Remerciements

Je remercie de tout coeur Michel Rigo qui m'a initiée aux numérations et aux langages formels et m'a transmis son goût de la recherche. Son enthousiasme permanent, son incroyable disponibilité et ses nombreux encouragements m'ont permis de mener cette aventure du doctorat à son terme.

J'adresse également de sincères remerciements à Valérie Berthé, Bernard Boigelot, Georges Hansoul, Pierre Lecomte et Jacques Sakarovitch de me faire l'honneur de faire partie de mon jury.

Durant les quelques années de ma recherche doctorale, j'ai eu l'occasion de collaborer avec d'autres chercheurs : Jason Bell, Aviezri Fraenkel, Tomi Kärki, Marion Le Gonidec et Wolfgang Steiner. À chaque fois, ces rencontres ont représenté pour moi une expérience extrêmement enrichissante et je les en remercie tous vivement.

Merci aussi à mes amis anglophones ou anglophiles, Ellen Harry, Christophe Henrard et Elizabeth L'Estrange, qui m'ont aidée par leurs réponses à mes nombreuses questions, leur relecture et même parfois traduction de cette dissertation.

J'ai eu la chance de réaliser cette recherche en tant qu'assistante au Département de mathématiques de l'Université de Liège, où il règne une chaleureuse ambiance de travail, notamment grâce à mes nombreux collègues assistants ou post-doctorants qui sont passés par là en même temps que moi. Merci à eux.

Ensuite, bien sûr, viennent la famille et les amis, et en particulier mon frère Nicolas. Par leurs attentions répétées, ils ont, eux aussi, contribué à la réalisation de ce travail. Je leur témoigne à tous ma plus grande affection.

Enfin, je remercie tout particulièrement Bernard Keris, mon compagnon. Par hasard, notre rencontre a coïncidé à peu près avec le début de ma thèse et même s'il ne s'agit pas de compétences scientifiques proprement dites, j'ai appris énormément à son contact. Il m'a soutenue et encouragée tout au long de ce processus de recherche mais aussi de construction personnelle. Je profite de la place que j'ai ici pour lui exprimer tout mon amour.

Contents

Introduction (version française)

Cette thèse se situe au carrefour de deux disciplines, proches certes mais distinctes : l'étude des numérations et la théorie des langages formels. Du côté des numérations, l'idée de base est de représenter les nombres par des *mots*, c'est-à-dire des juxtapositions de symboles, et d'ensuite étudier les propriétés arithmétiques de ces nombres en lien avec les propriétés syntaxiques de leurs *représentations*, c'est-à-dire les règles de construction des mots qui les représentent. Par exemple, dans la numération décimale, qui est celle utilisée par nous tous au quotidien, tout mot écrit sur l'alphabet $\{0, 1, \ldots, 9\}$ et ne commençant pas par 0 représente un nombre entier positif. Si la représentation se termine par 0, alors cet entier est divisible par 10, si la représentation se termine par $0, 2, 4, 6$ ou 8, alors il est pair, etc. Plus précisément, on peut voir un système de numération (sur les entiers dans un premier temps) comme une bijection $\text{rep}\colon \mathbb{N} \to L$ de l'ensemble des naturels dans un *langage*, c'est-à-dire un ensemble de mots. Ce langage, appelé *langage de la numération*, est l'ensemble des représentations valides des naturels. Chaque partie X de $\mathbb{N}$ est alors envoyée sur un sous-langage $\text{rep}(X)$ de L.

On s'intéresse assez naturellement aux parties de $\mathbb{N}$ qui correspondent à des sous-langages particulièrement élémentaires : les langages acceptés par les "machines" les plus simples de la hiérarchie de Chomsky. Ces machines sont les *automates finis* et les langages qu'elles acceptent sont dits *réguliers*. Une partie de $\mathbb{N}$ dont l'ensemble des représentations des éléments est un langage régulier sera dite *reconnaissable* pour la numération que l'on considère. On montre facilement que, pour la *numération standard en base entière* $b \geq 2$, dans laquelle un entier positif n est représenté par la suite de chiffres $c_\ell \cdots c_1 c_0$ apparaissant dans la *décomposition gloutonne*

$$n = \sum_{i=0}^{\ell} c_i b^i, \;\; c_\ell \neq 0, \;\; c_0, c_1, \ldots, c_\ell \in \{0, 1, \ldots, b-1\},$$

toute union finie de progressions arithmétiques est reconnaissable. En 1969, A. Cobham a prouvé que les seuls ensembles d'entiers reconnaissables dans toute numération en base entière sont précisément les unions finies de progressions arithmétiques [**Cob69**]. Ce résultat, dont la preuve est considérée comme difficile, est connu sous le nom de *théorème de Cobham* et constitue le point de départ de nombreuses recherches sur les possibles généralisations de celui-ci. Parmi celles-ci, on peut citer par exemple [**Sem77, Vil92a, Vil92b, BHMV94, Fab94, MV96, Bès97, PB97, Dur98, Han98, Dur02b, Bès00, Dur02a, HS03, Bel07, RW06, BB07, AB08, Dur08, BB09**]. Le théorème de Cobham a motivé notamment l'introduction des systèmes de numération non-standards et l'étude des ensembles reconnaissables d'entiers.

Il découle naturellement de cette notion de *reconnaissabilité* plusieurs types de problématiques avec notamment l'étude de la stabilité de la reconnaissabilité par opérations arithmétiques élémentaires. On sait (voir par exemple [**Ber79, Sak06**]) que la multiplication ne préserve pas la reconnaissabilité au sein d'une numération en base entière, et même plus généralement au sein d'une numération *de position*. Une numération de position est basée sur une suite strictement croissante d'entiers $U = (U_i)_{i\geq 0}$ de premier terme $U_0 = 1$ pour laquelle le quotient de deux éléments consécutifs est borné. Un entier positif n est alors représenté par la suite de chiffres $c_\ell \cdots c_1 c_0$ apparaissant dans la *décomposition gloutonne*

$$n = \sum_{i=0}^{\ell} c_i U_i,\ c_\ell \neq 0 \ \text{ et } \ \forall t \in \{0, \ldots, \ell\},\ \sum_{i=0}^{t} c_i\, U_i < U_{t+1}. \tag{1}$$

Par contre, la multiplication par une constante et l'addition, elles, sont des opérations qui préservent la reconnaissabilité pour les numérations de position dites "Pisot", c'est-à-dire les numérations de position basées sur des suites d'entiers qui satisfont une relation de récurrence linéaire dont le polynôme caractéristique est le polynôme minimum d'un nombre de Pisot [**BH97, Fro92**]. C'est donc en particulier le cas pour toute numération en base entière et la numération de Fibonacci. Ces propriétés peuvent être démontrées, par exemple, par le biais de la caractérisation logique des parties reconnaissables en termes d'ensembles définissables dans la structure $\langle \mathbb{N}, +, V_U \rangle$, où on pose $V_U(0) = U_0 = 1$ et, pour tout entier positif n, $V_U(n)$ est défini comme étant le plus petit terme U_i apparaissant dans la décomposition gloutonne (1) de n avec un coefficient non nul.

On s'est aussi attaché à chercher des conditions nécessaires et/ou suffisantes pour qu'une numération de position possède un langage de numération régulier, c'est-à-dire pour que l'ensemble $\mathbb{N}$ des entiers positifs ou nuls y soit reconnaissable [**Sha94, Lor95, Hol98**]. Une telle propriété pour un système de numération est souvent appréciée puisque dans ce cas, on peut tester en temps linéaire, grâce à un automate fini, si un mot donné est une représentation valide d'un entier ou non. Dans [**LR01**], en introduisant les *systèmes de numération abstraits*, P. Lecomte et M. Rigo ont choisi de contourner le problème en imposant *a priori* un langage de la numération régulier. En effet, par définition, un système de numération abstrait $S = (L, \Sigma, <)$ est la donnée d'un langage infini régulier L sur un alphabet totalement ordonné $(\Sigma, <)$. L'ordre sur cet alphabet induit un ordre total sur les mots du langage, appelé *ordre généalogique*. Un entier positif ou nul n est alors représenté par le $(n+1)$-ième mot du langage (0 étant représenté par le premier mot du langage). Cette notion généralise celle des numérations de position dont l'ordre sur les mots préserve l'ordre naturel des entiers et donnant lieu à un langage de la numération régulier. C'est en particulier le cas des systèmes de numération "Pisot" évoqués plus haut. Bien sûr, en procédant de la sorte, vu la grande généralité de ces systèmes, on perd *a priori* toute spécificité d'un système particulier. Mais d'un autre côté, leur intérêt réside justement dans cette généralité : on essaie plutôt de dégager les propriétés qui sont indépendantes du système de numération choisi, comme par exemple les propriétés liées à la complexité du langage de la numération.

Ainsi on arrive à la théorie des langages formels. En effet, d'un point de vue purement théorique, on peut regarder un système de numération abstrait comme un langage infini ordonné et s'intéresser aux propriétés des sous-langages de celui-ci. C'est par exemple le point de vue adopté par D. Krieger *et al.* dans [**KMR$^+$09**] en définissant et en étudiant la notion de *décimation* d'un langage. Néanmoins, même s'ils pourront éventuellement être réinterprétés différemment sous cet angle, tous les résultats de cette thèse seront présentés du point de vue des numérations.

La définition de la reconnaissabilité d'un ensemble d'entiers s'étend naturellement au contexte des numérations abstraites. Avec elle se généralisent également de nombreuses questions, analogues à celles posées dans le cadre des numérations de position. On montre par exemple que les unions finies de progressions arithmétiques sont reconnaissables pour tout système de numération abstrait [**LR01**]. Ce résultat n'est pas anodin car il a constitué une des motivations premières pour l'étude des numérations abstraites. Ainsi, M. Rigo a consacré sa thèse de doctorat à leur exploration [**Rig01a**]. Cette dissertation prend place dans la continuité de ces travaux. Nous y détaillons les résultats obtenus de plusieurs collaborations [**CRS08, CR08, BCFR09, CKR, CKR09, CLGR**].

Dans le premier chapitre, sont rappelées les notions de bases essentielles à la compréhension de cet ouvrage. Tout d'abord sont redéfinis les mots, les langages et les automates. Sont aussi rappelés quelques résultats utiles dans ce domaine. Ensuite sont introduits les systèmes de numération de position et les systèmes de numération abstraits. Enfin, sans prétendre à l'exhaustivité, nous dressons un portrait de l'état de la recherche autour de ces derniers systèmes depuis leur naissance en 2001.

Le deuxième chapitre traite de la préservation de la reconnaissabilité d'un ensemble d'entiers après multiplication par une constante au sein d'un système de numération abstrait construit sur un langage *borné*, c'est-à-dire de la forme

$$a_1^* a_2^* \cdots a_\ell^*,$$

où ℓ est un entier positif et $a_1, \ldots, a_\ell$ sont des lettres. A l'origine, l'intérêt pour cette question provient des étonnants résultats de [**LR01, Rig01b**] qui montrent que dans le cas d'un alphabet de deux lettres, c'est-à-dire le cas du langage a^*b^*, la multiplication par une constante préserve la reconnaissabilité si et seulement si cette constante est un carré impair. Notre souhait était initialement d'obtenir une extension de ce résultat liant le système de numération abstrait construit sur le langage $a^*b^*c^*$ aux cubes. Il s'est avéré par la suite qu'une telle généralisation était inenvisageable puisqu'en fait, nous montrons dans cette thèse que, dans le cas d'un alphabet de plus de deux lettres, la reconnaissabilité n'est jamais préservée par la multiplication par une constante au sein d'un tel système. Plus précisément, nous prouvons que, dans ce cas, pour toute constante entière λ, on peut toujours trouver un ensemble reconnaissable d'entiers X tel que l'ensemble correspondant λX n'est pas reconnaissable. La classe des langages réguliers se sépare en deux sous-classes selon le comportement de leurs fonctions de complexité[1] : celle des langages réguliers *exponentiels* et celle des langages réguliers *polynomiaux* [**SYZS92**]. Dans les numérations de position de type "Pisot", le langage de la numération est toujours un langage exponentiel. Par contre, un langage borné est toujours un langage polynomial. L'étude des systèmes abtraits construits sur de tels langages donne donc lieu à de nouveaux phénomènes. Les langages polynomiaux ont été caractérisés dans [**SYZS92**] : ce sont les unions finies de langages de la forme

$$xy_1^* z_1 y_2^* \cdots y_k^* z_k,$$

où k est un entier positif ou nul et $x, y_1, z_1, y_2, \ldots, z_k$ sont des mots finis. De plus, pour chaque entier positif ℓ, la fonction de complexité du langage borné $a_1^* a_2^* \cdots a_\ell^*$ est un polynôme de degré $\ell - 1$. Les langages bornés peuvent ainsi être vus comme des archétypes des langages polynomiaux. Dès lors, nous espérons que nos résultats sur les langages bornés contiennent les idées de ce qui se passe pour le cas des langages polynomiaux en général. Nous épinglons au passage quelques propriétés structurelles des langages bornés. Nous proposons notamment une caractérisation des parties reconnaissables en termes d'ensembles semi-linéaires de $\mathbb{N}^\ell$, où ℓ est le nombre de lettres du

[1] La fonction de complexité $\mathbf{u}_L : \mathbb{N} \to \mathbb{N}$ d'un langage L écrit sur un alphabet Σ compte le nombre de mots de chaque longueur de ce langage : $\mathbf{u}_L(n) = \mathrm{Card}(L \cap \Sigma^n)$.

système abstrait considéré. Nous étudions aussi l'action de la multiplication par une constante de la forme β^ℓ sur un mot $a_1^{n_1}a_2^{n_2}\cdots a_\ell^{n_\ell}$ quelconque.

Nous savons que les unions finies de progressions arithmétiques sont reconnaissables pour tout système de numération abstrait. Il est donc naturel de se demander si, étant donné un automate qui reconnaît un ensemble d'entiers, on peut décider[2] si cet ensemble est ou n'est pas une union finie de progressions arithmétiques. Dans [**Hon86**], J. Honkala répond par l'affirmative pour le cas des bases entières. Ensuite, dans [**Muc03**], A. Muchnik donne une procédure de décision pour tous les systèmes de numération de position ayant un langage de numération régulier à condition que l'addition y soit reconnaissable, c'est-à-dire que son graphe soit un langage régulier. À notre tour, nous proposons deux procédures de décision pour ce problème. Ceci fait l'objet du troisième chapitre. La première concerne les numérations de position telles que $\mathbb{N}$ y est reconnaissable et satisfaisant certaines hypothèses assez faibles. En particulier, nous englobons des cas de systèmes pour lesquels l'addition n'est pas reconnaissable. Nous nous intéressons au nombre de classes $N_U(m)$ visitées infiniment souvent par la suite réduite $(U_i \bmod m)_{i\geq 0}$, où $U = (U_i)_{i\geq 0}$ est une suite linéaire récurrente d'entiers. Si U est la base de la numération, notre procédure de décision repose sur la condition $\lim_{m\to+\infty} N_U(m) = +\infty$. Dans une section séparée, nous donnons une caractérisation algébrique de telles suites U satisfaisant cette condition. Sous certaines hypothèses du même type, nous proposons ensuite une deuxième procédure de décision dans le cas des numérations abstraites. En particulier, dans les deux cas, nous mettons en évidence quelques exemples non encore résolus jusqu'ici. Nous terminons ce chapitre par une brève discussion à propos du problème de périodicité des systèmes HD0L : étant donné un morphisme f prolongeable à partir d'une lettre a et un morphisme g, peut-on décider si le mot infini $g(f^\omega(a)) = \lim_{n\to+\infty} g(f^n(a))$ est ultimement périodique ou non ? Nous montrons que nos résultats se révèlent être un pas dans la direction de la résolution de ce célèbre problème, encore ouvert à ce jour. En effet, J. Honkala et M. Rigo ont démontré l'équivalence entre ce problème et le problème de décision qui nous intéresse dans ce chapitre mais étendu à tout système de numération abstrait [**HR04**]. Ce résultat provient du fait que les numérations abstraites sont en fait étroitement liées aux suites *morphiques*, c'est-à-dire aux mots infinis de la forme $g(f^\omega(a))$ évoqués plus haut [**RM02**].

[2] En informatique, un problème de décision est un problème ayant au moins un paramètre pouvant prendre une infinité de valeurs et auquel il convient de répondre par "oui" ou par "non". Décider un tel problème signifie qu'à partir de toute instance du problème, supposée implémentable, on peut, après un nombre fini (mais arbitrairement grand) d'opérations réalisables effectivement, répondre avec certitude au problème posé.

Dans le quatrième chapitre, nous généralisons au cas multidimensionnel un résultat d'A. Maes et de M. Rigo à propos des suites automatiques étendues aux systèmes de numération abstraits [**Rig00, RM02**]. Dans ce cas, on parle de suites S-*automatiques*. L'idée d'une telle généralisation provient naturellement de ce qui existait déjà dans le cas des bases entières. En effet, le résultat évoqué plus haut étend un théorème d'A. Cobham établissant une correspondance entre les mots b-automatiques et les mots morphiques obtenus à l'aide d'un morphisme uniforme de longueur b [**Cob72**]. O. Salon avait déjà étendu ce théorème aux mots multidimensionnels en considérant des morphismes par lesquels l'image d'une lettre est un hypercube de côté b [**Sal87a, Sal87b**]. Ainsi le théorème démontré dans ce chapitre comble la "case manquante" à ce diagramme d'extensions aux systèmes de numération abstraits et au cas multidimensionnel. Dans sa thèse de doctorat, A. Maes avait défini des mots multidimensionnels "shape-symmetric". Nous démontrons que dans le cas multidimensionnel, les mots S-automatiques correspondent aux images par un codage de mots purement morphiques "shape-symmetric". Un point crucial de la démonstration de ce théorème est de généraliser le résultat classique dans le cas unidimensionnel (voir par exemple [**Cob68, Pan83, AS03**]) selon lequel tout mot obtenu à partir d'un mot morphique en y effaçant toutes les occurrences d'une lettre déterminée est soit fini, soit morphique. Tout au long de ce chapitre, afin de rendre la présentation plus claire, nous illustrons les différents concepts introduits par de nombreux exemples.

Enfin, dans le cinquième et dernier chapitre, nous nous intéressons à la représentation des nombres réels dans le cadre général des systèmes de numération abstraits étendus à des langages quelconques, c'est-à-dire à des langages qui ne sont plus nécessairement réguliers. Le but de cette recherche était de proposer une approche unifiée à plusieurs systèmes de numération apparaissant dans la littérature [**AFS08, DT89, LR01, Lot02**]. Par exemple, les numérations en base rationnelle récemment introduites dans [**AFS08**] donnent lieu à des langages de numération non algébriques. Nous construisons, sous certaines hypothèses générales sur le langage de la numération, un formalisme pour la représentation des réels par des mots infinis limites de mots du langage de la numération. Nous illustrons ensuite ce formalisme à l'aide de trois exemples de numérations abstraites construites sur des langages non réguliers. L'un d'eux est basé sur le langage des préfixes des mots de Dyck. Dans chacun des cas, nous nous assurons que les systèmes étudiés vérifient nos hypothèses générales et, si cela est possible, nous calculons explicitement la valeur correspondant à un mot infini appartenant à l'ensemble des représentations valides.

Cette dissertation se termine avec quelques perpectives de recherches dans la continuité des travaux réalisés dans le cadre de cette thèse de doctorat.

Introduction (English version)

The framework of this doctoral dissertation encompasses two related but distinct domains: the study of numeration systems and formal language theory. From a numeration point of view, the basic approach is to represent numbers by *words*, *i.e.*, by concatenation of symbols, and then to study the arithmetic properties of these numbers in relation to the syntactical properties of their *representations*, *i.e.*, the construction rules for the words representing them. For example, in the decimal numeration system, which is the standard used in everyday life, any word written over the alphabet $\{0, 1, \ldots, 9\}$ and not beginning with 0 represents a positive integer. If the representation ends with 0, then this integer is divisible by 10, if the representation ends with $0, 2, 4, 6$ or 8, then it is even, and so on. More precisely, a *numeration system* (first for the integers) can be viewed as a bijection $\mathrm{rep}\colon \mathbb{N} \to L$ from the set of non-negative integers to a *language*, *i.e.*, a set of words. This language, which is called the *numeration language*, is the set of the valid representations of the non-negative integers. Each subset X of $\mathbb{N}$ is then mapped onto a sublanguage $\mathrm{rep}(X)$ of L.

Researchers are naturally interested in the subsets of $\mathbb{N}$ that correspond to especially simple sublanguages: those accepted by the simplest "machines" of Chomsky's hierarchy. These machines are *finite automata* and the languages accepted by them are said to be *regular*. A subset of $\mathbb{N}$ such that the representations of its elements form a regular language is said to be *recognizable* for the numeration system under consideration. It is easily shown that, for the *standard integer base $b \geq 2$ numeration system*, in which a positive integer n is represented by the sequence of digits $c_\ell \cdots c_1 c_0$ appearing in the *greedy decomposition*

$$n = \sum_{i=0}^{\ell} c_i b^i,\ c_\ell \neq 0,\ c_0, c_1, \ldots, c_\ell \in \{0, 1, \ldots, b-1\},$$

any finite union of arithmetic progressions is recognizable. In 1969, A. Cobham proved that the only sets of non-negative integers that are recognizable in all integer base numeration systems are precisely the finite unions of arithmetic progressions [**Cob69**]. This result, whose proof is considered to be difficult, is known as *Cobham's theorem* and has inspired a number of studies about the generalizations that can be drawn from it. Among these studies, one can mention, for example, [**Sem77, Vil92a, Vil92b, BHMV94, Fab94, MV96, Bès97, PB97, Dur98, Han98, Dur02b, Bès00, Dur02a, HS03, Bel07, RW06, BB07, AB08, Dur08, BB09**]. Most notably, Cobham's theorem motivated the introduction of non-standard numeration systems and the study of the recognizable sets of non-negative integers.

Several issues stem naturally from this notion of *recognizability*, notably the study of the stability of recognizability under elementary arithmetic operations. It is well known (for instance, see [**Ber79, Sak06**]) that multiplication does not preserve recognizability within any integer base numeration system, and even more generally, within any *positional* numeration system. A positional numeration system is based on an increasing sequence of integers $U = (U_i)_{i\geq 0}$ whose first term is $U_0 = 1$ and for which the quotient of two consecutive elements is bounded. A positive integer n is thus represented by the sequence of digits $c_\ell \cdots c_1 c_0$ appearing in the *greedy decomposition*

$$n = \sum_{i=0}^{\ell} c_i U_i,\ c_\ell \neq 0 \ \text{ and }\ \forall t \in \{0,\dots,\ell\},\ \sum_{i=0}^{t} c_i\, U_i < U_{t+1}. \qquad (2)$$

On the other hand, multiplication by a constant and addition are both operations that preserve recognizability for the "Pisot" numeration systems, *i.e.*, positional numeration systems based on sequences of integers satisfying a linear recurrence relation whose characteristic polynomial is the minimal polynomial of a Pisot number [**BH97, Fro92**]. In particular, this is the case for any integer base numeration system and for the Fibonacci numeration system. For example, these properties can be demonstrated thanks to the logical characterization of the recognizable sets in terms of sets definable in the structure $\langle \mathbb{N}, +, V_U \rangle$, where we set $V_U(0) = U_0 = 1$ and, for any positive integer n, $V_U(n)$ is defined to be the smallest term U_i appearing in the greedy decomposition (2) of n with a non-zero coefficient.

Other studies have focused on finding necessary and/or sufficient conditions so that a positional numeration system would have a regular numeration language, *i.e.*, so that the whole set $\mathbb{N}$ of non-negative integers would be recognizable within this system [**Sha94, Lor95, Hol98**]. Such a property for a numeration system is often desirable since, in this case, one can check in linear time, thanks to a finite automaton, whether a given word is a valid representation of a non-negative integer or not. In [**LR01**], by introducing the *abstract numeration systems*, P. Lecomte and M. Rigo chose to approach the problem from a different angle by assuming *a priori* a regular numeration language. Indeed, by definition, an abstract numeration system $S = (L, \Sigma, <)$ is given by an infinite regular language L over a totally ordered alphabet $(\Sigma, <)$. The order on this alphabet induces a total order on the words of the language, which is called *the genealogical order*. A non-negative integer n is thus represented by the $(n+1)$st word in the language (0 being represented by the first word in the language). This notion generalizes that of the positional numeration systems whose order on words preserves the natural order on integers and which have a regular numeration language. In particular, this is the case of the "Pisot" numeration systems mentioned above. Of course, by proceeding so, in view of the significant generality of these systems, we lose *a priori* any specificity of a particular system. Yet, their advantage also stems from this generality: current research on this subject strives to highlight the properties that are independent of the target numeration system, such as properties related to the complexity of the numeration language.

Thus, we reach formal language theory. Indeed, from a purely theoretical point of view, an abstract numeration system can be seen as an infinite ordered language and the properties of its sublanguages become the main interest. This is, for instance, the perspective taken by D. Krieger *et al.* in [**KMR$^+$09**] by defining and studying the notion of the *decimation* of a language. Nevertheless, even though the results of the present dissertation could be reinterpreted in another way from this angle, they will be presented from the numeration approach.

The definition of the recognizability of a set of non-negative integers extends naturally to the context of abstract numeration systems. With this definition, a number of questions can be applied as well, analogous to those asked in a positional numeration framework. It was shown, for example, that the finite unions of arithmetic progressions are recognizable within any abstract numeration system [**LR01**]. This outcome is not insignificant because it represents one of the primary motivations for the study of abstract numeration systems. So, M. Rigo devoted his doctoral dissertation to the study of these systems [**Rig01a**]. The context of the present dissertation is a continuation of this work. In this text the author will discuss the results obtained from several collaborations in detail [**CRS08, CR08, BCFR09, CKR, CKR09, CLGR**].

In the first chapter the author will recall basic notions necessary for a clear understanding of this work. First, words, languages and automata are redefined. In addition, several useful results in this domain will be recalled. Next, positional numeration systems and abstract numeration systems will be introduced. Without attempting to provide an exhaustive description, the author will portray the state of the art regarding research on the latter systems since their appearance in 2001.

The second chapter deals with the preservation of the recognizability of a set of non-negative integers under multiplication by a constant within an abstract numeration system built on a *bounded* language, *i.e.*, on a language of the form

$$a_1^* a_2^* \cdots a_\ell^*,$$

where ℓ is a positive integer and $a_1, \ldots, a_\ell$ are letters. Originally, the interest in this question arose from the surprising results of [**LR01, Rig01b**], which establish that, in the case of a two-letter alphabet, *i.e.*, the case of the language a^*b^*, multiplication by a constant preserves recognizability if and only if this constant is an odd square. Initially, our wish was to extend this result by linking the abstract numeration system built on the language $a^*b^*c^*$ to cubes. Afterwards, it turned out that such a generalization was not feasible since, in fact, we will show in this dissertation that, in the case of an alphabet containing more than two letters, recognizability is never preserved under multiplication by a constant within such a system. More precisely, we will prove that, in this case, for any integer constant λ, a recognizable set of integers X such that the corresponding set λX is not recognizable can always be found. The class of regular languages is divided into two subclasses according to the behavior of their counting functions[3]: the *exponential* regular languages and the *polynomial* regular languages [**SYZS92**]. In "Pisot" numeration systems the numeration language is always exponential. On the other hand, a bounded language is always polynomial. Therefore the study of abstract numeration systems built on such languages give rise to new phenomena. Regular polynomial languages were characterized in [**SYZS92**]: they are finite unions of languages of the form

$$xy_1^* z_1 y_2^* \cdots y_k^* z_k,$$

where k is a non-negative integer and $x, y_1, z_1, y_2, \ldots, z_k$ are finite words. Furthermore, for every positive integer ℓ, the counting function of the bounded language $a_1^* a_2^* \cdots a_\ell^*$ is a polynomial of degree $\ell - 1$. Bounded languages may thus be seen as archetypes of polynomial languages. Therefore we hope that our results on bounded languages will give an idea of what occurs in the case of polynomial languages in general. During the discussion, we will pinpoint structural properties of bounded languages. In particular, we will propose a characterization of recognizable sets in terms of semi-linear sets of $\mathbb{N}^\ell$, where ℓ is the number of letters of the target abstract numeration system. We will also study the action of multiplication by a constant of the form β^ℓ on any word $a_1^{n_1} a_2^{n_2} \cdots a_\ell^{n_\ell}$.

[3]The counting function $\mathbf{u}_L \colon \mathbb{N} \to \mathbb{N}$ of a language L over an alphabet Σ counts the number of words of each length in this language: $\mathbf{u}_L(n) = \mathrm{Card}(L \cap \Sigma^n)$.

We know that finite unions of arithmetic progressions are recognizable for any abstract numeration systems. Thus it is natural to wonder if, given an automaton that recognizes a set of integers, one can decide[4] whether or not this set is a finite union of arithmetic progressions. In [**Hon86**] J. Honkala answered positively for the case of the integer base numeration systems. Then, in [**Muc03**], A. Muchnik gave a decision procedure for all positional numeration systems with a regular numeration language, provided that the addition is recognizable therein, *i.e.*, that its graph is regular. For our part, we will propose two decision procedures for this problem. This will be the focus for the third chapter. The first procedure handles positional numeration systems in which $\mathbb{N}$ is recognizable, satisfying certain relatively weak conditions. In particular, we will incorporate systems for which addition is not recognizable. We will focus on the number of residue classes $N_U(m)$ visited infinitely often by the reduced sequence $(U_i \bmod m)_{i\geq 0}$, where $U = (U_i)_{i\geq 0}$ is a linear recurrence sequence of integers. If U is the basis of the numeration system, our decision procedure requires the condition $\lim_{m\to+\infty} N_U(m) = +\infty$. In a separate section we will give an algebraic characterization of such sequences U satisfying this condition. With similar hypotheses, we will then propose a second decision procedure in the case of abstract numeration systems. In particular, in both cases, we will highlight several examples that had not been resolved until now. We will end this chapter with a brief discussion of the HD0L periodicity problem: given a morphism f prolongable on a and a morphism g, is it decidable whether or not the infinite word $g(f^\omega(a)) = \lim_{n\to+\infty} g(f^n(a))$ is ultimately periodic? We will show that our results turn out to be a step in the direction of solving this famous problem, which is still open. Indeed, J. Honkala and M. Rigo proved the equivalence between this problem and the decision problem we are interested in throughout this chapter extended to any abstract numeration systems [**HR04**]. This result arises from the fact that abstract numeration systems are closely related to morphic sequences, *i.e.*, infinite words of the form $g(f^\omega(a))$ mentioned above [**RM02**].

In the fourth chapter we will generalize to the multidimensional case A. Maes and M. Rigo's result about automatic sequences extended to abstract numeration systems [**Rig00, RM02**]. In this case one refers to S-*automatic* sequences. The origin of such a generalization arises naturally from that which existed previously in the case of integer base numeration

[4]In computer science a decision problem is a problem that has at least one parameter able to take an infinite number of values and which can be answered by "yes" or "no". Deciding such a problem means that, given any instances of the problem assumed to be implementable, we can answer the problem with certainty, after a finite (but arbitrarily large) number of effectively computable operations.

systems. Indeed, the result mentioned above extends one of A. Cobham's theorems establishing a correspondence between b-automatic words and morphic words obtained thanks to a uniform morphism of length b [**Cob72**]. O. Salon had already extended this result to multidimensional words by considering morphisms under which the image of a letter is a hypercube of size b [**Sal87a, Sal87b**]. Thus, the theorem demonstrated in this chapter fills the "missing cell" of this diagram of extensions to the abstract numeration systems and to the multidimensional setting. In his doctoral dissertation A. Maes had introduced the definition of *shape-symmetric* multidimensional words. We will show that, in the multidimensional case, the S-automatic words correspond to the images under a coding of shape-symmetric pure morphic words. An essential point of the proof of this theorem is to generalize the standard result in the unidimensional case (for instance, see [**Cob68, Pan83, AS03**]) according to which any word obtained by erasing all occurrences of a fixed letter from a morphic word is either finite or morphic. Throughout this chapter, in order to provide a clear presentation, we illustrate the concepts under consideration thanks to a number of examples.

Finally, in the fifth and final chapter, we will concentrate on the representation of real numbers in the general framework of abstract numeration systems extended to any languages, *i.e.*, to languages which are no longer necessarily regular. The aim of this study was to provide a unified approach to several numeration systems encountered in the literature [**AFS08, DT89, LR01, Lot02**]. For example, the rational base numeration systems recently introduced in [**AFS08**] give rise to non-context-free numeration languages. We will construct, under certain general hypotheses on the numeration language, a formalism for the representation of the real numbers by infinite words, which are limits of words in the numeration language. We will then illustrate this formalism thanks to three examples of abstract numeration systems built on non-regular languages. One of these is based on the language of the prefixes of Dyck words. In each of these cases we will check if the systems under consideration verify our general hypotheses and, if possible, we will compute the value corresponding to an infinite word belonging to the set of valid representations explicitly.

This study will finish with several perspectives for future research continuing the work accomplished within this dissertation.

CHAPTER 1

Basics

This first chapter outlines the basic notions that are needed for a clear understanding of the present dissertation.

We start with some usual definitions and common results from automata theory. The interested reader can find many more details in [**Eil74, Sak03**].

Next, we define positional numeration systems and linear numeration systems. For instance, see [**Lot02**] for details. In particular, we give the definition of the so-called integer base numeration systems and the Fibonacci numeration system. We also introduce the notion of U-recognizability of a set of non-negative integers in this context.

Finally, we define abstract numeration systems as originally introduced in [**LR01**] and the corresponding notion of S-recognizability of a set of non-negative integers. In doing so we will revisit some of the first few results achieved in this area.

We will always refer to the set of non-negative integers $\{0, 1, 2, \ldots\}$ as $\mathbb{N}$. Moreover, for two integers i and j satisfying $i \leq j$, we let $[\![i, j]\!]$ denote the interval of integers $\{i, i+1, \ldots, j-1, j\}$.

1.1. Words and Languages

Definition 1.1.1. An *alphabet* is a non-empty finite set. The elements of an alphabet are called *letters*. A *word* over an alphabet Σ is a finite or infinite sequence of letters in Σ. The *empty word*, denoted by ε, is the empty sequence. The *minimal alphabet* of a word is the set of letters occurring in this word. The *length* of a finite word w, denoted by $|w|$, is the number of letters making up w. If w is a non-empty finite (resp. infinite) word, then for any $n \in [\![0, |w|-1]\!]$ (resp. $n \in \mathbb{N}$), we let $w[n]$ denote its $(n+1)$st letter. The *reversal* of a finite word w, denoted by $\widetilde{w}$, is the finite word defined by $\widetilde{w}[n] = w[|w|-n-1]$ for all $n \in [\![0, |w|-1]\!]$. The set of finite (resp. infinite) words over an alphabet Σ is denoted by Σ^* (resp. Σ^ω). For a unary alphabet $\{a\}$, we usually write a^* instead of $\{a\}^*$. A *language* (resp. ω-language) over an alphabet Σ is a subset of Σ^* (resp. Σ^ω).

Note that the letters of a word are indexed from 0. Thus the first letter of a non-empty word w is $w[0]$.

Example 1.1.2. Let $\Sigma = \{a, b, c\}$ be the alphabet composed of the three letters a, b, and c. Consider the finite word $w = bccba$ over Σ. Its length is $|w| = 5$, its 4th letter is $w[3] = b$, and its reversal is $\widetilde{w} = abccb$. The minimal alphabet of the word aba is $\{a, b\}$. Now, consider the infinite word $z = abcabcabcabc \cdots$ over Σ. Its 10th letter is $z[9] = a$.

Definition 1.1.3. If u and v are two finite words over an alphabet Σ, then the *concatenation of* u *and* v, denoted by $u \cdot v$ (or simply uv if there is no need to emphasize), is the finite word w satisfying $w[n] = u[n]$ for all $n \in [\![0, |u| - 1]\!]$ and $w[n] = v[n - |u|]$ for all $n \in [\![|u|, |u| + |v| - 1]\!]$. For a finite word u over an alphabet Σ and a non-negative integer n, we let u^n denote the concatenation of n copies of u, which is defined by induction by $u^0 = \varepsilon$ and $u^{n+1} = u^n u$ for all $n \in \mathbb{N}$.

Example 1.1.4. The concatenation of the words *wood* and *stock* produces the word *woodstock*.

Note that, embedded with the concatenation product of words, Σ^* is the free monoid generated by Σ having ε as neutral element. We can thus define morphisms from Σ^* to Δ^* for two alphabets Σ and Δ.

Definition 1.1.5. Let Σ and Δ be two alphabets. A *morphism* is a map $\mu \colon \Sigma^* \to \Delta^*$ satisfying $\mu(xy) = \mu(x)\mu(y)$ for all $x, y \in \Sigma^*$. Whenever we have $\Sigma = \Delta$, we say that μ is a morphism *on* Σ.

Definition 1.1.6. Let L and M be two languages. The *concatenation of* L *and* M is the language $LM = \{uv \mid u \in L, v \in M\}$. For all $n \in \mathbb{N}$, we let L^n denote the concatenation of n copies of L, which is defined by

$$L^0 = \{\varepsilon\} \text{ and } \forall n \in \mathbb{N} \setminus \{0\},\ L^n = \{u^{(1)} \cdots u^{(n)} \mid \forall i \in [\![1, n]\!],\ u^{(i)} \in L\}.$$

For all $n \in \mathbb{N}$, we define $L^{\leq n} = \bigcup_{i=0}^{n} L^i$. The *Kleene closure of* L is the language $L^* = \bigcup_{n \geq 0} L^n$. For a language $L = \{w\}$ containing only one element, we usually write w^* instead of $\{w\}^*$.

Definition 1.1.7. Let Σ be an alphabet, u be a finite word over Σ, and v be an infinite word over Σ. The *concatenation of* u *and* v, denoted by $u \cdot v$ (or simply uv if there is no need to emphasize), is the infinite word w defined by $w[n] = u[n]$ for all $n \in [\![0, |u| - 1]\!]$ and $w[n] = v[n - |u|]$ for all integers $n \geq |u|$.

Definition 1.1.8. Let w be a word over an alphabet Σ. A *factor* of w is a finite word u such that there exist $x \in \Sigma^*$ and $y \in \Sigma^* \cup \Sigma^\omega$ satisfying $w = xuy$. For any non-negative integers m and n satisfying $m \leq n$, we let $w[m, n]$ denote the factor $w[m] \cdots w[n]$ of w. For any $n \in \mathbb{N}$, the *prefix of length* n of w is the factor $w[0, n-1]$, where, by convention, we set $w[0, -1] = \varepsilon$. We let $\mathrm{Pref}(w)$ denote the set of all prefixes of w:

$$\mathrm{Pref}(w) = \{x \in \Sigma^* \mid \exists y \in \Sigma^* \cup \Sigma^\omega,\ w = xy\}.$$

Observe that we have $w[n] = w[n, n]$ for any non-empty word w and any non-negative integer n.

Definition 1.1.9. Let u be a finite word over an alphabet Σ. We let u^ω denote the concatenation of infinitely many copies of u, which is defined by $u^\omega[n\,|u|, (n+1)\,|u| - 1] = u$ for all $n \in \mathbb{N}$.

Definition 1.1.10. The *prefix-closure* of a language L over an alphabet Σ, which is denoted by $\mathrm{Pref}(L)$, is the language of the prefixes of its words:

$$\mathrm{Pref}(L) = \{x \in \Sigma^* \mid \exists y \in \Sigma^*,\ xy \in L\}.$$

A language L is *prefix-closed* if it satisfies $L = \mathrm{Pref}(L)$.

1.2. Orders on Words

If an alphabet Σ is endowed with a total order, then one can extend this order to Σ^* or to $\Sigma^\omega \cup \Sigma^*$. In this text two particular orders on words will essentially be used: the lexicographical order and the genealogical order.

Definition 1.2.1. Let $(\Sigma, <)$ be a totally ordered alphabet. The order $<$ on Σ extends to an order on Σ^ω, called the *lexicographical order*, as follows. If u and v are two infinite words over Σ, then u is said to be *lexicographically less* than v, and we write $u <_{\mathrm{lex}} v$, if there exist $p \in \Sigma^*$, $s, t \in \Sigma^\omega$, and $a, b \in \Sigma$ such that we have $u = pas$, $v = pbt$, and $a < b$. This order extends to $\Sigma^\omega \cup \Sigma^*$ by replacing finite words z over Σ by $z\#^\omega \in (\Sigma \cup \{\#\})^\omega$, where $\#$ is a letter not belonging to the alphabet Σ which is assumed to satisfy $\# < a$ for all a in Σ. We write $u \leq_{\mathrm{lex}} v$ for two words u and v satisfying either $u <_{\mathrm{lex}} v$ or $u = v$.

Note that the lexicographical order is the usual order used in any natural language dictionary (if, of course, accents and dashes are omitted).

Definition 1.2.2. Let $(\Sigma, <)$ be a totally ordered alphabet. The order $<$ on Σ extends to an order on Σ^*, called the *genealogical order*, as follows. If u

and v are two finite words over Σ, then u is said to be *genealogically less* than v, and we write $u <_{\text{gen}} v$, if they satisfy either $|u| = |v|$ and $u <_{\text{lex}} v$, or $|u| < |v|$. We write $u \leq_{\text{gen}} v$ for two finite words u and v satisfying either $u <_{\text{gen}} v$ or $u = v$.

In the literature some authors call radix order or military order what we call genealogical order.

Example 1.2.3. Consider the alphabet $\{a, b\}$ totally ordered by $a < b$. We have $aabb <_{\text{lex}} aba \; <_{\text{lex}} abaa$ but $aba <_{\text{gen}} aabb <_{\text{gen}} abaa$.

Definition 1.2.4. Let L be a language over an alphabet totally ordered by $<$. The *minimal (resp. maximal) language of L with respect to $<$* is the language of the smallest (resp. greatest) words of each length with respect to the lexicographical order:

$$\text{Min}_<(L) = \{w \in L \mid \forall z \in L,\, |z| = |w| \Rightarrow z \geq_{\text{lex}} w\};$$
$$\text{Max}_<(L) = \{w \in L \mid \forall z \in L,\, |z| = |w| \Rightarrow z \leq_{\text{lex}} w\}.$$

1.3. Automata

Automata can be viewed as the simplest model of computation. They will appear all along this dissertation.

Definition 1.3.1. A *deterministic automaton* is a 5-tuple

$$\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$$

where

- Q is a non-empty set, called the set of *states*;
- Σ is an alphabet;
- $\delta \colon Q \times \Sigma \to Q$ is the (possibly partial) *transition function*;
- q_0 is a distinguished element of Q, called the *initial state*;
- $F \subseteq Q$ is the set of *final states*.

The function δ naturally extends to a (possibly partial) function on $Q \times \Sigma^*$ by declaring $\delta(q, \varepsilon) = q$ and $\delta(q, aw) = \delta(\delta(q, a), w)$ for $q \in Q$, $a \in \Sigma$, and $w \in \Sigma^*$. When the context is clear, we use the notation $q \cdot w$ as a shorthand for $\delta(q, w)$. If the transition function is total, then the automaton is said to be *complete*. A deterministic automaton is *finite* (resp. *infinite*) if its set of states is finite (resp. infinite). We use DFA as a shorthand for "deterministic finite automaton". A finite word w over Σ is *accepted* (or *recognized*) *by* $\mathcal{A}$ if $\delta(q_0, w)$ belongs to F. The set of words accepted by $\mathcal{A}$, denoted by $L(\mathcal{A})$, is the *language accepted* (or *recognized*) *by* $\mathcal{A}$. The *language accepted from the* *state* $q \in Q$, denoted by $L_q(\mathcal{A})$, is the set of words accepted by the automaton $(Q, \Sigma, \delta, q, F)$.

Deterministic automata can be represented by oriented graphs. Nodes are states and, for all states p and q and all letters a satisfying $p \cdot a = q$, there is an edge from p to q labeled by a. The initial state is designated by an incoming arrow and final states are designated by outgoing arrows.

Example 1.3.2. Consider the DFA $\mathcal{A} = (\{1, 2, 3, 4\}, \{a, b, c\}, \delta, 1, \{1, 2, 3\})$ where the transition function δ is given by the following tables:

	a	b	c
1	1	2	3
2	2	3	4

	a	b	c
3	3	4	4
4	4	4	4

Since δ is a total function, $\mathcal{A}$ is a complete DFA. The transition graph of $\mathcal{A}$ is depicted in Figure 1.1.

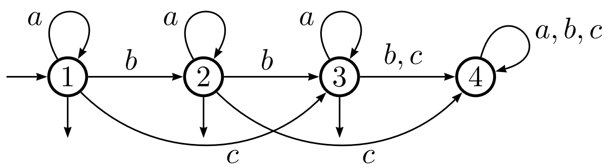

FIGURE 1.1. A deterministic finite automaton.

Definition 1.3.3. Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a deterministic automaton. A state q in Q is *accessible* if it can be reached from the initial state, *i.e.*, if there exists a word $w \in \Sigma^*$ such that we have $\delta(q_0, w) = q$ and is *coaccessible* if one can reach a final state from it, *i.e.*, if there exists $w \in \Sigma^*$ such that we have $\delta(q, w) \in F$. The automaton $\mathcal{A}$ is *accessible* (resp. *coaccessible*) if all its states are accessible (resp. coaccessible) and is *trim* if it is both accessible and coaccessible.

Example 1.3.4. The automaton depicted in Figure 1.2 is a trim deterministic automaton. Observe that its transition function is a partial function because, for instance, $2 \cdot c$ is undefined. Hence this DFA is not complete. Also, note that the automata of Figure 1.1 and Figure 1.2 recognize the same words.

Among all deterministic automata accepting a language, one can distinguish the minimal automaton of this language.

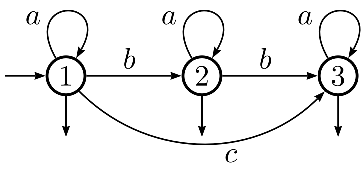

FIGURE 1.2. A trim deterministic automaton.

Definition 1.3.5. Let L be a language over an alphabet Σ. The *Myhill-Nerode equivalence relation*, denoted by $\sim_L$, is the relation on Σ^* defined as follows: $u \sim_L v$ means that, for all $w \in \Sigma^*$, we have $uw \in L \Leftrightarrow vw \in L$. If u is a finite word over Σ, then we define $u^{-1}L = \{w \in \Sigma^* \mid uw \in L\}$, that is, $u^{-1}L$ is the language of the finite words over Σ which, when concatenated with w, form a word that belongs to L.

Note that, with the notation of the previous definition, we have $u \sim_L v \Leftrightarrow u^{-1}L = v^{-1}L$.

Definition 1.3.6. The *minimal automaton* of a language L over an alphabet Σ is the deterministic automaton

$$\mathcal{A}_L = (Q_L, \Sigma, \delta_L, q_{0,L}, F_L)$$

with

- $Q_L = \{u^{-1}L \mid u \in \Sigma^*\}$;
- $\forall q \in Q_L, \forall a \in \Sigma,\ \delta_L(q, a) = a^{-1}\, q$;
- $q_{0,L} = \varepsilon^{-1}L = L$;
- $F_L = \{u^{-1}L \mid u \in L\}$.

The *trim minimal automaton* of a language L is the minimal automaton of L from which the only possible non-coaccessible state, the *sink state*, is removed.

Proposition 1.3.7. *The minimal automaton of a language accepts this language.*

The denomination "minimal automaton" is justified by the following proposition.

Proposition 1.3.8. *Let L be a language over an alphabet Σ and let $\mathcal{A}$ be a deterministic automaton accepting L having Q as set of states. Then we have* $\operatorname{Card} Q_L \leq \operatorname{Card} Q$.

Now, we introduce a second kind of automata, namely, the non-deterministic automata.

Definition 1.3.9. A *non-deterministic automaton* is a 5-tuple

$$\mathcal{A} = (Q, \Sigma, \Delta, I, F)$$

where

- Q, Σ, and F are defined as in a deterministic automaton;
- $\Delta \subseteq Q \times \Sigma^* \times Q$ is a non-empty set, called the *transition relation*;
- $I \subseteq Q$ is a non-empty set, called the *set of initial states*.

A non-deterministic automaton is *finite* (resp. *infinite*) if its set of states is finite (resp. infinite). We use NFA as a shorthand for "non-deterministic finite automaton". A word w is *accepted* (or *recognized*) *by* $\mathcal{A}$ if there exists a *path* from an initial state to a final state labeled by w, *i.e.*, if there exist a positive integer k, finite words $w_1, w_2, \ldots, w_k$ over Σ, and states $q_0, q_1, \ldots, q_k$ in Q with $q_0 \in I$ and $q_k \in F$ such that

$$(q_0, w_1, q_1),\ (q_1, w_2, q_2) \ldots,\ (q_{k-1}, w_k, q_k)$$

belong to Δ. The *language accepted* (or *recognized*) *by* $\mathcal{A}$, denoted by $L(\mathcal{A})$, is the set of words accepted by $\mathcal{A}$.

Example 1.3.10. Consider the non-deterministic finite automaton

$$\mathcal{A} = (\{1, 2, 3\}, \{a, b, c\}, \Delta, \{1, 2\}, \{1, 3\})$$

where the transition relation Δ is given by

$$\Delta = \{(1, a, 1), (1, a, 2), (1, ba, 2), (2, b, 2), (2, ab, 3), (3, a, 1), (3, c, 2), (3, c, 3)\}.$$

The transition graph of $\mathcal{A}$ is depicted in Figure 1.3. For all states p and q and all words w satisfying $(p, w, q) \in \Delta$, there is an edge from p to q labeled by w. Again, initial states are designated by incoming arrows and final states are designated by outgoing arrows.

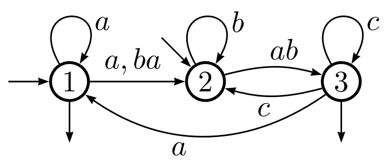

FIGURE 1.3. A non-deterministic finite automaton.

Finally, we introduce the notion of deterministic automata with output. We will use them in Chapter 4 to define S-automatic words.

Definition 1.3.11. A deterministic finite automaton *with output* (DFAO for short) is a 6-tuple

$$\mathcal{B} = (Q, \Sigma, \delta, q_0, \Gamma, \tau)$$

where

- Q, Σ, δ, q_0 are defined as in a DFA;
- Γ is the *output alphabet*;
- $\tau : Q \to \Gamma$ is the *output function*.

The *output* corresponding to the *input* $w \in \Sigma^*$ is $\tau(\delta(q_0, w))$.

The transition graph of a DFAO is represented as the one of the associated DFA with additional outgoing labeled arrows on each state to indicate the corresponding output. Usually, we only represent the useful part of a DFAO, that is, the states accessible by reading words in the language under consideration.

1.4. Regular Languages

A DFA is a particular NFA. Therefore one could naturally believe that the class of languages accepted by a NFA is larger than the class of languages accepted by a DFA. The following proposition shows that this is actually not the case.

Proposition 1.4.1. [**RS59**] *A language is accepted by a NFA if and only if it is accepted by a DFA.*

Now, we are ready to introduce the definition of a regular language, which is a central notion in this text.

Definition 1.4.2. A language is *regular* if it is accepted by a finite automaton.

The next two theorems are characterizations of regular languages. They are known as *Kleene's theorem* and *Myhill and Nerode's theorem* respectively.

Theorem 1.4.3. [**Kle56**] *The family of regular languages over an alphabet Σ is the least family of languages over Σ containing the empty set and the singletons, and closed under union, concatenation, and Kleene closure.*

Theorem 1.4.4. [**Ner58**] *A language L is regular if and only if the Myhill-Nerode equivalence relation $\sim_L$ is of finite index,* i.e., *if and only if its minimal automaton $\mathcal{A}_L$ is finite.*

Let us also recall the following stability result of the class of regular languages.

Proposition 1.4.5. *The class of regular languages is closed under concatenation, Kleene closure, union, intersection, complementation, reversal, and image under morphism.*

The next result is often used to reject the regularity of a language. One usually refers to it as the *pumping lemma.*

Proposition 1.4.6. *If L is a regular language over an alphabet Σ, then there exists a positive integer k such that any word w in L of length $|w| \geq k$ can be decomposed as $w = xyz$, where x, y, z are finite words over Σ satisfying $y \neq \varepsilon$, $|xy| \leq k$, and $xy^*z \subseteq L$.*

1.5. Counting Function

Definition 1.5.1. For any language L over an alphabet Σ and any non-negative integer n, we let

$$\mathbf{u}_L(n) = \mathrm{Card}(L \cap \Sigma^n)$$

denote the number of words of length n in L and

$$\mathbf{v}_L(n) = \sum_{i=0}^{n} \mathbf{u}_L(i) = \mathrm{Card}(L \cap \Sigma^{\leq n})$$

denote the number of words of length less than or equal to n in L. The map $\mathbf{u}_L \colon \mathbb{N} \to \mathbb{N}$ is called the *counting (or combinatorial complexity) function of L.*

Let us fix some asymptotic notation. Note that some authors do not use exactly the same definition for Ω as the one introduced in the next paragraph, but in this text we will always refer to this symbol as defined below.

Definition 1.5.2. Let f and g be functions taking values in $\mathbb{N}$. We say that f is $O(g)$, and we write $f = O(g)$, if there exist positive constants c and N such that, for all integers $n \geq N$, we have $f(n) \leq cg(n)$. We say that f is $\Omega(g)$, and we write $f = \Omega(g)$, if there exists a positive constant c and a strictly increasing sequence of positive integers $(n_i)_{i\geq 0}$ such that, for all $i \in \mathbb{N}$, we have $f(n_i) \geq cg(n_i)$. We say that f is $\Theta(g)$, and we write $f = \Theta(g)$, if f is both $O(g)$ and $\Omega(g)$. Morever, we say that f and g have *equivalent behaviors at infinity*, which is denoted by $f(n) \sim g(n)$ $(n \to +\infty)$ (or simply $f \sim g$ when the context is clear), if we have $\lim_{n\to+\infty} \frac{f(n)}{g(n)} = 1$.

Definition 1.5.3. A language L is *polynomial* if its counting function $\mathbf{u}_L(n)$ is $O(n^k)$ for some $k \in \mathbb{N}$ and *exponential* if it is $\Omega(\theta^n)$ for some $\theta > 1$.

The following theorem provides us with the general form of a polynomial regular language.

Theorem 1.5.4. [**SYZS92**] *Let L be a regular language over an alphabet Σ and k be a non-negative integer. The counting function $\mathbf{u}_L(n)$ of L is $O(n^k)$ if and only if L is a finite union of languages of the form*

$$xy_1^* z_1 y_2^* \cdots y_k^* z_k \tag{3}$$

with $x, y_i, z_i \in \Sigma^*$ for all $i \in [\![1, k]\!]$.

As shown by the next theorem, there is a gap between the class of polynomial regular languages and the one of exponential regular languages.

Theorem 1.5.5. [**SYZS92**] *Any regular language is either polynomial or exponential.*

Notation. When the context is sufficiently clear, if q is a state of a DFA $\mathcal{A}$ and n is a non-negative integer, then we write $\mathbf{u}_q(n)$ and $\mathbf{v}_q(n)$ instead of $\mathbf{u}_{L_q(\mathcal{A})}(n)$ and $\mathbf{v}_{L_q(\mathcal{A})}(n)$ respectively. So $\mathbf{u}_q(n)$ (resp. $\mathbf{v}_q(n)$) designates the number of words of length n (resp. less than or equal to n) accepted from the state q in $\mathcal{A}$.

The following standard result will often be involved in the remaining part of this dissertation. Before we state it, we need a definition. The terminology "strict" is taken from [**BR09**]. It will be used in Chapter 3.

Definition 1.5.6. Let R be a commutative ring. A sequence $U = (U_i)_{i \geq 0}$ in $R^{\mathbb{N}}$ is a *linear recurrence sequence over R* if it satisfies a *linear recurrence relation over R*, *i.e.*, if there exist a positive integer k and some coefficients $a_1, \ldots, a_k$ in R such that we have

$$\forall i \in \mathbb{N},\ U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i.$$

The *length* of the linear recurrence relation is k and its *characteristic polynomial* is

$$x^k - a_1 x^{k-1} - \cdots - a_k.$$

The linear recurrence relation is *strict* if its last coefficient a_k does not vanish. A linear recurrence sequence over a commutative ring R is *strict* if it satisfies a strict linear recurrence relation over R.

Proposition 1.5.7. *If L is a regular language, then the sequences $(\mathbf{u}_L(i))_{i \geq 0}$ and $(\mathbf{v}_L(i))_{i \geq 0}$ satisfy linear recurrence relations over $\mathbb{Z}$.*

Example 1.5.8. Consider the language $L = \{a, ab\}^* \cup \{c, cd\}^*$. We have $\mathbf{u}_L(0) = 1$ and $\mathbf{u}_L(i+1) = 2F_i$ for all $i \in \mathbb{N}$, where $(F_i)_{i\geq 0}$ is the Fibonacci sequence which is defined by $F_0 = 1$, $F_1 = 2$, and $F_{i+2} = F_{i+1} + F_i$ for all $i \in \mathbb{N}$. Consequently, we obtain

$$\forall i \in \mathbb{N},\ \mathbf{v}_L(i+1) = 1 + \sum_{j=1}^{i+1} \mathbf{u}_L(j) = 1 + 2\sum_{j=0}^{i} F_j.$$

For all $i \in \mathbb{N}$, we have $\mathbf{v}_L(i+1) - \mathbf{v}_L(i) = \mathbf{u}_L(i+1) = 2F_i$. This gives

$$\forall i \in \mathbb{N},\ \mathbf{v}_L(i+3) - \mathbf{v}_L(i+2) = (\mathbf{v}_L(i+2) - \mathbf{v}_L(i+1)) + (\mathbf{v}_L(i+1) - \mathbf{v}_L(i)).$$

Thus we obtain

$$\forall i \in \mathbb{N},\ \mathbf{v}_L(i+3) = 2\mathbf{v}_L(i+2) - \mathbf{v}_L(i),$$

with $\mathbf{v}_L(0) = 1, \mathbf{v}_L(1) = 3$, and $\mathbf{v}_L(2) = 7$.

Remark 1.5.9. The computation given in the previous example to obtain a linear recurrence relation for the sequence $(\mathbf{v}_L(i))_{i\geq 0}$ for any language L accepted by a given finite automaton can be carried on in general. Let L be a regular language and let q be a state of its minimal automaton $\mathcal{A}_L$. By Proposition 1.5.7, we know that the sequence $(\mathbf{u}_q(i))_{i\geq 0}$ satisfies a linear recurrence relation with integer coefficients. Thus there exist a positive integer k and $a_1, \ldots, a_k \in \mathbb{Z}$ such that we have

$$\forall i \in \mathbb{N},\ \mathbf{u}_q(i+k) = a_1\mathbf{u}_q(i+k-1) + \cdots + a_k\mathbf{u}_q(i).$$

Consequently, we have

$$\begin{aligned}\forall i \in \mathbb{N},\ \mathbf{v}_q(i+k+1) - \mathbf{v}_q(i+k) &= \mathbf{u}_q(i+k+1)\\ &= a_1(\mathbf{v}_q(i+k) - \mathbf{v}_q(i+k-1)) + \cdots + a_k(\mathbf{v}_q(i+1) - \mathbf{v}_q(i)).\end{aligned}$$

Therefore the sequence $(\mathbf{v}_q(i))_{i\geq 0}$ satisfies a linear recurrence relation over $\mathbb{Z}$ of length $k+1$.

Let us also recall here a standard result about (strict) linear recurrence sequences. For instance, see [**BR88, GKP94, BR09**]. Note that a stronger version of this proposition — Theorem 3.5.1 — will be given in Chapter 3.

