Fragments of first-order logic over infinite words

Volker Diekert   Manfred Kufleitner

Universität Stuttgart, FMI
Universitätsstraße 38
D-70569 Stuttgart, Germany

{diekert,kufleitner}@fmi.uni-stuttgart.de

Technical Report No. 2009/04, Computer Science

June 16th, 2009

Contents

Abstract

Abstract. We give topological and algebraic characterizations as well as language theoretic descriptions of the following subclasses of first-order logic FO[<] for ω-languages: Σ_2, FO^2, $\mathrm{FO}^2 \cap \Sigma_2$, and Δ_2 (and by duality Π_2 and $\mathrm{FO}^2 \cap \Pi_2$). These descriptions extend the respective results for finite words. In particular, we relate the above fragments to language classes of certain (unambiguous) polynomials. An immediate consequence is the decidability of the membership problem of these classes, but this was shown before by Wilke [28] and Bojańczyk [2] and is therefore not our main focus. The paper is about the interplay of algebraic, topological, and language theoretic properties.

1 Introduction

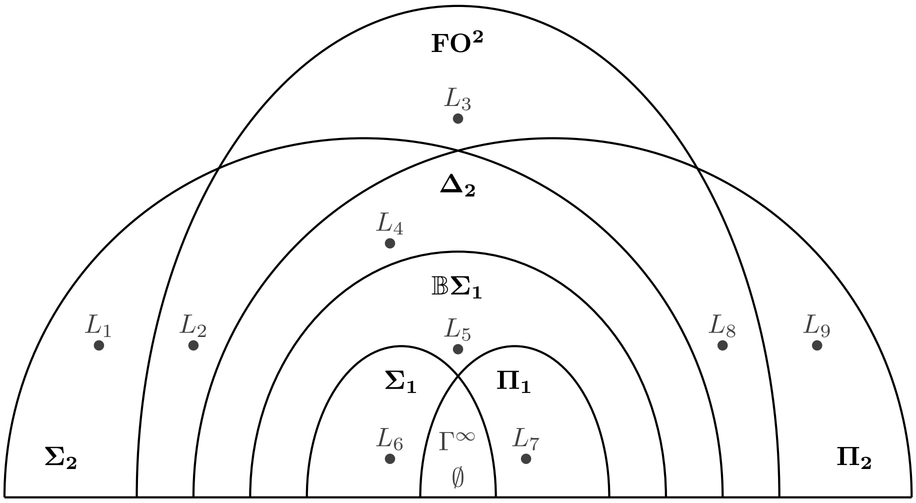
The algebraic approach is fundamental for the understanding of regular languages. It has been particularly fruitful for fragments of first-order logic over finite words. For example, a result of Wilke and Thérien is that FO^2 and Δ_2 have the same expressive power [22], where the latter class by definition denotes $\Sigma_2 \cap \Pi_2$. Further results are language theoretic and (very often decidable) algebraic characterizations of logical fragments, see e.g. [21] or [8] for surveys. Several results for finite words have been extended to other structures such as trees and other graphs, see [26] for a survey. More recently, FO^2, Δ_2, and Σ_2 have been characterized for Mazurkiewicz traces [9, 13]; Δ_2 and the Boolean closure of Σ_1 have been characterized for unranked trees [3, 4]. For some characterizations over finite words, it has been shown that they cannot be generalized; e.g. over unranked trees, it turned out that FO^2 and Δ_2 are incomparable [1]. For infinite words, the expressive power of FO^2 is not equal to Δ_2, since saying that letters a and b appear infinitely often, but c only finitely many times is FO^2-definable, but there is neither a Σ_2-formula nor a Π_2-formula specifying this language.

Our results deepen the understanding of first-order fragments over infinite words. A decidable characterization of the membership problem for FO^2 over infinite words has been given in the habilitation thesis of Wilke [28]. Recently, decidability for Σ_2 has been shown independently by Bojańczyk [2]. Language theoretic and decidable algebraic characterizations of the fragment Σ_1 and of its Boolean closure can be found in [15, 17].

We introduce two generalizations of the usual Cantor topology for infinite words. One of our first results is a characterization of languages $L \subseteq \Gamma^\infty$ being Σ_2-definable. This characterization consists of two components: The first one is an algebraic property of the syntactic monoid and the second part is requiring that L is open in some alphabetic topology. Both properties are decidable.

Our second result is that a regular language is FO^2-definable if and only if its syntactic monoid is in the variety **DA**. (The result is surprising in the sense that it contradicts an explicit statement in [28]). Moreover, we show that FO^2-definability can be characterized by being closed in some further refined alphabetic topology and in terms of weak recognition by some monoid in **DA**. In particular, weak recognition and strong recognition do not coincide for the variety **DA**. This seems to be a new result as well. We also contribute a language theoretic characterization of FO^2 in terms of unambiguous polynomials with additional constraints on the letters which occur infinitely often.

Other results of our paper are the characterization of $\mathrm{FO}^2 \cap \Sigma_2$ as the class of unambiguous polynomials and of Δ_2 in terms of unambiguous polynomials in some special form and also in terms of deterministic languages. It follows already from this description that Δ_2 is a strict subset of FO^2. Furthermore, we show that the equality of FO^2 and Δ_2 holds relativized to some fixed set of letters which occur infinitely often. If this set of letters is empty, we obtain the situation for finite words as a special case. Finally, we relate topological constructions such as *interior* and *closure* with membership in the fragments under consideration. Among other results, we are going to explain the following relations between the fragments FO^2, Σ_2, Π_2, and $\Delta_2 = \Sigma_2 \cap \Pi_2$ (for completeness we included the fragments Σ_1, Π_1, and their Boolean closure $\mathbb{B}\Sigma_1$ in the picture):



Here $\Gamma = \{a, b, c\}$ and

$$
\begin{aligned}
L_1 &= \text{“there exists a factor } ab\text{”} = \Gamma^* ab \Gamma^\infty, \\
L_2 &= \text{“finitely many } a\text{'s”} = \Gamma^* \{b, c\}^\infty, \\
L_3 &= \text{“finitely many } a\text{'s and infinitely many } b\text{'s”} = L_2 \cap L_8, \\
L_4 &= \text{“the first } a \text{ occurs before the first } b\text{”} = c^* a \Gamma^* b \Gamma^\infty, \\
L_5 &= \text{“some } a \text{ occurs before some } b \text{ but no } c \text{ occurs before some } a\text{”} \\
&= \Gamma^* a \Gamma^* b \Gamma^\infty \cap \{a, b\}^* a \{b, c\}^\infty = L_6 \cap L_7, \\
L_6 &= \text{“some } a \text{ occurs before some } b\text{”} = \Gamma^* a \Gamma^* b \Gamma^\infty, \\
L_7 &= \text{“no } c \text{ occurs before some } a\text{”} = \{a, b\}^* a \{b, c\}^\infty \cup \{b, c\}^\infty, \\
L_8 &= \text{“infinitely many } b\text{'s”} = (\Gamma^* b)^\omega, \\
L_9 &= \text{“there is no factor } ab\text{”} = \Gamma^\infty \setminus L_1.
\end{aligned}
$$

The intersection $\Delta_1 = \Sigma_1 \cap \Pi_1$ contains only the trivial languages $\emptyset$ and Γ^∞. It will turn out that L_8 is the closure of L_3 within some alphabetic topology, whereas L_2 is not the interior of L_3 since $L_3 \subsetneq L_2$. In fact, the interior of L_3 (as well as of any other language in Γ^ω) with respect to our topology is empty. A brief summary of the results for the various fragments can be found in Section 7 at the end of this paper.

For basic notions on languages of infinite words we refer to standard references such as [15, 24]. Most results of the present paper are from its conference version [10], but for lack of space they appeared in many cases without proof. The present journal version gives full proofs and some new material. In particular, we give a new characterization of ω-regular Δ_2-languages involving deterministic and complement-deterministic languages, cf. Corollary 6.9.

2 Preliminaries

Words Throughout, Γ is a finite alphabet, $A \subseteq \Gamma$ is a subset of the alphabet, u, v, w are finite words, and α, β, γ are finite or infinite words. If not specified otherwise, then in all examples we assume that Γ has three different letters a, b, c. By $u \leq \alpha$ we mean that u is a prefix of α. By $\mathrm{alph}(\alpha)$ we denote the *alphabet* of α, i.e., the letters occurring in the sequence α. As usual, Γ^* is the free monoid of finite words over Γ. The neutral element is the empty word 1. If L is a subset of a monoid, then L^* is the submonoid generated by L. For $L \subseteq \Gamma^*$ we let $L^\omega = \{u_1 u_2 \cdots \mid u_i \in L \text{ for all } i \geq 1\}$ be the set of infinite products. We also let $L^\infty = L^* \cup L^\omega$. A natural convention is $1^\omega = 1$. Thus, $L^\infty = L^\omega$ if and only if $1 \in L$.

