Rich ω-Words and Monadic Second-Order Arithmetic*

Ludwig Staiger**

Institut für Informatik
Martin-Luther-Universität Halle-Wittenberg
Kurt-Mothes-Straße 1
D-06120 Halle, Germany

* *This paper is to appear in the proceedings of the conference* **Computer Science Logic '97**, *Lecture Notes in Computer Science, Springer-Verlag, Berlin 1998*

** email: staiger@cantor.informatik.uni-halle.de

Table of Contents

Abstract. Rich ω-words are one-sided infinite strings which have every finite word as a subword (infix). Infix-regular ω-words are one-sided infinite strings for which the infix set of a suffix is a regular language. We show that for a regular ω-language F (a set of predicates definable in Büchi's restricted monadic second order arithmetic) the following conditions are equivalent:

1. F contains a rich ω-word.
2. F is of second Baire category in the Cantor space of ω-words.
3. F is a non-nullset for a class of measures (including the natural Lebesgue measure on Cantor space).
4. F has maximum Hausdorff dimension.

This shows that, although we cannot fully translate Compton's result (Theorem 1 below) on rich $\mathbb{Z}$-words (in the MSO theory of the integers) to MSO arithmetic on naturals, a set definable in MSO arithmetic and containing a rich ω-word is large in several respects simultaneously.

Moreover, we show under the assumption of an exchanging property for 'distinguishing' prefixes that two regular ω-words not necessarily being rich but having the same sets of infixes occurring infinitely often are indistinguishable by MSO formulas or, equivalently, by finite automata.

Restricted monadic second-order (MSO) arithmetic was shown to be decidable by J.R. Büchi [Bü60]. This famous result lead to a series of investigations of sets of predicates definable in MSO arithmetic. In his proof Büchi showed a strong correspondence between monadic second-order logic and the theory of finite automata. Put into the context of formal language theory it resulted in the theory of regular ω-languages, that is sets of one-sided infinite strings accepted by finite automata (see [Th90, 97], [St97]).

Along with MSO arithmetic the MSO theory of the integers $\mathbb{Z}$ has been considered. Here K. Compton proved the following theorem on predicates or so-called $\mathbb{Z}$-words (biinfinite strings over a finite alphabet X) of a certain type.

Theorem 1 [Co83, Corollary 4.5]. *Let Φ be a formula in the restricted MSO logic over the integers (without constant 0), and let E_Φ be the set of $\mathbb{Z}$-words satisfying Φ. Then the following are equivalent.*

1. *E_Φ is comeager.*
2. *E_Φ is of measure 1.*
3. *There is a rich $\mathbb{Z}$-word satisfying Φ.*
4. *All rich $\mathbb{Z}$-words satisfy Φ.*

As topology and measure Compton considered the (natural) product space and product measure of the discrete probability space X, and he called a $\mathbb{Z}$-word *rich* if it contains all finite words as subwords (infixes) infinitely often to the right and to the left. Compton's theorem exhibits a strong correspondence between category and measure for MSO-definable sets in the space of $\mathbb{Z}$-words. Subsequently a part of this theorem has been generalized to $\mathbb{Z}$-words not containing all finite words as subwords.

Theorem 2 [Se84, PS86]. *Let Φ be a formula in the restricted MSO logic over the integers (without constant 0) such that there is a recurrent $\mathbb{Z}$-word*[1] *satisfying Φ. Then every recurrent $\mathbb{Z}$-word having the same set of infixes satisfies Φ.*

The purpose of our paper is to derive analogous statements for MSO logic over the naturals. In contrast to $\mathbb{Z}$-words, however, predicates over the naturals $\mathbb{N}$ (ω-words) have a fixed scale. So it is easy to construct for each pair of ω-words ξ, η ($\xi \neq \eta$) an MSO-formula $\Phi_{\xi,\eta}$ which separates ξ and η, that is, $\Phi_{\xi,\eta}(\xi) \wedge \neg\Phi_{\xi,\eta}(\eta)$ holds true, because there is a distinguishing prefix w of ξ which is not a prefix of η. This shows that a direct translation of the above theorems to MSO arithmetic is not possible.

Thus we substitute the topological and measure-theoretical largeness conditions in Theorem 1 by slightly weaker conditions (cf. [Ox71]), and add a third one, that is, we consider the following three types of largeness:

- in the sense of measure: Large sets are sets of non-null measure.

[1] A $\mathbb{Z}$-word is called *recurrent* if it has every of its infixes infinitely often to the left and to the right.

– in the sense of topological density: Large sets are sets of second Baire category.
– in the sense of dimension: Large sets are sets of maximal Hausdorff dimension.

It is known that in general large sets with respect to measure (non-nullsets) are not necessarily large with respect to category and vice versa (cf. [Ox71]). A similar situation holds true in the relations between Hausdorff dimension and category or measure (cf. [Fa90]).

In the first part we prove that an MSO-definable predicate Ψ over the naturals which is satisfied by an ω-word which has all finite words as infixes (henceforth called a *rich* ω-word[2]) is satisfied by a 'large' (in each one of the above respects) set of ω-words. This proves also a similar strong coincidence of category, measure and dimension in the case of MSO-definable sets of ω-words (so called ω-languages). It should be noted that, in contrast to Theorem 1, we do not confine our considerations to the ordinary product measure, but to a larger class of measures which may also vanish on nonempty open sets. This requires also relativizing of topological density, but enables us to take into consideration ω-languages having arbitrary (non-maximal) Hausdorff dimension.

In the second part we turn to the situation (corresponding to Theorem 2) where the ω-words under consideration do not necessarily have all finite words as infixes. Here we show that, if we allow for an exchange of the 'distinguishing' prefixes, MSO arithmetic cannot 'distinguish' between two recurrent ω-words provided both words satisfy a certain regularity condition.