Proposition 1.5.10. *Let K be an algebraically closed field of characteristic zero and let $U = (U_i)_{i\geq 0} \in K^{\mathbb{N}}$ be a sequence satisfying*

$$\forall i \in \mathbb{N},\ U_{i+k} = a_1U_{i+k-1} + \cdots + a_kU_i,$$

for some $k \in \mathbb{N}\setminus\{0\}$ and $a_1, \ldots, a_k \in K$ with $a_k \neq 0$. Assume that $\alpha_1, \ldots, \alpha_t$ are the roots of the associated characteristic polynomial with respective multiplicities $m_1, \ldots, m_t$. Then there exist polynomials $P_1, \ldots, P_t$ in $K[x]$ of *degrees respectively less than* $m_1, \ldots, m_t$ *and depending only on the initial conditions* $U_0, \ldots, U_{k-1}$ *such that we have*

$$\forall i \in \mathbb{N},\ U_i = P_1(i)\,\alpha_1^i + \cdots + P_t(i)\,\alpha_t^i.$$

1.6. Positional Numeration Systems

Definition 1.6.1. A *positional numeration system* is given by a strictly increasing sequence $U = (U_i)_{i\geq 0}$ of integers such that we have $U_0 = 1$ and $C_U = \sup\{\lceil U_{i+1}/U_i \rceil \mid i \in \mathbb{N}\}$ is finite. The *greedy U-representation* of a positive integer n, denoted by $\mathrm{rep}_U(n)$, is the unique finite word w over the alphabet $\Sigma_U = [\![0, C_U - 1]\!]$ not beginning with 0 and satisfying

$$n = \sum_{i=0}^{|w|-1} \widetilde{w}[i]\,U_i \quad \text{and} \quad \forall t \in [\![0, |w|-1]\!],\ \sum_{i=0}^{t} \widetilde{w}[i]\,U_i < U_{t+1}.$$

Moreover, we set $\mathrm{rep}_U(0) = \varepsilon$. The elements in Σ_U are called *digits*. The set $\mathrm{rep}_U(\mathbb{N})$ is called the *numeration language*. If w is a finite word over any alphabet of integers, then the *U-numerical value* of w, denoted by $\mathrm{val}_U(w)$, is given by

$$\mathrm{val}_U(w) = \sum_{i=0}^{|w|-1} \widetilde{w}[i]\,U_i.$$

The following two examples are very important. The notation they introduce will be used throughout the text. Note that, if there is no need to emphasize, we usually make no distinction between the symbols $0, 1, 2, 3, \ldots$ and the integers they represent.

Example 1.6.2. Let $b \geq 2$ be an integer. The *integer base b numeration system* is the positional numeration system built on the sequence

$$U_b = (b^i)_{i\geq 0}.$$

In this case we have $\Sigma_{U_b} = [\![0, b-1]\!]$ and the numeration language is

$$\mathcal{L}_b = \mathrm{rep}_{U_b}(\mathbb{N}) = \{1, 2, \ldots, b-1\}\{0, 1, \ldots, b-1\}^* \cup \{\varepsilon\}.$$

Thus we find back the usual base 10 numeration system which is used to represent numbers in everyday life.

Example 1.6.3. Consider the sequence $F = (F_i)_{i\geq 0} = (1, 2, 3, 5, 8, 13, \ldots)$ defined by

$$F_0 = 1,\ F_1 = 2, \quad \text{and} \quad \forall i \in \mathbb{N},\ F_{i+2} = F_{i+1} + F_i.$$

The *Fibonacci numeration system* is built on this sequence F. It was proved in [**Zec72**] that we have $\Sigma_F = \{0,1\}$ and that the set of the greedy representations of non-negative integers, *i.e.*, the numeration language, is the set

$$\mathrm{rep}_F(\mathbb{N}) = 1\{0,01\}^* \cup \{\varepsilon\}$$

of the words over $\{0,1\}$ not containing the factor 11. For instance, we have $\mathrm{rep}_F(15) = 100010$ and $\mathrm{val}_F(101001) = 13 + 5 + 1 = 19$.

As stated by the next proposition, considering greedy representations allows us to work with *order-preserving* positional numeration systems. More precisely, in this case, the natural order on the set of integers corresponds to the genealogical order on the numeration language.

Proposition 1.6.4. *Let U be a positional numeration system. For all non-negative integers m and n, we have*

$$m < n \Leftrightarrow \mathrm{rep}_U(m) <_{gen} \mathrm{rep}_U(n)$$

where the genealogical order $<_{gen}$ is induced by the natural order of the alphabet $\Sigma_U \subseteq \mathbb{N}$.

Definition 1.6.5. Let U be a positional numeration system. A set X of non-negative integers is *U-recognizable* if the language $\mathrm{rep}_U(X)$ over Σ_U is regular.

It is often convenient to work with positional numeration systems such that the whole set $\mathbb{N}$ of non-negative integers is U-recognizable, *i.e.*, such that the numeration language is regular. As we shall see further on, a necessary condition for this is that the sequence U satisfies a linear recurrence relation.

Definition 1.6.6. A positional numeration system $U = (U_i)_{i\geq 0}$ is said to be *linear* if U is a linear recurrence sequence over $\mathbb{Z}$.

Example 1.6.7. Integer base numeration systems and the Fibonacci numeration system introduced in Example 1.6.2 and Example 1.6.3 respectively are linear numeration systems with regular numeration languages. The shortest linear recurrence relations they satisfy are of length 1 and 2 respectively.

The next proposition is a particular case of a theorem of Shallit [**Sha94**].

Proposition 1.6.8. *Let U be a positional numeration system. If $\mathbb{N}$ is U-recognizable, then the sequence U satisfies a linear recurrence relation over $\mathbb{Z}$,* i.e.*, U is a linear numeration system.*

The converse of Proposition 1.6.8 does not hold in general. Sufficient conditions on the linear recurrence relation satisfied by U for $\mathbb{N}$ to be U-recognizable are considered in [**Lor95, Hol98**]. Here is a counterexample from [**Sha94**].

Example 1.6.9. Consider the positional numeration system $U = (U_i)_{i\geq 0}$ defined by $U_i = (i+1)^2$ for all $i \in \mathbb{N}$. Since it satisfies

$$\forall i \in \mathbb{N},\ U_{i+3} = 3\,U_{i+2} - 3\,U_{i+1} + U_i,$$

it is a linear numeration system. We have

$$\mathrm{rep}_U(\mathbb{N}) \cap 10^*10^* = \{10^a10^b \mid b^2 < 2a+4\}.$$

Using the pumping lemma (see Proposition 1.4.6), we easily obtain that the latter set is not regular. Therefore, since the class of regular languages is closed under intersection, $\mathbb{N}$ cannot be U-recognizable either.

The following two examples show that, for a positional numeration system U, the U-recognizability of $\mathbb{N}$ does not imply that U is a strict linear recurrence sequence. The first one is obtained by modifying the initial conditions of the Fibonacci numeration system.

Example 1.6.10. Consider the linear numeration system $U = (U_i)_{i\geq 0}$ defined by

$$U_0 = 1,\ U_1 = 2,\ U_2 = 4,\ \text{and } \forall i \in \mathbb{N},\ U_{i+3} = U_{i+2} + U_{i+1}.$$

It is easily verified that the sequence U satisfies no linear recurrence relation of length shorter than 3. But, of course, it ultimately satisfies the strict linear recurrence relation $U_{i+2} = U_{i+1} + U_i$ of length 2. Furthermore, observe that $\mathbb{N}$ is U-recognizable since the associated numeration language is given by

$$\mathrm{rep}_U(\mathbb{N}) = 1\{0, 01\}^*\{\varepsilon, 011\} \cup \{\varepsilon, 11\},$$

which is of course a regular language.

We can do the same, for instance, with the integer base 2 numeration system, as shown by the next example.

Example 1.6.11. Consider the linear numeration system $U = (U_i)_{i\geq 0}$ defined by

$$U_0 = 1,\ U_1 = 2,\ U_2 = 3,\ U_3 = 4,\ \text{and } \forall i \in \mathbb{N},\ U_{i+4} = 2U_{i+3}.$$

Again, the sequence U satisfies no linear recurrence relation of length shorter than 4. But, of course, it ultimately satisfies the strict linear recurrence relation $U_{i+2} = 2U_{i+1}$ of length 1. Once again, observe that $\mathbb{N}$ is U-recognizable since the associated numeration language is given by

$$\mathrm{rep}_U(\mathbb{N}) = 1\{0,1\}^*\{000,001,010,100\} \cup \{\varepsilon,1,10,100\},$$

which is of course a regular language.

For the sake of completeness, we restate a well-known property of ultimately periodic sets. The interested reader could for instance have a look at [**Sak03**] for a prologue on Pascal's machine for integer base systems. First, we need a definition.

Definition 1.6.12. A set of integers X is *ultimately periodic* (or *eventually periodic*) if there exist $a, p \in \mathbb{N}$ with $p > 0$, such that, for all $i \geq a$, we have $i \in X \Leftrightarrow i + p \in X$. If the integers a and p are minimal for the latter property, then we say that they are the *preperiod* and the *period* of X respectively.

Lemma 1.6.13. *Let $U = (U_i)_{i\geq 0}$ be a linear numeration system and p, q be non-negative integers. The language*

$$\mathrm{val}_U^{-1}(p + \mathbb{N}\, q) = \{w \in \Sigma_U^* \mid \mathrm{val}_U(w) \in p + \mathbb{N}\, q\} \subseteq \Sigma_U^*$$

is regular and a DFA accepting this language can be obtained efficiently. In particular, if $\mathbb{N}$ is U-recognizable, then a DFA accepting $\mathrm{rep}_U(p+\mathbb{N}\, q)$ can be obtained efficiently and any ultimately periodic set is U-recognizable.

Before giving the proof, observe that, for any non-negative integer n, $\mathrm{val}_U^{-1}(n)$ is a finite set of words $\{x_1, \ldots, x_{t_n}\}$ over Σ_U such that, for all $i \in [\![1, t_n]\!]$, we have $\mathrm{val}_U(x_i) = n$. This set contains in particular the greedy U-representation $\mathrm{rep}_U(n)$ of n.

Proof. Since regularity is stable under finite modifications, that is, adding or removing a finite number of words in the language, we can assume $0 \leq p < q$. Since U is linear, the reduced sequence $(U_i \bmod q)_{i\geq 0}$ is ultimately periodic, say with preperiod ι and period π. The following DFA recognizes the reversal of the words in $\mathrm{val}_U^{-1}(p + \mathbb{N}\, q)$. The alphabet of the automaton is Σ_U. States are pairs (r, s) of integers satisfying $0 \leq r < q$ and $0 \leq s < \iota+\pi$. The initial state is $(0, 0)$. Final states are the ones with the first component equal to p. Transitions are defined as follows. For all $j \in \Sigma_U$, all $r \in [\![0, q-1]\!]$, and all $s \in [\![0, \iota + \pi - 2]\!]$, we have

$$(r, s) \xrightarrow{j} ((jU_s + r) \bmod q,\ s + 1)\,;$$

$$(r, \iota + \pi - 1) \xrightarrow{j} ((jU_{\iota+\pi-1} + r) \bmod q,\ \iota)\,.$$

Observe that this automaton does not check the greediness of the accepted words because the construction only relies on the U-numerical value of the words modulo q.

For the particular case, one has to consider the intersection of two regular languages: $\mathrm{rep}_U(\mathbb{N}) \cap \mathrm{val}_U^{-1}(p + \mathbb{N}\, q)$. □

1.7. Abstract Numeration Systems

A numeration system can be viewed as a bijection between the set of non-negative integers $\mathbb{N}$ and a language, called the numeration language. Otherwise stated, this is a way of representing numbers by words.

When the numeration language is accepted by a finite automaton, one can easily check whether or not a word is a valid representation of a number. Moreover, Lemma 1.6.13 states that in any positional numeration system, $\mathbb{N}$ is U-recognizable if and only if all ultimately periodic sets are U-recognizable. Also, note that regular languages are the simplest languages in terms of the Chomsky hierarchy, by opposition to recursively enumerable languages, *i.e.*, languages recognized by Turing machines; see for instance [**Sud06**] or [**Wol06**] for details. Therefore the recognizability of $\mathbb{N}$, that is, the regularity of the numeration language, is desirable and can be considered to be a natural expectation for any numeration system.

In view of the arguments above, P. Lecomte et M. Rigo introduced the following definition [**LR01**]. They considered the problem the other way around and the recognizability of $\mathbb{N}$ became their basic requirement. Instead of taking a sequence U of integers and looking for conditions that guarantee the U-recognizability of $\mathbb{N}$, they took an arbitrary infinite regular language L over an alphabet Σ to build a numeration system, this language L being viewed as the set of valid representations of all the integers.

Definition 1.7.1. An *abstract numeration system* is a triple

$$S = (L, \Sigma, <)$$

where L is an infinite regular language, called the *numeration language*, written over a totally ordered alphabet $(\Sigma, <)$. Enumerating the words in L using the genealogical order $<_{\mathrm{gen}}$ induced by the order $<$ on Σ gives a one-to-one correspondence $\mathrm{rep}_S\colon \mathbb{N} \to L$ mapping any non-negative integer n onto the $(n+1)$st word in L. The inverse map is denoted by $\mathrm{val}_S\colon L \to \mathbb{N}$. For all words w in L, we say that $\mathrm{val}_S(w)$ is the *numerical S-value* (or simply the *S-value*) of w.

Note that, in particular, 0 is sent onto the first word in the genealogically ordered numeration language.

One could relax the assumption about the regularity of L in the definition of an abstract numeration system $\mathcal{S} = (L, \Sigma, <)$. This would give a wider framework, but then we would loose the recognizability of $\mathbb{N}$. We shall consider this larger class of abstract numeration systems in Chapter 5. In this case, to avoid any confusion, we will speak about *generalized* abstract numeration systems.

Example 1.7.2. Take $S = (\{a, ba\}^*, \{a, b\}, a < b)$. The first few words in the numeration language enumerated with respect to the genealogical order are $\varepsilon, a, aa, ba, aaa, aba, baa, aaaa, aaba, abaa, baaa, baba, aaaaa$. So, for instance, we have $\mathrm{val}_S(ba) = 3$ and $\mathrm{rep}_S(8) = aaba$.

In view of Proposition 1.6.4, in any positional numeration system U, the map rep_U is order-preserving with respect to the genealogical order and the natural order of the integers. Therefore any positional numeration system having a regular numeration language is an abstract numeration system. Thus one can say that the framework of abstract numeration systems generalizes that of positional numeration systems having a regular numeration language.

Example 1.7.3. Let $b \geq 2$ be an integer. By considering the abstract numeration system S built on the language $\mathcal{L}_b$ with the natural order on the digits (see Example 1.6.2) one gets back the standard integer base b numeration system, that is, for all $n \in \mathbb{N}$, we have $\mathrm{rep}_S(n) = \mathrm{rep}_{U_b}(n)$.

Example 1.7.4. The abstract numeration system $S = (L, \{0, 1\}, 0 < 1)$ built on the language $L = 1\{0, 01\}^* \cup \{\varepsilon\}$ of the words over $\{0, 1\}$ that do not contain the factor 11 gives back the Fibonacci numeration system introduced in Example 1.6.3, *i.e.*, for all $n \in \mathbb{N}$, we have $\mathrm{rep}_S(n) = \mathrm{rep}_F(n)$.

The next example shows that the class of positional numeration systems for which $\mathbb{N}$ is recognizable is strictly included in the class of abstract numeration systems.

Example 1.7.5. Consider once again the language $L = \{a, ab\}^* \cup \{c, cd\}^*$ of Example 1.5.8. Let S be the abstract numeration system built on L with the order $a < b < c < d$ on the alphabet $\{a, b, c, d\}$. The first few words in L enumerated with respect to the genealogical order are given in Table 1.1. We show that there is no bijection $f : \{a, b, c, d\} \to \mathbb{N}$ between $\{a, b, c, d\}$ and a set of integers leading to a positional numeration system. Otherwise stated, the letters a, b, c, d cannot be identified with usual "digits". Assume

0	ε	5	cc	10	ccc	15	$aaba$	20	$ccdc$
1	a	6	cd	11	ccd	16	$abaa$	21	$cdcc$
2	c	7	aaa	12	cdc	17	$abab$	22	$cdcd$
3	aa	8	aab	13	$aaaa$	18	$cccc$	23	$aaaaa$
4	ab	9	aba	14	$aaab$	19	$cccd$	24	$aaaab$

TABLE 1.1. The first few words in $\{a, ab\}^* \cup \{c, cd\}$.

that there exists a sequence $U = (U_i)_{i\geq 0}$ of integers satisfying

$$\forall w \in L,\ \mathrm{val}_S(w) = \sum_{i=0}^{|w|-1} f(\widetilde{w}[i])\, U_i.$$

Since we have $\mathrm{val}_S(a) = 1$ and $\mathrm{val}_S(c) = 2$, we get $U_0 = 1$, $f(a) = 1$, and $f(c) = 2$. Moreover, we have $\mathrm{val}_S(aa) = 3 = f(a)U_1 + f(a)U_0$. So we get $U_1 = 2$. Therefore we obtain $\mathrm{val}_U(f(c)f(c)) = 2U_1 + 2U_0 = 6$. But we also have $\mathrm{val}_S(cc) = 5$, leading to a contradiction.

When the context is clear, if $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ is a DFA accepting the language L, then, for any state $q \in Q$, we usually write $\mathrm{val}_q \colon L_q \to \mathbb{N}$ instead of val_{S_q}, where S_q denotes the abstract numeration system $(L_q, \Sigma, <)$.

The following proposition provides a method for computing the function val_S. Recall that $\mathbf{1}_{q,q'}$ equals 1 if we have $q = q'$ and equals 0 otherwise.

Proposition 1.7.6. [**LR01**] *Let $S = (L, \Sigma, <)$ be an abstract numeration system and $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting L. We have*

$$\forall w \in L,\ \mathrm{val}_S(w) = \sum_{q\in Q} \sum_{i=0}^{|w|-1} \beta_{q,i}(w)\, \mathbf{u}_q(|w| - i - 1),$$

where we set

$$\beta_{q,i}(w) = \mathrm{Card}\{a < w[i] \mid \delta(q_0, w[0, i-1]\, a) = q\} + \mathbf{1}_{q,q_0} \tag{4}$$

for all $i \in [\![0, |w| - 1]\!]$.

Remark 1.7.7. Note that, with the notation of the previous proposition, if x and y are finite words over Σ with $x \neq \varepsilon$, then we have $\beta_{q,i}(xy) = \beta_{q,i}(x)$ for all $q \in Q$ and all $i \in [\![0, |x| - 1]\!]$.

The following definition extends Definition 1.6.5 to abstract numeration systems.

Definition 1.7.8. Let $S = (L, \Sigma, <)$ be an abstract numeration system. A subset X of $\mathbb{N}$ is *S-recognizable* if $\mathrm{rep}_S(X)$ is regular.

As in the framework of positional numeration systems, a number of natural questions stems from this notion of S-recognizability:

- Given an abstract numeration system S, can we describe the S-recognizable sets of non-negative integers?
- What are the sets of non-negative integers which are S-recognizable for all abstract numeration systems S?
- Are there some sets of non-negative integers that are never S-recognizable?
- Given a set of non-negative integers X, can we build an abstract numeration system S such that X is S-recognizable?
- For which abstract numeration systems do arithmetic operations like translation, multiplication by a constant, addition, or multiplication preserve S-recognizability?
- Given an abstract numeration system S, what are the operations which preserve S-recognizability?
- ...

Other kind of natural questions arises in this context of abstract numeration systems:

- How to represent the real numbers in an abstract numeration system?
- Can automatic sequences be defined in this context?
- Can we give a logical characterization of S-recognizable sets of non-negative integers?
- ...

Most of these questions are not answered yet. Nevertheless, some results exist and a few are presented below.

First of all, a noteworthy property of abstract numeration systems is that the arithmetic progressions are always recognizable.

Theorem 1.7.9. [**LR01**] *Let S be an abstract numeration system and let p and q be non-negative integers. The arithmetic progression $p + \mathbb{N}\, q$ is S-recognizable and a DFA accepting $\mathrm{rep}_S(p + \mathbb{N}\, q)$ is effectively computable. Consequently, any ultimately periodic set is S-recognizable.*

In view of this result, a new question naturally arises: given an abstract numeration system, is it decidable whether or not an S-recognizable set X, which is given through a DFA accepting $\mathrm{rep}_S(X)$, is ultimately periodic?

This question has not been solved yet, even in the case of a positional numeration system having a regular numeration language. We will come back to this particular problem in Chapter 3.

The next proposition shows that S-recognizability is stable under translation by a constant.

Proposition 1.7.10. [**LR01**] *Let S be an abstract numeration system and X a subset of $\mathbb{N}$. If X is S-recognizable, then $X + t$ is also S-recognizable for any $t \in \mathbb{N}$.*

Preservation under multiplication by a constant will be studied in Chapter 2 for abstract numeration systems of the form

$$S_\ell = (a_1^* a_2^* \cdots a_\ell^*, \{a_1, a_2, \ldots, a_\ell\}, a_1 < a_2 < \cdots < a_\ell).$$

In particular we will give a characterization of the S_ℓ-recognizable sets of non-negative integers.

The following theorem is widely known but it is still worth mentioning.

Theorem 1.7.11. [**Eil74**] *Let S be the abstract numeration system built on a^*. Then a set of non-negative integers is S-recognizable if and only if it is a finite union of arithmetic progressions.*

Given an abstract numeration system S, the following result provides particular S-recognizable sets of non-negative integers.

Proposition 1.7.12. [**Sha94**] *If L is a regular language over an alphabet totally ordered by $<$, then $Min_<(L)$ and $Max_<(L)$ are regular languages.*

Since, for all abstract numeration systems $S = (L, \Sigma, <)$, the S-value of the first word of length $n + 1$ is given by $\mathbf{v}_L(n)$, the next result directly follows from the previous proposition.

Corollary 1.7.13. *Let $S = (L, \Sigma, <)$ be an abstract numeration system. Then the set $\{\mathbf{v}_L(n) \mid n \in \mathbb{N}\}$ is S-recognizable.*

It is well known that the set of squares is never U_b-recognizable; see for instance [**Eil74**]. Nevertheless we can make the following observation.

Remark 1.7.14. The set $\{n^2 \mid n \in \mathbb{N}\}$ of squares is S-recognizable for the abstract numeration system $S = (a^*b^* \cup a^*c^*, \{a, b, c\}, a < b < c)$. More precisely, it is easy to see that we have $\mathbf{v}_L(n) = (n+1)^2$ for all $n \in \mathbb{N}$. Therefore we obtain $\mathrm{rep}_S(\{n^2 \mid n \in \mathbb{N}\}) = \mathrm{Min}_<(a^*b^* \cup a^*c^*) = a^*$.

This remark is a particular case of the following theorem. In view of Corollary 1.7.13, given an infinite set of non-negative integers X, we can look for an abstract numeration system $S = (L, \Sigma, <)$ for which X is S-recognizable by requiring $X = \{\mathbf{v}_L(n) \mid n \in \mathbb{N}\}$. Such an abstract numeration system exists when X has the form given in the following result.

Theorem 1.7.15. [**Rig02**] *Let t be a positive integer, $c_1, \ldots, c_t$ be non-negative integers, and $P_1, \ldots, P_t \in \mathbb{Q}[X]$ be polynomials satisfying $P_i(\mathbb{N}) \subseteq \mathbb{N}$ for all $i \in [\![1, t]\!]$. Set*

$$f\colon \mathbb{N} \to \mathbb{N},\ n \mapsto \sum_{i=1}^{t} P_i(n)\, c_i^n.$$

The image $f(\mathbb{N})$ is S-recognizable for some abstract numeration system S which can be effectively constructed.

We also mention this last negative result.

Proposition 1.7.16. [**Rig00**] *The set of prime numbers is never S-recognizable.*

We will not go into more details here. In Chapter 4 we shall discuss the notion of S-automatic words introduced in [**Rig00**] but extended to the case of multidimensional words. In Chapter 5 we shall consider a wider class of abstract numeration systems, that is, we shall not necessarily work with regular numeration languages, and we shall give a formalism for representing real numbers in these "generalized" abstract numeration systems.

CHAPTER 2

Multiplication by a Constant

2.1. Introduction

The problem addressed in this chapter deals with the preservation of recognizability under the operation of multiplication by a constant. It can be stated as follows.

Problem 1. Let

- $S = (L, \Sigma, <)$ be an abstract numeration system;
- X be any S-recognizable set of non-negative integers;
- λ be any positive integer.

What can be said about the S-recognizability of $\lambda X = \{\lambda x \mid x \in X\}$?

Let us mention here that the material of this chapter can be found in [**CRS08**].

This problem is a first step before handling more complex operations such as addition $X + Y = \{x + y \mid x \in X,\ y \in Y\}$ or multiplication $XY = \{xy \mid x \in X,\ y \in Y\}$ of two arbitrary recognizable sets of non-negative integers X and Y. Of course, if multiplication preserves S-recognizability, *i.e.*, if the set XY is recognizable for any recognizable sets X and Y, then so multiplication by a constant does, *i.e.*, the set λX is recognizable for any recognizable set X and any positive integer λ. It is well-know that, in the case of integer base numeration systems, multiplication never preserves recognizability; for example, see [**Ber79, Sak06**]. The relationship between multiplication by a constant and addition is not so obvious. A stronger requirement would be that addition is computable by a finite automaton, that is, that the graph of addition

$$\{(\mathrm{rep}_S(x), \mathrm{rep}_S(y), \mathrm{rep}_S(x + y))^{\#} \mid x, y \in \mathbb{N}\} \subseteq ((\Sigma \cup \{\#\})^3)^*$$

is regular, where $(\cdot, \cdot, \cdot)^{\#}$ denotes the operation of padding the shorter components of a triple of words with some symbol # to make three words of the same length.[1] In this case, addition and multiplication by a constant

[1] This *padding map* will be formally defined on page 90 in Chapter 4. Also, note that, for instance, the word $(a, ba, ab)^{\#} = (\#a, ba, ab) \in (\{a, b, \#\}^3)^*$ denotes the concatenation of the letters $(\#, b, a)$ and (a, a, b) in $\{a, b, \#\}^3$.

preserve S-recognizability. Note that the converse does not hold in general, as shown by the following example.

Example 2.1.1. Let us consider the addition within the abstract numeration system S built on a^*. For this particular system, a set of non-negative integers is S-recognizable if and only if it is ultimately periodic; see Theorem 1.7.11 on page 20. Therefore addition clearly preserves S-recognizability, *i.e.*, for all S-recognizable subsets X and Y of $\mathbb{N}$, the sum $X + Y$ is S-recognizable. On the other hand, its graph is given by

$$\{(a^m, a^n, a^{m+n})^\# \mid m, n \in \mathbb{N}\} = \{(\#^n a^m, \#^m a^n, a^{m+n}) \mid m, n \in \mathbb{N}\}.$$

By applying the pumping lemma to this language, we easily show that it cannot be regular.

The question of the preservation of recognizability under arithmetic operations is a very common interest when studying numeration systems. In the framework of linear numeration system, partial answers of Problem 1 are known; see for instance [**BH97**]. More precisely, for the case of a linear numeration system defined by a recurrence relation whose characteristic polynomial is the minimal polynomial of a Pisot number[2], it is well known that addition and multiplication by a constant are computable by a finite automaton. Note that usual positional numeration systems like integer base numeration systems or the Fibonacci numeration system are special cases of these "Pisot systems". In Chapter 3 we will come back to these considerations; see page 48 below.

Recall that the class of regular languages splits into two parts with respect to the behavior of the counting function of the language: the polynomial regular languages and the exponential regular languages; see Theorem 1.5.5 on page 10. In view of Proposition 1.5.10 on page 11, the numeration language of a "Pisot system" is always exponential. On the other hand, the case of polynomial languages has not been considered yet, except in [**LR01, Rig01b**]. This new framework leads to new phenomena.

Definition 2.1.2. For any positive integer ℓ, we let $\mathcal{B}_\ell = a_1^* a_2^* \cdots a_\ell^*$ denote the *bounded language* over the alphabet $\Sigma_\ell = \{a_1, a_2, \ldots, a_\ell\}$ and we let $S_\ell = (\mathcal{B}_\ell, \Sigma_\ell, <)$ denote the corresponding abstract numeration system where the total order $<$ on Σ_ℓ is given by $a_1 < a_2 < \cdots < a_\ell$.

[2]A Pisot number is an algebraic integer $\beta > 1$ such that all its Galois conjugates γ satisfy $|\gamma| < 1$.

In Figure 2.1 we have depicted the trim minimal automaton $\mathcal{A}_\ell$ of $\mathcal{B}_\ell$. It has $[\![1,\ell]\!]$ as set of states. Each state is final, 1 is initial and, for all integers i and j satisfying $1 \le i \le j \le \ell$, we have a transition $i \xrightarrow{a_j} j$. In

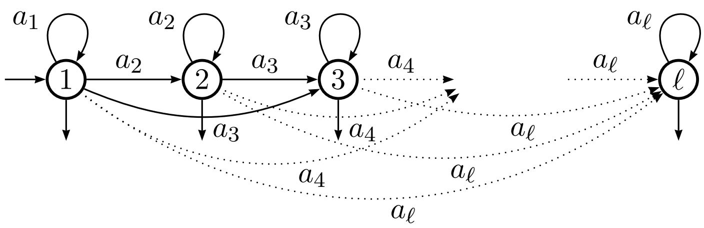

FIGURE 2.1. The trim minimal automaton $\mathcal{A}_\ell$ of $\mathcal{B}_\ell$.

examples, when considering cases $\ell = 2$ or $\ell = 3$, we shall sometimes use alphabets like $\{a, b\}$ or $\{a, b, c\}$ instead of $\{a_1, a_2\}$ or $\{a_1, a_2, a_3\}$. Note that these particular abstract numeration systems S_ℓ do not correspond to any positional numeration system as introduced in Definition 1.6.1 on page 12.

Example 2.1.3. Consider the alphabet $\Sigma_2 = \{a, b\}$ with $a < b$. The first few words of $\mathcal{B}_2 = a^*b^*$ enumerated with respect to the genealogical order are $\varepsilon, a, b, aa, ab, bb, aaa, aab, abb, bbb, aaaa$. For instance, we have $\mathrm{rep}_{S_2}(5) = bb$ and $\mathrm{val}_{S_2}(8) = abb$. Furthemore, observe that $\mathrm{val}_{S_2}(a^*) = \{0, 1, 3, 6, 10, 15, \ldots\}$ is a S_2-recognizable subset of $\mathbb{N}$ (formed of all triangular numbers).

Bounded languages are good candidates to start with. Indeed, a polynomial regular language is a finite union of languages of the form (3) given in Theorem 1.5.4 on page 10 and automata accepting these languages share the same properties as those accepting bounded languages. Therefore we hope that our results give the flavor of what could be expected for any polynomial languages.

For more details on bounded languages, see for instance [**GH64**]. Also, note that this map val_{S_ℓ} is a special case of diagonal function as considered for instance in [**LMSF96**] and defined as follows.

Definition 2.1.4. Let ℓ be a positive integer. A *diagonal function of dimension* ℓ is a bijection $f\colon \mathbb{N}^\ell \to \mathbb{N}$ such that, for all $\mathbf{m}, \mathbf{n} \in \mathbb{N}^\ell$, we have $f(\mathbf{m}) < f(\mathbf{n})$ whenever $s(\mathbf{m}) < s(\mathbf{n})$, with $s(\mathbf{n}) = \sum_{i=1}^{\ell} n_i$ for all $\mathbf{n} = (n_1, \ldots, n_\ell) \in \mathbb{N}^\ell$.

Further on, we will see that val_{S_ℓ} is even a diagonal polynomial, that is, a diagonal function which is a polynomial.

Since rep_{S_ℓ} is a one-to-one correspondence between $\mathbb{N}$ and $\mathcal{B}_\ell$, multiplication by a constant $\lambda \in \mathbb{N}$ can be viewed as a transformation $f_\lambda : \mathcal{B}_\ell \to \mathcal{B}_\ell$ acting on the language $\mathcal{B}_\ell$, the question being then to study the preservation of the regularity of the subsets of $\mathcal{B}_\ell$ under this transformation.

Example 2.1.5. Take $\ell = 2$, $\Sigma_2 = \{a, b\}$ and $\lambda = 25$. We have the following diagrams.

$$\begin{array}{ccc} 8 & \xrightarrow{\times 25} & 200 \\ {\scriptstyle \mathrm{rep}_{S_2}}\downarrow & & \downarrow{\scriptstyle \mathrm{rep}_{S_2}} \\ ab^2 & \xrightarrow[f_{25}]{} & a^9b^{10} \end{array} \qquad \begin{array}{ccc} \mathbb{N} & \xrightarrow{\times \lambda} & \mathbb{N} \\ {\scriptstyle \mathrm{rep}_{S_\ell}}\downarrow & & \downarrow{\scriptstyle \mathrm{rep}_{S_\ell}} \\ \mathcal{B}_\ell & \xrightarrow[f_\lambda]{} & \mathcal{B}_\ell \end{array}$$

Thus multiplication by $\lambda = 25$ induces a mapping f_λ onto $\mathcal{B}_2$ such that, for all $w, w' \in \mathcal{B}_2$, we have $f_\lambda(w) = w' \Leftrightarrow \mathrm{val}_{S_2}(w') = 25\ \mathrm{val}_{S_2}(w)$.

This chapter follows the organization given below. In the first section we recall a few results related to our main question. In particular, we characterize the recognizable sets of non-negative integers for abstract numeration systems whose language is slender, *i.e.*, has at most d words of each length for some constant d. We easily obtain that, in this situation, multiplication by a constant always preserves recognizability.

In Section 2.3 we compute $\mathrm{val}_{S_\ell}(a_1^{n_1} \cdots a_\ell^{n_\ell})$ for any positive integer ℓ and any non-negative integers $n_1, \ldots, n_\ell$. Then, we derive an easy proof of the fact that any non-negative integer can be decomposed in a unique way as

$$n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell-1} + \cdots + \binom{z_1}{1},$$

where $z_1, \ldots, z_\ell$ are integers satisfying $z_\ell > z_{\ell-1} > \cdots > z_1 \geq 0$. In [**Fra85**] A. Fraenkel called this system the *combinatorial numeration system* and referred to D. Lehmer [**Leh64**]. Even if this seems to be a folklore result, the only proof that we were able to trace out dates back to G. Katona [**Kat68**] who developed different (and quite long) arguments to obtain the same decomposition.

Next, in Section 2.4, we make the regular subsets of $\mathcal{B}_\ell$ explicit in terms of semi-linear subsets of $\mathbb{N}^\ell$ and we give an application to the S_ℓ-recognizability of arithmetic progressions.

In Section 2.5, first, we provide a formula that can be used to obtain estimates on the S_ℓ-representation of λn from the one of n, for any non-negative integers λ and n. Second, thanks to a counting argument and to the results from Section 2.4, we show that, if ℓ is an integer greater than or equal to 3, then, for any positive integer λ, there exists a S_ℓ-recognizable set X such that λX is not S_ℓ-recognizable.

In Section 2.6 we answer our main question about bounded languages and recognizability after multiplication by a constant. More precisely, our main result — Theorem 2.6.1 — in this chapter can be stated as follows:

Theorem. *Let ℓ and λ be positive integers with $\lambda \geq 2$. Then multiplication by λ preserves S_ℓ-recognizability if and only if either ℓ equals* 1 *or, ℓ equals* 2 *and λ is an odd square.*

Finally, we put in the last section some structural results regarding the effect of multiplication by a constant in the abstract numeration system built on $\mathcal{B}_\ell$.

2.2. First Results about S-Recognizability

In this section we collect a few results directly related with our problem. The first result states that only some constants λ are good candidates for multiplication within $\mathcal{B}_\ell$.

Theorem 2.2.1. [**Rig01b**] *Let L be an infinite regular language over an alphabet Σ such that is counting function $\mathbf{u}_L(n)$ is $\Theta(n^k)$ for some $k \in \mathbb{N}$, let $S = (L, \Sigma, <)$ be an abstract numeration system built on L, and let λ be a non-negative integer. Preservation of S-recognizability after multiplication by λ holds only if we have $\lambda = \beta^{k+1}$ for some $\beta \in \mathbb{N}$. Otherwise stated, if we have $\lambda \neq \beta^{k+1}$ for all $\beta \in \mathbb{N}$, then there exists an S-recognizable set $X \subseteq \mathbb{N}$ such that λX is not S-recognizable.*

Since we shall see in the next section that, for all positive integers ℓ, the counting function $\mathbf{u}_{\mathcal{B}_\ell}(n)$ of $\mathcal{B}_\ell$ is $\Theta(n^{\ell-1})$, thanks to this theorem, we only have to focus on multipliers of the form β^ℓ, with $\beta \in \mathbb{N}$. The particular case $\mathbf{u}_L(n) = O(1)$ is interesting in itself and is settled as follows. First, let us recall the definition from [**APDS93**] and the characterization from [**PS95, Sha94**] of slender regular languages.

Definition 2.2.2. Let d be a non-negative integer. A language L is said to be *d-slender* if, for all $n \in \mathbb{N}$, we have $\mathbf{u}_L(n) \leq d$. A language is said to be *slender* if it is d-slender for some $d \in \mathbb{N}$.

Definition 2.2.3. A language L is a *union of single loops* if, for some $k \in \mathbb{N}$ and words x_i, y_i, z_i, with $i \in [\![1, k]\!]$, we have

$$L = \bigcup_{i=1}^{k} x_i\, y_i^*\, z_i. \tag{5}$$

A language L is a *disjoint union of single loops* if the sets $x_i\, y_i^* z_i$ in the union (5) are pairwise disjoint.

Lemma 2.2.4. [**PS95, Sha94**] *Let L be a regular language. The following conditions are equivalent:*

- *L is slender;*
- *L is a union of single loops;*
- *L is a disjoint union of single loops.*

The following theorem already appeared in the doctoral dissertation of M. Rigo [**Rig01a**]. We restate the proof here for the sake of completeness.

Theorem 2.2.5. *Let S be an abstract numeration system built on a slender regular language over an alphabet Σ. A set $X \subseteq \mathbb{N}$ is S-recognizable if and only if it is a finite union of arithmetic progressions.*

Proof. Let L be an infinite slender regular language over an alphabet Σ and let $S = (L, \Sigma, <)$ be an abstract numeration system. From the above characterization of slender regular languages, we can write

$$L = \bigcup_{i=1}^{k} x_i\, y_i^* z_i \cup F$$

for some finite set $F \subseteq \Sigma^*$, some $k \in \mathbb{N} \setminus \{0\}$, and some words $x_i, y_i, z_i \in \Sigma^*$, with $i \in [\![1, k]\!]$, such that we have $y_i \neq \varepsilon$ for all $i \in [\![1, k]\!]$ and such that the sets $x_i\, y_i^* z_i$ are pairwise disjoint. The sequence $(\mathbf{u}_L(n))_{n\geq 0}$ is ultimately periodic with period $p = \operatorname{lcm}(|y_1|, \ldots, |y_k|)$. Furthermore, for large enough integers n, if $x_i\, y_i^n\, z_i$ is the m-th word of length $|x_i\, z_i| + n\,|y_i|$ with respect to the genealogical order, then $x_i\, y_i^{n+p/|y_i|}\, z_i$ is the m-th word of length $|x_i\, z_i| + n\,|y_i| + p$ with respect to the same order. Roughly speaking, for sufficiently large integers n, the structures of the ordered sets of words of length n and $n + p$ are the same.

The regular subsets of L are of the form

$$\bigcup_{j\in J} x_{i_j}\, (y_{i_j}^{\alpha_j})^* z_{i_j} \cup F' \tag{6}$$

with J a finite set, $i_j \in [\![1, k]\!]$ and $\alpha_j \in \mathbb{N} \setminus \{0\}$ for all $j \in J$, and F' a finite subset of L.

It is possible to conclude now. If X is S-recognizable, then $\operatorname{rep}_S(X)$ is a regular subset of L of the form (6). Hence, in view of the first part of the proof, X is ultimately periodic with period $\operatorname{lcm}(p, \operatorname{lcm}(|y_{i_j}^{\alpha_j}|, j \in J))$. The converse is straightforward by using Theorem 1.7.9 on page 19. □

Example 2.2.6. Consider the language $L = ab^*c \cup b(aa)^*c$. It contains exactly two words of each positive even length and one word of each odd length larger than 2, that is, $ab^{2i}c < ba^{2i}c$ and $ab^{2i+1}c$ with $i \in \mathbb{N}$. The sequence $(\mathbf{u}_L(n))_{n\geq 0}$ is ultimately periodic of period two: $0, 0, 2, 1, 2, 1, \ldots$.

Corollary 2.2.7. *Let S be an abstract numeration system built on a slender language and let X be a set of non-negative integers. If X is S-recognizable, then λX is S-recognizable for any $\lambda \in \mathbb{N}$.*

Corollary 2.2.8. *Let S be an abstract numeration system built on a slender language and let X and Y be sets of non-negative integers. If X and Y are S-recognizable, then $X + Y$ is S-recognizable.*

Theorem 2.2.1 shows that, for abstract numeration systems S built on polynomial languages, multiplication by a constant does not generally preserve S-recognizability. In general abstract numeration systems built on exponential languages with polynomial complement do not preserve S-recognizability after multiplication by a constant either.

Theorem 2.2.9. [**Rig01b**] *Let Σ be an alphabet with $\mathrm{Card}(\Sigma) \geq 2$, let L be an infinite polynomial regular language over Σ, and let S be an abstract numeration system built on its complement $\Sigma^* \setminus L$. Then there exists an S-recognizable set $X \subseteq \mathbb{N}$ and a positive integer t such that tX is not S-recognizable.*

Finally, for a bounded language over a binary alphabet, the case is completely settled too. The aim of this study was initially to extend the following result.

Theorem 2.2.10. [**LR01**] *Let β be a positive integer. For the abstract numeration system $S = (a^*b^*, \{a, b\}, a < b)$, multiplication by β^2 preserves S-recognizability if and only if β is odd.*

2.3. S_ℓ-Representation of Integers: Combinatorial Expansion

In this section we determine the number of words of a given length in $\mathcal{B}_\ell$ and we obtain an algorithm for computing $\mathrm{rep}_{S_\ell}(n)$ for any non-negative integer n. Interestingly, this algorithm is related to the decomposition of n as a sum of binomial coefficients of a specific form.

Let us recall that the binomial coefficient $\binom{i}{j}$ vanishes for integers $i < j$.

Lemma 2.3.1. *For all $\ell \in \mathbb{N} \setminus \{0\}$ and all $n \in \mathbb{N}$, we have*

$$\mathbf{u}_{\mathcal{B}_{\ell+1}}(n) = \mathbf{v}_{\mathcal{B}_\ell}(n) \tag{7}$$

and

$$\mathbf{u}_{\mathcal{B}_\ell}(n) = \binom{n+\ell-1}{\ell-1}. \tag{8}$$

Proof. Let ℓ be a positive integer. Relation (7) can be deduced from the fact that the set of words of length n belonging to $\mathcal{B}_{\ell+1}$ is partitioned as follows:

$$\mathcal{B}_{\ell+1} \cap \Sigma_\ell^n = \bigcup_{i=0}^{n} (a_1^* \cdots a_\ell^* \cap \Sigma_\ell^i) a_{\ell+1}^{n-i}.$$

To obtain (8), we proceed by induction on ℓ. For $\ell = 1$, it is clear that we have $\mathbf{u}_{\mathcal{B}_1}(n) = 1$ for all $n \in \mathbb{N}$. Assume that (8) is satisfied for ℓ and let us verify that it holds true for $\ell + 1$. Thanks to (7), we obtain

$$\forall n \in \mathbb{N},\ \mathbf{u}_{\mathcal{B}_{\ell+1}}(n) = \sum_{i=0}^{n} \mathbf{u}_{\mathcal{B}_\ell}(i) = \sum_{i=0}^{n} \binom{i+\ell-1}{\ell-1} = \binom{n+\ell}{\ell},$$

as desired. □

Lemma 2.3.2. *For all $\ell \in \mathbb{N} \setminus \{0\}$ and all $n_1, \ldots, n_\ell \in \mathbb{N}$, we have*

$$\mathrm{val}_{S_\ell}(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \sum_{i=1}^{\ell} \binom{n_i + \cdots + n_\ell + \ell - i}{\ell - i + 1}. \tag{9}$$

Proof. Let ℓ be a positive integer. It follows from the structure of the ordered language $\mathcal{B}_\ell$ that, for all $n \in \mathbb{N}$, the ordered list of words of length n in $\mathcal{B}_\ell$ contains an ordered copy of the words of length at most n in the language $a_2^* \cdots a_\ell^*$. To get this, we apply the erasing morphism $\varphi\colon \Sigma_\ell^* \to \Sigma_{\ell-1}^*$ defined by $\varphi(a_1) = \varepsilon$ and, for all $i \in [\![2, \ell]\!]$, $\varphi(a_i) = a_i$. Hence we obtain

$$\forall n_1, \ldots, n_\ell \in \mathbb{N},\ \mathrm{val}_{S_\ell}(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \mathrm{val}_{S_\ell}(a_1^{n_1+\cdots+n_\ell}) + \mathrm{val}_{S_{\ell-1}}(a_1^{n_2} \cdots a_{\ell-1}^{n_\ell}).$$

By iterating the latter decomposition, we obtain

$$\forall n_1, \ldots, n_\ell \in \mathbb{N},\ \mathrm{val}_{S_\ell}(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \sum_{i=1}^{\ell} \mathrm{val}_{S_{\ell-i+1}}(a_1^{n_i+\cdots+n_\ell}). \tag{10}$$

Since, for all $n \in \mathbb{N}$, the first word of length n with respect to the genealogical order is a_1^n, we have $\mathrm{val}_{S_\ell}(a_1^n) = \mathbf{v}_{\mathcal{B}_\ell}(n-1)$ where we set $\mathbf{v}_{\mathcal{B}_\ell}(-1) = 0$. We finish the proof by using relations (7) and (8). □

The example below illustrates the proof of Lemma 2.3.2 in the case $\ell = 3$.

Example 2.3.3. Consider the words of length 3 in the language $a^*b^*c^*$:

$$aaa < aab < aac < abb < abc < acc < bbb < bbc < bcc < ccc.$$

We have $\mathrm{val}_{S_3}(aaa) = \binom{5}{3} = 10$ and $\mathrm{val}_{S_3}(acc) = 15$. Applying the erasing morphism $\varphi : \{a,b,c\}^* \to \{b,c\}^*$ defined by $\varphi(a) = \varepsilon$, $\varphi(b) = b$ and $\varphi(c) = c$ on the words of length 3, we get

$$\varepsilon < b < c < bb < bc < cc < bbb < bbc < bcc < ccc.$$

So the ordered list of words of length 3 in $a^*b^*c^*$ contains an ordered copy of the words of length at most 3 in the language b^*c^*. In addition, to obtain $\mathrm{val}_{S_3}(acc)$, we just add to $\mathrm{val}_{S_3}(aaa)$ the position of the word cc in the ordered language b^*c^*, which is equal to $\mathrm{val}_{S_2}(bb)$ with our notation.

Corollary 2.3.4. *For any positive integer ℓ, the mapping $\mathrm{val}_{S_\ell} : \mathbb{N}^\ell \to \mathbb{N}$ is a diagonal polynomial of dimension ℓ.*

The following result is given in [**Kat68**]. Here we obtain a bijective proof only relying on the use of abstract numeration systems on a bounded language.

Corollary 2.3.5 (Combinatorial numeration system)**.** *Let ℓ be a positive integer. Any non-negative integer n can be uniquely written as*

$$n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell-1} + \cdots + \binom{z_1}{1}, \tag{11}$$

where $z_1, \ldots, z_\ell$ are integers satisfying $z_\ell > z_{\ell-1} > \cdots > z_1 \geq 0$.

Proof. The mapping $\mathrm{rep}_{S_\ell} : \mathbb{N} \to a_1^* \cdots a_\ell^*$ is a one-to-one correspondence. So any non-negative integer n has a unique representation of the form $a_1^{n_1} \cdots a_\ell^{n_\ell}$, with $n_1, \ldots, n_\ell \in \mathbb{N}$. Now, the result follows from Lemma 2.3.2. □

The general method given in [**LR01**, Algorithm 1] for computing the S-value functions val_S has a special form in the case of the language $\mathcal{B}_\ell$. We derive an algorithm computing the decomposition (11) or equivalently the S_ℓ-representation of any integer.

Algorithm. Let $\mathtt{n} \in \mathbb{N}$ and $\mathtt{l} \in \mathbb{N} \setminus \{0\}$. The following algorithm produces the integers `z(l)`, ..., `z(1)` corresponding to the z_i's appearing in the decomposition (11) of `n` given in Corollary 2.3.5.

```
For i=l,l-1,...,1 do
   if n>0,
      find t such that (t choose i) ≤ n < (t+1 choose i)
```

```
        z(i)←t
        n←n-(t i)
    otherwise, z(i)←i-1
```

Now, consider the following triangular system having $n_1, \dots, n_\ell \in \mathbb{N}$ as unknowns:

$$n_i + \cdots + n_\ell = \mathtt{z}(\ell - i + 1) - \ell + i, \quad i = 1, \dots, \ell.$$

We have $\mathrm{rep}_{S_\ell}(\mathtt{n}) = a_1^{n_1} \cdots a_\ell^{n_\ell}$.

Remark 2.3.6. To speed up the computation of $\mathtt{t}$ in the above algorithm, one can benefit from methods of numerical analysis. Indeed, for given $\mathtt{i}$ and $\mathtt{n}$, $\binom{\mathtt{t}}{\mathtt{i}} - \mathtt{n}$ is a polynomial in $\mathtt{t}$ of degree $\mathtt{i}$ and we are looking for the largest root z of this polynomial. Therefore we have $\mathtt{t} = \lfloor z \rfloor$.

Example 2.3.7. For $\ell = 3$, one gets for instance

$$12345678901234567890 = \binom{4199737}{3} + \binom{3803913}{2} + \binom{1580642}{1}.$$

Solving the system

$$\left.\begin{array}{rcl} n_1 + n_2 + n_3 & = & 4199737 - 2 \\ n_2 + n_3 & = & 3803913 - 1 \\ n_3 & = & 1580642 \end{array}\right\}$$

$$\Leftrightarrow (n_1, n_2, n_3) = (395823, 2223270, 1580642),$$

we find $\mathrm{rep}_{S_3}(12345678901234567890) = a^{395823} b^{2223270} c^{1580642}$.

2.4. Regular Subsets of $\mathcal{B}_\ell$

Let ℓ be a positive integer that will be kept constant throughout this section. To study the preservation of S_ℓ-recognizability after multiplication by a constant λ, one has to consider an arbitrary S_ℓ-recognizable subset X of $\mathbb{N}$ and verify whether or not λX is still S_ℓ-recognizable. To that end, let us observe that the regular subsets of $\mathcal{B}_\ell$ are the finite unions of languages of the form

$$a_1^{m_1}(a_1^{n_1})^* \cdots a_\ell^{m_\ell}(a_\ell^{n_\ell})^*, \tag{12}$$

with $m_i, n_i \in \mathbb{N}$ for all $i \in [\![1, \ell]\!]$.

Definition 2.4.1. For any word w over Σ_ℓ and $j \in [\![1, \ell]\!]$, we let $|w|_{a_j}$ denote the number of occurrences of the letter a_j in w. The *Parikh mapping* Ψ maps a word $w \in \Sigma_\ell^*$ onto the vector $\Psi(w) = (|w|_{a_1}, \dots, |w|_{a_\ell}) \in \mathbb{N}^\ell$.

Remark 2.4.2. In this setting of bounded languages, rep_{S_ℓ} and $\Psi_{|\mathcal{B}_\ell}$ are both one-to-one correspondences. Therefore, in what follows, we shall make no distinction between a non-negative integer $n \in \mathbb{N}$, its S_ℓ-representation $\mathrm{rep}_{S_\ell}(n) = a_1^{n_1} \cdots a_\ell^{n_\ell} \in \mathcal{B}_\ell$, and the corresponding Parikh vector

$$\Psi(\mathrm{rep}_{S_\ell}(n)) = (n_1, \ldots, n_\ell) \in \mathbb{N}^\ell.$$

Definition 2.4.3. A set $Z \subseteq \mathbb{N}^\ell$ is *linear* if there exist some $k \in \mathbb{N}$ and some $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_k \in \mathbb{N}^\ell$ such that we have

$$Z = \mathbf{p}_0 + \mathbb{N}\,\mathbf{p}_1 + \cdots + \mathbb{N}\,\mathbf{p}_k = \{\mathbf{p}_0 + \lambda_1\mathbf{p}_1 + \cdots + \lambda_k\mathbf{p}_k \mid \lambda_1, \ldots, \lambda_k \in \mathbb{N}\}.$$

The vectors $\mathbf{p}_1, \ldots, \mathbf{p}_k$ are said to be the *periods* of Z. The set Z is k-*dimensional* if it has exactly k linearly independent periods over $\mathbb{Q}$. A set is *semi-linear* if it is a finite union of linear sets. A vector is a *period of a semi-linear set* if it is a period of one of the corresponding linear sets.

Even if we shall make no use of if, it is probably worth recalling here the following directly related result from [**Par66**].

Theorem 2.4.4. [**Par66**] *If $L \subseteq \Sigma_\ell^*$ is a context-free language, then $\Psi(L)$ is a semi-linear subset of $\mathbb{N}^\ell$.*

For all $i \in [\![1, \ell]\!]$, we let $\mathbf{e}_i \in \mathbb{N}^\ell$ denote the canonical vector having 1 at the i-th component and 0 at the other components. In view of the general form (12) of the regular subsets of $\mathcal{B}_\ell$, the following result is obvious.

Lemma 2.4.5. *A set $X \subseteq \mathbb{N}$ is S_ℓ-recognizable if and only if $\Psi(\mathrm{rep}_{S_\ell}(X))$ is a semi-linear set whose periods are integer multiples of canonical vectors.*

With such a characterization, it is not difficult to obtain an alternative proof of Theorem 1.7.9 on page 19 restricted to the case of abstract numeration systems built on bounded languages.

Proposition 2.4.6. *Let p and q be non-negative integers. Then the set $\Psi(\mathrm{rep}_{S_\ell}(p + \mathbb{N}q)) \subseteq \mathbb{N}^\ell$ is a finite union of linear sets of the form*

$$\mathbf{n} + \mathbb{N}\,r\,\mathbf{e}_1 + \cdots + \mathbb{N}\,r\,\mathbf{e}_\ell$$

for some $\mathbf{n} \in \mathbb{N}^\ell$ and $r \in \mathbb{N}$.

Proof. If q is zero, then the result is straightforward. Now, assume $q > 0$. We make use of Equation (9). The sequences $\left(\binom{n}{\ell-i+1} \bmod q\right)_{n\geq 0}$ are (purely) periodic (see *e.g.* [**Zab56**]), say with period $\pi_i(q)$. Let us define $P = \mathrm{lcm}(\pi_1(q), \ldots, \pi_\ell(q))$. Then, for all $i \in [\![1, \ell]\!]$ and all $n_1, \ldots, n_\ell \in \mathbb{N}$, we have

$$\mathrm{val}_{S_\ell}(a_1^{n_1}\cdots a_i^{n_i}\cdots a_\ell^{n_\ell}) \equiv \mathrm{val}_{S_\ell}(a_1^{n_1}\cdots a_i^{n_i+P}\cdots a_\ell^{n_\ell}) \pmod q.$$

We have just shown that any $\mathbf{n}\in\mathbb{N}^\ell$ belongs to $\Psi(\mathrm{rep}_{S_\ell}(p+\mathbb{N}q))$ if and only if $\mathbf{n}+c_1P\mathbf{e}_1+\cdots+c_\ell P\mathbf{e}_\ell$ belongs to the same set for all $c_1,\ldots,c_\ell\in\mathbb{N}$. Now, the conclusion is straightforward: we obtain that $\Psi(\mathrm{rep}_{S_\ell}(p+\mathbb{N}q))$ is the finite union

$$\bigcup_{\substack{(n_1,\ldots,n_\ell)\in\Psi(\mathrm{rep}_{S_\ell}(p+\mathbb{N}q))\\ |\mathrm{rep}_S(p)|\le n_1+\ldots+n_\ell\le|\mathrm{rep}_S(p)|+\ell(P-1)}} ((n_1,\ldots,n_\ell)+\mathbb{N}P\mathbf{e}_1+\cdots+\mathbb{N}P\mathbf{e}_\ell)\,.$$

□

Example 2.4.7. In Figure 2.2 the x-axis (resp. y-axis) counts the number of letters a (resp. b) in a word. The empty word corresponds to the lower-left corner. A point in $\mathbb{N}^2$ of coordinates (i,j) has its color determined by the value of $\mathrm{val}_{S_2}(a^ib^j)$ modulo q (with $q=3,5,6$ and 8 respectively). Therefore there are q possible colors. In this figure we represent the words a^ib^j for $0\le i,j\le 19$.

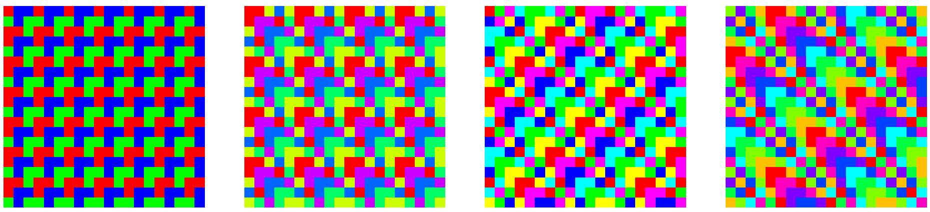

FIGURE 2.2. $\Psi(\mathrm{rep}_{S_2}(p+\mathbb{N}\,q))$ for $q=3,5,6,8$.

2.5. Multiplication by $\lambda=\beta^\ell$

In the case of a bounded language on ℓ letters, if multiplication by some constant preserves S_ℓ-recognizability, then, from Theorem 2.2.1 and Lemma 2.3.1, this constant must be a ℓ-th power of a non-negative integer β. Let us fix two integers ℓ and β for the whole section. Since the case $\ell=1$ corresponds to the slender case already discussed in the previous section, the case $\ell=2$ is the one of Theorem 2.2.10, and the case $\beta=1$ is trivial, we assume that the integers ℓ and β satisfy $\ell\ge 3$ and $\beta\ge 2$.

In this section we shall make use of the Stirling numbers of the first kind, which are defined as follows.

Definition 2.5.1. The Stirling numbers $s(i,j)$, for $i,j \in \mathbb{N}$, of the first kind are defined by

$$i!\binom{x+i-1}{i} = x(x+1)\cdots(x+i-1) = \sum_{j=1}^{i} s(i,j)x^j$$

and $s(i,j)=0$ for $i<j$ or $j=0$.

Also, recall the following useful fact.

Lemma 2.5.2. *The Stirling numbers of the first kind satisfy the recursion relation*

$$\forall i \in \mathbb{N}\setminus\{0\},\ \forall j \in [\![1,i]\!],\ s(i+1,j) = s(i,j-1) + i\, s(i,j).$$

Notation. For a polynomial $P = c_k x^k + \cdots + c_1 x + c_0$ of degree k and for all $i \in [\![0,k]\!]$, we let $[x^i]P$ denote the coefficient c_i of x^i in P.

The next lemma also holds for $\ell = 2$. Recall that the map f_λ is defined by $f_\lambda(w) = \operatorname{rep}_{S_\ell}(\lambda \operatorname{val}_{S_\ell}(w)$ for $w \in \mathcal{B}_\ell$.

Lemma 2.5.3. *For sufficiently large $q \in \mathbb{N}$, we have*

$$|f_{\beta^\ell}(a_1^q)| = \beta q + \left\lceil \frac{(\beta-1)(\ell+1)}{2} \right\rceil - \beta \quad \textit{and} \quad |f_{\beta^\ell}(a_\ell^q)| = \beta q + \left\lceil \frac{(\beta-1)(\ell+1)}{2} \right\rceil.$$

Proof. Set $c = (\beta-1)(\ell+1)/2 > 0$. Observe that since, for all non-negative integers q, we have $\operatorname{val}_{S_\ell}(a_\ell^q) = \operatorname{val}_{S_\ell}(a_1^{q+1}) - 1$, we only need to show that the following two inequalities are satisfied whenever q is large enough:

$$|f_{\beta^\ell}(a_1^q))| \le \beta q + \lceil c \rceil - \beta; \tag{13}$$

$$|f_{\beta^\ell}(a_\ell^q))| \ge \beta q + \lceil c \rceil. \tag{14}$$

Indeed, for all sufficiently large integers q, this implies

$$|f_{\beta^\ell}(a_\ell^q))| \le |f_{\beta^\ell}(a_1^{q+1}))| \le \beta(q+1) + \lceil c \rceil - \beta = \beta q + \lceil c \rceil;$$

$$|f_{\beta^\ell}(a_1^q))| \ge |f_{\beta^\ell}(a_\ell^{q-1}))| \ge \beta(q-1) + \lceil c \rceil = \beta q + \lceil c \rceil - \beta.$$

First, let us prove (14). A non-negative integer n has an S_ℓ-representation $\operatorname{rep}_{S_\ell}(n)$ of length q if and only if it satisfies $\operatorname{val}_{S_\ell}(a_1^q) \le n < \operatorname{val}_{S_\ell}(a_1^{q+1})$. Therefore, using Lemma 2.3.2, the inequality (14) is satisfied if and only if we have

$$\beta^\ell \operatorname{val}_{S_\ell}(a_\ell^q) = \beta^\ell \sum_{i=1}^{\ell} \binom{q+i-1}{i} \ge \operatorname{val}_{S_\ell}(a_1^{\beta q + \lceil c \rceil}) = \binom{\beta q + \lceil c \rceil + \ell - 1}{\ell}.$$

We easily compute

$$\beta^\ell \sum_{i=1}^{\ell}\binom{q+i-1}{i}-\binom{\beta q+\lceil c\rceil+\ell-1}{\ell}=(c+1-\lceil c\rceil)\frac{(\beta q)^{\ell-1}}{(\ell-1)!}+O(q^{\ell-2}).$$

Note that we are comparing polynomials in q. Hence (14) is proved.

