Available online at www.sciencedirect.com
ScienceDirect
European Journal of Combinatorics 28 (2007) 996–1022
European Journal of Combinatorics
www.elsevier.com/locate/ejc

The origins of combinatorics on words

Jean Berstel, Dominique Perrin

Institut Gaspard–Monge, Université de Marne–la–Vallée, F-77454 Marne-la-Yallé Cedex 2, France

Received 1 July 2005; accepted 26 July 2005
Available online 15 December 2005

Abstract

We investigate the historical roots of the field of combinatorics on words. They comprise applications and interpretations in algebra, geometry and combinatorial enumeration. These considerations gave rise to early results such as those of Axel Thue at the beginning of the 20th century. Other early results were obtained as a by-product of investigations on various combinatorial objects. For example, paths in graphs are encoded by words in a natural way, and conversely, the Cayley graph of a group or a semigroup encodes words by paths. We give in this text an account of this two-sided interaction.

1. Introduction

Combinatorics on words is a comparatively new area of discrete mathematics. The collective volumes written under the pseudonym of Lothaire give an account of it. Lothaire's first volume [74] appeared in 1983 and was reprinted with corrections in 1997 [75]. In the introduction to the first edition, Roger Lyndon stated that

> This is the first book devoted to broad study of the combinatorics of words, that is to say, of sequences of symbols called letters. This subject is in fact very ancient and has cropped up repeatedly in a wide variety of subjects.

In 2002, almost twenty years after the appearance of the first volume, a second volume was published [76], with a partially different set of authors, under the editorial guidance of the authors of this survey. A third volume [77], dedicated to the applications of this field, appeared in July 2005.

E-mail address: jean.berstel@univ-mlv.fr (J. Berstel).

0195-6698/$ - see front matter
doi:10.1016/j.ejc.2005.07.019

In this paper, we investigate the historical roots of this field. Needless to say, the material relies heavily on the Notes sections of each of the chapters of Lothaire's volumes. The aim of this form of presentation is to underline the interesting intertwinings among the early ideas that eventually combined to form the present state of the art. We will see that ideas from algebra in group theory, number theory and differential geometry have led to the elaboration of the concepts of today's combinatorial theory of words.

Our paper is written in such a way that a non-specialist reader can use it as a gentle introduction to the subject. We never give proofs and we stay rather informal, but we hope that the examples provided are explicit enough to explain the underlying problems.

Let us now outline the content of this paper. We start with a section on regularities in words. This begins with a description of the work of A. Thue on *square-free words* at the dawn of the 20th century. This is certainly the real beginning of our subject. To quote Roger Lyndon again:

> The systematic study of words seems to have been initiated by Axel Thue in three papers (...). Even more than for his theorems, we owe him for delineating this subject.

The next subsection describes automatic words which have, in the continuation of Thue's work, a connection with number theory. Let us mention at this point that we did not include a subject which is related to ours and concerns words as representation of integers in some basis (automatic words are linked with infinite expansions rather than finite ones). This is, nonetheless, part of the historical roots of combinatorics on words, since many recreational problems on integers are formulated in terms of their representation.

Section 2.3 deals with the famous topic of *unavoidable regularities* on words. This subject can be considered as a generalization of the idea of square-free words: the existence of infinite square-free words on three symbols means, in this terminology, that squares are avoidable. On the contrary, some regularities are unavoidable and this is linked to famous results in combinatorics such as the theorems of Ramsey and van der Waerden.

The next two sections (Sections 3.1 and 3.2) deal with topics which belong to the field of *symbolic dynamics* founded by Marston Morse and Gustav Hedlund, with ideas originating in the theory of dynamical systems, a branch of classical analysis (see [73] for an introduction to the field of symbolic dynamics). It is remarkable that they found an intersection with the work of Thue, exemplified in what is now called the Thue–Morse infinite word.

Section 4 presents the rich and curious history of necklaces, de Bruijn and Lyndon words, from Euler to modern problems of coding.

The next section (Section 5) deals with algebraic topics related to group and semigroup theory. It also introduces some famous undecidable problems formulated in terms of words, such as the Post correspondence problem.

The last section (Section 6) deals with the use of words to encode paths and nonlinear structures such as trees or curves in the plane.

We have added at the end a chronology showing the intertwining of the development of these ideas with some early references to ancestors such as Bernoulli, Hadamard and Poincaré. The bibliography does not pretend to be exhaustive, but represents a guide to old literature and also to other texts treating, as we do, the history of this subject.

Preliminary versions of this paper have been presented in particular at the Centre de Mathématiques Sociales of the Maison des Sciences de l'Homme, on the invitation of Pierre Rosenstiehl, and at the University of Rouen, at a workshop organized by Jean Néraud and Julien Cassaigne.

2. Regularities

Regularities in words, even if produced by random devices, constitute a paradoxical phenomenon. They have been an object of astonishment for many years. More precisely, regularities in the appearance and distribution of digits in the decimal expansion of familiar numbers such as π is an old subject of investigation. The study of infinite sequences of abstract symbols reveals the existence of a dichotomy between avoidable and unavoidable regularities.

2.1. The work of Thue and square-free words

The investigation of repetitions in words began with the work of Axel Thue (1863–1922). He wrote two papers on this subject, one in 1906 [129] and one in 1912 [131]. It is appropriate to quote here[1] the introduction of his 1912 paper:

> For the development of logical sciences it will be important, without consideration for possible applications, to find large domains for speculations about difficult problems. In this paper, we present some investigations in the theory of sequences of symbols, a theory which has some connections with number theory.

Nowadays, the concept of repetitions is familiar to combinatorialists and to biologists, for instance under the name of *tandem repeats*. This was not the case when Thue wrote the lines above, and his work remained largely unknown for forty years, as acknowledged by Hedlund in 1967 [59].

A *word* is a finite, infinite from left to right or two-sided infinite sequence of symbols taken in a finite set called *alphabet*. The letters of an alphabet are usually denoted by $a, b, c, \ldots$. The set of finite words over an alphabet A is denoted by A^*. The *length* of a finite word is the number of symbols it is composed of. The length of w is denoted by $|w|$. The *empty word*, denoted ε, is the unique word of length 0. A *factor* of a word x is a block of consecutive symbols of x. For instance, the word $abba$ is a factor of $baabbaab$.

A word (finite, infinite from left to right or two-sided infinite) is called *square-free* if it does not contain two adjacent identical factors. For instance, the word word is square-free whereas repetition is not.

Historically, the first infinite square-free word was given by Thue in his 1906 paper. To describe it, we introduce some terminology. A *substitution* is a mapping h which assigns to each symbol a word. It is extended to words by the rule $h(xy) = h(x)h(y)$, and to (right) infinite words $x = a_0a_1\ldots$ by $h(x) = h(a_0)h(a_1)\ldots$. In this sense, it is a monoid morphism; therefore a substitution is frequently called a *morphism*. A fixed point of a substitution h is an infinite word x such that $h(x) = x$.

The word defined by Thue is the (unique) fixed point of the substitution

$$a \mapsto adbcb$$
$$b \mapsto abdcb$$
$$c \mapsto abcdb$$
$$d \mapsto abcbd$$

[1] Thue wrote his papers in German, and says: "Für die Entwicklung der logischen Wissenschaften wird es, ohne Rücksicht auf etwaige Anwendungen, von Bedeutung sein, ausgedehnte Felder für Spekulation über schwierige Probleme zu finden. Wir werden hier in dieser Abhandlung einige Untersuchungen aus der Theorie über Zeichenreihen, die gewisse Berührungspunkte mit der Zahlentheorie darbietet, mitteilen."

Actually, this word reads

$$adbcb\ abcbd\ abdcb\ abcdb\ abdcb\ldots.$$

In his second paper, in 1912, Thue introduced what is now called the Thue–Morse word

$$t = abbabaab\ldots$$

obtained by iterating the substitution

$$a \mapsto ab, \qquad b \mapsto ba,$$

starting with the letter a. One obtains

$$abba\,baab\,baab\,abba\ baab\ abba\ abba\ baab\ldots.$$

Thue proved that t is *overlap-free*, in the sense that it has no factor of the form $uvuvu$ for some words u, v, with u nonempty. He also proved a nice relationship between square-free words on three letters and overlap-free words. Indeed, for any infinite overlap-free word x over two letters a, b, the inverse image of x under the substitution

$$a \mapsto abb, \qquad b \mapsto ab, \qquad c \mapsto a$$

is a square-free word on three letters a, b, c without the factors aba or cbc, and conversely (see [11] for a description of the content of Thue's papers). Starting from the Thue–Morse word t, one obtains the square-free word on three symbols

$$m = abcacbabcbac\ldots.$$

The study of repetitions in words was forgotten for a long time after Thue's initial work. There was a partial revival in the 1920's and 1940's, and much progress has been achieved in the last twenty years (see [76] for a survey). We shall meet again these problems in Section 2.3.

2.2. Number theory and automatic words

We shall see in what follows that the Thue–Morse word appears several times, mainly because Morse and many others encountered this fundamental sequence independently of Thue (see [3] for a systematic exposition of this curious phenomenon). Let us mention a connection with number theory here. Consider two sequences $(a_1, a_2, \ldots, a_r)$ and $(b_1, b_2, \ldots, b_r)$ of natural numbers such that

$$a_1^k + a_2^k + \cdots + a_r^k = b_1^k + b_2^k + \cdots + b_r^k, \qquad (1 \le k \le m).$$

It is easy to see that one has $r \ge m + 1$ if the solution is nontrivial, i.e., if the r-tuples are not permutations of each other. Prouhet [105] was the first to raise a particular case of this problem (namely $r = n^{m-1}$) in the following terms:

> n et m étant deux nombres entiers quelconques, il existe une infinité de suites de n^m nombres susceptibles de se partager en n groupes de n^{m-1} termes chacun, et tels que la somme des puissances k des termes soit la même pour tous les groupes, k étant un nombre entier inférieur à m.
>
> n^m nombres en progression arithmétique jouissent de la propriété précédente. Pour opérer le partage de ces nombres en groupes, on écrira en cercle les indices $0, 1, 2, \ldots, n-1$; on lira les indices en suivant le cercle et en ayant soin d'en passer

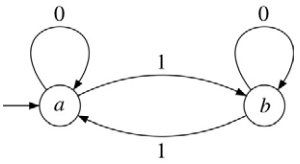

Fig. 1. The even system. The letters are 0 and 1. The state a is initial.

un à chaque tour; deux tous les n tours; trois tous les n^2 tours, et ainsi de suite. Ces indices, écrits à mesure qu'on les lit sous les termes de la progression, apprendront à quel groupe appartient chaque terme.