We write $\mathrm{im}(\alpha)$ for those letters in $\mathrm{alph}(\alpha)$ which have infinitely many occurrences in α. The notation has been introduced in the framework of so called *complex traces*, see e.g. [12] for a detailed discussion of this concept. The notation $\mathrm{im}(\alpha)$ refers to the *imaginary part* and we adopt it here, but for our purpose it might be also convenient to remember $\mathrm{im}(\alpha)$ as an abbreviation for letters which appear *infinitely many* times in α. A crucial role in our paper play sets of the form A^{im}. By definition, A^{im} is the set of words α such that $\mathrm{im}(\alpha) = A$. Note that $\Gamma^* = \emptyset^{\mathrm{im}}$. The set Γ^∞ is the disjoint union over all A^{im}.

Logic and regular sets We assume that the reader is familiar with basic concepts in formal language theory. Our focus is on *regular* languages. If $L \cap \Gamma^\infty$ is regular, then we may think that its finitary part $L \cap \Gamma^*$ is specified by some NFA and that its infinitary part $L \cap \Gamma^\omega$ is specified by some Büchi automaton. For a unified model to accept regular languages in Γ^∞ it is convenient to consider an *extended Büchi automaton* which has a finite set of states Q and two types of accepting states, a set of *final* states $F \subseteq Q$ for accepting finite words and a set of *repeated* states $R \subseteq Q$ for accepting infinite words. Thus, this model yields also a natural definition of *deterministic regular* languages in Γ^∞, see below for more details.

We focus on regular languages which are given by first-order sentences in FO[<]. Thus, atomic predicates are $\lambda(x) = a$ and $x < y$ saying that position x in a word α is labeled with $a \in \Gamma$ and position x is smaller than y, respectively. By FO^2 we mean FO[<]-sentences which use at most two names x and y as variables or the class of languages specified by such formulas. It is well-known that three variable are sufficient to express any FO[<]-property (see e.g. [7]), whereas FO^2 is a strict subclass. Similarly, Σ_2 means FO[<]-sentences which are in prenex normal form and which start with a block of existential quantifiers, followed by a block of universal quantifiers and a Boolean combination of atomic formulas. A Π_2-formula means a negation of a Σ_2-formula. The notations Σ_2 and Π_2 refer also to the corresponding language classes. The class Δ_2 means the class of Σ_2-formulas which have an equivalent Π_2-formula. But the notion of equivalence depends on the set of models we use.

If the models are finite words, then a result of Thérien and Wilke [22] states $\mathrm{FO}^2 = \Delta_2$. Moreover, FO^2 is the class of regular languages in Γ^* which are recognized by some finite monoid in the variety **DA** and a classical result of Schützenberger shows that **DA** also coincides with unambiguous polynomials [18]. The variety **DA** has been baptized this way because it means *$\mathcal{D}$-classes are aperiodic*. More precisely, **DA** contains those finite monoids, where all regular $\mathcal{D}$-classes are aperiodic semigroups. We refer to [20, 8] for more background on the class **DA**. It is also the class of finite monoids defined e.g. by equations of type $(xy)^\omega = (xy)^\omega y (xy)^\omega$. Another characterization says that **DA** is defined by finite monoids M satisfying $e = ese$ for all idempotents e (i.e., $e^2 = e$) and for all $s = s_1 \cdots s_n$ where $e \in M s_i M$ for each i, see e.g. [5, 25]. This is the definition which we use below.

Saying that formulas are equivalent if they agree on all finite and infinite words refines the notion of equivalence for formulas and changes the picture. This is actually the starting point of this work. So, in this paper models are finite and infinite words. We are mainly interested in infinite words, but it does no harm to include finite words, and this makes the situation more uniform and the results on finite words reappear as special cases. See e.g. Theorem 5.10 which means $\mathrm{FO}^2 = \Delta_2$ for finite words by choosing $A = \emptyset$.

Recognizability by finite monoids By M we denote a finite monoid. We always assume that M is equipped with a partial order $\leq$ being compatible with the multiplication, i.e., $u \leq v$ implies $sut \leq svt$ for all $s, t, u, v \in M$. If not specified otherwise, we may choose $\leq$ to be the identity relation.

For an idempotent element $e \in M$ we define $M_e = \{s \in M \mid e \in MsM\}^*$, i.e., M_e is the submonoid of M which is generated by factors of e. If M has a generating set Γ, then M_e is generated by $\{a \in \Gamma \mid e \in MaM\}$. We can think of this set as the maximal alphabet of the idempotent e. We say that an idempotent e is *locally top* (*locally bottom*, resp.) if $ese \leq e$ ($ese \geq e$, resp.) for all $s \in M_e$. By **DA** we denote the class of finite monoids such that $ese = e$ for all idempotents $e \in M$ and all $s \in M_e$. Thus, it is the class of finite monoids where idempotents are locally top and locally bottom.

Let $L \subseteq \Gamma^\infty$ be a language. The *syntactic preorder* $\leq_L$ over Γ^* is defined as follows. We let

$u \leq_L v$ if for all $x, y, z \in \Gamma^*$ we have both implications:

$$xvyz^\omega \in L \;\Rightarrow\; xuyz^\omega \in L \qquad \text{and} \qquad x(vy)^\omega \in L \;\Rightarrow\; x(uy)^\omega \in L.$$

Let us recall that $1^\omega = 1$. Two words $u, v \in \Gamma^*$ are syntactically equivalent, written as $u \equiv_L v$, if both $u \leq_L v$ and $v \leq_L u$. This is a congruence and the congruence classes $[u]_L = \{v \in \Gamma^* \mid u \equiv_L v\}$ form the *syntactic monoid* $\mathrm{Synt}(L)$ of L. The preorder $\leq_L$ on words induces a partial order $\leq_L$ on congruence classes, and $(\mathrm{Synt}(L), \leq_L)$ becomes an ordered monoid. It is a well-known classical result that the syntactic monoid of a regular language $L \subseteq \Gamma^\infty$ is finite, see e.g. [15, 24]. Moreover, in this case L can be written as a finite union of languages of type $[u]_L\,[v]_L^\omega$ where $u, v \in \Gamma^*$ with $uv \equiv_L u$ and $v^2 \equiv_L v$. In contrast to finite words, there exist non-regular languages in Γ^∞ with a finite syntactic monoid.

Now, let $h : \Gamma^* \to M$ be any surjective homomorphism onto a finite ordered monoid M and let $L \subseteq \Gamma^\infty$. If the reference to h is clear, then we denote by $[s]$ the set of finite words $h^{-1}(s)$ for $s \in M$. We use the following notation.

- $(s, e) \in M \times M$ is a *linked pair*, if $se = s$ and $e^2 = e$.
- h *weakly recognizes* L, if

$$L = \bigcup \{[s][e]^\omega \mid (s, e) \text{ is a linked pair and } [s][e]^\omega \subseteq L\}$$

- h *strongly recognizes* L (or simply *recognizes* L), if

$$L = \bigcup \{[s][e]^\omega \mid (s, e) \text{ is a linked pair and } [s][e]^\omega \cap L \neq \emptyset\}$$

- L is *downward closed (on finite prefixes)* for h, if $[s][e]^\omega \subseteq L$ implies $[t][e]^\omega \subseteq L$ for all $s, t, e \in M$ where $t \leq s$.

If L is regular, then the syntactic homomorphism h_L strongly recognizes L.

Lemma 2.1 *Let $L \subseteq \Gamma^\infty$ be a regular language and let $h_L : \Gamma^* \to \mathrm{Synt}(L)$ be its syntactic homomorphism. Then for all $s, t, e, f \in M$ such that $t \leq s$, $f \leq e$, and $[s][e]^\omega \subseteq L$ we have $[t][f]^\omega \subseteq L$. In particular, L is downward closed (on finite prefixes) for h_L.*

Proof: Let $u \in [s]$, $x \in [e]$ and let $v \in [t]$, $y \in [f]$. Now, $ux^\omega \in L$ implies $vx^\omega \in L$, which in turn implies $vy^\omega \in L$. Since L is regular, h_L strongly recognizes L; and we obtain $[t][f]^\omega \subseteq L$, because $vy^\omega \in [t][f]^\omega \cap L$. □

Deterministic, complement-deterministic, and arrow languages Intuitively, the best way to define *deterministic languages* is to say that a language is *deterministic*, if it is recognized by a deterministic extended Büchi automaton with final and repeated states as described above. Therefore, a regular language $L \subseteq \Gamma^\infty$ is deterministic if and only if its ω-regular part $L \cap \Gamma^\omega$ can be accepted by some deterministic Büchi automaton in the usual sense.

There is also a well-known tight connection to what we call here *arrow languages* $\overrightarrow{W}$: For $W \subseteq \Gamma^*$ we define

$$\overrightarrow{W} = \{\alpha \in \Gamma^\infty \mid \text{for every prefix } u \leq \alpha \text{ there exists } uv \leq \alpha \text{ with } uv \in W\}.$$

Using Büchi automata, we see that a regular language $L \subseteq \Gamma^\infty$ is deterministic if and only if we can write $L \cap \Gamma^\omega = \overrightarrow{W} \cap \Gamma^\omega$ for some regular $W \subseteq \Gamma^*$. Actually, a classical result of Landweber yields a more precise statement: If $L \subseteq \Gamma^\omega$ is ω-regular and $L = \overrightarrow{W} \cap \Gamma^\omega$ for some set $W \subseteq \Gamma^*$, then W can be chosen to be regular, too (which means L is deterministic) see e.g. [24]. Therefore it is justified to take the weakest condition as a formal definition here. Moreover, as we have not formally defined Büchi automata, we use the Landweber characterization as our working definition: If we speak about a deterministic language then we are content with L being regular and $L \cap \Gamma^\omega = \overrightarrow{W} \cap \Gamma^\omega$ for some set $W \subseteq \Gamma^*$. It is called *complement-deterministic,* if $\Gamma^\infty \setminus L$ is deterministic. It is well-known and easy to see (e.g. with our working definition) that deterministic languages are closed under finite union and finite intersection.