Second, let us prove (13). Using the same arguments as above, this happens if and only if we have

$$\beta^\ell \operatorname{val}_{S_\ell}(a_1^q)=\beta^\ell\binom{q+\ell-1}{\ell}<\operatorname{val}_{S_\ell}(a_1^{\beta q+\lceil c\rceil-\beta+1})=\binom{\beta q+\lceil c\rceil-\beta+\ell}{\ell}.$$

We get

$$\binom{\beta q+\lceil c\rceil-\beta+\ell}{\ell}-\beta^\ell\binom{q+\ell-1}{\ell}=(\lceil c\rceil-c)\frac{(\beta q)^{\ell-1}}{(\ell-1)!}+O(q^{\ell-2}).$$

If we have $c\notin\mathbb{N}$, then we are done. Now, assume $c\in\mathbb{N}$. Hence we have $\lceil c\rceil=c$ and, in the last expression, the coefficient of $q^{\ell-1}$ vanishes. Let us compute the coefficient of $q^{\ell-2}$. Set $d=c-\beta$. By involving Stirling numbers we can write

$$[q^{\ell-2}]\binom{q+\ell-1}{\ell}=\frac{s(\ell,\ell-2)}{\ell!}$$

and

$$\begin{aligned}
[q^{\ell-2}]\quad&\binom{\beta q+d+\ell}{\ell}=[q^{\ell-1}]\frac{1}{\beta}\left((\ell+1)\binom{\beta q+d+\ell}{\ell+1}-d\binom{\beta q+d+\ell}{\ell}\right)\\
=\quad&[q^{\ell-1}]\frac{1}{\beta}\left(\sum_{i=1}^{\ell+1}\frac{s(\ell+1,i)}{\ell!}(\beta q+d)^i-d\binom{\beta q+d+\ell}{\ell}\right)\\
=\quad&[q^{\ell-1}]\frac{1}{\beta}\left(\sum_{i=1}^{\ell+1}\frac{s(\ell+1,i)}{\ell!}\sum_{k=0}^{i}\binom{i}{k}d^{i-k}\beta^k q^k-d\binom{\beta q+d+\ell}{\ell}\right)\\
=\quad&[q^{\ell-1}]\frac{1}{\beta}\left(\sum_{k=1}^{\ell+1}\left(\beta^k\sum_{i=k}^{\ell+1}\frac{s(\ell+1,i)}{\ell!}\binom{i}{k}d^{i-k}\right)q^k-d\binom{\beta q+d+\ell}{\ell}\right)\\
=\quad&\beta^{\ell-2}\sum_{i=\ell-1}^{\ell+1}\frac{s(\ell+1,i)}{\ell!}\binom{i}{\ell-1}d^{i-\ell+1}-\frac{d\beta^{\ell-2}}{\ell!}\sum_{i=1}^{\ell}(d+i)\\
=\quad&\frac{\beta^{\ell-2}}{\ell!}\left(s(\ell+1,\ell-1)+\frac{\ell^2(\ell+1)}{2}d+\frac{\ell(\ell+1)}{2}d^2-\ell d^2-\frac{\ell(\ell+1)}{2}d\right)\\
=\quad&\frac{\beta^{\ell-2}}{\ell!}\left(s(\ell+1,\ell-1)+\frac{(\ell-1)\ell(\ell+1)}{2}d+\frac{(\ell-1)\ell}{2}d^2\right).
\end{aligned}$$

Then, by using Lemma 2.5.2, we easily find

$$s(\ell+1,\ell-1)=s(\ell,\ell-2)+\frac{(\ell-1)\ell^2}{2}=\frac{(3\ell+2)(\ell+1)\ell(\ell-1)}{24}. \quad (15)$$

Finally we obtain

$$\binom{\beta q + c - \beta + \ell}{\ell} - \beta^\ell \binom{q + \ell - 1}{\ell} = \frac{c(\beta+1)}{12} \frac{(\beta q)^{\ell-2}}{(\ell-2)!} + O(q^{\ell-3}),$$

which shows that the inequality (13) also holds in this case whenever q is sufficiently large. $\square$

The following corollary provides a relationship between the lengths of the S_ℓ-representations of n and $\beta^\ell n$, roughly by a factor β.

Corollary 2.5.4. *For sufficiently large $n \in \mathbb{N}$, we have*

$$|\operatorname{rep}_{S_\ell}(\beta^\ell n)| = \beta\, |\operatorname{rep}_{S_\ell}(n)| + \left\lceil \frac{(\beta-1)(\ell+1)}{2} \right\rceil - i$$

for some $i \in [\![0, \beta]\!]$.

In certain cases we can provide a formula for the entire expansion of $\beta^\ell \operatorname{val}_{S_\ell}(a_\ell^q)$ for all large enough integers q. Using the same kind of computation we could also obtain a formula for the entire expansion of $\beta^\ell \operatorname{val}_{S_\ell}(a_1^q)$ or even of any $\beta^\ell n$ with $|\operatorname{rep}_{S_\ell}(n)| = q$.

Definition 2.5.5. Define $c_\ell, c_{\ell-1}, \ldots, c_1$ recursively by

$$c_{k+1} = k!\,(\beta^{\ell-k} - 1) \sum_{i=k}^{\ell} \frac{s(i,k)}{i!} - \sum_{i=k+2}^{\ell} \sum_{j=k+1}^{i} \frac{s(i,j)}{i!} \frac{j!}{(j-k)!} c_i^{j-k}$$

for all $k \in [\![0, \ell-1]\!]$.

Example 2.5.6. For $\ell = 3$, we have $c_3 = 2(\beta-1)$, $c_2 = 2(\beta-1) - (\beta^2-1)/6$, and

$$c_1 = -\frac{c_2}{2} - \frac{c_2^2}{2} - \frac{c_3}{3} - \frac{c_3^2}{2} - \frac{c_3^3}{6} = 2(\beta-1) - \frac{(\beta^2-1)^2}{72} - (\beta^3-1) - \frac{\beta^2-1}{4}.$$

Lemma 2.5.7. *For all $q \in \mathbb{N}$, we have*

$$\beta^\ell \operatorname{val}_{S_\ell}(a_\ell^q) = \sum_{i=1}^{\ell} \binom{\beta q + c_i + i - 1}{i}. \tag{16}$$

Furthermore, if the c_k's are integers satisfying $c_\ell \geq c_{\ell-1} \geq \cdots \geq c_1$, then we have

$$f_{\beta^\ell}(a_\ell^q) = \operatorname{rep}_{S_\ell}(\beta^\ell \operatorname{val}_{S_\ell}(a_\ell^q)) = a_1^{c_\ell - c_{\ell-1}} a_2^{c_{\ell-1} - c_{\ell-2}} \cdots a_{\ell-1}^{c_2 - c_1} a_\ell^{\beta q + c_1} \tag{17}$$

for all integers $q \geq -c_1/\beta$. In this case $f_{\beta^\ell}(a_\ell^*)$ is regular.

PROOF. For the second part of the lemma, observe that, if the c_k's are integers satisfying $c_\ell \geq c_{\ell-1} \geq \cdots \geq c_1$, then (17) is a straightforward consequence of (16) and Lemma 2.3.2. Thus, in this case, the language $f_{\beta^\ell}(a_\ell^*)$ is regular because it can be obtained by a finite number of modifications of the language

$$a_1^{c_\ell - c_{\ell-1}} a_2^{c_{\ell-1}-c_{\ell-2}} \cdots a_{\ell-1}^{c_2-c_1} a_\ell^{c_1} (a_\ell^\beta)^*,$$

which is of course regular. Therefore we only have to show (16). By involving Stirling numbers, we can write

$$\begin{aligned}\beta^\ell \operatorname{val}_{S_\ell}(a_\ell^q) = \beta^\ell \sum_{i=1}^{\ell} \binom{q+i-1}{i} &= \beta^\ell \sum_{i=1}^{\ell}\sum_{k=1}^{i} \frac{s(i,k)}{i!} q^k \\ &= \sum_{k=1}^{\ell} \left(\beta^\ell \sum_{i=k}^{\ell} \frac{s(i,k)}{i!} \right) q^k\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^{\ell} \binom{\beta q + c_i + i - 1}{i} &= \sum_{i=1}^{\ell}\sum_{j=1}^{i} \frac{s(i,j)}{i!} (\beta q + c_i)^j \\ &= \sum_{i=1}^{\ell}\sum_{j=1}^{i} \frac{s(i,j)}{i!} \sum_{k=0}^{j} \binom{j}{k} c_i^{j-k} \beta^k q^k \\ &= \sum_{i=1}^{\ell}\sum_{j=1}^{i} \frac{s(i,j)}{i!} c_i^j + \sum_{k=1}^{\ell} \left(\beta^k \sum_{i=k}^{\ell}\sum_{j=k}^{i} \frac{s(i,j)}{i!} \binom{j}{k} c_i^{j-k} \right) q^k.\end{aligned}$$

Then relation (16) is satisfied if and only if we have

$$\forall k \in [\![1,\ell]\!],\ \beta^{\ell-k} \sum_{i=k}^{\ell} \frac{s(i,k)}{i!} = \sum_{i=k}^{\ell}\sum_{j=k}^{i} \frac{s(i,j)}{i!} \binom{j}{k} c_i^{j-k};$$

$$0 = \sum_{i=1}^{\ell}\sum_{j=1}^{i} \frac{s(i,j)}{i!} c_i^j.$$

Since the last equation is satisfied by the definition of c_1 and since we have

$$\forall k \in [\![1,\ell]\!],\ \beta^{\ell-k} \sum_{i=k}^{\ell} \frac{s(i,k)}{i!} = \sum_{i=k}^{\ell} \frac{s(i,k)}{i!} + \frac{c_{k+1}}{k!} + \sum_{i=k+2}^{\ell}\sum_{j=k+1}^{i} \frac{s(i,j)}{i!} \binom{j}{k} c_i^{j-k}$$

by the definition of c_{k+1}, the lemma is proved. □

Remark 2.5.8. The formula for c_k can be simplified by using

$$\sum_{i=k}^{\ell} \frac{s(i,k)}{i!} = \begin{cases} s(\ell+1,k+1)/\ell!, & \text{if } k \geq 1; \\ 0, & \text{if } k = 0. \end{cases}$$

Note that c_ℓ is the constant c in the proof of Lemma 2.5.3:

$$c_\ell = (\beta - 1)\frac{s(\ell+1,\ell)}{\ell} = \frac{(\beta-1)(\ell+1)}{2}.$$

By using (15), we also obtain

$$\begin{aligned} c_{\ell-1} &= (\beta^2-1)\frac{(3\ell+2)(\ell+1)}{24} - \frac{\ell-1}{2}c_\ell - \frac{1}{2}c_\ell^2 \\ &= c_\ell\left(1 - \frac{\beta+1}{12}\right) = \frac{(\beta-1)(\ell+1)}{2} - \frac{(\beta^2-1)(\ell+1)}{24}. \end{aligned}$$

Now, we turn to our main counting argument that will be used to obtain that S_ℓ-recognizability is not preserved through multiplication by a constant.

Lemma 2.5.9. *Let A be a k-dimensional linear subset of $\mathbb{N}^\ell$ for some k in $[\![0, \ell-1]\!]$ and $B = \Psi^{-1}(A) \cap \mathcal{B}_\ell$ be the corresponding subset of $\mathcal{B}_\ell$. Assume that $\Psi(f_{\beta^\ell}(B))$ contains a sequence $\mathbf{x}^{(n)} = (x_1^{(n)}, \ldots, x_\ell^{(n)})$ such that we have*

$$\min\{x_{j_1}^{(n)}, x_{j_2}^{(n)}, \ldots, x_{j_{k+1}}^{(n)}\} \to +\infty \text{ as } n \to +\infty$$

for some integers $j_1, \ldots, j_{k+1}$ satisfying $1 \le j_1 < j_2 < \cdots < j_{k+1} \le \ell$. Then $f_{\beta^\ell}(B)$ is not regular.

Proof. Since A is a k-dimensional linear subset of $\mathbb{N}^\ell$, we have

$$\mathrm{Card}(B \cap \Sigma_\ell^{\le q}) = \mathrm{Card}\{(x_1, \ldots, x_\ell) \in A \mid \sum_{i=1}^{\ell} x_i \le q\} = \Theta(q^k).$$

Using Corollary 2.5.4, for all sufficiently large integers q, we obtain

$$\mathrm{Card}\left(B \cap \Sigma_\ell^{\le \lfloor (q-c)/\beta \rfloor}\right) \le \mathrm{Card}(f_{\beta^\ell}(B) \cap \Sigma_\ell^{\le q}) \le \mathrm{Card}\left(B \cap \Sigma_\ell^{\le \lceil (q-c)/\beta \rceil + 1}\right)$$

with $c = \lceil (\beta-1)(\ell+1)/2 \rceil$. This implies $\mathrm{Card}(f_{\beta^\ell}(B) \cap \Sigma_\ell^{\le q}) = \Theta(q^k)$. Thus $f_{\beta^\ell}(B)$ is regular if and only if $\Psi(f_{\beta^\ell}(B))$ is a finite union of at most k-dimensional sets as in Lemma 2.4.5. Since the sequence $\mathbf{x}^{(n)}$ cannot occur in such a finite union, $f_{\beta^\ell}(B)$ is not regular. □

The coefficients c_ℓ and $c_{\ell-1}$ (explicitely given in Remark 2.5.8) are rational numbers. In the next two propositions we will discuss the fact that these coefficients could be integers and we rule out all the possible cases.

Proposition 2.5.10. *If we have $\frac{(\beta-1)(\ell+1)}{2} \notin \mathbb{N}$ or $\frac{(\beta^2-1)(\ell+1)}{24} \notin \mathbb{N}$, then $f_{\beta^\ell}(a_\ell^*)$ is not regular.*

Proof. From Lemma 2.3.2 and Corollary 2.3.5 we have

$$\forall q \in \mathbb{N},\ \beta^\ell \operatorname{val}_{S_\ell}(a_\ell^q) = \beta^\ell \sum_{i=1}^{\ell} \binom{q+i-1}{i} = \sum_{i=1}^{\ell} \binom{y_i+i-1}{i}$$

for some integers $y_1, \ldots, y_\ell$ satisfying $y_\ell \geq y_{\ell-1} \geq \cdots \geq y_1 \geq 0$ (depending on q). It follows

$$\beta^\ell \left(\frac{q^\ell}{\ell!} + \frac{(\ell+1)\, q^{\ell-1}}{2\,(\ell-1)!} + O(q^{\ell-2}) \right) = \frac{y_\ell^\ell}{\ell!} + \frac{y_\ell^{\ell-1}}{2\,(\ell-2)!} + \frac{y_{\ell-1}^{\ell-1}}{(\ell-1)!} + O(y_\ell^{\ell-2}). \tag{18}$$

From Lemma 2.5.3 we obtain $y_\ell = |\operatorname{rep}_{S_\ell}(\beta^\ell \operatorname{val}_{S_\ell}(a_\ell^q)| = \beta q + \lceil c_\ell \rceil$ for all sufficiently large integers q. In particular, this implies $y_\ell = \beta q + O(1)$. Hence, by using (18), it follows

$$\frac{\beta(\ell+1)}{2} q^{\ell-1} + O(q^{\ell-2}) = \left((y_\ell - \beta q) + \frac{\ell-1}{2} + \left(\frac{y_{\ell-1}}{\beta q}\right)^{\ell-1} \right) q^{\ell-1};$$

$$y_\ell = \beta q + c_\ell + 1 - \left(\frac{y_{\ell-1}}{\beta q}\right)^{\ell-1} + O\left(\frac{1}{q}\right).$$

First case: $c_\ell = \frac{(\beta-1)(\ell+1)}{2} \notin \mathbb{N}$

We have $y_\ell = \beta q + c_\ell + 1/2$ for all large enough integers q. So we obtain

$$y_{\ell-1} = (1/2)^{\frac{1}{\ell-1}} \beta q + O\left(q^{1-\frac{1}{\ell-1}}\right).$$

Hence we get

$$|f_{\beta^\ell}(a_\ell^q)|_{a_1} = y_\ell - y_{\ell-1} = \left(1 - (1/2)^{\frac{1}{\ell-1}}\right) \beta q + o(q);$$

$$\sum_{j=2}^{\ell} |f_{\beta^\ell}(a_\ell^q)|_{a_j} = y_{\ell-1} = (1/2)^{\frac{1}{\ell-1}} \beta q + o(q).$$

Consequently, $f_{\beta^\ell}(a_\ell^*)$ is not regular from Lemma 2.5.9.

Second case: $c_\ell = \frac{(\beta-1)(\ell+1)}{2} \in \mathbb{N}$

We have $y_\ell = \beta q + c_\ell$ for all large enough integers q. Since we have $y_{\ell-1} \leq y_\ell$, we obtain $y_{\ell-1} = \beta q + O(1)$. By comparing the coefficients of $q^{\ell-2}$ in (18), by using (15) and the definition of $c_{\ell-1}$, we obtain

$$\left(\frac{\beta^2 s(\ell, \ell-2)}{\ell(\ell-1)} + \frac{\beta^2(\ell-2)}{2} + \beta^2 \right) q^{\ell-2} + O(q^{\ell-3})$$

$$= \left(\frac{c_\ell^2}{2} + \frac{c_\ell(\ell-1)}{2} + \frac{s(\ell,\ell-2)}{\ell(\ell-1)} + (y_{\ell-1} - \beta q) + \frac{\ell-2}{2} + \left(\frac{y_{\ell-2}}{\beta q}\right)^{\ell-2} \right) q^{\ell-2};$$

$$y_{\ell-1} = \beta q + c_{\ell-1} + 1 - \left(\frac{y_{\ell-2}}{\beta q}\right)^{\ell-2} + O\left(\frac{1}{q}\right).$$

Since in this case $c_{\ell-1} = \frac{(\beta-1)(\ell+1)}{2} - \frac{(\beta^2-1)(\ell+1)}{24}$ is not an integer and since $y_{\ell-2} \le y_{\ell-1}$ implies $y_{\ell-2} = \beta q + O(1)$, we obtain $y_{\ell-1} = \beta q + \lceil c_{\ell-1} \rceil$ for all large enough integers q and

$$y_{\ell-2} = d^{\frac{1}{\ell-2}} \beta q + O\left(q^{1-\frac{1}{\ell-2}}\right)$$

with $d = 1 - (\lceil c_{\ell-1} \rceil - c_{\ell-1}) \in (0,1)$. Hence we obtain

$$\begin{aligned} |f_{\beta^\ell}(a_\ell^q)|_{a_2} &= y_{\ell-1} - y_{\ell-2} = (1 - d^{\frac{1}{\ell-2}})\beta q + o(q); \\ \sum_{j=3}^{\ell} |f_{\beta^\ell}(a_\ell^q)|_{a_j} &= y_{\ell-2} = d^{\frac{1}{\ell-2}} \beta q + o(q). \end{aligned}$$

Therefore $f_{\beta^\ell}(a_\ell^*)$ is not regular from Lemma 2.5.9. □

Proposition 2.5.11. *If we have* $\frac{(\beta-1)(\ell+1)}{2} \in \mathbb{N}$ *and* $\frac{(\beta^2-1)(\ell+1)}{24} \in \mathbb{N}$*, then* $f_{\beta^\ell}(a_1^* a_\ell^*)$ *is not regular.*

Proof. Using Lemma 2.3.2, if we choose q large enough with respect to p, *e.g.*, $q \ge p^3$, then we get

$$\begin{aligned} \beta^\ell \operatorname{val}_{S_\ell}(a_1^p a_\ell^q) &= \beta^\ell \left(\binom{p+q+\ell-1}{\ell} + \sum_{i=1}^{\ell-1} \binom{q+i-1}{i} \right) \\ &= \binom{\beta(p+q)+c_\ell+\ell-1}{\ell} + \binom{\beta q - (\beta-1)\beta p + c_{\ell-1} + \ell - 2}{\ell-1} \\ &\qquad + \binom{\beta q - \frac{(\beta-1)\beta}{2}(\beta p)^2 + dp + e}{\ell-2} + O\big(q^{\ell-3}\big) \end{aligned}$$

for some constants d and e. Indeed, this equation holds for $p = 0$ by Lemma 2.5.7. Therefore the coefficients of $q^\ell p^0$, $q^{\ell-1}p^0$ and $q^{\ell-2}p^0$ on the left-hand side are equal to those on the right-hand side. It is easy to see that the same holds for $q^{\ell-1}p^1$, $q^{\ell-2}p^2$, $q^{\ell-3}p^3$, and $q^{\ell-4}p^4$. By considering the coefficients of $q^{\ell-2}p^1$ and $q^{\ell-3}p^2$ and multiplying by $(\ell-2)!/\beta^{\ell-1}$ and $(\ell-3)!/\beta^{\ell-1}$ respectively, we obtain the following two equations:

$$\begin{aligned} \beta\frac{\ell-1}{2} &= c_{\ell-1} + \frac{\ell-1}{2} - (\beta-1); \\ \beta\frac{\ell-1}{4} &= \frac{c_{\ell-1}}{2} + \frac{\ell-1}{4} + \frac{(\beta-1)^2}{2} - \frac{(\beta-1)\beta}{2}. \end{aligned}$$

Therefore the same holds for $q^{\ell-2}p^1$ and $q^{\ell-3}p^2$ too. Finally we can choose the constant d so that the coefficient of $q^{\ell-3}p^1$ vanishes as well and the term $O\left(q^{\ell-3}\right)$ remains.

From Corollary 2.3.5 we also have

$$\forall p, q \in \mathbb{N},\ \beta^\ell \operatorname{val}_{S_\ell}(a_1^p a_\ell^q) = \sum_{i=1}^{\ell} \binom{y_i + i - 1}{i}$$

for some integers $y_1, \ldots, y_\ell$ satisfying $y_\ell \geq y_{\ell-1} \geq \cdots \geq y_1 \geq 0$ (depending on p and q). Set

$$P(p, q) = \beta^\ell \operatorname{val}_{S_\ell}(a_1^p a_\ell^q) - \binom{\beta(p+q) + c_\ell + \ell - 1}{\ell}.$$

For $q \geq p^3$, the dominant term in P is $\frac{(\beta q)^{\ell-1}}{(\ell-1)!} > 0$. Therefore, in this case, we obtain $y_\ell \geq \beta(p+q) + c_\ell$. Since, by assumption, we have $c_\ell \in \mathbb{N}$, by using Corollary 2.5.4 we then obtain $y_\ell = |\beta^\ell \operatorname{val}_{S_\ell}(a_1^p a_\ell^q)| = \beta(p+q) + c_\ell$. Next, let us show

$$\binom{\beta q - (\beta - 1)\beta p + c_{\ell-1} + \ell - 2}{\ell - 1} \leq P(p, q) < \binom{\beta q - (\beta - 1)\beta p + c_{\ell-1} + \ell - 1}{\ell - 1}$$

for all sufficiently large p and $q \geq p^3$. Thanks to the foregoing, observe that, for $q \geq p^3$, we have

$$[q^{\ell-2}] \left(P(p, q) - \binom{\beta q - (\beta - 1)\beta p + c_{\ell-1} + \ell - 2}{\ell - 1} \right) = \frac{\beta^{\ell-2}}{(\ell-2)!} > 0;$$

$$[q^{\ell-3} p^2] \left(P(p, q) - \binom{\beta q - (\beta - 1)\beta p + c_{\ell-1} + \ell - 1}{\ell - 1} \right) = -\frac{(\beta - 1)\beta^\ell}{2(\ell-3)!} < 0,$$

which are in both cases the dominant term to consider, *i.e.*, all the possible greater terms vanish. Since we have $c_{\ell-1} \in \mathbb{Z}$ by hypothesis, we obtain $y_{\ell-1} = \beta q - (\beta - 1)\beta p + c_{\ell-1}$ and $y_{\ell-2} = \beta q - \frac{(\beta-1)\beta}{2}(\beta p)^2 + dp + O(1)$ for all large enough p and $q \geq p^3$. In the latter conditions, we have

$$|f_{\beta^\ell}(a_1^p a_\ell^q)|_{a_1} = \beta^2 p + O(1);$$

$$|f_{\beta^\ell}(a_1^p a_\ell^q)|_{a_2} = \frac{(\beta - 1)\beta^3}{2} p^2 + O(p);$$

$$\sum_{j=3}^{\ell} |f_{\beta^\ell}(a_1^p a_\ell^q)|_{a_j} = \beta q + O(p^2).$$

Hence $f_{\beta^\ell}(a_1^* a_\ell^*)$ is not regular from Lemma 2.5.9. □

Example 2.5.12. We illustrate some of the above computations. Let us continue Example 2.5.6. If we have $\beta \equiv \pm 1 \pmod 6$, then c_3, c_2, and c_1 are all integers. Hence this gives

$$f_{\beta^3}(a_3^q) = a_1^{c_3 - c_2} a_2^{c_2 - c_1} a_3^{\beta q + c_1}$$

for all sufficiently large integers q. In particular, the latter formula shows that a_3^* cannot be used to prove that multiplication by β^3 does not preserve recognizability when we have $\beta \equiv \pm 1 \pmod 6$. Thanks to Proposition 2.5.10, $f_{\beta^3}(a_3^*)$ is regular if and only if we have $\beta \equiv \pm 1 \pmod 6$. On the other hand, for large enough p and $q \geq p^3$, we find

$$f_{\beta^3}(a_1^p a_3^q) = a_1^{\beta^2 p + c_3 - c_2} a_2^{\frac{(\beta-1)\beta}{2}(\beta p)^2 - (\beta - 1 + d)\beta p + c_2 - c_1} a_3^{\beta q - \frac{(\beta-1)\beta}{2}(\beta p)^2 + d\beta p + c_1},$$

with $d = -(\beta - 1)(\beta^2 - 2\beta + 6)/6$. This shows that $f_{\beta^3}(a_1^* a_3^*)$ is not regular.

Otherwise, *i.e.*, if we have $1 - \beta^2 \equiv j \pmod 6$ with $j \in \{1, 3, 4\}$ and $k = 1 - \frac{j}{6} > 0$, then, for all large enough integers q, we have

$$f_{\beta^3}(a_3^q) = a_1^{c_3 - \lceil c_2 \rceil} a_2^{k\beta q + \lceil c_2 \rceil - c_1 + kc_2 + \frac{k(k+1)}{2}} a_3^{\frac{j}{6}\beta q + c_1 - kc_2 - \frac{k(k+1)}{2}}.$$

2.6. Main Result

By collecting results from Theorems 2.2.1 and 2.2.10, Corollary 2.2.7, and Propositions 2.5.10 and 2.5.11, we obtain our main result about multiplication by a constant.

Theorem 2.6.1. *Let ℓ, λ be positive integers with $\lambda \geq 2$. For the abstract numeration system*

$$S = (a_1^* \cdots a_\ell^*, \{a_1, \ldots, a_\ell\}, a_1 < \cdots < a_\ell),$$

multiplication by λ preserves S-recognizability if and only if one of the following conditions is satisfied:

- *$\ell = 1$;*
- *$\ell = 2$ and λ is an odd square.*

Proof. The case $\ell = 1$ is ruled out by Corollary 2.2.7 and the case $\ell = 2$ is given by Theorem 2.2.10. Consider $\ell \geq 3$. Thanks to Theorem 2.2.1, it is only necessary to consider λ of the form β^ℓ. Then the result can be deduced from Propositions 2.5.10 and 2.5.11. □

2.7. Structural Properties of $\mathcal{B}_\ell$ Seen Through f_{β^ℓ}

In this independent section we closely inspect how a word is transformed by applying f_{β^ℓ}. To that end, $\mathcal{B}_\ell$ (or equivalently $\mathbb{N}$) is partitioned into regions where f_{β^ℓ} acts differently. Thanks to our discussion, we are able to detect some kind of patterns periodically occurring within these regions. To have a flavor of the computations involved in this section, the reader could first have a look at Example 2.7.5. Let ℓ and β be two positive integers with $\ell \geq 2$ that will be fixed throughout this section.

According to Corollary 2.5.4, we define a partition of $\mathbb{N}$.

Definition 2.7.1. For all $i \in [\![0, \beta]\!]$ and large enough $q \in \mathbb{N}$, we define

$$\mathcal{R}_{i,q} = \{n \in \mathbb{N} \mid |\operatorname{rep}_{S_\ell}(n)| = q,\, |\operatorname{rep}_{S_\ell}(\beta^\ell n)| = \beta q + \left\lceil \frac{(\beta-1)(\ell+1)}{2} \right\rceil - i\}$$

and we let $m_{i,q} = \min \mathcal{R}_{i,q}$ denote the smallest integer in $\mathcal{R}_{i,q}$.

Lemma 2.7.2. *If we have $\beta = \prod_{i=1}^{k} p_i^{u_i}$ where $k, u_1, \ldots, u_k$ are positive integers and $p_1, \ldots, p_k$ are prime numbers greater than ℓ, then for any integer $x \geq \ell$, we have*

$$\binom{x}{\ell} \equiv \binom{x+\beta^\ell}{\ell} \pmod{\beta^\ell}.$$

Proof. Let x, y be integers satisfying $x, y \geq \ell$. We have

$$\binom{y}{\ell} - \binom{x}{\ell} = \frac{y(y-1)\cdots(y-\ell+1) - x(x-1)\cdots(x-\ell+1)}{\ell!}.$$

The numerator on the right-hand side is an integer divisible by $\ell!$. Since it can be written as $P(y) - P(x)$ for some polynomial P, it is also divisible by $y - x$. Then, for $y = x + \beta^\ell$, the corresponding numerator is divisible by $\ell!$ and also by β^ℓ. But since any prime factor of β is larger than ℓ, $\ell!$ and β^ℓ are relatively prime. Consequently, the corresponding numerator is divisible by $\beta^\ell \ell!$ and the lemma is proved. □

An inspection of multiplication by β^ℓ using the partition induced by Corollary 2.5.4 provides us with the following observation.

Proposition 2.7.3. *For all large enough integers q, we have*

$$|\operatorname{rep}_{S_\ell}(\beta^\ell m_{\beta,q+\beta^{\ell-1}})| = |\operatorname{rep}_{S_\ell}(\beta^\ell m_{\beta,q})| + \beta^\ell$$

and $m_{\beta,q} = \operatorname{val}_S(a_1^q)$. If β satisfies the condition of Lemma 2.7.2, then, for all $i \in [\![0, \beta-1]\!]$ and all large enough integers q, we have

$$\forall j \in \{2, \ldots, \ell\},\ |\operatorname{rep}_{S_\ell}(\beta^\ell m_{i,q+\beta^{\ell-1}})|_{a_j} = |\operatorname{rep}_{S_\ell}(\beta^\ell m_{i,q})|_{a_j},$$
$$|\operatorname{rep}_{S_\ell}(\beta^\ell m_{i,q+\beta^{\ell-1}})|_{a_1} = |\operatorname{rep}_{S_\ell}(\beta^\ell m_{i,q})|_{a_1} + \beta^\ell,$$

and $m_{i,q} = \left\lceil \frac{C_i(q)}{\beta^\ell} \right\rceil$ with

$$C_i(q) = \operatorname{val}_{S_\ell}\left(a_1^{\beta q + \frac{(\beta-1)(\ell+1)}{2} - i}\right) = \binom{\beta q + \frac{(\beta-1)(\ell+1)}{2} - i + \ell - 1}{\ell}.$$

Proof. For $i = \beta$, the first part of the lemma is a straightforward consequence of Lemma 2.5.3 (which also holds for $\ell = 2$). Now, assume that β satisfies the condition of Lemma 2.7.2 and choose some $i \in [\![0, \beta-1]\!]$. From

Lemma 2.5.3 it follows $\mathrm{val}_{S_\ell}(a_1^q) \in \mathcal{R}_{\beta,q}$ and $\mathrm{val}_{S_\ell}(a_\ell^q) \in \mathcal{R}_{0,q}$ for all sufficiently large integers q. Hence we can choose an integer q large enough so that the regions $\mathcal{R}_{j,r}$ are non-empty for all $j \in [\![0,\beta]\!]$ and all integers $r \geq q$. Note that $(\beta-1)(\ell+1)$ is even since β satisfies the condition of Lemma 2.7.2. By definition we have $C_i(q) \leq \beta^\ell m_{i,q}$. Since we have $m_{i,q}-1 \in \mathcal{R}_{i+1,q}$, we also obtain $\beta^\ell m_{i,q} < C_i(q)+\beta^\ell$. Therefore we find $m_{i,q} = \lceil C_i(q)/\beta^\ell \rceil$. Moreover, there exists a unique integer $\mu_i(q)$ that satisfies

$$\beta^\ell m_{i,q} = C_i(q) + \mu_i(q) \quad \text{and} \quad 0 \leq \mu_i(q) < \beta^\ell.$$

In particular, we have

$$\beta^\ell m_{i,q+\beta^{\ell-1}} = C_i(q+\beta^{\ell-1}) + \mu_i(q+\beta^{\ell-1}) \quad \text{and} \quad 0 \leq \mu_i(q+\beta^{\ell-1}) < \beta^\ell.$$

From Lemma 2.7.2 it follows $C_i(q) \equiv C_i(q+\beta^{\ell-1}) \pmod{\beta^\ell}$. Consequently we obtain $\mu_i(q) = \mu_i(q+\beta^{\ell-1})$. If we have $\mathrm{rep}_{S_{\ell-1}}(\mu_i(q)) = a_1^{n_1}\cdots a_{\ell-1}^{n_{\ell-1}}$, then, from Lemma 2.3.2, we deduce

$$\mathrm{rep}_{S_\ell}(\beta^\ell m_{i,q}) = a_1^t a_2^{n_1} \cdots a_\ell^{n_{\ell-1}}$$

and

$$\mathrm{rep}_{S_\ell}(\beta^\ell m_{i,q+\beta^{\ell-1}}) = a_1^{t+\beta^\ell} a_2^{n_1} \cdots a_\ell^{n_{\ell-1}},$$

where t is the integer defined by $|\,\mathrm{rep}_{S_\ell}(\beta^\ell m_{i,q})| = \beta\, q + \frac{(\beta-1)(\ell+1)}{2} - i$. The lemma is proved now. □

Remark 2.7.4. In the previous proposition we were interested in the first word in $\mathcal{R}_{i,q}$. Actually, it is even possible to describe how multiplication by β^ℓ affects representations inside $\mathcal{R}_{i,q}$. With the notation of the previous proof, for any $n \in \mathcal{R}_{i,q}$ (and q large enough), we have

$$\mathrm{rep}_{S_\ell}(\beta^\ell n) = a_1^t a_2^{n_1} \cdots a_\ell^{n_{\ell-1}}$$

where t and $n_1, \ldots, n_{\ell-1}$ are the integers defined by

$$|\,\mathrm{rep}_{S_\ell}(\beta^\ell n)| = \beta q + \frac{(\beta-1)(\ell+1)}{2} - i$$

and

$$\mathrm{rep}_{S_{\ell-1}}\Big(\mu_i(q) + \beta^\ell(n-m_{i,q})\Big) = a_1^{n_1}\cdots a_{\ell-1}^{n_{\ell-1}}$$

respectively.

Example 2.7.5. Take $\ell = 3$ and $\beta = 5$. We thus have $c_\ell = 8$. The number 171717 (resp. 172739) is the first element belonging to $\mathcal{R}_{4,100}$ (resp. $\mathcal{R}_{3,100}$). We have

$$\mathrm{rep}_{S_3}(171717) = a^{95}b^3c^2 \quad \text{and} \quad \mathrm{rep}_{S_3}(5^3 \cdot 171717) = a^{490}\underline{b^{14}c^0};$$
$$\mathrm{rep}_{S_3}(172739) = a^{55}b^{41}c^4 \quad \text{and} \quad \mathrm{rep}_{S_3}(5^3 \cdot 172739) = a^{493}\underline{b^0c^{12}}.$$

Therefore, with the notation of the previous proof, we obtain $\mu_4(100) = \mathrm{val}_{S_2}(\phi(b^{14})) = 105$ (resp. $\mu_3(100) = \mathrm{val}_{S_2}(\phi(c^{12})) = 90$), where ϕ is the morphism on $\{a, b, c\}^*$ defined by $\phi(a) = \varepsilon$, $\phi(b) = a$, and $\phi(c) = b$. The number 333396 (resp. 334986) is the smallest element in $\mathcal{R}_{4,125}$ (resp. $\mathcal{R}_{3,125}$). The corresponding S_3-representations are given by

$$\mathrm{rep}_{S_3}(333396) = a^{119}b^6c^0 \quad \text{and} \quad \mathrm{rep}_{S_3}(5^3 \cdot 333396) = a^{615}\underline{b^{14}c^0};$$

$$\mathrm{rep}_{S_3}(334986) = a^{69}b^{41}c^{15} \quad \text{and} \quad \mathrm{rep}_{S_3}(5^3 \cdot 334986) = a^{618}\underline{b^0c^{12}}.$$

We have $\mathrm{Card}(\mathcal{R}_{4,100}) = 1022$ and $\mathrm{Card}(\mathcal{R}_{4,125}) = 1590$. We can then build the following table.

j	$\Psi(\mathrm{rep}_{S_3}(5^3(m_{4,100}+j)))$	$\Psi(\mathrm{rep}_{S_3}(5^3(m_{4,125}+j)))$
0	$(490, 14, 0)$	$(615, 14, 0)$
1	$(484, 0, 20)$	$(609, 0, 20)$
2	$(478, 22, 4)$	$(603, 22, 4)$
$\vdots$	$\vdots$	$\vdots$
1021	$(0, 34, 470)$	$(125, 34, 470)$
1022	$\times$	$(124, 415, 90)$
$\vdots$	$\vdots$	$\vdots$
1589	$\times$	$(0, 34, 595)$

CHAPTER 3

A Decidability Problem

3.1. Introduction

In this chapter we mainly address the following decidability question and its extension to abstract numeration systems.

Problem 2. Let

- U be a linear numeration system such that $\mathbb{N}$ is U-recognizable;
- Let X be any U-recognizable set of non-negative integers, which is given through a DFA accepting $\mathrm{rep}_U(X)$.

Is it decidable whether or not X is ultimately periodic, *i.e.*, whether or not X is a finite union of arithmetic progressions?

Note that the regularity of $\mathrm{rep}_U(\mathbb{N})$ in the previous statement ensures that there exists a set $X \subseteq \mathbb{N}$ such that $\mathrm{rep}_U(X)$ is regular; see Lemma 1.6.13 on page 15 and Remark 3.2.16 below.

The material of this chapter was first introduced in [**CR08**] and then was developed in [**BCFR09**].

Ultimately periodic sets of integers play a special role. On the one hand, such infinite sets are coded by a finite amount of information. On the other hand, the famous theorem of Cobham stated below asserts that these sets are the only ones that are recognizable in all integer base numeration systems. This is the reason why they are also referred to in the literature as *recognizable* sets of integers (the recognizability being in that case independent of the base). Furthermore, Cobham's theorem has been extended to various situations and, in particular, to numeration systems given by substitutions [**Dur02b**].

Definition 3.1.1. Two integers $p, q \geq 2$ are said *multiplicatively independent* if, for all positive integers m and n, we have $p^m \neq q^n$.

Theorem 3.1.2. [**Cob69**] *Let $p, q \geq 2$ be multiplicatively independent integers. A set X of non-negative integers is both U_p-recognizable and U_q-recognizable if and only if it is ultimately periodic.*

If we restrict ourselves to usual integer base numeration systems, then several results are known. J. Honkala showed in [**Hon86**] that Problem 2 turns out to be decidable in this case. Let us also mention [**Ale04**], where the number of states of the minimal automaton accepting numbers written in a given integer base $b \geq 2$ and divisible by a given positive integer is explicitly computed. J.-P. Allouche and J. Shallit asked in [**AS03**] if one can obtain a polynomial time decision procedure for integer base numeration systems. Using the logic formalism of the Presburger arithmetic, a positive answer to this question is given by J. Leroux in [**Ler05**] even when considering subsets of $\mathbb{Z}^d$, where d is a positive integer. In dimension one ultimately periodic sets are exactly the sets definable in the Presburger arithmetic $\langle \mathbb{N}, + \rangle$.

A. Muchnik showed that Problem 2 turns out to be decidable for any linear numeration system U for which both $\text{rep}_U(\mathbb{N})$ and addition are recognizable by finite automata [**Muc03**]. Still, it is a difficult question to characterize numeration systems U for which addition is computable by a finite automaton.[1] For details in this area, as already mentioned on page 24, see [**BH97, Fro92**], in which the authors mainly considered positional numeration systems defined by a linear recurrence relation whose characteristic polynomial is the minimal polynomial of a Pisot number. In [**Fro97**] the sequentiality of the successor function, *i.e.*, the action of adding 1, is studied. If addition is computable by a finite automaton, so the successor function is, but the converse does not hold in general. In particular, some examples of linear numeration systems for which addition is not computable by a finite automaton are given in [**Fro97**]: for example, the sequence $(U_i)_{i\geq 0}$ defined by $U_{i+4} = 3U_{i+3} + 2U_{i+2} + 3U_i$ for all $i \in \mathbb{N}$ with any integer initial conditions satisfying $1 = U_0 < U_1 < U_2 < U_3$. So the decision techniques from [**Ler05, Muc03**] cannot be applied to that numeration system. Nevertheless, as we shall see in Example 3.5.14, our decision procedure can be applied to this system. Also, note that, in the extended framework of abstract numeration systems, one can exhibit systems such that multiplication by a constant does not preserve S-recognizability. For a discussion on this topic, see Theorems 2.2.1 and 2.2.10, and Propositions 2.5.10 and 2.5.11 in Chapter 2, which led to the proof of Theorem 2.6.1 on page 43. Therefore the powerful tools from logic discussed above cannot be applied in that context either.

The question studied in this chapter was raised by J. Sakarovitch during the "*Journées de Numération*" in Graz, May 2007. The question was initially asked for a larger class of numeration systems than the one treated here, namely for any abstract numeration systems built on an infinite regular language.

[1] See Chapter 2 on page 23 for the definition of "computable by a finite automaton".

This decision problem for all abstract numeration systems is equivalent to the famous (and unsolved) *HD0L periodicity problem*: given a morphism f and a coding g such that f is prolongable[2] on a letter a, decide whether or not the infinite word $g(f^\omega(a))$ is ultimately periodic; see [**HR04, RM02**]. For the restricted case of the *D0L periodicity problem*, where only the morphism f is considered, decision procedures are well-known [**HL86, Pan86**].

Finally, questions related to those addressed here have independently and recently gained interest [**ARS09**]. In particular, a simple proof of J. Honkala's original result based on the construction of some non-deterministic automata is given in that paper. As for the logical approach considered by A. Muchnik and J. Leroux, the arguments given in [**ARS09**] rely on the recognizability of addition by a finite automaton (which can be done for the standard integer base numeration systems but not necessarily for an arbitrary linear numeration system).

This chapter follows the organization described hereafter. The structure of Section 3.2 is the same as the one of [**Hon86**]. First, we give an upper bound on the admissible periods of a U-recognizable set X of non-negative integers when it is assumed to be ultimately periodic. Then, an upper bound on the admissible preperiods is obtained. These bounds depend essentially on the number of states of the minimal automaton recognizing $\mathrm{rep}_U(X)$. Finally, a finite number of such periods and preperiods have to be checked. For each of them, we have to build automata accepting the corresponding ultimately periodic sets. In particular, this implies that $\mathbb{N}$ has to be recognizable; see Lemma 1.6.13 on page 15. Though the structure is the same, our arguments and techniques are quite different from [**Hon86**]. They rely on the study of the quantity $N_U(m)$, which is defined as the number of residue classes that appear infinitely often in the reduced sequence $(U_i \bmod m)_{i\geq 0}$, where m is a positive integer. Our main result — Theorem 3.2.15 — in this section can be stated as follows:

Theorem. *Let $U = (U_i)_{i\geq 0}$ be a linear numeration system such that $\mathbb{N}$ is U-recognizable and satisfying $\lim_{i\to+\infty} U_{i+1} - U_i = +\infty$. If we have $\lim_{m\to+\infty} N_U(m) = +\infty$, then Problem 2 is decidable.*

Actually our techniques cannot be applied to integer base numeration systems, which is the case treated by J. Honkala [**Hon86**], because in that case we have $N_U(m) \not\to +\infty$ as $m \to +\infty$; see Remark 3.2.19.

[2] This notion will be formally defined on page 87 in Chapter 4.

In Section 3.3 and Section 3.4 we recall some background about p-adic numbers and finitely generated Abelian groups respectively. These notions will be used in Section 3.5.

Then, in Section 3.5, we give a characterization of the linear numeration systems $U = (U_i)_{i\geq 0}$ which satisfy $\lim_{m\to+\infty} N_U(m) = +\infty$. To do so we use p-adic methods leading to a study of the sequences $(U_i \bmod p^v)_{i\geq 0}$ for all positive integers v and well-chosen prime numbers p.

In Section 3.6 we consider again the same decision problem but restated in the framework of abstract numeration systems [**LR01**]. We successfully apply the same kind of techniques to a large class of abstract numeration systems. For instance, an example consisting of two copies of the Fibonacci system is considered. The corresponding decision procedure is given by Theorem 3.6.4.

In the last section we use results from [**HR04**] to show that Theorem 3.6.4 provides a decision procedure for particular instances of the HD0L periodicity problem.

All along this chapter, whenever it is possible, we try to state results in their most general form, even if later on we have to restrict ourselves to particular cases. For instance, results about the admissible preperiods do not require any particular assumption on the numeration system except linearity.

3.2. A Decision Procedure for a Class of Linear Numeration Systems

We will often consider positional numeration systems $U = (U_i)_{i\geq 0}$ satisfying the following condition:

$$\lim_{i\to+\infty} U_{i+1} - U_i = +\infty. \tag{19}$$

Note that it is a weak requirement. Usually, the sequence U has an exponential growth, that is, $U_i \simeq \beta^i$ for some $\beta > 1$, and therefore, condition (19) is trivially satisfied. For example, it is the case for the numeration systems that will be considered in Remark 3.2.21 on page 60 or the Fibonacci numeration system, which was introduced in Remark 1.6.3 on page 12.

The following lemma ensures that if a word w is a greedy U-representation, then the words $10^r w$ are also greedy U-representations for all large enough integers r.

Lemma 3.2.1. *Let $U = (U_i)_{i\geq 0}$ be a positional numeration system satisfying condition* (19). *Then, for all $j \in \mathbb{N}$, there exists $L > 0$ such that, for all integers $\ell \geq L$, the words*

$$10^{\ell-|\operatorname{rep}_U(t)|}\operatorname{rep}_U(t),\ t = 0,\ldots,U_j - 1$$

are greedy U-representations.

PROOF. Choose $j \in \mathbb{N}$. Note that we have $\mathrm{rep}_U(U_j) = 10^j$. Hence $\mathrm{rep}_U(U_j - 1)$ is the greatest word of length j in $\mathrm{rep}_U(\mathbb{N})$ with respect to the genealogical order. By hypothesis, there exists $L > 0$ such that, for all integers $\ell \geq L$, we have $U_{\ell+1} - U_\ell > U_j - 1$. Therefore, for all integers $\ell \geq L$, the word $10^{\ell-j}\,\mathrm{rep}_U(U_j - 1)$ is the greedy U-representation of the number $U_\ell + U_j - 1 < U_{\ell+1}$. The result easily follows. □

Example 3.2.2. Consider the positional numeration system $U = (U_i)_{i\geq 0}$ defined by $U_0 = 1$, $U_1 = 2$, $U_2 = 3$, and $U_{3i+r} = 3^{i+1} + r$ for all $i \in \mathbb{N} \setminus \{0\}$ and all $r \in \{0, 1, 2\}$. This system does not satisfy condition (19) because we have $U_{i+1} - U_i = 1$ for infinitely many integers i. We have $\mathrm{rep}_U(2) = 10$, but, for all $i \in \mathbb{N}$, the word $10^{3i+1}10$ is not a greedy U-representation. Indeed, for all $i \in \mathbb{N}$, the number $\mathrm{val}_U(10^{3i+1}10) = U_{3(i+1)} + 2 = U_{3i+5}$ has 10^{3i+5} as greedy U-representation.

Remark 3.2.3. In the lemma above one cannot exchange the order of the quantifiers about j and L. For example, consider the positional numeration system $U = (U_i)_{i\geq 0}$ defined by $U_i = (i + 1)(i + 2)/2$ for all $i \in \mathbb{N}$. This sequence satisfies condition (19). Moreover, it is a linear numeration system since it satisfies the linear recurrence relation $U_{i+3} = 3U_{i+2} - 3U_{i+1} + U_i$ for all $i \in \mathbb{N}$. Observe that, for all non-negative integers i, we have

$$\begin{aligned}
\mathrm{val}_U(10^n 10^i) \geq U_{n+i+2} &\Leftrightarrow U_{n+i+1} + U_i \geq U_{n+i+2} \\
&\Leftrightarrow U_{n+i+2} - U_{n+i+1} \leq U_i \\
&\Leftrightarrow n + i + 3 \leq U_i \\
&\Leftrightarrow n < U_i - i - 2.
\end{aligned}$$

It follows that the greedy U-representations of the form $10^n 10^i$ are exactly those for which we have $n \geq U_i - i - 2$, which grows with i.

Remark 3.2.4. Numeration systems associated with real numbers $\beta > 1$ are defined as follows. In this case, one usually refers to *β-numeration systems.* Set $\Delta_\beta = [\![0, \lceil\beta\rceil - 1]\!]$. Any $x \in [0, 1]$ can be written as

$$x = \sum_{i=1}^{+\infty} c_i\, \beta^{-i}, \text{ with } c_i \in \Delta_\beta\ \forall i \in \mathbb{N} \setminus \{0\}.$$

The sequence $(c_i)_{i\geq 1}$ is said to be a *β-representation* of x. For all $x \in [0, 1]$, we let $d_\beta(x)$ denote the maximal β-representation of x with respect to the lexicographical order, which is called the *β-development* of x. It is obtained by the greedy algorithm. More details can be found, for instance, in [**Lot02**, Chap. 8]. We say that a β-development $(c_i)_{i\geq 1}$ is *finite* if there exists an integer N such that we have $c_i = 0$ for all integers $i \geq N$. If there exists a positive integer m such that we have $d_\beta(1) = t_1 \cdots t_m$ with $t_m \neq 0$, then we set $d^*_\beta(1) = (t_1 \cdots t_{m-1}(t_m - 1))^\omega$. Otherwise $d_\beta(1)$ is infinite and we set $d^*_\beta(1) = d_\beta(1)$.[3]

Now, we are able to define a positional numeration system $U_\beta = (U_i)_{i\geq 0}$ associated with β; see [**BM89**]. If we have $d^*_\beta(1) = (t_i)_{i\geq 1}$, then we define

$$U_0 = 1 \text{ and } \forall i \in \mathbb{N} \setminus \{0\},\ U_i = t_1 U_{i-1} + \cdots + t_i U_0 + 1. \qquad (20)$$

Such systems are called *Bertrand numeration systems*. If β is a Parry number, *i.e.*, if $d_\beta(1)$ is finite or ultimately periodic, then one can derive from (20) that the sequence U_β satisfies a linear recurrence relation. Then, as a consequence of a theorem of A. Bertrand [**BM89**] linking together greedy U_β-representations and finite factors occurring in β-developments, the language $\text{rep}_{U_\beta}(\mathbb{N})$ of the greedy U_β-representations is regular. The trim minimal automaton accepting these representations is well-known [**FS96**] and has a special form: all states are final and from all these states, an edge of label 0 goes back to the initial state. Therefore we have the following property which is much stronger than the previous lemma. If x and y are greedy U_β-representations, then $x0y$ is also a greedy U_β-representation.

Example 3.2.5. The Fibonacci system is the Bertrand system associated with the golden ratio $(1 + \sqrt{5})/2$. Since greedy representations in the Fibonacci system are the words not containing two consecutive occurrences of 1 [**Zec72**], we have $x0y \in \text{rep}_F(\mathbb{N})$ for all $x, y \in \text{rep}_F(\mathbb{N})$.

Recall (see Definition 1.6.12 on page 15) that a set of integers X is said to be ultimately periodic if there exist $a, p \in \mathbb{N}$ with $p > 0$, such that, for all $i \geq a$, we have $i \in X \Leftrightarrow i + p \in X$. Moreover, if the integers a and p are minimal for this property, then we say that they are the preperiod and the period of X respectively. The minimality of the period chosen to represent an ultimately periodic set is used in the next lemma.

Lemma 3.2.6. *Let $X \subseteq \mathbb{N}$ be an ultimately periodic set of period p_X and preperiod a_X and let i, j be integers satisfying $i, j \geq a_X$ and $i \not\equiv j \pmod{p_X}$. Then there exists $t \in [\![0, p_X - 1]\!]$ such that we have either $i + t \in X$ and $j + t \notin X$, or $i + t \notin X$ and $j + t \in X$.*

[3] Note that, when β is an integer, we have $d^*_\beta(1) = (\beta - 1)^\omega$. In that case, the β-representations of real numbers ending with $d^*_\beta(1)$ are the "improper" representations.

Proof. Without loss of generality we may assume $i < j$ and $p_X > 1$. Proceed by contradiction and assume that, for all $t \in [\![0, p_X - 1]\!]$, we have $i + t \in X \Leftrightarrow j + t \in X$. Let $p \in [\![1, p_X - 1]\!]$ be defined by $p \equiv j - i \pmod{p_X}$ and let n be an integer satisfying $n \geq i$. We can write $n \equiv i + r \pmod{p_X}$ with $r \in [\![0, p_X - 1]\!]$. Then we have $n + p \equiv j + r \pmod{p_X}$. Therefore we obtain $n + p \in X \Leftrightarrow j + r \in X \Leftrightarrow i + r \in X \Leftrightarrow n \in X$. This leads to a contradiction since p_X is minimal for this property. □

Definition 3.2.7. For a sequence $(U_i)_{i \geq 0}$ of integers and a positive integer m, we let $N_U(m) \in [\![1, m]\!]$ denote the number of values that are taken infinitely often by the reduced sequence $(U_i \bmod m)_{i \geq 0}$.

Example 3.2.8. If $F = (F_i)_{i \geq 0}$ is the Fibonacci sequence, then we have, for instance,

$$(F_i \bmod 4) = (1, 2, 3, 1, 0, 1, 1, 2, 3, \ldots);$$
$$(F_i \bmod 11) = (1, 2, 3, 5, 8, 2, 10, 1, 0, 1, 1, 2, 3, \ldots).$$

This shows that we have $N_F(4) = 4$ and $N_F(11) = 7$.

A number of papers are specifically intended to the study of the distribution of the Fibonacci sequence reduced modulo m; for instance, see [**Wal60, Mam61, Sha68, Bru70, KS72, Yal73, And74, Cat74, Tur74, DL77, Vin78, Ehr89, Dar01, SC05**]. More generally, for the study of linear recurrence sequences reduced modulo m, see, for instance, [**Eng31, War33, Rau64, Rob66, Vin81, Nar84, Pin93, Wad96, Her04, WY06**].

Proposition 3.2.9. *Let $U = (U_i)_{i \geq 0}$ be a positional numeration system satisfying condition* (19) *and let $X \subseteq \mathbb{N}$ be an ultimately periodic U-recognizable set of period p_X. Then any DFA accepting* $\mathrm{rep}_U(X)$ *has at least $N_U(p_X)$ states.*

Proof. Let a_X be the preperiod of X. From Lemma 3.2.1 there exists an integer $L > 0$ such that for any integer $h \geq L$, the words

$$10^{h - |\mathrm{rep}_U(t)|}\,\mathrm{rep}_U(t),\ t = 0, \ldots, p_X - 1$$

are greedy U-representations. The sequence $(U_i \bmod p_X)_{i \geq 0}$ takes infinitely often $N := N_U(p_X)$ different values. Let $h_1, \ldots, h_N \geq L$ be integers satisfying

$$\forall i, j \in [\![1, N]\!],\ U_{h_i} \geq a_X \text{ and } (i \neq j \Rightarrow U_{h_i} \not\equiv U_{h_j} \pmod{p_X}))$$

From Lemma 3.2.6, for all distinct $i, j \in [\![1, N]\!]$, there exists $t_{i,j} \in [\![0, p_X - 1]\!]$ such that we have either $U_{h_i} + t_{i,j} \in X$ and $U_{h_j} + t_{i,j} \notin X$, or $U_{h_i} + t_{i,j} \notin X$

and $U_{h_j} + t_{i,j} \in X$. Therefore

$$w_{i,j} = 0^{|\operatorname{rep}_U(p_X-1)|-|\operatorname{rep}_U(t_{i,j})|}\operatorname{rep}_U(t_{i,j})$$

is a word such that we have either

$$10^{h_i-|\operatorname{rep}_U(p_X-1)|}w_{i,j} \in \operatorname{rep}_U(X) \quad \text{and} \quad 10^{h_j-|\operatorname{rep}_U(p_X-1)|}w_{i,j} \notin \operatorname{rep}_U(X),$$

or

$$10^{h_i-|\operatorname{rep}_U(p_X-1)|}w_{i,j} \notin \operatorname{rep}_U(X) \quad \text{and} \quad 10^{h_j-|\operatorname{rep}_U(p_X-1)|}w_{i,j} \in \operatorname{rep}_U(X).$$

Therefore the N words $10^{h_1-|\operatorname{rep}_U(p_X-1)|}, \ldots, 10^{h_N-|\operatorname{rep}_U(p_X-1)|}$ are pairwise non-equivalent for the Myhill-Nerode equivalence relation $\sim_{\operatorname{rep}_U(X)}$. Then, from Definition 1.3.6 on page 6, the minimal automaton of $\operatorname{rep}_U(X)$ has at least $N = N_U(p_X)$ states. □

The previous proposition has an immediate consequence for getting a bound on the period of an eventually periodic set accepted by a given DFA.

Corollary 3.2.10. *Let $U = (U_i)_{i\geq 0}$ be a positional numeration system satisfying condition* (19)*. Assume*

$$\lim_{m\to+\infty} N_U(m) = +\infty.$$

Then the period of an ultimately periodic set $X \subseteq \mathbb{N}$ such that $\operatorname{rep}_U(X)$ is accepted by a DFA with d states is bounded by the smallest non-negative integer s such that, for all integers $m \geq s$, we have $N_U(m) > d$.

A result similar to the previous corollary (in the sense that it makes possible to give an upper bound on the period) can be stated as follows.

Proposition 3.2.11. *Let $U = (U_i)_{i\geq 0}$ be a positional numeration system satisfying condition* (19)*, let X be an ultimately periodic U-recognizable set of non-negative integers of period p_X, and let c be a divisor of p_X. If* 1 *occurs infinitely many times in $(U_i \bmod c)_{i\geq 0}$, then any DFA accepting $\operatorname{rep}_U(X)$ has at least c states.*

Proof. Let a_X denote the preperiod of X. By applying several times Lemma 3.2.1, there exist non-negative integers $n_1, \ldots, n_c$ such that

$$10^{n_c}10^{n_{c-1}}\cdots 10^{n_1}0^{|\operatorname{rep}_U(p_X-1)|-|\operatorname{rep}_U(t)|}\operatorname{rep}_U(t),\ t = 0, \ldots, p_X - 1$$

are greedy U-representations. Moreover, since 1 occurs infinitely many times in the sequence $(U_i \bmod c)_{i\geq 0}$, the integers $n_1, \ldots, n_c$ can be chosen such that we have

$$\forall j \in [\![1, c]\!],\ \mathrm{val}_U(10^{n_j} \cdots 10^{n_1 + |\operatorname{rep}_U(p_X - 1)|}) \equiv j \pmod{c};$$
$$\mathrm{val}_U(10^{n_1 + |\operatorname{rep}_U(p_X - 1)|}) \geq a_X.$$

For all distinct $i, j \in [\![1, c]\!]$, from Lemma 3.2.6 and since c divides p_X, the words

$$10^{n_i} \cdots 10^{n_1} \text{ and } 10^{n_j} \cdots 10^{n_1}$$

are non-equivalent for the relation $\sim_{\operatorname{rep}_U(X)}$. We show this by concatenating some word of the kind $0^{|\operatorname{rep}_U(p_X-1)| - |\operatorname{rep}_U(t)|} \operatorname{rep}_U(t)$ with $t \in [\![0, p_X - 1]\!]$, as in the proof of Proposition 3.2.9. This concludes the proof. □

Definition 3.2.12. Let $U = (U_i)_{i \geq 0}$ be a sequence of integers and m a positive integer. If the sequence $(U_i \bmod m)_{i \geq 0}$ is ultimately periodic, then we let $\iota_U(m)$ denote its minimal preperiod and $\pi_U(m)$ denote its minimal period.

