HAL
open science

On the Contribution of WORDS to the Field of Combinatorics on Words

Jean Néraud

To cite this version:

Jean Néraud. On the Contribution of WORDS to the Field of Combinatorics on Words. Combinatorics on Words, pp.27-34, 2015. hal-02123623

HAL Id: hal-02123623

https://hal.science/hal-02123623v1

Submitted on 8 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Contribution of WORDS to the Field of Combinatorics on Words

Jean Néraud

Université de Rouen, France
* jean.neraud@univ-rouen.fr, neraud.jean@free.fr, neraud.jean@gmail.com
url: neraud.jean.free.fr

The representation of numbers by sequences of symbols is inherent in mathematics. A noticeable step has certainly been reached by the introduction of the decimal representation. This notion appears at the tenth century in the documents written by the arabe mathematician Al-Uqlidisi, which was interested in the indian system of numeration. In the occident, the fractional representation of numbers delayed the introduction of the decimal representation until the seventeenth century, when the belgian mathematician Simon Stevin recommends it as a performing tool of calculation. In his document "La Disme", he predicts that these methods of calculation will be extended to unrestricted representations and even be applied to the so-called incommensurable numbers.

A systematic study of words as formal mathematical objects appeared at the beginning of the twentieth century, when three now famous papers were published by the norvegian mathematician Axel Thue [17–19]. Presently, thanks to Jean Berstel [1] and James F. Power [16], we get translations of these papers in a more recent terminology and in relation with more contemporary directions of research.

Actually, from the end of the nineteenth century, words have taken an important role in different domains of mathematics such as Groups, Semigroups, Formal Languages, Number Theory, Ergodic Theory. In another hand, constituting a unified treatment of words was more and more in demand: such a request was especially stimulated by Max Paul Schützenberger, who presented a series of challenging questions concerning the topic in his lectures of 1966 .

In the growing importance of the field of Combinatorics on Words, a new fundamental step was reached in the eighties when were published the series of Lothaire's books [9, 10] and the famous "Theory of Codes" from Jean Berstel and Dominique Perrin [2] (cf also [3]). The importance of the topic has been also supported by the publication of the third book from Lothaire [11], which testifies that words have fundamental applications in many domains of computer science. Another important step was reached in 1991, when the terminology "Combinatorics on Words" has been introduced in the famous "Mathematics Subject Classification" as a subfield of "Discrete Mathematics in Relation to Computer Science" (68R15).

In view of sharing scientific results, international workshops play a complementary part beside the publication of books and full papers. They quickly

* litis, Université de Rouen, UFR Sciences et Techniques, Campus du Madrillet, 76800 Saint Etienne du Rouvray, France

provide a picture of the state-of-the-art and are special meeting places for the communauty. Actually, until the end of the nineties, due to the numerous varied topics in theoretical computer science, in most of the international conferences only a few sessions could be granted to Combinatorics on Words. As regard to the growing number of papers concerned by the topic, a specific international conference essentially devoted to words was strongly required.

The organization of "WORDS" was the response to such a request: with Aldo de Luca and Antonio Restivo, we drew the main features of the project, and opted in particular for a bi-annual series of meetings. The first conference WORDS was planned in Rouen, France, during September 1997. Thanks are due to the many researchers that supported that event and the subsequent conferences (they will recognize themselves); a special thought is due to Jean Berstel, for his invaluable investment in the project. A series of ten international workshops were organized:

WORDS 1997, Rouen, France, Jean Néraud chair,
WORDS 1999, Rouen, France, Jean Néraud chair,
WORDS 2001, Palermo, Italia, Felipo Mignosi chair,
WORDS 2003, Turku, Finland, Juhani Karhumäki chair,
WORDS 2005, Montréal, Canada, Srecko Brlek and Christophe Reutenauer co-chairs,
WORDS 2007, Marseille, France, Srecko Brlek and Julien Cassaigne co-chairs,
WORDS 2009, Salerno, Italia, Arturo Carpi and Clelia De Felice co-chairs,
WORDS 2011, Prague, Czech Republic, Štěpán Holub and Edita Pelandova co-chairs,
WORDS 2013, Turku, Finland, Juhani Karhumäki and Luca Zamboni co-chairs,
WORDS 2015, Kiel, Germany, Dirk Nowotka, chair.