Si l'on applique la règle et le théorème précédents aux 27 premiers nombres de la suite naturelle, on arrive aux identités suivantes:

$$\begin{aligned} 1+6+8+12+14+16+20+22+27 &\\ = 2+4+9+10+15+17+21+23+25 &\\ = 3+5+7+11+13+18+19+24+26 &\\ 1^2+6^2+8^2+\cdots = 2^2+4^2+9^2+\cdots = 3^2+5^2+7^2+\cdots. & \end{aligned}$$

Lorsque $n = 10$ et que la progression commence en 0, tous les nombres dont la somme des chiffres, divisée par 10, laisse le même reste, appartiennent à la même classe.

In the special case $n = 2$, Prouhet's solution for $m \geq 1$ with $r = 2^m$ is obtained by partitioning the integers $\{0, 1, \ldots, 2^r - 1\}$ into two sets of a_k and b_k according to the parity of the sum of their digits in base 2. For $m = 2$, we obtain

$$\begin{aligned} 0+3+5+6 &= 1+2+4+7 \\ 0^2+3^2+5^2+6^2 &= 1^2+2^2+4^2+7^2. \end{aligned}$$

This is the Thue–Morse sequence again, writing a at the positions corresponding to indices a_i, and b at b_i. Indeed, the Thue–Morse word $t = t_0t_1t_2\ldots$ can be defined alternatively as follows. Let $b(n)$ be the sum of the bits in the binary expansion of n. Then

$$t_n = \begin{cases} a & \text{if } b(n) \text{ is even} \\ b & \text{otherwise} \end{cases}$$

The Thue–Morse word is sometimes also called the Prouhet–Thue–Morse word, in recognition of the precedence of Prouhet. The number-theoretic problem raised by Prouhet was studied later by Tarry [128] and Escott and many other authors. Dickson [31] devotes a chapter of his book to the seventy papers on this topics, at that time. He also explains the connection of the problem to the efficient computation of numerical values of logarithms. This is the purpose of Escott's papers [40,39]. Nowadays, the so-called Prouhet–Tarry–Escott problem consists in the determination of the optimal value of k for each m (see [58]).

An interesting point is that Prouhet's definition of the Thue–Morse word has a connection with finite automata, as noted by Cobham. A *finite automaton* (see e.g. Fig. 1) is a finite graph with edges labeled with letters in some alphabet. The nodes are usually called states. There is a distinguished *initial* state. The state *reached* by a word w is the state obtained from the initial state by following the path composed of the edges labeled by the letters of w. One assumes here that the automaton is deterministic, that is that only one such path may exist. For instance, the state reached by 1010 in Fig. 1 is a.

As an equivalent characterization, t_n is the output of a finite automaton reading the binary expansion of the integer n. The output is defined according to the state reached. Such a word (or sequence) is called *automatic* (more precisely k-automatic for a choice of the base k) or k-recognizable, in Eilenberg's terminology [36]. The idea of *automatic words* appears in the work of Cobham [24] who proved that an infinite word is k-automatic iff it is the image, under a length-preserving morphism, of a fixed point of a substitution of constant length k. As an example of another automatic word, the *Rudin–Shapiro* word

$$r = aaabaabaaaabbbab\ldots$$

is defined by

$$r_n = \begin{cases} a & \text{if } d_{11}(n) = 0 \bmod 2 \\ b & \text{if } d_{11}(n) = 1 \bmod 2 \end{cases}$$

where $d_{11}(n)$ is the number of factors 11 in the binary expansion of n. It is also the image, under the morphism

$$0, 1 \mapsto a, \qquad 2, 3 \mapsto b$$

of the fixed point of the substitution

$$0 \mapsto 01, \qquad 1 \mapsto 02, \qquad 2 \mapsto 31, \qquad 3 \mapsto 32$$

starting with 0 (here again, the alphabet is composed of the symbols 0, 1, 2, 3). See [4] for a systematic exposition of the theory of automatic words.

2.3. Unavoidable regularities

The existence of infinite square-free words on three letters can be seen as an example of an avoidable pattern. In general, we say that the pattern p *occurs* in the word w if there is a non-erasing substitution h such that $h(p)$ is a factor of w. A substitution is *non-erasing* or *length-increasing* if it maps no word into the empty word. In the case p does not occur in w, we say that w *avoids* the pattern p. Equivalently, one may call the *pattern language* of p over an alphabet A the set of all words $h(p)$ where h is a non-erasing substitution into A^*. The w avoid p if and only if no factor of w is in the pattern language of p.

Thus a word is square-free if it avoids the pattern xx. It is overlap-free if it avoids the pattern $xyxyx$. The patterns $\alpha, \alpha\beta\alpha, \alpha\beta\alpha\gamma\alpha\beta\alpha, \ldots$ are called *sesquipowers*. They are all unavoidable, a fact which has been used to formulate finiteness conditions in algebra (see [62] for example, where sesquipowers are called m-sequences). In Coudrain and Schützenberger [26], the unavoidability of sesquipowers is used to formulate a finiteness condition for semigroups (see also Chapter 4 of [76]).

Finding infinite words that avoid repetitions has its roots in the work of Thue and has been pursued, in particular in connection with problems of algebra. The general idea of unavoidable patterns was introduced independently by Bean et al. [8] and by Zimin [144]. One of the striking theorems obtained by Zimin shows that sesquipowers (also called *Zimin words*) are essentially all unavoidable patterns: a pattern p is unavoidable if and only if there exists a sesquipower which has a factor in the pattern language of p. For example, the pattern $p = \alpha\beta\gamma\beta\alpha$ is unavoidable, consistent with the fact that the sesquipower $\alpha\beta\alpha\gamma\alpha\beta\alpha$ has the factor $\beta\alpha\gamma\alpha\beta$ in which p obviously occurs. These results were used in algebra, for example to exhibit varieties of algebras which cannot be defined by a finite set of identities (see [117]).

The general idea of unavoidable regularities is actually much wider and does not have a precise definition. One direction is that initiated by Ramsey in 1930 [110]. His famous theorem states that, given integers k, r, for any coloring of the k element subsets of an infinite set X in r colors, there exists an infinite subset Y of X such that all k element subsets of Y have the same color (this is the infinite version of Ramsey's theorem which has also a finite version; see [53] for a survey). Applied to infinite words, the theorem implies that if we color the factors of an infinite word x in r colors, there is a factorization $x = vu_0u_1u_2\ldots$ such that $u_0, u_1, u_2, \ldots$, and even all factors $u_i \ldots u_j$ with $i \leq j$ have the same color. Indeed, each occurrence of a factor in x is a 2 element subset of the set of all positions of letters in x. So there exists an infinite subset of positions in x such that each occurrence of a factor delimited by 2 positions in this subset has the same color. A related unavoidable regularity is given by van der Waerden's theorem [138], which says that if the positive integers are partitioned in r classes, then at least one of the classes must contain arbitrary long arithmetic progressions. Actually, as related in [139], this theorem had been conjectured by I. Schur. Van der Waerden heard of the conjecture through Baudot, a student at Göttingen at the time, and referred to his result as Baudot's conjecture. This applies directly to words, with the equivalent formulation that for any infinite word x there are arbitrary large integers k such that for some n, m, we have $x_n = x_{n+m} = \cdots = x_{n+km}$.

Van der Waerden's theorem admits several generalizations. One of them, the Hales–Jewett theorem, puts it in the form of a Ramsey type theorem (see [53]). Another generalization had been conjectured by Erdös and Turan in 1936 [38]: if A is a set of positive integers with positive upper density, i.e., such that $\limsup |A \cap [1, n]|/n > 0$, then A contains arbitrary long arithmetic progressions. This theorem was proved by Szemerédi in 1975 [127]. A proof based on ergodic theory was given by Fürstenberg in 1977 [46].

3. Symbolic dynamics

We are going to meet the Thue–Morse word again in this section. It was indeed rediscovered, ten years after Thue, by Marston Morse, the founder of the field called symbolic dynamics.

3.1. Recurrence and minimality

The study of dynamical systems is derived from the work of Newton on the laws of motion applied in particular to the planetary system. The motion of a dynamical system, in this formulation, is governed by a system of differential equations satisfied by the parameters of the system as a function of time. To quote H. Fürstenberg [47] in his introduction called 'From Differentiable to Abstract Dynamics':

> The approach most natural for the eighteen and nineteen century analysts from Euler to Jacobi was to extract as much information as one could by analytic manipulation of the specific differential equations.

The point of view was decisively shifted by Poincaré in *Méthodes Nouvelles de la Mécanique Céleste* from the individual solution curves to the set of all possible curves and their interrelation [100]. Birkhoff made this transition more explicit and two theories evolved from these developments: ergodic theory and topological dynamics. The first one studies the action of a group of transformations on a measure space, the second on a topological space.

A fundamental notion is that of *recurrence* or stability and, to quote Fürstenberg again:

> A meta-theorem of dynamical theory states that whenever the underlying space X of the system is appropriately bounded, the orbits of the motion will exhibit some form of recurrence or return close to their initial position.

Poincaré's theorem asserting that this holds when T is a measure preserving transformation on a space X of finite measure, is the first result of this kind. When X is a topological space and T is a continuous transformation on X, a point $x \in X$ is called recurrent if for any neighborhood V of x there exists an $n \geq 1$ such that $T^n x \in V$ (older literature speaks of an orbit 'stable in the sense of Poisson'). The point x is called *uniformly recurrent* if the set of integers n such that the above condition holds has bounded gaps (other writers, such as Birkhoff and Morse, speak about 'recurrence' and some modern ones use the term 'almost periodicity').

Words come in with what was called by Marston Morse (1892–1977) and Gustav Hedlund a *symbolic flow*. The elements, called *symbolic trajectories*, are just infinite words (actually doubly or simply infinite). The idea goes back to Hadamard [56], who first used sequences of symbols to describe qualitatively the infinite geodesic curves on a surface. Morse proved the existence of a non-periodic uniformly recurrent point [92] by considering the Thue–Morse word. It is indeed easy to prove that the Thue–Morse word is uniformly recurrent.

In the foundational paper of Morse and Hedlund in 1938 on symbolic dynamics [94], the notion of what is now known as a subshift is introduced somewhat informally. It is the set of two-sided infinite words (the 'symbolic trajectories') subject to a restriction defined by a set of forbidden blocks. Morse and Hedlund observe that this condition also applies to define the reduced words in the free group (see Section 5.2). Fürstenberg uses in his book [47] the term of a *Bebutov system* to describe the more general notion of a subsystem of Λ^G where Λ is a compact metric space and G a countable group or semigroup. These were introduced by Bebutov in 1940 [9]. The case where Λ is finite and G is either $\mathbb{N}$ or $\mathbb{Z}$ is called a symbolic flow. A simple construction of non-periodic uniformly recurrent words is also described, as an alternative to the use of the Thue–Morse word or other word obtained by iterating a substitution. It has the general form of an infinite sesquipower

$$\omega = [(aw^{(1)}a)w^{(2)}(aw^{(1)}a)]w^{(3)}[(aw^{(1)}a)w^{(2)}(aw^{(1)}a)]\dots$$

for some letter a and a sequence $w^{(1)}, w^{(2)}, w^{(3)}, \dots$ of words. A sufficient condition for the word ω to be uniformly recurrent is that the words $w^{(n)}$ have bounded length. This observation makes the existence of non-periodic uniformly recurrent points in symbolic systems relatively obvious, a fact first established by Robbins in 1937 [114] and of course by Morse using the Thue–Morse sequence.