In the previous definition we have chosen the notation ι to allude to the word "index" which is equally used as preperiod in the literature.

Remark 3.2.13. Observe that for any sequence of integers $U = (U_i)_{i \geq 0}$ ultimately satisfying a linear recurrence relation of length k of the kind

$$U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i,\ \text{with } a_1, \ldots, a_k \in \mathbb{Z},\ a_k \neq 0, \tag{21}$$

we have

$$\forall m \in \mathbb{N} \setminus \{0\},\ N_U(m) \leq \pi_U(m) \leq (N_U(m))^k.$$

Therefore we have $\lim_{m \to +\infty} N_U(m) = +\infty \Leftrightarrow \lim_{m \to +\infty} \pi_U(m) = +\infty$. In particular, observe that it is the case for any linear numeration system; see Definition 1.6.6 and Examples 1.6.10 and 1.6.11 on page 13. Also, note that, for $m = p \cdot q$, with $\gcd(p, q) = 1$, we have $\pi_U(m) = \operatorname{lcm}(\pi_U(p), \pi_U(q))$.

Now, we want to obtain an upper bound on the preperiod of any ultimately periodic U-recognizable set recognized by a given DFA.

Proposition 3.2.14. *Let $U = (U_i)_{i \geq 0}$ be a linear numeration system and let X be an ultimately periodic U-recognizable set of non-negative integers of period p_X and preperiod a_X. Then any DFA accepting $\operatorname{rep}_U(X)$ has at least $|\operatorname{rep}_U(a_X - 1)| - \iota_U(p_X)$ states.*

The arguments of the following proof are similar those found in [**Hon86**].

Proof. Since U is linear, the sequence $(U_i \bmod p_X)_{i\geq 0}$ is ultimately periodic with preperiod $\iota_U(p_X)$ and period $\pi_U(p_X)$. Without loss of generality, we may assume $|\operatorname{rep}_U(a_X - 1)| - \iota_U(p_X) > 0$. Proceed by contradiction and assume that $\mathcal{A}$ is a DFA accepting $\operatorname{rep}_U(X)$ with less than $|\operatorname{rep}_U(a_X-1)|-\iota_U(p_X)$ s. Then there exist words w, w_4 such that the greedy U-representation of $a_X - 1$ can be factorized as

$$\operatorname{rep}_U(a_X - 1) = ww_4$$

with $|w| = |\operatorname{rep}_U(a_X - 1)| - \iota_U(p_X)$. By the pumping lemma w can be decomposed as $w = w_1w_2w_3$ with $w_2 \neq \varepsilon$ so that we have

$$\forall i \in \mathbb{N},\ w_1w_2^iw_3w_4 \in \operatorname{rep}_U(X) \Leftrightarrow w_1w_2w_3w_4 \in \operatorname{rep}_U(X).$$

By the minimality of a_X and p_X, we must have either $a_X - 1 \in X$ and, for all positive integers n, $a_X + np_X - 1 \notin X$, or $a_X - 1 \notin X$ and, for all positive integers n, $a_X + np_X - 1 \in X$. Using the ultimate periodicity of $(U_i \bmod p_X)_{i\geq 0}$ and since we have $|w_4| = \iota_U(p_X)$, we get

$$\begin{aligned}\forall i \in \mathbb{N},\ &\operatorname{val}_U(w_1w_2^{i\pi_U(p_X)}w_2w_3w_4)\\ &\equiv \operatorname{val}_U(w_1w_2w_3w_4) + i\operatorname{val}_U(w_2^{\pi_U(p_X)}0^{|w_2w_3w_4|}) \pmod{p_X}.\end{aligned}$$

Therefore, repeating the factor $w_2^{\pi_U(p_X)}$ of length multiple of $\pi_U(p_X)$ exactly p_X times does not change the value modulo p_X. Hence we get

$$\operatorname{val}_U(w_1w_2^{p_X\pi_U(p_X)}w_2w_3w_4) \equiv a_X - 1 \pmod{p_X},$$

leading to a contradiction. □

Theorem 3.2.15. *Let $U = (U_i)_{i\geq 0}$ be a linear numeration system such that $\mathbb{N}$ is U-recognizable and satisfying condition* (19)*. Assume*

$$\lim_{m\to+\infty} N_U(m) = +\infty.$$

Then it is decidable whether or not a U-recognizable set is ultimately periodic.

Proof. The sequence U ultimately satisfies a strict linear recurrence relation of length k of the kind (21), *i.e*, there exists $\ell \in \mathbb{N}$ such that U satisfies $U_{i+k} = a_1U_{i+k-1} + \cdots + a_kU_i$ for all integers $i \geq \ell$. Let the prime decomposition of $|a_k|$ be $|a_k| = p_1^{u_1}\cdots p_r^{u_r}$ with $u_j > 0$ for all $j \in [\![1, r]\!]$. Consider a U-recognizable set $X \subseteq \mathbb{N}$ that is given through a DFA $\mathcal{A}$ accepting $\operatorname{rep}_U(X)$ and let d be the number of states of this automaton.

Assume that X is ultimately periodic with period

$$p_X = p_1^{v_1}\cdots p_r^{v_r}c$$

with $v_1, \ldots, v_r, c \in \mathbb{N}$ and $\gcd(a_k, c) = 1$. Then, from Proposition 3.2.9, we obtain $d \geq N_U(p_X)$.

First, let us give a bound on the part c coprime with a_k. Using Remark 3.2.13, we obtain

$$N_U(c) \le \pi_U(c) \le \pi_U(p_X) \le (N_U(p_X))^k \le d^k.$$

For all integers $i \ge \ell$, the term U_{i+k} is determined by the k previous terms $U_{i+k-1}, \dots, U_i$. Moreover, $\gcd(a_k, c) = 1$ implies that a_k is invertible modulo c. Hence, for all integers $i \ge \ell$, the term $U_i \bmod c$ is also determined by the k following terms $U_{i+1} \bmod c, \dots, U_{i+k} \bmod c$. Therefore the shifted reduced sequence $(U_i \bmod c)_{i\ge\ell}$ is purely periodic. Then, from Definition 3.2.7, this sequence takes exactly $N_U(c)$ different values because any term appears infinitely often. Let $\alpha\colon \mathbb{N} \to \mathbb{N}$ be the function mapping any $m \in \mathbb{N}$ onto the smallest index i such that we have $U_i \ge m$. Since the sequence U is increasing, the map α is non-decreasing and we have $\lim_{m\to+\infty} \alpha(m) = +\infty$. If we have $\alpha(c) \le \ell$, then c is bounded by the first integer c_1 which satisfies $\alpha(c_1) > \ell$. Otherwise we have $U_\ell < \cdots < U_{\alpha(c)-1} < c$. From this observation and the pure periodicity of the sequence $(U_i \bmod c)_{i\ge\ell}$, it follows $N_U(c) \ge \alpha(c) - \ell$. Thus we obtain

$$\alpha(c) \le N_U(c) + \ell \le d^k + \ell.$$

Therefore, in this case, c is bounded by the first integer c_2 that satisfies $\alpha(c_2) > d^k + \ell$. These constants c_1 and c_2 are effectively computable and c is bounded by $c_0 := \max\{c_1, c_2\}$.

Now, take $j \in [\![1, r]\!]$. Using Remark 3.2.13 once again, we obtain

$$N_U(p_j^{v_j}) \le d^k.$$

The assumption $\lim_{m\to+\infty} N_U(m) = +\infty$ implies $\lim_{v\to+\infty} N_U(p_j^v) = +\infty$. Observe that the map $v \mapsto N_U(p_j^v)$ is non-decreasing. Consequently, the exponent v_j occurring in the decomposition of p_X is bounded by s_j where s_j is the smallest non-negative integer such that, for all integers $v \ge s_j$, we have $N_U(p_j^v) > d^k$. This bound s_j can be effectively computed as follows. For any non-negative integer v, we can find $N_U(p_j^v)$ in a finite number of operations by inspecting the first values of $(U_i \bmod p_j^v)_{i\ge0}$ and looking for two identical k-tuples made of k consecutive elements. Once the period is determined, we immediately obtains the values that are repeated infinitely often. Since the map $v \mapsto N_U(p_j^v)$ is non-decreasing, we have to compute

$$N_U(p_j) \le N_U(p_j^2) \le N_U(p_j^3) \le \cdots$$

until finding the first value s_j such that we have $N_U(p_j^{s_j}) > d^k$.

We have just proved that if X is ultimately periodic, then the admissible periods are bounded by the constant

$$P = p_1^{s_1} \cdots p_r^{s_r} c_0,$$

which is effectively computable. Then, using Proposition 3.2.14, the admissible preperiods a_X must satisfy

$$|\operatorname{rep}_U(a_X - 1)| \leq d + \max\{\iota_U(p) \mid p \in [\![1, P]\!]\}.$$

Since $m \mapsto |\operatorname{rep}_U(m)|$ is a non-decreasing map and the preperiods $\iota_U(p)$ are computable for any positive integer p, a bound on the admissible preperiods of X can effectively be given.

Consequently the sets of admissible preperiods and periods we have to check are finite. To each pair (a, p) of admissible preperiods and periods corresponds at most $2^a 2^p$ distinct ultimately periodic sets. Using Lemma 1.6.13, one can build an automaton for each of them and then compare the language L accepted by this automaton with $\operatorname{rep}_U(X)$. Since testing whether or not we have $L \setminus \operatorname{rep}_U(X) = \emptyset$ and $\operatorname{rep}_U(X) \setminus L = \emptyset$ is decidable algorithmically, this completes the proof. □

Remark 3.2.16. In the statement of the previous theorem, the assumption about the U-recognizability of $\mathbb{N}$ is of particular interest. Recall that, for an arbitrary linear numeration system, $\mathbb{N}$ is in general *not* U-recognizable. If $\mathbb{N}$ is U-recognizable, then U satisfies a linear recurrence relation over $\mathbb{Z}$ (see Theorem 1.6.8 on page 13) but the converse does not hold.

In view of the previous result, it is natural to wish to characterize linear recurrence sequences U that satisfy $\lim_{m\to+\infty} N_U(m) = +\infty$. This is exactly the aim of Section 3.5. However, we will already prove here the following proposition in this direction, which provides an interesting particular case of Theorem 3.2.15.

Proposition 3.2.17. *Let $U = (U_i)_{i\geq 0}$ be an increasing sequence of integers eventually satisfying a linear recurrence relation of length k of the kind* (21). *The following assertions are equivalent:*

(i) $\lim_{m\to+\infty} N_U(m) = +\infty$;
(ii) *for all prime divisors p of a_k,* $\lim_{v\to+\infty} N_U(p^v) = +\infty$.

In particular, $a_k = \pm 1$ implies $\lim_{m\to+\infty} N_U(m) = +\infty$.

Proof. Of course, it is enough to show that (ii) implies (i). Let the prime decomposition of $|a_k|$ be $|a_k| = p_1^{u_1} \cdots p_r^{u_r}$ with $u_1, \ldots, u_r > 0$. Observe that if a positive integer m can be decomposed as $m = p_1^{v_1} \cdots p_r^{v_r} c$, with $v_1, \ldots, v_r, c \in \mathbb{N}$ and $\gcd(a_k, c) = 1$, then we have

$$\pi_U(m) = \operatorname{lcm}(\pi_U(p_1^{v_1}), \ldots, \pi_U(p_r^{v_r}), \pi_U(c)).$$

Take $j \in [\![1, r]\!]$. By assumption, we have $\lim_{v\to+\infty} N_U(p_j^v) = +\infty$. Hence, in view of Remark 3.2.13, we get $\lim_{v\to+\infty} \pi_U(p_j^v) = +\infty$. Therefore $\pi_U(m)$ gets larger than any constant by considering integers m which are divisible by a sufficiently large power of p_j. Using Remark 3.2.13 again, the same conclusion holds for $N_U(m)$.

Let $C = \{c_0, c_1, c_2, \ldots\}$ be the set of natural numbers coprime to a_k and assume $c_0 < c_1 < c_2 < \cdots$. By hypothesis, there exists $\ell \in \mathbb{N}$ such that the sequence $(U_i \bmod c)_{i\geq\ell}$ is purely periodic. As in the proof of Theorem 3.2.15, let us consider the non-decreasing function $\alpha\colon \mathbb{N} \to \mathbb{N}$ mapping any $m \in \mathbb{N}$ onto the smallest index i satisfying $U_i \geq m$. Since U is increasing, we have $\lim_{m\to+\infty} \alpha(m) = +\infty$. Therefore there exists an integer N such that, for all integers $n \geq N$, we have $\alpha(c_n) > \ell$. Similarly to what was done in the proof of Theorem 3.2.15, we find $N_U(c_n) \geq \alpha(c_n) - \ell$ for all integers $n \geq N$. Consequently, we obtain

$$\lim_{n\to+\infty} N_U(c_n) = +\infty.$$

Any large enough integer m contains either a large power of p_j for some $j \in [\![1, r]\!]$ or some large c prime to a_k. Consequently (i) is satisfied. □

Corollary 3.2.18. *Let $U = (U_i)_{i\geq 0}$ be a linear numeration system such that $\mathbb{N}$ is U-recognizable. Assume that U eventually satisfies a linear recurrence relation of length k of the kind* (21) *with $a_k = \pm 1$ and condition* (19)*. Then it is decidable whether or not a U-recognizable set is ultimately periodic.*

Remark 3.2.19. We have thus obtained a decision procedure for Problem 2 when the linear numeration system U has a regular numeration language and satisfies (19) and $\lim_{m\to+\infty} N_U(m) = +\infty$; hence, in particular, when U ultimately satisfies a linear recurrence relation of the kind (21) with $a_k = \pm 1$. On the other hand, whenever $\gcd(a_1, \ldots, a_k) = g \geq 2$ holds in (21), we have $U_i \equiv 0 \bmod g^n$ for all $n \in \mathbb{N}$ and all large enough integers i. Therefore the only value taken infinitely often by the reduced sequence $(U_i \bmod g^n)_{i\geq 0}$ is 0. This implies that $N_U(m)$ equals 1 for infinitely many values of m. Hence, in this case, the assumption about $N_U(m)$ in Theorem 3.2.15 does not hold. In particular, note that the same observation can be made for the usual integer base numeration systems: for all non-negative integers n and b with $b \geq 2$, the only value taken infinitely often by the sequences $(b^i \bmod b^n)_{i\geq 0}$ is 0.

To conclude this section, we make a small digression by showing how to use a result of H. Engstrom about preperiods [**Eng31**] to get some particular linear numeration systems U satisfying $\lim_{m\to+\infty} N_U(m) = +\infty$. In his paper H. Engstrom was interested in the problem of finding a general period associated with a given linear recurrence relation modulo m for *any* initial conditions. Let us also mention the related work [**War33**] of M. Ward, in which he considered the problem where the initial conditions are fixed and then the period of the linear recurrence sequence reduced modulo m has to be determined.

Theorem 3.2.20. [**Eng31**, Theorem 9] *Let $U = (U_i)_{i\geq 0}$ be a strict linear recurrence sequence of length $k \geq 2$ satisfying* (21) *and p be a prime divisor of a_k. If there exists $s \in [\![1, k-1]\!]$ satisfying $a_k, \dots, a_{k-s+1} \equiv 0 \pmod p$ and $a_{k-s} \not\equiv 0 \pmod p$, then we have $\iota_U(p^v) \leq vs$ for all $v \in \mathbb{N}$.*

Remark 3.2.21. Assume that we are dealing with a strict linear numeration system $U = (U_i)_{i\geq 0}$ satisfying (21) and that the assumptions of the previous theorem hold for all prime divisors p of a_k, which is equivalent to have $\gcd(a_1, \dots, a_k) = 1$. So, for all prime divisors p of a_k, there are s_p in $[\![1, k-1]\!]$ such that we have $a_k, \dots, a_{k-s_p+1} \equiv 0 \pmod p$ and $a_{k-s_p} \not\equiv 0 \pmod p$. Let $\chi_U = x^k - a_1 x^{k-1} - \cdots - a_k$ denote the characteristic polynomial of the linear recurrence relation satisfied by U. Assume that $\beta > 1$ is a root of multiplicity $\ell \geq 1$ of χ_U satisfying

- $\beta > |\gamma|$ for any other root $\gamma \in \mathbb{C}$ of χ_U;
- $\beta < p^{1/s_p}$ for all prime divisors p of a_k.

From Proposition 1.5.10 there exists a constant $c > 0$ such that we have $U_i \sim c\, i^{\ell-1}\beta^i$ $(i \to +\infty)$. Let p be a prime divisor of a_k and let $j_p(v)$ denote the largest index j such that we have $U_j < p^v$. Let $t > s_p$ be a real number such that we have $\beta < p^{1/t} < p^{1/s_p}$. For all large enough integers v, we have $U_{\lfloor vt \rfloor} < p^v$, hence, we also have $j_p(v) \geq \lfloor vt \rfloor$. From the previous theorem it follows $\iota_U(p^v) \leq vs_p$ for all $v \in \mathbb{N}$. Therefore, for all large enough integers v, $U_{\iota_U(p^v)} < \cdots < U_{j_p(v)}$ are the first terms of the periodic part of the sequence $(U_i \bmod p^v)_{i\geq 0}$ and we obtain $N_U(p^v) \geq \lfloor vt \rfloor - vs_p + 1 \geq v(t - s_p)$. This means that, for all prime divisors p of a_k, we have $N_U(p^v) \to +\infty$ as $v \to +\infty$. Therefore, using Proposition 3.2.17, we obtain $\lim_{m\to+\infty} N_U(m) = +\infty$ and we can apply our decision procedure given by Theorem 3.2.15 whenever $\mathbb{N}$ is U-recognizable.

Example 3.2.22. Consider the linear recurrence sequence given by

$$\forall i \in \{0, 1, 2\},\ U_i = i + 1 \quad \text{and} \quad \forall i \in \mathbb{N},\ U_{i+3} = U_{i+1} + 3U_i.$$

The first few terms of the sequence are

$$1, 2, 3, 5, 9, 14, 24, 41, 66, 113, 189, 311, 528, 878, 1461, 2462, 4095.$$

With the above notation, we have $s_3 = 1$ and $\beta \simeq 1.6717 < 3$, and the other two complex roots have a modulus close to 1.34. Thanks to Theorem 3.2.20, the preperiod $\iota_U(3^v)$ is bounded by v. On the other hand, we have $U_i \sim c\,\beta^i$ for some $c > 0$. Note that we have $\beta < 3^{1/2} < 3$. Therefore, for all large enough integers v, we get $U_{2v} \sim c\,\beta^{2v} < 3^v$. Consequently, we obtain that the elements $U_v < \cdots < U_{2v}$ appear in the periodic part of $(U_i \bmod p^v)_{i\geq 0}$. In Figure 3.1 these elements have been underlined.

v	preperiod	period
3	$1, 2, 3$	$(\underline{5, 9, 14, 24}, 14, 12, 5, 0, 14, 15, 14, 3, 5, 18, 14, 6, 14, 21)$
4	$1, 2, 3, 5$	$(\underline{9, 14, 24, 41, 66}, 32, 27, 68, 42, 68, 3, 32, 45, 41, 60, 14, \ldots)$
5	$1, 2, 3, 5, 9$	$(\underline{14, 24, 41, 66, 113, 189}, 68, 42, 149, 3, 32, 207, 41, 60, \ldots)$

TABLE 3.1. The preperiod of $(U_i \bmod 3^v)_{i\geq 0}$ is bounded by v.

3.3. Background on the p-adic Numbers

The aim of this section is to recap some background information on the p-adic numbers. For more details in this area, see, for instance, the handbooks [**Kob84, Gou97, Rob00**] or the lecture notes [**Bak09**]. In particular, the proofs of all the result mentioned in this section can be found in those texts. This material will be used in Section 3.5 to characterize linear numeration systems U satisfying the condition $\lim_{m\to+\infty} N_U(m) = +\infty$. Remember that these systems are those for which we have exhibited a decision procedure for Problem 2; see Theorem 3.2.15 above.

Let p be a fixed prime number throughout this section. We can put an absolute value $|\cdot|_p$ on $\mathbb{Z}$ as follows.

Definition 3.3.1. For each integer $n \neq 0$, we can write $|n| = p^v\ell$ for non-negative integers v and ℓ with $\gcd(p, \ell) = 1$. With these notation, the *p-adic absolute value on* $\mathbb{Z}$ is defined by

$$|n|_p = \begin{cases} p^{-v}, & \text{if } n \neq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Example 3.3.2. We have

$$|7|_3 = |1|_3 = 1,\ |7|_7 = \frac{1}{7},$$
$$|15|_3 = |6|_3 = \frac{1}{3},\ |15|_5 = \frac{1}{5},\ |15|_7 = 1,$$
$$|9|_3 = |18|_3 = \frac{1}{9},\ |18|_2 = \frac{1}{2},\ |18|_5 = 1.$$

In particular, note that we have $|n|_p \leq 1$ for all $n \in \mathbb{Z}$. This absolute value extends to $\mathbb{Q}$ in a natural way. Formally, we have the following definition.

Definition 3.3.3. The p-adic absolute value on $\mathbb{Z}$ extends to $\mathbb{Q}$ by declaring

$$\forall m, n \in \mathbb{Z} \text{ with } n \neq 0, \; \left|\frac{m}{n}\right|_p = \frac{|m|_p}{|n|_p}.$$

In particular, for all $a, b \in \mathbb{Q}$, we have $|a \cdot b|_p = |a|_p \cdot |b|_p$. Furthermore, the p-adic absolute value satisfies

$$\forall a, b \in \mathbb{Q}, \; |a + b|_p \leq \max\{|a|_p, |b|_p\},$$

i.e, it is *non-Archimedean.*

Example 3.3.4. We have

$$\left|\frac{7}{3}\right|_3 = 3, \; \left|\frac{7}{3}\right|_7 = \frac{1}{7},$$
$$\left|\frac{15}{21}\right|_3 = \left|\frac{15}{3}\right|_3 = 1, \; \left|\frac{15}{3}\right|_5 = \frac{1}{5},$$
$$\left|\frac{9}{14}\right|_3 = \left|\frac{18}{7}\right|_3 = \frac{1}{9}, \; \left|\frac{18}{7}\right|_2 = \frac{1}{2}, \; \left|\frac{18}{7}\right|_7 = 7,$$
$$\left|\frac{9}{14} + \frac{7}{3}\right|_3 = \left|\frac{3^3 + 2 \cdot 7^2}{3 \cdot 14}\right|_3 = \frac{1}{\frac{1}{3}} = 3 = \max\left\{\left|\frac{9}{14}\right|_3, \left|\frac{7}{3}\right|_3\right\}.$$

Proposition 3.3.5. *There are Cauchy sequences in $\mathbb{Q}$ that do not have limit with respect to the p-adic absolute value. Hence the field $\mathbb{Q}$ is not complete with respect to the p-adic absolute value.*

In view of this result, it is natural to consider the completion of $\mathbb{Q}$.

Definition 3.3.6. The field of *p-adic numbers*, denoted by $\mathbb{Q}_p$, is the completion of $\mathbb{Q}$ with respect to the p-adic absolute value.

Recall that, in particular, this implies the following properties.

- The constructed set $\mathbb{Q}_p$ is an extension field of $\mathbb{Q}$.
- There exists a unique absolute value N on $\mathbb{Q}_p$ that satisfies
$$\forall a \in \mathbb{Q}, \; N(a) = |a|_p.$$
- The set $\mathbb{Q}$ is dense in $\mathbb{Q}_p$ with respect to N.
- This absolute value N is non-Archimedean if and only if the p-adic absolute value $|\cdot|_p$ on $\mathbb{Q}$ is.

Again, we denote this extended absolute value by $|\cdot|_p$.

Remark 3.3.7. Note that, since $\mathbb{Q}_p$ is a complete non-Archimedean field with respect to the p-adic absolute value, a series $\sum_{i\geq 0}\gamma_i$ converges in $\mathbb{Q}_p$ if and only if we have $\lim_{i\to+\infty}|\gamma_i|_p = 0$.

Definition 3.3.8. The closed unit ball

$$\mathbb{Z}_p := \{x \in \mathbb{Q}_p \mid |x|_p \leq 1\}$$

is called the set of *p-adic integers*.

It is easily verified that the p-adic integers form a subring of $\mathbb{Q}_p$. Therefore we will usually refer to $\mathbb{Z}_p$ as the ring of p-adic integers.

Proposition 3.3.9. *The set of ordinary integers $\mathbb{Z}$ is dense in $\mathbb{Z}_p$ with respect to the p-adic absolute value. Conversely, every Cauchy sequence in $\mathbb{Z}^{\mathbb{N}}$ has a limit in $\mathbb{Z}_p$ with respect to the p-adic absolute value.*

Another (equivalent) way to define the p-adic numbers is to consider them to be the formal expressions of the form

$$c_{-N}p^{-N} + \cdots + c_{-1}p^{-1} + c_0 + c_1p + c_2p^2 + \cdots ,$$

with $N \in \mathbb{Z}$ and $c_j \in [\![0, p-1]\!]$ for all integers $j \geq -N$. The p-adic integers are then identified with the formal expressions involving only non-negative powers of p. Such expressions are called the *p-adic developments* of p-adic numbers. In particular, a noteworthy property from this representation of the p-adic numbers is that the p-adic development of an ordinary non-negative integer is always finite and corresponds to its usual integer base p decomposition.

Proposition 3.3.10. *The field $\mathbb{Q}_p$ is not algebraically closed.*

In view of the previous proposition, we consider naturally the algebraic closure of $\mathbb{Q}_p$, which is denoted by $\overline{\mathbb{Q}_p}$. The next proposition shows that the absolute value $|\cdot|_p$ on $\mathbb{Q}_p$ extends to this algebraic closure.

Proposition 3.3.11. *The p-adic absolute value $|\cdot|_p$ on $\mathbb{Q}_p$ extends to a unique non-Archimedean absolute value N on $\overline{\mathbb{Q}_p}$, i.e., there exists a unique non-Archimedean absolute value N on $\overline{\mathbb{Q}_p}$ satisfying $N(x) = |x|_p$ whenever x belongs to $\mathbb{Q}_p$.*

Once again, we will denote the extended absolute value on $\overline{\mathbb{Q}_p}$ by $|\cdot|_p$. The algebraic closure is not complete, however, as shown by the following result.

Proposition 3.3.12. *There are Cauchy sequences in* $\overline{\mathbb{Q}_p}$ *that do not have limit with respect to the* p*-adic absolute value. In other words, the field* $\overline{\mathbb{Q}_p}$ *is not complete with respect to the* p*-adic absolute value.*

Again, in view of this negative result, it is natural to consider the completion of this algebraic closure $\overline{\mathbb{Q}_p}$.

Definition 3.3.13. The completion of $\overline{\mathbb{Q}_p}$ with respect to the p-adic absolute value is denoted by $\mathbb{C}_p$.

As previously, this new field $\mathbb{C}_p$ has a unique absolute value, still denoted by $|\cdot|_p$, that restricts to the p-adic absolute value on $\overline{\mathbb{Q}_p}$. Finally, the next theorem, which is sometimes called the *fundamental theorem of algebra for* p*-adic numbers*, shows that we have reached a complete algebraically closed field.

Theorem 3.3.14. *The field* $\mathbb{C}_p$ *is algebraically closed.*

3.4. Some Material about Finitely Generated Abelian Groups

As in the previous section, we recall here some useful material for the proof of Theorem 3.5.8 below. More details on this subject can be found, for instance, in [**BJN94**] or in [**Lan04**].

Definition 3.4.1. A *torsion Abelian group* is an Abelian group such that each of its elements has a finite order, that is, generates a finite subgroup.

Definition 3.4.2. A *torsion-free Abelian group* is an Abelian group such that none of its elements, except the neutral element, has a finite order.

Lemma 3.4.3. *Any finitely generated torsion Abelian group is finite.*

Definition 3.4.4. Let G be an Abelian group and let $(g_i)_{i\in I}$ be a family of elements of G. If every x in G can be uniquely decomposed as $x = \sum_{i\in I} x_i g_i$ with $x_i \in \mathbb{Z}$ and $x_i = 0$ for almost all i, then $(g_i)_{i\in I}$ is said to be a *basis* of G. A *free Abelian group* is an Abelian group that has a basis.

Lemma 3.4.5. *Any free Abelian group is torsion-free.*

Lemma 3.4.6. *Every finitely generated free Abelian group is isomorphic to* $\mathbb{Z}^e$ *for some* $e \in \mathbb{N}$*. In that case we say that* e *is the* rank *of the free Abelian group.*

The next theorem is central in the study of finitely generated Abelian groups. It will be used in the proof of Theorem 3.4.8. Note that this result does not hold if the hypothesis "finitely generated" is removed from the statement.

Theorem 3.4.7. *Any finitely generated torsion-free Abelian group is a free Abelian group.*

The following theorem is usually referred to as the *fundamental theorem of finitely generated Abelian groups.*

Theorem 3.4.8. *Any finitely generated Abelian group G is the direct summand of a free Abelian group F and a torsion Abelian group T, that is,*

$$G = F \oplus T \cong \mathbb{Z}^e \oplus T, \text{ for some } e \in \mathbb{N}.$$

3.5. Linear Recurrence Sequences and Residue Classes

As it was observed in Remark 3.2.19, since our approach to solving Problem 2 requires the condition $\lim_{m\to+\infty} N_U(m) = +\infty$, it can only be applied to sequences which eventually satisfy a linear recurrence relation of the kind (21) with $\gcd(a_1, \ldots, a_k) = 1$. In this section our aim is to determine which linear recurrence sequences of integers U have the property $\lim_{m\to+\infty} N_U(m) = +\infty$. To that end, in view of Proposition 3.2.17, it is clear that we only have to focus on the behavior of $N_U(p^v)$ for any prime p dividing a_k.

First, let us recall some characterizations and some stability results of linear recurrence sequences; for instance, see the books [**GKP94, EvdPSW03, BR09**]. A part of the following theorem has already been given in Chapter 1; see Proposition 1.5.10 on page 11.

Theorem 3.5.1. *Let K be a field of characteristic zero, k be a positive integer, and $a_1, \ldots, a_k$ be elements of K with $a_k \neq 0$. The following assertions are equivalent.*

- *The sequence $(U_i)_{i\geq 0} \in K^{\mathbb{N}}$ satisfies the following strict linear recurrence relation of length k:*

$$\forall i \in \mathbb{N},\ U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i.$$

- *The general term of the sequence $(U_i)_{i\geq 0} \in K^{\mathbb{N}}$ is given by*

$$\forall i \in \mathbb{N},\ U_i = P_1(i)\ \alpha_1^i + \cdots + P_s(i)\ \alpha_s^i,$$

where $\alpha_1, \ldots, \alpha_s$ are the roots of the polynomial $x^k - a_1 x^{k-1} - \cdots - a_k$ with respective multiplicities $m_1, \ldots, m_s$ and $P_1, \ldots, P_s$ are polynomials of degree less than $m_1, \ldots, m_s$ respectively.

Note that, in the third point of this statement, the roots $\alpha_1, \ldots, \alpha_s$ and the coefficients of the polynomials $P_1, \ldots, P_s$ are contained in some extension field of the coefficient field K.

Theorem 3.5.2. *Let K be a field, k be a positive integer, and $a_1, \ldots, a_k$ be elements of K. The following assertions are equivalent.*

- *The sequence $(U_i)_{i\geq 0} \in K^{\mathbb{N}}$ satisfies the following linear recurrence relation of length k:*

$$\forall i \in \mathbb{N},\ U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i.$$

- *The formal power series $\sum_{i\geq 0} U_i\, x^i \in K[[x]]$ is rational of the following form:*

$$\sum_{i\geq 0} U_i\, x^i = \frac{\sum_{i=0}^{k-1} U_i\, x^i - \sum_{i+j<k} a_i U_j\, x^{i+j}}{1 - a_1 x - \cdots - a_k x^k}. \tag{22}$$

The following proposition is known as *Fatou's lemma*.

Proposition 3.5.3. [**Fat06**] *A rational power series S with integer coefficients can always be written as $S = P/Q$ with $P, Q \in \mathbb{Z}[x]$ and $Q(0) = 1$.*

Lemma 3.5.4. *If a sequence $(U_i)_{i\geq 0}$ of integers satisfies a linear recurrence relation over $\mathbb{Q}$, then it also satisfies a linear recurrence relation over $\mathbb{Z}$. Furthermore, the shortest linear recurrence relation over $\mathbb{Q}$ satisfied by $(U_i)_{i\geq 0}$ has integer coefficients.*

Definition 3.5.5. Let R be a ring and $U = (U_i)_{i\geq 0}$ and $V = (V_i)_{i\geq 0}$ be sequences over R. The *sum* of U and V is the sequence $(U_i + V_i)_{i\geq 0}$. For all $c \in R$, the *scalar multiplication by c* of U is the sequence $(c\, U_i)_{i\geq 0}$. The *Hadamard product* of U and V is the sequence $(U_i\, V_i)_{i\geq 0}$. The *Cauchy product* of U and V is the sequence $(\sum_{j=0}^{i} U_j\, V_{i-j})_{i\geq 0}$.

Proposition 3.5.6. *Let R be a commutative ring. The class of linear recurrence sequences over R (resp. strict linear recurrence sequences over R) is closed under sum, scalar multiplication, the Hadamard product and the Cauchy product.*

Throughout this section, we let $U = (U_i)_{i\geq 0}$ be a sequence of integers satisfying the following strict linear recurrence relation of length k:

$$\forall i \in \mathbb{N},\ U_{i+k} = a_1 U_{i+k-1} + \cdots + a_k U_i,\ \text{with } a_1, \ldots, a_k \in \mathbb{Z},\ a_k \neq 0. \quad (23)$$

Note that this implies no loss of generality because, for any non-negative integer ℓ, if $U^{(\ell)}$ denotes the shifted sequence $(U_i)_{i\geq \ell}$, then we have

$$\lim_{m\to+\infty} N_U(m) = +\infty \Leftrightarrow \lim_{m\to+\infty} N_{U^{(\ell)}}(m) = +\infty.$$

Furthermore, we assume that U satisfies no shorter linear recurrence relation with coefficients in $\mathbb{Q}$. It is well known that this is equivalent to assume that k is the greatest integer satisfying

$$\det \begin{pmatrix} U_0 & \cdots & U_{k-1} \\ \vdots & & \vdots \\ U_{k-1} & \cdots & U_{2k-2} \end{pmatrix} \neq 0.$$

This result is sometimes referred to as *Kronecker's theorem*; for instance, see [**BR09**]. We let

$$\chi_U = x^k - a_1 x^{k-1} - \cdots - a_k$$

denote the associated characteristic polynomial and we define

$$P_U = x^k \chi_U(1/x) = 1 - a_1 x - \cdots - a_k x^k.$$

Remark 3.5.7. Since we have assumed $a_k \neq 0$, observe that if $\alpha_1, \ldots, \alpha_s$ are the roots of χ_U, then their reciprocals $1/\alpha_1, \ldots, 1/\alpha_s$ are exactly the roots of P_U. Note that, in particular, 0 cannot be a root of χ_U. Moreover, we have assumed that the linear recurrence relation (23) satisfied by U has shortest length k. Therefore the numerator and the denominator of the right-hand side in (22) corresponding to the generating function $\sum_{i\geq 0} U_i\, x^i$ of $(U_i)_{i\geq 0}$ are relatively prime. Hence the poles of $\sum_{i\geq 0} U_i\, x^i$ are exactly the roots of P_U, that is, the reciprocals of the roots of χ_U.

Our goal is to prove the following result.

Theorem 3.5.8. *Let p be a prime number. We have $N_U(p^v) \not\to +\infty$ as $v \to +\infty$ if and only if P_U can be factorized as $P_U = A \cdot B$ with $A, B \in \mathbb{Z}[x]$ satisfying:*

(i) $B \equiv 1 \pmod{p\,\mathbb{Z}[x]}$*;*

(ii) *A has no repeated roots and all its roots are roots of unity.*

Furthermore, in that case, we have $A(0) = B(0) = 1$.

We note that one direction is fairly simple. Therefore, for the sake of clarity, we have split the proof into two parts.

First part of the proof of Theorem 3.5.8. Assume that P_U has such a factorization $P_U = A \cdot B$. Of course we must have $A(0) = B(0) = 1$. From (ii) there is a natural number d such that A divides $x^d - 1$. From theorem 3.5.2 there exist polynomials Q and R such that we have

$$(x^d - 1) \sum_{i \geq 0} U_i \; x^i = \frac{(x^d - 1)\; Q}{P_U} = \frac{(x^d - 1)\; Q}{A \cdot B} = \frac{R}{B}.$$

Note that Q is an integer polynomial, and since we have $A \in \mathbb{Z}[x]$ by hypothesis, R must be an integer polynomial as well. Next, from (i), there exists an integer polynomial C such that we have $B = 1 - p\,C$. Hence we obtain

$$(x^d - 1) \sum_{i \geq 0} U_i \; x^i = \frac{R}{1 - p\,C} = \sum_{i \geq 0} p^i R\, C^i.$$

In particular, this implies that, for any fixed non-negative integer v, the series $(x^d - 1) \sum_{i \geq 0} U_i \; x^i$ is congruent to a polynomial modulo p^v. This means $U_{i+d} \equiv U_i \pmod{p^v}$ for all sufficiently large integers i. Therefore there are at most d values that occur infinitely often modulo p^v, that is, we have $N_U(p^v) \leq d$ for every $v \in \mathbb{N}$. □

Since the second part of the proof is much longer than the first one, we have separated the following technical lemmas from it. The first two can probably be considered to be well-known results in linear recurrence theory. Nevertheless, we restate them here for the sake of being thorough.

Lemma 3.5.9. *For all positive integers a and all $b \in [\![0,\, a-1]\!]$, the sequence $(U_{ai+b})_{i \geq 0}$ satisfies a strict linear recurrence relation over $\mathbb{Z}$ and the poles of its generating function $\sum_{i \geq 0} U_{ai+b}\, x^i$ are a^{th} powers of the poles of the generating function $\sum_{i \geq 0} U_i\, x^i$ of $(U_i)_{i \geq 0}$. Furthermore, the shortest linear recurrence relation over $\mathbb{Q}$ satisfied by $(U_{ai+b})_{i \geq 0}$ has integer coefficients.*

Proof. We know from Theorem 3.5.1 that we have

$$\forall i \in \mathbb{N},\; U_i = \sum_{j=1}^{s} P_j(i)\, \alpha_j^i,$$

where the α_j's are the roots of χ_U with multiplicities m_j and the P_j's are polynomials of degree at most $m_j - 1$. Take a positive integer a and take $b \in [\![0,\, a-1]\!]$. Then we obtain

$$\forall i \in \mathbb{N},\; U_{ai+b} = \sum_{j=1}^{s} Q_j(i)\, (\alpha_j^a)^i,$$

where $Q_j(i) = \alpha_j^b\, P_j(ai + b)$ is a polynomial of degree at most $m_j - 1$. Therefore, from Theorem 3.5.1, it follows that $(U_{ai+b})_{i \geq 0}$ satisfies the linear recurrence relation of characteristic polynomial $(x-\alpha_1^a)^{m_1}\cdots(x-\alpha_s^a)^{m_s}$. This polynomial can be rewritten as $(x-\beta_1^a)\cdots(x-\beta_k^a)$, where $\beta_1,\dots,\beta_k$ are the roots of χ_U, although they may be repeated. This polynomial has its coefficients in $\mathbb{Z}$ since it is a symmetric polynomial in $\beta_1,\dots,\beta_k$ and

$$\chi_U = (x-\alpha_1)^{m_1}\cdots(x-\alpha_s)^{m_s} = (x-\beta_1)\cdots(x-\beta_k)$$

has its coefficients in $\mathbb{Z}$.[4] From Theorems 3.5.1 and 3.5.2 the poles of $\sum_{i\geq 0} U_{ai+b}\, x^i$ are contained in the set $\{1/\alpha_1^a,\dots,1/\alpha_s^a\}$. The second part of the statement is a straightforward consequence of Lemma 3.5.4. □

Lemma 3.5.10. *For all positive integers a and for each pole γ of $\sum_{i\geq 0} U_i\, x^i$, there exists $b\in [\![0,\, a-1]\!]$ such that γ is an a^{th} root of a pole of $\sum_{i\geq 0} U_{ai+b}\, x^i$.*

Proof. Take a positive integer a. From Lemma 3.5.9 we know that, for all $b\in [\![0,\, a-1]\!]$, the shortest linear recurrence relation over $\mathbb{Q}$ satisfied by $(U_{ai+b})_{i\geq 0}$ has coefficients in $\mathbb{Z}$. So, from Theorem 3.5.2, we have

$$\forall b\in [\![0,\, a-1]\!],\ \sum_{i\geq 0} U_{ai+b}\, x^i = \frac{Q^{(b)}}{P_U^{(b)}} \text{ and } \sum_{i\geq 0} U_i x^i = \frac{Q}{P_U}$$

for some relatively prime polynomials $Q^{(b)}$ and $P_U^{(b)}$ in $\mathbb{Z}[x]$ and some polynomial Q in $\mathbb{Z}[x]$ coprime to P_U. Then we have

$$\begin{aligned}\forall b\in [\![0,\, a-1]\!],\ \frac{Q}{P_U} &= \sum_{b=0}^{a-1}\sum_{i\geq 0} U_{ai+b}\, x^{ai+b}\\ &= \sum_{b=0}^{a-1} x^b \sum_{i\geq 0} U_{ai+b}\,(x^a)^i\\ &= \sum_{b=0}^{a-1} x^b \frac{Q^{(b)}(x^a)}{P_U^{(b)}(x^a)}.\end{aligned}$$

We obtain that P_U divides $\prod_{b=0}^{a-1} P_U^{(b)}(x^a)$. So, if γ is a pole of $\sum_{i\geq 0} U_i\, x^i$, *i.e.*, a root of P_U, then there exists $b\in [\![0,\, a-1]\!]$ such that γ^a is a root of $P_U^{(b)}$, that is, a pole of $\sum_{i\geq 0} U_{ai+b}\, x^i$. This finishes the proof. □

Lemma 3.5.11. *Suppose that the multiplicative subgroup of $\mathbb{C}_p^\times$ generated by the poles of $\sum_{i\geq 0} U_i\, x^i$ is a free Abelian group. Take $b_1,\dots,b_d\in\mathbb{C}_p$ and set $V_i=\prod_{j=1}^d (U_i-b_j)$ for all $i\in\mathbb{N}$. If the rational power series $\sum_{i\geq 0} V_i\, x^i$*

[4]It is a well-known result that a symmetric polynomial in $A[\beta_1,\dots,\beta_k]$, where A is a commutative ring, is a polynomial in $A[e_0,\dots,e_{k-1}]$, where we define $(x-\beta_1)\cdots(x-\beta_k) = x^k - e_{k-1}x^{k-1} - \cdots e_0$. For instance, see [**Sti94, Lan04**].

in $\mathbb{C}_p[[x]]$ *has poles* $\beta_1, \ldots, \beta_r$ *in* $\mathbb{C}_p$ *satisfying* $|\beta_j|_p > 1$ *for all* $j \in [\![1,\, r]\!]$, *then every pole* $\gamma \in \mathbb{C}_p$ *of* $\sum_{i\geq 0} U_i\, x^i$ *satisfies either* $|\gamma|_p > 1$ *or* $\gamma = 1$.

Proof. Let $\alpha_1, \ldots, \alpha_s \in \mathbb{C}_p$ be the (distinct) roots of χ_U. Then, in view of Remark 3.5.7, $1/\alpha_1, \ldots, 1/\alpha_s$ are the poles of $\mathsf{U} := \sum_{i\geq 0} U_i\, x^i$. We first claim that we have $|\alpha_j|_p \leq 1$ for all $j \in [\![1,\, s]\!]$. To see this, note that, to be a pole, each $1/\alpha_j$ must satisfy $P_U(1/\alpha_j) = 0$, that is,

$$1 - a_1/\alpha_j - \cdots - a_k/{\alpha_j}^k = 0.$$

Consequently we have

$$\forall j \in [\![1,\, s]\!],\ \left| a_1/\alpha_j + \cdots + a_k {\alpha_j}^k \right|_p = |1|_p = 1.$$

Then, by the non-Archimedean property of the p-adic absolute value, for all $j \in [\![1,\, s]\!]$, there exists $\ell \in [\![1,\, k]\!]$ such that we have $|a_\ell/{\alpha_j}^\ell|_p \geq 1$ and so $(|\alpha_j|_p)^\ell \leq |a_\ell|_p$. Since we have $a_\ell \in \mathbb{Z}$, it comes $|a_\ell|_p \leq 1$, which was the first claim. So the poles of U have p-adic absolute value greater than or equal to 1. Without loss of generality, we may assume $|\alpha_1|_p = \cdots = |\alpha_t|_p = 1$ for some $t \in [\![0, s]\!]$ and $|\alpha_j|_p < 1$ for all $j \in [\![t+1, s]\!]$. If t is zero, then the lemma holds. Therefore we may assume $t \geq 1$.

Next, using Theorem 3.5.1, there exist polynomials $P_1, \ldots, P_s \in \mathbb{C}_p[x]$ such that we have

$$\forall i \in \mathbb{N},\ U_i = \sum_{j=1}^{s} P_j(i)\, \alpha_j^i.$$

Moreover, we have

$$\forall i \in \mathbb{N},\ V_i = \prod_{j=1}^{d} (U_i - b_j) = c_d U_i^d + c_{d-1} U_i^{d-1} + \cdots + c_0$$

for some $c_0,\, \ldots,\, c_{d-1},\, c_d \in \mathbb{C}_p$ with $c_d = 1$. Hence, by the multinomial theorem, we obtain

$$\begin{aligned}
\forall i \in \mathbb{N},\ V_i &= \sum_{j=0}^{d} c_j \sum_{\substack{j_1+\cdots+j_s=j \\ j_1,\ldots,j_s\in\mathbb{N}}} \binom{j}{j_1, \ldots, j_s} \prod_{\ell=1}^{s} (P_\ell(i)\, \alpha_\ell^i)^{j_\ell} \\
&= \sum_{j=0}^{d} c_j \sum_{\substack{j_1+\cdots+j_s=j \\ j_1,\ldots,j_s\in\mathbb{N}}} \binom{j}{j_1, \ldots, j_s} \prod_{\ell=1}^{s} (P_\ell(i))^{j_\ell} \left(\prod_{\ell=1}^{s} \alpha_\ell^{j_\ell} \right)^i .
\end{aligned}$$

The sequence $(V_i)_{i\geq 0}$ satisfies a strict linear recurrence relation over $\mathbb{C}_p$ by Proposition 3.5.6. Since the roots of the characteristic polynomial associated with the shortest linear recurrence relation satisfied by $(V_i)_{i\geq 0}$ are the reciprocals of the poles of the corresponding rational power series, from Theorems 3.5.1 and 3.5.2, it follows that the set of poles of $\mathsf{V} := \sum_{i\geq 0} V_i\, x^i$ is contained in the set

$$\left\{\prod_{\ell=1}^{s} \alpha_\ell^{-j_\ell} \mid j_1,\ldots,j_s \in \mathbb{N},\ j_1+\cdots+j_s \leq d\right\}.$$

From the assumptions the poles of V all have p-adic absolute values strictly greater than 1. Note that $\left|\prod_{\ell=1}^{s} \alpha_\ell^{-j_\ell}\right|_p = \left|\prod_{\ell=1}^{t} \alpha_\ell^{-j_\ell}\right|_p \left|\prod_{\ell=t+1}^{s} \alpha_\ell^{-j_\ell}\right|_p > 1$ holds if and only if we have $j_\ell > 0$ for some $\ell \in [\![t+1,s]\!]$. Therefore we can conclude that the possible poles of V only supported by products on $\alpha_1,\ldots,\alpha_t$ do not occur. So, for instance, α_1^{-d} is not a pole of V.

Let G denote the multiplicative subgroup of $\mathbb{C}_p^\times$ that is generated by $1/\alpha_1,\ldots,1/\alpha_t$. By assumption, G is a subgroup of a finitely generated free Abelian group. Hence we have $G \cong \mathbb{Z}^e$ for some $e \in \mathbb{N}$. To conclude the proof, it is enough to show $e = 0$ because, in that case, one can conclude that the only possible pole γ of U such that $|\gamma|_p = 1$ is 1. To see this, suppose $e > 0$ and let $\gamma_1,\ldots,\gamma_e$ be generators for a free Abelian group of rank e. Note that we must have $e \leq t$. Then we can write

$$\forall j \in [\![1,t]\!],\ \alpha_j = \prod_{i=1}^{e} \gamma_i^{b_{i,j}},$$

where $b_{i,j}$ are integers, and these decompositions are unique. We relabel if necessary so that we obtain

$$|b_{1,1}| = \max\{|b_{1,j}| \mid j \in [\![1,t]\!]\} > 0;$$
$$\forall i \in [\![2,e]\!],\ |b_{i,1}| = \max\{|b_{i,j}| \mid j \in [\![1,t]\!] \text{ and } \forall \ell \in [\![1,i-1]\!],\ b_{\ell,j} = b_{\ell,1}\}.$$

By construction, α_1^d cannot be written as a different word over $\alpha_1,\ldots,\alpha_t$ of length at most d. Then the expression for V_i above has an occurrence of

$$c_d\,(P_1(i))^d\,(\alpha_1^d)^i$$

that cannot be canceled by any other pole, by our selection of α_1. Consequently, α_1^{-d} should be a pole of V which contradicts the conclusion of the previous paragraph. So we have $e = 0$ and the lemma is proved. □

The following example illustrates the selection of α_1 in the previous proof.

Example 3.5.12. We keep the same notation as in the proof of the previous lemma. Suppose that U has 8 poles $1/\alpha_1,\ldots,1/\alpha_8 \in \mathbb{C}_p$ having modulus 1 and that these poles generates a multiplicative subgroup that is isomorphic to $\mathbb{Z}^3$. Let γ_1, γ_2, γ_3 be generators of this group. Assume

$$\forall j \in [\![1,8]\!],\ \alpha_j = \prod_{i=1}^{3} \gamma_i^{b_{i,j}},$$

where the $b_{i,j}$'s are the integers given by the matrix

$$B = (b_{i,j})_{\substack{i\in[\![1,3]\!]\\ j\in[\![1,8]\!]}} = \begin{pmatrix} -9 & -9 & 8 & 4 & 9 & 7 & -9 & 2 \\ 4 & 3 & 0 & -6 & 9 & 2 & 4 & 9 \\ 7 & 2 & 1 & 2 & -1 & 5 & 6 & 0 \end{pmatrix}.$$

For instance, we have $\alpha_1 = \gamma_1^{-9}\gamma_2^4\gamma_3^7$ and $\alpha_4 = \gamma_1^4\gamma_2^{-6}\gamma_3^2$. Since γ_1, γ_2, γ_3 are generators, the equation $\alpha_1^d = \alpha_1^{\ell_1}\cdots\alpha_8^{\ell_8}$ is equivalent to the system of equations

$$B\ (\ell_1,\dots,\ \ell_8)^t = (-9d,\ 4d,\ 7d)^t,$$

where x^t denotes the transposed vector of the vector x. Observe that two columns of B cannot be equal because the poles $1/\alpha_1,\dots,1/\alpha_8$ are distinct. We are looking for solutions $(\ell_1,\dots,\ \ell_8) \in \mathbb{N}^8$ satisfying $\ell_1 + \cdots + \ell_8 \le d$. Since we have

$$9 = |b_{1,1}| = \max\{|b_{1,j}| \mid j \in [\![1,8]\!]\},$$

the first equation of this system implies $\ell_3 = \ell_4 = \ell_6 = \ell_8 = 0$. This is because if $\ell_j > 0$ holds for some $j \in \{3,4,6,8\}$, then we would get

$$9d = |9\ell_1 + 9\ell_2 - 8\ell_3 - 4\ell_4 - 9\ell_5 - 7\ell_6 - 9\ell_7 - 2\ell_8| < 9(\ell_1 + \cdots + \ell_8),$$

which is impossible. Hence we obtain $\ell_1 + \ell_2 - \ell_5 + \ell_7 = d$ which, together with the equation $\ell_1 + \ell_2 + \ell_5 + \ell_7 \le d$, implies $\ell_5 = 0$. Then, by the same reasoning, since we have

$$\begin{aligned} 4 = |b_{2,1}| &= \max\{|b_{2,j}| \mid j \in [\![1,8]\!],\ b_{1,j} = b_{1,1}\} \\ &= \max\{|b_{2,j}| \mid j \in \{1,2,7\}\} = \max\{4,3\}, \end{aligned}$$

the second equation implies $\ell_2 = 0$. Finally, since we have

$$\begin{aligned} 7 = |b_{3,1}| &= \max\{|b_{3,j}| \mid j \in [\![1,8]\!],\ b_{1,j} = b_{1,1},\ b_{2,j} = b_{2,1}\} \\ &= \max\{|b_{2,j}| \mid j \in \{1,7\}\} = \max\{7,6\}, \end{aligned}$$

we get $\ell_7 = 0$ and $\ell_1 = d$ using the third equation. This means that α_1^d cannot be written as another word over $\alpha_1,\dots,\alpha_8$ of length less than or equal to d.

Lemma 3.5.13. *Assume $N_U(p^v) \not\to +\infty$ as $v \to +\infty$ and any pole γ of the rational series $\sum_{i\ge0} U_i\, x^i$ either satisfies $|\gamma|_p > 1$ or is a root of unity. Then the poles of $\sum_{i\ge0} U_i\, x^i$ that are roots of unity are simple.*

PROOF. We first note that $v \mapsto N_U(p^v)$ is a non-decreasing function, *i.e.*, for all $v, w \in \mathbb{N}$, we have $N_U(p^w) \geq N_U(p^v)$ whenever $w \geq v$. In particular, since one has $N_U(p^v) \not\to +\infty$, there are integers N and d such that, for all integers $v \geq N$, we have $N_U(p^v) = d$. Hence, for all integers $v \geq N$, we can pick integers $a_{1,v}, \ldots, a_{d,v}$ such that $U_i \equiv a \pmod{p^v}$ for infinitely many i implies $a \equiv a_{j,v} \pmod{p^v}$ for some $j \in [\![1, d]\!]$. Since we have

$$\{a_{1,w} \bmod p^v, \ldots, a_{d,w} \bmod p^v\} = \{a_{1,v} \bmod p^v, \ldots, a_{d,v} \bmod p^v\}$$

for all integers $v, w \geq N$ satisfying $w \geq v$, there is no loss of generality to assume that in these conditions, we have

$$\forall j \in [\![1, d]\!],\ a_{j,w} \equiv a_{j,v} \pmod{p^v}.$$

It follows that, for each $j \in [\![1, d]\!]$, the integer sequence $(a_{j,v})_{v \geq N}$ is Cauchy in $\mathbb{Z}_p$. Hence, by Proposition 3.3.9, there exist $b_1, \ldots, b_d \in \mathbb{Z}_p$ such that we have

$$\forall j \in [\![1, d]\!],\ a_{j,v} \xrightarrow{|\cdot|_p} b_j \text{ as } v \to +\infty. \tag{24}$$

Let $\alpha_1, \ldots, \alpha_s \in \mathbb{C}_p$ be the distinct roots of χ_U. Then $1/\alpha_1, \ldots, 1/\alpha_s$ are the poles of $\mathsf{U} := \sum_{i \geq 0} U_i\, x^i$; see Remark 3.5.7. By hypothesis, we may assume that there exists $t \in [\![0, s]\!]$ such that $\alpha_1, \ldots, \alpha_t$ are roots of unity and that, for all $j \in [\![t+1, s]\!]$, we have $|\alpha_j|_p < 1$. Then, from Theorem 3.5.1, there exist polynomials $P_1, \ldots, P_s \in \mathbb{C}_p[x]$ such that we have

$$\forall i \in \mathbb{N},\ U_i = \sum_{j=1}^{s} P_j(i)\, \alpha_j^i = \underbrace{\sum_{j=1}^{t} P_j(i)\, \alpha_j^i}_{T_i :=} + \underbrace{\sum_{j=t+1}^{s} P_j(i)\, \alpha_j^i}_{W_i :=}.$$

Since, for all $j \in [\![t+1, s]\!]$, we have $|\alpha_j|_p < 1$ and since, for any polynomial Q in $\mathbb{C}_p[x]$, the set $\{|Q(i)|_p \mid i \in \mathbb{N}\}$ is bounded by a constant, we get

$$|U_i - T_i|_p = |W_i|_p \to 0 \text{ as } i \to +\infty. \tag{25}$$

Since $\alpha_1, \ldots, \alpha_t$ are roots of unity, there exists a positive integer a such that we have $\alpha_j^a = 1$ for all $j \in [\![1, t]\!]$. We define

$$\forall b \in [\![0,\, a-1]\!],\ \forall i \in \mathbb{N},\ T_i^{(b)} = T_{ai+b}.$$

Thus

$$T_i^{(b)} = \sum_{j=1}^{t} P_j(ai+b)\, \alpha_j^{ai+b} = \sum_{j=1}^{t} \alpha_j^b\, P_j(ai+b),\ b = 0, \ldots, a-1,$$

are polynomials in i with coefficients in $\mathbb{C}_p$. For all $b \in [\![0,\, a-1]\!]$, we let Q_b denote this polynomial, that is, $Q_b(x) = T_x^{(b)} \in \mathbb{C}_p[x]$.

Choose $\varepsilon > 0$. By the definition of the b_j's, for all large enough integers i, there exists $\ell(i) \in [\![1, d]\!]$ such that we have $|U_i - b_{\ell(i)}|_p < \varepsilon$. In view of (25), we obtain

$$|(T_i - b_1)\cdots(T_i - b_d)|_p \leq \prod_{j=1}^{d} (|T_i - U_i|_p + |U_i - b_{\ell(i)}|_p + |b_{\ell(i)} - b_j|_p) \to 0$$

as $i \to +\infty$ because every factor is bounded by a constant and one tends to zero. Consequently, since we have $T_i^{(b)} = Q_b(i)$ for all $i \in \mathbb{N}$ and $b \in [\![0, a-1]\!]$, we obtain

$$\forall b \in [\![0, a-1]\!],\ |(Q_b(i) - b_1)\cdots(Q_b(i) - b_d)|_p \to 0 \text{ as } i \to +\infty.$$

For each $b \in [\![0, a-1]\!]$, consider the polynomial R_b defined by

$$\forall i \in \mathbb{N},\ R_b(i) = (Q_b(i) - b_1)\cdots(Q_b(i) - b_d).$$

From the binomial theorem, for each positive integer i, we have

$$|R_b(i + p^v) - R_b(i)|_p \leq Cp^{-v} \to 0 \text{ as } v \to +\infty,$$

where C is the maximum of the p-adic absolute values of the coefficients of the polynomial R_b. So, since we also have $|R_b(i)|_p \to 0$ as $i \to +\infty$, we obtain that, for all integers i, we have

$$|R_b(i)|_p \leq |R_b(i) - R_b(i + p^v)|_p + |R_b(i + p^v)|_p \to 0 \text{ as } v \to +\infty.$$

We have thus shown that, for each $b \in [\![0, a-1]\!]$, the polynomial R_b vanishes at all positive integers i. So each Q_b is a constant polynomial since it takes infinitely often the same value amongst $b_1, \ldots, b_d$. It follows that T_i is purely periodic with period a:

$$\forall i \in \mathbb{N},\ T_{i+a} = T_i.$$

Consequently, since the p-adic absolute value in non-Archimedean and in view of Remark 3.3.7, the rational series $(x^a - 1)\,\mathsf{U}$ has no pole on the closed unit disc because we have

$$|U_{i+a} - U_i|_p \leq |T_{i+a} - T_i|_p + |W_{i+a} - W_i|_p \to 0 \text{ as } i \to +\infty.$$

In particular, it has no pole on the unit circle. From the minimality assumption on the length k of the linear recurrence relation satisfied by U, we have $\mathsf{U} = Q/P_U$ with Q and P_U relatively prime; see Remark 3.5.7. Let $m_1, \ldots, m_t$ denote the respective multiplicities of $1/\alpha_1, \ldots, 1/\alpha_t$ as poles of U. The corresponding factor $(1-\alpha_1 x)^{m_1}\cdots(1-\alpha_t x)^{m_t}$ of P_U must divide $x^a - 1$ since the rational series $(x^a - 1)\,Q/P_U$ has no pole on the unit circle. Consequently, the multiplicities $m_1, \ldots, m_t$ are all equal to 1. This concludes the proof. □

Now, we come back to the proof of Theorem 3.5.8.