For each meeting, papers selected to be presented are collected in written acts; moreover among them, some of the most stricking are published in a special issue of an international scientific revue [13–15, 6, 5, 4, 12, 8].

The tenth conference WORDS provides the opportunity to take a panoramic look at all the fascinating papers which were presented. Of course, drawing up an exhaustive list would be an impossible task: in what follows, only a limited number of results will be mentioned: according to the frequency of their presentations, we have opted for a classification into three thematic areas.

The topic of Unavoidable Patterns

The famous infinite word of The-Morse have the fundamental property that given a letter a, no factor of type $avava$ may appear in the word. It is also well-known that this word allows to construct an infinite cube-free word on a three letter alphabet. These properties have naturally opened a more general problem: given finite alphabet does an infinite word exists such that none of its factors may be a *repetition* of type $(uv)^k u$, with $u \neq \varepsilon$ and $k \geq 2$.

- *Avoidance of patterns* is a central question in the topic. In the meeting of 1997, Roman Kolpakov, Gregory Kucherov and Yuri Tarannikov present some properties of repetition-free binary words of minimal density. During that of 2003, James Currie draws an extensive survey concentrating on open problems. In the same workshop, Ina N. Rampersad, Jeffrey Shallit and Ming-Wei Wang are interested in infinite words avoiding long squares. In the conference of 2007, Arturo Carpi and Valerio D'Alonzo present a word avoiding near repeats. In their contribution to WORDS 2011, Elena A. Petrova and Arseny M. Shur will focuse on binary cube-free words that cannot be infinitely extended preserving cube-freeness: they prove that such words exist but can have arbitrarily long finite cube-free extensions both to one side and two sides. In WORDS 2013, Tero Harju presents an infinite square-free word on a three-letter alphabet that can be shuffled with itself to produce another infinite square-free word. In the same conference, Tomi Kärki gives an overview of the results concerning repetition-freeness in connection with the so-called *similarity relations*, which are relations on words of equal length induced by a symmetric and reflexive relation on letters.

- A natural question consists in examining the number of patterns that may appear in a finite word. In their contribution to WORDS 1997, the exact number of squares in the Fibonacci words is examined by Aviezri S. Fraenkel and Jamie Simpson: they prove that a word of length n contains at most $2n$ squares. The bound will be refined as $2n - O(\log n)$ by Lucian Illie in his talk of WORDS 2005.

- In their talk of WORDS 2009, Maxime Crochemore, Lucian Ilie and Liviu Tinta define *runs* as patterns of type $uu\alpha$ that are maximal in the sense where they cannot be extended from left or right to obtain the same type pattern. Such objects play an important role in a lot of string matching algorithms: the authors show that, given a word of length n, the number of its runs is up-bounded by $1.029n$.

- The *repetition threshold* for k letters, which we denote by $RT(k)$, is the shortest word x such that there exists an infinite x^+-free word over a k-letter alphabet. Actually, repetitions of the Thue-Morse sequence have exponent at most 2 and we have $RT(2) = 2$. In the seventies, Françoise Dejean conjectured that for every $k > 2$ the following holds:

$$RT(k) = \begin{cases} 7/4 & \text{if} \quad k = 3 \\ 7/5 & \text{if} \quad k = 4 \\ k/k-1 & \text{otherwise.} \end{cases}$$

A stronger version of this conjecture has been stated by Pascal Ochem in his contribution to WORDS 2005:

- For every $k \geq 5$, an infinite $(k/k-1)^+$-free word over k letters exists with letter frequency $1/k+1$
- For every $k \geq 6$,an infinite $(k/k-1)^+$-free word over k letters exists with letter frequency $1/k-1$.