For example, if $W = \Gamma^* a$, then $\overrightarrow{W} \cap \Gamma^\omega$ is the deterministic ω-regular language of words having infinitely many a's. Its complement is not deterministic (if $|\Gamma| \geq 2$). Hence *infinitely many a's* is not complement-deterministic. In particular, deterministic languages do not form a Boolean algebra, whereas the class of languages which are simultaneously deterministic and complement-deterministic does. Note that the class of arrow languages is not closed under finite intersection: $\overrightarrow{\Gamma^* a} \cap \overrightarrow{\Gamma^* b}$ is deterministic but no arrow language (in our sense) because the intersection is not empty, e.g., it contains $(ab)^\omega$, but it does not contain any finite word.

Our definitions differ slightly from the notation used elsewhere, where $\overrightarrow{W}$ is commonly used as the ω-language of those infinite words with infinitely many prefixes in W, which is the set $\overrightarrow{W} \cap \Gamma^\omega$ in our notation. In our definition we have however a closure operator: $W \subseteq \overrightarrow{W} = W \cup (\overrightarrow{W} \cap \Gamma^\omega)$. Moreover, the characterization of Δ_2-languages is more natural in our definition. Also note that if $L = \overrightarrow{W}$, then $W = L \cap \Gamma^*$. If we only have $L \cap \Gamma^\omega = \overrightarrow{W} \cap \Gamma^\omega$, then there are uncountably many choices for W, in general.

Finite ω-semigroups The notion of an ω-semigroup has been introduced as a tool for language varieties of finite and infinite words. Their use is justified by the existence of an Eilenberg-type theorem, see [15, 27]. This leads to another possible framework to express most of our results. In the present paper a reformulation in terms of ω-semigroups would mean however to introduce another technical concept which is not needed. Our focus is to transfer results from finite words to infinite words using topology, so the classical theory of recognition by finite monoids is perfectly suitable for our purposes.

To some extent it is a matter of taste to use one or another formalism to express the results. So, our choice is to work with less technical prerequisites in order to understand the contents of the paper. A conversion of our results to the terminology of ω-semigroups is left to the interested readers familiar with the theory ω-semigroups. We refer to the textbook [15], where the theory has been nicely presented in detail.

3 The alphabetic topology and polynomials

Topological information is crucial in our characterization results. Recall that a *topology* on a set X is given by a family of subsets (called *open subsets*) such that a finite intersection and an arbitrary union of open subsets is open. We define the *alphabetic topology* on the set Γ^∞ by its basis, which is given by all sets of the form uA^∞ for $u \in \Gamma^*$ and $A \subseteq \Gamma$. Thus, a set $L \subseteq \Gamma^\infty$ is *open* if and only if for each $A \subseteq \Gamma$ there is a set of finite words $W_A \subseteq \Gamma^*$ such that $L = \bigcup W_A A^\infty$. By definition, a set is *closed*, if its complement is open; and it is *clopen*, if it is both open and closed. For example, the sets uA^∞ are clopen. In particular, the sets A^∞ are clopen, too. A set of the form A^{im} is not open unless $A = \emptyset$, it is not closed unless $A = \Gamma$.

Note that in the alphabetic topology every singleton $u \in \Gamma^*$ is open since $u\emptyset^\infty = u\{1\} = \{u\}$. Thus, Γ^* is an open, discrete, and dense subset of Γ^∞. The alphabetic topology is a refinement of the usual Cantor topology, where the languages $\{u\}$ and $u\Gamma^\infty$ form a basis of (Cantor-)open subsets for $u \in \Gamma$. The Cantor space Γ^∞ is compact. As soon as Γ has at least two letters more sets are open in the alphabetic topology than in the Cantor topology. For example, the sets uA^∞ being clopen in the alphabetic topology are neither open nor closed in the Cantor topology for $\emptyset \neq A \neq \Gamma$.

Remark 3.1 *The space Γ^∞ with the alphabetic topology is Hausdorff. It is compact if and only if $|\Gamma| \leq 1$. To see that it is not compact for $\Gamma = \{a, b\}$ note that Γ^∞ is covered by a^∞ together with open sets of the form $ub\Gamma^\infty$ with $u \in \Gamma^*$. But for no finite subset $F \subseteq \Gamma^*$ we have $\Gamma^\infty = a^\infty \cup Fb\Gamma^\infty$.*

For a language L, its *closure* $\overline{L}$ is the intersection of all closed sets containing L. A word $\alpha \in \Gamma^\infty$ belongs to $\overline{L}$ if for all open subsets $U \subseteq \Gamma^\infty$ with $\alpha \in U$ we have $U \cap L \neq \emptyset$. The *interior* of L is the union of all open sets contained in L. It can be constructed as the complement of the closure of its complement. For languages L and K we define the right quotient as a language of finite words by $L/K = \{u \in \Gamma^* \mid u\alpha \in L \text{ for some } \alpha \in K\}$. In particular, we have

$$L/A^\infty = \{u \in \Gamma^* \mid u\alpha \in L \text{ for some } \alpha \in A^\infty\}\,.$$

The following proposition gives a description of the closure in the alphabetic topology in terms of arrow languages $\overrightarrow{W}$ plus some alphabetic restrictions.

Proposition 3.2 *In the alphabetic topology we have* $\overline{A^{\mathrm{im}}} = \bigcup_{A \subseteq B} B^{\mathrm{im}}$ *and*

$$\overline{L} = \bigcup_{A \subseteq \Gamma} \left(\overrightarrow{L/A^\infty} \cap A^{\mathrm{im}}\right) = \bigcup_{A \subseteq \Gamma} \left(\overrightarrow{L/A^\infty} \cap \overline{A^{\mathrm{im}}}\right).$$

Proof: It is straightforward to show $\overline{A^{\mathrm{im}}} = \bigcup_{A \subseteq B} B^{\mathrm{im}}$. We first show $\overline{L} \subseteq \bigcup_{A \subseteq \Gamma} \left(\overrightarrow{L/A^\infty} \cap A^{\mathrm{im}}\right)$. Let $\alpha \in \overline{L}$ with $\alpha \in A^{\mathrm{im}}$. For all prefixes u of α we find v such that $\alpha \in uvA^\infty$. We have $uvA^\infty \cap L \neq \emptyset$; and thus $uv \in L/A^\infty$. This shows $\alpha \in \overrightarrow{L/A^\infty}$.

The inclusion $\bigcup_{A \subseteq \Gamma} \left(\overrightarrow{L/A^\infty} \cap A^{\mathrm{im}}\right) \subseteq \bigcup_{A \subseteq \Gamma} \left(\overrightarrow{L/A^\infty} \cap \overline{A^{\mathrm{im}}}\right)$ is trivial.

Let now $\alpha \in \overrightarrow{L/A^\infty} \cap B^{\mathrm{im}}$ with $A \subseteq B$. Since $L/A^\infty \subseteq L/B^\infty$, we have $\alpha \in \overrightarrow{L/B^\infty} \cap B^{\mathrm{im}}$. Let $u \in \Gamma^*$ with $\alpha = u\beta$ and $\beta \in B^\infty$. We have to show $uB^\infty \cap L \neq \emptyset$. Since $\alpha \in \overrightarrow{L/B^\infty}$ there is some $v \in \Gamma^*$ with $uv \leq \alpha$ and $uv \in L/B^\infty$. This means $uv\gamma \in L$ for some $\gamma \in B^\infty$. Since $\beta \in B^\infty$ we have $v \in B^*$. Hence $v\gamma \in B^\infty$ and thus $uv\gamma \in uB^\infty \cap L \neq \emptyset$ as desired. □

The following corollary generalizes a well-known fact for the Cantor topology to the (finer) alphabetic topology. This result will be used in Section 6.

Corollary 3.3 *Let $L \subseteq \Gamma^\infty$ be a regular language. Then its closure in the alphabetic topology $\overline{L}$ is deterministic.*

Proof: Deterministic languages are closed under finite union and finite intersection. For a letter a the language $\overline{\{a\}^{\mathrm{im}}}$ is deterministic as it is the language of words having infinitely many a's. Hence $\overline{A^{\mathrm{im}}} = \bigcap_{a \in A} \overline{\{a\}^{\mathrm{im}}}$ is deterministic, too. The result follows. □

Corollary 3.4 *Given a regular language $L \subseteq \Gamma^\infty$, we can decide whether L is closed (open resp., clopen resp.).*

Proof: We may assume that L is specified by some NFA for $L \cap \Gamma^*$ and by some Büchi automaton for $L \cap \Gamma^\omega$. The construction of an NFA recognizing L/A^∞ is standard. Since $L/A^\infty \subseteq \Gamma^*$ we can assume that the NFA is deterministic, and we can view it as a (deterministic) Büchi automaton recognizing $\overrightarrow{L/A^\infty} \cap \Gamma^\omega$. Intersection with A^{im} yields a Büchi automaton for $\overline{L} \cap A^{\mathrm{im}}$ and $A \neq \emptyset$. Thus, we can test $\overline{L} \cap A^{\mathrm{im}} \subseteq L$ for all A. This implies that we can test $L = \overline{L}$. The result for *open* and *clopen* follows since regular languages are effectively closed under complementation. □

Actually, we have a more precise statement than pure decidability. In the following, PSPACE denotes as usual the class of problems which can be decided by some polynomially space bounded (deterministic) Turing machine.