The terms of *subshift* and the definition of a *subshift of finite type* which is the basic notion of modern symbolic dynamics were introduced by Smale [122] only in 1967. The connection with finite automata went largely ignored for a long time and the identification of subshifts of finite type as a particular case of subshifts definable by a finite semigroup was introduced by Weiss only in 1973 [141] (he introduced the term of *sofic shift* using a Hebrew word meaning "finite").

3.2. Sturmian words

An infinite word over a binary alphabet is called *Sturmian* if for all $n \geq 0$, the number of its factors of length n is $n + 1$. As an example, the *Fibonacci word*

$$f = abaababaabaab\dots$$

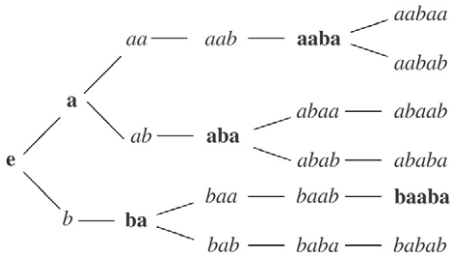

Fig. 2. The factors of the Fibonacci word.

which is defined as the unique fixed point of the substitution $(a \mapsto ab, b \mapsto a)$ is Sturmian. For each length $n \geq 0$, there is exactly one factor of length n which can be followed by both letters (and thus contributes to increase by 1 the number of factors of length $n+1$). Such a factor is called *right special*. The right special factors of length at most 5 of the Fibonacci word are printed in boldface on Fig. 2.

These words were introduced by Morse and Hedlund in 1940 [95] and named after François Sturm (1803–1855), born in Geneva, who taught at the École Polytechnique in Paris from 1840. He is famous for his rule to solve an algebraic equation by considering the sign changes of the corresponding polynomial. Actually, his name is used here in relation with the zeros of solutions of a homogeneous second order differential equation

$$y'' + \phi(x)y = 0$$

where ϕ is continuous of period 1. If k_n is the number of zeros of a solution in the interval $[n, n+1)$, then the infinite word $ab^{k_0}ab^{k_1}\ldots$ is Sturmian (or eventually periodic).

There is also an alternative definition of the Fibonacci word using approximations of irrationals by rationals. Let α be some irrational with $0 < \alpha < 1$, and let $s(\alpha) = (s_n)$ be the sequence

$$s_n = \begin{cases} a & \text{if } \lfloor (n+1)\alpha \rfloor = \lfloor n\alpha \rfloor, \\ b & \text{otherwise} \end{cases}$$

where $\lfloor x \rfloor$ denotes the lower integral part of x. For $\alpha = 2/(3+\sqrt{5})$, one has $s(\alpha) = af$. This formula shows that the symbols s_n can be interpreted as the approximation of a line of slope α by points with integer coordinates (see Fig. 3). It is a theorem due to Morse and Hedlund that Sturmian words can be defined equivalently by a formula as above with α an irrational number. Words defined in such a way are now called *mechanical* (see the survey written by Jean Berstel and Patrice Séébold as chapter 20 [76], or the collective work [107] or also the book of Allouche and Shallit [4]). They have historical roots in the work of the astronomer Jean Bernoulli III [10] who studied these words in connection with continued fractions. The book of Venkov [140] describes early work by Christoffel [23] and by Markov [88].

The connection with continued fractions can be summarized as follows. Let us denote by $[n_0, n_1, n_2, \ldots]$ the continued fraction

$$n_0 + \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cdots}}$$

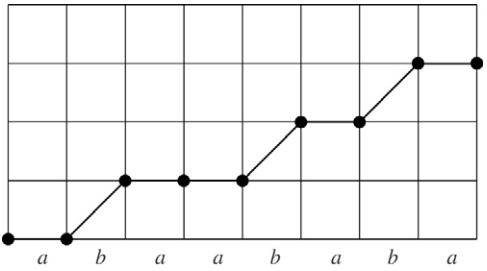

Fig. 3. The graphical representation of the Fibonacci word.

Let $[0, 1 + d_1, d_2, \ldots]$ be the continued fraction expansion of some irrational α with $0 < \alpha < 1$. Let w_n be the sequence of words defined by

$$w_{-1} = b, \qquad w_0 = a, \qquad w_n = w_{n-1}^{d_n} w_{n-2} \quad (n \geq 1).$$

Then $s(\alpha) = \lim w_n$. In particular, let us consider the case of the Fibonacci word. We have $s(\alpha) = af$ with $\alpha = 2/(3 + \sqrt{5})$. Actually, $2/(3 + \sqrt{5}) = [0, 2, 1, 1, \ldots]$ in accordance with the fact that the sequence w_n is the sequence of Fibonacci words.

4. Necklaces

The occurrence of circular sequences or necklaces is certainly very old, since any periodic discrete phenomenon gives rise to such a sequence. Fields such as music or astronomy are places where such phenomena are commonplace.

4.1. Enumeration of necklaces

A circular word, or *necklace*, is the equivalence class of a word under circular shift (see Fig. 4). A necklace of length n is primitive if its period is not a proper divisor of n. The enumeration of necklaces of length n on k symbols has been known for a long time. It appears explicitly in a paper of MacMahon of 1892 [85]. The formula

$$M(n, k) = \frac{1}{n} \sum_{d|n} k^d \mu(n/d)$$

for the number $M(n, k)$ of primitive necklaces where μ is the Möbius function, is often called *Witt's formula* (see e.g.[86]). It was actually proved by E. Witt in 1937 in connection with the theorem on free Lie algebras now called the Poincaré–Birkhoff–Witt theorem [143]. The same formula holds for the enumeration of polynomials of degree n on a field with k elements, when k is a prime. This was proved by Gauss [48], and can be used to prove the existence of an irreducible polynomial of each degree on a finite field of k elements. No natural bijection is known between Lyndon words and irreducible polynomials.

The formula for the total number of necklaces of length n on k symbols

$$N(n, k) = \frac{1}{n} \sum_{d|n} k^d \varphi(n/d)$$

where φ is Euler's function, is called *MacMahon's* formula in the book by Graham et al. [52]. In Lucas' book "Théorie des Nombres" [80], page 503, it is credited to M. le colonel C. Moreau.

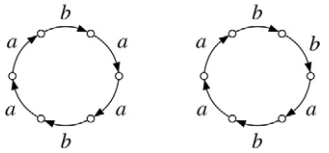

Fig. 4. A necklace with period 3, and a primitive necklace.

A famous result is the existence, for each $n, k \geq 1$, of a de Bruijn cycle of length k^n on k letters. This is a necklace such that each word of length n over k letters appears exactly once as a factor. For example,

$$aaaaabaaabbaababaabbbababbabbbbb$$

is a de Bruijn cycle of length 32 for $k = 2, n = 5$. This result has a curious and interesting history told by Nicolaas Govert de Bruijn in [30] (and also by Knuth in [66]). First, it was actually explicitly obtained many years before by Flye Sainte-Marie in 1894 [44]. He proved that the number of such necklaces of length 2^n on a binary alphabet is $2^{2^{n-1}-n}$. For example, there are $2^{11} = 2048$ binary de Bruijn necklaces of length 2^5. The result had been conjectured by A. de Rivière the same year [113] as question 48:

> Si l'on considère tous les arrangements n à n qu'on peut former avec deux objets,[2] il est toujours possible de trouver un arrangement de 2^n termes (formé avec les mêmes deux objets) $a_1, a_2, \ldots, a_{2^n}$, tel que les groupes
>
> $$a_1, a_2, \ldots, a_n; \qquad a_2, a_3 \ldots, a_{n+1}; \ldots$$
> $$a_{2^n-1}, a_{2^n}, a_1, \ldots, a_{n-2}; \qquad a_{2^n}, a_1, a_2, \ldots, a_{n-1};$$
>
> représentent tous les arrangements n à n dont le nombre est évidemment 2^n. Cette proposition est vérifiée expérimentalement jusqu'à des limites suffisantes pour en présager l'exactitude. Est-elle déjà connue? Pourrait-on en donner une démonstration? Y a-t-il en général plus d'une espèce de solutions et dans ce cas combien?

The same problem was reintroduced by Martin in 1934 [90], who first had the idea of producing the least possible such word by a greedy algorithm. Independently, the problem was raised again in 1943 by Klaas Posthumus (1902–1990), a radio engineer working at the Philips Research Laboratories. He found that the number of de Bruijn cycles for $n = 1, 2, 3, 4, 5$ was $1, 1, 2, 16, 2048$. The cycles with $n = 5$ were of technical interest. Indeed, the Baudot code, patented in 1874 by Emile Baudot (1845–1903), use 5-bit words to encode 32 characters. The encoding is obtained through the rotation of a wheel with 32 electric contacts producing the 32 possible codewords by rotation. The existence of a binary de Bruijn sequence for $n = 5$ was thus an essential element of the encoding procedure for this code. Baudot's code replaced the Morse code before being itself replaced by the ASCII 8-bit code almost a century later. The number of cycles was, however, not of any practical significance, and the interest in proving the formula seems to have been purely *pour le sport*. N.G. de Bruijn worked at this problem when he began to work himself at the Philips Laboratories in 1944. He worked out by hand the case $n = 6$, finding 67108864 cycles and used the techniques he had developed to achieve

[2] in other terms, binary words of length n.

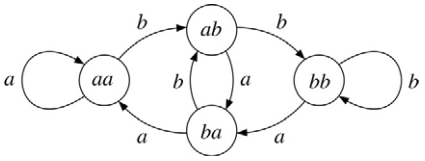

Fig. 5. The de Bruijn graph of order $n = 3$.

the computational prowess to prove a formula answering the question of Posthumus. Around 1970, Richard Stanley point out that this formula had been obtained previously by Camille Flye Sainte-Marie.