Similar to the preceding papers [Co83] and [PS86], which use Ehrenfeucht-Fraïsse games or semigroup theory to prove their results, the proofs of our results are also not based directly on monadic second order logic, but make use of the above mentioned strong correspondence between MSO arithmetic and the theory of finite automata on (infinite) ω-words (the theory of regular ω-languages) (cf. [Ei74], [Th90, 97], [St97]). In the latter theory several results linking metric (topological), measure theoretical and dimension theoretical properties of regular ω-languages are already at hand (see e.g. [St80, 93] and [MS94]) and need not be developed. Therefore, we introduce also the notation used there.

[2] We follow here Compton's terminology, in [JS83] such ω-words were called *disjunctive*. In the case of ω-words it is apparent that if all words appear at least once as infix then they appear also infinitely often as infix.

1 Notation and Definitions

Before proceeding to the formulation of our results we introduce some necessary notation, and we give a proper specification of the topology, measures and dimension investigated throughout the paper. For more detailed considerations on measure, topology and Hausdorff dimension the interested reader is referred to standard textbooks as e.g. [Ku66], [Ox71], [Fa90], and [Ed90].

We consider, for a finite alphabet X of cardinality $r := \mathrm{card}\, X \geq 2$, the space X^ω of all ω-words over X, and as usual we denote by X^* the set of finite words over X. For $w \in X^*$ and $b \in X^* \cup X^\omega$ let $w \cdot b$ be their concatenation. If $b = w \cdot b'$ the word w is called a *prefix* of b (short: $w \sqsubseteq b$), and in this case $b' = b/w$ is the *left derivative* of b by the word w.
These notations generalize in an obvious way to $w \cdot B$, $W \cdot B$ and B/w for subsets $W \subseteq X^*$ and $B \subseteq X^* \cup X^\omega$. For a finite word $w \in X^*$ we denote by $|w|$ its length. A *prefix code* is a set $W \subseteq X^*$ for which no word $w \in W$ is a proper prefix ($w \sqsubset w'$) of another word $w' \in W$.

The set X^ω is usually considered as a metric space (Cantor space) with metric ϱ defined by

$$\varrho(\eta, \xi) = \inf \{r^{-|w|} : \ w \sqsubseteq \eta \text{ and } w \sqsubseteq \xi\} \ .$$

The family of open balls is given by $\{w \cdot X^\omega : w \in X^*\}$ where $\mathrm{diam}\, w \cdot X^\omega = r^{-|w|}$ is the *diameter* of the ball $w \cdot X^\omega$. Thus a subset $E \subseteq X^\omega$ is *open* iff, for some $W \subseteq X^*$, it has the form $E = W \cdot X^\omega$.

A set $F \subseteq X^\omega$ is called *nowhere dense* provided for every ball $w \cdot X^\omega$ there is a nonempty subball $wv \cdot X^\omega$ such that $F \cap wv \cdot X^\omega = \emptyset$. A countable union $\bigcup_{i \in \mathbb{N}} F_i$ of nowhere dense sets F_i is called *meager* or of *first Baire category*. A set which is not of first Baire category is referred to as of *second Baire category*.[3]

In order to define Hausdorff dimension in the metric space (X^ω, ϱ) we introduce

$$\mathbf{L}_\alpha(F) := \lim_{\delta \to 0} \Big(\inf \Big\{ \sum_{v \in V} (\mathrm{diam}\, v \cdot X^\omega)^\alpha : V \cdot X^\omega \supseteq F \wedge \sup_{v \in V} \mathrm{diam}\, v \cdot X^\omega < \delta \Big\} \Big)$$

as the *α-dimensional outer measure* of $F \subseteq X^\omega$. Then as usual (cf. [Ed90], [Fa90]) the *Hausdorff dimension* $\dim F$ is given by

$$\dim F := \sup \{\alpha : \mathbf{L}_\alpha(F) = \infty\} = \inf\{\alpha : \mathbf{L}_\alpha(F) = 0\} \ .$$

Thus the Hausdorff dimension assigns to each subset $F \subseteq X^\omega$ a real number $\alpha = \dim F$ between 0 and 1 which defines in some sense a size of F (cf. [Fa90], [St93], [MS94]).

For every $w \in X^*$ the ball $w \cdot X^\omega$ is a disjoint union of the balls $wx \cdot X^\omega$ ($x \in X$). Thus $\mu(w \cdot X^\omega) = \sum_{x \in X} \mu(wx \cdot X^\omega)$ for every measure μ on X^ω. As

[3] In X^ω and $X^{\mathbb{Z}}$ every complement of a meager set (a comeager set) is of second Baire category, but not vice versa.

measures μ on X^ω we consider finite measures ($\mu(X^\omega) < \infty$) having the following property that the mass of a subball $wx \cdot X^\omega$ with non-null mass does not deviate too much from $\mu(w \cdot X^\omega)$:

Balance condition There is a constant $c_\mu > 0$ depending only on μ such that for all words $w \in X^*$ and every $x \in X$ we have $\mu(wx \cdot X^\omega) = 0$ or $c_\mu \cdot \mu(w \cdot X^\omega) \leq \mu(wx \cdot X^\omega)$.

The *support* of a measure μ on X^ω, $\mathbf{supp}(\mu)$, is the smallest closed subset of X^ω such that $\mu(\mathbf{supp}(\mu)) = \mu(X^\omega)$.

2 Main results

The following characterization, which shows that rich ω-words are contained only in the largest (with respect to all three sizes) MSO-definable sets, is the first main result of our paper.

Theorem 3. *Let Φ be some MSO formula, and let $F_\Phi := \{\xi : \xi \in X^\omega \wedge \xi \text{ satisfies } \Phi\}$. Then the following conditions are equivalent:*

1. *Some ω-word $\zeta \in F_\Phi$ is a rich ω-word.*
2. *F_Φ is of second Baire category.*
3. *For all measures μ with $\mathbf{supp}(\mu) = X^\omega$ satisfying the balance condition it holds $\mu(F_\Phi) > 0$.*
4. *There is a measure μ with $\mathbf{supp}(\mu) = X^\omega$ satisfying the balance condition such that $\mu(F_\Phi) > 0$.*
5. $\dim F_\Phi = \dim X^\omega$.