Theorem 3.5 *The following problem is* PSPACE*-complete:*
Input: A Büchi automaton $\mathcal{A}$ with $L(\mathcal{A}) \subseteq \Gamma^\omega$.
Question: Is the regular language $L(\mathcal{A})$ closed?

Proof: We can check in PSPACE whether a regular language $L \subseteq \Gamma^\omega$ is closed: Let $L = L(\mathcal{A})$ for some non-deterministic Büchi automaton $\mathcal{A}$. We verify $L = \overline{L}$ using the characterization of $\overline{L}$ given in Proposition 3.2. We can check in PSPACE whether two Büchi automata are equivalent, see [19]. In particular, we can check in PSPACE whether $L \cap A^{\mathrm{im}} = \overrightarrow{L/A^\infty} \cap A^{\mathrm{im}}$ for all $A \subseteq \Gamma$.

It is PSPACE-hard to decide whether a regular language $L \subseteq \Gamma^\omega$ is closed: We use a reduction of the problem whether $L(\mathcal{A}) = \Gamma^*$ for some NFA $\mathcal{A}$, see [14]. We can assume that $1 \in L(\mathcal{A})$. Let $c \notin \Gamma$ be a new letter. We can construct a non-deterministic Büchi automaton $\mathcal{B}$ such that $L(\mathcal{B}) = \{w_1 c w_2 c \cdots \in (\Gamma \cup \{c\})^\omega \mid \exists i\colon w_i \in L(\mathcal{A})\}$. The closure of $L(\mathcal{B})$ is $K = \{w_1 c w_2 c \cdots \in (\Gamma \cup \{c\})^\omega \mid \forall i\colon w_i \in \Gamma^*\} = (\Gamma^* c)^\omega$. Hence, $L(\mathcal{A}) = \Gamma^*$ if and only if $L(\mathcal{B}) = K$ if and only if $L(\mathcal{B})$ is closed. □

According to Proposition 3.2 the alphabetic closure is a union over languages of type $\overrightarrow{L/A^\infty}$ or $\overrightarrow{L/A^\infty} \cap \overline{A^{\mathrm{im}}}$. But these pieces need not to be closed, as we can see in the following example.

Example 3.6 Let $A = \{a\}$, $B = \{a, b\}$, and $L = a^*(ab)^*ba^\omega$. Then $L/A^\infty = a^*(ab)^*ba^*$ and L/B^∞ is the set of all finite prefixes of words in L. We have $\overrightarrow{L/A^\infty} = a^*(ab)^*ba^\infty$ and $\overrightarrow{L/A^\infty} \cap \overline{A^{\mathrm{im}}} = a^*(ab)^*ba^\omega = L$. The language $\overrightarrow{L/A^\infty}$ is open but neither $\overrightarrow{L/A^\infty}$ nor $\overrightarrow{L/A^\infty} \cap \overline{A^{\mathrm{im}}}$ is closed in the alphabetic topology, because $(ab)^\omega$ belongs to both closures. We have $\overrightarrow{L/B^\infty} = a^*(ab)^*ba^\infty \cup a^*(ab)^\omega$ and $\overrightarrow{L/B^\infty} \cap B^{\mathrm{im}} = a^*(ab)^\omega$. Both sets are closed. Actually, $\overline{L} = L \cup a^*(ab)^\omega$ in the alphabetic topology.

The alphabetic closure $\overline{L}$ is not closed in the Cantor topology since $a^\omega \notin \overline{L}$, but every Cantor-open neighborhood of a^ω contains a word $a^n(ab)^\omega$ for some $n \in \mathbb{N}$. ◇

Frequently we apply the closure operator to polynomials. A *polynomial* is a finite union of monomials. A *monomial* (of *degree k*) is a language of the form $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ with $a_i \in \Gamma$ and $A_i \subseteq \Gamma$. In particular, $A_1^* a_1 \cdots A_k^* a_k$ is a monomial with $A_{k+1} = \emptyset$. The set A^* is a polynomial since $A^* = \emptyset^\infty \cup \bigcup_{a \in A} A^* a$. It is not hard to see that polynomials are closed under intersection. Thus, $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^* = A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty \cap \Gamma^*$ is in our language a polynomial, but not a monomial unless $A_{k+1} = \emptyset$. A monomial $P = A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ is *unambiguous* if for every $\alpha \in P$ there exists a unique factorization $\alpha = u_1 a_1 \cdots u_k a_k \beta$ such that $u_i \in A_i^*$ and $\beta \in A_{k+1}^\infty$. A polynomial is *unambiguous* if it is a finite union of unambiguous monomials.

It follows from the definition of the alphabetic topology that polynomials are open. Actually, it is the coarsest topology with this property. The crucial observation is that we have a syntactic description of the closure of a polynomial as a finite union of other polynomials. For later use we make a more precise statement by considering the closure with respect to different subsets B at infinity.

Lemma 3.7 *Let $P = A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ be a monomial and $L = P \cap B^{\mathrm{im}}$ for some $B \subseteq A_{k+1}$. Then the closure of L is given by*

$$\overline{L} = \bigcup_{\{a_i, \ldots, a_k\} \cup B \subseteq A \subseteq A_i} A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty \cap A^{\mathrm{im}}.$$

Proof: First consider an index i with $1 \leq i \leq k+1$ such that $\{a_i, \ldots, a_k\} \cup B \subseteq A \subseteq A_i$. Let $\alpha \in A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty \cap A^{\mathrm{im}}$. We have to show that α is in the closure of L. Let $\alpha = u\beta$ with $u \in A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^*$ and $\beta \in A^\infty \cap A^{\mathrm{im}}$. We show that $uA^\infty \cap L \neq \emptyset$. Choose some $\gamma \in B^\infty \cap B^{\mathrm{im}}$. As $B \subseteq A_{k+1}$ holds by hypothesis, we see that $u a_i \cdots a_k \gamma \in P$, and hence $u a_i \cdots a_k \gamma \in uA^\infty \cap L$.

Let now $\alpha \in \overline{L}$ and write $\alpha \in u v_1 \cdots v_{k+1} A^\infty \cap A^{\mathrm{im}}$ with $\mathrm{alph}(v_j) = A$. There exists $\gamma \in A^\infty$ such that $u v_1 \cdots v_{k+1} \gamma \in P \cap B^{\mathrm{im}}$. This implies $B \subseteq A$. Since $u v_1 \cdots v_{k+1} \gamma \in A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ there are some $1 \leq i, j \leq k+1$ such that $u v_1 \cdots v_{j-1}$ belongs to $A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^*$, $v_j \in A_i^*$, and $v_{j+1} \cdots v_{k+1} \gamma \in A_i^* a_i \cdots A_k^* a_k A_{k+1}^\infty \cap A^\infty$. Therefore $\{a_i, \ldots, a_k\} \subseteq A \subseteq A_i$, too. It follows that $\alpha \in A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty \cap A^{\mathrm{im}}$. □

As usual, let $L \subseteq \Gamma^\infty$ be a regular language. Let us define $tf^\omega \leq_L se^\omega$ for linked pairs (s,e), (t,f) by the implication:

$$[s][e]^\omega \subseteq L \;\Rightarrow\; [t][f]^\omega \subseteq L.$$

With this notation we can give an algebraic characterization of being open.

Lemma 3.8 *A regular language $L \subseteq \Gamma^\infty$ is open in the alphabetic topology if and only if for all linked pairs (s,e), (t,f) of $M = \mathrm{Synt}(L)$ with $t, f \in M_e$ we have $stf^\omega \leq_L se^\omega$.*

Proof: Let L be open and $\alpha \in [s][e]^\omega \subseteq L$. We find a finite prefix $u \in [s]$ of α such that $\alpha \in uA^\infty \subseteq L$. Since $t, f \in M_e$ we may assume $\mathrm{alph}(vw) \subseteq A$ for some $v \in [t]$ and $w \in [f]$. Hence, $uvw^\omega \in [st][f]^\omega \subseteq L$. This shows $stf^\omega \leq_L se^\omega$.