Second part of the proof of Theorem 3.5.8. To do the other direction is a little more work and we use p-adic methods. First, we define $b_1, \dots, b_d \in \mathbb{Z}_p$ satisfying (24) the same way as in the first part of the proof of Lemma 3.5.13. We set

$$\forall i \in \mathbb{N},\ V_i = \prod_{j=1}^{d} (U_i - b_j).$$

We have $V_i \in \mathbb{Z}_p$ for all $i \in \mathbb{N}$. By construction, for any fixed non-negative integer v, one has $|V_i|_p \leq p^{-v}$ for all sufficiently large integers i. Hence we have $|V_i|_p \to 0$ as $i \to +\infty$. From Proposition 3.5.6 the sequence $(V_i)_{i\geq 0}$ satisfies a strict linear recurrence relation over $\mathbb{Z}_p$. Therefore, from Theorem 3.5.2, the power series

$$\mathsf{V} := \sum_{i\geq 0} V_i x^i$$

is a rational power series in $\mathbb{Q}_p[[x]]$. Moreover, V converges on the closed unit disc $\mathbb{Z}_p$ since we have $|V_i|_p \to 0$, which is enough to guarantee convergence; see Remark 3.3.7. Since V is a rational series and it converges on the unit disc, its poles $\beta_1, \dots, \beta_r \in \mathbb{C}_p$ must satisfy $|\beta_j|_p > 1$ for all $j \in [\![1, r]\!]$.

To continue the proof, we will make use of Lemma 3.5.11. We note that the statement of this lemma is very close to what we already have, but it makes the additional assumption that the poles of $\mathsf{U} := \sum_{i\geq 0} U_i x^i$ generate a free Abelian multiplicative subgroup. In general, the poles of U generate a finitely generated Abelian multiplicative subgroup of $\mathbb{C}_p^\times$. From the so-called fundamental theorem of finitely generated Abelian groups, that is Theorem 3.4.8, this group is isomorphic to $\mathbb{Z}^e \otimes T$, for some finite Abelian group T and integer $e \geq 0$.

Let us show how to get rid of the torsion group T to be able to invoke Lemma 3.5.11. Set $a = \operatorname{Card} T$. We may assume $a \geq 1$. For any $b \in [\![0, a-1]\!]$, instead of taking the sequence $(U_i)_{i\geq 0}$, consider the sequence $(U_{ai+b})_{i\geq 0}$. In view of Lemma 3.5.9, the shortest linear recurrence relation satisfied by this sequence has integer coefficients and the poles of its generating function $\mathsf{U}^{(b)}$ are a^{th} powers of the poles of U. Consequently, the poles of $\mathsf{U}^{(b)}$ generate a finitely generated torsion-free Abelian group, which is necessarily a free Abelian group by Theorem 3.4.7. By the same reasoning as in the proof of Lemma 3.5.9, the generating function

$$\mathsf{V}^{(b)} := \sum_{i\geq 0} V_{ai+b}\, x^i$$

is a rational power series over $\mathbb{Z}_p$ and its poles are a^{th} powers of the poles $\beta_1, \dots, \beta_r$ of V. Since the poles of V satisfy $|\beta_j|_p > 1$ for all $j \in [\![1, r]\!]$, we obtain $|{\beta_j}^a|_p = |\beta_j|_p^a > 1$ for all $j \in [\![1, r]\!]$.

Now, we can invoke Lemma 3.5.11 applied to any sequence $(U_{ai+b})_{i\geq 0}$ with $b \in [\![0, a-1]\!]$. We deduce that, for all $b \in [\![0, a-1]\!]$, any pole γ of $\mathsf{U}^{(b)}$ satisfies either $|\gamma|_p > 1$ or $\gamma = 1$. In particular, each pole of U, which is an a^{th} root of a pole of $\mathsf{U}^{(b)}$ for some $b \in [\![0, a-1]\!]$ by Lemma 3.5.10, is either strictly greater than 1 in p-adic absolute value or a root of unity. Remember that the poles of U are precisely the roots of P_U; see Remark 3.5.7. Hence we factor

$$P_U = (1 - \delta_1 x) \cdots (1 - \delta_k x),$$

where $1/\delta_1, \ldots, 1/\delta_k \in \mathbb{C}_p$ corresponds to the poles of U, although they may be repeated. Let us factor P_U as $P_U = A \cdot B$ with

$$A = \prod_{\substack{|\delta_j|_p = 1 \\ j \in [\![1,k]\!]}} (1 - \delta_j x) \quad \text{and} \quad B = \prod_{\substack{|\delta_j|_p < 1 \\ j \in [\![1,k]\!]}} (1 - \delta_j x).$$

By assumption we have $P_U \in \mathbb{Z}[x]$. Moreover, if K is a splitting field of P_U over $\mathbb{Q}$, then any automorphism of K must permute the set of the δ_j's with $|\delta_j|_p < 1$, since the automorphism permutes the entire set of the δ_j's and it must send roots of unity to roots of unity. From the fundamental theorem of Galois theory, the rationals are exactly the elements in K that are fixed by every automorphism of K. Thus B is a rational polynomial. Furthermore, we have $B(0) = 1$. Note that, for $n > 0$, the coefficient of x^n in B is given by a sum of products of n elements in the set $\{\delta_j \mid j \in [\![1, k]\!], |\delta_j|_p < 1\}$. The set of algebraic integers is a subring of $\mathbb{C}_p$ and the only rationals that are algebraic integers are in fact integers. Since the δ_j's are algebraic integers, we get that B is an integer polynomial. Moreover, since the p-adic absolute value is non-Archimedean, the coefficient of x^n in B, with $n > 0$, has p-adic absolute value strictly less than 1. Note that an integer m satisfying $|m|_p < 1$ is necessarily a multiple of p. Hence we obtain $B \equiv 1 \pmod{p\mathbb{Z}[x]}$.

Now, let us turn to the polynomial A. We have $A(0) = 1$ and $A \in \mathbb{Z}[x]$ by the same reasoning as before. Moreover, the roots of A are roots of unity. From Lemma 3.5.13 the poles of U that are roots of unity are simple. Consequently, A has no repeated roots, which completes the proof. □

As it was stressed in the introduction of this section, we are now able to check whether or not the number of values taken infinitely often by a linear recurrence sequence modulo p^v tends to infinity as $v \to +\infty$. Let us consider the following example.

Example 3.5.14. Let us consider once again the linear recurrence sequence $U = (U_i)_{i\geq 0}$ given in [**Fro97**] and defined by $U_{i+4} = 3U_{i+3} + 2U_{i+2} + 3U_i$ for all $i \in \mathbb{N}$ and $U_i = i + 1$ for all $i \in [\![0, 3]\!]$. As shown in [**Fro97**], addition within this linear numeration system is not computable by a finite automaton. Nevertheless, we can show $N_U(3^v) \to +\infty$ as $v \to +\infty$ by applying the previous theorem. We have $P_U = 1 - 3x - 2x^2 - 3x^4$ and it is easily verified that this polynomial cannot be factorized as $A \cdot B$ with two factors satisfying the hypotheses of Theorem 3.5.8. This example shows that our decision procedure given by Theorem 3.2.15 can take care of numeration systems not handled by [**ARS09, Ler05, Muc03**].

Example 3.5.15. Consider a sequence $U = (U_i)_{i\geq 0}$ satisfying the following linear recurrence relation:

$$\forall i \in \mathbb{N},\ U_{i+5} = 6U_{i+4} + 3U_{i+3} - U_{i+2} + 6U_{i+1} + 3U_i.$$

With the above notation, we have $\chi_U = x^5 - 6x^4 - 3x^3 + x^2 - 6x - 3$ and

$$P_U = 1 - 6x - 3x^2 + x^3 - 6x^4 - 3x^5 = \underbrace{(x^3+1)}_{A}\underbrace{(-3x^2-6x+1)}_{B}.$$

With the initial conditions $U_i = i + 1$ for $i \in [\![0,4]\!]$, the corresponding sequence does not satisfy any shorter linear recurrence relation over $\mathbb{Q}$ since we have

$$\det \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 54 \\ 3 & 4 & 5 & 54 & 359 \\ 4 & 5 & 54 & 359 & 2344 \\ 5 & 54 & 359 & 2344 & 15129 \end{pmatrix} = 8458240 \neq 0.$$

Even if the greatest common divisor of the coefficients of the linear recurrence relation is 1, since P_U satisfies the assumptions of Theorem 3.5.8 for $p = 3$, we obtain $N_U(3^v) \not\to +\infty$ as $v \to +\infty$. The following table gives the first values of the function $v \mapsto N_U(3^v)$.

v	period	$N_U(3^v)$
1	$(1, 0, 1, 2, 0, 2)$	3
2	$(4, 0, 1, 5, 0, 8)$	5
3	$(22, 9, 19, 5, 18, 8)$	6
4	$(49, 63, 19, 32, 18, 62)$	6
5	$(211, 225, 19, 32, 18, 224)$	6
⋮	⋮	⋮

3.6. A Decision Procedure for a Class of Abstract Numeration Systems

We know that any ultimately periodic set of non-negative integers is S-recognizable for any abstract numeration system S and that a DFA accepting the set $\mathrm{rep}_S(X)$ of the S-representations in such a set X can effectively be obtained; see Theorem 1.7.9 on page 19. Then it makes sense to consider the analogous to Problem 2 in the larger framework of abstract numeration systems. Remember that abstract numeration systems are a generalization of positional numeration systems $U = (U_i)_{i\geq 0}$ for which $\mathbb{N}$ is U-recognizable. Thus, in this section, we consider, with some extra hypotheses on the abstract numeration system, the following decidability question.

Problem 3. Let

- S be an abstract numeration system;
- X be any S-recognizable set of non-negative integers, which is given through a DFA accepting $\mathrm{rep}_S(X)$.

Is it decidable whether or not X is ultimately periodic, *i.e.*, whether or not X is a finite union of arithmetic progressions?

As shown by Proposition 1.7.6 on page 18, when computing the map val_S in an abstract numeration system $S = (L, \Sigma, <)$, the different sequences $(\mathbf{u}_q(i))_{i\geq 0}$, for $q \in Q_L$, play the role of the single sequence $(U_i)_{i\geq 0}$ defining a positional numeration system as in Definition 1.6.1 on page 12. Moreover, recall that the sequences $(\mathbf{u}_q(i))_{i\geq 0}$ and $(\mathbf{v}_q(i))_{i\geq 0}$ satisfy linear recurrence relations with integer coefficients; see Proposition 1.5.7 on page 10.

We have then the following proposition analogous to Proposition 3.2.9.

Proposition 3.6.1. *Let $S = (L, \Sigma, <)$ be an abstract numeration system such that all states q of the trim minimal automaton $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ of L satisfy*

$$\lim_{i\to+\infty} \mathbf{u}_q(i) = +\infty$$

and such that we have $\mathbf{u}_L(i) > 0$ for all $i \in \mathbb{N}$. Let $X \subseteq \mathbb{N}$ be an ultimately periodic set of period p_X. Then any DFA accepting $\mathrm{rep}_S(X)$ has at least $\lceil N_{\mathbf{v}}(p_X)/\operatorname{Card} Q\rceil$ states, with $\mathbf{v} = (\mathbf{v}_L(i))_{i\geq 0}$.

Proof. Let a_X be the preperiod of X. By hypothesis there exists a minimal constant $J > 0$ such that, for all $i \geq J$ and all states q of the trim minimal automaton of L, we have $\mathbf{u}_q(i) \geq p_X$. For any $i \in \mathbb{N}$, consider the word

$$w_i = \mathrm{rep}_S(\mathbf{v}_L(i))$$

which corresponds to the first word of length $i + 1$ in the genealogically ordered language L. Consequently, for any integer $i \geq J - 1$, the word w_i is factorizable as $w_i = x_i y_i$ with $|y_i| = J$ and we define $q_i = \delta_L(q_0, x_i)$. Note that each y_i is the smallest word of length J accepted from q_i. By the definition of J for each integer $i \geq J - 1$, there are at least p_X words of length J leading from q_i to a final state. We order them using the genealogical order and we let

$$y_i = y_{i,0} < y_{i,1} < \cdots < y_{i,p_X-1}$$

denote the first p_X of them. Note that we have

$$\forall t \in [\![0, p_X - 1]\!],\ \mathrm{val}_S(x_i y_{i,t}) = \mathrm{val}_S(x_i y_i) + t = \mathbf{v}_L(i) + t.$$

The sequence $\mathbf{v} := (\mathbf{v}_L(i))_{i\geq 0}$ satisfies a linear recurrence relation over $\mathbb{Z}$ by Proposition 1.5.7. Hence the sequence $(\mathbf{v}_L(i) \bmod p_X)_{i\geq 0}$ is ultimately periodic and takes infinitely often $N := N_{\mathbf{v}}(p_X)$ different values. Let $h_1, \ldots, h_N$ be integers greater than J such that we have

$$\forall i, j \in [\![1, N]\!],\ \mathbf{v}_L(h_i) \geq a_X \text{ and } (i \neq j \Rightarrow \mathbf{v}_L(h_i) \not\equiv \mathbf{v}_L(h_j) \pmod{p_X}).$$

In particular, we have

$$\forall i \in [\![1, N]\!],\ \mathrm{rep}_S(\mathbf{v}_L(h_i)) = w_{h_i} = x_{h_i} y_{h_i} \text{ and } q_{h_i} = \delta_L(q_0, x_{h_i}).$$

The elements in the set $\{q_{h_1}, \ldots, q_{h_N}\}$ can take only $\mathrm{Card}\, Q$ different values. So at least $\sigma := \lceil N / \mathrm{Card}\, Q \rceil$ of them are the same. For the sake of simplicity, assume that they are $q_{h_1}, \ldots, q_{h_\sigma}$. For all $i, j \in [\![1, \sigma]\!]$ and all $t \in [\![0, p_X - 1]\!]$, we thus have $y_{h_i,t} = y_{h_j,t}$. By Lemma 3.2.6, for all distinct i, j in $[\![1, \sigma]\!]$, there exists $t_{i,j} \in [\![0, p_X - 1]\!]$ such that we have either $\mathbf{v}_L(h_i) + t_{i,j} \in X$ and $\mathbf{v}_L(h_j) + t_{i,j} \notin X$, or $\mathbf{v}_L(h_i) + t_{i,j} \notin X$ and $\mathbf{v}_L(h_j) + t_{i,j} \in X$. Therefore the words x_{h_i} and x_{h_j} do not belong to the same equivalence class for the Myhill-Nerode equivalence relation $\sim_{\mathrm{rep}_S(X)}$. This can be shown by concatenating the word $y_{h_i,t_{i,j}} = y_{h_j,t_{i,j}}$. Hence, from Definition 1.3.6 on page 6, the minimal automaton of $\mathrm{rep}_S(X)$ has at least σ states. The conclusion easily follows. □

Corollary 3.6.2. *Let $S = (L, \Sigma, <)$ be an abstract numeration system having the same properties as in Proposition 3.6.1. Assume that the sequence $\mathbf{v} = (\mathbf{v}_L(i))_{i\geq 0}$ satisfies*

$$\lim_{m\to+\infty} N_{\mathbf{v}}(m) = +\infty.$$

Then the period of an ultimately periodic set $X \subseteq \mathbb{N}$ such that $\mathrm{rep}_S(X)$ is accepted by a DFA with d states is bounded by the smallest non-negative integer s such that, for all integers $m \geq s$, we have $N_{\mathbf{v}}(m) > d\, \mathrm{Card}\, Q$.

Proposition 3.6.3. *Let $S = (L, \Sigma, <)$ be an abstract numeration system, let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be the trim minimal automaton of L, and let $X \subseteq \mathbb{N}$ be an ultimately periodic set of period p_X and preperiod a_X. Define $\mathbf{u}_q = (\mathbf{u}_q(i))_{i\geq 0}$ and $I = \max\{\iota_{\mathbf{u}_q}(p_X) \mid q \in Q\}$. Then any DFA accepting $\mathrm{rep}_S(X)$ has at least $\lceil (|\mathrm{rep}_S(a_X - 1)| - I) / \mathrm{Card}\, Q \rceil$ states.*

Proof. For all $q \in Q_L$, the sequence $\mathbf{u}_q$ satisfies a linear recurrence relation over $\mathbb{Z}$ by Proposition 1.5.7. The sequences $(\mathbf{u}_q(i) \bmod p_X)_{i\geq 0}$ are thus ultimately periodic and I is well defined. Now, define $P = \mathrm{lcm}_{q\in Q}\, \pi_{\mathbf{u}_q}(p_X)$. Let $\mathcal{B} = (R, \Sigma, \lambda, r_0, G)$ be a DFA accepting $\mathrm{rep}_S(X)$. Our aim is to show

$$|\,\mathrm{rep}_S(a_X - 1)| \leq \mathrm{Card}\, R \cdot \mathrm{Card}\, Q + I.$$

If we have $|\,\mathrm{rep}_S(a_X - 1)| \leq \mathrm{Card}\, R \cdot \mathrm{Card}\, Q$, then we are done. Thus, from now on, assume $|\,\mathrm{rep}_S(a_X - 1)| > \mathrm{Card}\, R \cdot \mathrm{Card}\, Q$. By applying the pumping lemma to the product automaton[5] $\mathcal{A} \times \mathcal{B}$, there exist finite words x, y, z over Σ with $y \neq \varepsilon$, such that we have

$$\begin{aligned} \mathrm{rep}_S(a_X - 1) &= xyz; \\ |xy| &\leq \mathrm{Card}\, R \cdot \mathrm{Card}\, Q; \\ \delta(q_0, x) &= \delta(q_0, xy); \\ \lambda(r_0, x) &= \lambda(r_0, xy); \\ \forall n \in \mathbb{N},\ xy^n z \in \mathrm{rep}_S(X) &\Leftrightarrow xyz \in \mathrm{rep}_S(X). \end{aligned} \tag{26}$$

Therefore it suffices to show $|z| \leq I$. Proceed by contradiction and assume that we have $|z| > I$. Using Proposition 1.7.6 on page 18 and the periodicity of the sequences $(\mathbf{u}_q(i) \bmod p_X)_{i\geq 0}$, we obtain

$$\mathrm{val}_S(xy^{p_X P} yz) \equiv \mathrm{val}_S(xyz) \pmod{p_X}. \tag{27}$$

Let us give some extra details on how we derive identity (27). Set $r = |x|$, $s = |y|$, and $t = |z|$. For all positive integers n, using Proposition 1.7.6 for $w = xy^n z$, we get $|w| = r + ns + t$ and

$$\begin{aligned} \mathrm{val}_S(xy^n z) = \sum_{q\in Q} \Bigg(& \sum_{i=0}^{r-1} \beta_{q,i}(w)\mathbf{u}_q(|w| - i - 1) \\ & + \sum_{i=r}^{r+s-1} \beta_{q,i}(w)\mathbf{u}_q(|w| - i - 1) + \cdots + \sum_{i=r+(n-1)s}^{r+ns-1} \beta_{q,i}(w)\mathbf{u}_q(|w| - i - 1) \\ & + \sum_{i=r+ns}^{r+ns+t-1} \beta_{q,i}(w)\mathbf{u}_q(|w| - i - 1) \Bigg), \end{aligned}$$

[5]The product automaton $\mathcal{A} \times \mathcal{B}$ is defined as follows. For any state (q, r) in the set of states $Q \times R$, when reading $a \in \Sigma$, one reaches in $\mathcal{A} \times \mathcal{B}$ the state $(\delta(q, a), \lambda(r, a))$. The initial state is (q_0, r_0) and the set of final states is $F \times G$. Roughly speaking, the product automaton mimics the behavior of both automata $\mathcal{A}$ and $\mathcal{B}$.

where the first (resp. second, third) line corresponds, as explained below, to the contribution of x (resp. y^n, z). From the definition (4) of the coefficients $\beta_{q,i}(w)$ on page 18 for $w = xy^nz$ with $n \in \mathbb{N} \setminus \{0\}$, we obtain

$$\begin{aligned}\beta_{q,r+\ell s+j}(w) &= \mathrm{Card}\{a < w[r+\ell s+j] \mid \delta(q_0, w[0, r+\ell s+j-1]\,a) = q\}\\ &= \mathrm{Card}\{a < y[j] \mid \delta(q_0, xy^\ell(y[0,j-1])a) = q\}\\ &= \mathrm{Card}\{a < y[j] \mid \delta(q_0, x(y[0,j-1])a) = q\}\\ &= \mathrm{Card}\{a < w[r+j] \mid \delta(q_0, w[0, r+j-1]\,a) = q\}\\ &= \beta_{q,r+j}(w)\end{aligned}$$

for all $q \in Q$, all $\ell \in [\![0, n-1]\!]$, and all $j \in [\![0, s-1]\!]$. Consequently, the previous expansion becomes

$$\begin{aligned}\mathrm{val}_S(xy^nz) \quad = \quad & \sum_{q\in Q}\Bigg(\sum_{i=0}^{r-1}\beta_{q,i}(w)\,\mathbf{u}_q(|w|-i-1)\\ &+ \sum_{i=r}^{r+s-1}\beta_{q,i}(w)\sum_{\ell=0}^{n-1}\mathbf{u}_q(|w|-i-\ell s-1)\\ &+ \sum_{i=r+ns}^{r+ns+t-1}\beta_{q,i}(w)\,\mathbf{u}_q(|w|-i-1)\Bigg).\end{aligned}$$

Now, choose $n = p_X P + 1$. Hence we have $|w| = r + p_X Ps + s + t$. For $q \in Q$ and $i \in [\![r, r+s-1]\!]$, we have

$$\sum_{\ell=0}^{n-1}\mathbf{u}_q(|w|-i-\ell s-1) = \mathbf{u}_q(|w|-i-1) + \sum_{\ell=1}^{p_X P}\mathbf{u}_q(|w|-i-\ell s-1)$$

and the second term is congruent to 0 modulo p_X by the definitions of P and I since we are considering $|z| = t > I$. Consequently, we have

$$\begin{aligned}\mathrm{val}_S(xy^nz) \quad \equiv \quad & \sum_{q\in Q}\Bigg(\sum_{i=0}^{r-1}\beta_{q,i}(w)\,\mathbf{u}_q(|w|-i-1)\\ &+ \sum_{i=r}^{r+s-1}\beta_{q,i}(w)\,\mathbf{u}_q(|w|-i-1)\\ &+ \sum_{i=r+ns}^{r+ns+t-1}\beta_{q,i}(w)\,\mathbf{u}_q(|w|-i-1)\Bigg) \pmod{p_X}.\end{aligned}$$

By the same reasoning as previously, observe that, for all $j \in [\![0, t-1]\!]$, we have $\beta_{q,r+ns+j}(xy^nz) = \beta_{q,r+s+j}(xyz)$. Then we easily derive (27).

Now, we make use of the minimality of a_X to obtain a contradiction. Assume $a_X - 1 \in X$, the other case being similar. Therefore $a_X + np_X - 1$ is not in X for all positive integers n. From (26) we obtain $xy^{p_X P}yz \in \mathrm{rep}_S(X)$

but from (27) this word represents a number of the kind $a_X + np_X - 1$ with $n \in \mathbb{N} \setminus \{0\}$, which cannot belong to X. This completes the proof. □

Theorem 3.6.4. *Let $S = (L, \Sigma, <)$ be an abstract numeration system such that all states q of the trim minimal automaton of L satisfy*

$$\lim_{i \to +\infty} \mathbf{u}_q(i) = +\infty$$

and such that we have $\mathbf{u}_L(i) > 0$ for all $i \in \mathbb{N}$. Moreover, assume

$$\lim_{m \to +\infty} N_{\mathbf{v}}(m) = +\infty,$$

with $\mathbf{v} = (\mathbf{v}_L(i))_{i \geq 0}$. Then it is decidable whether or not an S-recognizable set is ultimately periodic.

Proof. This proof is essentially the same as the one of Theorem 3.2.15. Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be the trim minimal automaton of L. Then consider an S-recognizable set $X \subseteq \mathbb{N}$ given by a DFA with d states. The sequence $\mathbf{v} := (\mathbf{v}_L(i))_{i \geq 0}$ ultimately satisfies a linear recurrence relation of the kind (21) having $a_k \neq 0$ as last coefficient. Moreover, it is an increasing sequence, since we have $\mathbf{u}_L(i) > 0$ for all $i \in \mathbb{N}$. Let the prime decomposition of $|a_k|$ be $|a_k| = p_1^{u_1} \cdots p_r^{u_r}$ with $u_1, \ldots, u_r > 0$.

Assume that X is periodic with period

$$p_X = p_1^{v_1} \cdots p_r^{v_r} c$$

with $v_1, \ldots, v_r, c \in \mathbb{N}$ and $\gcd(a_k, c) = 1$. From Proposition 3.6.1 we obtain $N_{\mathbf{v}}(p_X) \leq d \operatorname{Card} Q$. Therefore, by using Remark 3.2.13, we obtain $N_{\mathbf{v}}(c) \leq (d \operatorname{Card} Q)^k$ and $N_{\mathbf{v}}(p_j^{v_j}) \leq (d \operatorname{Card} Q)^k$ for all $j \in [\![1, r]\!]$. We have exactly the same reasoning as in the proof of Theorem 3.2.15 to find upper bounds on c and on the exponents v_j, for $j \in [\![1, r]\!]$. Hence, if X is ultimately periodic, then its period p_X is bounded by a constant P that can effectively be estimated. Now, define $\mathbf{u}_q := (\mathbf{u}_q(i))_{i \geq 0}$ for all $q \in Q$ and $I(p) = \max\{\iota_{\mathbf{u}_q}(p) \mid q \in Q\}$ for all $p \in [\![1, P]\!]$. Then, using Proposition 3.6.3, the admissible preperiods a_X for X to be ultimately periodic must satisfy

$$|\operatorname{rep}_S(a_X - 1)| \leq d \operatorname{Card} Q + \max\{I(p) \mid p \in [\![1, P]\!]\}.$$

Since $m \mapsto |\operatorname{rep}_S(m)|$ is a non-decreasing map and the quantities $I(p)$ can be effectively computed for any positive integer p, the latter observation provides a computable bound on the preperiod a_X of X.

Consequently the sets of admissible periods and preperiods we have to check are finite. Thanks to Theorem 1.7.9 on page 19, we are able to build an automaton for each ultimately periodic set corresponding to a pair of admissible preperiods and periods. Then we only have to compare the accepted languages with $\operatorname{rep}_S(X)$. This finishes the proof. □

Example 3.6.5. Consider the abstract numeration system defined in Example 1.7.5 once again: $S = (\{a, ab\}^* \cup \{c, cd\}^*, \{a, b, c, d\}, a < b < c < d)$. The trim minimal automaton of the numeration language is depicted in Figure 3.1. Remember from Example 1.7.5 that there is no positional numer-

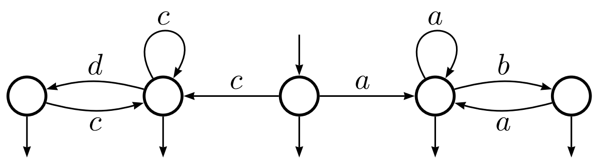

FIGURE 3.1. The trim minimal automaton of $\{a, ab\}^* \cup \{c, cd\}^*$.

ation system which corresponds to this abstract numeration system. Moreover, it was shown in Example 1.5.8 that the numeration language L satisfies $\mathbf{v}_L(i+3) = 2\mathbf{v}_L(i+2) - \mathbf{v}_L(i)$ for all $i \in \mathbb{N}$. Then, using Proposition 3.2.17, it is clear that S satisfies all the assumptions of the previous theorem.

3.7. Connection with the HD0L Periodicity Problem

In this last section we show how Theorem 3.6.4 can be used to decide particular instances of the HD0L periodicity problem. First, let us define what is a HD0L system.

Definition 3.7.1. A D0L system is a triple $G = (\Delta, f, w)$ where

- Δ is an alphabet;
- $f\colon \Delta^* \to \Delta^*$ is a morphism;
- w is a finite word over Δ.

Definition 3.7.2. A HD0L system is a 5-tuple $G = (\Delta, \Gamma, f, g, w)$ where

- (Δ, f, w) is a D0L system;
- Γ is an alphabet;
- $g\colon \Delta^* \to \Gamma^*$ is a morphism.

Let $G = (\Delta, \Gamma, f, g, w)$ be a HD0L system. If w is a prefix of $f(w)$ and if $g(f^\omega(w))$ is an infinite word over Γ, where $f^\omega(w)$ denotes the limit $\lim_{n\to+\infty} f^n(w)$, then we define *the infinite word* (or *ω-word*) *generated by* G to be

$$\omega(G) = g(f^\omega(w)).$$

The question is to decide whether or not the infinite word $\omega(G)$ is ultimately periodic. This problem is called the *HD0L periodicity problem.* From [**HR04**] we know that we may assume that w is a letter. Furthermore, it is well known [**Cob72, Pan83, AS03**] that we can assume that f is a non-erasing morphism and g is a coding, that is, $f(a) \neq \varepsilon$ for all $a \in \Delta$ and $g(\Delta) \subseteq \Gamma$. In [**HR04**] J. Honkala and M. Rigo showed that the HD0L periodicity problem is equivalent to Problem 3. Also, see Theorem 4.1.15 below.

Thanks to [**RM02**], given a HD0L system G, one can canonically build an abstract numeration system $S = (L, \Sigma, <)$ and a DFAO

$$\mathcal{M} = (Q, \Sigma, \delta, q_0, \Gamma, \tau)$$

such that we have

$$\forall n \in \mathbb{N},\ \omega(G)[n] = \tau(\delta(q_0, \mathrm{rep}_S(n))).$$

Such a word is said to be an S-automatic word. This notion will be more precisely studied in the next chapter. From [**Rig00**] we know that the sets

$$X_a = \{n \in \mathbb{N} \mid \omega(G)[n] = a\},\ \text{for } a \in \Gamma,$$

are S-recognizable. Hence, if S satisfies the assumptions of Theorem 3.6.4, then one can decide whether or not these sets X_a are ultimately periodic. Then, observe that the infinite word $\omega(G)$ is ultimately periodic if and only if, for each $a \in \Gamma$, the set of non-negative integers X_a is ultimately periodic. Therefore, if G is such that the associated abstract numeration system S satisfies the assumptions of Theorem 3.6.4, then one can decide whether or not $\omega(G)$ is ultimately periodic.

Example 3.7.3. Consider the HD0L system $G = (\{a, b, c\}, \{a, b\}, f, g, a)$ where the morphisms f and g are defined by

$$f\colon \begin{cases} a & \mapsto & ab \\ b & \mapsto & ac \\ c & \mapsto & b \end{cases} \quad \text{and} \quad g\colon \begin{cases} a & \mapsto & a \\ b & \mapsto & b \\ c & \mapsto & a. \end{cases}$$

The canonically associated DFAO is depicted in Figure 3.2. The details of such a construction are given in Chapter 4; see Definition 4.6.2. Thus this example should probably be read again in the light of the further developments of Chapter 4. Let M be the language accepted by the corresponding

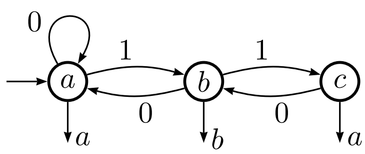

FIGURE 3.2. The DFAO associated with G.

DFA where all states are final. Then the canonically associated abstract numeration system is $S = (L, \{0, 1\}, 0 < 1)$ with

$$L = M \setminus 0\{0, 1\}^* = \{\varepsilon, 1, 10, 11, 100, 101, 110, 1000, 1001, 1010, 1011, \ldots\}.$$

We compute

$$\mathbf{v}_L(0) = 1,\ \mathbf{v}_L(1) = 2,\ \mathbf{v}_L(2) = 4,$$
$$\forall i \in \mathbb{N},\ \mathbf{v}_L(i+3) = \mathbf{v}_L(i+2) + 2\mathbf{v}_L(i+1) - \mathbf{v}_L(i).$$

From Proposition 3.2.17 it follows $\lim_{m\to+\infty} N_{\mathbf{v}}(m) = +\infty$. Hence we easily verify that the hypotheses of Theorem 3.6.4 are all satisfied. Thus we can decide whether or not

$$\omega(G) = g(f^{\omega}(a)) = abaaabbabaaaaabaaabbabbabaa\cdots$$

is ultimately periodic.

CHAPTER 4

Multidimensional S-Automatic Words and Morphisms

4.1. Introduction

In this chapter, in the vein of A. Maes's doctoral dissertation [**Mae99b**], we study the relationship between infinite words generated by finite automata and infinite words generated by morphisms but extended to the framework of multidimensional infinite words. The content of this chapter can be found in [**CKR, CKR09**].

Nowadays, the notion of automatic sequences is widely known. One of the many interests of these sequences, or infinite words, is that they link together automata theory and number theory. Indeed, automatic sequences are a powerful tool to obtain results in number theory using methods from automata theory. The reader interested in these automatic sequences and their generalizations can find more details in [**AS03**].

Definition 4.1.1. Let $b \geq 2$ be an integer. An infinite word x over an alphabet Γ is *b-automatic* if, for all non-negative integers n, its $(n+1)$st letter $x[n]$ is obtained by "feeding" a DFAO $\mathcal{A} = (Q, [\![0, b-1]\!], \delta, q_0, \Gamma, \tau)$ with the U_b-representation of n:

$$\forall n \in \mathbb{N},\ \tau(\delta(q_0, \mathrm{rep}_{U_b}(n))) = x[n].$$

In his seminal paper [**Cob72**] A. Cobham proved the following characterization of b-automatic infinite words. Before stating it, let us recall some basic definitions.

Definition 4.1.2. If μ is a morphism on an alphabet Σ and a is a letter in Σ such that the image $\mu(a)$ begins with a, then we say that μ is *prolongable on a*.

If a morphism μ is prolongable on a letter a, then the limit $\lim_{n\to+\infty} \mu^n(a)$ is well defined. As usual, we denote this limit by $\mu^\omega(a)$. Furthermore, this limit word is a fixed point of μ. Observe that it is an infinite word if and only if there is a letter b occurring in $\mu(a)$ that satisfies $\mu^n(b) \neq \varepsilon$ for all non-negative integers n.

Definition 4.1.3. An infinite word is said to be *pure morphic* if it can be written as $\mu^\omega(a)$ for some morphism μ prolongable on a letter a. It is said to be *morphic* if it the image under a coding, *i.e.*, a letter-to-letter morphism, of some pure morphic word.

Definition 4.1.4. Let Σ and Δ be two alphabets and b be a positive integer. A morphism $\mu\colon \Sigma^* \to \Delta^*$ is said to be *b-uniform* if the images of the letters in Σ are all words of length b.

Theorem 4.1.5. [**Cob72**] *Let $b \geq 2$ be an integer. An infinite word is b-*automatic *if and only if it is the image under a coding of a fixed point of a b-uniform morphism.*

Let us illustrate Cobham's characterization thanks to the following two examples.

Example 4.1.6. Consider the 2-automatic word x generated by the DFAO depicted in Figure 4.1. The beginning of the computation of x is given in the next table.

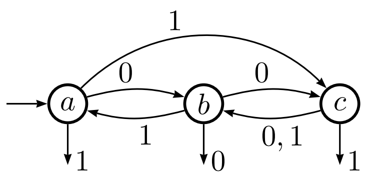

FIGURE 4.1. A DFAO generating x.

n	0	1	2	3	4	5	6	7	8	$\cdots$
$\text{rep}_{U_2}(n)$	ε	1	10	11	100	101	110	111	1000	$\cdots$
$a \cdot \text{rep}_{U_2}(n)$	a	c	b	b	c	a	c	a	b	$\cdots$
x	1	1	0	0	1	1	1	1	0	$\cdots$

We want to build a 2-uniform morphism f prolongable on some letter α and a coding g satisfying $g(f^\omega(\alpha)) = x$. To that end, we observe that the DFAO depicted in Figure 4.2 generates the same 2-automatic word and has a loop of label 0 at the initial state. The images under the morphism f associated with this DFAO are all of length 2 and are defined by $f(\sigma)[i] = \sigma \cdot i$ for $i = 0, 1$. The coding g mimics the output function. Thus we have

$$f\colon \begin{cases} \alpha & \mapsto \alpha c \\ a & \mapsto bc \\ b & \mapsto ca \\ c & \mapsto bb \end{cases} \quad \text{and} \quad g\colon \begin{cases} \alpha & \mapsto 1 \\ a & \mapsto 1 \\ b & \mapsto 0 \\ c & \mapsto 1 \end{cases}.$$

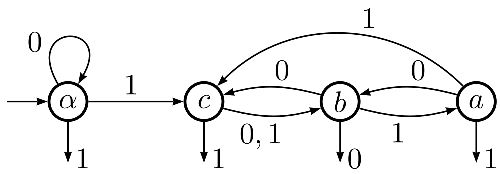

FIGURE 4.2. Another DFAO generating x.

Now, it is easily verified that we have $g(f^\omega(\alpha)) = x$, as desired.

Example 4.1.7. Consider the 2-uniform morphism f prolongable on the letter a and the coding g defined by

$$f\colon \begin{cases} a & \mapsto & ab \\ b & \mapsto & ac \\ c & \mapsto & cb \end{cases} \quad \text{and} \quad g\colon \begin{cases} a & \mapsto & 1 \\ b & \mapsto & 1 \\ c & \mapsto & 0 \end{cases}.$$

We have $g(f^\omega(a)) = 11101101111001101110\cdots$. We associate the DFAO depicted in Figure 4.3 in the following way. The states are the letters a, b, c and a is the initial state. The alphabet is $\{0,1\}$. The output function is g. Transitions are defined by $\sigma \cdot i = f(\sigma)[i]$ for $i = 0, 1$. The 2-automatic word

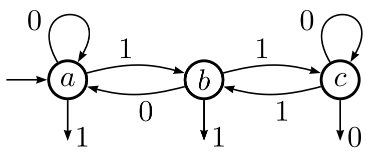

FIGURE 4.3. The DFAO associated with f and g.

generated by this DFAO is $g(f^\omega(a))$; see the following table.

n	0	1	2	3	4	5	6	7	8	$\cdots$
$\text{rep}_{U_2}(n)$	ε	1	10	11	100	101	110	111	1000	$\cdots$
$a \cdot \text{rep}_{U_2}(n)$	a	b	a	c	a	b	c	b	a	$\cdots$
$g(f^\omega(a))$	1	1	1	0	1	1	0	1	1	$\cdots$

An extension of Theorem 4.1.5 obtained by O. Salon deals with a multidimensional setting and an integer base b numeration system. In this framework images of letters under a morphism are hypercubes of constant size b. To work with d-tuples of words of the same length, we introduce the following map.[1]

[1]We already used this map, in an informal way, on page 23 in the introduction of Chapter 2.

Definition 4.1.8. Let d be a positive integer. If $w^{(1)}, \dots, w^{(d)}$ are finite words over an alphabet Σ, the *padding map* $(\cdot)^{\#}\colon (\Sigma^*)^d \to ((\Sigma \cup \{\#\})^d)^*$ is defined by

$$(w^{(1)}, \dots, w^{(d)})^{\#} = (\#^{m-|w^{(1)}|} w^{(1)}, \dots, \#^{m-|w^{(d)}|} w^{(d)}),$$

where we set $m = \max\{|w^{(1)}|, \dots, |w^{(d)}|\}$. In what follows we will use the notation $\Sigma_{\#}$ as a shorthand for the alphabet $\Sigma \cup \{\#\}$.

Example 4.1.9. We have $(ab, bbaa)^{\#} = (\#\#ab, bbaa)$.

Note that, with the same notation as in the previous definition, a d-tuple $(w^{(1)}, \dots, w^{(d)})^{\#}$ in $((\Sigma_{\#})^d)^*$ never begins with $(\#, \dots, \#)$.

Definition 4.1.10. Let d be a positive integer. A *d-dimensional infinite word* x over an alphabet Γ is a map from $\mathbb{N}^d$ to Γ, that is, $x\colon \mathbb{N}^d \to \Gamma$. We use notation such as $x_{n_1,\dots,n_d}$ or $x[n_1, \dots, n_d]$ to denote the value of x at $(n_1, \dots, n_d)$.

The following definition was given in [**Sal87a**] except that the words were read from right to left.

Definition 4.1.11. Let $b \geq 2$ and $d \geq 1$ be integers. A d-dimensional infinite word x over an alphabet Γ is said to be *b-automatic* if there exists a DFAO $\mathcal{A} = (Q, [\![0, b-1]\!]^d, \delta, q_0, \Gamma, \tau)$ such that we have

$$\forall n_1, \dots, n_d \in \mathbb{N},\ \tau(\delta(q_0, (\mathrm{rep}_{U_b}(n_1), \dots, \mathrm{rep}_{U_b}(n_d))^0)) = x_{n_1,\dots,n_d}.$$

Theorem 4.1.12. [**Sal87a, Sal87b**] *Let $b \geq 2$ and $d \geq 1$ be integers. A d-dimensional infinite word is b-automatic if and only if it is the image under a coding of a fixed point of a b-uniform d-dimensional morphism.*

We give an example from [**Sal87a**] illustrating the previous notions.

Example 4.1.13. The *Thue-Morse word* is the fixed point t beginning with 0 of the 2-uniform morphism μ over $\{0, 1\}$ defined by $\mu(0) = 01$ and $\mu(1) = 10$, that is, $t = \mu^\omega(0) = 0110100110010110\cdots$. This word t is 2-automatic from Theorem 4.1.5. Now, define a bidimensional infinite word u by $u_{m,n} = t[m + n]$ for all $m, n \in \mathbb{N}$. This word is depicted in Figure 4.4. First, we determine a 2-uniform bidimensional morphism and a coding which generate u and second, we exhibit an DFAO reading pairs of letters which produces u. Let Σ denote the alphabet $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$. It is clear that u is generated by the morphism f over the alphabet Σ prolongable

	0	1	2	3	4	5	6	7	8	⋯
0	0	1	1	0	1	0	0	1	1	⋯
1	1	1	0	1	0	0	1	1	0	
2	1	0	1	0	0	1	1	0	0	
3	0	1	0	0	1	1	0	0	1	
4	1	0	0	1	1	0	0	1	0	
5	0	0	1	1	0	0	1	0	1	
6	0	1	1	0	0	1	0	1	1	
7	1	1	0	0	1	0	1	1	0	
8	1	0	0	1	0	1	1	0	1	
⋮	⋮									⋱

FIGURE 4.4. A bidimensional word built on the Thue-Morse word.

on $(0,1)$ defined by

$$(0,1) \mapsto \begin{array}{|c|c|}\hline (0,1) & (1,1) \\ \hline (1,1) & (1,0) \\ \hline \end{array}, \quad (1,1) \mapsto \begin{array}{|c|c|}\hline (1,0) & (0,1) \\ \hline (0,1) & (1,0) \\ \hline \end{array},$$

$$(1,0) \mapsto \begin{array}{|c|c|}\hline (1,0) & (0,0) \\ \hline (0,0) & (0,1) \\ \hline \end{array}, \quad (0,0) \mapsto \begin{array}{|c|c|}\hline (0,1) & (1,0) \\ \hline (1,0) & (0,1) \\ \hline \end{array},$$

and the coding g over the alphabet Σ defined by

$$g\colon \left\{ \begin{array}{lcl} (0,1),(0,0) & \mapsto & 0 \\ (1,1),(1,0) & \mapsto & 1 \end{array} \right. .$$

Indeed, in u, each row corresponds to the shifted previous row. Therefore, in the fixed point $f^\omega((0,1))$, the second components of the pairs occurring in a row anticipate the first components occurring in the next row. By applying the coding g, we then obtain $g(f^\omega((0,1))) = u$. Next, we easily check that the DFAO over Σ depicted in Figure 4.5 generates u as a 2-automatic bidimensional infinite word. For instance, the letter 1 at position $(3,5)$ is computed as

$$A \cdot (11,101)^0 = A \cdot (011,101) = B \cdot (11,01) = A \cdot (1,1) = D$$

and the output corresponding to the state D is 1.

Another possible extension of Theorem 4.1.5 in the unidimensional setting is to relax the terms of the hypothesis on the uniformity of the morphism. In this case, Cobham's result still holds but integer base numeration systems are replaced by abstract numeration systems.

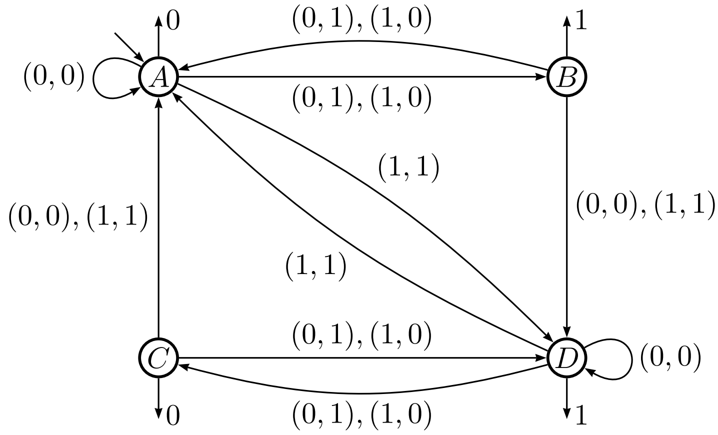

FIGURE 4.5. A DFAO generating u.

Definition 4.1.14. Let S be an abstract numeration system. An infinite word x over an alphabet Γ is *S-automatic* if, for all non-negative integers n, its $(n+1)$st letter $x[n]$ is obtained by "feeding" a DFAO $\mathcal{A} = (Q, \Sigma, \delta, q_0, \Gamma, \tau)$ with the S-representation of n:

$$\forall n \in \mathbb{N},\ \tau(\delta(q_0, \mathrm{rep}_S(n))) = x[n].$$

Theorem 4.1.15. [**Rig00, RM02**] *An infinite word is S-automatic for some abstract numeration system S if and only if it is morphic.*

From Remark 1.7.14 on page 20 we know that the set of squares is S-recognizable for some abstract numeration system S. This implies that its characteristic word is S-automatic, as shown by the following example.

Example 4.1.16. Consider once again the abstract numeration system

$$S = (a^*b^* \cup a^*c^*, \{a, b, c\}, a < b < c)$$

of Remark 1.7.14. Since the set $X = \{n^2 \mid n \in \mathbb{N}\}$ of squares is S-recognizable, its characteristic word, *i.e.*, the infinite word

$$\chi_X = 11001000010000001000\cdots$$

over $\{0, 1\}$ defined by $\chi_X[n] = 1$ for $n \in X$ and $\chi_X[n] = 0$ for $n \notin X$, is S-automatic. Indeed, it is computed by the DFAO depicted in Figure 4.6. This DFAO is built on the trim minimal automaton of the numeration language. Since we have $\mathrm{rep}_S(X) = a^*$, the output function τ is simply defined by $\tau(a) = 1$ and $\tau(b) = \tau(c) = 0$. This word χ_X is also morphic because it is generated by the morphism f iterated from α and the coding g that are

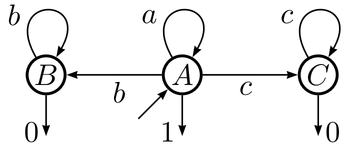

FIGURE 4.6. A DFAO generating the squares.

defined by

$$f\colon \begin{cases} \alpha & \mapsto & \alpha ABB \\ A & \mapsto & ABB \\ B & \mapsto & B \end{cases} \text{ and } g\colon \begin{cases} \alpha & \mapsto & 1 \\ A & \mapsto & 1 \\ B & \mapsto & 0 \end{cases}.$$

Note that this morphism f is not directly associated with the DFAO of Figure 4.6. Actually, to associate a morphism with this DFAO, we have to add a sink state D to make a complete graph. We then obtain the morphism r over the 4-letter alphabet $\{A, B, C, D\}$ defined by

$$r\colon \beta \mapsto \beta ABC,\ A \mapsto ABC,\ B \mapsto DBD,\ C \mapsto DDC,\ D \mapsto DDD.$$

Next, we have to erase all occurrences of the letter D (which correspond to words not belonging to the numeration language) in the infinite fixed point $r^\omega(\alpha)$. This process does not give a real coding since it produces a morphism s which maps D to the empty word:

$$s\colon \beta \mapsto 1,\ A \mapsto 1,\ B \mapsto 0,\ C \mapsto 0,\ D \mapsto \varepsilon.$$

However, the generated word $s(r^\omega(\beta))$ is morphic (for instance, see [**AS03**, Theorem 7.7.4]), *i.e.*, we can build a morphism μ prolongable on some letter γ and a coding ν generating the same word:

$$\nu(\mu^\omega(\gamma)) = s(r^\omega(\beta)) = g(f^\omega(\alpha)) = \chi_X.$$

The mentioned result from [**AS03**] will be illustrated in Example 4.5.2 below.

The aim of this chapter is to extend Theorem 4.1.15 to a multidimensional setting. The notion of an S-automatic multidimensional word is simple to define. However, difficulties occur when one has to iterate a morphism under which the images of the letters are no longer hypercubes of constant size.

It is worth mentioning the work of [**ABS04**], where a different notion of bidimensional morphisms is introduced in connection with problems arising in discrete geometry. Also, see [**Pey87**] for questions related to frequencies of letters in bidimensional automatic words and [**NR07**] for their generalization to the S-automatic case. In [**DFNR**] bidimensional S-automatic sequences turn out to be useful in the context of combinatorial game theory. Indeed, they play a central role in finding new characterizations of P-positions for the famous Wythoff game and some of its variations. Another motivation for studying the set of multidimensional S-automatic words w over $\{0,1\}$ is to consider them to be characteristic words of subsets P_w of $\mathbb{N}^d$, to extend the Presburger structure $\langle \mathbb{N}; < \rangle$ by the corresponding predicates P_w and to study the decidability of the corresponding first-order theory. In addition, see [**CT02**] for the relationship with second-order monadic theory.

The organization of this chapter is given below. Our main result — Theorem 4.6.1— can be precisely stated as follows.

Theorem. *Let $d \geq 1$ be an integer. The d-dimensional infinite word x is S-automatic for some abstract numeration system $S = (L, \Sigma, <)$ with $\varepsilon \in L$ if and only if it is the image under a coding of a shape-symmetric pure morphic word.*

Our first task is to present the different concepts occurring in this statement. In Sections 4.2 and 4.3 we define multidimensional S-automatic words and multidimensional morphisms respectively. Then, in Section 4.4, we define the notion of a shape-symmetric word as originally introduced by A. Maes.

In particular, to prove our main result, we have to handle erasing morphisms in a multidimensional setting. To that aim, we need to generalize the well-known result (see [**AS03**, Theorem 7.7.4]) that states that a word obtained by erasing from a morphic word over an alphabet Σ all occurrences of the letters in a subset of Σ is either finite or morphic. In Section 4.5, first, we give an example illustrating this result in the unidimensional case. Second, we define how to erase a hyperplane from a d-dimensional array.

Finally, in Section 4.6, we prove our main result. Throughout the proof, we strive to clarify the concepts under consideration with a number of examples.

4.2. Multidimensional S-Automatic Words

The following notion was introduced in [**RM02**] as a natural generalization of the multidimensional b-automatic sequences introduced in [**Sal87a, Sal87b**]; see Definition 4.1.11 above.

Definition 4.2.1. Let d be a positive integer. A d-dimensional infinite word x over an alphabet Γ is said to be *S-automatic* for some abstract numeration system $S = (L, \Sigma, <)$ if there exists a DFAO $\mathcal{A} = (Q, (\Sigma_{\#})^d, \delta, q_0, \Gamma, \tau)$ such that we have

$$\forall n_1, \ldots, n_d \in \mathbb{N},\ \tau(\delta(q_0, (\mathrm{rep}_S(n_1), \ldots, \mathrm{rep}_S(n_d))^{\#})) = x_{n_1,\ldots,n_d}.$$

Example 4.2.2. Consider the abstract numeration system

$$S = (\{a, ba\}^*\{\varepsilon, b\}, \{a, b\}, a < b)$$

and the DFAO depicted in Figure 4.7. The first few words in the numeration

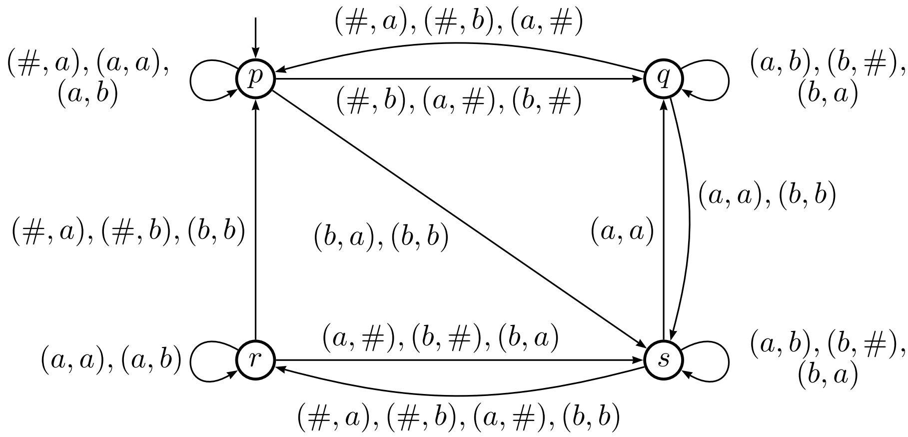

FIGURE 4.7. A deterministic finite automaton with output.

language enumerated with respect to the genealogical order are

$$\varepsilon, a, b, aa, ab, ba, aaa, aab, aba, baa, bab, aaaa.$$

We do not consider the transitions on input $(\#, \#)$ since this automaton is fed with entries of the form $(\text{rep}_S(n_1), \text{rep}_S(n_2))^{\#}$. If the outputs of the DFAO are considered to be the states themselves, then we produce the bidimensional S-automatic word given in Figure 4.8.

	ε	a	b	aa	ab	ba	aaa	aab	$\cdots$
ε	p	p	q	p	q	p	p	q	$\cdots$
a	q	p	p	p	p	s	p	p	
b	q	s	s	s	s	q	s	s	
aa	p	s	q	p	p	p	p	p	
ab	q	q	s	s	s	s	s	s	
ba	p	s	q	q	s	q	q	s	
aaa	q	p	p	q	s	s	p	p	
aab	q	s	s	s	r	q	s	s	
$\vdots$	$\vdots$								$\ddots$

FIGURE 4.8. A bidimensional S-automatic word.

4.3. Multidimensional Morphisms

Let d be a positive integer fixed throughout this section.

Notation. For all d-tuples $\mathbf{n}$ in $\mathbb{N}^d$ and all $i \in [\![1, d]\!]$, we let n_i denote the ith component of $\mathbf{n}$ and $\mathbf{n}_{\hat{i}}$ denote the $(d-1)$-tuple $(n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_d)$ in $\mathbb{N}^{d-1}$. Let $\mathbf{m}$ and $\mathbf{n}$ be two d-tuples in $\mathbb{N}^d$. We write $\mathbf{m} \leq \mathbf{n}$ (resp. $\mathbf{m} < \mathbf{n}$) if we have $m_i \leq n_i$ (resp. $m_i < n_i$) for all $i \in [\![1, d]\!]$. In particular, we set $\mathbf{0} = (0, \ldots, 0)$ and $\mathbf{1} = (1, \ldots, 1)$.

Definition 4.3.1. A *d-dimensional array* over an alphabet Σ is a map $x\colon [\![0, s_1 - 1]\!] \times \cdots \times [\![0, s_d - 1]\!] \to \Sigma$, where $s_1, \ldots, s_d$ are positive integers or $+\infty$. As a rule, we set $[\![0, s-1]\!] = \mathbb{N}$ for $s = +\infty$. Let x be such a d-dimensional array. If the d-tuple $\mathbf{n} = (n_1, \ldots, n_d)$ belongs to the domain of x, then we use the notation $x_{n_1,\ldots,n_d}$, $x_{\mathbf{n}}$, $x[n_1, \ldots, n_d]$, or $x[\mathbf{n}]$ indifferently. The *shape* of x, denoted by $|x|$, is the d-tuple $\mathbf{s} = (s_1, \ldots, s_d)$ in $(\mathbb{N} \cup \{+\infty\})^d$. We extend the definition of d-dimensional arrays to *empty* d-dimensional arrays, which we will denote by $\varepsilon_{\mathbf{s}}$, *i.e.*, to d-dimensional arrays of shapes $\mathbf{s} \in \mathbb{N}^d$ having at least a zero component. If, for all $i \in [\![1, d]\!]$, we have $|x|_i < +\infty$, then x is said to be *bounded*. We let $B_d(\Sigma)$ denote the set of d-dimensional bounded arrays over Σ. If x is bounded of shape $|x| = (c, c, \ldots, c)$ for some $c \in \mathbb{N}$, then it is said to be a *square* of *size* c.

A d-dimensional word is a d-dimensional array of shape $(+\infty, \ldots, +\infty)$, that is, infinite in all directions.

Definition 4.3.2. Let x be a d-dimensional array. If $\mathbf{s}, \mathbf{t}$ are d-tuples in $\mathbb{N}^d$ satisfying $\mathbf{0} \leq \mathbf{s} \leq \mathbf{t} \leq |x| - \mathbf{1}$, then $x[\mathbf{s}, \mathbf{t}]$ is said to be a *factor* of x and is defined to be the array y of shape $\mathbf{t} - \mathbf{s} + \mathbf{1}$ given by $y[\mathbf{n}] = x[\mathbf{n} + \mathbf{s}]$ for all $\mathbf{n} \in \mathbb{N}^d$ satisfying $\mathbf{n} \leq \mathbf{t} - \mathbf{s}$. For any $\mathbf{u} \in \mathbb{N}^d$, we let $\mathrm{Fact}_{\mathbf{u}}(x)$ denote the set of factors of x of shape $\mathbf{u}$.

Note that, for $d = 1$, the notation $|x|$ (resp. $x[n]$, $x[m, n]$ for $n, m \in \mathbb{N}$ with $n \leq m$) is compatible with that used to denote the length (resp. the $(n+1)$st letter, the factor $x[m] \cdots x[n]$) of a (unidimensional) word. Also note that, in the unidimensional setting, there is only one empty word $\varepsilon_{\mathbf{0}} = \varepsilon_0 = \varepsilon$, since, of course, there is only one component that can vanish.

Example 4.3.3. Consider the bidimensional bounded array

$$x = \begin{array}{|c|c|c|c|c|} \hline a & b & a & a & b \\ \hline c & d & b & c & d \\ \hline \end{array}$$

of shape $(2, 5)$. We have

$$x[(0,0),(1,1)] = \begin{array}{|c|c|}\hline a & b \\ \hline c & d \\ \hline \end{array} \quad \text{and} \quad x[(0,2),(1,4)] = \begin{array}{|c|c|c|}\hline a & a & b \\ \hline b & c & d \\ \hline \end{array}.$$

For instance, we have $\mathrm{Fact}_{\mathbf{1}}(x) = \{a, b, c, d\}$ and

$$\mathrm{Fact}_{(2,3)}(x) = \left\{ \begin{array}{|c|c|c|}\hline a & b & a \\ \hline c & d & b \\ \hline \end{array}, \begin{array}{|c|c|c|}\hline b & a & a \\ \hline d & b & c \\ \hline \end{array}, \begin{array}{|c|c|c|}\hline a & a & b \\ \hline b & c & d \\ \hline \end{array} \right\}.$$

Now, we would like to concatenate arrays. Note that this is not always possible. To do this safely, we need the following definition.

Definition 4.3.4. Let x, y be two d-dimensional arrays. If we have

$$|x|_{\hat{i}} = |y|_{\hat{i}} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_d) > \mathbf{0}$$

for some $i \in [\![1, d]\!]$, then we define the *concatenation* of x and y *in the direction* i to be the d-dimensional array $x \odot^i y$ of shape

$$(s_1, \dots, s_{i-1}, |x|_i + |y|_i, s_{i+1}, \dots, s_d)$$

satisfying

$$x = (x \odot^i y)[\mathbf{0}, |x| - \mathbf{1}];$$
$$y = (x \odot^i y)[(0, \dots, 0, |x|_i, 0, \dots, 0), (0, \dots, 0, |x|_i, 0, \dots, 0) + |y| - \mathbf{1}].$$

Let $\varepsilon_{\mathbf{s}}$ be a d-dimensional empty word of shape

$$\mathbf{s} = (s_1, \dots, s_{i-1}, 0, s_{i+1}, \dots, s_d) \geq \mathbf{0}$$

having a zero component at the ith position for some $i \in [\![1, d]\!]$. We extend the definition to the concatenation of $\varepsilon_{\mathbf{s}}$ and any d-dimensional word x of shape

$$(s_1, \dots, s_{i-1}, s_i, s_{i+1}, \dots, s_d), \quad \text{for } s_i \in \mathbb{N},$$

in the direction i by

$$\varepsilon_{\mathbf{s}} \odot^i x = x \odot^i \varepsilon_{\mathbf{s}} = x.$$

In particular, we have $\varepsilon_{\mathbf{s}} \odot^i \varepsilon_{\mathbf{s}} = \varepsilon_{\mathbf{s}}$.