It should be noted that already for a slightly larger class of ω-languages our theorem will not be true. This is explained best using one-counter-automata:

Example 1. Let the language $V_3 \subseteq \{a, b\}^*$ be given by the equation

$$V_3 = \{a\} \cup \{b\} \cdot V_3^3 \ .$$

This language is a prefix code and can be defined by a deterministic one-counter-automaton. Accordingly, the ω-language $E := \bigcap_{n=0}^{\infty} V_3^n \cdot \{a, b\}^\omega = \{\xi : \exists (w_i)_{i=1}^{\infty} (w_i \in V_3 \wedge \xi = w_1 \cdots w_i \cdots)\}$ can be defined also by a deterministic one-counter-automaton. The ω-language E is a $\mathbf{G}_\delta$-set dense in the space $\{a, b\}^\omega$, hence of second Baire category (cf. [Ku66]), but it has $\mu(E)_= = 0$, for the (equidistribution) measure defined by $\mu_=(w \cdot \{a, b\}^\omega) := 2^{-|w|}$ (cf. [St80]) and $\dim E < \dim\{a, b\}^\omega$ (cf. [St93, Example 6.3]). □

As it was announced in the introduction we are going to prove our Theorem 3 as a special case of its relativized version presented in Theorem 4 below. This gives further evidence that the above mentioned notions density, measure or dimension are in some sense strongly equivalent for MSO-definable sets.

To this end, we introduce a concept of complexity of infinite sequences ξ which is intimately related to rich ω-words. This concept is based solely on the sets of subwords (infixes) of an ω-word $\xi \in X^\omega$, $\mathbf{T}(\xi) := \{w : w \in X^* \wedge \exists u(u \cdot w \sqsubset \xi)\}$.

For a language $W \subseteq X^*$ let $H_W := \limsup_{n\to\infty} \frac{\log_r \operatorname{card} W \cap X^n}{n}$ be its *entropy* (e.g. [Ku70], [St93]).

We will refer to $\tau(\xi) := H_{\mathbf{T}(\xi)}$ as the *subword complexity* of the word $\xi \in X^\omega$. Recall that $r = \operatorname{card} X$. Thus $0 \le \tau(\xi) \le 1$. Moreover, ξ is rich iff $\tau(\xi) = 1$.

We recall that a set $F \subseteq X^\omega$ has *finite α-measure* iff $\mathbf{L}_\alpha(F) < \infty$, and F has *locally positive α-measure* iff $0 < \mathbf{L}_\alpha(F \cap w \cdot X^\omega)$ whenever $F \cap w \cdot X^\omega \neq \emptyset$. Clearly, a set F having finite and locally positive α-measure has Hausdorff dimension $\dim F = \alpha$.

For $\alpha = 0$ a set of finite α-measure is itself finite; this simple case will not be considered in the following theorem.

Theorem 4. *Let μ be a non-null measure on X^ω satisfying the balance condition, and let $\mathbf{supp}(\mu)$ be MSO-definable. Moreover, let for $\alpha := \dim \mathbf{supp}(\mu) > 0$ the support $\mathbf{supp}(\mu)$ have finite and locally positive α-measure. Then for every MSO-definable subset $F \subseteq \mathbf{supp}(\mu)$ the following conditions are equivalent:*

1. *For all $\xi \in F$, $\tau(\xi) < \alpha$.*
2. *F is of first Baire category in $\mathbf{supp}(\mu)$.*
3. *$\mu(F) = 0$*
4. *$\mathbf{L}_\alpha(F) = 0$*
5. *$\dim F < \alpha$*

Crucial steps in the proof of Theorem 3 and 4 will be the results of the subsequent section relating measure, density and dimension for regular ω-languages.

Before proceeding to this section we present our second main theorem. This needs some preparatory considerations. Let $\mathbf{T}_\infty(\xi) := \{w : w \in X^* \wedge \forall n \exists u(|u| \ge n \wedge u \cdot w \sqsubset \xi)\}$ be the set of infixes occurring infinitely often in ξ.

We call an ω-word $\xi \in X^\omega$ as *recurrent* if there is a prefix $w \sqsubset \xi$ such that $\mathbf{T}(\xi/w) = \mathbf{T}_\infty(\xi)$[4], and we call an ω-word $\xi \in X^\omega$ *infix-regular* provided there is a prefix $w \sqsubset \xi$ such that $\mathbf{T}(\xi/w)$ is a regular language.

The following lemma yields a connection between infix-regular ω-words and recurrent ω-words.

Lemma 5. *If $\xi \in X^\omega$ is a infix-regular ω-word then ξ is also recurrent.*

[4] In view of $\mathbf{T}_\infty(\xi) = \mathbf{T}_\infty(\xi/w)$ an ω-word ξ is recurrent iff for some of its tails ξ/w every infix occurs infinitely often as an infix, this in some sense resembles the recurrence of $\mathbb{Z}$-words.

Proof. As $\mathbf{T}(\xi/v) \supseteq \mathbf{T}(\xi/v \cdot u)$ and $\mathbf{T}(\xi)/v \subseteq \mathbf{T}(\xi/v) = \mathbf{T}(\mathbf{T}(\xi)/v)$ the family $(\mathbf{T}(\xi/v))_{v \sqsubset \xi}$ is a decreasing family of subsets of X^*. Since $\mathbf{T}(\xi/w)$ is a regular language for some prefix $w \sqsubset \xi$, this family consists of only finitely many distinct members. The assertion now follows from $\mathbf{T}_\infty(\xi) = \bigcap_{v \sqsubset \xi} \mathbf{T}(\xi/v)$. □

It should be noted that not every ω-word ξ for which $\mathbf{T}_\infty(\xi)$ is a regular language is recurrent.