The corresponding problem for an alphabet with k letters seems to have been first raised and solved in [1]. The number is $k^{-n}(k!)^{n^{k-1}}$. The proof, as in the original one by Flye Sainte-Marie and de Bruijn, uses what is now called the de Bruijn graph of order n. The set of vertices is the set of words of length $n-1$ on a k letter alphabet A. The set of edges is the set of triples (au, b, ub) for $a, b \in A$ and $u \in A^{n-2}$. This graph is Eulerian since each vertex has indegree k and outdegree k. By Euler's theorem [41] it has an Eulerian cycle.[3] The label of an Eulerian cycle is a de Bruijn word. This gives an easy (and natural) proof of the existence of de Bruijn cycles. As for their enumeration, the first proofs operate by induction, going from the graph of order n to the graph of order $2n$. The formula has been generalized by a theorem sometimes called the BEST theorem proved independently by van Aardenne-Ehrenfest and de Bruijn [1] and by Smith and Tutte [123]. It enumerates all Eulerian cycles in an Eulerian graph G with n vertices $v_1, v_2, \ldots, v_n$ by the formula

$$t_i(G) \prod_{j=1}^{n} (d(v_j) - 1)! \tag{1}$$

where $t_i(G)$ is the number of spanning trees rooted at v_i and with edges oriented towards v_i and $d(v)$ is the outdegree of the vertex v. The role played by spanning trees in this formula comes from the nice combinatorial property associating with each Eulerian cycle the tree of edges used to leave a vertex for the last time. As an example, Fig. 6 represents the two possible spanning trees rooted in bb in the de Bruijn graph of order 3. Following the Eulerian path starting and ending at the root, we obtain the two possible de Bruijn words

$$aaababbb \text{ and } abaaabbb.$$

The introduction of the number of spanning trees of a graph and its characterization in terms of the adjacency matrix of the graph is itself much older, since it appears with the work of Kirchhoff on electrical networks [64]. The number of spanning trees with a given root in a graph is computed by a determinant related to the adjacency graph of the matrix. For example, let us consider the de Bruijn graph of order 3 (Fig. 5). Let us consider the matrix $M = D - A$ where D is the diagonal matrix of degrees and A is the adjacency matrix (we do not take loops into

[3] A graph is *Eulerian* if there is a cycle going through all edges exactly once. Euler's theorem states that a graph is Eulerian if and only if it is connected and every vertex has the same indegree and outdegree. This theorem, published in 1736, can be considered the first theorem of graph theory. It is supposed to have been motivated by the puzzle of the seven bridges on the Pregel, in the city of Königsberg, now Kaliningrad.

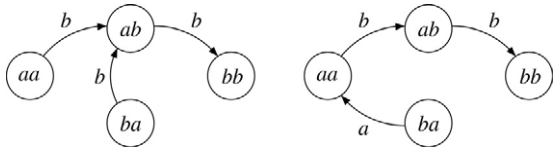

Fig. 6. The two spanning trees of de Bruijn graph of order $n = 3$ with root bb.

account). We have

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, M = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 2 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

Then each principal minor of order 3 of this matrix has a determinant equal to 2, which is precisely the number of spanning trees with a given root. This is what Tutte calls the *Matrix-Tree theorem* [137]. It appears first in Borchardt in 1860 [15]. The derivation of the formula giving the explicit number of de Bruijn cycles from Formula (1) using the determinant was carried out in [28] and, in a more general way, in [65].

The existence of a de Bruijn cycle (as we continue to call them) was actually rediscovered several times. In particular, Good [51] used the Eulerian graph method to prove the existence of a de Bruijn cycle. Earlier, Mantel [87] seems to have been the first one to use, for a prime value of k a linear recurrence relation corresponding to a primitive polynomial[4] of degree n over the field with k elements to define a de Bruijn cycle of length k^n. For example, for $k = 2$ and $n = 3$, using the relation $u_n = u_{n-2} + u_{n-3}$ corresponding to the primitive polynomial $1 + x + x^3$, we obtain from the initial value $u_0 = 0, u_1 = 0, u_2 = 1$ the sequence of length 7 (0, 0, 1, 0, 1, 1, 1). Adding a final 0 gives a de Bruijn cycle. This construction was rediscovered later in [112]. The computation of a de Bruijn cycle in this form is often called a shift-register sequence, as a reference to the implementation of the recurrence relation in hardware using a register (see [50] for a general reference and [68] for a recent one).

4.2. Lyndon words

A *Lyndon word* is a primitive word that is minimal in its conjugacy class.[5] These words were introduced under the name *standard lexicographic sequences* in [82]. For example, the list of Lyndon words of length 6 on the alphabet $\{a, b\}$ reads

$$aaaaab, aaaabb, aaabab, aaabbb, aababb, aabbab, aabbbb, ababbb, abbbbb.$$

The number of Lyndon words of length n on k symbols is, of course, $M(n, k)$. One of the basic properties of Lyndon words is that any Lyndon word x that is not a letter can be written $x = yz$ where y, z are Lyndon words with $y < z$. This factorization is not unique since, for example $(a)(abb) = (aab)(b)$ but there is a unique one, called *standard*, that has the following property. If $x = (y, z)$ is a standard factorization, and t is a Lyndon word such that $t \leq z$, then (x, t) is

[4] A *primitive* polynomial over a finite field is the minimal polynomial of a generator of the multiplicative group formed by the non-zero elements of the field.

[5] The *conjugacy class* of a word is the set of all its circular shifts. The minimality is understood with respect to some fixed lexicographic order.

standard. It can be shown (see [74]) that the standard factorization also corresponds to the choice of the longest possible second term.

Lyndon words give rise to commutators through a process of iterated dichotomy, using the notion of standard factorization. For example, the Lyndon word $aababb$ with standard factorization $(a)(ababb)$ gives rise to the commutator $[a, [[a, b], [[a, b], b]]]$. These commutators can be interpreted either as elements of the free group with $[xy] = xyx^{-1}y^{-1}$, or as elements of the free Lie algebra with $[xy] = xy - yx$. In both cases, Lyndon words give rise to a basis of some algebra.

A famous theorem concerning Lyndon words asserts that any word w can be factorized in a unique way as a nonincreasing product of Lyndon words, i.e., written $w = x_1x_2 \dots x_n$ with $x_1 \geq x_2 \geq \cdots \geq x_n$. This theorem has actually an imprecise origin. It is usually credited to Chen, Fox and Lyndon, following the paper of Schützenberger [120] in which it appears as an example of a factorization of free monoids. Actually, as pointed out to one of us by D. Knuth in 2004, the reference [21] does not contain explicitly this statement. However, it can be recovered from their results.

There is a surprising connection with de Bruijn cycles, that was discovered in 1978 by Fredericksen and Maiorana [45]. The concatenation in increasing order of Lyndon words of length dividing n is, for any $n \geq 1$, a de Bruijn word (which is actually the first one in lexicographic order and coincides with the word produced by Martin [90]). For example, for $n = 4$, we obtain

$$a\,aaab\,aabb\,ab\,abbbb.$$

Since Lyndon words can be generated efficiently, this characterization also provides a linear-time algorithm that requires only logarithmic additional space for computing one de Bruijn word. Knuth gives an account of the relationship between Lyndon words, shift-register sequences and de Bruijn words (see [68]).

5. Words in algebra

Abstract algebra is, in some sense, the birthplace of words. Indeed, basic algebraic objects, such as free groups for example, are defined explicitly using words. This has motivated, as we shall see here, the study of many important notions on words, such as Nielsen transformations for example.

5.1. Post correspondence problem

The concept of computability was formalized with the works of Gödel [49] in 1931, Turing [136], and Post [101] in 1936. The definition of computability uses one of the equivalent formalisms of recursive functions, Turing machines or other formal systems. Cantor's diagonal argument is used to prove that some sets are recursively enumerable without being recursive, thus providing an undecidable problem. This allows one to prove that the problem of whether a Turing machine will eventually halt is undecidable. Other problems are then shown to be undecidable by *reducing* them to the halting problems. This is done by showing that their solution would imply a solution of the halting problem for Turing machines. Among these, the *Post correspondence problem* consists in answering the following question: given two substitutions $f, g : A^* \to B^*$, does there exist a word $x \in A^*$ such that $f(x) = g(x)$? This problem was proved by Post [102] to be undecidable.

The proof consists, as for all undecidability results, in showing that it is possible to simulate a Turing machine using the data of the problem, in this case a pair of morphisms on words. Many problems on words can in turn be proved undecidable by reduction to the Post correspondence problem. For example, the undecidability of the equivalence of context-free grammars is easily shown in this way.

The decidability of the Post correspondence problem for small cardinalities of the alphabet has been a stimulating challenge. Amusingly, it was actually proposed by Hopcroft and Ullman in 1969 [61] as an exercise (Exercise 14.3 on page 231) to prove the decidability of the problem for $\mathrm{Card}(A) \leq 3$. The case $\mathrm{Card}(A) = 2$ was proved to be decidable in 1982 by Ehrenfeucht et al. [35]. On the other side, Matiyasevich and Senizergues have proved that the case $\mathrm{Card}(A) = 7$ is undecidable. The cases $3 \leq \mathrm{Card}(A) \leq 6$ are, to our knowledge, still open.

5.2. *Word problems in groups*

The combinatorial theory of groups using the representation of group elements by words appeared relatively late in the theory of groups. It is considered to begin with the papers that Walther Franz Anton von Dyck (1856–1934) published in 1882 and 1883 [33,34], while the idea of permutation groups appears in the work of Lagrange on polynomial equations in 1771. The word "group" itself appears in the 'Mémoire sur les conditions de résolubilité des équations par radicaux' by Galois in 1831 (see [126]).

The free group on a set A of generators is formed of the words on the alphabet $A \cup \bar{A}$ of the generators and their inverses which have no occurrences of factors of the form $a\bar{a}$ or $\bar{a}a$ for $a \in A$. These are called *reduced words*, and any word can be reduced to a unique reduced word by using a sequence of cancellations of factors $a\bar{a}$ or $\bar{a}a$ for $a \in A$. The identity of the group is the empty word ε. The term *Dyck language* for the class of the empty word appears for the first time in the paper of Chomsky and Schützenberger in 1963 [22]. The term *Dyck language* is now used both for the two-sided case as above, and for the one-sided case, which only considers cancellations of factors of the form $a\bar{a}$. In this case, the words of the Dyck language are well-formed systems of parentheses, with as many sorts of parentheses as there are letters in the alphabet. As we shall see in Section 6.2, these words are in bijection with trees.

The definition of a group by generators and relations in the form $G = \langle A \mid R \rangle$ which defines the group G as formed of the quotient of the free group over the generators $a \in A$ by the normal subgroup generated by the relators $r \in R$ is known as Dyck's theorem [125], although the name of Dyck is strangely not mentioned in the book which is the first systematic reference of Combinatorial Group theory, by Magnus et al. [86] and neither in the part of M. Hall's book on group theory [57] dedicated to free groups. Stillwell makes the interesting observation that

> The explanation of relations in terms of normal subgroups and quotients suggests a reconstruction of combinatorial group theory in more conventional algebraic terms. This can indeed be done, including the definition of free groups themselves, but it proves to be an object lesson of the impotence of abstract algebra. All substantial theorems in combinatorial group theory still require honest toil with words and relations and the best labor-saving device seems to be the topological interpretation (...), rather than algebra.