Example 4.3.5. Consider the two bidimensional arrays

$$x = \begin{array}{|c|c|}\hline a & b \\ \hline c & d \\ \hline \end{array} \text{ and } y = \begin{array}{|c|c|c|}\hline a & a & b \\ \hline b & c & d \\ \hline \end{array}$$

of shape $|x| = (2, 2)$ and $|y| = (2, 3)$ respectively. Since $|x|_{\hat{2}} = |y|_{\hat{2}} = 2$, we obtain

$$x \odot^2 y = \begin{array}{|c|c|c|c|c|}\hline a & b & a & a & b \\ \hline c & d & b & c & d \\ \hline \end{array}.$$

However, $x \odot^1 y$ is not defined because we have $2 = |x|_{\hat{1}} \neq |y|_{\hat{1}} = 3$.

Let x be a d-dimensional array over an alphabet Σ and $\mu\colon \Sigma \to B_d(\Sigma)$ be a map. Note that $\mu(x)$ is not always well-defined. Depending on the shapes of the images by μ of the letters in Σ, when trying to build $\mu(x)$ by concatenating the images $\mu(x_{\mathbf{n}})$, we can obtain "holes" or "overlaps". Thus μ cannot necessarily be iterated. Therefore, in order to define morphisms, we must introduce some restrictions on μ. First, we need two quite technical definitions.

Definition 4.3.6. Let x be a d-dimensional array of shape $\mathbf{s} = (s_1, \ldots, s_d)$. For all $i \in [\![1, d]\!]$ and $k \in [\![0, s_i - 1]\!]$, we let $x_{|i,k}$ denote the $(d-1)$-dimensional array of shape

$$|x|_{\widehat{i}} = \mathbf{s}_{\widehat{i}} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_d)$$

defined by setting the ith coordinate equal to k in x, that is,

$$x_{|i,k}[n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_d] = x[n_1, \ldots, n_{i-1}, k, n_{i+1}, \ldots, n_d]$$

for all $n_j \in [\![0, s_j - 1]\!]$ with $j \in [\![1, d]\!] \setminus \{i\}$.

Definition 4.3.7. Let $\mu\colon \Sigma \to B_d(\Sigma)$ be a map and x be a d-dimensional array which satisfies

$$\forall i \in [\![1, d]\!],\ \forall k \in [\![0, |x|_i - 1]\!],\ \forall a, b \in \mathrm{Fact}_{\mathbf{1}}(x_{|i,k}),\ |\mu(a)|_i = |\mu(b)|_i. \quad (28)$$

Then $\mu(x)$ is the d-dimensional array defined by

$$\mu(x) = \odot^1_{0 \le n_1 < |x|_1} \left(\cdots \left(\odot^d_{0 \le n_d < |x|_d} \mu(x_{n_1, \ldots, n_d}) \right) \cdots \right).$$

Note that the order of the products in the different directions is unimportant.

Example 4.3.8. Consider the map μ given by

$$a \mapsto \begin{array}{|c|c|}\hline a & a \\ \hline b & d \\ \hline\end{array},\ b \mapsto \begin{array}{|c|}\hline c \\ \hline b \\ \hline\end{array},\ c \mapsto \begin{array}{|c|c|}\hline a & a \\ \hline\end{array},\ d \mapsto \begin{array}{|c|}\hline d \\ \hline\end{array}.$$

Take

$$x = \begin{array}{|c|c|}\hline a & b \\ \hline c & d \\ \hline\end{array}.$$

We have $|\mu(a)|_1 = |\mu(b)|_1 = 2$, $|\mu(c)|_1 = |\mu(d)|_1 = 1$, $|\mu(a)|_2 = |\mu(c)|_2 = 2$, and $|\mu(b)|_2 = |\mu(d)|_2 = 1$. Thus $\mu(x)$ is well-defined and given by

$$\mu(x) = \begin{array}{|c|c|c|}\hline a & a & c \\ \hline b & d & b \\ \hline a & a & d \\ \hline\end{array}.$$

However, $\mu^2(x)$ is not well-defined.

Now, we are ready to introduce the definition of a d-dimensional morphism.

Definition 4.3.9. Let $\mu\colon \Sigma \to B_d(\Sigma)$ be a map. If, for all $a \in \Sigma$ and all positive integer n, $\mu^n(a)$ is inductively well-defined from $\mu^{n-1}(a)$, *i.e.*, if $\mu^{n-1}(a)$ satisfies (28) for all positive integer n, then μ is said to be a *d-dimensional morphism*.

Remark 4.3.10. A. Maes showed that determining whether or not a map $\mu\colon \Sigma \to B_d(\Sigma)$ is a d-dimensional morphism is a decidable problem [**Mae98, Mae99a, Mae99b**].

The usual notion of a prolongable morphism can also be given in this multidimensional setting; see Definition 4.1.2.

Definition 4.3.11. If μ is a d-dimensional morphism and a is a letter such that we have $(\mu(a))_{\mathbf{0}} = a$, then we say that μ is *prolongable on a*.

Observe that, if μ is a d-dimensional morphism prolongable on the letter a, the limit $\mu^\omega(a) = \lim_{n\to+\infty} \mu^n(a)$ is well defined and it is a fixed point of μ beginning with a. It is a d-dimensional infinite word if and only if we have

$$\forall i \in [\![1, d]\!],\ \exists b \in \mathrm{Fact}_{\mathbf{1}}(\mu(a)),\ \forall n \in \mathbb{N},\ |\mu^n(b)|_i \neq 0.$$

In this case, it is the only fixed point of μ beginning with a. Also, note that, if a d-dimensional infinite word x is a fixed point of a d-dimensional morphism μ, then (28) implies

$$\forall i \in [\![1, d]\!],\ \forall k \in \mathbb{N},\ \forall a, b \in \mathrm{Fact}_{\mathbf{1}}(x_{|i,k}),\ |\mu(a)|_i = |\mu(b)|_i.$$

Definition 4.3.12. A d-dimensional infinite word x is said to be *pure morphic* if it can be written as $x = \mu^\omega(a)$ for a d-dimensional morphism μ prolongable on the letter a. It is said to be *morphic* if it is the image under a coding of a pure morphic word.

4.4. Shape-Symmetric Morphic Words

The so-called property of shape-symmetry that we introduce now is a natural generalization of uniform morphisms under which all the images of letters are squares of the same size [**Sal87a, Sal87b**]. This property was first introduced by A. Maes, and was mainly used in connection with logical questions about the decidability of first-order theories where $\langle \mathbb{N}; < \rangle$ is extended by some morphic predicate [**Mae98, Mae99a, Mae99b**]. Again, we fix an integer $d \geq 2$ for the whole section.

Definition 4.4.1. Let $\mu\colon \Sigma \to B_d(\Sigma)$ be a d-dimensional morphism having the infinite word x as a fixed point. If the images $\mu(x_{n,\ldots,n})$, for $n \in \mathbb{N}$, of the letters on the diagonal of x are all squares, then x is said to be *shape-symmetric* (with respect to μ).

Remark 4.4.2. Two equivalent formulations of shape-symmetry are given as follows. Let $\mu\colon \Sigma \to B_d(\Sigma)$ be a d-dimensional morphism having the d-dimensional infinite word x as a fixed point. This word is shape-symmetric if and only if we have

$$\forall i, j \in [\![1, d]\!],\ \forall k \in \mathbb{N},\ \forall a \in \mathrm{Fact}_{\mathbf{1}}(x_{|i,k}),\ \forall b \in \mathrm{Fact}_{\mathbf{1}}(x_{|j,k}),\ |\mu(a)|_i = |\mu(b)|_j,$$

or, if and only if, for any permutation f of $[\![1, d]\!]$, we have

$$\begin{aligned}\forall n_1, \ldots, n_d \in \mathbb{N},\ |\mu(x_{n_1,\ldots,n_d})| &= (s_1, \ldots, s_d)\\ &\Rightarrow\ |\mu(x_{n_{f(1)},\ldots,n_{f(d)}})| = (s_{f(1)}, \ldots, s_{f(d)}).\end{aligned}$$

Example 4.4.3. The following bidimensional morphism μ has a fixed point $\mu^\omega(a)$ which is shape-symmetric:

$$\mu(a) = \mu(f) = \begin{array}{|c|c|}\hline a & b\\\hline c & d\\\hline\end{array},\ \mu(b) = \begin{array}{|c|}\hline e\\\hline c\\\hline\end{array},\ \mu(c) = \begin{array}{|c|c|}\hline e & b\\\hline\end{array},\ \mu(d) = \begin{array}{|c|}\hline f\\\hline\end{array},$$

$$\mu(e) = \begin{array}{|c|c|}\hline e & b\\\hline g & d\\\hline\end{array},\ \mu(g) = \begin{array}{|c|c|}\hline h & b\\\hline\end{array},\ \mu(h) = \begin{array}{|c|c|}\hline h & b\\\hline c & d\\\hline\end{array}.$$

In Figure 4.9 we have represented the beginning of the array. Some elements are underlined for the use of Example 4.6.10.

Definition 4.4.4. Let $\mu\colon \Sigma \to B_d(\Sigma)$ be a d-dimensional morphism having the d-dimensional infinite word x as a fixed point. The *shape sequence* of x with respect to μ *in the direction* $i \in [\![1, d]\!]$ is the sequence

$$\mathrm{Shape}_{\mu,i}(x) = (|\mu(x_{|i,k})|_i)_{k\geq 0}.$$

For a unidimensional morphism μ having the infinite word x as a fixed point, the *shape sequence* of x with respect to μ is $\mathrm{Shape}_\mu(x) = (|\mu(x[k])|)_{k\geq 0}$.

Remark 4.4.5. Let $\mu\colon \Sigma \to B_d(\Sigma)$ be a d-dimensional morphism having the d-dimensional infinite word x as a fixed point. Note that x is shape-symmetric if and only if we have $\mathrm{Shape}_{\mu,1}(x) = \cdots = \mathrm{Shape}_{\mu,d}(x)$.

	ε	1	10	100	101	1000	1001	1010	10000	10001	$\cdots$
ε	$\underline{a}$	$\underline{b}$	e	e	b	e	b	e	e	b	$\cdots$
1	c	d	$\underline{c}$	g	d	g	d	c	g	d	
10	e	b	f	e	$\underline{b}$	h	b	f	h	b	
100	e	b	e	a	b	e	b	e	h	b	
101	g	d	c	c	d	g	d	$\underline{c}$	c	d	
1000	e	b	e	e	b	a	b	e	e	b	
1001	g	d	c	g	d	c	d	c	g	d	
1010	h	b	f	e	b	e	b	f	h	b	
10000	e	b	e	e	b	e	b	e	a	b	
10001	g	d	c	g	d	g	d	c	c	d	
$\vdots$	$\vdots$										$\ddots$

FIGURE 4.9. A infinite fixed point of μ.

4.5. Erasing Hyperplanes from Multidimensional Arrays

To prove our main result, we will have to generalize the following well-known theorem to the multidimensional setting. A. Cobham already foretold this result. Then, J.-J. Pansiot gave a proof in [**Pan83**]. In their book [**AS03**] J.-P. Allouche and J. Shallit proposed another proof, which follows Cobham's original ideas.

Theorem 4.5.1. *If x is a morphic word over an alphabet Σ, then the word obtained by erasing from x all occurrences of the letters in a subset of Σ is either finite or morphic.*

In this section we give an example illustrating the proof from [**AS03**] in the unidimensional case. Then, we define how to erase hyperplanes from a multidimensional array. It will essentially be used in the second part of the proof of Theorem 4.6.1.

Example 4.5.2. Define a morphism f over the alphabet $\Sigma = \{a, b, c, d, e\}$ by

$$f\colon a \mapsto abce,\ b \mapsto aded,\ c \mapsto c,\ d \mapsto cc,\ e \mapsto aeb.$$

We have $f^\omega(a) = abceadedcaebabceccaebcccabceaebaded\cdots$. Set $\Gamma = \{c, e\}$ and let us show that the word $y = abaddababababadd\cdots$ obtained by erasing from $f^\omega(a)$ the letters in Γ is still morphic. First, observe that the letter c is *dead* with respect to the morphism f and the subset Γ, *i.e.*, it satisfies $f^n(c) \in \Gamma^*$ for all $n \in \mathbb{N}$. Hence, by erasing this letter from the images by f of any other letters, we obtain a morphism

$$g\colon a \mapsto abe,\ b \mapsto aded,\ d \mapsto \varepsilon,\ e \mapsto aeb$$

over the alphabet $\Sigma \setminus \{c\} = \{a, b, d, e\}$ such that the infinite word obtained by erasing from $g^\omega(a)$ all occurrences of the letters in $\Lambda = \Gamma \setminus \{c\} = \{e\}$ is also y. Next, the letter d is *moribund* with respect to the morphism g and the subset Λ, *i.e.*, there exists a non-negative integer N such that $g^N(d)$ contains at least one letter not belonging to Λ and we have $g^n(d) \in \Lambda^*$ for all integers $n > N$. Here, we can take $N = 0$. Now, we factor the images of the other letters, which are called *robust* letters, so that we obtain factors having only one letter in $\Sigma \setminus \Lambda = \{a, b, d\}$:

$$g(a) = \underbrace{a}_{w(a,1)} \cdot \underbrace{be}_{w(a,2)},\ g(b) = \underbrace{a}_{w(b,1)} \cdot \underbrace{de}_{w(b,2)} \cdot \underbrace{d}_{w(b,3)},\ g(e) = \underbrace{ae}_{w(e,1)} \cdot \underbrace{b}_{w(e,2)}.$$

With each of these factors $w(\sigma, i)$, we associated a new symbol σ_i, so that we obtain the new alphabet $\Delta = \{a_1, a_2, b_1, b_2, b_3, e_1, e_2\}$. Now, define a morphism μ over this alphabet by

$$\mu\colon a_1, b_1 \mapsto a_1a_2,\ a_2 \mapsto b_1b_2b_3e_1e_2,\ b_2 \mapsto e_1e_2,\ b_3 \mapsto \varepsilon,$$
$$e_1 \mapsto a_1a_2e_1e_2,\ e_2 \mapsto b_1b_2b_3.$$

The image of e_1 is $a_1a_2e_1e_2$ because the corresponding factor in $g(e)$ is $w(e, 1) = ae$ and we have $g(ae) = g(a) \cdot g(e) = a \cdot be \cdot ae \cdot b$, where the factor $a = w(a, 1)$ (resp. $be = w(a, 2)$, $ae = w(e, 1)$, $b = w(e, 2)$) corresponds to a_1 (resp. a_2, e_1, e_2). The image of b_3 is ε because the corresponding factor in $g(b)$ is $w(b, 3) = d$, which contains no robust letter. The other images are computed by using the same reasoning. Next, we define the coding ν over Δ by

$$\nu\colon a_1, b_1, e_1 \mapsto a,\ a_2, e_2 \mapsto b,\ b_2, b_3 \mapsto d.$$

The image by ν of a letter $\sigma_i \in \Delta$ is the letter in $\Sigma \setminus \Lambda = \{a, b, d\}$ occurring in the corresponding factor $w(\sigma, i)$. From these computations we have obtained a morphism μ prolongable on the letter a_1 and a coding ν satisfying $\nu(\mu^\omega(a_1) = y$. In particular, this shows that y is a morphic word.

Definition 4.5.3. Let $d \geq 2$ be an integer and x be a d-dimensional array over $\Sigma \cup \{e\}$ of shape $(s_1, s_2, \ldots, s_d) > \mathbf{0}$, where e does not belong to Σ. For any $i \in [\![1, d]\!]$ and $k \in [\![0, s_i - 1]\!]$, the $(d-1)$-dimensional array $x_{|i,k}$ is called an *e-hyperplane* of x if each letter in $x_{|i,k}$ is equal to e. Then *erasing* an e-hyperplane $x_{|i,k}$ of x means replacing x with a d-dimensional array $x' = y \odot^i z$ with

$$y = \begin{cases} x[\mathbf{0}, (s_1, \ldots, s_{i-1}, k, s_{i+1}, \ldots, s_d) - \mathbf{1}], & \text{if } k \geq 1; \\ \varepsilon_{(s_1,\ldots,s_{i-1},0,s_{i+1},\ldots,s_d)}, & \text{otherwise} \end{cases}$$

and

$$z = \begin{cases} x[(0,\ldots,0,k+1,0,\ldots,0),|x|-\mathbf{1}], & \text{if } k < s_i - 1; \\ \varepsilon_{(s_1,\ldots,s_{i-1},0,s_{i+1},\ldots,s_d)}, & \text{otherwise.} \end{cases}$$

We let ρ_e denote the map which associates with any d-dimensional array x over $\Sigma \cup \{e\}$ the array $\rho_e(x)$ obtained by erasing every e-hyperplane of x iteratively. Furthermore, we say that x is *e-erasable* if the array $\rho_e(x)$ does not contain the letter e as a factor.

Observe that a d-dimensional array x is e-erasable if, for each position $\mathbf{n}$ such that we have $x[\mathbf{n}] = e$, there exists an integer $i \in [\![1,d]\!]$ such that $x_{|i,n_i}$ is an e-hyperplane.

Example 4.5.4. Consider the bidimensional array

$x =$

a	b	a	e	e	a
e	e	e	e	e	e
a	a	e	e	e	b
e	e	a	e	b	b
b	a	b	e	e	a

of shape $(5,6)$. Clearly, $x_{|2,3}$ is an e-hyperplane. By erasing $x_{|2,3}$ from x, we obtain the bidimensional array $x' = y \odot^2 z$ of shape $(5,5)$, with $y = x[(0,0),(4,2)]$ and $z = x[(0,4),(4,5)]$. Then $x'_{|1,1}$ is an e-hyperplane of x'. By erasing $x'_{|1,1}$ from x', we obtain the bidimensional array $x'' = y' \odot^1 z'$ of shape $(4,5)$, with $y' = x'[(0,0),(0,4)]$ and $z' = x'[(2,0),(4,4)]$. The erased arrays x' and x'' are depicted in Figure 4.10. Furthermore, we have

$x' = y \odot^i z =$

a	b	a	e	a
e	e	e	e	e
a	a	e	e	b
e	e	a	b	b
b	a	b	e	a

and $x'' = y' \odot^i z' =$

a	b	a	e	a
a	a	e	e	b
e	e	a	b	b
b	a	b	e	a

FIGURE 4.10. The successive e-erased arrays from x.

$\rho_e(x) = x''$ since there is no e-hyperplane in x''. Because the letter e still occurs in x'', the bidimensional array x is not e-erasable.

4.6. Characterization of S-Automatic Arrays

Let us recall that our goal is to prove the following result.

Theorem 4.6.1. *Let d be a positive integer. The d-dimensional infinite word x is S-automatic for some abstract numeration system $S = (L, \Sigma, <)$ with $\varepsilon \in L$ if and only if x is the image under a coding of a shape-symmetric pure morphic word.*

The case $d = 1$ is given by Theorem 4.1.15. The main difficulties in generalizing the proof of Theorem 4.1.15 appear in the step from the unidimensional case to the bidimensional case. Therefore, and also for the sake of clarity, we have chosen to present the proof in the case $d = 2$. We have split the proof into two parts.

Part 1. Assume $x = \nu(\mu^\omega(a))$, where $\mu\colon \Sigma \to B_2(\Sigma)$ is a bidimensional morphism prolongable on a and $\nu\colon \Sigma^* \to \Gamma^*$ is a coding such that $y = \mu^\omega(a)$ is shape-symmetric. In this part we show that x is S-automatic for some abstract numeration system $S = (L, \Sigma, <)$ with $\varepsilon \in L$.

Let $Y_1 = (y_{n,0})_{n\geq 0}$ be the first column of y. This word Y_1 is a unidimensional infinite word over a subset Σ_1 of Σ. It is clear that Y_1 is generated by a unidimensional morphism μ_1 derived from μ: one only has to consider the first column occurring in the images by μ of the letters in Σ.

Definition 4.6.2. With each (unidimensional) morphism $\mu\colon \Sigma^* \to \Sigma^*$ and with each letter $a \in \Sigma$, we can canonically associate a DFA, denoted by $\mathcal{A}_{\mu,a}$, and defined as follows. Define $r_\mu = \max_{b\in\Sigma} |\mu(b)|$. The alphabet of $\mathcal{A}_{\mu,a}$ is $[\![0, r_\mu - 1]\!]$. The set of states is Σ. The initial state is a and every state is final. The (partial) transition function δ_μ is defined by $\delta_\mu(b, i) = \mu(b)[i]$ for all $b \in \Sigma$ and $i \in [\![0, |\mu(b)| - 1]\!]$. Removing the words having 0 as a prefix from the accepted language, we obtain the *directive language* of (μ, a). We let $L_{\mu,a}$ denote this directive language.

Note that $L_{\mu,a}$ is a prefix-closed language since all states in $\mathcal{A}_{\mu,a}$ are final. In particular, we have $\varepsilon \in L_{\mu,a}$. The reason why the directive language of a morphism is called in this way will be made clear; see Lemma 4.6.5 and Corollary 4.6.6.

Example 4.6.3. Let us consider the morphism μ of Example 4.4.3 again. Then we obtain $\Sigma_1 = \{a, c, e, g, h\}$ and

$$\mu_1\colon a \mapsto ac,\ c \mapsto e,\ e \mapsto eg,\ g \mapsto h,\ h \mapsto hc.$$

Furthermore, we have $Y_1 = \mu_1^\omega(a) = aceegegheghhceghhchceegh\cdots$. The DFA associated with (μ_1, a) is depicted in Figure 4.11. The first few words in $L_{\mu_1,a}$ ordered with respect to the genealogical order are

$$\varepsilon, 1, 10, 100, 101, 1000, 1001, 1010, 10000.$$

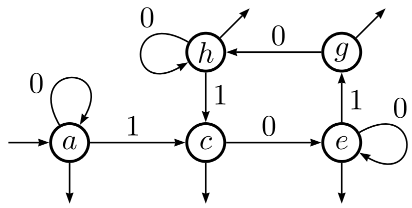

FIGURE 4.11. The automaton $\mathcal{A}_{\mu_1,a}$.

Lemma 4.6.4. *Let $\mu\colon \Sigma^* \to \Sigma^*$ be a morphism prolongable on $a \in \Sigma$. The counting function of the associated directive language is given by*

$$\mathbf{u}_{L_{\mu,a}}(s) = \begin{cases} |\mu^s(a)| - |\mu^{s-1}(a)|, & \textit{if } s \in \mathbb{N} \setminus \{0\}; \\ 1, & \textit{if } s = 0. \end{cases}$$

PROOF. Since $L_{\mu,a}$ is a prefix-closed language, it must contain ε. Therefore we have $\mathbf{u}_{L_{\mu,a}}(0) = 1$. Now, take a positive integer s. The adjacency matrix $M = (M_{b,c})_{b,c\in\Sigma}$ in $\mathbb{N}^{\Sigma\times\Sigma}$ of $\mathcal{A}_{\mu,a}$ is defined by

$$\forall b, c \in \Sigma,\ M_{b,c} = \text{Card}\{i \in [\![0, |\mu(b)| - 1]\!] \mid \delta_\mu(b, i) = c\}.$$

We know that $[M^s]_{b,c}$ is the number of paths of length s from b to c in $\mathcal{A}_{\mu,a}$; for example, see [**Ber70, GR01**]. Since all states are final, the number N_s of words of length s accepted by $\mathcal{A}_{\mu,a}$ is obtained by summing up all the entries in the row of M^s corresponding to a. Because μ is prolongable on a, the automaton $\mathcal{A}_{\mu,a}$ has a loop of label 0 at a. Therefore the number of words of length s accepted by $\mathcal{A}_{\mu,a}$ and starting with 0 is equal to the number N_{s-1} of words of length $s-1$ accepted by $\mathcal{A}_{\mu,a}$. Consequently, the number of words of length s in the directive language $L_{\mu,a}$ is exactly $N_s - N_{s-1}$. The matrix M can also be related to the morphism μ because $M_{b,c}$ is also the number of occurrences of c in $\mu(b)$. In particular, summing up all entries in the row of M^s corresponding to a gives $|\mu^s(a)|$. This completes the proof. □

Lemma 4.6.5. *Let $\mu\colon \Sigma^* \to \Sigma^*$ be a morphism prolongable on $a \in \Sigma$ and S be the abstract numeration system defined by*

$$S = (L_{\mu,a}, [\![0, r_\mu - 1]\!], 0 < \cdots < r_\mu - 1).$$

Write $\mu^\omega(a) = y_0y_1y_2\cdots$ with $y_n \in \Sigma$ for all $n \in \mathbb{N}$. Then, for each $n \in \mathbb{N}$, the $(n+1)$st letter of $\mu^\omega(a)$ is given by

$$y_n = \delta_\mu(a, \text{rep}_S(n))$$

and, by setting $\mathrm{val}_S(0) = 0$*, its image by* μ *is*

$$\mu(y_n) = \mu^\omega(a)[\mathrm{val}_S(\mathrm{rep}_S(n)0), \mathrm{val}_S(\mathrm{rep}_S(n)(|\mu(y_n)| - 1))],$$

i.e., the factor of $\mu^\omega(a)$ *from the position indexed by the* S*-value of the concatenation* $\mathrm{rep}_S(n)0 \in L_{\mu,a}$ *to the position indexed by the* S*-value of the concatenation* $\mathrm{rep}_S(n)(|\mu(y_n)| - 1) \in L_{\mu,a}$*. The latter formula is equivalent to*

$$\forall i \in [\![0, |\mu(y_n)| - 1]\!],\ \mu(y_n)[i] = y_{\mathrm{val}_S(\mathrm{rep}_S(n)i)}.$$

Proof. Proceed by induction on the length s of the words in $L_{\mu,a}$. The only word of length 0 in $L_{\mu,a}$ is $\mathrm{rep}_S(0) = \varepsilon$. Since μ is prolongable on a, we have $y_0 = a = \delta_\mu(a, \mathrm{rep}_S(0))$. Furthermore, for any $i \in [\![0, |\mu(a)| - 1]\!]$, we have $\mathrm{val}_S(i) = i$. So, for any $i \in [\![0, |\mu(a)| - 1]\!]$, we obtain $\mu(y_0)[i] = y_i$.

Now, take a positive integer s and assume that the lemma holds for all integers m which satisfy $|\mathrm{rep}_S(m)| \in [\![0, s-1]\!]$. Take a positive integer n satisfying $|\mathrm{rep}_S(n)| = s$. Let us write $\mathrm{rep}_S(n) = wk$ with $|w| = s - 1$ and $k \in [\![0, |\mu(\delta_\mu(a, w))| - 1]\!]$. Since $L_{\mu,a}$ is prefix-closed, there exists an integer m such that we have $w = \mathrm{rep}_S(m)$. Hence we obtain

$$\begin{aligned}
\delta_\mu(a, \mathrm{rep}_S(n)) &= \delta_\mu(a, wk) \\
&= \delta_\mu(\delta_\mu(a, \mathrm{rep}_S(m)), k) \\
&= \delta_\mu(y_m, k) \text{ (by the induction hypothesis)} \\
&= \mu(y_m)[k] \text{ (by the definition of } \delta_\mu) \\
&= y_{\mathrm{val}_S(\mathrm{rep}_S(m)k)} \text{ (by the induction hypothesis)} \\
&= y_n.
\end{aligned}$$

We have thus shown $y_\ell = \delta_\mu(a, \mathrm{rep}_S(\ell))$ for all $\ell \in [\![0, |\mu^s(a)| - 1]\!]$. Furthermore, from Lemma 4.6.4, we obtain

$$|\mathrm{rep}_S(n)| = t \Leftrightarrow n \in [\![|\mu^{t-1}(a)|, |\mu^t(a)| - 1]\!]. \tag{29}$$

Therefore we can write

$$\mu^{s+1}(a) = \underbrace{\mu^{s-1}(a)uy_nv}_{\mu^s(a)}\mu(u)\mu(y_n)\mu(v)$$

for some finite words u, v. Hence we obtain

$$\forall i \in [\![0, |\mu(y_n)| - 1]\!],\ \mu(y_n)[i] = y_{|\mu^s(a)|+|\mu(u)|+i}.$$

By the definition of $L_{\mu,a}$ we have

$$\forall i \in [\![0, |\mu(y_n)| - 1]\!],\ \mathrm{val}_S(\mathrm{rep}_S(n)i) = \mathrm{val}_S(\mathrm{rep}_S(n)0) + i.$$

Hence it suffices to show $\mathrm{val}_S(\mathrm{rep}_S(n)0) = |\mu^s(a)| + |\mu(u)|$ to conclude the proof.

From relation (29) we know that $|\mu^s(a)|$ is the S-value of the first word of length $s+1$ in $L_{\mu,a}$ with respect to the genealogical order. Next, from the definition of $L_{\mu,a}$ and from the first part of the proof, it follows that $L_{\mu,a}$ contains exactly $|\mu(\delta_\mu(a, \mathrm{rep}_S(\ell)))| = |\mu(y_\ell)|$ words of the form $\mathrm{rep}_S(\ell)j$, where ℓ belongs to $[\![0, |\mu^s(a)|-1]\!]$ and j is a letter. Since $\mathrm{rep}_S(|\mu^{s-1}(a)|)$ is the first word of length s in $L_{\mu,a}$ with respect to the genealogical order, we get that $|\mu(u)| = \sum_{\ell=|\mu^{s-1}(a)|}^{n-1} |\mu(y_\ell)|$ is exactly the number of words of length $s+1$ in $L_{\mu,a}$ of the form $\mathrm{rep}_S(\ell)j$ with $|\,\mathrm{rep}_S(\ell)| = s$ and $\ell < n$, *i.e.*, the number of words in $L_{\mu,a}$ of length $s+1$ less than $\mathrm{rep}_S(n)0$ with respect to the genealogical order. This finishes the proof. □

Corollary 4.6.6. *Let $\mu\colon \Sigma^* \to \Sigma^*$ be a morphism prolongable on $a \in \Sigma$ and S be the abstract numeration system defined by*

$$S = (L_{\mu,a}, [\![0, r_\mu - 1]\!], 0 < \cdots < r_\mu - 1).$$

Assume $\mu^\omega(a) = y_0y_1y_2\cdots$ and $rep_S(n) = w_0\cdots w_\ell$ with $n \in \mathbb{N}$, where the y_i's and the w_i's are letters. Define $z^{(0)} = \mu(a)$ and $z^{(j+1)} = \mu(z^{(j)}[w_j])$ for all $j \in [\![0, \ell-1]\!]$. Then we have $y_n = z^{(\ell)}$.

Example 4.6.7. Let us continue Example 4.6.3. The fixed point Y_1 of μ_1 starts with

$$aceegegh = y_0\cdots y_7$$

and $\mathrm{rep}_S(7) = 1010$. From Lemma 4.6.5 the letter $y_7 = h$ has been generated by applying μ_1 to the letter in the position $\mathrm{val}_S(101) = 4$, that is, $y_4 = g$. We have $y_7 = \mu_1(g)[0]$. In turn, the letter y_4 occurs in the image under μ_1 of the letter in the position $\mathrm{val}_S(10) = 2$, that is, $y_2 = e$. Thus we have $y_4 = \mu_1(e)[1]$. Then the letter y_2 appears in the image of the letter in the position $\mathrm{val}_S(1) = 1$, that is, $y_1 = c$. Finally we have $y_2 = \mu_1(c)[0]$.

The following result is self-evident.

Lemma 4.6.8. *Let x and y be two infinite words and λ and μ be two morphisms such that there exist two letters a and b such that we have $x = \lambda^\omega(a)$ and $y = \mu^\omega(b)$. The directive languages $L_{\lambda,a}$ and $L_{\mu,b}$ are equal if and only if we have* $\mathrm{Shape}_\lambda(x) = \mathrm{Shape}_\mu(y)$.

Example 4.6.9. If one considers the morphism μ_2 defined by $\mu_2\colon a \mapsto ab$, $b \mapsto e$, $e \mapsto eb$ (which is derived from the first row of the bidimensional morphism in Example 4.4.3), then we obtain the DFA $\mathcal{A}_{\mu_2,a}$ depicted in Figure 4.12. The automata in Figure 4.11 and Figure 4.12 clearly accept the same language (the second one being minimal).

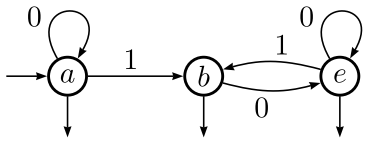

FIGURE 4.12. The automaton $\mathcal{A}_{\mu_2,a}$.

Let $Y_2 = (y_{0,n})_{n\geq 0}$ be the first row of y. This word Y_2 is a unidimensional infinite word over a subset Σ_2 of Σ. It is clear that Y_2 is generated by a morphism μ_2 derived from μ. Since y is shape-symmetric, thanks to Remark 4.4.5 and to Lemma 4.6.8, we can define

$$L_{\mu_1,a} = L_{\mu_2,a} =: L \text{ and } r_{\mu_1} = r_{\mu_2} =: r_\mu.$$

We consider the abstract numeration system built upon this language L (with the natural order on digits). With the above discussion and in particular in view of Lemma 4.6.5, it is clear that, for all $m, n \in \mathbb{N}$, all $u, v \in [\![0, r_\mu - 1]\!]^*$, and all $b, c \in [\![0, r_\mu - 1]\!]$, we have

$$(\mu(y_{\mathrm{val}_S(u),\mathrm{val}_S(v)}))_{b,c} = y_{m,n} \quad \Leftrightarrow \quad ub = \mathrm{rep}_S(m) \text{ and } vc = \mathrm{rep}_S(n). \qquad (30)$$

Example 4.6.10. Consider the letter c occurring in the position $(4, 7)$ in the fixed point y of μ underlined in Figure 4.9. We have $(\mathrm{val}_S(101), \mathrm{val}_S(1010)) = (4, 7)$. If we consider the pair $(\mathrm{val}_S(10), \mathrm{val}_S(101)) = (2, 4)$, then we obtain $(\mu(y_{2,4}))_{1,0} = (\mu(b))_{1,0} = c = y_{4,7}$. In other words, the letter $y_{4,7}$ comes from the letter $y_{2,4}$. We can continue in this way. Since we have $(\mathrm{val}_S(1), \mathrm{val}_S(10)) = (1, 2)$, we also have $b = y_{2,4} = (\mu(y_{1,2}))_{0,1}$. Then we get $y_{1,2} = c = (\mu(y_{0,1}))_{1,0}$ because we have $(\mathrm{val}_S(\varepsilon), \mathrm{val}_S(1)) = (0, 1)$. Finally we obtain $y_{0,1} = b = (\mu(y_{0,0}))_{0,1} = (\mu(a))_{0,1}$ because we have $(\mathrm{val}_S(\varepsilon), \mathrm{val}_S(\varepsilon)) = (0, 0)$.

Now, we extend Definition 4.6.2 to the multidimensional case.

Definition 4.6.11. For each d-dimensional morphism $\mu\colon \Sigma \to B_d(\Sigma)$ and for each letter $a \in \Sigma$, define a DFA $\mathcal{A}_{\mu,a}$ over the alphabet $[\![0, r_\mu - 1]\!]^d$, with

$$r_\mu = \max\{|\mu(b)|_i \mid b \in \Sigma,\ i \in [\![1, d]\!]\}.$$

The set of states is Σ, the initial state is a and all states are final. The (partial) transition function is defined by

$$\forall b \in \Sigma,\ \forall \mathbf{n} \leq |\mu(b)|,\ \delta_\mu(b, \mathbf{n}) = (\mu(b))_{\mathbf{n}}.$$

Thanks to (30), the automaton $\mathcal{A}_{\mu,a}$ satisfies

$$\forall m, n \in \mathbb{N},\ \delta_\mu(a, (\mathrm{rep}_S(m), \mathrm{rep}_S(n))^0) = y_{m,n},$$

where we have padded the shortest word with enough 0's to make two words of the same length; see Definition 4.1.8. If we consider the coding ν as the output function, the corresponding DFAO generates x as an S-automatic word. Note that padding with 0's works correctly since 0 is the lexicographically smallest letter and the directive language L does not contain any words starting with 0. This concludes the first part of the proof.

Example 4.6.12. Again, consider the bidimensional morphism μ of Example 4.4.3 and its shape-symmetric fixed point $\mu^\omega(a)$ depicted in Figure 4.9. If $S = (L, \{0,1\}, 0 < 1)$ is the abstract numeration system built on the directive language $L = \{\varepsilon, 1, 10, 100, 101, 1000, 1001, 1010, 10000, \ldots\}$, then the corresponding DFAO generates $\mu^\omega(a)$ as an S-automatic word. This DFAO is depicted in Figure 4.13, where the output function is the identity. For instance, if we continue Example 4.6.10, then, by reading the bidimensional word $(\mathrm{rep}_S(4), \mathrm{rep}_S(7))^0 = (0101, 1010)$, we obtain

$$y_{0,0} = a \overset{(0,1)}{\rightarrow} y_{0,1} = b \overset{(1,0)}{\rightarrow} y_{1,2} = c \overset{(0,1)}{\rightarrow} y_{2,4} = b \overset{(1,0)}{\rightarrow} y_{4,7} = c.$$

The letters appearing in this sequence of transitions correspond to the underlined ones in Figure 4.9.

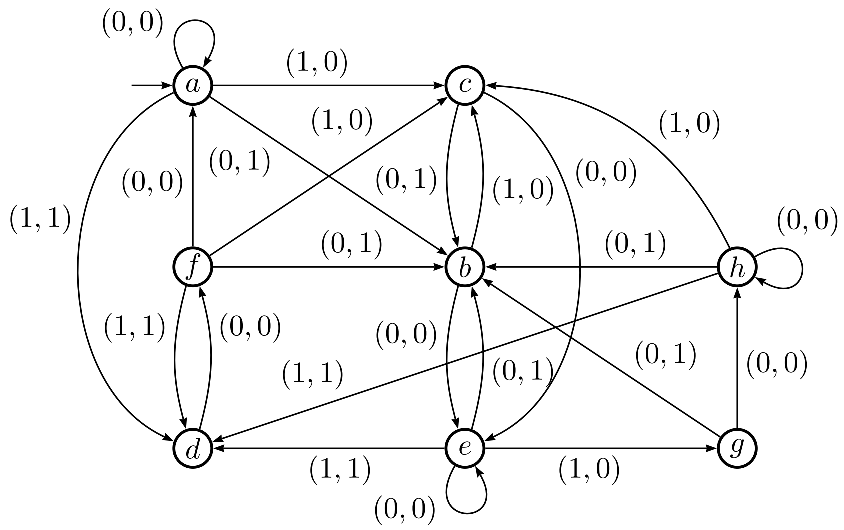

FIGURE 4.13. A DFAO generating $\mu^\omega(a)$ as an S-automatic word.

Part 2. Assume that $x = (x_{m,n})_{m,n\geq 0}$ is a bidimensional S-automatic infinite word over an alphabet Γ for some abstract numeration system

$$S = (L, \Sigma, <), \text{ with } \varepsilon \in L.$$

Assume $\Sigma = \{a_1, \ldots, a_r\}$ and $a_1 < \cdots < a_r$. Let

$$\mathcal{A} = (Q_\mathcal{A}, (\Sigma_\#)^2, \delta_\mathcal{A}, q_0, \Gamma, \tau_\mathcal{A})$$

be a DFAO generating x. We assume that $\# =: a_0$ is a symbol not belonging to Σ and $a_0 < a_1$. Recall that this means

$$\forall m, n \in \mathbb{N},\ \tau_\mathcal{A}(\delta_\mathcal{A}(q_0, (\mathrm{rep}_S(m), \mathrm{rep}_S(n))^\#)) = x_{m,n}.$$

Without loss of generality, we suppose $\delta_\mathcal{A}(q, (\#, \#)) = q$ for all $q \in Q_\mathcal{A}$.

Example 4.6.13. Consider once again the automaton given by Figure 4.7. Then the word x introduced in this proof corresponds to the bidimensional infinite word depicted in Figure 4.8.

In this part we prove that x can be represented as the image under a coding of a shape-symmetric bidimensional pure morphic word. We present the proof in three steps. First, we show that x can be obtained by applying an erasing map to a fixed point of a uniform bidimensional morphism. In the second step we prove that x is morphic. The generating morphism μ and the coding ν are obtained using a construction based on a unidimensional construction from [**AS03**]. Finally, we show that the fixed point of μ under consideration is shape-symmetric.

Definition 4.6.14. Let d be a positive integer and $(\Sigma, <)$ be the totally ordered alphabet $(\{a_0, a_1, \ldots, a_r\}, a_0 < \cdots < a_r)$. Any complete DFA of the form $\mathcal{A} = (Q, \Sigma^d, \delta, q_0, F)$ can be canonically associated with a d-dimensional morphism denoted by $\mu_\mathcal{A} \colon Q \to B_d(Q)$ and defined as follows. The image of a letter q in Q is a d-dimensional square x of size $r+1$ defined by $x_\mathbf{n} = \delta(q, (a_{n_1}, \ldots, a_{n_d}))$ for all $\mathbf{n} \in \mathbb{N}^d$ satisfying $\mathbf{n} \leq (r, \ldots, r)$.

Example 4.6.15. Consider the alphabet $\Sigma = \{\#, a, b\}$ with $\# < a < b$ and the automaton $\mathcal{A}$ depicted in Figure 4.7, with added loops on inputs $(\#, \#)$ at all states. Then we obtain

$$\mu_\mathcal{A}(p) = \begin{array}{|c|c|c|}\hline p & p & q \\ \hline q & p & p \\ \hline q & s & s \\ \hline\end{array}\,,\quad \mu_\mathcal{A}(q) = \begin{array}{|c|c|c|}\hline q & p & p \\ \hline p & s & q \\ \hline q & q & s \\ \hline\end{array}\,,$$

$$\mu_\mathcal{A}(r) = \begin{array}{|c|c|c|}\hline r & p & p \\ \hline s & r & r \\ \hline s & s & p \\ \hline\end{array}\,,\quad \mu_\mathcal{A}(s) = \begin{array}{|c|c|c|}\hline s & r & r \\ \hline r & q & s \\ \hline s & s & r \\ \hline\end{array}\,,$$

and ${\mu_\mathcal{A}}^\omega(p)$ is the bidimensional infinite word depicted in Figure 4.14. Note that ${\mu_\mathcal{A}}^\omega(p)$ is different from the S-automatic word given in Figure 4.8. However, by erasing certain rows and columns in Figure 4.14, namely the ones corresponding to words not belonging to $L = \{a, ba\}^*\{\varepsilon, b\}$, we obtain the same word as in Figure 4.8.

$L \times L$	$\underline{\varepsilon}$	$\underline{a}$	$\underline{b}$	$a\#$	$\underline{aa}$	$\underline{ab}$	$b\#$	$\underline{ba}$	bb	$a\#\#$	$a\#a$	$a\#b$	$aa\#$	$\underline{aaa}$	$\underline{aab}$	$ab\#$	$\cdots$
$\underline{\varepsilon}$	p	p	q	p	p	q	q	p	p	p	p	q	p	p	q	q	$\cdots$
$\underline{a}$	q	p	p	q	p	p	p	s	q	q	p	p	q	p	p	p	
$\underline{b}$	q	s	s	q	s	s	q	q	s	q	s	s	q	s	s	q	
$a\#$	q	p	p	p	p	q	p	p	q	q	p	p	p	p	q	p	
$\underline{aa}$	p	s	q	q	p	p	q	p	p	p	s	q	q	p	p	q	
$\underline{ab}$	q	q	s	q	s	s	q	s	s	q	q	s	q	s	s	q	
$b\#$	q	p	p	s	r	r	s	r	r	q	p	p	s	r	r	s	
$\underline{ba}$	p	s	q	r	q	s	r	q	s	p	s	q	r	q	s	r	
bb	q	q	s	s	s	r	s	s	r	q	q	s	s	s	r	s	
$a\#\#$	q	p	p	p	p	q	p	p	q	p	p	q	p	p	q	q	
$a\#a$	p	s	q	q	p	p	q	p	p	q	p	p	q	p	p	p	
$a\#b$	q	q	s	q	s	s	q	s	s	q	s	s	q	s	s	q	
$aa\#$	p	p	q	s	r	r	q	p	p	q	p	p	p	p	q	p	
$\underline{aaa}$	q	p	p	r	q	s	p	s	q	p	s	q	q	p	p	q	
$\underline{aab}$	q	s	s	s	s	r	q	q	s	q	q	s	q	s	s	q	
$ab\#$	q	p	p	q	p	p	s	r	r	q	p	p	s	r	r	s	
$\vdots$	$\vdots$																$\ddots$

FIGURE 4.14. The fixed point $\mu_{\mathcal{A}}{}^{\omega}(p)$.

From the assumption L is a regular language over Σ. Hence there exists a DFA accepting L and we may easily modify it to obtain a DFA

$$\mathcal{L} = (Q_{\mathcal{L}}, \Sigma_{\#}, \delta_{\mathcal{L}}, \ell_0, F_{\mathcal{L}})$$

accepting $\#^*L$ and satisfying $\delta_{\mathcal{L}}(\ell_0, \#) = \ell_0$. Note that ℓ_0 is a final state since ε belongs to L. Next, let us define a "product" automaton

$$\mathcal{P} = (Q, (\Sigma_{\#})^2, \delta, p_0, F)$$

imitating the behavior of $\mathcal{A}$ and two copies of the automaton $\mathcal{L}$, one for each dimension. The set of states of $\mathcal{P}$ is $Q = Q_{\mathcal{A}} \times Q_{\mathcal{L}} \times Q_{\mathcal{L}}$, where the initial state p_0 is (q_0, ℓ_0, ℓ_0). The transition function $\delta\colon Q \times (\Sigma_{\#})^2 \to Q$ is defined by

$$\delta((q, k, \ell), (a, b)) = (\delta_{\mathcal{A}}(q, (a, b)), \delta_{\mathcal{L}}(k, a), \delta_{\mathcal{L}}(\ell, b)),$$

where (q, k, ℓ) belongs to Q and (a, b) is a pair of letters in $(\Sigma_{\#})^2$. The set of final states is $F = Q_{\mathcal{A}} \times F_{\mathcal{L}} \times F_{\mathcal{L}}$. Let $y = (y_{m,n})_{m,n \geq 0}$ be the infinite word satisfying

$$\forall m, n \in \mathbb{N},\ \delta(p_0, (\mathrm{rep}_S(m), \mathrm{rep}_S(n))^{\#}) = y_{m,n}.$$

Note that both the first and the second component of $(\mathrm{rep}_S(m), \mathrm{rep}_S(n))^{\#}$ belong to the language $\#^*L$. Therefore $\delta(p_0, (\mathrm{rep}_S(m), \mathrm{rep}_S(n))^{\#})$ is a final state. Let us define a coding $\tau\colon F \to \Gamma$ by

$$\forall (q, k, \ell) \in F,\ \tau((q, k, \ell)) = \tau_{\mathcal{A}}(q).$$

By construction, it is clear that we have $\tau(y) = x$.

We consider the canonically associated morphism $\mu_{\mathcal{P}}\colon Q \to B_2(Q)$ given in Definition 4.6.14. Note that $\mu_{\mathcal{P}}$ is prolongable on p_0 since we have

$$\begin{aligned}(\mu_{\mathcal{P}}(p_0))_{\mathbf{0}} &= \delta(p_0, (a_0, a_0))\\ &= (\delta_{\mathcal{A}}(q_0, (\#, \#)), \delta_{\mathcal{L}}(\ell_0, \#), \delta_{\mathcal{L}}(\ell_0, \#))\\ &= (q_0, \ell_0, \ell_0) = p_0.\end{aligned}$$

Example 4.6.16. Let us continue Example 4.2.2 and once again consider the abstract numeration system $S = (\{a, ba\}^*\{\varepsilon, b\}, \{a, b\}, a < b)$ and the DFAO depicted in Figure 4.7, with additional loops on inputs $(\#, \#)$ at all states. The minimal automaton of $\#^*\{a, ba\}^*\{\varepsilon, b\}$ is depicted in Figure 4.15. If $\mathcal{P}$ is the corresponding product automaton, then the fixed point

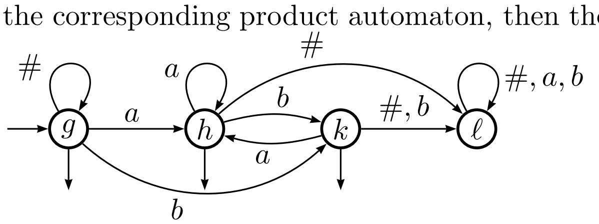

FIGURE 4.15. The minimal automaton accepting $\#^*\{a, ba\}^*\{\varepsilon, b\}$.

${\mu_{\mathcal{P}}}^{\omega}((p, g, g))$ of $\mu_{\mathcal{P}}$ is the bidimensional infinite word depicted in Figure 4.16.

Let e be a new symbol. Remember that ρ_e is the erasing map defined in Section 4.5. Let ρ denote the map $\rho_e \circ \lambda$, where λ is the morphism on $(Q \cup \{e\})^*$ defined by

$$\lambda(p) = \begin{cases} e, & \text{if } p \notin F;\\ p, & \text{otherwise.}\end{cases}$$

We first claim that $y = \rho({\mu_{\mathcal{P}}}^{\omega}(p_0))$. Observe that the infinite word $\lambda({\mu_{\mathcal{P}}}^{\omega}(p_0))$ is e-erasable. Namely, all letters in a fixed row Y of the bidimensional infinite word ${\mu_{\mathcal{P}}}^{\omega}(p_0)$ are of the form (q, k, ℓ) where the second component k is fixed. If k does not belong to $F_{\mathcal{L}}$, the word $\lambda(Y)$ is a unidimensional e-hyperplane of $\lambda({\mu_{\mathcal{P}}}^{\omega}(p_0))$. Thus the map ρ erases all rows where the second component k does not belong to $F_{\mathcal{L}}$. The same holds for

(p,g,g)	(p,g,h)	(q,g,k)	(p,g,ℓ)	(p,g,h)	(q,g,k)	(q,g,ℓ)	$\cdots$
(q,h,g)	(p,h,h)	(p,h,k)	(q,h,ℓ)	(p,h,h)	(p,h,k)	(p,h,ℓ)	
(q,k,g)	(s,k,h)	(s,k,k)	(q,k,ℓ)	(s,k,h)	(s,k,k)	(q,k,ℓ)	
(q,ℓ,g)	(p,ℓ,h)	(p,ℓ,k)	(p,ℓ,ℓ)	(p,ℓ,h)	(q,ℓ,k)	(p,ℓ,ℓ)	
(p,h,g)	(s,h,h)	(q,h,k)	(q,h,ℓ)	(p,h,h)	(p,h,k)	(q,h,ℓ)	
(q,k,g)	(q,k,h)	(s,k,k)	(q,k,ℓ)	(s,k,h)	(s,k,k)	(q,k,ℓ)	
(q,ℓ,g)	(p,ℓ,h)	(p,ℓ,k)	(s,ℓ,ℓ)	(r,ℓ,h)	(r,ℓ,k)	(s,ℓ,ℓ)	
(p,h,g)	(s,h,h)	(q,h,k)	(r,h,ℓ)	(q,h,h)	(s,h,k)	(r,h,ℓ)	
(q,ℓ,g)	(q,ℓ,h)	(s,ℓ,k)	(s,ℓ,ℓ)	(s,ℓ,h)	(r,ℓ,k)	(s,ℓ,ℓ)	
(q,ℓ,g)	(p,ℓ,h)	(p,ℓ,k)	(p,ℓ,ℓ)	(p,ℓ,h)	(q,ℓ,k)	(p,ℓ,ℓ)	
$\vdots$							$\ddots$

FIGURE 4.16. The fixed point ${\mu_{\mathcal{P}}}^{\omega}((p,g,g))$.

columns and third components ℓ of the letters in Q. Hence the bidimensional infinite word $\rho({\mu_{\mathcal{P}}}^{\omega}(p_0))$ only contains letters belonging to F. By the construction of the morphism $\mu_{\mathcal{P}}$, those letters come from the automaton $\mathcal{P}$ by "feeding" it with words belonging to $((\Sigma_{\#})^2)^* \cap (\#^* L)^2$. More precisely, all rows and columns corresponding to words not belonging to L are erased and $(\rho({\mu_{\mathcal{P}}}^{\omega}(p_0)))_{m,n}$ is equal to $\delta(p_0, (\mathrm{rep}_S(m), \mathrm{rep}_S(n))^{\#}) = y_{m,n}$. Hence, by defining $\vartheta = \tau \circ \rho$, we get a map from Σ to Γ such that we have $x = \vartheta({\mu_{\mathcal{P}}}^{\omega}(p_0))$.

Example 4.6.17. We continue Example 4.6.16. This time we consider the bidimensional S-automatic word depicted in Figure 4.8. This word corresponds to the bidimensional infinite word obtained by erasing all rows and columns corresponding to words not belonging to $L = \{a, ba\}^*\{\varepsilon, b\}$ from the bidimensional infinite word ${\mu_{\mathcal{A}}}^{\omega}(p)$ depicted in Figure 4.14. In the latter figure the elements in L have been underlined. By the previous construction we obtain that this word is also the bidimensional infinite word obtained by first erasing all rows with ℓ as the second component and all columns with ℓ as the third component from the bidimensional infinite word ${\mu_{\mathcal{P}}}^{\omega}((p,g,g))$ depicted in Figure 4.16 and then applying the coding τ.

Next, we show that x is morphic by getting rid of the erasing map ρ. We construct a morphism μ prolongable on some letter α and a coding ν such that we have $x = \nu(\mu^{\omega}(\alpha))$. We follow the guidelines of [**AS03**, Theorem 7.7.4]. Also, see Example 4.5.2 above, which illustrates the proof from [**AS03**] in the unidimensional case. First, we need the following definitions.

Definition 4.6.18. Let μ be a morphism on some finite alphabet Σ and let Ψ be a subset of Σ. We say that a letter $a \in \Sigma$ is

(i) (μ, Ψ)-*dead* if we have $\mu^n(a) \in \Psi^*$ for every $n \in \mathbb{N}$;
(ii) (μ, Ψ)-*moribund* if there exists $m \in \mathbb{N}$ such that the word $\mu^m(a)$ contains at least one letter in $\Sigma \setminus \Psi$ and, for every integer $n > m$, we have $\mu^n(a) \in \Psi^*$;
(iii) (μ, Ψ)-*robust* if there exist infinitely many $n \in \mathbb{N}$ such that the word $\mu^n(a)$ contains at least one letter in $\Sigma \setminus \Psi$.

The following lemma from [**AS03**, Lemma 7.7.3] is also valid for multidimensional morphisms, since the proof is only based on the finiteness of the alphabet Σ.

Lemma 4.6.19. *Let μ be a morphism on some finite alphabet Σ and let Ψ be a subset of Σ. Then there exists a positive integer T such that the morphism $\varphi = \mu^T$ satisfies:*

(a) *If a is (φ, Ψ)-moribund, then we have $\varphi^n(a) \in \Psi^*$ for all $n \in \mathbb{N}\setminus\{0\}$ and $a \in \Sigma \setminus \Psi$.*
(b) *If a is (φ, Ψ)-robust, then the word $\varphi^n(a)$ contains at least one letter in $\Sigma \setminus \Psi$ for all $n \in \mathbb{N} \setminus \{0\}$.*

Remark 4.6.20. Note that, from Lemma 4.6.19, a letter in Ψ is either (φ, Ψ)-dead or (φ, Ψ)-robust. Furthermore, a letter in $\Sigma \setminus \Psi$ is either (φ, Ψ)-moribund or (φ, Ψ)-robust.

We may assume, by taking a power of $\mu_{\mathcal{P}}$ if necessary, that $\mu_{\mathcal{P}}$ satisfies the properties (a) and (b) listed for φ in Lemma 4.6.19 with $\Psi = F^c = Q\setminus F$. For the sake of simplicity, we use the words *dead*, *moribund* and *robust* instead of $(\mu_{\mathcal{P}}, F^c)$-dead, $(\mu_{\mathcal{P}}, F^c)$-moribund and $(\mu_{\mathcal{P}}, F^c)$-robust from now on.

We classify the states of $Q_{\mathcal{L}}$ and Q into four categories. The *type of a state* $k \in Q_{\mathcal{L}}$ is

$$T_k = \begin{cases} \Delta, & \text{if } k \notin F_{\mathcal{L}} \text{ and } \forall a \in \Sigma_{\#},\ \delta_{\mathcal{L}}(k,a) \notin F_{\mathcal{L}}; \\ M, & \text{if } k \in F_{\mathcal{L}} \text{ and } \forall a \in \Sigma_{\#},\ \delta_{\mathcal{L}}(k,a) \notin F_{\mathcal{L}}; \\ R_{F^c}, & \text{if } k \notin F_{\mathcal{L}} \text{ and } \exists a \in \Sigma_{\#},\ \delta_{\mathcal{L}}(k,a) \in F_{\mathcal{L}}; \\ R_F, & \text{if } k \in F_{\mathcal{L}} \text{ and } \exists a \in \Sigma_{\#},\ \delta_{\mathcal{L}}(k,a) \in F_{\mathcal{L}}. \end{cases}$$

The *type of a state* $p = (q, k, \ell) \in Q$ is

$$T_p = \begin{cases} \Delta, & \text{if } p \text{ is dead;} \\ M, & \text{if } p \text{ is moribund;} \\ R_{F^c}, & \text{if } p \in F^c \text{ and } p \text{ is robust;} \\ R_F, & \text{if } p \in F \text{ and } p \text{ is robust.} \end{cases}$$

From these definitions it is clear that the type of $(q,k,\ell) \in Q$ only depends on the types of k and $\ell \in Q_{\mathcal{L}}$ according to Figure 4.17. Note that, using the properties (a) and (b) of Lemma 4.6.19, it suffices to consider transitions $\delta_{\mathcal{L}}(k,a)$ by each letter $a \in \Sigma_{\#}$ instead of transitions $\delta_{\mathcal{L}}(k,w)$ by all words w in $(\Sigma_{\#})^*$. For instance, if the type of k is R_{F^c} and the type of ℓ is R_F, then we have $k \notin F_{\mathcal{L}}$. Hence (q,k,ℓ) belongs to F^c for all $q \in Q_{\mathcal{A}}$. Furthermore, there exist $m,n \in [\![0,r]\!]$ such that we have $\delta_{\mathcal{L}}(k,a_m) \in F_{\mathcal{L}}$ and $\delta_{\mathcal{L}}(\ell,a_n) \in F_{\mathcal{L}}$. This means that $(\mu_{\mathcal{P}}((q,k,\ell)))_{m,n}$ belongs to F. Hence, from Lemma 4.6.19 and Remark 4.6.20, (q,k,ℓ) is robust.

T_k \ T_ℓ	Δ	M	R_{F^c}	R_F
Δ	Δ	Δ	Δ	Δ
M	Δ	M	Δ	M
R_{F^c}	Δ	Δ	R_{F^c}	R_{F^c}
R_F	Δ	M	R_{F^c}	R_F

FIGURE 4.17. Type T_p of a letter $p = (q,k,\ell) \in Q$.