Example 2. Consider $\xi_0 := \prod_{i=1}^{\infty} a^i \cdot b$. Then $\mathbf{T}_\infty(\xi_0) = a^* \cdot b \cdot a^*$, but $\mathbf{T}(\xi_0/w) \not\subseteq a^* \cdot b \cdot a^*$ for arbitrary $w \sqsubset \xi_0$. □

Now we can formulate the second main result of this paper.

Theorem 6. *Let $\xi \in X^\omega$ be an infix-regular ω-word and let $F \subseteq X^\omega$ be a regular ω-language containing ξ. Then there is a prefix $w_F \sqsubset \xi$ such that for every infix-regular ω-word η with $\mathbf{T}_\infty(\xi) = \mathbf{T}_\infty(\eta)$ there is a prefix $v_F \sqsubset \eta$ such that $w_F \cdot (\eta/v_F) \in F$.*

Cast in the language of automata Theorem 6 asserts that a finite automaton $\mathcal{A}$ cannot distinguish the ω-words ξ and η in the following sense: either $\mathcal{A}$ does not accept both of them or if $\mathcal{A}$ accepts ξ (say) then there are prefixes $w_\mathcal{A} \sqsubset \xi$ and $v_\mathcal{A} \sqsubset \eta$ such that $\mathcal{A}$ after reading $w_\mathcal{A}$ cannot distinguish the tails $\xi/w_\mathcal{A}$ and $\eta/v_\mathcal{A}$.

The proof of Theorem 6 relies on relationships between sub-word complexity of ω-words and the α-dimensional measure of ω-languages containing them, and is presented in Section 4.

3 Measure, category and dimension in X^ω

In this section we derive a proof of the equivalences $(2) \leftrightarrow (3) \leftrightarrow (4) \leftrightarrow (5)$ in Theorem 3 and 4. The proof of the equivalences $(1) \leftrightarrow (5)$ will be postponed to Section 4 where we deal in more detail with the subword complexity of ω-words.

The announced equivalences can be easily derived from the subsequent Theorems 7 and 8. Our first theorem concerns measure and category where, in contrast to the case of arbitrary sets which is thoroughly treated in Oxtoby's monograph [Ox71], regular ω-languages behave quite 'regular', that is, for a large class of measures, the nullsets are exactly the sets of first Baire category.

Theorem 7. *Let S be a nonempty closed regular subset of the space (X^ω, ϱ), and let μ be a measure on X^ω satisfying $0 < \mu(S \cap w \cdot X^\omega) < \infty$ whenever $S \cap w \cdot X^\omega \neq \emptyset$. If there is a constant $c > 0$ such that for all words $w \in X^*$ and all $x \in X$ the balance condition*

$$\mu(wx \cdot X^\omega \cap S) = 0 \text{ or } c \cdot \mu(w \cdot X^\omega \cap S) \leq \mu(wx \cdot X^\omega \cap S)$$

holds true then a regular subset $E \subseteq S$ is of first Baire category in S if and only if $\mu(E) = 0$.

The second case where regular subsets of X^ω behave quite 'regular' is connected with the Hausdorff dimension of subsets of X^ω. Here it is seen that the Hausdorff dimension is indeed a measure of the relative (topological) density of regular ω-languages.

Theorem 8. *Let $S \subseteq X^\omega$ be nonempty and regular, $0 < \alpha = \dim S$, $0 < \mathbf{L}_\alpha(S) < \infty$ and $0 < \mathbf{L}_\alpha(S \cap w \cdot X^\omega)$ whenever $S \cap w \cdot X^\omega \neq \emptyset$. Then for every regular $E \subseteq S$ the following conditions are equivalent:*

1. *E is of first category in S,*
2. *$\mathbf{L}_\alpha(E) = 0$, and*
3. $\dim E < \dim S$.

Before we proceed to the proofs we explain when a subset $F \subseteq X^\omega$ is of first and second Baire category in another subset $S \subseteq X^\omega$. To this end let $\mathcal{C}(F)$ denote the smallest closed subset of X^ω containing $F \subseteq X^\omega$, and let $\mathbf{A}(B) := \{w : w \in X^* \wedge \exists b(b \in B \wedge w \sqsubseteq b)\}$ be the set of all finite prefixes of strings in $B \subseteq X^* \cup X^\omega$.

As usual we call a set F *nowhere dense in* S provided $\mathcal{C}(S \setminus \mathcal{C}(F)) = \mathcal{C}(S)$, that is, if $\mathcal{C}(F)$ does not contain a nonempty subset of the form $S \cap w \cdot X^\omega$. Then as above, a subset F is referred to as of *first Baire category in* S if F is a countable union of sets nowhere dense in S, and F is of second Baire category in S if it is not of first Baire category in E. As usual, a countable union of closed sets is called an $\mathbf{F}_\sigma$-set and a countable intersection of open sets is called a $\mathbf{G}_\delta$-set.

The proof of Theorems 7 proceeds as follows.
It is well-known (cf. [St97], [Th90]) that every regular ω-language is a countable union of regular $\mathbf{G}_\delta$-sets (cf. Fact 3 below). Thus we may prove our result inductively in three steps:

First we prove Theorem 7 for closed regular ω-languages. Utilizing Facts 4 and 5, which refer to some general topological properties of $\mathbf{G}_\delta$-sets, we show the result for regular $\mathbf{F}_\sigma$-sets and to regular $\mathbf{G}_\delta$-sets. Finally, the structural results of Facts 2 and 3 allow us to extend the result to all regular ω-languages.

Then Theorem 8 is derived from Theorem 7.

We start with the case of (topologically) closed regular ω-languages. First we mention that the closure $\mathcal{C}(E)$ of a regular ω-language $E \subseteq X^\omega$ is again regular, and that the class of regular ω-languages is closed under Boolean operations (cf. [SW74], [Th90], [Wa79]).