For the converse, suppose that for all linked pairs $(s,e),(t,f)$ of $M = \mathrm{Synt}(L)$ with $t,f \in M_e$ we have $stf^\omega \leq_L se^\omega$. Let $\alpha \in [s][e]^\omega \subseteq L$. Write $\alpha = u\beta$ with $u \in [s]$ and $\beta \in [e]^\omega \cap A^\infty \cap A^{\mathrm{im}}$. Now, any $\gamma \in A^\infty$ can be written as $\gamma \in [t][f]^\omega$ for some linked pair with $t, f \in M_e$. Indeed, we have $A^* \subseteq [M_e]$: consider $a \in A$ and let $p, q \in A^*$ such that $paq \in [e]$. Then $a \in [M_e]$ and therefore $A \subseteq [M_e]$. Since M_e is a submonoid, $[M_e]$ is a submonoid of Γ^* and hence $A^* \subseteq [M_e]$. By assumption $u\gamma \in [st][f]^\omega \subseteq L$. It follows $uA^\infty \subseteq L$, i.e., L is open. □

4 The fragment Σ_2

By a (slight extension of a) result of Thomas [23] on ω-languages we know that a language $L \subseteq \Gamma^\infty$ is definable in Σ_2 if and only if L is a polynomial. However, this statement alone does not yield decidability. It turns out that we obtain decidability by a combination of an algebraic and a topological criterion. (This decidability result has also been shown independently by Bojańczyk [2] using different techniques.) We know that polynomials are open. Therefore, we concentrate on algebra.

Lemma 4.1 *If $L \subseteq \Gamma^\infty$ is a polynomial, then all idempotents of* $\mathrm{Synt}(L)$ *are locally top.*

Proof: By h_L we denote the syntactic homomorphism $\Gamma^* \to \mathrm{Synt}(L)$. Let $n \in \mathbb{N}$ such that L is a finite union of monomials of degree less than n. Let $h_L(e)$ be idempotent; in particular $e^n \equiv_L e$. For $e \equiv_L f$ we may assume that $\mathrm{alph}(f) \subseteq \mathrm{alph}(e)$. This means we take the maximal possible alphabet for e. Let $s \in \mathrm{alph}(e)^*$. We want to show that $xeseyz^\omega \in L$ if $xeyz^\omega \in L$.

Suppose $u = xe^nyz^\omega \in A_1^*a_1 \cdots A_k^*a_kA_{k+1}^\infty \subseteq L$ and $k < n$. Since there are at most $n-1$ letters a_i, some factor e of u lies completely within one of the A_i^* or within A_{k+1}^∞, i.e., $\mathrm{alph}(e) \subseteq A_i$ for some $1 \leq i \leq k+1$. Hence, $ese \in A_i^*$ and $xe^{n_1}se^{n_2}yz^\omega \in A_1^*a_1 \cdots A_k^*a_kA_{k+1}^\infty \subseteq L$ for some $n_1, n_2 \geq 1$. Since $h_L(e)$ is idempotent, it follows that $xeyz^\omega \in L$ implies $xeseyz^\omega \in L$. Similarly, $x(ey)^\omega \in L$ implies $x(esey)^\omega \in L$ and therefore $ese \leq_L e$ for all $s \in \mathrm{alph}(e)^*$, i.e., $h_L(e)$ is locally top. □

Theorem 4.2 *Let $L \subseteq \Gamma^\infty$ be a regular language. The following five assertions are equivalent:*

1. *L is Σ_2-definable.*
2. *L is a polynomial.*
3. *L is open in the alphabetic topology and all idempotents of* $\mathrm{Synt}(L)$ *are locally top.*
4. *The syntactic monoid $M = \mathrm{Synt}(L)$ and the syntactic order $\leq_L$ satisfy:*
   (a) *For all linked pairs (s,e), (t,f) with $t, f \in M_e$ we have $stf^\omega \leq_L se^\omega$.*
   (b) *$e = e^2$ and $s \in M_e$ implies $ese \leq_L e$.*
5. *The following three conditions hold for some homomorphism $h : \Gamma^* \to M$ which weakly recognizes L:*

(a) L is open in the alphabetic topology.

(b) All idempotents of M are locally top.

(c) L is downward closed (on finite prefixes) for h.

Proof: "1 ⇔ 2": This is a slight modification of a result by Thomas [23].

"2 ⇒ 3": By definition, polynomials are open in the alphabetic topology. In Lemma 4.1 it has been shown that all idempotent elements are locally top.

"3 ⇔ 4": The equivalence of L being open and "4a" is Lemma 3.8. Property "4b" is the definition of all elements being locally top.

"4 ⇒ 5": Let $h = h_L$ be the syntactic homomorphism onto the syntactic monoid $M = \mathrm{Synt}(L)$. Since L is regular, the homomorphism h strongly recognizes L. Applying Lemma 3.8, property "5a" follows from "4a" and "5b" trivially follows from "4b". The condition "5c" holds for $\mathrm{Synt}(L)$ by Lemma 2.1.

"5 ⇒ 2": Consider $\alpha \in L$ with $\mathrm{im}(\alpha) = A$. By "5a" the language L is open. Hence, there exists a prefix u of α such that $\alpha \in uA^\infty \subseteq L$. From the case of finite words and the hypothesis "5b" on M, we know that $P = \{v \in \Gamma^* \mid h(v) \leq h(u)\}$ is a polynomial. We can assume that all monomials in P end with a letter. We define the polynomial $P_\alpha = PA^\infty$. Clearly, $L \subseteq \bigcup\{P_\alpha \mid \alpha \in L\}$ and this union is finite since M is finite. It remains to show that $P_\alpha \subseteq L$ for $\alpha \in L$. Let $v \in P$ and $\beta \in A^\infty$. We know $u\beta \in L$ and there exists a linked pair (s, e) such that $u\beta \in [s][e]^\omega \subseteq L$. Now, there exists $w\gamma = \beta$ such that $uw \in [s]$ and $\gamma \in [e]^\omega$. By definition of P, we have $h(v) \leq h(u)$ and therefore $t = h(vw) \leq h(uw) = s$. It follows $v\beta = vw\gamma \in [t][e]^\omega \subseteq L$ by "5c". This shows $P_\alpha \subseteq L$ and thus $L = \bigcup\{P_\alpha \mid \alpha \in L\}$. □

Corollary 4.3 *It is decidable whether a regular language is Σ_2-definable.*

Proof: The syntactic congruence is computable and the conditions in "3" (or "4") of Theorem 4.2 are decidable. □

Remark 4.4 *An ω-language $L \subseteq \Gamma^\omega$ is Σ_2-definable, if $L = \{\alpha \in \Gamma^\omega \mid \alpha \models \varphi\}$ for some $\varphi \in \Sigma_2$. This is equivalent with $L \cup \Gamma^*$ being Σ_2-definable as a subset of Γ^∞. Thus, the decidability of Corollary 4.3 transfers to ω-regular languages.*

Of course, complementation yields dual results for the fragment Π_2. In particular, Π_2-definable languages are closed in the alphabetic topology.

5 Two variable first-order logic

Etessami, Vardi, and Wilke have given a characterization of FO^2 in terms of unary temporal logic [11]. In the same paper, they considered the satisfiability problem for FO^2. In this section, we continue the study of FO^2 over infinite words.

The following lemma can be proved essentially in the same way as for finite words. The result is also (implicitly) stated in the habilitation thesis of Wilke [28].

Lemma 5.1 *Let $L \subseteq \Gamma^\infty$ be FO^2-definable. Then the syntactic monoid $\mathrm{Synt}(L)$ is in $\mathbf{DA}$.*

Proof: Let $L = L(\varphi)$ for some FO^2-formula of quantifier depth n. Let $e^2 = e \in M = \mathrm{Synt}(L)$ and let $s \in M_e$. We can choose words $v, w \in \Gamma^*$ such that $h_L(v) = s$, $h_L(w) = e$, and, moreover, $\mathrm{alph}(v) \subseteq \mathrm{alph}(w)$. Now, consider words of the form $\alpha = xw^n vw^n yz^\omega$, $\alpha' = xw^n yz^\omega$ and $\beta = x(w^n vw^n y)^\omega$, $\beta' = x(w^n y)^\omega$. An Ehrenfeucht-Fraïssé-game for FO^2 shows that $\alpha \in L$ if and only if $\alpha' \in L$. Analogously, $\beta \in L$ if and only if $\beta' \in L$. Thus, $\mathrm{Synt}(L) \in \mathbf{DA}$. □

A set like A^{im} is FO^2-definable, but it is neither open nor closed in the alphabetic topology, in general. Therefore, we need a refinement of the alphabetic topology. As a basis for the *strict alphabetic topology* we take all sets of the form $uA^\infty \cap A^{\mathrm{im}}$. Thus, more sets are open (and closed) than in the alphabetic topology. Another way to define the strict alphabetic topology is to say that it is the coarsest topology on Γ^∞ where all sets of the form $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty \cap B^{\mathrm{im}}$ are open. The strict alphabetic topology is not used outside this section, but it is essential here in order to prove the converse of Lemma 5.1.