Let us define two morphisms λ_Δ and λ_M on $(Q \cup \{e\})^*$ in a similar way as λ was defined above:

$$\lambda_\Delta(p) = \begin{cases} e, & \text{if } p \text{ is dead;} \\ p, & \text{otherwise} \end{cases} \quad \text{and} \quad \lambda_M(p) = \begin{cases} e, & \text{if } p \text{ is moribund;} \\ p, & \text{otherwise.} \end{cases}$$

From the property (b) of Lemma 4.6.19 we know that if p is robust, then $\mu_{\mathcal{P}}(p)$ contains at least one letter in F. Since every dead letter must belong to F^c, for all robust $p \in Q$, the word $\lambda_\Delta(\mu_{\mathcal{P}}(p))$ contains at least one letter in F. For any $\ell \in Q_{\mathcal{L}}$, let us define a sequence $(d_\ell(i))_{0 \le i \le h_\ell}$ such that we have $d_\ell(0) = 0$, $d_\ell(h_\ell) = r+1$, and, for all $i \in [\![0, h_\ell - 1]\!]$, we have $d_\ell(i) < d_\ell(i+1)$ and there exists exactly one index $n \in [\![d_\ell(i), d_\ell(i+1)-1]\!]$ satisfying

$$\delta_{\mathcal{L}}(\ell, a_n) \in F_{\mathcal{L}}. \tag{31}$$

Note that h_ℓ is the number of letters $a_n \in \Sigma_{\#}$ satisfying condition (31). Hence, for each robust letter $p = (q,k,\ell)$, we obtain $h_k, h_\ell \ge 1$. Thus we may define the factorization

$$\lambda_\Delta(\mu_{\mathcal{P}}(p)) = \boxed{\begin{matrix} w_p(0,0) & w_p(0,1) & \cdots & w_p(0,h_\ell-1) \\ w_p(1,0) & w_p(1,1) & \cdots & w_p(1,h_\ell-1) \\ \vdots & \vdots & \ddots & \vdots \\ w_p(h_k-1,0) & w_p(h_k-1,1) & \cdots & w_p(h_k-1,h_\ell-1) \end{matrix}},$$

where each bidimensional array is

$$w_p(i,j) = \lambda_\Delta(\mu_{\mathcal{P}}(p))[(d_k(i), d_\ell(j)), (d_k(i+1)-1, d_\ell(j+1)-1)]$$

contains exactly one letter in F.

Example 4.6.21. Let us continue Example 4.6.17. Recall that the product automaton $\mathcal{P}$ is produced from the automaton $\mathcal{A}$ depicted in Figure 4.7 and the automaton $\mathcal{L}$ depicted in Figure 4.15. Note that the type of the state ℓ in $\mathcal{L}$ is $T_\ell = \Delta$ and all other states have type R_F. From Figure 4.16 we see that we have

$$\mu_{\mathcal{P}}(p,g,g) = \boxed{\begin{matrix} (p,g,g) & (p,g,h) & (q,g,k) \\ (q,h,g) & (p,h,h) & (p,h,k) \\ (q,k,g) & (s,k,h) & (s,k,k) \end{matrix}}$$

and

$$\mu_{\mathcal{P}}(q,h,g) = \boxed{\begin{matrix} (q,\ell,g) & (p,\ell,h) & (p,\ell,k) \\ (p,h,g) & (s,h,h) & (q,h,k) \\ (q,k,g) & (q,k,h) & (s,k,k) \end{matrix}}\,.$$

By definition, h_ℓ is the number of letters a_n in $\Sigma_\#$ such that $\delta_{\mathcal{L}}(\ell, a_n)$ belongs to $F_{\mathcal{L}}$. Hence we obtain $h_g = 3$ and $h_h = 2$. Then, from Figure 4.17, we find $\lambda_\Delta(\mu_{\mathcal{P}}(p,g,g)) = \mu_{\mathcal{P}}(p,g,g)$ and

$$\lambda_\Delta(\mu_{\mathcal{P}}(q,h,g)) = \boxed{\begin{matrix} e & e & e \\ (p,h,g) & (s,h,h) & (q,h,k) \\ (q,k,g) & (q,k,h) & (s,k,k) \end{matrix}}\,.$$

Since all letters in $\mu_{\mathcal{P}}(p,g,g)$ belong to F, the array $w_{(p,g,g)}(i,j)$ is a square of size 1 for every $(i,j) \in [\![0, h_g - 1]\!] \times [\![0, h_g - 1]\!]$. We also obtain

$$w_{(q,h,g)}(0,0) = \boxed{\begin{matrix} e \\ (p,h,g) \end{matrix}}\,, \quad w_{(q,h,g)}(0,1) = \boxed{\begin{matrix} e \\ (s,h,h) \end{matrix}}\,,$$

$$w_{(q,h,g)}(0,2) = \boxed{\begin{matrix} e \\ (q,h,k) \end{matrix}}\,, \quad w_{(q,h,g)}(1,0) = \boxed{(q,k,g)}\,,$$

$$w_{(q,h,g)}(1,1) = \boxed{(q,k,h)}\,, \quad w_{(q,h,g)}(1,2) = \boxed{(s,k,h)}\,.$$

Next, we show that if p is a robust state in Q, then the bidimensional array $\lambda_M(\lambda_\Delta(\mu_{\mathcal{P}}(p)))$ is e-erasable. Proceed by contradiction and assume that $v := \lambda_M(\lambda_\Delta(\mu_{\mathcal{P}}(p)))$ is not e-erasable. Then there must exist $m, n \in \mathbb{N}$ such that we have $v_{m,n} = e$, $v_{m,n'} \neq e$ for some n', and $v_{m',n} \neq e$ for some m'. By construction, the letter $p' = (\mu_{\mathcal{P}}(p))_{m,n} = (q,k,\ell)$ is mapped onto e if we have either $T_{p'} = \Delta$ or $T_{p'} = M$. For the same reason, the letters $v_{m,n'} = (q',k,\ell')$ and $v_{m',n} = (q'',k',\ell)$ must be robust. Thus there exist letters $a_{m''}$, $a_{n''} \in \Sigma_\#$ such that we have $\delta_{\mathcal{L}}(k, a_{m''}) \in F_{\mathcal{L}}$ and $\delta_{\mathcal{L}}(\ell, a_{n''}) \in F_{\mathcal{L}}$. Hence it follows that $p' = (q,k,\ell)$ is robust since the letter $(\mu_{\mathcal{P}}(p'))_{m'',n''}$ belongs to F, which is a contradiction.

Now, we are ready to introduce a bidimensional morphism μ on a new alphabet Ξ and a coding $\nu'\colon \Xi \to Q$ such that we have $y = \nu'(\mu^\omega(\alpha))$ for some letter $\alpha \in \Xi$. The alphabet of the new symbols is

$$\Xi = \{\alpha(p,i,j) \mid p = (q,k,\ell) \in Q \text{ is robust, } (i,j) \in [\![0,h_k-1]\!] \times [\![0,h_\ell-1]\!]\}.$$

For each letter $\alpha(p,i,j)$ in Ξ, if we have

$$(\rho_e(\lambda_M(w_p(i,j))))_{m,n} = (q',k',\ell') = p',$$

for suitable (m,n) (that is, $(m,n) < |\rho_e(\lambda_M(w_p(i,j)))|$), then we define $u_{p,i,j}(m,n)$ to be the array of shape $(h_{k'},h_{\ell'})$ satisfying

$$(u_{p,i,j}(m,n))_{i',j'} = \alpha((q',k',\ell'),i',j') = \alpha(p',i',j')$$

for all $(i',j') \in [\![0,h_{k'}-1]\!] \times [\![0,h_{\ell'}-1]\!]$. The image of $\alpha(p,i,j)$ under the morphism $\mu\colon \Xi \to B_2(\Xi)$ is defined to be the array

$$\boxed{\begin{array}{cccc} u_{p,i,j}(0,0) & u_{p,i,j}(0,1) & \cdots & u_{p,i,j}(0,s_2-1) \\ u_{p,i,j}(1,0) & u_{p,i,j}(1,1) & \cdots & u_{p,i,j}(1,s_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ u_{p,i,j}(s_1-1,0) & u_{p,i,j}(s_1-1,1) & \cdots & u_{p,i,j}(s_1-1,s_2-1) \end{array}}\,,$$

with $(s_1,s_2) = |\rho_e(\lambda_M(w_p(i,j)))|$. Note that the above concatenation of the arrays $u_{p,i,j}(m,n)$ is well defined. Since all letters occurring in a row of $w_p(i,j)$ are of the form (q',k',ℓ') where the second component k' is fixed, it means that letters occurring in a same row of $\rho_e(\lambda_M(w_p(i,j)))$ have the same second component k'. Hence we have $|u_{p,i,j}(m,n)|_{\widehat{2}} = |u_{p,i,j}(m,n')|_{\widehat{2}} = h_{k'}$. Thus the words $u_{p,i,j}(m,n)$ and $u_{p,i,j}(m,n')$ can be concatenated in the direction 2. The same holds for $u_{p,i,j}(m,n)$ and $u_{p,i,j}(m',n)$ in the direction 1. The coding $\nu'\colon \Xi \to Q$ is defined by

$$\nu'(\alpha(p,i,j)) = \rho(w_p(i,j)). \tag{32}$$

Note that, from the definition of $w_p(i,j)$, there is only one letter belonging to F in $w_p(i,j)$. Hence the array $\lambda(w_p(i,j))$ is e-erasable, since only one letter is different from e.

Following the proof of [**AS03**, Theorem 7.7.4], we may prove by induction on n that, for all robust letters $p = (q,k,\ell)$ and for all $n \in \mathbb{N}$, we have

$$\nu' \circ \mu^n(z) = \rho(\mu_{\mathcal{P}}^{n+1}(p)), \tag{33}$$

with

$$z = \boxed{\begin{array}{cccc} \alpha(p,0,0) & \alpha(p,0,1) & \cdots & \alpha(p,0,h_\ell-1) \\ \alpha(p,1,0) & \alpha(p,1,1) & \cdots & \alpha(p,1,h_\ell-1) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha(p,h_k-1,0) & \alpha(p,h_k-1,1) & \cdots & \alpha(p,h_k-1,h_\ell-1) \end{array}}\,.$$

Since $\mu_{\mathcal{P}}$ is prolongable on p_0 and $x = \vartheta({\mu_{\mathcal{P}}}^{\omega}(p_0))$ is a bidimensional infinite word, the letter p_0 must be robust. Furthermore, we have $p_0 \in F$. Therefore we obtain $(w_{p_0}(0,0))_{0,0} = (\rho_e(\lambda_M(w_{p_0}(0,0))))_{0,0} = p_0$. Then, we find $(u_{p_0,0,0}(0,0))_{0,0} = \alpha(p_0,0,0)$. Consequently, the morphism μ is prolongable on $\alpha := \alpha(p_0,0,0)$. From (33) it follows

$$\begin{aligned} \forall n \in \mathbb{N},\ \nu'(\mu^{n+1}(\alpha)) &= \begin{bmatrix} \nu'(\mu^n(u_{p_0,0,0}(0,0))) & U \\ V & W \end{bmatrix} \\ &= \begin{bmatrix} \rho(\mu_{\mathcal{P}}^{n+1}(p_0)) & U \\ V & W \end{bmatrix}, \end{aligned}$$

where U, V and W are bidimensional arrays. Since $\rho(\mu_{\mathcal{P}}^{n+1}(p_0))$ tends to y as n tends to infinity, we have

$$\nu'(\mu^{\omega}(\alpha)) = \rho(\mu_{\mathcal{P}}^{\omega}(p_0)) = y.$$

Hence, by defining the coding $\nu\colon \Xi \to \Gamma$ as $\nu = \tau \circ \nu'$, we obtain

$$\nu(\mu^{\omega}(\alpha)) = \tau(y) = x.$$

This concludes the second step of the proof.

Example 4.6.22. Let us continue Example 4.6.21. We obtain

$$\rho_e(\lambda_M(\lambda_\Delta(\mu_{\mathcal{P}}(q,h,g)))) = \boxed{\begin{array}{ccc} (p,h,g) & (s,h,h) & (q,h,k) \\ (q,k,g) & (q,k,h) & (s,k,k) \end{array}}.$$

Since $w_{(p,g,g)}(i,j)$ is a square of size 1 for every $(i,j) \in [\![0,2]\!] \times [\![0,2]\!]$, we have

$$|\rho_e(\lambda_M(w_{(p,g,g)}(i,j)))| = (1,1)$$

and

$$(\rho_e(\lambda_M(w_{(p,g,g)}(i,j))))_{0,0} = w_{(p,g,g)}(i,j) = (\mu_{\mathcal{P}}(p,g,g))_{i,j}.$$

In particular, we have

$$\begin{aligned} (\rho_e(\lambda_M(w_{(p,g,g)}(0,0))))_{0,0} &= (p,g,g); \\ (\rho_e(\lambda_M(w_{(p,g,g)}(1,0))))_{0,0} &= (q,h,g). \end{aligned}$$

Hence, $u_{(p,g,g),0,0}(0,0)$ is an array of shape $(h_g,h_g) = (3,3)$ satisfying

$$(u_{(p,g,g),0,0}(0,0))_{i',j'} = \alpha((p,g,g),i',j')$$

for all $(i',j') \in [\![0,2]\!] \times [\![0,2]\!]$. We obtain that the image $\mu(\alpha((p,g,g),0,0))$ is

$$u_{(p,g,g),0,0}(0,0) = \boxed{\begin{array}{ccc} \alpha((p,g,g),0,0) & \alpha((p,g,g),0,1) & \alpha((p,g,g),0,2) \\ \alpha((p,g,g),1,0) & \alpha((p,g,g),1,1) & \alpha((p,g,g),1,2) \\ \alpha((p,g,g),2,0) & \alpha((p,g,g),2,1) & \alpha((p,g,g),2,2) \end{array}}.$$

Similarly, we obtain $|u_{(p,g,g),1,0}(0,0)| = (h_h,h_g) = (2,3)$ and

$$(u_{(p,g,g),1,0}(0,0))_{i',j'} = \alpha((q,h,g),i',j')$$

for all $(i', j') \in [\![0,1]\!] \times [\![0,2]\!]$. Therefore the image $\mu(\alpha((p,g,g),1,0))$ is

$$u_{(p,g,g),1,0}(0,0) = \boxed{\begin{matrix} \alpha((q,h,g),0,0) & \alpha((q,h,g),0,1) & \alpha((q,h,g),0,2) \\ \alpha((q,h,g),1,0) & \alpha((q,h,g),1,1) & \alpha((q,h,g),1,2) \end{matrix}}\,.$$

Next, we apply the coding ν to the images above. In view of (32), we get

$$\nu'(\mu(\alpha((p,g,g),0,0))) = \mu_{\mathcal{P}}(p,g,g)$$

and

$$\nu'(\mu(\alpha((p,g,g),1,0))) = \boxed{\begin{matrix} (p,h,g) & (s,h,h) & (q,h,k) \\ (q,k,g) & (q,k,h) & (s,k,k) \end{matrix}}\,.$$

Since we have $\nu = \tau \circ \nu'$, the infinite word $\nu(\mu^\omega(\alpha((p,g,g),0,0)))$ begins with

$$\nu\left(\mu(\alpha((p,g,g),0,0)) \odot^1 \mu(\alpha((p,g,g),1,0))\right) = \boxed{\begin{matrix} p & p & q \\ q & p & p \\ q & s & s \\ p & s & q \\ q & q & s \end{matrix}}\,,$$

which corresponds to the upper left corner of the infinite word depicted in Figure 4.8.

Finally, we have to show that $w = \mu^\omega(\alpha)$ is shape-symmetric, that is, that $|\mu(w_{n,n})|$ is a square for all $n \in \mathbb{N}$. First, observe that since we have $\alpha = \alpha(p_0, 0, 0)$, where the second and the third component of $p_0 = (q_0, \ell_0, \ell_0)$ are equal, the letters $(\mu^\omega(\alpha))_{n,n}$ must be of the form $\alpha((q,k,k),i,i)$. Second, if $p = (q,k,k)$ is a robust letter belonging to Q, then $\mu(\alpha(p,i,i))$ is a square for all $i \in [\![0, h_k - 1]\!]$. This completes the proof.

CHAPTER 5

Representing Real Numbers

5.1. Introduction

In [**LR02**] P. Lecomte and M. Rigo showed how to represent an interval of real numbers in an abstract numeration system built over an exponential regular language satisfying some suitable conditions. In this chapter we provide a wider framework and we show that their results can be extended to abstract numeration systems built on a language that is not necessarily regular. Our aim is to provide a unified approach for representing real numbers in various numeration systems encountered in the literature [**AFS08, DT89, LR01, Lot02**]. The material of this chapter can be found in [**CLGR**].

We will follow the structure given below. In Section 5.2 we consider generalized abstract numeration systems S, that is, having a numeration language L which is not necessarily regular. In particular, we extend the computation of the numerical S-value of words in L to these generalized abstract numeration systems.

Then, in Section 5.3, we show that the infinite words obtained as limits of words of a language are exactly the infinite words having all its prefixes in the corresponding prefix-closure. In view of this result, to represent real numbers, we shall only consider abstract numeration systems built on a prefix-closed language.

In Section 5.4 we show how to represent an interval $[s_0, 1]$ of real numbers in a generalized abstract numeration system built on a language satisfying some general hypotheses. To that aim, we divide $[s_0, 1]$ into particular subintervals I_y for each word y that is a prefix of infinitely many words in L. Our hypotheses are satisfied if and only if these subintervals I_y are well defined and become smaller and smaller as the length of the corresponding prefixes y becomes larger and larger. We also note that our formalism to represent real numbers generalizes that of usual positional numeration systems like integer base numeration systems defined in Example 1.6.2 on page 12 or β-numerations introduced in Remark 3.2.4 on page 51.

In Section 5.5 we show that the methods developed in Section 5.4 for representing real numbers generalize those from [**LR02**], where converging sequences of words were mainly considered.

Next, in Section 5.6, we apply some results from [**BB97**] to show that, if the numeration language is context-free, then the representations of the limit points of the subintervals I_y must be ultimately periodic.

Finally, in Section 5.7, we give three applications of our methods, which were not settled by the results of [**LR02**]. First, we consider a non-regular language L such that its prefix-closure $\text{Pref}(L)$ is regular. In a second part we illustrate the representation of real numbers in the generalized abstract numeration system built on the language of the prefixes of Dyck words. In this case neither the Dyck language D nor its prefix-closure $\text{Pref}(D)$ is recognized by a finite automaton. We compute the complexity functions of the latter language and we show that we can apply our results to the corresponding abstract numeration system. The third application that we consider is the abstract numeration system built on the language $L_{3/2}$ recently introduced in [**AFS08**]. We show that our method, up to some scaling factor, leads to the same representation of real numbers as that given in [**AFS08**].

5.2. Generalized Numeration Systems

In this chapter we consider abstract numeration systems $S = (L, \Sigma, <)$ where L is not necessarily regular. In this case, to avoid any confusion, we will refer to them as *generalized abstract numeration systems*.

Example 5.2.1. Let us consider the language

$$L = \{w \in \{a, b\}^* \colon \ ||w|_a - |w|_b| \leq 1\}$$

and the generalized abstract numeration system $S = (L, \{a, b\}, a < b)$. The minimal automaton of L is given in Figure 5.1. Note that L is context-free

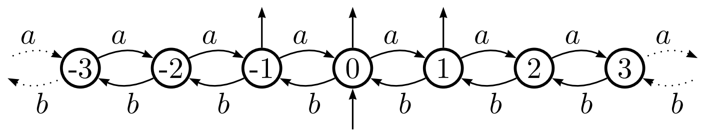

FIGURE 5.1. The minimal automaton of L.

but not regular. Therefore its minimal automaton must be infinite. The first few words in L are $\varepsilon, a, b, ab, ba, aab, aba, abb, baa, bab, bba, aabb, abab, abba$. For instance, we have $\text{rep}_S(5) = aab$ and $\text{val}_S(aabb) = 11$.

The next result is another formulation of Proposition 1.7.6 on page 18 but extended to any language.

Proposition 5.2.2. *Let $S = (L, \Sigma, <)$ be a generalized abstract numeration system and let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a deterministic automaton recognizing L. The S-value of a word w in L is given by*

$$\mathrm{val}_S(w) = \mathbf{v}_{q_0}(|w| - 1) + \sum_{i=0}^{|w|-1} \sum_{a<w[i]} \mathbf{u}_{q_0 \cdot w[0,i-1]a}(|w| - i - 1).$$

5.3. Languages with Uncountable Adherence

One can endow $\Sigma^\omega \cup \Sigma^*$ with a metric space structure as follows.

Definition 5.3.1. Let Σ be an alphabet. The usual distance d over Σ^ω is defined as follows. For two distinct infinite words x and y over Σ, we define $d(x, y) = 2^{-\ell(x,y)}$, where $\ell(x, y) = \inf\{i \in \mathbb{N} \mid x[i] \neq y[i]\}$ denotes the length of the largest common prefix between x and y. Furthermore, we set $d(x, x) = 0$ for all $x \in \Sigma^\omega$. This distance can be extended to $\Sigma^\omega \cup \Sigma^*$ by replacing finite words z over Σ by $z\#^\omega \in (\Sigma \cup \{\#\})^\omega$, where $\#$ is a new letter, not belonging to the alphabet Σ. A sequence $(w^{(n)})_{n\geq 0}$ of words over Σ *converges* to a word w over Σ if $d(w^{(n)}, w)$ tends to 0 as n tends to $+\infty$. We use notation such as $w^{(n)} \to w$ or $\lim_{n\to+\infty} w^{(n)} = w$ to mean that $(w^{(n)})_{n\geq 0}$ converges to w.

This distance d is *ultrametric*[1], that is, it satisfies

$$\forall x, y, z \in \Sigma^\omega \cup \Sigma^*,\ d(x, y) \leq \max\{d(x, z), d(y, z)\}.$$

Example 5.3.2. Consider the finite words $x = 0012012$ and $y = 00021$, and the infinite word $z = 0002010101\cdots$ over the alphabet $\{0, 1, 2\}$. We have $d(x, y) = d(x, z) = 2^{-2} = \frac{1}{4}$ and $d(y, z) = 2^{-4} = \frac{1}{16}$. The sequence $(aac(db)^n ca)_{n\geq 0}$ of words over $\{a, b, c, d\}$ converges to the infinite word $aac(db)^\omega$.

The notion of adherence was introduced in [**Niv78**] and was extensively studied in [**BN80**].

Definition 5.3.3. Let L be a language over an alphabet Σ. The *adherence of L*, denoted by $\mathrm{Adh}(L)$, is the ω-language of the infinite words over Σ whose prefixes are the prefixes of words in L:

$$\mathrm{Adh}(L) = \{w \in \Sigma^\omega \mid \mathrm{Pref}(w) \subseteq \mathrm{Pref}(L)\}.$$

[1]This notion is closely related to that of non-Archimedean absolute value mentioned in Section 3.3 on page 62. Indeed, if N is a non-Archimedean absolute value, then the distance $d(x, y) = N(x - y)$ induced is ultrametric.

Note that $\mathrm{Adh}(L)$ is empty if and only if L is finite. Indeed, if L if infinite, from König's lemma[2], there is at least an infinite word belonging to $\mathrm{Adh}(L)$. For the usual topology on $\Sigma^\omega \cup \Sigma^*$, which has been defined above, the closure $\bar{L}$ of a language L over Σ satisfies the equality $\bar{L} = L \cup \mathrm{Adh}(L)$.

The following lemma provides a characterization of the adherence of a language. This result was proved in [**BN80**] but we restate the proof for the sake of thoroughness.

Lemma 5.3.4. *Let L be a language over an alphabet Σ. The adherence of L is the ω-language of the infinite words over Σ that are limits of words in L:*

$$\mathrm{Adh}(L) = \{w \in \Sigma^\omega \mid \exists (w^{(n)})_{n\geq 0} \in L^{\mathbb{N}},\, w^{(n)} \to w\}.$$

Proof. Take an infinite word w in $\mathrm{Adh}(L)$. Then, from the definition, for all $n \in \mathbb{N}$, we have $w[0, n-1] \in \mathrm{Pref}(L)$. Thus, for all $n \in \mathbb{N}$, there exists a finite word $z^{(n)}$ over Σ such that $w^{(n)} := w[0, n-1]z^{(n)}$ belongs to L. Obviously we have $w^{(n)} \to w$. Consequently, w belongs to the right-hand side set in the statement. Conversely, take an infinite word w which is the limit of a sequence $(w^{(n)})_{n\geq 0}$ of words in L. Then, for all $\ell \in \mathbb{N}$, there exists $n \in \mathbb{N}$ such that we have $w[0, \ell - 1] \in \mathrm{Pref}(w^{(n)}) \subseteq \mathrm{Pref}(L)$. This shows that w belongs to $\mathrm{Adh}(L)$. □

The notion of center of a language can be found in [**BN80**].

Definition 5.3.5. Let L be a language over an alphabet Σ. The *center of* L, denoted by $\mathrm{Center}(L)$, is the prefix-closure of the adherence of L:

$$\mathrm{Center}(L) = \mathrm{Pref}(\mathrm{Adh}(L)).$$

Note that the center of a language contains only finite words. The next lemma gives a characterization of the center of a language. This result was proved in [**BN80**] but, again, we have chosen to restate the proof here in order to be thorough.

Lemma 5.3.6. *Let L be a language over an alphabet Σ. The center of L is the language of the prefixes of an infinite number of words in L:*

$$\mathrm{Center}(L) = \{w \in \mathrm{Pref}(L) \mid w^{-1}L \textit{ is infinite}\}.$$

[2]König's lemma states that, in any infinite tree in which any node has a finite number of sons, there is at least an infinite path. In other words, if L is an infinite prefix-closed language over an alphabet Σ, then there is an infinite word over Σ having all its prefixes in L. For instance, see the handbooks [**Ber70, GR01**].

Proof. Take a word w in Center(L). By definition there exists a infinite word z over Σ such that wz belongs to Adh(L). Then, for all $n \in \mathbb{N}$, the word $w \cdot z[0, n-1]$ belongs to Pref(L). Thus, for all $n \in \mathbb{N}$, there exists a finite word $y^{(n)}$ over Σ such that $w^{(n)} := w \cdot z[0, n-1] \cdot y^{(n)}$ belongs to L. Furthermore, there are infinitely many such words $w^{(n)}$. This shows that w belongs to the right-hand side set in the statement. Conversely, let w be a prefix of infinitely many words in L. Hence there exists a letter a in Σ such that wa is the prefix of infinitely many words in L. By iterating this argument, there exists a sequence $(a_n)_{n\geq 0}$ of letters in Σ such that $wa_0 \cdots a_n$ belongs to Pref(L) for all $n \in \mathbb{N}$. This implies that the infinite word $wa_0a_1 \cdots$ belongs to Adh(L). Hence w belongs to Center(L). □

Definition 5.3.7. If L is a language over an alphabet Σ, then we let

$$L_\infty = \{w \in \Sigma^\omega \mid \exists^\infty n \in \mathbb{N},\, w[0, n-1] \in L\}$$

denote the ω-language of infinite words over Σ having infinitely many prefixes in L, where the notation $\exists^\infty n$ means "there exist infinitely many n".

Again, observe that L_∞ is empty if and only if L is finite. The following lemma is self-evident.

Lemma 5.3.8. *For any language L, we have $L_\infty \subseteq \mathrm{Adh}(L)$. Furthermore, if L is a prefix-closed language, then we have*

$$L_\infty = \mathrm{Adh}(L) = \{w \in \Sigma^\omega \mid \forall n \in \mathbb{N},\, w[0, n-1] \in L\}.$$

Let us recall two results from [**LR02**].

Proposition 5.3.9. *Let L be a regular language. The adherence* Adh(L) *is uncountably infinite if and only if, in any DFA accepting L, there exist at least two distinct cycles $(p_1, \ldots, p_r, p_1)$ and $(q_1, \ldots, q_s, q_1)$, with $r, s \geq 2$, starting from the same accessible and coaccessible state $p_1 = q_1$.*

Proposition 5.3.10. *Let L be a regular language. The ω-language L_∞ is uncountably infinite if and only if, in any DFA accepting L, there exist at least two distinct cycles $(p_1, \ldots, p_r, p_1)$ and $(q_1, \ldots, q_s, q_1)$, with $r, s \geq 2$, starting from the same accessible state $p_1 = q_1$ and such that each of them contains at least a final state.*

Once again, let us recall that the class of regular languages splits into two parts: the exponential regular languages and the polynomial regular languages. From Theorem 1.5.4 on page 10 we know that the polynomial regular languages over an alphabet Σ are exactly those that are finite unions of languages of the form

$$xy_1^*z_1y_2^*\cdots y_k^*z_k,$$

with $k\in\mathbb{N}$ and $x, y_i, z_i\in\Sigma^*$ for all $i\in[\![1,k]\!]$. The following result is thus a straightforward consequence of Proposition 5.3.9.

Corollary 5.3.11. *If L is a regular language, then the following assertions are equivalent:*

- $\mathrm{Adh}(L)$ *is uncountable;*
- L *is exponential;*
- $\mathrm{Pref}(L)$ *is exponential.*

If the language L is not regular, then only the sufficient conditions of Proposition 5.3.9 and Proposition 5.3.10 hold true.

Proposition 5.3.12. *Let $\mathcal{A}$ be a (possibly infinite) deterministic automaton accepting a language L. If there exist in $\mathcal{A}$ at least two distinct cycles $(p_1,\ldots,p_r,p_1)$ and $(q_1,\ldots,q_s,q_1)$, with $r,s\geq 2$, starting from the same accessible and coaccessible state $p_1=q_1$, then $\mathrm{Adh}(L)$ is uncountably infinite and L is exponential.*

Proposition 5.3.13. *Let $\mathcal{A}$ be a (possibly infinite) deterministic automaton accepting a language L. If there exist in $\mathcal{A}$ at least two distinct cycles $(p_1,\ldots,p_r,p_1)$ and $(q_1,\ldots,q_s,q_1)$, with $r,s\geq 2$, starting from the same accessible state $p_1=q_1$ and such that each of them contains at least a final state, then L_∞ is uncountably infinite and L is exponential.*

There exist non-regular exponential languages L with an uncountable associated ω-language L_∞, and consequently also with an uncountable adherence $\mathrm{Adh}(L)$, that are recognized by deterministic automata without distinct cycles satisfying the conditions of Proposition 5.3.12. For instance, see Example 5.7.3 of Section 5.7 regarding the $\frac{3}{2}$-number system. Note that the corresponding trim minimal automaton depicted in Figure 5.7 has an infinite number of final states. In fact, by considering automata having a finite set of final states, we recover the necessary condition of Proposition 5.3.10.

Proposition 5.3.14. *Let L be a language recognized by a (possibly infinite) deterministic automaton $\mathcal{A}$ having a finite set of final states. The ω-language L_∞ is uncountably infinite if and only if there exist in $\mathcal{A}$ at least two distinct cycles $(p_1,\ldots,p_r,p_1)$ and $(q_1,\ldots,q_s,q_1)$, with $r,s\geq 2$, starting* *from the same accessible state $p_1 = q_1$ and such that each of them contains at least a final state.*

Proof. In view of Proposition 5.3.13, we only have to show that the condition is necessary. We let q_0 denote the initial state of $\mathcal{A}$. Since there is only a finite number of final states, if w belongs to L_∞, then there exist a final state f and infinitely many integers n such that we have $q_0 \cdot w[0, n-1] = f$. If the automaton $\mathcal{A}$ does not contain such distinct cycles, then this implies that any word in L_∞ is of the form xy^ω, where x, y are finite words. Since there is a countable number of such words, L_∞ would be a countable set. By contraposition, the conclusion follows. □

Corollary 5.3.15. *Let L be a language recognized by a (possibly infinite) deterministic automaton $\mathcal{A}$ having a finite set of final states. If the ω-language L_∞ is uncountably infinite, then L is exponential.*

Remark 5.3.16. Any deterministic automaton recognizing a non-regular prefix-closed language has an infinite number of final states. This is because, in such an automaton, all coaccessible states are final.

There exist exponential (and prefix-closed) languages L with a countable, and even finite, adherence $\mathrm{Adh}(L)$. We give an example of such a language.

Example 5.3.17. Let us consider

$$L = \{w \in \{a, b\}^* \mid \exists u \in \{a, b\}^* : \; w = a^{\lfloor |w|/2 \rfloor} u\}.$$

The minimal automaton of this language is depicted in Figure 5.2. Note that this automaton contains no distinct cycles as in Proposition 5.3.12. We have

$$\mathbf{u}_L(n) = \begin{cases} 2^{\frac{n}{2}}, & \text{if } n \equiv 0 \pmod 2; \\ 2^{\frac{n+1}{2}}, & \text{if } n \equiv 1 \pmod 2. \end{cases}$$

Also, we have $\mathrm{Adh}(L) = L_\infty = \{a^\omega\}$.

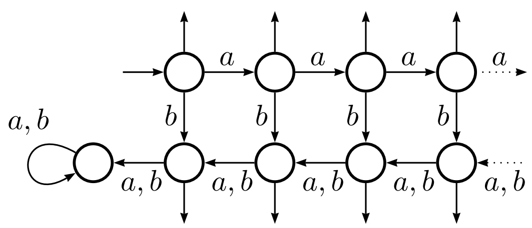

Figure 5.2. The minimal automaton of L.

5.4. Representation of Real Numbers

In the framework of [**LR02**] a real number is represented as a limit of a sequence of words in a regular language L. Observe that in this context, thanks to Lemma 5.3.4, the set of possible representations is $\mathrm{Adh}(L)$. Therefore one could consider abstract numeration systems built on the prefix-closure of the numeration language instead of the one built on the numeration language itself; see also Remarks 5.4.4 and 5.4.5 below. This point of view is relevant if we compare it with the framework of the standard integer base numeration systems. In these systems the numeration language is of the form

$$\mathcal{L}_b = \{1, 2, \ldots, b-1\}\{0, 1, \ldots, b-1\}^* \cup \{\varepsilon\}, \text{ for } b \geq 2,$$

which is, of course, a prefix-closed language. Note that this is also the case for non-standard numeration systems like β-numeration systems and substitutive numeration systems [**DT89, Lot02**]. Adopting this new framework, we will only consider abstract numeration systems built on prefix-closed languages. Therefore, to represent real numbers, we will no longer distinguish abstract numeration systems built on two distinct languages L and M having the same prefix-closure $\mathrm{Pref}(L) = \mathrm{Pref}(M)$.

Let $S = (L, \Sigma, <)$ be a generalized abstract numeration system built on a prefix-closed language L. Let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be an accessible deterministic automaton recognizing L. We fix these notation once and for all throughout this section. Furthermore, we make the following three assumptions:

Hypotheses.

(H1) The set $\mathrm{Adh}(L)$ is uncountable;
(H2) $\forall w \in \Sigma^*$, $\exists r_w \geq 0$, $\lim_{n\to+\infty} \frac{\mathbf{u}_{q_0 \cdot w}(n-|w|)}{\mathbf{v}_{q_0}(n)} = r_w$;
(H3) $\forall w \in \mathrm{Adh}(L)$, $\lim_{\ell\to+\infty} r_{w[0,\ell-1]} = 0$.

Observe that, for all $w \notin \mathrm{Center}(L)$, we have $r_w = 0$. Also, note that, since we have assumed that L is a prefix-closed language, from Lemma 5.3.8 we have $\mathrm{Adh}(L) = L_\infty$.

Notation. We set $r_0 = r_\varepsilon$ and

$$s_0 = 1 - r_0 = \lim_{n\to+\infty} \frac{\mathbf{v}_{q_0}(n-1)}{\mathbf{v}_{q_0}(n)}.$$

By convention, for any word w and any state q, we set $w[0, n] = \varepsilon$ and $\mathbf{v}_q(n) = 0$ for $n < 0$.

Remark 5.4.1. In [**LR02**] the authors considered regular languages L with uncountably infinite $\mathrm{Adh}(L)$ such that, for each state q of a DFA recognizing L, either L_q is finite, or we have $\mathbf{u}_q(n) \sim P_q(n)\theta_q^n$ $(n \to +\infty)$, with $P_q \in \mathbb{R}[x]$ and $\theta_q \geq 1$. One can note that such languages satisfy the hypotheses (H1), (H2) and (H3) above. For all states q and all $\ell \in \mathbb{N}$, it can be shown that we have

$$\lim_{n\to+\infty} \frac{\mathbf{u}_q(n-\ell)}{\mathbf{v}_{q_0}(n)} = \frac{(\theta_{q_0}-1)\, a_q}{\theta_{q_0}^{\ell+1}}$$

with $\theta_{q_0} > 1$ and $a_q := \lim_{n\to+\infty} \frac{\mathbf{u}_q(n)}{\mathbf{u}_{q_0}(n)}$. Since there is only a finite number of states, this is sufficient to verify our assumptions. Also, note that, for the integer base numeration systems, the three hypotheses are directly satisfied.

We shall represent real numbers by infinite words w in $\mathrm{Adh}(L)$ by considering the corresponding numerical limit

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S(w[0,n-1])}{\mathbf{v}_{q_0}(n)}. \tag{34}$$

Our aim is to show that, for all $w \in \mathrm{Adh}(L)$, the limit (34) exists; this is Proposition 5.4.10.

Remark 5.4.2. This way of representing integers generalizes the integer base case. The decimal representation of $\frac{11}{13}$ is $0.(846153)^\omega$. It is obtained by considering the following consecutive approximations:

$$\frac{8}{10}, \frac{84}{100}, \frac{846}{1000}, \frac{8461}{10000}, \frac{84615}{100000}, \ldots.$$

Note that, for all integers $b \geq 2$ and all $n \in \mathbb{N}$, we have $\mathbf{v}_{\mathcal{L}_b}(n) = b^n$. Thus, the denominator and the numerator of the nth fraction correspond to $\mathbf{v}_{\mathcal{L}_{10}}(n)$ and to the numerical value in base 10 of the prefix of length n of the infinite word $(846153)^\omega$ respectively. The binary representation of $\frac{11}{13}$ is $0.(110110001001)^\omega$. It is obtained by considering the following consecutive approximations:

$$\frac{1}{2}, \frac{3}{4} = \frac{6}{8}, \frac{13}{16}, \frac{27}{32} = \frac{54}{64} = \frac{108}{128} = \frac{216}{256}, \frac{433}{512} = 0.845703125, \ldots.$$

Again, the nth denominator of this sequence of approximations is $\mathbf{v}_{\mathcal{L}_2}(n)$ and the corresponding numerator is the numerical value in base 2 of the prefix of length n of the infinite word $(110110001001)^\omega$. For instance, we have $108 = \mathrm{val}_{U_2}(1101100)$.

Note that this also generalizes the formalism of β-numeration systems. Take a real number $\beta > 1$. With the same notation as in Remark 3.2.4 on page 51, remember that the numeration alphabet is $\Delta_\beta = [\![0, \lceil\beta\rceil - 1]\!]$. Hence we have $\mathbf{v}_{\Delta_\beta}(n) = \lceil\beta\rceil^n$ for all $n \in \mathbb{N}$. Next, let $(c_i)_{i\geq 1}$ be a β-representation of a real number $x \in [0,1]$, that is, a sequence $(c_i)_{i\geq 1}$ that satisfies

$$x = \sum_{i=1}^{+\infty} c_i\,\beta^{-i}, \text{ with } c_i \in \Delta_\beta \ \forall i \in \mathbb{N}\setminus\{0\}.$$

Define $\mathrm{val}_\beta(w) = \sum_{i=0}^{|w|-1} w[i]\,\beta^{|w|-i-1}$ for all $w \in \Delta_\beta^*$. Again, the numerical value of x can be obtained by considering the consecutive approximations:

$$\frac{\mathrm{val}_\beta(c_1)}{\beta}, \frac{\mathrm{val}_\beta(c_1c_2)}{\beta^2}, \frac{\mathrm{val}_\beta(c_1c_2c_3)}{\beta^3}, \ldots.$$

In the next remark and in Example 5.7.2 we will make use of the following well-known result from [**Bou07**, Ch. V.4].

Proposition 5.4.3. *Let $(x_n)_{n\geq 0}$ and $(y_n)_{n\geq 0}$ be two sequences of non-negative real numbers. If the series $\sum_{n=0}^{+\infty} x_n$ is not bounded, then we have*

$$x_n \sim y_n \Rightarrow \sum_{i=0}^{n} x_i \sim \sum_{i=0}^{n} y_i \ (n \to +\infty). \tag{35}$$

If the series $\sum_{n=0}^{+\infty} x_n$ is bounded, then we have

$$x_n \sim y_n \Rightarrow \sum_{i=n}^{+\infty} x_i \sim \sum_{i=n}^{+\infty} y_i \ (n \to +\infty). \tag{36}$$

Remark 5.4.4. If the abstract numeration system is built on a language that is not prefix-closed, we cannot guarantee that the limit (34) exists. For instance, consider the abstract numeration system built on the language L in Example 5.2.1, which is not prefix-closed. The sequences $((ab)^n)_{n\geq 0}$ and $((ab)^n a)_{n\geq 0}$ of words in L converge to the same infinite word $(ab)^\omega$ but the corresponding numerical sequences do not converge to the same real number. More precisely, using the notation of Example 5.2.1, we show

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S((ab)^n)}{\mathbf{v}_0(2n)} = \frac{3}{4} \quad\text{and}\quad \lim_{n\to+\infty} \frac{\mathrm{val}_S((ab)^n a)}{\mathbf{v}_0(2n+1)} = \frac{3}{5}. \tag{37}$$

Consequently, the limit

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S((ab)^\omega[0,n-1])}{\mathbf{v}_0(n)}$$

does not exist. Let us prove (37). This fact essentially comes from the staircase behavior of $(\mathbf{u}_0(n))_{n\geq 0}$:

$$\mathbf{u}_0(n) = \begin{cases} \binom{n}{\frac{n}{2}}, & \text{if } n \equiv 0 \pmod 2; \\ 2\binom{n}{\frac{n-1}{2}}, & \text{if } n \equiv 1 \pmod 2. \end{cases}$$

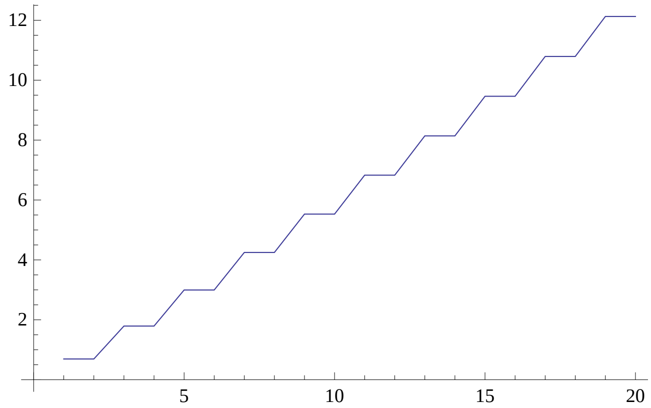

FIGURE 5.3. The first 20 values of $\mathbf{u}_0(n)$ up to a logarithmic scaling.

In particular this implies that $\lim_{n\to+\infty} \frac{\mathbf{v}_0(n-1)}{\mathbf{v}_0(n)}$ does not exist. Using Stirling's formula[3], we obtain

$$\mathbf{u}_0(2n) \sim \frac{1}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \quad \text{and} \quad \mathbf{u}_0(2n+1) \sim \frac{4}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \ (n \to +\infty). \qquad (38)$$

Then, using (35) and $\sum_{i=0}^{n} i^{-1/2} 4^i \sim \frac{1}{3} n^{-1/2} 4^{n+1}$ $(n \to +\infty)$, it follows

$$\mathbf{v}_0(2n) \sim \frac{8}{3\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \quad \text{and} \quad \mathbf{v}_0(2n-1) \sim \frac{5}{3\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \ (n \to +\infty). \qquad (39)$$

Hence we obtain

$$\lim_{n\to+\infty} \frac{\mathbf{v}_0(2n-1)}{\mathbf{v}_0(2n)} = \frac{5}{8} \quad \text{and} \quad \lim_{n\to+\infty} \frac{\mathbf{v}_0(2n)}{\mathbf{v}_0(2n+1)} = \frac{2}{5}.$$

From Proposition 5.2.2 we then obtain

$$\frac{\mathrm{val}_S((ab)^n)}{\mathbf{v}_0(2n)} = \frac{\mathbf{v}_0(2n-1)}{\mathbf{v}_0(2n)} + \frac{\sum_{i=0}^{n-1} \mathbf{u}_2(2i)}{\mathbf{v}_0(2n)},$$
$$\frac{\mathrm{val}_S((ab)^n a)}{\mathbf{v}_0(2n+1)} = \frac{\mathbf{v}_0(2n)}{\mathbf{v}_0(2n+1)} + \frac{\sum_{i=0}^{n-1} \mathbf{u}_2(2i+1)}{\mathbf{v}_0(2n+1)}$$

for all $n \in \mathbb{N}$. Observe that we have

$$\mathbf{u}_2(2n) = \binom{2n}{n-1} \sim \mathbf{u}_0(2n) \ (n \to +\infty),$$
$$\mathbf{u}_2(2n+1) = \binom{2n+1}{n} + \binom{2n+1}{n-1} \sim \mathbf{u}_0(2n+1) \ (n \to +\infty).$$

Therefore, in view of (35), (38), and (39), it follows

$$\lim_{n\to+\infty} \frac{\sum_{i=0}^{n-1} \mathbf{u}_2(2i)}{\mathbf{v}_0(2n)} = \frac{1}{8} \quad \text{and} \quad \lim_{n\to+\infty} \frac{\sum_{i=0}^{n-1} \mathbf{u}_2(2i+1)}{\mathbf{v}_0(2n+1)} = \frac{1}{5}$$

[3] $n! \sim \sqrt{2\pi}\, e^{-n}\, n^{n+\frac{1}{2}}$ $(n \to +\infty)$

and we obtain the two limits of (37).

Remark 5.4.5. Considering prefix-closed languages not only avoids numerical convergence problems as in Remark 5.4.4 but also makes it possible to get rid of problems arising from languages L such that there are infinitely many non-negative integers n for which we have $L \cap \Sigma^n = \emptyset$ as discussed in [**LR02**, Remark 4].

Definition 5.4.6. If w is an infinite word in $\mathrm{Adh}(L)$ and x is a real number satisfying

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S\,(w[0,n-1])}{\mathbf{v}_{q_0}(n)} = x,$$

then we say that w is an *S-representation* of x.

Example 5.4.7. Let us consider the generalized abstract numeration system built on the prefix-closure of the Dyck language that will be described in Example 5.7.2. In Table 5.1 we give some numerical approximations. Further on, we will be able to compute

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S((aab)^\omega[0,n-1])}{\mathbf{v}_{q_0}(n)} = \frac{39}{49} = 0.79592\cdots.$$

w	$\mathrm{val}_S(w)$	$\mathbf{v}_{q_0}(\lvert w\rvert)$	$\frac{\mathrm{val}_S(w)}{\mathbf{v}_{q_0}(\lvert w\rvert)}$
a	1	2	0.50000
aa	2	4	0.50000
aab	5	7	0.71429
$aaba$	9	13	0.69231
$aabaa$	17	23	0.73913
$aabaab$	32	43	0.74419
$aabaaba$	60	78	0.76923
$aabaabaa$	112	148	0.75676
$aabaabaab$	213	274	0.77737
$aabaabaaba$	404	526	0.76806
$aabaabaabaa$	771	988	0.78036
$aabaabaabaab$	1479	1912	0.77354
$aabaabaabaaba$	2841	3628	0.78308
$aabaabaabaabaa$	5486	7060	0.77705
$aabaabaabaabaab$	10591	13495	0.78481

TABLE 5.1. Some numerical approximations.

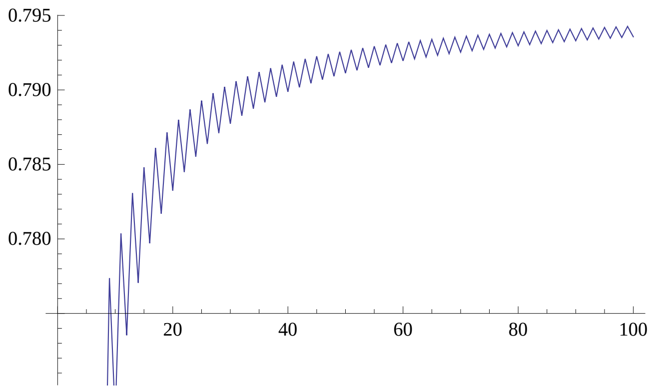

FIGURE 5.4. The first 100 values of $\frac{\text{val}_S((aab)^\omega[0,n-1])}{\mathbf{v}_{q_0}(n)}$.

Note that, for all $w \in \text{Adh}(L)$ and $n \in \mathbb{N}$, we have

$$\text{val}_S(w[0, n-1]) \in [\mathbf{v}_{q_0}(n-1), \mathbf{v}_{q_0}(n) - 1].$$

Therefore the represented real numbers x must belong to the interval $[s_0, 1]$. Let us divide $[s_0, 1]$ into some subintervals I_y for all prefixes y of infinitely many words in L, *i.e.*, for all $y \in \text{Center}(L)$. First, observe that, for all $\ell \in \mathbb{N}$, if the integer $n \geq \ell$ is large enough, then all words in L of length n have a prefix in $\text{Center}(L) \cap \Sigma^\ell$. Therefore, for all $y \in \text{Center}(L)$ and for every large enough integer $n \geq |y|$, the number of words of length n in L having a prefix of length $|y|$ lexicographically less than y is equal to

$$\sum_{\substack{x<y \\ x \in \text{Center}(L)\cap\Sigma^{|y|}}} \mathbf{u}_{q_0 \cdot x}(n - |y|).$$

Consequently, in the latter conditions, the S-value of the first word of length n in L having y as a prefix is equal to

$$\mathbf{v}_{q_0}(n-1) + \sum_{\substack{x<y \\ x \in \text{Center}(L)\cap\Sigma^{|y|}}} \mathbf{u}_{q_0 \cdot x}(n - |y|).$$

Now, for each $y \in \text{Center}(L)$ and each integer $n \geq |y|$, define

$$\alpha_{y,n} = \frac{\mathbf{v}_{q_0}(n-1)}{\mathbf{v}_{q_0}(n)} + \sum_{\substack{x<y \\ x \in \text{Center}(L)\cap\Sigma^{|y|}}} \frac{\mathbf{u}_{q_0 \cdot x}(n - |y|)}{\mathbf{v}_{q_0}(n)}$$

and

$$I_{y,n} = \left[\alpha_{y,n}, \alpha_{y,n} + \frac{\mathbf{u}_{q_0 \cdot y}(n - |y|)}{\mathbf{v}_{q_0}(n)}\right].$$

In view of Hypothesis (H2), for all $y \in \mathrm{Center}(L)$, we may define

$$\alpha_y = \lim_{n\to+\infty} \alpha_{y,n} = s_0 + \sum_{\substack{x<y \\ x\in \mathrm{Center}(L)\cap\Sigma^{|y|}}} r_x.$$

Then, for all $y \in \mathrm{Center}(L)$, we let I_y denote the limit interval:

$$I_y = \lim_{n\to+\infty} I_{y,n} = [\alpha_y, \alpha_y + r_y].$$

Furthermore, we set $I_y = \emptyset$ for all $y \in L \setminus \mathrm{Center}(L)$. From [**LR02**] we know that

$$\forall \ell \in \mathbb{N},\ [s_0, 1] = \bigcup_{y\in \mathrm{Center}(L)\cap\Sigma^\ell} I_y$$

and

$$\forall\, y, z \in \Sigma^*,\ I_{yz} \subseteq I_y. \tag{40}$$

More precisely, if $a_1, \ldots, a_k$ are the letters in Σ and if the order on Σ is given by $a_1 < \cdots < a_k$, then, for all $y \in \mathrm{Center}(L)$ and all $j \in [\![1, k]\!]$ such that we have $ya_j \in \mathrm{Center}(L)$, one has

$$I_{ya_j} = \left[\alpha_y + \sum_{i=1}^{j-1} r_{ya_i},\ \alpha_y + \sum_{i=1}^{j} r_{ya_i}\right]. \tag{41}$$

Remark 5.4.8. Take $y \in \mathrm{Center}(L)$ and let w be a word in Σ^* having y as a prefix and such that $|w|$ is large enough so that every word of length $|w|$ has a prefix in $\mathrm{Center}(L) \cap \Sigma^{|y|}$. Then, from Proposition 5.2.2, it follows

$$\begin{aligned}\mathrm{val}_S(w) = \mathbf{v}_{q_0}(|w|-1) + &\sum_{\substack{x<y \\ x\in \mathrm{Center}(L)\cap\Sigma^{|y|}}} \mathbf{u}_{q_0\cdot x}(|w|-|y|) \\ &+ \sum_{i=|y|}^{|w|-1} \sum_{a<w[i]} \mathbf{u}_{q_0\cdot w[0,i-1]a}(|w|-i-1).\end{aligned} \tag{42}$$

Lemma 5.4.9. *For all $w \in \mathrm{Adh}(L)$, the following limit exists*

$$\lim_{\ell\to+\infty} \alpha_{w[0,\ell-1]}.$$

Proof. First, note that, for all non-negative integers $\ell \in \mathbb{N}$, we have $w[0, \ell-1] \in \mathrm{Center}(L)$ from the definition of the center of a language. On the one hand, observe that (40) implies $\alpha_{w[0,\ell-1]} \le \alpha_{w[0,\ell]}$ for all $\ell \in \mathbb{N}$. On the other hand, we also have $\alpha_{w[0,\ell-1]} \le 1$ for all $\ell \in \mathbb{N}$. Hence $(\alpha_{w[0,\ell-1]})_{\ell\ge0}$ is a bounded and non-decreasing sequence. So it must converge. □

Notation. For all $w \in \mathrm{Adh}(L)$, we define $\alpha_w = \lim_{\ell\to+\infty} \alpha_{w[0,\ell-1]}$.

Note that we have $\alpha_w \geq \alpha_{w[0,\ell-1]}$ for all $\ell \in \mathbb{N}$. Now, we prove that the limit (34) exists.

Proposition 5.4.10. *For all $w \in \mathrm{Adh}(L)$, we have*

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S(w[0,n-1])}{\mathbf{v}_{q_0}(n)} = \alpha_w.$$

PROOF. Take $w \in \mathrm{Adh}(L)$. In view of Remark 5.4.8 and from the definition of the $\alpha_{y,n}$'s, we obtain that, for all $\ell \in \mathbb{N}$ and all integers $n \geq \ell$ large enough with respect to ℓ, we have

$$\alpha_{w[0,\ell-1],n} \leq \frac{\mathrm{val}_S(w[0,n-1])}{\mathbf{v}_{q_0}(n)} < \alpha_{w[0,\ell-1],n} + \frac{\mathbf{u}_{q_0\cdot w[0,\ell-1]}(n-\ell)}{\mathbf{v}_{q_0}(n)}. \tag{43}$$

Now, choose any $\varepsilon > 0$. From the definition of the α_y's and from Hypothesis (H2), we know that, for all $\ell \in \mathbb{N}$, there exists $N(\ell) \geq \ell$ such that, for all integers $n \geq N(\ell)$, we have

$$\alpha_{w[0,\ell-1]} - \frac{\varepsilon}{2} < \frac{\mathrm{val}_S(w[0,n-1])}{\mathbf{v}_{q_0}(n)} < \alpha_{w[0,\ell-1]} + r_{w[0,\ell-1]} + \frac{\varepsilon}{2}.$$

From Hypothesis (H3) and Lemma 5.4.9 there also exists $k \in \mathbb{N}$ such that, for all integers $\ell \geq k$, we have

$$r_{w[0,\ell-1]} < \frac{\varepsilon}{2} \quad \text{and} \quad 0 \leq \alpha_w - \alpha_{w[0,\ell-1]} < \frac{\varepsilon}{2}.$$

It follows that, for all integers $n \geq N(k)$, we have

$$\alpha_w - \varepsilon < \alpha_{w[0,k-1]} - \frac{\varepsilon}{2} < \frac{\mathrm{val}_S(w[0,n-1])}{\mathbf{v}_{q_0}(n)} < \alpha_w + \varepsilon.$$

Hence the proposition is proved. □

Now, we are ready to introduce the S-value of an infinite word in $\mathrm{Adh}(L)$. Thus, the definition of the map $\mathrm{val}_S \colon L \to \mathbb{N}$ extends to $L \cup \mathrm{Adh}(L)$.

Definition 5.4.11. The application $\mathrm{val}_S \colon \mathrm{Adh}(L) \to [s_0, 1] \colon w \mapsto \alpha_w$ is called the *numerical S-value function.* For all words w in $\mathrm{Adh}(L)$, we say that $\mathrm{val}_S(w)$ is the *numerical S-value* (or simply the *S-value*) of w.

Proposition 5.4.12. *For all $w, z \in \mathrm{Adh}(L)$ satisfying $w \leq_{lex} z$, one has $\mathrm{val}_S(w) \leq \mathrm{val}_S(z)$.*

PROOF. Let w, z be infinite words in $\mathrm{Adh}(L)$ satisfying $w <_{\mathrm{lex}} z$. We deduce from (41) that if we have $k = \inf\{i \in \mathbb{N} \mid w[i] < z[i]\}$, then, for all integers $\ell \geq k$, we have $\alpha_{w[0,\ell-1]} \leq \alpha_{z[0,\ell-1]}$. By letting ℓ tend to infinity in both sides of this inequality, we have the conclusion. □

Now, let us recall a result from [**BB97**].

Lemma 5.4.13. [**BB97**] *If K is an infinite language over a totally ordered alphabet, then* $\mathrm{Adh}(K)$ *contains a minimal element with respect to the lexicographical order.*

This leads to the following definition.

Definition 5.4.14. For all $y \in \mathrm{Center}(L)$, we let m_y (resp. M_y) denote the *minimal* (resp. *maximal*) *word* in $\mathrm{Adh}(L)$ with respect to the lexicographical order having y as a prefix.

Note that, for all $y \in \mathrm{Center}(L)$, we have $m_y = yu$ (resp. $M_y = yv$), where u (resp. v) is the minimal (resp. maximal) word in $\mathrm{Adh}(y^{-1}L)$ with respect to the lexicographical order.

Example 5.4.15. Continuing Example 5.2.1, we have $m_{aab} = aaba^\omega$ and $M_{aab} = aab^\omega$. Further on, we will see that for the Dyck language, we have $m_{aab} = aaba^\omega$ and $M_{aab} = aabb(ab)^\omega$; see Example 5.7.2.

For each $y \in \mathrm{Center}(L)$, these minimal and maximal words m_y and M_y respectively are representations of the limit points of the corresponding interval I_y, as shown by the following lemma.

Lemma 5.4.16. *For all $y \in \mathrm{Center}(L)$, we have*

$$\mathrm{val}_S(m_y) = \alpha_y \quad \textit{and} \quad \mathrm{val}_S(M_y) = \alpha_y + r_y.$$

Proof. Take $y \in \mathrm{Center}(L)$. From (41) it follows that, for all integers $\ell \geq |y|$, we have $\alpha_{m_y[0,\ell-1]} = \alpha_y$ and $\alpha_{M_y[0,\ell-1]} + r_{M_y[0,\ell-1]} = \alpha_y + r_y$. Therefore we obtain that, for all integers $\ell \geq |y|$, we have

$$\begin{aligned} \alpha_y \leq \mathrm{val}_S(m_y) &\leq \alpha_y + r_{m_y[0,\ell-1]};\\ \alpha_y + r_y - r_{M_y[0,\ell-1]} \leq \mathrm{val}_S(M_y) &\leq \alpha_y + r_y. \end{aligned}$$

We conclude by using Hypothesis (H3). □

Proposition 5.4.17. *The S-value function* $\mathrm{val}_S\colon \mathrm{Adh}(L) \to [s_0, 1]$ *is uniformly continuous.*

Proof. Take $w, z \in \mathrm{Adh}(L)$. Assume $d(w, z) = 2^{-\ell}$. We thus have $w[0, \ell-1] = z[0, \ell-1]$. Then, from Lemma 5.4.16, the S-values $\mathrm{val}_S(w)$ and

$\operatorname{val}_S(z)$ belong to $I_{w[0,\ell-1]}$. Hence, from Hypothesis (H3), we have

$$|\operatorname{val}_S(w) - \operatorname{val}_S(z)| \le r_{w[0,\ell-1]} \to 0 \text{ as } \ell \to +\infty.$$

This finishes the proof. □

Using Lemma 5.4.16, we are able to give an expression of the numerical S-value of a word in $\operatorname{Adh}(L)$.

Proposition 5.4.18. *For all $w \in \operatorname{Adh}(L)$, we have*

$$\operatorname{val}_S(w) = s_0 + \sum_{i=0}^{+\infty} \sum_{a<w[i]} r_{w[0,i-1]a}.$$

Proof. Take $w \in \operatorname{Adh}(L)$. From the definition of the adherence, the prefixes $w[0,\ell-1]$ belong to $\operatorname{Center}(L)$ for all $\ell \in \mathbb{N}$. Recall that we have $r_w = 0$ for all $w \notin \operatorname{Center}(L)$. Therefore, using (41), we obtain

$$\begin{aligned} \alpha_{w[0,\ell-1]} &= s_0 + \sum_{\substack{x<w[0,\ell-1] \\ x\in \operatorname{Center}(L)\cap\Sigma^\ell}} r_x \\ &= s_0 + \sum_{i=0}^{\ell-1} \sum_{a<w[i]} \sum_{|y|=\ell-i-1} r_{w[0,i-1]ay} \\ &= s_0 + \sum_{i=0}^{\ell-1} \sum_{a<w[i]} r_{w[0,i-1]a} \end{aligned}$$

for all $\ell \in \mathbb{N}$. From Lemma 5.4.9 and by letting ℓ tend to infinity in the latter equality, we obtain the expected result. □

5.5. Link with Converging Sequences of Words

The following proposition links together the framework of [**LR02**], where converging sequences of words were mainly considered, and the framework that has been developed in the previous section to represent real numbers.