Utilizing the identity $F \cap w \cdot X^\omega = w \cdot (F/w)$ and the equivalence $F \cap w \cdot X^\omega \neq \emptyset$ iff $w \in \mathbf{A}(F)$, we may reformulate the condition that F is nowhere dense in S in terms of their languages of finite prefixes.

Proposition 9. *A subset F of X^ω is nowhere dense in $S \subseteq X^\omega$ iff for every $w \in \mathbf{A}(S)$ there is a $v \in \mathbf{A}(S/w)$ such that $v \notin \mathbf{A}(F/w)$.* □

Then for regular sets S and F we get a condition bounding the length of the word v.

Theorem 10. *Let S, F be regular subsets of X^ω and let F be nowhere dense in S. Then for every $k \in \mathbb{N}$ with $k \geq \mathrm{card}\{S/w : w \in X^*\} \cdot \mathrm{card}\{F/w : w \in X^*\}$ it holds the following:*

$$\forall w (w \in \mathbf{A}(S) \to \exists v (v \in X^* \wedge |v| < k \wedge w \cdot v \in \mathbf{A}(S) \setminus \mathbf{A}(F))) \ .$$

The proof is an immediate consequence of Proposition 9 and the fact, well-known from automata theory, that for every nonempty language $W \subseteq X^*$ accepted by an automaton with at most k states there is a word $v \in W$ such that $|v| < k$. Take here $W := \mathbf{A}(S/w) \setminus \mathbf{A}(F/w)$. □

For measures μ on X^ω the assumption $0 < \mu(S \cap w \cdot X^\omega)$ whenever $S \cap w \cdot X^\omega \neq \emptyset$ in Theorem 7 is equivalent to $S \subseteq \mathbf{supp}(\mu)$. Since we are not interested in the behaviour of the measure μ on $X^\omega \setminus S$, in the sequel we assume $\mu(X^\omega \setminus S) = 0$, that is, $S = \mathbf{supp}(\mu)$.

A first obvious fact relating topological and measure theoretical behaviour is the following one:

Fact 1 *If $F \subseteq \mathbf{supp}(\mu)$ is closed and $\mu(F) = 0$ then F is nowhere dense in $\mathbf{supp}(\mu)$.*

For regular sets we obtain the following converse.

Lemma 11. *Let $S = \mathbf{supp}(\mu)$ be a nonempty, closed and regular subset of X^ω, and let $\mu(X^\omega) < \infty$. If μ satisfies the balance condition then every regular set F which is nowhere dense in S has $\mu(F) = 0$.*

Proof. If $F = \emptyset$ the assertion is true. Let F be nonempty, regular and nowhere dense in $S = \mathbf{supp}(\mu)$. Then according to Theorem 10 there is a $k \in \mathbb{N}$, such that for all $w \in \mathbf{A}(S)$ there is a $u \in \mathbf{A}(S/w)$ with $|u| \leq k$ and $w \cdot u \notin \mathbf{A}(F)$.

Set $\mathbf{m}(F) := \sup\{\mu(v \cdot (F/v))/\mu(v \cdot X^\omega) : v \in \mathbf{A}(F)\}$ and choose $w \in \mathbf{A}(F)$ such that $\mu(w \cdot (F/w))/\mu(w \cdot X^\omega) \geq \mathbf{m}(F) \cdot (1 - \delta)$ where $c^k > \delta > 0$.

Since $w \cdot (F/w) = F \cap w \cdot X^\omega$, we have

$$(1 - \delta) \cdot \mathbf{m}(F) \leq \frac{\mu(w \cdot (F/w))}{\mu(w \cdot X^\omega)} = \sum_{|v|=k} \frac{\mu(w \cdot v \cdot X^\omega)}{\mu(w \cdot X^\omega)} \cdot \frac{\mu(w \cdot v \cdot (F/w \cdot v))}{\mu(w \cdot v \cdot X^\omega)} \ ,$$

and $1 = \sum_{|v|=k} \mu(w \cdot v \cdot X^\omega)/\mu(w \cdot X^\omega)$, because of $S = \mathbf{supp}(\mu)$ and $w \in \mathbf{A}(S)$.

By the hypothesis, $\mu(w \cdot v \cdot X^\omega)/\mu(w \cdot X^\omega) \geq c^k$ whenever $v \in \mathbf{A}(S/w)$ and $|v| = k$. Since there is at least one word $u \in \mathbf{A}(S)/w \setminus \mathbf{A}(F)/w$ with $|u| = k$, we get the inequality $(1-\delta) \cdot \mathbf{m}(F) \leq \mu(w \cdot (F/w))/\mu(w \cdot X^\omega) \leq (1 - c^k) \cdot \mathbf{m}(F)$. Thus, $\delta < c^k$ implies $\mathbf{m}(F) = 0$, hence, $\mu(F) = 0$. □

From Fact 1 and Lemma 11 the following is immediate.

Corollary 12. *Let $S = \mathbf{supp}(\mu)$ be a nonempty, closed and regular subset of X^ω, and let $\mu(X^\omega) < \infty$. If μ satisfies the balance condition then a closed regular ω-language F is nowhere dense in S iff $\mu(F) = 0$.*

Next, we recall some facts about the topological nature of regular subsets of X^ω (cf. [SW74], [Th90], [Wa79]):

Fact 2 *Every regular $\mathbf{F}_\sigma$-subset $F \subseteq X^\omega$ is a countable union of regular and closed subsets of X^ω.*

Fact 3 *Every regular subset $E \subseteq X^\omega$ is a countable union of regular $\mathbf{G}_\delta$-subsets.*

It should be noted that the converse of Fact 3 is not true.

We proceed with a simple property linking sets of first Baire category in $S = \mathbf{supp}(\mu)$ and μ-nullsets.

Proposition 13. *Let $\{F_i : i \in \mathbb{N}\}$ be a family of subsets of X^ω such that for every $i \in \mathbb{N}$ the set F_i is of first Baire category in S if and only if $\mu(F_i) = 0$. Then $\bigcup_{i\in\mathbb{N}} F_i$ is of first category in S iff $\mu(\bigcup_{i\in\mathbb{N}} F_i) = 0$.*

And we obtain our assertion for the case of $\mathbf{F}_\sigma$-sets.