An important question concerning groups presented by generators and relations is the *word problem*. It consists in answering the question whether $u = v$ modulo the defining relations for two elements u, v of the free group or semigroup. It was first proved by Post [103] and by

Markov[6] [89] simultaneously in 1947 that the question concerning semigroups is undecidable. This problem had also been studied by Thue in two papers in 1910 and 1914 [130,132] in which he was able to solve some particular cases positively. The idea that the problem could be undecidable had perhaps come to his mind. Indeed, as pointed out by Büchi in the book [17] published after his death and edited by D. Siefkes, Thue suggests the possibility that a mathematical problem can be algorithmically unsolvable. This sentence from his 1910 paper [130], reproduced in the historical note of Steinby and Thomas [124], says:

> A solution of this problem in the most general case may perhaps be connected with insurmountable difficulties.

Stillwell [125] points out the anteriority of Tietze [133], who claimed informally in 1908 the nonexistence of an algorithm to decide the isomorphism of two finitely presented groups (this was actually proved formally by Rabin in 1958 [109]). What is now called a *Thue system* is actually the same as a semigroup presentation $\langle A \mid R\rangle$ with a set A of generators and a set R of relators which are pairs (u, v) of words on A. The problem in groups originated in the "contractibility problem" for curves on a surface. It consists in deciding whether a given closed curve contracts to a point. This problem had already been considered by Jordan in 1866 and was solved by Dehn in 1912 with a method now called Dehn's algorithm, which applies to one relator group. It was proved by Novikov in 1955 that the general problem is undecidable [99].

One of the notions which may, in our view, be considered as a historical landmark of combinatorics on words is that of a *Nielsen transformation*. This is the transformation used by Nielsen to prove that any subgroup of a free group is free and provides a procedure to compute a basis [98]. Given a set $X = \{x_1, x_2, \ldots, x_n\}$ of elements of a free group F, the Nielsen transformations on X are the compositions of three types of elementary transformations:

(i) replace some x_i by x_i^{-1}.
(ii) replace some x_i by $x_i x_j$ for some $j \neq i$.
(iii) erase x_i if $x_i = 1$.

These transformations do not change the subgroup H generated by X and allow one to transform any set X into a set which is Nielsen reduced (in the sense that no element can completely cancel in a product of elements of the set and their inverses) and therefore is a basis of the subgroup H. For example, the set $x = aa$, $y = ab$, $z = ba$ is not reduced since $zx^{-1}y = bb$, a product in which x disappears completely. A Nielsen reduced form is obtained by changing z into $zx^{-1} = ba^{-1}$.

The notion of a Nielsen transformation is a fundamental one since these transformations (applied to the generating set of a finitely generated free group F) generate the automorphism group of F. It is remarkable that there is also a link with Sturmian words, as observed by Wen and Wen [142]. In this paper, it is shown that the positive automorphisms of the free group on two generators are exactly the morphisms that preserve Sturmian words (see chapter 2 of [76]). As an illustration, the Fibonacci morphism $a \mapsto ba$, $b \mapsto a$ of Section 3.2 is clearly a Nielsen transformation.

A completely different method to prove Nielsen's theorem was found by Schreier in [118]. It is based upon the notion of a *Schreier system*. It is just a prefix-closed subset of the free group

[6] A.A. Markov, born on September 1903, is the son of the probabilist (1856–1922) with the same initials after whom Markov chains are named.

F on a set of generators A (viewed as the set of reduced words on $A \cup \bar{A}$). It can be shown that for any subgroup H of F, it is possible to choose a system S of representatives of the left cosets which is a Schreier system. Then, H is freely generated by the elements of the form uav^{-1} for all $u, v \in S$ and $a \in A$ such that $Hua = Hv$. This proof is certainly much more simple than the one obtained by Nielsen's method. It gives however less information on the reduction process. Algorithmically, if one starts with a subgroup H generated by a set X, an algorithm to obtain by Schreier's method a set of free generators consists in building a Schreier system for H from the set X. This amounts to applying an algorithm now known as the Todd–Coxeter algorithm, proposed by Todd and Coxeter in 1936 [134]. This algorithm gives no information on the possible overlaps among the words of X. In this sense it gives less information and the later developments of combinatorial group theory, like the theory of small cancellation for example (see [83]), use Nielsen's method. On the other hand, it is worth mentioning that Schreier's method is really an ancestor of the construction of a deterministic automaton. This point of view is systematically adopted in some presentations of group theory, like the book of Sims [121]. It is also worth mentioning that the interaction between group theory and automata has known many interesting developments, notably with the theory of automatic groups (see [37] for an introduction).

It was observed very early that Nielsen's theorem does not hold in free semigroups. An abstract characterization of free semigroups was proposed by Levi in [72]. Later, Schützenberger [119] characterized a free subsemigroup H of a free semigroup F by the following property: for $x \in F$ and $y, z \in H$, $xy, zx \in H$ imply $x \in H$. This interesting characterization, which directly implies several closure properties of this family of subsemigroups, was independently discovered several times (in particular by Paul Moritz Cohn in [25]). It was shown by Kolotov [70] that the property above is also satisfied by free subalgebras of a free algebra. The converse does not hold, however (see chapter 9 of [76]).

5.3. Burnside groups

In 1902, Burnside [18] raised the following famous problem: "Is every group with a finite number of generators and satisfying an identical relation $x^n = 1$ finite?". The problem was solved negatively in 1968 by Adjan and Novikov (see [2]). An interesting point for us is that the proof uses at its end the existence of an infinite cube-free word. It rests on a construction due to Aršon [6] who built in 1937 an infinite cube-free word independently of Thue's work (once again). The binary word he gives is precisely the Thue–Morse sequence.

The Burnside problem for semigroups has been also studied, starting with a paper of Green and Rees in 1952 [54]. They proved that a finitely generated semigroup satisfying the identity $x^{r+1} = x$ is finite provided any finitely generated group satisfying the identity $x^r = 1$ is finite. In particular the free idempotent semigroup, i.e., satisfying the identity $x^2 = x$ is finite. Actually, the case $r = 1$ admits another natural generalization with the identity $x^{n+1} = x^n$ for $n \geq 1$. The existence of infinite square-free words shows that the corresponding semigroups are not finite for $n \geq 2$ but Brzozowski et al. [16] raised the question of whether the classes of the congruence generated by the relations $x^{n+1} = x^n$ are regular, i.e. recognizable by a finite automaton. The case $n = 2$ was solved positively in [16]. Even, more generally, one may raise the same problem for the congruence generated by the relations $x^{n+m} = x^n$ for some $n, m \geq 1$. Many cases have been solved, including a proof by V. Guba that positively settles the case $n \geq 3$, $m \geq 1$ (see [79]). The case $n = 2$ and $m = 1$ is still unsolved.

5.4. Equations in words

There is a problem dual to that of group or semigroup presentations. Instead of looking at the groups or semigroups satisfying a set of relations between the generators, one may look, in a given group or semigroup, for the set of elements that satisfy a given set of identities. For example, in a group, the pairs of elements x, y that satisfy the equation $xz = zy$ for some z are all pairs of conjugate elements. These *equations* have been considered in particular for unknowns or *variables* which are to be substituted by words, either in the free group or in the free semigroup (see [83] for an introduction to equations in free groups).

The earliest reference to equations in groups or semigroups seems to be relatively recent. Neumann [96] has studied the problem of adjunction of elements to groups, in analogy to the corresponding problem in the theory of fields, asking, for example, whether one can find a solution of the equation $x^2 = g$ for some element g of the group.

In this context, many results have the following form: if $x_1, x_2, \ldots$ are elements of a free group which satisfy an equation of some type, then they all belong to a cyclic subgroup. For example, Lyndon and Schützenberger [84] showed that if x, y, z are elements of a free group such that $x^m y^n z^p = 1$ with $m, n, p \geq 2$, then x, y, z are all contained in a cyclic subgroup.

The first systematic studies of equations in words (i.e., in the free semigroup) appear in the 1970's with the volumes of Lentin [71] and, on the Russian school side, of Hmelevskii [60]. Lentin's book [71] contains many results and collects references to earlier work. An account is given in Combinatorics on Words [74] in the chapter written by C. Choffrut. In particular, it contains a study of *quadratic equations*, i.e., such that each variable has at most two occurrences. This subject has a close connection with the study of *quadratic sets* of words (see [83]). These sets are such that each variable has at most two occurrences in total in the set of words. The idea goes back to Nielsen and is connected with the classification of compact 2-manifolds [97].

One of the important results is Makanin's theorem proving the decidability of the existence of a solution for a finite system of word equations (see chapter 12 of [76] for a full exposition).

A very useful tool in the study of equations in words is the *periodicity lemma* which describes what happens when words have several periods. In its optimal form, it is due to Fine and Wilf [43]. It says that if the words x^ω and y^ω have a common prefix of length $|x| + |y| - \gcd(|x|, |y|)$, then x and y are actually powers of a common word.[7] The bound is the best possible. Indeed, any pair of consecutive Fibonacci words x, y is such that x^ω and y^ω have a common prefix of length $|x| + |y| - \gcd(|x|, |y|) - 1$. For example, the words $x = abaab$ and $y = abaababa$ are such that x^3 and y^2 have a common prefix of length $5 + 8 - 2 = 11$. It is interesting from the historical point of view that we adopt here, to note that the periodicity lemma is closely related to Euclid's algorithm (as noted in [69]). In particular, the worst case of Euclid's algorithm is given by pairs of consecutive Fibonacci numbers, as noted by Lamé in 1845 (see [67]).

Another point of view is the decidability of the theory of a family of groups. The (*elementary*) *theory* of a group G is the set of all first order sentences in the language of group theory which are true in G. These sentences are build upon atomic formulas which are equations or their negations, i.e. of the form $u(X_1, X_2, \ldots, X_n) = 1$ or $u(X_1, X_2, \ldots, X_n) \neq 1$ where u is a word in the n variables $X_1, X_2, \ldots, X_n$.

The theory of abelian groups is decidable by the well-known theorem of Presburger [104], while the theory of free non-abelian semigroups is undecidable by a theorem of Quine [108]

[7] We write x^ω for the infinite repetition $xxx\ldots$.

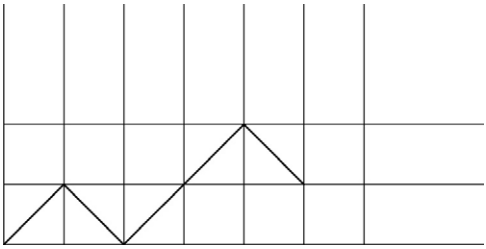

Fig. 7. A path labeled *abaab*.

showing that one can encode the theory of integers with addition and multiplication (known to be undecidable by Gödel's theorem) in the elementary theory of the free semigroup on two generators. In contrast, the elementary theory of the free group has been conjectured long ago to be decidable by Tarski. This conjecture was recently announced to be true by Kharlampovich and Myasnikov [63].

Fragments of the theory of free groups and semigroups are decidable. In particular, Makanin's result allows us to prove the decidability of the existential theory of the free semigroup.