Lemma 5.2 *If $L \subseteq \Gamma^\infty$ is strongly recognized by some homomorphism $h : \Gamma^* \to M \in \mathbf{DA}$, then L is clopen in the strict alphabetic topology.*

Proof: Since h also strongly recognizes $\Gamma^\infty \setminus L$ as well, it is enough to show that L is open. Let $\alpha \in L$ with $\alpha \in [s][e]^\omega$ for some linked pair (s, e) and let $A = \mathrm{im}(\alpha)$. We show that $[s]A^\infty \cap A^{\mathrm{im}} \subseteq L$. Indeed, let $\beta \in [s]A^\infty \cap A^{\mathrm{im}}$. Then we have $\beta = uv\gamma$ with $h(u) = s$, $h(v) = r$, $\gamma \in [f]^\omega$ where $v \in A^*$, $\mathrm{alph}(\gamma) = \mathrm{im}(\gamma) = A$, and (r, f) is a linked pair. Since $M \in \mathbf{DA}$, we obtain $s = se = serfe = srfe$ and $efe = e$ and $fef = f$. We have $[sr][fef]^\omega \cap [srfe][efe]^\omega \neq \emptyset$ and $[srfe][efe]^\omega = [s][e]^\omega \subseteq L$. Since h strongly recognizes L, we have $[sr][f]^\omega = [sr][fef]^\omega \subseteq L$, too. In particular, $\beta \in L$. □

Lemma 5.3 *If L is closed in the strict alphabetic topology and if L is weakly recognized by some homomorphism $h : \Gamma^* \to M \in \mathbf{DA}$, then L is a finite union of languages $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty \cap A_{k+1}^{\mathrm{im}}$, where each $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ is an unambiguous monomial.*

Proof: Let $\alpha \in L$. Write $\alpha = u\beta$ with $\beta \in A^\infty \cap A^{\mathrm{im}}$ for some $A \subseteq \Gamma$. There is a linked pair (s, e) with $\alpha \in [s][e]^\omega \subseteq L$ and we may assume $h(u) = s$ and $\beta \in [e]^\omega$. For $A = \emptyset$ we have $[s] \subseteq L$ and, using our knowledge about the finite case, we may include $[s]$ in our finite union of unambiguous polynomials. Therefore, let $A \neq \emptyset$. We may choose an unambiguous monomial $P = A_1^* a_1 \cdots A_k^* a_k \subseteq [s]$ such that $u \in P$ and each last position of every letter $a \in \{a_1, \ldots, a_k\} \cup A_1 \cup \cdots \cup A_k$ occurs explicitly as some a_j in the expression P. Note that $[s]$ is a finite union of such monomials. Moreover, we may assume that $uv \in P$ for infinitely many prefixes $v \leq \beta$. Each such uv can uniquely be written as $uv = v_1 a_1 \cdots v_k a_k$ with $v_i \in A_i^*$. This yields a vector in $\mathbb{N}^k$ by $(|v_1 a_1|, |v_1 a_1 v_2 a_2|, \ldots, |v_1 a_1 \cdots v_k a_k|)$ for every $uv \in P$. By Dickson's Lemma [6], we may assume that this sequence of vectors is in no component decreasing when v gets longer. Hence (after removing finitely many v's) we may assume there is some i such that the component $|v_1 a_1 \cdots v_i a_i|$ is constant and $|v_1 a_1 \cdots v_i a_i v_{i+1} a_{i+1}|$ is strictly increasing. It follows that we may assume $\{a_{i+1}, \ldots, a_k\} \subseteq \mathrm{alph}(v_{i+1}) = A \subseteq A_{i+1}$. In particular, $\alpha \in A_1^* a_1 \cdots A_i^* a_i A^\infty \cap A^{\mathrm{im}}$. It is clear that this expression is unambiguous.

It remains to show $A_1^* a_1 \cdots A_i^* a_i A^\infty \cap A^{\mathrm{im}} \subseteq L$. Consider $u'\gamma$ with $u' \in A_1^* a_1 \cdots A_i^* a_i$ and $\gamma \in A^\infty \cap A^{\mathrm{im}}$. Since L is closed, it is enough to show that $u'\gamma$ belongs to the closure of L in the strict alphabetic topology. Choose any prefix $w \leq \gamma$. It is enough to show that $u'wA^\infty \cap A^{\mathrm{im}} \cap L \neq \emptyset$. Let $z \in \Gamma^*$ with $\mathrm{alph}(z) = A$ and $h(z) = e$. Since $w \in A^* \subseteq A_{i+1}^*$, we have $u'wa_{i+1} \cdots a_k \in P \subseteq [s]$. Hence $u'wa_{i+1} \cdots a_k z^\omega \in [s][e]^\omega \subseteq L$. □

The next statement follows again as in the case of finite words.

Lemma 5.4 *Every language A^{im} and every unambiguous monomial $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ is FO^2-definable.*

Proof: The language of non-empty words in A^{im} is defined by the FO^2-sentence

$$\bigwedge_{a \in A} \forall x \exists y\colon x < y \,\wedge\, \lambda(y) = a \;\wedge\; \bigwedge_{b \notin A} \exists x \forall y\colon x < y \,\wedge\, \lambda(y) \neq b.$$

We use induction on k in order to show that $P = A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ is FO^2-definable. Clearly, for $k = 0$ this is true. Let now $k \geq 1$. By unambiguity, we cannot have $\{a_1, \ldots, a_k\} \subseteq A_1 \cap A_{k+1}$

since for $(a_1 \cdots a_k)^2$ there would exist two different factorizations. First, suppose $a_i \notin A_{k+1}$. Let $\alpha = \alpha_1 a_i \alpha_2 \in P$ where $a_i \notin \mathrm{alph}(\alpha_2)$. There are two possibilities: the last a_i of α could be one of the a_j's, $i \leq j \leq k$, and then

$$\alpha_1 \in A_1^* a_1 \cdots A_j^*, \quad a_i = a_j, \quad \alpha_2 \in A_{j+1}^* a_{j+1} \cdots A_k^* a_k A_{k+1}^\infty$$

or it matches some A_j^*, $i < j < k+1$ and then

$$\alpha_1 \in A_1^* a_1 \cdots A_j^*, \quad a_i \in A_j, \quad \alpha_2 \in A_j^* a_j \cdots A_k^* a_k A_{k+1}^\infty.$$

In any case, the remaining four polynomials are unambiguous and their degree is strictly smaller than k. Hence, by induction we have FO^2-formulas describing them. Obviously, we can also express intersections with languages of the form B^* or B^∞ for $B \subseteq \Gamma$. So there is a finite list of FO^2-formulas such that for each $\alpha \in P$ there are formulas φ and ψ from the list and a letter $a \in \Gamma$ with $\alpha \in L(\varphi)aL(\psi) \subseteq P$ and $L(\psi) \subseteq (\Gamma \setminus \{a\})^\infty$. Now, the last a-position x in every $\alpha \in L(\varphi)aL(\psi)$ is uniquely defined by

$$\xi(x) \quad = \quad \lambda(x) = a \wedge \forall y \colon x < y \Rightarrow \lambda(y) \neq a.$$

Using relativization techniques, we now define FO^2-sentences $\varphi_{<a}$ and $\psi_{>a}$ such that $L(\varphi)aL(\psi) = L\big(\varphi_{<a} \wedge \exists x \colon \xi(x) \wedge \psi_{>a}\big)$. We give the inductive construction for $\psi_{>a}$. The other one for $\varphi_{<a}$ is symmetric. Atomic formulas are unchanged and Boolean connectives are straightforward. Existential quantification is as follows: $(\exists x \colon \zeta)_{>a} \ = \ \exists x \colon (\exists y \colon y < x \wedge \xi(y)) \wedge \zeta_{>a}$.

The case $a_i \notin A_1$ is similar (using a factorization of α at the first a_i-position). □

Theorem 5.5 *Let $L \subseteq \Gamma^\infty$. The following assertions are equivalent:*

1. *L is FO^2-definable.*
2. *L is regular and $\mathrm{Synt}(L) \in \mathbf{DA}$.*
3. *L is strongly recognized by some homomorphism $h : \Gamma^* \to M \in \mathbf{DA}$.*
4. *L is closed in the strict alphabetic topology and L is weakly recognized by some homomorphism $h : \Gamma^* \to M \in \mathbf{DA}$.*
5. *L is a finite union of sets of the form $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty \cap A_{k+1}^{\mathrm{im}}$, where each language $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ is an unambiguous monomial.*

Proof: "1 ⇒ 2": First-order definable languages are regular; $\mathrm{Synt}(L) \in \mathbf{DA}$ by Lemma 5.1. "2 ⇒ 3": Trivial, since $\mathrm{Synt}(L)$ strongly recognizes L. "3 ⇒ 4": Strong recognition implies weak recognition; closure in the strict alphabetic topology follows by Lemma 5.2. "4 ⇒ 5": Lemma 5.3. "5 ⇒ 1": Lemma 5.4. □

Recall that if a language $L \subseteq \Gamma^\infty$ is weakly recognizable by some finite monoid, then it is also strongly recognizable by a finite monoid. The same holds for aperiodic monoids, but Theorem 5.5 suggests that this fails for $\mathbf{DA}$. Indeed, we have the following example.

Example 5.6 Let $\Gamma = \{a, b, c\}$. Consider the congruence of finite index such that each class $[u]$ is defined by the set of words v where u and v agree on all suffixes of length at most 2. The quotient monoid of Γ^* by this congruence is in $\mathbf{DA}$. In fact, it is a very simple monoid within $\mathbf{DA}$ since it is $\mathcal{L}$-trivial (where $\mathcal{L}$ is one of Green's relations, see e.g. [15]). Let $L = [ab]^\omega = (\Gamma^* ab)^\omega$. Then, by definition, L is weakly recognizable in $\mathbf{DA}$. But L is the language of all α which contain infinitely many factors of the form ab. This is however not closed for the strict alphabetic topology since $(acb)^\omega \notin L$, but $(acb)^\omega$ belongs to the strict alphabetic closure of L since every open set U with $(acb)^\omega \in U$ contains some $(acb)^m (cab)^\omega$ and $[(acb)^m (cab)] = [ab]$ for all $m \geq 0$. ◇

5.1 Unambiguous polynomials and the fragment $\mathbf{FO^2} \cap \boldsymbol{\Sigma_2}$

In this section, we show that the intersection of FO^2 and Σ_2 has very natural descriptions involving topological notions or unambiguous polynomials.