Corollary 14. *Let F be a countable union of regular and closed subsets of X^ω. Then F is of first category in S if and only if $\mu(F) = 0$.*

The following facts about $\mathbf{G}_\delta$-sets are known from topology (cf. [Ku66]):

Fact 4 *A $\mathbf{G}_\delta$-set which is of first category in S is nowhere dense in S.*

Fact 5 *If $E \subseteq S$ is a $\mathbf{G}_\delta$-set then the $\mathbf{F}_\sigma$-set $\mathcal{C}(E) \setminus E$ is of first category in S.*

Utilizing these facts we derive our assertion for regular $\mathbf{G}_\delta$-subsets of X^ω.

Lemma 15. *Let $S = \mathbf{supp}(\mu)$ be a nonempty, closed and regular subset of X^ω, and let $\mu(X^\omega) < \infty$. If μ satisfies the balance condition then a regular $\mathbf{G}_\delta$-set E is nowhere dense in S iff $\mu(E) = 0$.*

Proof. Lemma 11 shows that $E \subseteq S$ is a nullset if E is nowhere dense in S.

Conversely, let $\mu(E) = 0$ and consider the ω-language $F := (\mathcal{C}(E) \cap S) \setminus E = (\mathcal{C}(E) \setminus E) \cap S$. According to the above mentioned properties of regular ω-languages, F is a regular $\mathbf{F}_\sigma$-subset of X^ω. By Fact 5 the set F, and hence also $\mathcal{C}(E) \setminus E$ is of first category in S. Thus, $\mu(\mathcal{C}(E)) = 0$, and Lemma 11 shows that $\mathcal{C}(E)$ and also E are nowhere dense in S. □

Now, utilizing Lemma 15, Fact 3 and Proposition 13 stated above, we derive a slight generalization of Theorem 7.

Theorem 16. *Let $S = \mathbf{supp}(\mu)$ be a nonempty, closed and regular subset of X^ω, $\mu(X^\omega) < \infty$ and let $E = \bigcup_{i\in\mathbb{N}} E_i$ be a countable union of regular $\mathbf{G}_\delta$-sets. Then E is of first category in S iff $\mu(E) = 0$.*

In particular, Theorem 16 holds for all open (not necessarily regular) subsets of X^ω, because they are of the form $\bigcup_{w\in W} w \cdot X^\omega$ $(W \subseteq X^*)$. This fact may be used to explain that we cannot transfer the assertion of Theorem 16 to complements, that is, to the class of countable intersections of regular $\mathbf{F}_\sigma$-subsets of X^ω. We give a particular simple example.

Example 3 (continuation of Example 1). Consider the same language $V_3 \subseteq \{a,b\}^*$ as in Example 1 and the same measure $\mu_=$. Since $E := \bigcap_{n=0}^{\infty} V_3^n \cdot \{a,b\}^\omega$ is a nullset, its complement $\{a,b\}^\omega \setminus E = V_3^* \cdot (\{a,b\}^\omega \setminus V_3 \cdot \{a,b\}^\omega)$ has measure 1, and one obtains that the closed and nowhere dense (nonregular) ω-language $\{a,b\}^\omega \setminus V_3 \cdot \{a,b\}^\omega$ has measure $\mu_=(\{a,b\}^\omega \setminus V_3 \cdot \{a,b\}^\omega) = \frac{3-\sqrt{5}}{2} > 0$.

According to the preceding discussion, as a closed subset of $\{a,b\}^\omega$ the set $\{a,b\}^\omega \setminus V_3 \cdot \{a,b\}^\omega$ is a countable intersection of regular $\mathbf{F}_\sigma$-subsets. □

Finally, we derive Theorem 8 from Theorem 7.

Proof of Theorem 8. We define the measure μ via $\mu(E) := \mathbf{L}_\alpha(E \cap S)$. Then $\mathbf{supp}(\mu) = \mathcal{C}(S)$ is a regular and closed subset of X^ω.

It suffices to verify that for every $w \cdot x \in \mathbf{A}(S)$ where $x \in X$ there is a constant $c > 0$ such that $c \cdot \mu(w \cdot X^\omega) \leq \mu(w \cdot x \cdot X^\omega)$.

To this end we observe that the following relations hold true: $\mu(w \cdot X^\omega) = \mathbf{L}_\alpha(w \cdot X^\omega \cap S) = \mathbf{L}_\alpha(w \cdot (S/w)) = r^{-\alpha\cdot|w|} \cdot \mathbf{L}_\alpha(S/w)$ for all $w \in X^*$. Moreover, $q_S := \inf\{\mathbf{L}_\alpha(S/w)/\mathbf{L}_\alpha(S/v) : w, v \in \mathbf{A}(S)\} > 0$, because the set $\{S/w : w \in \mathbf{A}(S)\}$ is finite and $0 < \mathbf{L}_\alpha(S/w) < \infty$ for all $w \in \mathbf{A}(S)$.

Thus, $\mu(w \cdot x \cdot X^\omega) = r^{-\alpha|w\cdot x|} \cdot \mathbf{L}_\alpha(S/(w \cdot x)) \geq (r^{-\alpha} \cdot q_S) \cdot r^{-\alpha\cdot|w|} \cdot \mathbf{L}_\alpha(S/w) = (r^{-\alpha} \cdot q_S) \cdot \mu(w \cdot X^\omega)$, and μ satisfies the hypothesis of Theorem 7.

Consequently, a regular subset $E \subseteq S$ is of first category in $\mathcal{C}(S) = \mathbf{supp}(\mu)$ iff $\mu(E) = \mathbf{L}_\alpha(E) = 0$.