6. Path encoding

We have already met words encoding paths in symbolic dynamics in Section 3. In a somehow different spirit, words can be used to encode paths and thus be used for path enumeration. We shall see here how this is linked with classical results of probability theory and with the description of knots using so-called Gauss codes. For a modern exposition of path encodings, see the chapter by Poulalhon and Schaeffer in [77].

6.1. Paths in the plane

The enumeration of paths in the plane was studied early in connection with probability theory. The first significant result, a solution to the *ballot problem*, is due to Bertrand in 1887 (the note of Bertrand [12] was immediately followed by two other notes, by Émile Barbier [7] and Désiré André [5]. See [32] for more on the history of this subject). It says [42][8]

> Suppose that, in a ballot, candidate P scores p votes and candidate Q scores q votes, where $p > q$. The probability that throughout the counting there are always more votes for P than for Q equals $(p - q)/(p + q)$.

The proof of this result consists in enumerating the paths going from the origin to the point (x, y) using up and down diagonal moves and staying in the first quadrant as on Fig. 7, with $x = p + q$, $y = p - q$. The proof uses the *reflexion principle* credited to Désiré André in 1887. This principle says that, given two points A, B in the first quadrant, the number of paths going from A to B which touch or cross the x-axis is equal to the total number of paths from A' to B

[8] Émile Barbier gives the following abstract:"M. Bertrand a trouvé que, si deux candidats A et B ont obtenu m et n voix dans un scrutin de ballotage, $\frac{m-n}{m+n}$ est la probabilité que, pendant le dépouillement du scrutin, le nombre de voix de A ne cessera pas une seule fois de surpasser celles de son concurrent." This is the same statement with A, B, m, n replaced by P, Q, p, q.

where A' is symmetrical to A with respect to the x-axis.[9] Using this lemma, it is easy to prove the ballot theorem. Indeed, let $N_{x,y} = \binom{x}{p}$. The number of paths from the origin to (x, y) that never touch the x-axis (except for the origin) is the same as the number of paths going from the point $(1, 1)$ to the point (x, y) which never touch the x-axis. By the reflexion lemma, this number is $N_{x-1,y-1} - N_{x-1,y+1} = y/x N_{x,y}$, and thus the probability of the paths staying above the x-axis is y/x.

This result can be used to count the number f_{2n} of paths from the origin to the point $(2n, 0)$, or equivalently the words of length $2n$ of the Dyck language D, which is the set of words over $\{a, b\}$ such that both a and b occur n times and all the prefixes have more a's than b's (we have already met the Dyck language in Section 5.2). Indeed, such a word ends with a b and their number is also the number of paths from the origin to the point $(2n - 1, 1)$ which stay above the x-axis which, by the ballot theorem, is equal to

$$f_{2n} = \frac{1}{2n-1}\binom{2n-1}{n-1} = \frac{1}{n}\binom{2n-2}{n-1}. \tag{2}$$

Let us remark that since $D = aD^*b$, the number of words of D^* of length $2n$ is

$$u_{2n} = \frac{1}{n+1}\binom{2n}{n}. \tag{3}$$

We shall soon meet these numbers, known as *Catalan numbers*, once more, with a completely different proof of this formula using power series. It is of interest to remark that another combinatorial proof of Formula (3) is possible. It goes along the following lines. For a word w on $\{a, b\}$, let $\varphi(w)$ be the difference between the number of occurrences of a and of b. Let L be the set of words w such that $\varphi(w) = -1$ and $\varphi(u) \geq 0$ for any proper prefix u of w. It is easy to see that $L = D^*b$, or in other terms that the words of L are those of D^* with an additional b at the end. The set L is actually the Lukasiewicz language used to denote expressions in polish notation with a as a symbol of operand and b as a symbol of operator. Then, as first observed explicitly by Raney [111], any word w such that $\varphi(w) = -1$ has exactly one conjugate in L, whence Formula (3).

6.2. Words and trees

A number of early contributions to combinatorics on words appear in the context of coding a nonlinear structure such as a tree by a linear one which is a word. We shall see another example in the next section with the coding of paths on a curve.

The coding of trees by words has its roots in the need for writing an algebraic term, which is actually a tree, as a word. This appeared as a necessity at the beginning of formal logic. It was also a concern for formulas of chemistry. It was in this context that the *polish notation*, credited to Lukasiewicz [81], was introduced. It allows us to encode a binary tree without using parentheses. This is a small combinatorial miracle not always given its real value. See again the chapter by Poulalhon and Schaeffer in [77].

The enumeration of trees is a subject with an interesting history which gave rise to a flood of papers in the *Journal de Mathématiques Pures et Appliquées* during the years 1838–1839.

[9] Actually André does not state the reflexion principle in this form in his paper. He only states it for the paths originating in $A = (1, 1)$, which is the case needed for the proof of the ballot theorem. Also, his proof does not use a reflexion but a conjugacy.

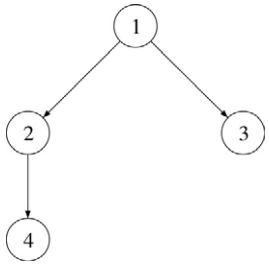

Fig. 8. A tree with Prüfer code 1, 2, 1.

Catalan [19] is credited for the formula which gives the number of binary trees with n internal vertices as the *Catalan number*

$$T_n = \frac{1}{n+1}\binom{2n}{n}.$$

This is also the number of planar trees with $n+1$ vertices and the number of (one-sided) Dyck words with $2n$ parentheses. Actually, as indicated in [66], the formula goes back to Euler. Binet indicated in [13] that Euler had communicated the formula to Segner, although he did not publish a proof. The object of the enumeration at that time was not binary trees, but the equivalent notion of a triangulation of a polygon. The problem was considered by a number of analysts including Rodrigues, Lamé and Catalan who used all their skills to transform the nonlinear recurrence

$$T_{n+1} = \sum_{i=0}^{n} T_i T_{n-i} \tag{4}$$

into the simple

$$T_{n+1} = \frac{2n+2}{n+2} T_n.$$

The paper of Binet uses the generating series $T(z) = \sum_{n\geq 0} T_n z^n$ to derive the solution via the formula

$$T(z) = \frac{1}{2}(1 - \sqrt{1-4z}).$$

Another classical result obtained early on in the enumeration of trees, is the formula giving the number of labeled trees with n vertices as n^{n-1}. The formula itself was discovered by Carl Wilhelm Borchardt (1817–1880) [15] and was known to Cayley [20]. The derivation of this formula through the bijection now known as the Prüfer code (see also Fig. 8) was obtained independently much later [106].

6.3. Gauss codes

A closed curve in the plane is *normal* if it has only finitely many self-intersections and these are transverse double points. Label the intersections of such a curve with distinct symbols from an alphabet A. The *Gauss code* of the curve is the word obtained by proceeding along the curve and noting each crossing point label as it is traversed. The word obtained is really a conjugacy class. For example, the Gauss code of the curve of Fig. 9 is $abba$. Gauss codes were introduced by Carl

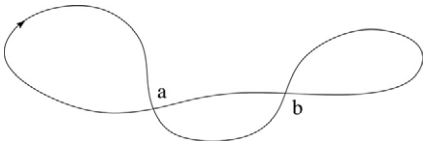

Fig. 9. A closed curve.

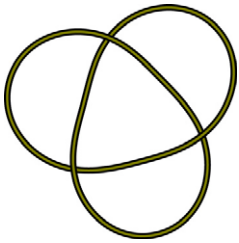

Fig. 10. The trefoil knot.

Friedrich Gauss in 1838 (see [48]). He observed that the distance between the two occurrences of each symbol in such a code is always an even integer. This property, however, is not characteristic. Several characterizations of Gauss codes have been given (see [135,91,78,115]).

Rosenstiehl gives parity properties of pairs of occurrences of the symbols, and thus completely answers the problem originally posed by Gauss. Furthermore, its algebraic formulation generalizes to words associated with closed curves on any surfaces [27].

For a history of the subject, see [55] or [116]. Today, Gauss codes are used as a succinct description of knots, with the addition of a sign to the letters to indicate whether the crossing is above or not. The website knotilus[10] proposes to draw a knot from a Gauss code. For example, typing $-1, 3, -2, 1, -3, 2$ produces the trefoil knot represented on Fig. 10.

7. Chronology

1736 Euler's characterization of "Eulerian" graphs.
1772 Jean Bernoulli's paper "Sur une nouvelle espèce de calcul" [10].
1838 Gauss' "parity condition" for Gauss codes[48].
1851 Prouhet's sequence [105].
1882 Dyck's paper "Gruppentheoretischen Studien" [33].
1892 McMahon's formula for the number of necklaces [85].
Poincaré's first volume of "Les Méthodes Nouvelles de la Mécanique Céleste" [100].
1894 C. Flye Sainte-Marie enumerates binary de Bruijn sequences [44].
1898 Hadamard's "Les surfaces à courbure opposées et leurs géodésiques" [56].
1902 Burnside states the finiteness problem for groups of finite exponent [18].
1906 Axel Thue's first paper proving the existence of an infinite square-free word [129].
1912 Axel Thue's second paper introducing the Thue–Morse sequence and overlap-free words [131].
1921 Morse's paper "Recurrent geodesics on a surface of negative curvature" [92].
Nielsen's paper proving that finitely generated subgroups of free groups are free [98].

[10] http://srankin.math.uwo.ca/cgi-bin/retrieve.cgi/html/start.html

1927 Birkhoff's book "Dynamical Systems" [14].
van der Waerden's theorem [138].
1930 Ramsey's theorem [110].
1936 Erdös and Turan's conjecture [38].
Post's definition of computability [101].
Todd–Coxeter algorithm [134].
1937 Witt's paper on free Lie algebras [143].
Morse: Lectures on symbolic dynamics [93].
Aršon's infinite square-free word [6].
1938 Morse and Hedlund's first paper on symbolic dynamics [94].
1940 Morse and Hedlund's second paper on symbolic dynamics: Sturmian sequences [95].
1946 Enumeration of binary de Bruijn words by de Bruijn [29].
1947 Post proves the unsolvability of the word problem in semigroups [103].
1954 Lyndon's paper "On Burnside problem I" [82].
1955 Novikov proves the unsolvability of the word problem in groups [99].
1963 Chomsky and Schützenberger's algebraic theory of context-free languages [22].
1965 Schützenberger's theorem on factorizations of free monoids [120]
1968 Adjan and Novikov's solution of Burnside's problem [2].

Acknowledgments

Several people helped us with valuable suggestions and information. N.G. de Bruijn has kindly given detailed information about the genesis of what are now called the de Bruijn words. Thanks are due to Donald E. Knuth for his interest and comments (we used extensively the historical notes included in "The Art of Computer Programming"). J.-P. Allouche, V. Diekert, J. Shallit and G. Skordev helped us in particular with filling gaps in the bibliography. J. Shallit also has improved our English. P. Séébold has corrected many mistakes, and C. Reutenauer has suggested several improvements. The anonymous referee made a great number of concrete suggestions for improvements.

