. (2021).