Theorem 5.7 *Let $L \subseteq \Gamma^\infty$. The following assertions are equivalent:*

1. *L is both FO^2-definable and Σ_2-definable.*
2. *L is FO^2-definable and open in the alphabetic topology.*
3. *L is a finite union of unambiguous monomials of the form $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$.*
4. *L is the interior of some FO^2-definable language.*

Proof: "1 ⇒ 2": Theorem 4.2.

"2 ⇒ 3": Let $\alpha \in L \in \mathrm{FO}^2 \cap \Sigma_2$. By Theorem 5.5 we choose an unambiguous monomial $P = A_1^* a_1 \cdots A_k^* a_k$ (from a given finite set depending on L) and $A \subseteq \Gamma$ such that $PA^\infty \cap A^{\mathrm{im}}$ is unambiguous and $\alpha \in PA^\infty \cap A^{\mathrm{im}} \subseteq L$. W.l.o.g. $A \neq \emptyset$. Let $A = \{b_1, \ldots, b_m\}$ and $B_i = A \setminus \{b_i\}$ and $R = B_1^* b_1 \cdots B_m^* b_m$. Let L be strongly recognized by $h : \Gamma^* \to M$. By Ramsey's Theorem there exists $r \in \mathbb{N}$ such that for every sequence $v_1 \cdots v_r$ with $v_i \in M$ there are $1 \leq j \leq \ell \leq r$ with $h(v_j \cdots v_\ell) = e = e^2$ in M. Trivially, we have $\alpha \in PR^r A^\infty$. The monomial $PR^r A^\infty$ is unambiguous and for some fixed language L we consider only finitely many of them. We claim that $PR^r A^\infty \subseteq L$. Let $\beta \in PR^r A^\infty$ and write $\beta = uv_1 \cdots v_r \gamma$ with $u \in P$, $v_i \in R$, and $\gamma \in A^\infty$. Choose $v_j \cdots v_\ell = v$ such that $h(v)$ is idempotent. Then $uv_1 \cdots v_\ell v^\omega \in PA^\infty \cap A^{\mathrm{im}} \subseteq L$. Since L is open and $\mathrm{alph}(v) = A$ we have $uv_1 \cdots v_\ell v^s A^\infty \subseteq L$ for some $s \in \mathbb{N}$. By strong recognition and by idempotency of $h(v)$ we see that $\beta \in uv_1 \cdots v_\ell A^\infty \subseteq L$. Therefore, $PR^r A^\infty \subseteq L$.

"3 ⇒ 1": Theorem 4.2 and Theorem 5.5.

"1 ⇔ 4": This is the dual statement of Theorem 5.8. The proof of this theorem in turn uses "2 ⇒ 1", but this has just been shown. □

5.2 The fragment $\mathbf{FO^2} \cap \boldsymbol{\Pi_2}$

We have the following characterization of the class $\mathrm{FO}^2 \cap \Pi_2$ which also yields the missing part "1 ⇔ 4" in Theorem 5.7.

Theorem 5.8 *Let $L \subseteq \Gamma^\infty$ be a regular language. The following assertions are equivalent:*

1. *L is both FO^2-definable and Π_2-definable.*
2. *L is FO^2-definable and closed in the alphabetic topology.*
3. *L is the closure of some FO^2-definable language.*

Proof: "1 ⇒ 2": This is the dual statement of "1 ⇒ 2" in Theorem 5.5.

"2 ⇒ 3" is trivial.

"3 ⇒ 1": By Theorem 5.5 we may assume that L is the closure of $P \cap B^{\mathrm{im}}$ where $P = A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ is an unambiguous monomial and $B = A_{k+1}$. By Lemma 3.7 we obtain

$$L = \bigcup_{\{a_i, \ldots, a_k\} \cup B \subseteq A \subseteq A_i} A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty \cap A^{\mathrm{im}}.$$

Every monomial $A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty$ is unambiguous, hence L and its complement are FO^2-definable. The complement of L is open. Thus, the complement is Σ_2-definable by Theorem 5.7, "2 ⇒ 1", and therefore L is Π_2-definable. □

We also have a characterization when certain unambiguous monomials are closed:

Proposition 5.9 *Let $A_1^* a_1 \cdots A_k^* a_k A^\infty$ be unambiguous with $A_i \subseteq \{a_i, \ldots, a_k\}$ for all $1 \leq i \leq k$ and let $P = A_1^* a_1 \cdots A_k^* a_k A^\infty \cap B^{\mathrm{im}}$ for some $B \subseteq A$. The following assertions are equivalent:*

1. *There is no $1 \leq i \leq k$ such that $B \subseteq \{a_i, \ldots, a_k\} \subseteq A_i$.*
2. *The unambiguous monomial $P = A_1^* a_1 \cdots A_k^* a_k A^\infty \cap B^{\mathrm{im}}$ is closed in the alphabetic topology.*

Proof: "1 ⇒ 2": Assume by contradiction that P is not closed. Let $\alpha \notin P$ with $\mathrm{im}(\alpha) = C$ such that α is in the closure of P. Then, by Lemma 3.7, there is some $1 \leq i \leq k$ such that $\{a_i, \ldots, a_k\} \cup B \subseteq C \subseteq A_i$. Thus, $\{a_i, \ldots, a_k\} = C = A_i$ since by hypotheses $A_i \subseteq \{a_i, \ldots, a_k\}$. Since α is in the closure of P we have $B \subseteq C = \{a_i, \ldots, a_k\} = A_i$. This is a contradiction to "1".

"2 ⇒ 1": Assume by contradiction that $B \subseteq \{a_i, \ldots, a_k\} \subseteq A_i$ for some $1 \leq i \leq k$. We have $a_1 \cdots a_{i-1}(a_i \cdots a_k)^m B^\infty \cap B^{\mathrm{im},} \subseteq P$ for all $m \geq 1$ because $B \subseteq A$. As P is closed and $B \subseteq \{a_i, \ldots, a_k\}$ we see $a_1 \cdots a_{i-1}(a_i \cdots a_k)^\omega \in P$ and hence $\{a_i, \ldots, a_k\} \subseteq A$. But this is a contradiction to the fact that P is unambiguous since $\{a_i, \ldots, a_k\} \subseteq A_i \cap A$ implies that $a_1 \cdots a_{i-1}(a_i \cdots a_k)^2$ has two different factorizations. □

Theorem 5.8 is not fully satisfactory since we do not have any direct characterization in terms of polynomials. We might wish that if L is closed (and $L \in \mathrm{FO}^2 \cap \Pi_2$), then it is a finite union of languages $K \cap B^{\mathrm{im}}$ where each $K \cap B^{\mathrm{im}}$ is closed. But this is not true: Let $L = \Gamma^* a \cup \Gamma^\omega$, then L is closed and in $\mathrm{FO}^2 \cap \Pi_2$, but cannot be written in this form because $L = \Gamma^* a$ is not closed. We also note that the closure of a language L in $\mathrm{FO}^2 \cap \Sigma_2$ needs not to be in Δ_2. A counter-example is the language $L = \Gamma^* abc$. By Lemma 3.7, the closure of L is $\overline{L} = L \cup \Gamma^{\mathrm{im}}$ which is not Σ_2-definable.

5.3 The relation between $\mathbf{FO^2}$ and $\mathbf{\Sigma_2 \cap \Pi_2}$

For finite words we have the well-known theorem that FO^2-definability is equivalent to Δ_2-definability. However, this does not transfer to ω-words where Δ_2 forms a proper subclass of FO^2. Consider $L = \{a, b\}^{\mathrm{im}}$, then L is neither open nor closed, in general. Hence $L \in \mathrm{FO}^2 \setminus (\Sigma_2 \cup \Pi_2)$. The result for finite words is therefore somewhat misleading. The correct translation for the general case is:

Theorem 5.10 *For all $A \subseteq \Gamma$ the following assertions are equivalent:*

1. *$L \cap A^{\mathrm{im}}$ is FO^2-definable.*
2. *There are languages $L_\sigma \in \mathrm{FO}^2 \cap \Sigma_2$ and $L_\pi \in \mathrm{FO}^2 \cap \Pi_2$ such that*
$$L \cap A^{\mathrm{im}} = L_\sigma \cap A^{\mathrm{im}} = L_\pi \cap A^{\mathrm{im}}.$$
3. *There are languages $L_\sigma \in \Sigma_2$ and $L_\pi \in \Pi_2$ such that*
$$L \cap A^{\mathrm{im}} = L_\sigma \cap A^{\mathrm{im}} = L_\pi \cap A^{\mathrm{im}}.$$