References

[1] Tanja van Aardenne-Ehrenfest, Nicolaas Govert de Bruijn, Circuits and trees in oriented linear graphs, Simon Stevin 28 (1951) 203–217.
[2] Sergei I. Adjan, The Burnside Problem and Identities in Groups, Springer, 1979.
[3] Jean-Paul Allouche, Jeffrey Shallit, The ubiquitous Prouhet–Thue–Morse sequence, in: C. Ding, T. Helleseth, H. Niederreiter (Eds.), Sequences and their Applications, Proceedings of SETA'98, Springer Verlag, 1999, pp. 1–16.
[4] Jean-Paul Allouche, Jeffrey Shallit, Automatic Sequences, Cambridge University Press, 2003.
[5] Désiré André, Solution directe du problème résolu par M. Bertrand, Comptes Rendus Acad. Sci. Paris 105 (1887) 436–437.
[6] Solomon Efimovitch Aršon, Proof of the existence of infinite asymmetric sequences, Mat. Sb. 44 (1937) 769–777 (in Russian).
[7] Émile Barbier, Généralisation du problème résolu par M.J. Bertrand, Comptes Rendus Acad. Sci. Paris 105 (1887) 407.
[8] Dwight R. Bean, Andrzej Ehrenfeucht, George McNulty, Avoidable patterns in strings of symbols, Pacific J. Math. 85 (1979) 261–294.
[9] M.V. Bebutov, On dynamical systems in the space of continuous functions, Bull. Mos. Gos. Univ. Math. 2 (1940).
[10] Jean Bernoulli, Sur une nouvelle espèce de calcul, in: Recueil pour les Astronomes, vol. 1, Berlin, 1772, pp. 255–284.

[11] Jean Berstel, Axel Thue's work on repetitions in words, in: P. Leroux, C. Reutenauer (Eds.), Séries Formelles et Combinatoire Algébrique, Publications du LaCIM, Université du Québec à Montreal, 1992, pp. 65–80.
[12] Joseph Bertrand, Solution d'un problème, Comptes Rendus Acad. Sci. Paris 105 (1887) 369.
[13] Jacques Binet, Réflexions sur le problème de déterminer le nombre de manières dont une figure rectiligne peut être partagée en triangles au moyen de ses diagonales, J. Math. Pures. Appl. 4 (1839) 91–94.
[14] George D. Birkhoff, Dynamical Systems, in: Amer. Math. Soc. Colloq. Publi., vol. 9, Amer. Math. Soc., 1927.
[15] Carl Wilhelm Borchardt, Ueber eine der Interpolation entsprechende Darstellung der Eliminations-Resultante, J. Reine Angew. Math. 57 (1860) 111–121.
[16] John Brzozowski, Karel Culik, A. Gabrielian, Classification of noncounting events, J. Comput. System Sci. 5 (1971) 41–53.
[17] Richard Büchi, in: D. Siefkes (Ed.), Finite Automata, their Algebras and Grammars, Springer-Verlag, 1989.
[18] William Burnside, On an unsettled question in the theory of discontinuous groups, Quart. J. Pure Appl. Math. 33 (1902) 230–238.
[19] Eugène Catalan, Addition à la note sur une équation aux différences finies, insérée dans le volume précédent, J. Math. Pures Appl. 4 (1839) 95–99.
[20] Arthur Cayley, A theorem on trees, Quart. J. Pure Appl. Math. 23 (1889) 376–378.
[21] Kuo Tsai Chen, Ralph H. Fox, Roger C. Lyndon, Free differential calculus, Ann. of Math. 68 (1958) 81–95.
[22] Noam Chomsky, Marcel Paul Schützenberger, The algebraic theory of context-free languages, in: P. Braffort, D. Hirshberg (Eds.), Computer Programming and Formal Systems, North-Holland, 1963.
[23] Elwyn Bruno Christoffel, Observatio arithmetica, Annali di Mathematica 6 (1875) 145–152.
[24] Alan Cobham, Uniform tag sequences, Math. Syst. Theory 6 (1972) 164–192.
[25] Paul Moritz Cohn, On subsemigroups of free semigroups, Proc. Amer. Math. Soc. 63 (1962) 347–351.
[26] M. Coudrain, M.-P. Schützenberger, Une condition de finitude des monoïdes finiment engendrés, Comptes Rendus Acad. Sci. Paris 262 (1966) 1149–1151.
[27] Henry Crapo, Pierre Rosenstiehl, On lacets and their manifolds, Discrete Math. 233 (2001) 299–320.
[28] Reed Dawson, Irving J. Good, Exact Markov probabilities from oriented linear graphs, Ann. Math. Stat. 28 (1957) 946–956.
[29] Nicolaas Govert de Bruijn, A combinatorial problem, Nederl. Akad. Wetensch. Proc 49 (1946) 758–764.
[30] Nicolaas Govert de Bruijn, Acknowledgement of priority to C. Fly Sainte-Marie on the counting of circular arrangements of 2^n zeros and ones that show each n-letter word exactly once, Technical Report, Technische Hogeschool Eindhoven, 1975.
[31] Leonard Eugene Dickson, History of the Theory of Numbers, vol. 2, Washington, 1920.
[32] Aryeh Dvoretsky, Theodore S. Motzkin, A problem in arrangements, Duke Math. J. 14 (1947) 305–313.
[33] Walther Franz Anton von Dyck, Gruppentheoretische Studien, Math. Ann. 20 (1882) 1–44.
[34] Walther Franz Anton von Dyck, Gruppentheoretische Studien II. Ueber die Zusammensetzung einer Gruppe discreter Operationen, über ihre Primitivität und Transitivität, Math. Ann. 22 (1883) 70–108.
[35] Andrzej Ehrenfeucht, Juhani Karhumäki, Grzegorz Rozenberg, On binary equality sets and a solution to the Ehrenfeucht conjecture in the binary case, Theoret. Comput. Sci. 21 (1982) 119–144.
[36] Samuel Eilenberg, Automata, Languages and Machines, vol. A, Academic Press, 1974.
[37] David B. Epstein, J. Cannon, D. Hold, S. Levy, M. Paterson, W. Thurston, Word Processing in Groups, Jones and Bartlett, 1992.
[38] Paul Erdös, Paul Turan, On some sequences of integers, J. London Math. Soc. 11 (1936) 261–264.
[39] Edward B. Escott, The calculation of logarithms, Quart. J. Math. 41 (1910) 147–167.
[40] Edward B. Escott, Logarithmic series, Quart. J. Math. 41 (1910) 141–156.
[41] Leonard Euler, Solutio problematis ad geometrian situs pertinentis, Comm. Acad. Sci. Imper. Petropol. 8 (1736) 128–140.
[42] William Feller, An Introduction to Probability Theory and its Applications, vol. I, third ed., Wiley & Sons, 1968.
[43] Nathan J. Fine, Herbert S. Wilf, Uniqueness theorems for periodic functions, Proc. Amer. Math. Soc. 16 (1965) 109–114.
[44] Camille Flye Sainte-Marie, Question 48, L'intermédiaire des mathématiciens 1 (1894) 107–110.
[45] Harold Fredricksen, James Maiorana, Necklaces of beads in k colors and k-ary de Bruijn sequences, Discrete Math. 23 (3) (1978) 207–210.
[46] Harry Fürstenberg, Ergodic behavior of diagonal measures and a theorem of Szemeredi on arithmetic progressions, J. Anal. Math. 31 (1977) 204–256.
[47] Harry Fürstenberg, Recurrence in Ergodic Theory and Combinatorial Number Theory, Princeton University Press, 1981.

[48] Carl Friedrich Gauss, Werke, Teubner, Leipzig, 1900.
[49] Kurt Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, Monatsh. Math. Phys. 38 (1931) 173–198.
[50] Solomon Golomb, Shift Register Sequences, Holden Day, 1967.
[51] Irving J. Good, Normal recurring decimals, J. London Math. Soc. 21 (1946) 167–169.
[52] Ronald L. Graham, Donald E. Knuth, Oren Pataschnik, Concrete Mathematics, Addison Wesley, 1988.
[53] Ronald L. Graham, Bruce Rothshild, Joel Spencer, Ramsey Theory, Wiley, 1980.
[54] James A. Green, David Rees, On semigroups in which $x^r = x$, Math. Proc. Camb. Phil. Soc. 48 (1952) 35–40.
[55] Branko Grünbaum, Arrangements and spreads, in: Conference Board of the Math. Sciences Regional Conf. Ser. in Math., vol. 10, Amer. Math. Soc., 1972.
[56] Jacques Hadamard, Les surfaces à courbures opposées et leurs géodésiques, J. Math. Pures Appl. 4 (1898) 27–73.
[57] Marshall Hall, Theory of Groups, Chelsea, 1959.
[58] Godfrey Harold Hardy, Edward Maitland Wright, An Introduction to the Theory of Numbers, Oxford University Press, 1979.
[59] Gustav A. Hedlund, Remarks on the work of Axel Thue on sequences, Normat 15 (1967) 148–150.
[60] Yu.I. Hmelevskii, Equations in free semigroups, Proc. Steklov Inst. Math. 107 (1971) (Amer. Math. Soc. translation 1976, p. 270).
[61] John E. Hopcroft, Jeffrey Ullman, Formal Languages and their Relation to Automata, Addison Wesley, 1969.
[62] Nathan Jacobson, Structure of Rings, in: American Math. Soc. Colloquium Publ., vol. 37, revised edition, American Math. Soc., 1964.
[63] Olga Kharlampovich, Alexei Myasnikov, Tarki's problem about the elementary theory of free groups has a positive solution, Electron. Res. Announc. 4 (1998) 101–108.
[64] Gustav Kirchhoff, Ueber die Auflösung der Gleichungen auf welche man bei der untersuchung der Lineare Vertheilung galvanischer Ströme gführt wird, Ann. Phys. Chem. 72 (1847) 497–508.
[65] Donald E. Knuth, Oriented subtrees of an arc digraph, J. Combin. Theory 3 (1967) 309–314.
[66] Donald E. Knuth, The Art of Computer Programming, Volume 1, Fundamental Algorithms, Addison Wesley, 1968, Second edition, 1973.
[67] Donald E. Knuth, The Art of Computer Programming, Volume 2, Seminumerical Algorithms, Addison Wesley, 1969.
[68] Donald E. Knuth, The Art of Computer Programming Fascicule 2 : Generating All Tuples And Permutations, Addison Wesley, 2005.
[69] Donald E. Knuth, J.H. Morris, V.R. Pratt, Fast pattern matching in strings, SIAM J. Comput. 6 (1977) 323–350.
[70] A.T. Kolotov, Free subalgebras of free associative algebras, Sibirsk. Math. Z 19 (1978) 328–335 (in Russian). English translation: Siberian Math. J. 19 (1978) 229–234.
[71] André Lentin, Equations dans les Monoïdes Libres, Gauthier-Villars, Paris, 1972.
[72] Frank W. Levi, On semigroups, Bull. Calcutta Math. Soc. 36 (1944) 141–146.
[73] Douglas Lind, Brian Marcus, Symbolic Dynamics and Coding, Cambridge University Press, 1995.
[74] M. Lothaire, Combinatorics on Words, in: Encyclopedia of Mathematics and its Applications, vol. 17, Addison-Wesley, Reading, Mass., 1983.
[75] M. Lothaire, Combinatorics on Words, Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997. Corrected reprint of the 1983 original.
[76] M. Lothaire, Algebraic Combinatorics on Words, in: Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002.
[77] M. Lothaire, Applied Combinatorics on Words, Cambridge University Press, 2005.
[78] Lazlo Lovasz, Morris L. Marx, A forbidden substructure characterization of Gauss codes, Bull. Amer. Math. Soc. 82 (1976) 121–122.
[79] Aldo De Luca, Stefano Varricchio, Finiteness and Regularity in Semigroups and Formal Languages, Springer-Verlag, 1999.
[80] Edouard Lucas, Théorie des Nombres, Gauthier-Villars, 1891, reprinted by Albert Blanchard, 1961.
[81] Jan Lucasiewicz, Aristotle's Syllogistic from the Standpoint of Modern Formal Logic, Oxford University Press, 1951.
[82] Roger C. Lyndon, On Burnside problem I, Trans. Amer. Math. Soc. 77 (1954) 202–215.
[83] Roger C. Lyndon, Paul Schupp, Combinatorial Group Theory, Springer-Verlag, 1977.
[84] Roger C. Lyndon, Marcel-Paul Schützenberger, The equation $a^m = b^n c^p$ in a free group, Michigan Math. J. 9 (1962) 289–298.