Proposition 5.5.1. *Let $S = (\operatorname{Pref}(L), \Sigma, <)$ be a generalized abstract numeration system built on the prefix-closure of a language L over an alphabet Σ and let $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ be a deterministic automaton recognizing $\operatorname{Pref}(L)$. Assume that $\operatorname{Pref}(L)$ satisfies Hypotheses (H1), (H2), and (H3) above. Then, for all sequences of words $(w^{(n)})_{n\ge 0} \in L^{\mathbb{N}}$ converging to w, we have*

$$\lim_{n\to+\infty} \frac{\operatorname{val}_S(w^{(n)})}{\mathbf{v}_{q_0}(|w^{(n)}|)} = \operatorname{val}_S(w).$$

Proof. Let $(w^{(n)})_{n\ge 0} \in L^{\mathbb{N}}$ be a sequence of words converging to w. Thanks to Lemma 5.3.4, this implies $\mathrm{Pref}(w) \subseteq \mathrm{Pref}(L)$, that is, w belongs to $\mathrm{Adh}(L)$. For any $\ell \in \mathbb{N}$, there exists $N(\ell)$ such that, for all integers $n \ge N(\ell)$, we have $w^{(n)}[0,\ell-1] = w[0,\ell-1]$ and any word in L of length greater than or equal to $|w^{(n)}|$ has a prefix in $\mathrm{Center}(L)\cap\Sigma^{\ell}$. Then, in view of Remark 5.4.8 and (43), for all $\ell \in \mathbb{N}$ and for all integers $n \ge N(\ell)$, we obtain

$$\left|\frac{\mathrm{val}_S\left(w[0,|w^{(n)}|-1]\right)}{\mathbf{v}_{q_0}(|w^{(n)}|)} - \frac{\mathrm{val}_S\left(w^{(n)}\right)}{\mathbf{v}_{q_0}(|w^{(n)}|)}\right| \le \frac{\mathbf{u}_{q_0\cdot w[0,\ell-1]}(|w^{(n)}|-\ell)}{\mathbf{v}_{q_0}(|w^{(n)}|)}.$$

Choose any $\varepsilon > 0$. From Hypothesis (H2), for all $\ell \in \mathbb{N}$, there exists $M(\ell)$ such that, for all integers $n \ge M(\ell)$, we have

$$\frac{\mathbf{u}_{q_0\cdot w[0,\ell-1]}(|w^{(n)}|-\ell)}{\mathbf{v}_{q_0}(|w^{(n)}|)} < r_{w[0,\ell-1]} + \frac{\varepsilon}{2}.$$

From Hypothesis (H3) there exists $k \in \mathbb{N}$ such that, for all integers $\ell \ge k$, we have $r_{w[0,\ell-1]} < \frac{\varepsilon}{2}$. Then, for all integers $n \ge \max\{N(k), M(k)\})$, we have

$$\left|\frac{\mathrm{val}_S\left(w[0,|w^{(n)}|-1]\right)}{\mathbf{v}_{q_0}(|w^{(n)}|)} - \frac{\mathrm{val}_S\left(w^{(n)}\right)}{\mathbf{v}_{q_0}(|w^{(n)}|)}\right| < \varepsilon.$$

This completes the proof. □

5.6. Ultimately Periodic Representations

In this section we use results from [**BB97**] to obtain syntactical properties about the representations of the endpoints of the intervals I_y, for $y \in \mathrm{Center}(L)$, whenever L is a language satisfying the assumptions of Section 5.4.

Proposition 5.6.1. [**BB97**] *If L is an infinite context-free language over a totally ordered alphabet, then the minimal word of* $\mathrm{Adh}(L)$ *is ultimately periodic and can be effectively computed.*

Of course, this result can be adapted to the case of the maximal word of the adherence of a language. Thus we have the following corollary.

Corollary 5.6.2. *Let L be an infinite context-free language over a totally ordered alphabet. For all $y \in \mathrm{Center}(L)$, the infinite words m_y and M_y are ultimately periodic and can be effectively computed. In particular, the languages* $\mathrm{Pref}(m_y)$ *and* $\mathrm{Pref}(M_y)$ *are regular.*

PROOF. Take $y \in \text{Center}(L)$. We only show the "minimal" case, the other one being similar. By definition, m_y is the minimal word of the adherence $\text{Adh}(y(y^{-1}L))$, where $y(y^{-1}L)$ is the language of the words in L beginning with y. Since L is an infinite context-free language and y belongs to $\text{Center}(L)$, $y(y^{-1}L)$ is an infinite context-free language as well. Then, from Proposition 5.6.1, m_y is an ultimately periodic infinite word. For the second part of the statement, observe that if an infinite word w is ultimately periodic, then its prefix language $\text{Pref}(w)$ is regular. □

Note that, in general, there exist ultimately periodic representations that are not endpoints of any interval I_y, for $y \in \text{Center}(L)$. For instance, in the integer base 10 numeration system, the U_{10}-representation of $\frac{1}{3}$ is $0.33333\cdots$ and $\frac{1}{3}$ is not the endpoint of any interval $\left[\frac{k}{10^\ell}, \frac{k+1}{10^\ell}\right]$, with ℓ in $\mathbb{N} \setminus \{0\}$ and k in $[\![0, 10^\ell - 1]\!]$.

Let us also mention the following interesting result from [**BB97**]. Remember that the minimal language $\text{Min}_<(L)$ of a language L over a totally ordered alphabet is the set of the minimal words of each length with respect to the induced lexicographical order; see Definition 1.2.4 on page 4.

Proposition 5.6.3. [**BB97**] *Assume that L is an infinite language that satisfies $L = \text{Center}(L)$. Then we have $Min_<(L) = \text{Pref}(m_\varepsilon)$.*

Corollary 5.6.4. *If L is a context-free language satisfying $L = \text{Center}(L)$, then $Min_<(L)$ is regular.*

To finish this section, it is probably worth mentioning here the main theorem from [**BB97**], which has become well-known today.

Theorem 5.6.5. [**BB97**] *If L is a context-free language, then so $Min_<(L)$ is.*

Again note that all these results can easily be adapted to the maximal language of a language L.

5.7. Applications

In this section we apply our techniques to three examples to represent real numbers in situations that were not settled in [**LR02**]. The first one demonstrates how it can be easier to consider the prefix-closure of a language instead of the language itself; compare this example with Remark 5.4.4.

Example 5.7.1. Once again, consider the language

$$L = \{w \in \{a,b\}^* \mid ||w|_a - |w|_b| \leq 1\}$$

from Example 5.2.1. This language is not prefix-closed. Observe that we have $\mathrm{Pref}(L) = \{a, b\}^*$, which is, of course, a regular language. It is easy to check that for the abstract numeration system $S = (\mathrm{Pref}(L), \{a, b\}, a < b)$, the hypotheses (H1), (H2) and (H3) are satisfied. More precisely, for all words w in $\{a, b\}^*$, we have $r_w = 2^{-|w|-1}$. Using the same notation as in Example 5.2.1, we have

$$\lim_{n\to+\infty} \frac{\mathbf{v}_0(n-1)}{\mathbf{v}_0(n)} = \frac{1}{2}.$$

Therefore we represent the interval $[\frac{1}{2}, 1]$. We have $\mathrm{Center}(L) = \{a, b\}^*$ and, for all $\ell \in \mathbb{N}$, the intervals corresponding to words of length ℓ are exactly the intervals

$$\left[\frac{1}{2} + \frac{k}{2^{\ell+1}}, \frac{1}{2} + \frac{k+1}{2^{\ell+1}}\right], \text{ for } k \in [\![0, 2^\ell - 1]\!].$$

The second example illustrates the case of a non-regular language with a non-regular prefix-closure.

Example 5.7.2. The *Dyck language* is the language

$$D = \{w \in \{a, b\}^* \,|\, |w|_a = |w|_b \text{ and } \forall u \in \mathrm{Pref}(w),\ |u|_a \geq |u|_b\}$$

of well-parenthesized words over two letters. Its minimal automaton $\mathcal{A}_D$ is depicted in Figure 5.5. For each $m \in \mathbb{N}$, we let d_m denote the state $d_m = (a^m)^{-1}D = \{w \in \{a, b\}^* \,|\, a^m w \in D\}$ of $\mathcal{A}_D$ and we set $d_{-1} = \emptyset$, so that the set of states of $\mathcal{A}_D$ is $Q_D = \{d_m \,|\, m \in \mathbb{N}\} \cup \{d_{-1}\}$; see Definition 1.3.6. Note that, in Figure 5.5, the states d_m are simply denoted by m.

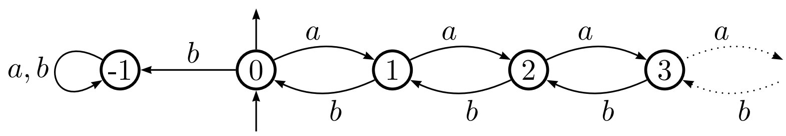

FIGURE 5.5. The minimal automaton of D.

It was proved in [**LG08**] that, for all $m, n \in \mathbb{N}$, we have

$$\mathbf{u}_{d_m}(n) = \begin{cases} 0, & \text{if } n < m \quad \text{or} \quad m \not\equiv n \pmod 2; \\ \frac{m+1}{n+1}\binom{n+1}{\frac{n-m}{2}}, & \text{if } n \geq m \quad \text{and} \quad m \equiv n \pmod 2. \end{cases}$$

From Stirling's formula we obtain that, for all $m \in \mathbb{N}$, we have

$$\mathbf{u}_{d_{2m}}(2n) \sim \frac{2m+1}{\sqrt{\pi}} n^{-\frac{3}{2}} 4^n \ (n \to +\infty); \tag{44}$$

$$\mathbf{u}_{d_{2m+1}}(2n+1) \sim \frac{2(2m+2)}{\sqrt{\pi}} n^{-\frac{3}{2}} 4^n \ (n \to +\infty). \tag{45}$$

The Dyck language is not prefix-closed. Hence we consider the generalized abstract numeration system $S = (\text{Pref}(D), \{a, b\}, a < b)$ built on the language

$$\text{Pref}(D) = \{w \in \{a, b\}^* \,|\, \forall u \in \text{Pref}(w),\ |u|_a \geq |u|_b\}$$

of the prefixes of Dyck words. The minimal automaton $\mathcal{A}_{\text{Pref}(D)}$ of $\text{Pref}(D)$ is depicted in Figure 5.6. Since the graphs of the minimal automaton $\mathcal{A}_{\text{Pref}(D)}$ of $\text{Pref}(D)$ and the minimal automaton $\mathcal{A}_D$ of D are nearly the same, we rename the states of $\mathcal{A}_{\text{Pref}(D)}$ by p_m so that the set of states of $\mathcal{A}_{\text{Pref}(D)}$ is $Q_{\text{Pref}(D)} = \{p_m \,|\, m \in \mathbb{N}\} \cup \{p_{-1}\}$. Formally, for all $m \in \mathbb{N}$, we have $p_m = (a^m)^{-1}\,\text{Pref}(D) = \{w \in \{a, b\}^* \,|\, a^m w \in \text{Pref}(D)\}$ and we set $p_{-1} = \emptyset$. Hence the $\mathbf{u}_{d_m}$'s denote the complexity functions of $\mathcal{A}_D$ and the $\mathbf{u}_{p_m}$'s denote the complexity functions of $\mathcal{A}_{\text{Pref}(D)}$. Once again, in Figure 5.6, the states p_m are simply denoted by m. From Proposition 5.3.12 the set $\text{Adh}(\text{Pref}(D)) = \text{Adh}(D)$ is uncountable and Hypothesis (H1) is satisfied.

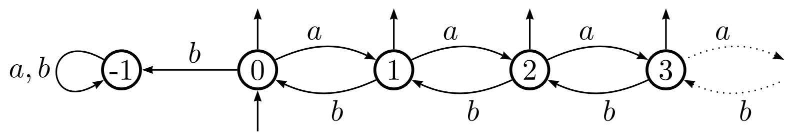

FIGURE 5.6. The minimal automaton of $\text{Pref}(D)$.

Observe that, for all $m, n \in \mathbb{N}$, we have

$$\mathbf{u}_{p_m}(n) = \begin{cases} 2^n, & \text{if } n \leq m; \\ 2\mathbf{u}_{p_m}(n-1) - \mathbf{u}_{d_m}(n-1), & \text{if } n > m. \end{cases}$$

Hence we obtain

$$\forall\, m, n \in \mathbb{N},\ \mathbf{u}_{p_m}(n) = 2^n - \sum_{i=m}^{n-1} \mathbf{u}_{d_m}(i)\, 2^{n-i-1}.$$

We claim that, for all $m \in \mathbb{N}$, we have

$$\mathbf{u}_{p_m}(2n) \sim \frac{m+1}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \ (n \to +\infty); \tag{46}$$

$$\mathbf{u}_{p_m}(2n+1) \sim \mathbf{v}_{p_m}(2n) \sim \frac{2(m+1)}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \ (n \to +\infty); \tag{47}$$

$$\mathbf{v}_{p_m}(2n+1) \sim \frac{4(m+1)}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \ (n \to +\infty). \tag{48}$$

We will only prove (46), since the same techniques can be applied to obtain (47) and (48). First, let us show that, for all $m \in \mathbb{N}$, we have

$$\sum_{i=m}^{+\infty} \mathbf{u}_{d_{2m}}(2i)\, 4^{-i} = 2 \qquad \text{and} \qquad \sum_{i=m}^{+\infty} \mathbf{u}_{d_{2m+1}}(2i+1)\, 4^{-i} = 4. \tag{49}$$

We will only compute the first sum; the second one can be treated in a similar way. In view of (36) and (44), for all $m \in \mathbb{N}$, we have

$$\sum_{i=n}^{+\infty} \mathbf{u}_{d_{2m}}(2i)4^{-i} \sim \frac{2m+1}{\sqrt{\pi}} \sum_{i=n}^{+\infty} i^{-\frac{3}{2}} \ (n \to +\infty).$$

This shows that the series

$$\sum_{i=m}^{+\infty} \mathbf{u}_{d_{2m}}(2i)4^{-i}$$

is convergent. Consequently, for all $m \in \mathbb{N}$, the series

$$\sum_{i=m}^{+\infty} \mathbf{u}_{d_{2m}}(2i)\, z^{i}$$

is uniformly convergent over $\{z \in \mathbb{C} \,|\, |z| \leq \frac{1}{4}\}$ because, for all integers p and q satisfying $q \geq p \geq m$, we have

$$\sup_{|z| \leq \frac{1}{4}} \left| \sum_{i=p}^{q} \mathbf{u}_{d_{2m}}(2i)\, z^{i} \right| \leq \sum_{i=p}^{q} \mathbf{u}_{d_{2m}}(2i)4^{-i}.$$

Then, observe that, for all $m \in \mathbb{N}$ and all integers $i \geq m$ satisfying $i \equiv m \pmod 2$, $\mathbf{u}_{d_m}(i)$ is given by

$$\begin{aligned} &\mathrm{Card}\{w^{(0)}bw^{(1)}b\cdots bw^{(m)} \mid \forall j \in [\![0, m]\!],\, w^{(j)} \in D,\ \sum_{j=0}^{m} |w^{(j)}| = i - m\} \\ &= \sum_{\ell_0+\cdots+\ell_m = \frac{i-m}{2}} \left(\prod_{j=0}^{m} \mathcal{C}_{\ell_j} \right) = \left[z^{\frac{i-m}{2}} \right] \left(\sum_{n=0}^{+\infty} \mathcal{C}_n\, z^n \right)^{m+1}, \end{aligned}$$

where $\mathcal{C}_n := \mathbf{u}_{d_0}(2n) = \frac{1}{2n+1}\binom{2n+1}{n}$ is the nth Catalan number and $[z^n]f$ denotes the coefficient of z^n in the power series f. It is well known (for instance, see [**GKP94, Lan03**]) that we have

$$\sum_{n=0}^{+\infty} \mathcal{C}_n\, z^n = \frac{1 - \sqrt{1-4z}}{2z}, \text{ for} |z| < \frac{1}{4}.$$

Hence, for all $m \in \mathbb{N}$ and for all $z \in \mathbb{C}$ satisfying $|z| < \frac{1}{4}$, we obtain

$$\sum_{i=m}^{+\infty} \mathbf{u}_{d_{2m}}(2i)\, z^{i} = z^m \left(\sum_{n=0}^{+\infty} \mathcal{C}_n\, z^n \right)^{2m+1} = \frac{(1 - \sqrt{1-4z})^{2m+1}}{2 \cdot 4^m\, z^{m+1}}.$$

Therefore we obtain the desired first sum of (49) by letting z tend to $\frac{1}{4}$ in the latter formula. Now, let us come back to (46). For all $m, n \in \mathbb{N}$

satisfying $m < n$, we have

$$\mathbf{u}_{p_{2m}}(2n) = 4^n - \frac{1}{2}\sum_{i=m}^{n-1} \mathbf{u}_{d_{2m}}(2i)\,4^{n-i} = \frac{1}{2}\,4^n \sum_{i=n}^{+\infty} \mathbf{u}_{d_{2m}}(2i)\,4^{-i}$$

and

$$\mathbf{u}_{p_{2m+1}}(2n) = 4^n - \frac{1}{4}\sum_{i=m}^{n-1} \mathbf{u}_{d_{2m+1}}(2i+1)\,4^{n-i} = \frac{1}{4}\,4^n \sum_{i=n}^{+\infty} \mathbf{u}_{d_{2m+1}}(2i+1)\,4^{-i}.$$

Then, using $\sum_{i=n}^{+\infty} i^{-\frac{3}{2}} \sim 2n^{-\frac{1}{2}}$ $(n \to +\infty)$, we obtain that, for all $m \in \mathbb{N}$, we have

$$\mathbf{u}_{p_{2m}}(2n) \sim \frac{2m+1}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \quad \text{and} \quad \mathbf{u}_{p_{2m+1}}(2n) \sim \frac{2m+2}{\sqrt{\pi}} n^{-\frac{1}{2}} 4^n \ (n \to +\infty),$$

proving (46).

Now, let us verify that the language $\mathrm{Pref}(D)$ satisfies our three hypotheses. From the previous reasoning we obtain

$$\forall m, \ell \in \mathbb{N}, \ \lim_{n \to +\infty} \frac{\mathbf{u}_{p_m}(n-\ell)}{\mathbf{v}_{p_0}(n)} = (m+1)\,2^{-\ell-1}.$$

In this example we have $\mathrm{Center}(D) = \mathrm{Center}(\mathrm{Pref}(D)) = \mathrm{Pref}(D)$. For all $w \in \mathrm{Pref}(D)$, we have $r_w = (m(w)+1)\,2^{-|w|-1}$ where $m(w) \in \mathbb{N}$ is defined by $p_0 \cdot w = p_{m(w)}$ and, for all $w \notin \mathrm{Pref}(D)$, we have $r_w = 0$. Hence Hypothesis (H2) is satisfied. Now, take an infinite word w in $\mathrm{Adh}(D)$. Observe that we have $m(w[0, \ell-1]) \leq \ell$ for all $\ell \in \mathbb{N}$. Therefore we have

$$r_{w[0,\ell-1]} \leq (\ell+1)2^{-\ell-1} \to 0 \text{ as } \ell \to +\infty.$$

Hence Hypothesis (H3) is also satisfied.

Since we have

$$\lim_{n \to +\infty} \frac{\mathbf{v}_{p_0}(n-1)}{\mathbf{v}_{p_0}(n)} = \frac{1}{2},$$

we represent the interval $[\frac{1}{2}, 1]$. Any word in $\mathrm{Pref}(D)$ begins with a. So we have $I_a = [\frac{1}{2}, 1]$. We have $\mathrm{Center}(D) \cap \{a,b\}^2 = \{aa, ab\}$ and I_a is partitioned into two subintervals:

$$I_{aa} = \left[\frac{1}{2}, \frac{7}{8}\right] \quad \text{and} \quad I_{ab} = \left[\frac{7}{8}, 1\right].$$

Then, we have $\mathrm{Center}(D) \cap \{a,b\}^3 = \{aaa, aab, aba\}$. Therefore we obtain $I_{ab} = I_{aba}$ and I_{aa} is partitioned into two new subintervals:

$$I_{aaa} = \left[\frac{1}{2}, \frac{3}{4}\right], \ I_{aab} = \left[\frac{3}{4}, \frac{7}{8}\right], \ I_{aba} = \left[\frac{7}{8}, 1\right].$$

Then, we have $\text{Center}(D) \cap \{a,b\}^4 = \{aaaa, aaab, aaba, aabb, , abaa, abab\}$ and we obtain

$$I_{aaaa} = \left[\frac{1}{2}, \frac{21}{32}\right],\ I_{aaab} = \left[\frac{21}{32}, \frac{3}{4}\right],\ I_{aaba} = \left[\frac{3}{4}, \frac{27}{32}\right],$$
$$I_{aabb} = \left[\frac{27}{32}, \frac{7}{8}\right],\ I_{abaa} = \left[\frac{7}{8}, \frac{31}{32}\right],\ I_{abab} = \left[\frac{31}{32}, 1\right].$$

As stated in Corollary 5.6.2, since the language D is context-free, the representations of the endpoints of the intervals I_y are ultimately periodic. For all $x \in [\frac{1}{2}, 1]$, let Q_x denote the set of the representations of x. We have $Q_{1/2} = \{a^\omega\}$ and $Q_1 = \{(ab)^\omega\}$. Now, let $x \in (1/2, 1)$ be an endpoint of some interval. This means that we have $x = \inf I_w = \sup I_z$ for some w, z in $\text{Center}(D) \cap \{a,b\}^\ell$ with $\ell \in \mathbb{N}$. We obtain $Q_x = \{\bar{w}(ab)^\omega, za^\omega\}$, where $\bar{w}$ is defined to be the smallest Dyck word having w as a prefix. Also, note that we have $\text{Pref}(m_\varepsilon) = a^* = \text{Min}_<(\text{Pref}(D))$ and $\text{Pref}(M_\varepsilon) = (ab)^* \cup (ab)^*a = \text{Max}_<(\text{Pref}(D))$, which are both regular languages. The latter observation is relevant to Corollary 5.6.2 and Proposition 5.6.3.

The third example illustrates the case of a generalized abstract numeration system generating endpoints of the intervals I_y which never have ultimately periodic S-representations. It also shows that our methods for representing real numbers generalize those involved in the $\frac{3}{2}$-number system and, by extension, in the rational base number systems as well.

Example 5.7.3. Consider the language $L_{3/2}$ recognized by the deterministic automaton $\mathcal{A} = (\mathbb{N} \cup \{-1\}, \{0,1,2\}, \delta, 0, \mathbb{N})$ where the transition function δ is defined as follows: $\delta(n,a) = \frac{1}{2}(3n+a)$ if $n \in \mathbb{N}$ and $a \in \{0,1,2\}$ satisfy $\frac{1}{2}(3n+a) \in \mathbb{N}\backslash\{0\}$ and $\delta(n,a) = -1$ otherwise. This language was introduced and extensively studied in [**AFS08**]. In particular, it has been shown that the automaton $\mathcal{A}$ is the minimal automaton of $L_{3/2}$, that $L_{3/2}$ is a non-context-free prefix-closed language and that $\text{Adh}(L_{3/2})$ is uncountable. Furthermore, no element in $\text{Adh}(L_{3/2})$ is ultimately periodic. The corresponding trim minimal automaton is depicted in Figure 5.7, where all states are final.

Let $(G_n)_{n\geq 0}$ be the sequence of integers defined by

$$G_0 = 1 \text{ and } \forall n \in \mathbb{N},\ G_{n+1} = \left\lceil \frac{3}{2} G_n \right\rceil.$$

It was shown in [**AFS08**] that we have $G_n = \lfloor K \left(\frac{3}{2}\right)^n \rfloor$, where $K := K(3) = 1.6222705\cdots$ is the constant discussed in [**OW91, HH97, Ste03**]. From [**AFS08**] again, we find

$$\mathbf{u}_0(0) = 1 \text{ and } \forall n \in \mathbb{N},\ \mathbf{u}_0(n+1) = G_{n+1} - G_n.$$

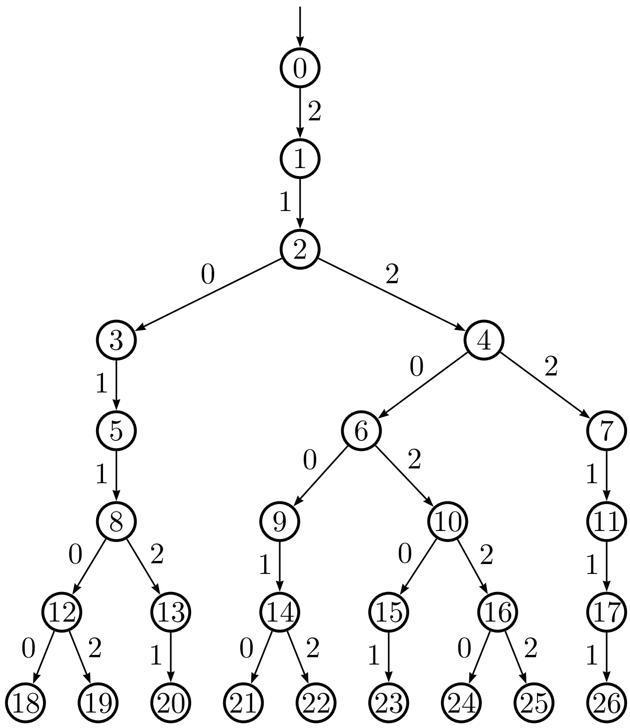

FIGURE 5.7. The first few levels of the trim minimal automaton of $L_{\frac{3}{2}}$.

Hence we obtain

$$\forall n \in \mathbb{N},\ \mathbf{v}_0(n) = G_n = \left\lfloor K \left(\frac{3}{2}\right)^n \right\rfloor .$$

Consider the generalized abstract numeration system

$$S = (L_{\frac{3}{2}}, \{0, 1, 2\}, 0 < 1 < 2)$$

built on this language. Clearly, we have $\delta(0, w) = \text{val}_S(w)$ for all w in $L_{3/2}$. Furthermore, from [**AFS08**], we know that, for all $w \in L_{3/2}$, we have

$$\text{val}_S(w) = \frac{1}{2} \sum_{i=0}^{|w|-1} w[i] \left(\frac{3}{2}\right)^{|w|-i-1} .$$

Consequently, for all $w \in \text{Adh}(L_{3/2})$, we obtain

$$\text{val}_S(w) = \lim_{n \to +\infty} \frac{\text{val}_S(w[0, n-1])}{\mathbf{v}_0(n)} = \frac{1}{3K} \sum_{i=0}^{+\infty} w[i] \left(\frac{3}{2}\right)^{-i} .$$

Now, let us verify whether $L_{3/2}$ satisfies Hypothesis (H2) and (H3). For all y in $\text{Center}(L_{3/2})$, remember that M_y (resp. m_y) denotes the maximal (resp. minimal) word in $\text{Adh}(L_{3/2})$ with respect to the lexicographical order having y as a prefix. Clearly, in this example, we have $L_{3/2} = \mathrm{Center}(L_{3/2})$. For all $y \in L_{3/2}$ and for all integers $n \geq |y|$, we have

$$\mathbf{u}_{0\cdot y}(n-|y|) = \mathrm{val}_S(M_y[0,n-1]) - \mathrm{val}_S(m_y[0,n-1]) + 1.$$

Therefore we obtain

$$\forall y \in L_{3/2},\ \frac{\mathbf{u}_{0\cdot y}(n-|y|)}{\mathbf{v}_0(n)} \sim \frac{\frac{1}{2}\sum_{i=0}^{n-1}(M_y[i]-m_y[i])(\frac{3}{2})^{n-i-1}}{K\left(\frac{3}{2}\right)^n}\ (n \to +\infty).$$

It follows

$$\begin{aligned}
\forall y \in L_{3/2},\ r_y &= \lim_{n\to+\infty} \frac{\mathbf{u}_{0\cdot y}(n-|y|)}{\mathbf{v}_0(n)} \\
&= \lim_{n\to+\infty} \frac{1}{3K} \sum_{i=|y|}^{n-1} (M_y[i]-m_y[i]) \left(\frac{3}{2}\right)^{-i} \\
&= \frac{1}{3K}\left(\frac{3}{2}\right)^{-|y|} \sum_{i=0}^{+\infty}(M_y[i+|y|]-m_y[i+|y|])\left(\frac{3}{2}\right)^{-i} \geq 0.
\end{aligned}$$

Note that this is consistent with Lemma 5.4.16. Furthermore, we have $r_y = 0$ for all $y \notin L_{3/2}$. Hence Hypothesis (H2) is satisfied. For all $y \in L_{3/2}$, since we have $M_y[i] - m_y[i] \leq 2$ for all $i \in \mathbb{N}$, we obtain

$$r_y \leq \frac{2}{K}\left(\frac{3}{2}\right)^{-|y|} \to 0 \text{ as } |y| \to +\infty.$$

Therefore we have $\lim_{\ell\to+\infty} w[0,\ell-1] = 0$ for all infinite words w in the adherence $\mathrm{Adh}(L_{3/2})$. Consequently, Hypothesis (H3) is also satisfied.

In this case, since the language $L_{3/2}$ is not context-free, we cannot use Corollary 5.6.2 to deduce any syntactical properties about the representations of the endpoints of the intervals. Actually, effectively computing the intervals I_y seems difficult to undertake. Nevertheless, for any given $y \in L_{3/2}$, it is possible to approximate the corresponding interval I_y as close as desired.

Perspectives

In this dissertation we studied a few questions related to abstract numeration systems, which were introduced by M. Rigo and P. Lecomte in [**LR01**]. We are happy to have slightly furthered the knowledge in this fascinating research topic. Still, a great deal of work remains to be achieved in this field, which is good news for future research.

In the context of the chapters developed in this dissertation, some natural questions arise. We will point out some of them below.

Multiplication by a Constant for Polynomial Languages

1. If S is an abstract numeration system built on any polynomial language L such that its counting function $\mathbf{u}_L(n)$ is $\Theta(n)$, then can we prove or refute that multiplication by a constant λ preserves S-recognizability if and only if λ is an odd square. Otherwise stated, does Theorem 2.2.10 on page 29 extend to any polynomial language L such that its counting function $\mathbf{u}_L(n)$ is $\Theta(n)$.

2. If S is an abstract numeration system built on any polynomial language L such that its counting function $\mathbf{u}_L(n)$ is $\Theta(n^k)$, with $k \geq 2$, and if λ is any fixed integer, can we always find an S-recognizable set X of non-negative integers such that λX is not S-recognizable? In other words, provided that the previous problem is solved, does Theorem 2.6.1 on page 43 extend to any polynomial language?

Other Decision Problems

3. Consider the linear numeration system $U = (U_i)_{i \geq 0}$ defined by

$$U_0 = 1, \ U_1 = 2, \ \text{and} \ \forall i \in \mathbb{N}, \ U_{i+2} = 2U_{i+1} + 2U_i.$$

Since it is a "Pisot" numeration system, the numeration language must be regular [**FS96**]. Hence, Problem 2 on page 47 is valid in this case. It is easily shown that we have $N_U(2^k) = 1$ for all non-negative integers k, so that we obtain $N_U(m) \not\to +\infty$. Therefore our decision procedure given in Theorem 3.2.15 cannot be applied. However, J. Honkala's decision produre for the integer base numeration systems [**Hon86**] does not require the hypothesis $N_U(m) \to +\infty$. So can we adapt J. Honkala's method for such a sequence U?

4. More generally, if progress is achieved for the previous problem, can we extend Theorem 3.2.15 on page 56 to positional numeration systems U such that $\mathbb{N}$ is U-recognizable but no longer satisfying the condition $\lim_{m\to+\infty} N_U(m) = +\infty$? Otherwise stated, can we give a decision procedure for Problem 2 in general?

5. Even more generally, can we give a decision procedure for Problem 3 on page 78 in general, *i.e.* for any abstract numeration system? Recall that answering this question would solve the HD0L periodicity problem!

About Multidimensional S-automatic words

For the sake of clarity, we have written statements in the bidimensional case but each of them can be more generally stated in the d-dimensional case for any integer $d \geq 2$.

6. If S and T are two abstract numeration systems, one can easily adapt Definition 4.2.1 on page 94 to define the notion of (S, T)-automatic word, *i.e.*, we define

$$\forall m, n \in \mathbb{N},\ x_{m,n} = \tau(\delta(q_0, (\text{rep}_S(m), \text{rep}_T(n))^{\#})).$$

   Can these (S, T)-automatic words be characterized by using morphisms?

7. Peano enumeration of $\mathbb{N}^2$ is defined by

$$P\colon \mathbb{N}^2 \to \mathbb{N}\colon (m, n) \mapsto \frac{1}{2}(m+n)(m+n+1) + n.$$

   Hence, to any bidimensional word $(a_{m,n})_{m,n\geq 0}$ corresponds a unidimensional word $(b_\ell)_{\ell\geq 0}$ defined by $b_\ell = a_{m,n}$ for $\ell = P(m, n)$. In their book [**AS03**, Chapter 14] the authors proposed, as an exercise, to show that if $(a_{m,n})_{m,n\geq 0}$ is b-automatic, then $(b_\ell)_{\ell\geq 0}$ need not necessarily to be b-automatic too. Now, suppose that $(a_{m,n})_{m,n\geq 0}$ is S-automatic for some abstract numeration system S. Does this imply that $(b_\ell)_{\ell\geq 0}$ is T-automatic for some abstract numeration system T, and conversely? For example, if we consider the S-automatic bidimensional infinite word depicted in Figure 4.8 on page 95, then we obtain the unidimensional word

$$pqpqpqpsppqsspqpqqspp\cdots.$$

   Can we deduce that this word is T-automatic for some T? One could also consider other primitive recursive enumerations of $\mathbb{N}^2$.

8. For any integer $b \geq 2$, the *b-kernel* of an infinite word w is the set

$$\{(w_{b^e n+r})_{n\geq 0} \mid e \in \mathbb{N},\ r \in [\![0, k^e - 1]\!]\}.$$

It is well known that an infinite word is b-automatic if and only if its b-kernel is finite [**Cob72**]. The notion of kernel has been extended to the framework of abstract numeration systems by A. Maes and M. Rigo [**RM02**]. In this case, one refers to *S-kernels*. They proved that an infinite word is S-automatic if and only if its S-kernel is finite. Does this characterization hold in the multidimensional setting?

9. Multidimensional b-automatic words were characterized in logical terms; for instance, see [**BHMV94**]. Furthermore, the (multidimensional) recognizable sets within a "Pisot" numeration system $U = (U_i)_{i\geq 0}$ were characterized in terms of sets definable in the structure $\langle \mathbb{N}, +, V_U\rangle$, where we set $V_U(0) = U_0 = 1$ and, for any positive integer n, $V_U(n)$ is defined to be the smallest term U_i appearing in the greedy decomposition $n = \sum_{i=0}^{\ell} c_i U_i$ with a non-zero coefficient [**BH97**]. Can such characterizations be extended to the framework of abstract numeration systems? What would be the correct logical structure to consider in this extended case? Even in the unidimensional case, this question is still open.

10. Any bidimensional word $a = (a_{m,n})_{m,n\geq 0}$ over an alphabet Σ which is embedded into a finite field $\mathbb{F}_q$ can be associated with a formal power series:
$$F(a) = \sum_{m,n\geq 0} a_{m,n} x^m y^n \in \mathbb{F}_q[[x,y]].$$
In [**Sal87a, Sal87b**], generalizing Christol's theorem [**CKMFR80**] to the multidimensional setting, O. Salon proved that, if p is a prime number, then a bidimensional infinite word a over $[\![0, p-1]\!]$ is p-automatic if and only if the associated formal power series $F(a)$ is algebraic over the field of rational functions $\mathbb{F}_p(x,y)$, *i.e.*, if and only if there exist $\ell \in \mathbb{N} \setminus \{0\}$ and $P_0, P_1, \ldots, P_\ell \in \mathbb{F}_p[x,y]$, not all zero, such that we have
$$P_\ell(F(a))^\ell + \cdots + P_2(F(a))^2 + P_1 F(a) + P_0 = 0.$$
If a is an S-automatic bidimensional infinite word, can we derive some algebraic properties of $F(a)$, and conversely? Again, even in the unidimensional case, this question is still open.

11. The following result of P. Deligne [**Del84**] is widely known; also, see [**DL87, Sal87a, Sal87b**] for other proofs: if the double series
$$\sum_{m,n\geq 0} a_{m,n} x^m y^n \in \mathbb{F}_q[[x,y]]$$
is algebraic over $\mathbb{F}_q(x,y)$, then its diagonal
$$\sum_{m\geq 0} a_{m,m} x^m \in \mathbb{F}_q[[x]]$$

is algebraic over $\mathbb{F}_q(x)$. So, as a particular case of the previous question, if a is an S-automatic bidimensional infinite word over an alphabet Σ which is embedded into a finite field $\mathbb{F}_q$, can something be said about the diagonal $D(a) = \sum_{m \geq 0} a_{m,m} x^m \in \mathbb{F}_q[[x]]$?

Bibliography

[AB08] B. Adamczewski and J. Bell. Function fields in positive characteristic: expansions and Cobham's theorem. *J. Algebra*, 319(6):2337–2350, 2008.

[ABS04] P. Arnoux, V. Berthé, and A. Siegel. Two-dimensional iterated morphisms and discrete planes. *Theoret. Comput. Sci.*, 319(1-3):145–176, 2004.

[AFS08] S. Akiyama, Ch. Frougny, and J. Sakarovitch. Powers of rationals modulo 1 and rational base number systems. *Israel J. Math.*, 168:53–91, 2008.

[Ale04] B. Alexeev. Minimal DFA for testing divisibility. *J. Comput. Syst. Sci.*, 69(2):235–243, 2004.

[And74] A. Andreassian. Fibonacci sequences modulo m. *Fibonacci Quart.*, 12:51–64, 1974.

[APDS93] M. Andraşiu, Gh. Păun, J. Dassow, and A. Salomaa. Language-theoretic problems arising from Richelieu cryptosystems. *Theoret. Comput. Sci.*, 116(2):339–357, 1993.

[ARS09] J.-P. Allouche, N. Rampersad, and J. Shallit. Periodicity, repetitions, and orbits of an automatic sequence. *Theoret. Comput. Sci.*, 410:2795–2803, 2009.

[AS03] J.-P. Allouche and J. Shallit. *Automatic Sequences. Theory, Applications, Generalizations.* Cambridge University Press, Cambridge, 2003.

[Bak09] A. Baker. An introduction to p-adic numbers and p-adic analysis. Lecture Notes, Department of Mathematics, University of Glasgow, 2009.

[BB97] J. Berstel and L. Boasson. The set of minimal words of a context-free language is context-free. *J. Comput. System Sci.*, 55(3):477–488, 1997.

[BB07] B. Boigelot and J. Brusten. A generalization of Cobham's theorem to automata over real numbers. In *Automata, languages and programming*, volume 4596 of *Lecture Notes in Comput. Sci.*, pages 813–824. Springer, Berlin, 2007.

[BB09] B. Boigelot and J. Brusten. A generalization of Cobham's theorem to automata over real numbers. *Theoret. Comput. Sci.*, 410(18):1694–1703, 2009.

[BCFR09] J. Bell, E. Charlier, A. S. Fraenkel, and M. Rigo. A decision problem for ultimately periodic sets in non-standard numeration systems. *Internat. J. Algebra Comput.*, 19(6):809–839, 2009.

[Bel07] J. P. Bell. A generalization of Cobham's theorem for regular sequences. *Sém. Lothar. Combin.*, 54A:Art. B54Ap, 15 pp. (electronic), 2005/07.

[Ber70] C. Berge. *Graphes et hypergraphes.* Dunod, Paris, 1970. Monographies Universitaires de Mathématiques, No. 37.

[Ber79] J. Berstel. *Transductions and context-free languages*, volume 38 of *Leitfäden der Angewandten Mathematik und Mechanik [Guides to Applied Mathematics and Mechanics].* B. G. Teubner, Stuttgart, 1979.

[Bès97] A. Bès. Undecidable extensions of Büchi arithmetic and Cobham-Semënov theorem. *J. Symbolic Logic*, 62(4):1280–1296, 1997.

[Bès00] A. Bès. An extension of the Cobham-Semenov theorem. *J. Symbolic Logic*, 65(1):201–211, 2000.

[BH97] V. Bruyère and G. Hansel. Bertrand numeration systems and recognizability. *Theoret. Comput. Sci.*, 181(1):17–43, 1997. Latin American Theoretical Informatics (Valparaíso, 1995).

[BHMV94] V. Bruyère, G. Hansel, Ch. Michaux, and R. Villemaire. Logic and p-recognizable sets of integers. *Bull. Belg. Math. Soc. Simon Stevin*, 1(2):191–238, 1994. Journées Montoises (Mons, 1992).

[BJN94] P. B. Bhattacharya, S. K. Jain, and S. R. Nagpaul. *Basic Abstract Algebra*. Cambridge University Press, Cambridge, second edition, 1994.

[BM89] A. Bertrand-Mathis. Comment écrire les nombres entiers dans une base qui n'est pas entière. *Acta Math. Hungar.*, 54(3-4):237–241, 1989.

[BN80] L. Boasson and M. Nivat. Adherences of languages. *J. Comput. System Sci.*, 20(3):285–309, 1980.

[Bou07] N. Bourbaki. *Fonctions d'une variable réelle.* Springer Berlin Heidelberg, 2007.

[BR88] J. Berstel and Ch. Reutenauer. *Rational series and their languages*, volume 12 of *EATCS Monographs on Theoretical Computer Science*. Springer-Verlag, Berlin, 1988.

[BR09] J. Berstel and Ch. Reutenauer. *Noncommutative rational series with applications*. New edition of *Rational series and their languages* to be published in Cambridge University Press, 2009.

[Bru70] G. Bruckner. Fibonacci sequence modulo a prime $p \equiv 3 \pmod 4$. *Fibonacci Quart.*, 8(2):217–220, 1970.

[Cat74] P. A. Catlin. A lower bound for the period of the Fibonacci series modulo m. *Fibonacci Quart.*, 12:349–350, 1974.

[CKMFR80] G. Christol, T. Kamae, M. Mendès France, and G. Rauzy. Suites algébriques, automates et substitutions. *Bull. Soc. Math. France*, 108(4):401–419, 1980.

[CKR] E. Charlier, T. Kärki, and M. Rigo. Multidimensional generalized automatic sequences and shape-symmetric morphic words. To appear in *Discrete Math.*

[CKR09] E. Charlier, T. Kärki, and M. Rigo. A characterization of multidimensional S-automatic sequences. In *Numeration: Mathematics and Computer Science*, volume 1 of *Actes des rencontres du CIRM*, pages 23–28. CIRM, Marseille, 2009.

[CLGR] E. Charlier, M. Le Gonidec, and M. Rigo. Representing reals numbers in a generalized numeration system. Submitted.

[Cob68] A. Cobham. On the Hartmanis-Staerns problem for a class of tag machines. In *IEEE Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory*, pages 51–60. IEEE Computer Society, 1968.

[Cob69] A. Cobham. On the base-dependence of sets of numbers recognizable by finite automata. *Math. Systems Theory*, 3:186–192, 1969.

[Cob72] A. Cobham. Uniform tag sequences. *Math. Systems Theory*, 6:164–192, 1972.

[CR08] E. Charlier and M. Rigo. A decision problem for ultimately periodic sets in non-standard numeration systems. In *Mathematical Foundations of Computer Science 2008*, volume 5162 of *Lecture Notes in Comput. Sci.*, pages 241–252. Springer, Berlin, 2008.

[CRS08] E. Charlier, M. Rigo, and W. Steiner. Abstract numeration systems on bounded languages and multiplication by a constant. *Integers*, 8:A35, 19, 2008.

[CT02] O. Carton and W. Thomas. The monadic theory of morphic infinite words and generalizations. *Inform. and Comput.*, 176(1):51–76, 2002.

[Dar01] G. Darvasi. More on the distribution of the Fibonacci numbers modulo $5c$. *Acta Math. Acad. Paedagog. Nyházi. (N.S.)*, 17(1):19–24 (electronic), 2001.

[Del84] P. Deligne. Intégration sur un cycle évanescent. *Invent. Math.*, 76(1):129–143, 1984.

[DFNR] E. Duchêne, A. S. Fraenkel, R. Nowakowski, and M. Rigo. Extensions and restrictions of Wythoff's game preserving Wythoff's sequence as set of P positions. To appear in J. Combinat. Theory Ser. A.

[DL77] M. J. De Leon. Fibonacci primitive roots and the period of the Fibonacci numbers modulo p. *Fibonacci Quart.*, 15(4):353–355, 1977.

[DL87] J. Denef and L. Lipshitz. Algebraic powers series and diagonals. *J. Number Theory*, 26(1):46–67, 1987.

[DT89] J.-M. Dumont and A. Thomas. Systèmes de numération et fonctions fractales relatifs aux substitutions. *Theoret. Comput. Sci.*, 65(2):153–169, 1989.

[Dur98] F. Durand. Sur les ensembles d'entiers reconnaissables. *J. Théor. Nombres Bordeaux*, 10(1):65–84, 1998.

[Dur02a] F. Durand. Combinatorial and dynamical study of substitutions around the theorem of Cobham. In *Dynamics and randomness (Santiago, 2000)*, volume 7 of *Nonlinear Phenom. Complex Systems*, pages 53–94. Kluwer Acad. Publ., Dordrecht, 2002.

[Dur02b] F. Durand. A theorem of Cobham for non-primitive substitutions. *Acta Arith.*, 104(3):225–241, 2002.

[Dur08] F. Durand. Cobham-Semenov theorem and $\mathbb{N}^d$-subshifts. *Theoret. Comput. Sci.*, 391(1-2):20–38, 2008.

[Ehr89] A. Ehrlich. On the periods of the Fibonacci sequence modulo m. *Fibonacci Quart.*, 27(1):11–13, 1989.

[Eil74] S. Eilenberg. *Automata, Languages, and Machines*, volume A. Academic Press, New York, 1974. Pure and Applied Mathematics, Vol. 58.

[Eng31] H. T. Engstrom. On sequences defined by linear recurrence relations. *Trans. Amer. Math. Soc.*, 33(1):210–218, 1931.

[EvdPSW03] G. Everest, A. van der Poorten, I. Shparlinski, and T. Ward. *Recurrence Sequences*, volume 104 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2003.

[Fab94] S. Fabre. Une généralisation du théorème de Cobham. *Acta Arith.*, 67(3):197–208, 1994.

[Fat06] P. Fatou. Séries trigonométriques et séries de Taylor. *Acta Math.*, 30(1):335–400, 1906.

[Fra85] A. S. Fraenkel. Systems of numeration. *Amer. Math. Monthly*, 92(2):105–114, 1985.

[Fro92] Ch. Frougny. Representations of numbers and finite automata. *Math. Systems Theory*, 25(1):37–60, 1992.

[Fro97] Ch. Frougny. On the sequentiality of the successor function. *Inform. and Comput.*, 139(1):17–38, 1997.

[FS96] Ch. Frougny and B. Solomyak. On the representation of integers in linear numeration systems. In *Ergodic theory of Z_d actions (Warwick, 1993–1994)*, volume 228 of *London Math. Soc. Lecture Note Ser.*, pages 345–368. Cambridge Univ. Press, Cambridge, 1996.

[GH64] S. Ginsburg and Spanier E. H. Bounded ALGOL-like languages. *Trans. Amer. Math. Soc.*, 113:333–368, 1964.

[GKP94] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics. A Foundation for Computer Science*. Addison-Wesley Publishing Company, Reading, MA, second edition, 1994.

[Gou97] F. Q. Gouvêa. *p-adic Numbers, An introduction*. Universitext. Springer-Verlag, Berlin, second edition, 1997.

[GR01] Ch. Godsil and G. Royle. *Algebraic graph theory*, volume 207 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2001.

[Han98] G. Hansel. Systèmes de numération indépendants et syndéticité. *Theoret. Comput. Sci.*, 204(1-2):119–130, 1998.

[Her04] T. Herendi. Uniform distribution of linear recurring sequences modulo prime powers. *Finite Fields Appl.*, 10(1):1–23, 2004.

[HH97] L. Halbeisen and N. Hungerbühler. The Josephus problem. *J. Théor. Nombres Bordeaux*, 9(2):303–318, 1997.

[HL86] T. Harju and M. Linna. On the periodicity of morphisms on free monoids. *RAIRO Inform. Théor. Appl.*, 20(1):47–54, 1986.

[Hol98] M. Hollander. Greedy numeration systems and regularity. *Theory Comput. Syst.*, 31(2):111–133, 1998.

[Hon86] J. Honkala. A decision method for the recognizability of sets defined by number systems. *RAIRO Inform. Theor. Appl.*, 20(4):395–403, 1986.

[HR04] J. Honkala and M. Rigo. Decidability questions related to abstract numeration systems. *Discrete Math.*, 285(1-3):329–333, 2004.

[HS03] G. Hansel and T. Safer. Vers un théorème de Cobham pour les entiers de Gauss. *Bull. Belg. Math. Soc. Simon Stevin*, 10(suppl.):723–735, 2003.

[Kat68] G. Katona. A theorem on finite sets. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 187–207. Academic Press, New York, 1968.

[Kle56] S. C. Kleene. Representation of events in nerve nets and finite automata. In *Automata studies*, Annals of mathematics studies, no. 34, pages 3–41. Princeton University Press, Princeton, N. J., 1956.

[KMR$^+$09] D. Krieger, A. Miller, N. Rampersad, B. Ravikumar, and J. Shallit. Decimations of languages and state complexity. *Theoret. Comput. Sci.*, 410(24-25):2401–2409, 2009.

[Kob84] N. Koblitz. *p-adic numbers, p-adic analysis, and zeta-functions*, volume 58 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1984.

[KS72] L. Kuipers and J. S. Shiue. A distribution property of the sequence of Fibonacci numbers. *Fibonacci Quart.*, 10(4):375–376, 392, 1972.

[Lan03] S. K. Lando. *Lectures on generating functions*, volume 23 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2003. Translated from the 2002 Russian original by the author.

[Lan04] S. Lang. *Algèbre : cours et exercices*. Dunod, Paris, third edition, 2004.

[Leh64] D. H. Lehmer. The machine tools of combinatorics. In E. F. Beckenbach, editor, *Applied Combinatorial Mathematics*, University of California Engineering and Physical Sciences Extension Series, pages 5–31. John Wiley and Sons, Inc., New York-London-Sydney, 1964.

[Ler05] J. Leroux. A polynomial time Presburger criterion and synthesis for number decision diagrams. In *20th IEEE Symposium on Logic in Computer Science*, pages 147–156. IEEE Computer Society, Chicago, IL, USA, 2005.

[LG08] M. Le Gonidec. On complexity of infinite words associated with generalized Dyck languages. *Theoret. Comput. Sci.*, 407(1-3):117–133, 2008.

[LMSF96] J. S. Lew, L. B. Morales, and A. Sánchez-Flores. Diagonal polynomials for small dimensions. *Math. Systems Theory*, 29(3):305–310, 1996.

[Lor95] N. Loraud. β-shift, systèmes de numération et automates. *J. Théor. Nombres Bordeaux*, 7(2):473–498, 1995.

[Lot02] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.

[LR01] P. B. A. Lecomte and M. Rigo. Numeration systems on a regular language. *Theory Comput. Syst.*, 34(1):27–44, 2001.

[LR02] P. Lecomte and M. Rigo. On the representation of real numbers using regular languages. *Theory Comput. Syst.*, 35(1):13–38, 2002.

[Mae98] A. Maes. Decidability of the first-order theory of $\langle \mathbb{N}; <, P\rangle$ for morphic predicates P. Preprint 9806, Inst. für Informatik und Praktische Math., Christian-Albrechts-Univ. Kiel, 1998.

[Mae99a] A. Maes. An automata-theoretic decidability proof for first-order theory of $\langle \mathbb{N}, <, P\rangle$ with morphic predicate P. *J. Autom. Lang. Comb.*, 4(3):229–245, 1999. Journées Montoises d'Informatique Théorique (Mons, 1998).

[Mae99b] A. Maes. *Morphic Predicates and Applications to the Decidability of Arithmetic Theories*. PhD thesis, Univ. Mons-Hainaut, 1999.

[Mam61] S. E. Mamangakis. Remarks on the Fibonacci series modulo m. *Amer. Math. Monthly*, 68:648–649, 1961.

[Muc03] An. A. Muchnick. The definable criterion for definability in Presburger arithmetic and its applications. *Theoret. Comput. Sci.*, 290(3):1433–1444, 2003.

[MV96] Ch. Michaux and R. Villemaire. Presburger arithmetic and recognizability of sets of natural numbers by automata: new proofs of Cobham's and Semenov's theorems. *Ann. Pure Appl. Logic*, 77(3):251–277, 1996.

[Nar84] W. Narkiewicz. *Uniform distribution of sequences of integers in residue classes*, volume 1087 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1984.

[Ner58] A. Nerode. Linear automaton transformations. *Proc. Amer. Math. Soc.*, 9:541–544, 1958.

[Niv78] M. Nivat. Sur les ensembles de mots infinis engendrés par une grammaire algébrique. *RAIRO Inform. Théor.*, 12(3):259–278, 1978.

[NR07] S. Nicolay and M. Rigo. About the frequency of letters in generalized automatic sequences. *Theoret. Comp. Sci.*, 374(1-3):25–40, 2007.

[OW91] A. M. Odlyzko and H. S. Wilf. Functional iteration and the Josephus problem. *Glasgow Math. J.*, 33(2):235–240, 1991.

[Pan83] J.-J. Pansiot. Hiérarchie et fermeture de certaines classes de tag-systèmes. *Acta Inform.*, 20(2):179–196, 1983.

[Pan86] J.-J. Pansiot. Decidability of periodicity for infinite words. *RAIRO Inform. Théor. Appl.*, 20(1):43–46, 1986.

[Par66] R. J. Parikh. On context-free languages. *J. Assoc. Comput. Mach.*, 13:570–581, 1966.

[PB97] F. Point and V. Bruyère. On the Cobham-Semenov theorem. *Theory Comput. Syst.*, 30(2):197–220, 1997.

[Pey87] J. Peyrière. Fréquence des motifs dans les suites doubles invariantes par une substitution. *Ann. Sci. Math. Québec*, 11(1):133–138, 1987.

[Pin93] R. G. E. Pinch. Recurrent sequences modulo prime powers. In *Cryptography and coding, III (Cirencester, 1991)*, volume 45 of *Inst. Math. Appl. Conf. Ser. New Ser.*, pages 297–310. Oxford Univ. Press, New York, 1993.

[PS95] Gh. Păun and A. Salomaa. Thin and slender languages. *Discrete Appl. Math.*, 61(3):257–270, 1995.

[Rau64] G. Rauzy. Relations de récurrence modulo m. In *Séminaire Delange-Pisot-Poitou, Théorie des nombres*, volume 5, 1963–1964.

[Rig00] M. Rigo. Generalization of automatic sequences for numeration systems on a regular language. *Theoret. Comput. Sci.*, 244(1-2):271–281, 2000.

[Rig01a] M. Rigo. *Abstract Numeration Systems on a Regular Language and Recognizability*. PhD thesis, Univ. Liège, 2001.

[Rig01b] M. Rigo. Numeration systems on a regular language: Arithmetic operations, recognizability and formal power series. *Theoret. Comput. Sci.*, 269(1-2):469–498, 2001.

[Rig02] M. Rigo. Construction of regular languages and recognizability of polynomials. *Discrete Math.*, 254(1-3):485–496, 2002.

[RM02] M. Rigo and A. Maes. More on generalized automatic sequences. *J. Autom., Lang. and Comb.*, 7(3):351–376, 2002.

[Rob66] D. W. Robinson. A note on linear recurrent sequences modulo m. *Amer. Math. Monthly*, 73:619–621, 1966.

[Rob00] A. M. Robert. *A course in p-adic analysis*, volume 198 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.

[RS59] M. O. Rabin and D. Scott. Finite automata and their decision problems. *IBM J. Res. Develop.*, 3:114–125, 1959.

[RW06] M. Rigo and L. Waxweiler. A note on syndeticity, recognizable sets and Cobham's theorem. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, (88):169–173, 2006.

[Sak03] J. Sakarovitch. *Élements de théorie des automates*. Vuibert, Paris, 2003.

[Sak06] J. Sakarovitch. Finite automata and number systems. In *International School and Conference on Combinatorics, Automata and Number Theory*. preprint of the Dep. of Math. of the University of Liège, 2006.

[Sal87a] O. Salon. Suites automatiques à multi-indices. In *Séminaire de théorie des nombres de Bordeaux*, volume 4, 1986-1987.

[Sal87b] O. Salon. Suites automatiques à multi-indices et algébricité. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(12):501–504, 1987.

[SC05] W. C. Shiu and C. Chu. Distribution of the Fibonacci numbers modulo 3^k. *Fibonacci Quart.*, 43(1):22–28, 2005.

[Sem77] A. L. Semenov. The Presburger nature of predicates that are regular in two number systems. *Siberian Math. J.*, 18(2):289–299, 1977.

[Sha68] A. P. Shah. Fibonacci sequence modulo m. *Fibonacci Quart.*, 6(2):139–141, 1968.

[Sha94] J. Shallit. Numeration systems, linear recurrences, and regular sets. *Inform. and Comput.*, 113(2):331–347, 1994.

[Ste03] R. Stephan. On a sequence related to the Josephus problem. *ArXiv e-prints*, May 2003. http://arxiv.org/abs/math/0305348v1.

[Sti94] J. Stillwell. *Elements of algebra*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1994. Geometry, numbers, equations.

[Sud06] T. A. Sudkamp. *Languages and Machines*. Addison Wesley, third edition, 2006.

[SYZS92] A. Szilard, S. Yu, K. Zhang, and J. Shallit. Characterizing regular languages with polynomial densities. In *Mathematical Foundations of Computer Science 1992 (Prague, 1992)*, volume 629 of *Lecture Notes in Comput. Sci.*, pages 494–503. Springer, Berlin, 1992.

[Tur74] M. R. Turner. Certain congruence properties (modulo 100) of Fibonacci numbers. *Fibonacci Quart.*, 12:87–91, 1974.

[Vil92a] R. Villemaire. Joining k- and l-recognizable sets of natural numbers. In *STACS 92 (Cachan, 1992)*, volume 577 of *Lecture Notes in Comput. Sci.*, pages 83–94. Springer, Berlin, 1992.

[Vil92b] R. Villemaire. The theory of $\langle \mathbf{N}, +, V_k, V_l \rangle$ is undecidable. *Theoret. Comput. Sci.*, 106(2):337–349, 1992.

[Vin78] A. Vince. The Fibonacci sequence modulo N. *Fibonacci Quart.*, 16(5):403–407, 1978.

[Vin81] A. Vince. Period of a linear recurrence. *Acta Arith.*, 39(4):303–311, 1981.

[Wad96] M. E. Waddill. Properties of a k-order linear recursive sequence modulo m. In *Applications of Fibonacci numbers, Vol. 6 (Pullman, WA, 1994)*, pages 505–519. Kluwer Acad. Publ., Dordrecht, 1996.

[Wal60] D. D. Wall. Fibonacci series modulo m. *Amer. Math. Monthly*, 67:525–532, 1960.

[War33] M. Ward. The arithmetical theory of linear recurring series. *Trans. Amer. Math. Soc.*, 35(3):600–628, 1933.

[Wol06] P. Wolper. *Introduction à la calculabilité*. Dunod, Paris, third edition, 2006.

[WY06] M. Wu and Y. M. Yang. On linear recurrent sequence modulo q. *Chinese Ann. Math. Ser. A*, 27(5):561–570, 2006.

[Yal73] C. C. Yalavigi. Fibonacci series modulo m. *Math. Education*, 7:A48–A54, 1973.

[Zab56] Ś. Zabek. Sur la périodicité modulo m des suites de nombres $\binom{n}{k}$. *Ann. Univ. Mariae Curie-Sklodowska. Sect. A*, 10:37–47, 1956.

[Zec72] E. Zeckendorf. Représentation des nombres naturels par une somme des nombres de Fibonacci ou de nombres de Lucas. *Bull. Soc. Roy. Sci. Liège*, 41:179–182, 1972.

Index