Since $E \subseteq S$, E is of first category in $\mathcal{C}(S)$ iff E is of first category in S, and the equivalence of the first two conditions is proved. The remaining equivalence follows from the fact (cf. [St93, Theorem 4.7]) that for nonempty regular ω-languages it holds $\dim E \leq \alpha$ iff $\mathbf{L}_\alpha(E) > 0$. □

4 Subword complexity and α-dimensional measure

In this section we derive the results which prove the equivalences (1) $\leftrightarrow$ (5) in Theorems 3 and 4 and Theorem 6.

The results of Section 5 in [St93] show that the subword complexity[5] $\tau(\xi)$ of an ω-word $\xi \in X^\omega$ is closely related to the Hausdorff dimension of the regular ω-languages containing ξ. We summarize them in the following proposition.

[5] It holds also $\tau(\xi) = H_{\mathbf{T}_\infty(\xi)}$.

Proposition 17. *Let F be a regular ω-language. Then $\tau(\xi) \leq \dim F$ for every $\xi \in F$. More precisely,*

$$\tau(\xi) = \inf\{\dim E : \xi \in E \wedge E \subseteq X^\omega \wedge E \text{ is regular}\} \ .$$

Moreover, for every nonempty regular ω-language $\emptyset \neq F \subseteq X^\omega$ there is an ω-word $\xi \in F$ such that $\tau(\xi) = \dim F$.

Now, the equivalences $(1) \leftrightarrow (5)$ are apparent.

In order to prove Theorem 6 we need some further properties of the α-dimensional measure $\mathbf{L}_\alpha$ which can be found e.g. in [MS94, Theorem 10, Corollary 13, and Theorem 16] or [St93, Theorem 4.7].

We call a set $B \subseteq X^* \cup X^\omega$ *strongly connected* if for every $w \in \mathbf{A}(B)$ there is a $v \in X^*$ such that $B/(w \cdot v) = B$.

Proposition 18. *Let $\emptyset \neq F \subseteq X^\omega$ be a regular ω-language and let $\alpha := \dim F$. Then the following holds:*

1. $\mathbf{L}_\alpha(F) > 0$,
2. $\exists w(\mathbf{L}_\alpha(F/w) \geq 1)$, *and*
3. *if F is closed and strongly connected then $\mathbf{L}_\alpha(F/w) \leq 1$ for all $w \in X^*$.*

A second result relating subword complexity to the α-dimensional measure is the following one.

Lemma 19. *Let $F \subseteq X^\omega$ be regular and $\xi \in F$. If $\alpha = \tau(\xi) = \dim F$ then for every $n \in \mathbb{N}$ there is a prefix $w_1 \sqsubset \xi$ such that $|w_1| \geq n$ and $\mathbf{L}_\alpha(F/w_1) \geq 1$.*

Proof. From the theory of regular ω-languages we know that $\xi \in w \cdot V^\omega \subseteq F$ for some $w \in X^*$ and a regular prefix code $V \subseteq X^*$ (cf. [Ei74]). Without loss of generality we may assume that $\mathbf{T}(\xi/w) = \mathbf{T}_\infty(\xi)$. According to Proposition 17 it follows $\dim w \cdot V^\omega = \dim V^\omega = \alpha$.

Next we show $V^* \subseteq \mathbf{T}_\infty(\xi)$. Assume that there are an $m > 0$ and a $v_0 \in V^m$ with $v_0 \notin \mathbf{T}_\infty(\xi)$. Then $\xi \in w \cdot (V^m \setminus \{v_0\})^\omega$. For the regular prefix code V^m one easily calculates $H_{V^*} = H_{(V^m)^*} > H_{(V^m \setminus \{v_0\})^*}$ (cf. [Ei74], [Ku70]). Now, from the identity $\dim W^\omega = H_{W^*}$ [St93, Eq. (6.2)] we obtain $\tau(\xi) = \alpha > \dim(V^m \setminus \{v_0\})^\omega$ which contradicts Proposition 17.

Thus for every $u \in \mathbf{A}(V^*)$ there is a $v \in \mathbf{A}(V^*)$ such that $w \cdot v \sqsubset \xi$ and $V^\omega/u = V^\omega/v$. Now Proposition 18 shows $\mathbf{L}_\alpha(V^\omega/u) \geq 1$ for some $u \in \mathbf{A}(V^*)$ and $\alpha := \tau(\xi)$. Hence $w_1 := w \cdot v$ fulfills $\mathbf{L}_\alpha(F/w_1) \geq 1$. □

After having derived these prerequisites the proof of Theorem 6 proceeds as follows.

Proof of Theorem 6. If ξ and η are infix-regular ω-words having $\mathbf{T}_\infty(\xi) = \mathbf{T}_\infty(\eta)$ then there are words u_ξ, u_η such that $\xi/u_\xi, \eta/u_\eta \in E := \{\zeta : \forall w(w \sqsubset \zeta \to w \in \mathbf{T}_\infty(\xi))\}$ (e.g. $u_\xi := w$ with $\mathbf{T}(\xi/w) = \mathbf{T}_\infty(\xi)$).

The ω-language E has the following properties:

1. E is closed,
2. $E/u \supseteq E/w \cdot u$ for all $w, u \in X^*$, and
3. $\dim E/v = \dim E$ if $E/v \neq \emptyset$.

E is a regular ω-language, because $\mathbf{T}_\infty(\xi)$ is a regular language. Thus the set of left derivatives $\{E/v : v \in X^*\}$ is finite, and contains, therefore, minimal (w.r.t. "$\subseteq$") nonempty derivatives. If E/u'_E is a minimal nonempty derivative of E we have $E/w \cdot u'_E = E/u'_E$ for all $w \in \mathbf{T}_\infty(\xi)$ such that also $w \cdot u'_E \in \mathbf{T}_\infty(\xi)$.