Dejean's conjecture had been partially solved by several authors. The final proof was completed in 009 by James Currie and Narad Rampersad for $15 \leq n \leq 26$, and independently by Michaël Rao for $8 \leq k \leq 38$. In WORDS 2009, Michaël Rao will present his proof. Moreover the technic that he applies allow also to prove Ochem's stronger version of the conjecture for $9 \leq k \leq 38$.

- Abelian patterns are also concerned: an *abelian square* consists in a pattern of type xy, where the word y is obtained by applying a permutation on the letters of x. In 1992, Veikko Keränen had solved a famous open problem in constructing an *abelian square free* word over a four-letter alphabet. In the meeting of 2007, he will present new abelian square-free morphisms and a powerful substitution over 4 letters.

Two words u, v are *k-abelian equivalents* if every word of length at most k occurs as a factor in u as many times as in v. A word is *strongly k-abelian nth-power* if it is k-abelian equivalent to a nth-power. In WORDS 2013, Mari Huova and Aleksi Saarela prove that strongly k-abelian nth-powers are unavoidable on any alphabet.

- Pattern avoidance by palindromes was the subject of the talk from Inna A. Mikhailova and Mikhail Volkov, in WORDS 2007.

A word is a *pseudopalindrome* if it is the fixed point of some involutary antimorphism ϕ of the free monoid (i.e. $\phi^2 = id$, $\phi(uv) = \phi(v)\phi(u)$) and the *pseudopalindromic closure* of a word w is the shortest pseudopalindrome having w as a prefix. In their contribution to WORDS 2009, Damien Jamet, Genevieve Paquin, Gwenaël Richomme and Laurent Vuillon present some several combinatorial properties of the fixed points under iterated by pseudopalindromic closure.

Factorization of words. Equations

Other important informations may be obtained by decomposing words into a convenient sequence of consecutive factors: $w = w_1 \cdots w_n$.

- In their talk of WORDS 1997, Juhani Karhumäki, Wojciech Plandowski and Wojciech Rytter investigate the properties of the so-called $\mathcal{F}$-factorization, where the preceding sequence $(w_1, \cdots, w_n)$ was assigned to satisfy a given property $\mathcal{F}$. From an algorithmic point of view, they examine the behavior of three fundamental properties of such factorization, namely *completeness*, *uniqueness* and *synchronization*.

- *Periodicity* is also clearly concerned by the notion of factorization. If for an integer $n \geq 2$, all the preceding words $w_1, \cdots, w_{n-1}$ are equal, the word w_n being one of their prefix, we say that the length of w_1 is a *period* of w. The famous theorem of Fine and Wilf states that if some powers of two words x, y have a common prefix of length $|x| + |y| - \gcd(|x|, |y|)$, then x and y themselves are powers of the same word. In WORDS 1997, Maria Gabriella Castelli, Filippo Mignosi and Antonio Restivo present an extension of Fine and Wilf's theorem for three periods. In their contribution to WORDS 2003, Sorin Constantinescu and Lucian Ilie prove a new extension of that theorem for arbitrary number of periods, and in WORDS 2007, Vesa Halava, Tero Harju, Tomi Krki, and Luca

Q. Zamboni will study the so-called relational Fine and Wilf words. At least, in his contribution to WORDS 2007, Kalle Saari, examine periods of the factors of the Fibonacci word.

- A word w is *quasiperiodic* if another word x exists such that any position in w falls in an occurrence of x as a factor of w (unformaly, w may be completely "covered" by a set of occurences of the factor x). In the conference of 2013, Florence Levé and Gwenaël Richomme extend the work that they presented in WORDS 2007: in particular they present algorithms for deciding whether a morphism is strongly quasiperiodic on finite and infinite words.