[85] Percy A. MacMahon, Application of a theory of permutations in circular procession to the theory of numbers, Proc. London Math. Soc. 23 (1892) 305–313.
[86] Wilhelm Magnus, Abraham Karass, Donald Solitar, Combinatorial Group Theory: presentation of groups in terms of generators and relations, Dover, 1966.
[87] W. Mantel, Resten van wederkerige reeksen, Nieuw Arch. Wisk. 1 (1895) 172–184.
[88] Andrei Andreievich Markov, Sur une question de Jean Bernoulli, Math. Ann. 19 (1882) 27–36.
[89] Andrei Andreievich Markov, On the impossibility of certain algorithms in the theory of associative systems, Dokl. Akad. Nauk. 55 (1947) 583–586; 58 353–356 (in Russian).
[90] Monroe H. Martin, A problem in arrangements, Bull. Amer. Math. Soc. 40 (1934) 859–864.
[91] Morris L. Marx, The Gauss realizability problem, Proc. Amer. Math. Soc. 22 (1969) 610–613.
[92] Marston Morse, Recurrent geodesics on a surface of negative curvature, Trans. Amer. Math. Soc. 22 (1921) 84–100.
[93] Marston Morse, Symbolic dynamics. Lectures at Princeton university, notes by Rufus Oldenburger, 1937.
[94] Marston Morse, Gustav A. Hedlund, Symbolic dynamics, Amer. J. Math. 60 (1938) 815–866.
[95] Marston Morse, Gustav A. Hedlund, Symbolic dynamics II: Sturmian sequences, Amer. J. Math. 62 (1940) 1–42.
[96] Bernhard H. Neumann, Adjunction of elements to groups, J. London Math. Soc. 18 (1943) 12–20.
[97] Jakob Nielsen, Die Isomorphismen der allgemeinen unendlichen Gruppe mit zwei Erzeugenden, Math. Ann. 78 (1918) 385–397.
[98] Jakob Nielsen, Om Regning med ikke kommutative Faktorer og dens Andvendelse i Gruppenteorien. Mat. Tidsskrift B, (1921) 77–94.
[99] Petr Sergeevich Novikov, On the algorithmic unsolvability of the word problem in groups, Tr. Mat. Inst. Steklova 55 (1955).
[100] Henri Poincaré, Méthodes Nouvelles de la Mécanique Céleste, vol. I, II, III, Gauthier-Villars, 1892, 1893, 1899.
[101] Emil Leon Post, Finite combinatory processes—formulation 1, J. Symbolic Logic 1 (3) (1936) 103–105.
[102] Emil Leon Post, A variant of a recursively unsolvable problem, Bull. Amer. Math. Soc. 52 (1946) 264–268.
[103] Emil Leon Post, Recursive unsolvability of a problem of Thue, J. Symbolic Logic 12 (1947) 1–11.
[104] Mojzesz Presburger, Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt, in: Comptes Rendus 1er Congr. Math. Pays Slaves, Warszawa, 1929, pp. 92–101.
[105] Eugène Prouhet, Mémoire sur quelques relations entre les puissances des nombres, C.R. Acad. Sci. Paris 33 (1851) 255.
[106] Heinz Prüfer, Neuer Beweis eines Satzes über Permutationen, Arch. Math. u. Phys. 27 (1918) 142–144.
[107] N. Pytheas Fogg, in: V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel (Eds.), Substitutions in Dynamics, Arithmetics and Combinatorics, in: Lecture Notes in Mathematics, vol. 1794, Springer-Verlag, 2002.
[108] Willard V. Quine, Concatenation as a basis for arithmetic, J. Symbolic Logic 4 (1946) 105–114.
[109] Michael O. Rabin, Recursive unsolvability of group-theoretic problems, Ann. Math. 67 (1958) 172–194.
[110] Frank Plumpton Ramsey, On a problem of formal logic, Proc. London Math. Soc. 30 (1930) 264–286.
[111] George Raney, Functional composition patterns and power series reversion, Trans. Amer. Math. Soc. 94 (1960) 441–451.
[112] David Rees, Note on a paper by I. J. Good, J. London Math. Soc. 21 (1947) 169–172.
[113] A. de Rivière, Question 48, L'intermédiaire des mathématiciens 1 (1894) 19–20.
[114] Herbert E. Robbins, On a class of recurrent sequences, Bull. Amer. Math. Soc. 43 (1937) 413–417.
[115] Pierre Rosenstiehl, Solution algébrique du problème de Gauss sur la permutation des points d'intersection d'une ou plusieurs courbes fermées du plan, C. R. Acad. Sci. Paris 283 (1976) 417–419.
[116] Pierre Rosenstiehl, A new proof of the Gauss interlace conjecture, Adv. in Appl. Math. 23 (1999) 3–13.
[117] Mark V. Sapir, Problems of Burnside type and the finite basis property in varieties of semigroups, Izv. Akad. Nauk SSSR Ser. Mat. 51 (2) (1987) 319–340, 447.
[118] Oscar Schreier, Die Untergruppen der freien Gruppen, Abh. Math. Sem. Hamburg 5 (1927) 161–183.
[119] Marcel-Paul Schützenberger, Une théorie algébrique du codage, in: Séminaire Dubreil-Pisot 1955–56, 1955. Exposé N^o. 15.
[120] Marcel-Paul Schützenberger, On a factorization of free monoids, Proc. Amer. Math. Soc. 16 (1965) 21–24.
[121] Charles Sims, Computation with finitely presented groups, Cambridge University Press, 1994.
[122] Stephen Smale, Differentiable dynamical systems, Bull. Amer. Math. Soc. 73 (1967) 747–817.
[123] Cedric A. Smith, William T. Tutte, On unicursal paths in a network of degree 4, Amer. Math. Monthly 48 (1941).
[124] Magnus Steinby, Wolfgang Thomas, Trees and term rewriting in 1910: on a paper by Axel Thue, Bull. EATCS 72 (2000) 256–269.

[125] John Stillwell, Classical Topology and Combinatorial Group Theory, Springer-Verlag, 1980.
[126] John Stillwell, Mathematics and its History, Springer-Verlag, 1989.
[127] Endre Szemeredi, On sets of integers containing no k elements in arithmetic progression, Acta Arith. 27 (1975) 199–245.
[128] Gaston Tarry, Question 4100, L'intermédiaire des mathématiciens 19 (1912) 200. Answers to the question were given by E. Barbette, E. Miot, and Welsch in vol. 20 (1913) 68–70.
[129] Axel Thue, Über unendliche Zeichenreihen, Norske Vid. Selsk. Skr. I Math-Nat. Kl. 7 (1906) 1–22.
[130] Axel Thue, Die Lösung eines Spezialfalles eines generellen logischen Problems, Norske Vid. Selsk. Skr. I Math-Nat. Kl. Chris. 8 (1910).
[131] Axel Thue, Über die gegenseitige Loge gleicher Teile gewisser Zeichenreihen, Norske Vid. Selsk. Skr. I Math-Nat. Kl. Chris. 1 (1912) 1–67.
[132] Axel Thue, Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln, Norske Vid. Selsk. Skr. I Math-Nat. Kl. Chris. 10 (1914).
[133] Heinrich Tietze, Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten, Monat. Math. Phys. 19 (1908) 1–118.
[134] J.A. Todd, H.S.M. Coxeter, A practical method for enumerating the cosets of a finite abstract group, Proc. Edinburgh Math. Soc. 5 (1936) 25–34.
[135] Leon Bruce Treybig, A characterization of the double point structure of the projection of a polygonal knot in regular position, Trans. Amer. Math. Soc. (1968) 223–247.
[136] Alan M. Turing, On computable numbers, with an application to the Entscheidungsproblem, Proc. London Math. Soc. 42 (1936) 230–265.
[137] William T. Tutte, Graph Theory, Cambridge University Press, 2001.
[138] Bartel Leendert van der Waerden, Beweis einer Baudet'schen Vermutung, Nieuw Arch. Wiskd. 15 (1927) 212–216.
[139] Bartel Leendert van der Waerden, Wie der Beweis der Vermutung von Baudet gefunden wurde, in: Abhandlungen des Mathematischen Seminars des Hanseatischen Universität Hamburg, 1965, pp. 6–15 (also published as: [How the proof of Baudet's conjecture was found, Studies in Pure Mathematics, Academic Press, 1971, pp. 251–260).
[140] Boris A. Venkov, Elementary Number Theory, Wolters-Noordhoff, Groningen, 1970.
[141] Benjamin Weiss, Subshifts of finite type and sofic systems, Monatsh. Math. 77 (1973) 462–474.
[142] Zhi-Xiong Wen, Zhi-Ying Wen, Local isomorphisms of invertible substitutions, Comptes Rendus Acad. Sci. 318 (1994) 299–304.
[143] Ernst Witt, Treue Darstellung Lieschen Ringe, J. Reine Angew. Math. 177 (1937) 152–160.
[144] A.I. Zimin, Blocking sets of terms, Mat. Sb. 119 (3) (1982) 363–375 (in Russian). English translation in Math. USSR Sbornik 47 (1984) 353–364.