Since for every $u \in \mathbf{T}_\infty(\xi)$ there is a $v \in \mathbf{T}_\infty(\xi)$ such that $u \cdot v \cdot u \in \mathbf{T}_\infty(\xi)$, this implies that E/u'_E is strongly connected. Thus, following Proposition 18, $\mathbf{L}_\alpha(E/u'_E \cdot v) \leq 1$ for all $v \in X^*$. Again by Proposition 18, there is a $u_E \sqsupseteq u'_E$ with $\mathbf{L}_\alpha(E/u_E) = 1$.

Without loss of generality let now $F \subseteq (u_\xi \cdot w' \cdot u_E) \cdot E/u_E$ for a suitable w' such that $u_\xi \cdot w' \cdot u_E \sqsubset \xi$. Then according to Lemma 19 there is some $w \sqsubset \xi$ with $u_\xi \cdot w' \cdot u_E \sqsubseteq w \sqsubset \xi$ and $\mathbf{L}_\alpha(F/w) \geq 1$.

By virtue of $\mathbf{L}_\alpha(E/u'_E \cdot v) \leq 1$ for all $v \in X^*$ and $u_E \sqsupseteq u'_E$ we have $\mathbf{L}_\alpha(E/u_E \cdot w'') \leq 1$ for the word w'' satisfying $w = u_\xi \cdot w' \cdot u_E \cdot w''$. Thus $\mathbf{L}_\alpha(F/w) = \mathbf{L}_\alpha(E/u_E \cdot w'') = 1$ and $\xi/w \in F/w \subseteq E/u_E \cdot w''$. In particular, we have $\mathbf{L}_\alpha(E/u_E \cdot w'' \setminus F/w) = 0$, that is, $\dim(E/u_E \cdot w'' \setminus F/w) < \alpha$, because $E/u_E \cdot w'' \setminus F/w$ is a regular ω-language.

For w'' we have $u_E \cdot w'' \in \mathbf{T}_\infty(\xi) = \mathbf{T}_\infty(\eta)$. Consequently, there is a v' such that $u_\eta \cdot v' \cdot u_E \cdot w'' \sqsubset \eta$, whence $\eta/(u_\eta \cdot v' \cdot u_E \cdot w'') \in E/u_E \cdot w''$. Since $\tau(\eta) = \alpha$ and $\dim(E/u_E \cdot w'' \setminus F/w) < \alpha$, the assertion $\eta/(u_\eta \cdot v' \cdot u_E \cdot w'') \in F/w$ follows from Proposition 17.

Now let $w_F := w = u_\xi \cdot v' \cdot u_E \cdot w''$ and $v_F := u_\eta \cdot v' \cdot u_E \cdot w''$. □

References

[Bü60] J.R. Büchi, On a decision method in restricted second order arithmetic. Proc. 1960 Int. Congr. for Logic, Stanford Univ. Press, Stanford 1962, 1–11.

[Co83] K. Compton, On rich words, in: *Combinatorics on Words*. (L.J. Cummings ed.), Academic Press, Toronto 1983, 39–61.

[Ei74] S. Eilenberg, *Automata, Languages and Machines*. Vol. A, Academic Press, New York 1974.

[Ed90] G.A. Edgar, *Measure, Topology, and Fractal Geometry*. Springer-Verlag, New York 1990.

[Fa90] K.J. Falconer, *Fractal Geometry*. Wiley, Chichester 1990.

[JS83] H. Jürgensen, H.-J. Shyr and G. Thierrin, Disjunctive ω-languages, Elektron. Informationsverarb. Kybernetik EIK Vol. 19 (1983) No. 6, 267–278.

[Ku66] K. Kuratowski, *Topology I*. Academic Press, New York 1966.

[Ku70] W. Kuich, On the entropy of context-free languages, Inform. Control Vol. 16 (1970) No. 2, 173–200.

[MS94] W. Merzenich and L. Staiger, Fractals, dimension, and formal languages. RAIRO–Inform. Théor., Vol. 28 (1994), No. 3–4, 361–386.

[Ox71] J.C. Oxtoby, *Measure and Category*. Springer-Verlag, Berlin 1971.

[PS86] D. Perrin and P.E. Schupp, Automata on the integers, recurrence distinguishability, and the equivalence and decidability of monadic theories, in: Proc. 1st Symp. Logic in Computer Science, IEEE Press, 1986, 301–304

[Se84] A. Semenov, Decidability of monadic theories, in: Math. Found. of Comput. Sci. (M. Chytil and V. Koubek eds.) Lect. Notes Comput. Sci. 176, Springer-Verlag, Berlin 1984, 162–175.

[St80] L. Staiger, Measure and category in X^ω, in: Topology and Measure II (J. Flachsmeyer, Z. Frolik and F. Terpe, eds.), Part 2, Wiss. Beitr. Ernst–Moritz–Arndt–Univ. Greifswald, 1980, 129–136.

[St93] L. Staiger, Kolmogorov Complexity and Hausdorff Dimension, *Inform. and Comput*. Vol. 102 (1993), No. 2, 159–194.

[St97] L. Staiger, ω-Languages, in: *Handbook of Formal Languages*, G. Rozenberg and A. Salomaa (Eds.), Vol. 3, 339–387, Springer-Verlag, Berlin 1997.

[SW74] L. Staiger und K. Wagner, Automatentheoretische und automatenfreie Charakterisierungen topologischer Klassen regulärer Folgenmengen. Elektron. Informationsverarb. Kybernetik EIK Vol. 10 (1974) No. 7, 379–392.

[Th90] W. Thomas, Automata on Infinite Objects, in: *Handbook of Theoretical Computer Science*, J. Van Leeuwen (Ed.), Vol. B, 133–191, Elsevier, Amsterdam, 1990.

[Th97] W. Thomas, Languages, automata, and logic, in: *Handbook of Formal Languages*, G. Rozenberg and A. Salomaa (Eds.), Vol. 3, 389–455, Springer-Verlag, Berlin 1997.

[Wa79] K. Wagner, On ω-regular sets. Inform. and Control Vol. 43 (1979), 123–177.

This article was processed using the LaTeX macro package with LLNCS style