- In WORDS 1999 Juhani Karhumäki and Ján Maňuch prove that if a non-periodic bi-infinite word possesses three disjoint factorizations on the words of a prefix-free set X, then a set Y with cardinality at most $|X| - 2$ exists such that $X \subseteq Y^*$.

Such a type of *defect effect* is strongly connected to *independent systems of equations*, as illustrated by the paper presented by Tero Harju and Dirk Nowotka in WORDS 2001, where the case of equations in three variables is investigated. Some properties of infinite systems of equations are also presented by Štěpán Holub and Juha Kortelainen in WORDS 2005.

- The famous Post Correspondence Problem (PCP for short) is also connected to decomposition of words. Given two morphisms h, g, it consists in examining wether the existence of a non-empty solution for the equation $h(x) = g(x)$ is decidable. In the meeting of 2005 Vesa Halava, Tero Harju, Juhani Karhumäki and Michel Latteux provide an extension of the decidability of the marked PCP to instances with unique blocks. The properties of new variants such as the circular-PCP, and the n-permutation PCP are also examined by Vesa Halava in his presentation of WORDS 2013. In the same meeting, in the topic of the *Dual-PCP*, the so-called periodicity forcing words is the aim of the talk from Joel D. Day, Daniel Reidenbach and Johannes C. Schneider.

Complexity issues

In the literature, with a word several notions of complexity can be associated, the most famous being certainly the *factor complexity*: given a word w, this complexity measures the number $p_w(n)$ of different factors of length n occuring in w. The famous characterization of Morse-Hedlund for ultimately periodic words has led to introduce the infinite *Surmian words* whose complexity is $p_w(n) = n + 1$, the best known example of them being certainly the famous *Fibonacci word*.

- Other notions of complexity for infinite words are defined by Sébastien Ferenczi and Zoltán Kása in their paper of WORDS 1997: the behavior of *upper (lower) total finite-word complexity* and *upper (lower) maximal finite-word complexity* are compared to the classical factor complexity, moreover new characterizations of Sturmian sequences are obtained.

-The *recurrence function* has been introduced by Morse and Hedlund: given a factor u, it associates with every non-negative integer n the size $R_u(n)$ of

the smallest window that contains every factor of length n of u. During its talk of WORDS 1997, Julien Cassaigne introduces the *recurrence quotient* as $\limsup_{n\to\infty} \frac{R(n)}{n}$, moreover he computes it for Sturmian sequences.

- The notion of *special factor* allows to obtain a performing characterization of Sturmian words. In the case of finite words, it leads to introduce two parameters, namely R, L which, given a finite word w, represent the least integer such no right (left) special factors of lenght $\geq R$ ($\geq L$) may occur in w. In his paper of WORDS 1997, Aldo de Luca studies the connections between these parameters and the classical factor complexity.

- The study of the ratio $p(n)/n$ brings also noticeable informations on infinite words. In WORDS 1999, Alex Heinis shows that if $p(n)/n$ has a limit, then it is either equal to 1, or highter than and equal to 2. By using the Rauzy graphs, in WORDS 2001, Ali Aberkane will present characterizations of the words such that the limit is 1.

- An words is *balanced* if for any pairs (u, v), of factors with same length, and for any letter a, we have $||u|_a - |v|_a| \leq 1$ (where $|u|_a$ stands for the number of occurrences of the letter a in u). In the paper he presents in WORDS 2001, Boris Adamczewski defines the *balance function* as $\max_{a\in A} \max_{u,v\in F(w)} \{||u|_a - |v|_a|\}$: as regard to the so-called *primitive substitutions*, he investigates the connections between the asymptotic behavior of the balance function and the incidence matrix of such a substitution. In the workshop of 2007, several contributions to the topic are also presented: Nicolas Bédaride, Eric Domenjoud, Damien Jamet and Jean-Luc Remy study the number of balanced words of given length and height on a binary alphabet. During WORDS 2013, two talks relating investigations of the property of balancedness of the Arnoux-Rauzy word are presented: one is from Julien Cassaigne and the other one from Vincent Delecroix, Tomáš Hejda and Wolfgang Steiner.

- The *arithmetical complexity* of an infinite sequence is the number of all words of a given length whose symbols occur in the sequence at positions which constitute an arithmetical progression. The study of arithmetical complexity will also appear in the contribution of Julien Cassaigne and Anna Frid of WORDS 2005: they give a uniform $O(n^3)$ upper bound for the arithmetical complexity of a Sturmian word and provide explicit expressions for the arithmetical complexity of Sturmian words of slope between 1/3 and 2/3 (this is, in particular, the case of the infinite Fibonacci word). In this case the difference between the genuine arithmetical complexity function and the precedingly mentioned upper bound is itself bounded and ultimately 2-periodic.

- The *palindromic complexity* of an infinite word is the function which counts the number $P(n)$ of different palindromes of each length occurring as factors in the word. In their talk of WORDS 2005, Peter Bali, Zuzana Masḱová and Edita Pelantovà provide an estimate of the palindromic complexity $P(n)$ for uniformly recurrent words; denoting by $p(n)$ the classical factor complexity this estimation is based on the equation: $P(n) + P(n + 1) = p(n + 1) - p(n) + 2$.

- The *m-binomial complexity* of an infinite word w maps an integer n to the number of m-binomial equivalence classes of factors of length n that occur in w. This relation of binomial equivalence is defined as follows: two words u, v are m-equivalent if, for any word x of length at most m, x appears in u and v with the same number of occurrences. In their contribution to WORDS 2013 Michel Rigo and Pavel Salimov compute the m-binomial complexity of famous words: the Sturmian words and the Thue-Morse word.

- Given a word v with several occurrences in an infinite word, the set of *return words* of v has for elements all distinct words beginning with an occurrence of v and ending just before the next occurrence of v. It had been proved that a word is Sturmian if and only if each of its factors has two returns: in the meeting of 2011, Svetlana Puzynina and Luca Zamboni prove that a word is Sturmian if and only if each of its factors has two or three abelian returns.

- Complexity is also implicitely present in the talk given by Shuo Tan and Jeffrey Shallit in WORDS 2013. Given an alphabet A, a subset X of A^n is *representable* if it occurs as the set of all factors of length n of a finite word. Clearly, the set A^n itself is represented by any De Bruinj word of order n. One of the questions which are presented by the authors consists in examining the length of a word needed to represent a given representable set X: they provide a lower and a upper boundsof the form α^{2^n} with $\alpha = \sqrt{2}$ for the lower bound and $\alpha = \sqrt[10]{4}$ for the upper one.

Many other interesting topics were presented during the conference: in the sequel we can only mention some of them.

As regard to sets of words, in WORDS 1999, Jean Berstel and Luc Boasson prove that, given a finite set of words S, at most one (normalized) multiset P may exists such that S is the shuffle of the words in P, the multiset P being effectively computable. Codes were also the subject of talks from Véronique Bruyère and Dominique Perrin (WORDS 1997), Jean Néraud and Carla Selmi (WORDS 2001), Fabio Burderi (WORDS 2011).

Extensions of the classical concept of words were also the subject of a lot of presentations: the notion of partial word has been introduced in WORDS 1997 by Jean Berstel and Luc Boasson; in WORDS 2009, in the framework of binary words, Francine Blanchet Sadri and Brian Shirey will examine the relationship between such a notion and periodicity. Multidimensional words were the feature of the talks from Valérie Berthé and Robert Tijdeman in WORDS 1999, and in the meeting of 2005, they were also the subject of two presentations: one was given by Pierre Arnoux, Valérie Berthé, Thomas Fernique and Damien Jamet and the other one by Jean-Pierre Borel.

Connections with Semigroups Theory were the subject of the lectures of Sergei I. Adian in WORDS 2003.

Words in connection with Number Theory and Numeration Systems were also the feature of very interesting papers from Tom Brown (WORDS 1999), Petr Ambrož and Christiane Frougny (WORDS 2005), Daniel Dombek (WORDS 2011), Shigeki Akiyama, Victor Marsault and Jacques Sakarovitch (WORDS

2013).

Anyway each of the numerous results which will be presented in the ten conferences plays a noticeable part in the state-of-the-art. We will have reached our goal if the preceding notes may testify to the involvement of WORDS in the development of the field of Combinatorics on Words.

References

1. Jean Berstel, Axel Thue's papers on repetitions in words: a translation, http://igm.univ-mlv.fr/∼berstel/Articles/1994ThueTranslation.pdf.
2. Jean Berstel and Dominique Perrin, Theory of Codes, Academic Press (1985).
3. Jean Berstel, Dominique Perrin and Christophe Reutenauer, Codes an Automata, Encyclopedia of Mathematics and its Applications 129, Cambridge University Press, (2010).
4. Arturo Carpi and Clelia De Felice, Guest Editors, Combinatorics on Words (WORDS 2009), 7th International Conference on Words, Fisciano, Italy, *Theoretical Computer Science 412* (27) (2011) 2909-3032.
5. Srecko Brlek and Christophe Reutenauer, Guest Editors, Combinatorics on words, *Theoretical Computer Science 380* (3) (2007) 219-410.
6. Tero Harju, Juhani Karhumäki and Antonio Restivo, Guest Editors, Combinatorics on words, *Theoretical Computer Science 339* (1) (2005) 1-166.
7. Juhani Karhumäki, Arto Lepistö and Luca Zamboni (Eds) Combinatorics on Words, 9th International Conference, WORDS 2013, Turku, Finland, September 2013 Proceedings,*Lecture Notes in Computer Science* 8079 (2013) 1-263.
8. Juhani Karhumäki, Arto Lepistö and Luca Zamboni, Guest Editors, Combinatorics on Words, *Theoretical Computer Science*, **601** (2015) 1–72.
9. M. Lothaire, Combinatorics on Words, Cambridge University Press, second edition, 1997.
10. M. Lothaire, Algebraic Combinatorics on Words, Encyclopedia of Mathametics and its Applications 90, Cambridge University Press, 2002 (First edition 1983).
11. M. Lothaire, Applied Combinatorics on Words, Encyclopedia of Mathematics and its Applications 105, Cambridge University Press, 2005.
12. Zuzana Masáková and Štěpán Holub, Special Issue WORDS 2011, *International Jounal of Foundations of Computer Science 23* (8) (2012) 1579-1728.
13. Jean Néraud, Guest Editor, WORDS, *Theoretical Computer Science 218* (1) (1999) 1–216.
14. Jean Néraud, Guest Editor, WORDS, *Theoretical Computer Science 273* (1-2) (2002) 1–306.
15. Jean Néraud, Guest Editor, WORDS, *Theoretical Computer Science 307* (1) (2003) 1-216.
16. James F. Power, Thues 1914 paper: a translation, http://arxiv.org/pdf/1308.5858.pdf
17. Axel Thue, *Über unendliche Zeichenreihen* Norske Vid. Selsk ; Skr. I. Mat. Kl., Christiana 1906 (7) 1-22.
18. Axel Thue, *Über die gegenseitige Lager gleicher Teile gewisser Zeichenreihen*, Skr. I. Mat. Kl., Christiana 1912 (1) 1-67.
19. Axel Thue, *Probleme über Verängerungen von Zeichenreihen nach gegebenen Regeln*, Christiana Videnskabs-Selskabs Skrifter, I. Math. naturv. Kl. 10 (1914). Reprinted in [Nagel and al. 77, pp. 493-524] .