Proof: "1 ⇒ 2": By Theorem 5.5 we see that $L \cap A^{\mathrm{im}}$ is a finite union of unambiguous monomials $A_1^* a_1 \cdots A_k^* a_k A^\infty \cap A^{\mathrm{im}}$. We let L_σ be the finite union of the monomials $A_1^* a_1 \cdots A_k^* a_k A^\infty$; by Theorem 5.7 we obtain $L_\sigma \in \mathrm{FO}^2 \cap \Sigma_2$. Let K be the complement of $L \cap A^{\mathrm{im}}$. Then K and $K \cap A^{\mathrm{im}}$ are FO^2-definable. Thus, $K \cap A^{\mathrm{im}} = K_\sigma \cap A^{\mathrm{im}}$ for some $K_\sigma \in \mathrm{FO}^2 \cap \Sigma_2$. Let L_π be the complement of K_σ. Then $L_\pi \in \mathrm{FO}^2 \cap \Pi_2$ and $L \cap A^{\mathrm{im}} = L_\pi \cap A^{\mathrm{im}}$. "2 ⇒ 3": Trivial. "3 ⇒ 1": If $L = L_\sigma \cap A^{\mathrm{im}}$, then a slight modification of the proof for Lemma 4.1 shows that all idempotents in $\mathrm{Synt}(L)$ are locally top. Identically, if $L = L_\pi \cap A^{\mathrm{im}}$, then all idempotents in $\mathrm{Synt}(L)$ are locally bottom. Thus $\mathrm{Synt}(L) \in \mathbf{DA}$, and by Theorem 5.5 we see that L is FO^2-definable. □

6 The fragment $\mathbf{\Delta_2 = \Sigma_2 \cap \Pi_2}$

6.1 Clopen unambiguous monomials

Languages in Σ_2 are open and languages in Π_2 are closed. Hence, a language in Δ_2 must be clopen in the alphabetic topology. The first step towards a convenient characterization of Δ_2 is therefore a description of clopen unambiguous monomials.

Lemma 6.1 *Let $P = A_1^* a_1 \cdots A_k^* a_k A^\infty$ be an unambiguous monomial. The following assertions are equivalent:*

1. *There is no $1 \le i \le k$ such that $\{a_i, \ldots, a_k\} \subseteq A_i$.*
2. *P is closed in the alphabetic topology.*
3. *P is clopen in the alphabetic topology.*

Proof: "$1 \Rightarrow 2$": For a moment let $A_{k+1} = A$. By Lemma 3.7 we see that the closure of P is:

$$\bigcup_{\{a_i,\ldots,a_k\} \subseteq B \subseteq A_i} A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty \cap B^{\mathrm{im}}.$$

Since there is no $\{a_i, \ldots, a_k\} \subseteq A_i$ for $1 \le i \le k$, we see that this union is just P itself. Therefore, P is closed. "$2 \Rightarrow 3$": is clear, because P is open. "$3 \Rightarrow 1$": Assume by contradiction that $\{a_i, \ldots, a_k\} \subseteq A_i$ for some $1 \le i \le k$. We have $a_1 \cdots a_{i-1}(a_i \cdots a_k)^m \in P$ for all $m \ge 1$. As P is closed we see $a_1 \cdots a_{i-1}(a_i \cdots a_k)^\omega \in P$ and hence $\{a_i, \ldots, a_k\} \subseteq A$. But this is a contradiction to the fact that P is unambiguous since $\{a_i, \ldots, a_k\} \subseteq A_i \cap A$ implies that $a_1 \cdots a_{i-1}(a_i \cdots a_k)^2 \in P$ has two different factorizations. □

Lemma 6.2 *Let $L \subseteq \Gamma^\infty$ be a closed polynomial. For every unambiguous monomial*

$$P = A_1^* a_1 \cdots A_k^* a_k A^\infty \subseteq L$$

there exist closed unambiguous monomials $Q_1, \ldots, Q_\ell$ such that $P \subseteq Q_1 \cup \cdots \cup Q_\ell \subseteq L$, i.e., there exists a finite covering of P with closed unambiguous monomials in L.

Proof: We start with a normalization procedure in which we begin with making the last appearances of the letters in A_i^* explicit. We have $B^* = (B \setminus \{b\})^* \cup B^* b \, (B \setminus \{b\})^*$ for every $b \in B$. This yields the substitution rule of replacing A_i^* in P by $(A_i \setminus \{a\})^*$ and also by $A_i^* a (A_i \setminus \{a\})^*$ which gives two new monomials. After iterating this substitution rule a finite number of times, we obtain unambiguous monomials of the form $P_i' = B_1^* b_1 \cdots B_s^* b_s A^\infty$ such that $P = \bigcup P_i'$ and $B_i \subseteq \{b_i, \ldots, b_s\}$ for every $1 \le i \le s$. In the next phase of the normalization procedure we make the first appearances of the letters in A^∞ explicit. We have $B^\infty = (B \setminus \{b\})^\infty \cup (B \setminus \{b\})^* \, b B^\infty$ for every $b \in B$. As above, this yields a substitution rule and after a finite number of applications to the P_i' we obtain unambiguous monomials of the form $P_i'' = B_1^* b_1 \cdots B_s^* b_s B_{s+1}^* b_{s+1} \cdots B_t^* b_t A^\infty$ such that $P = \bigcup P_i''$ and the following properties hold:

- $B_i \subseteq \{b_i, \ldots, b_t\}$ for every $1 \le i \le s$.
- $\{b_i, \ldots, b_t\} \not\subseteq B_i$ for all $s+1 \le i \le t$.
- $A = \{b_{s+1}, \ldots, b_t\}$.

It suffices to prove the lemma for $P = B_1^* b_1 \cdots B_s^* b_s B_{s+1}^* b_{s+1} \cdots B_t^* b_t A^\infty$ with the above properties. If P is not closed, then by Lemma 6.1 there exists $1 \le i \le s$ such that $B_i \supseteq \{b_i, \ldots, b_t\}$, and hence $A \subseteq B_i = \{b_i, \ldots, b_t\}$ due to the normalization procedure. We fix the minimal index i with this property.

Next, we use a Ramsey argument. Let L be strongly recognized by $h : \Gamma^* \to M$ and let $r = r(M)$ be the Ramsey number such that every complete edge-colored graph with r nodes and using at most $|M|$ colors contains a monochromatic triangle. We have $B_i^* = (B_i \setminus \{b_j\})^* \cup (B_i \setminus \{b_j\})^* b_j B_i^*$ and $B_i \setminus \{b_j\}$ is no longer a superset of $\{b_i, \ldots, b_t\}$. Therefore, we only have to consider the case where we replace the factor $b_{i-1} B_i^* b_i$ in P by $b_{i-1}(B_i \setminus \{b_j\})^* b_j B_i^* b_i$ for some $i \leq j \leq t$. Repeating this procedure we are left with a situation where we have replaced $b_{i-1} B_i^* b_i$ in P by $b_{i-1} R^r B_i^* b_i$ in P where

$$R = (B_i \setminus \{b_i\})^* b_i (B_i \setminus \{b_{i+1}\})^* b_{i+1} \cdots (B_i \setminus \{b_t\})^* b_t.$$

Note that the resulting monomial $\widetilde{P}$ is unambiguous and that the alphabet of every word in R is $B_i = \{b_i, \ldots, b_t\}$.

Now consider $\alpha = uv_1 \cdots v_r \in B_1^* b_1 \cdots B_{i-1}^* b_{i-1} R^r$, with $v_j \in R$ for all $1 \leq j \leq r$. By the choice of r being the Ramsey number for triangles we find some $j_1 \leq j_2 < j_3$ such that $h(v_{j_1} \cdots v_{j_2}) = h(v_{j_2+1} \cdots v_{j_3}) = h(v_{j_1} \cdots v_{j_3})$ is idempotent in the monoid M. Since L is closed we see that

$$uv_1 \cdots v_{j_1-1}(v_{j_1} \cdots v_{j_2})^\omega \in L.$$

This is clear because for each prefix $w_m = uv_1 \cdots v_{j_1-1}(v_{j_1} \cdots v_{j_2})^m$ we have $\mathrm{alph}(v_{j_1}) = \{b_i, \ldots, b_t\} = B_i$ and $w_m b_i \cdots b_t \in P \subseteq L$.

Since L is open, there is some m such that $w_m B_i^\infty \subseteq L$. This follows again because $\mathrm{alph}(v_{j_1}) = B_i$. Since h strongly recognizes L and since $h(w_m) = h(uv_1 \cdots v_{j_2})$ by idempotency of $h(v_{j_1} \cdots v_{j_2})$, we have $uv_1 \cdots v_{j_2} B_i^\infty \subseteq L$. In particular, $uv_1 \cdots v_r B_i^\infty \subseteq L$.

This is true for all $\alpha \in B_1^* b_1 \cdots B_{i-1}^* b_{i-1} R^r$, hence

$$B_1^* b_1 \cdots B_{i-1}^* b_{i-1} R^r B_i^\infty \subseteq L.$$

By construction, $Q = B_1^* b_1 \cdots B_{i-1}^* b_{i-1} R^r B_i^\infty$ is a closed unambiguous monomial and due to the normalization, we have $B_i^* b_i \cdots B_t^* b_t A^\infty \subseteq B_i^\infty$ and hence $P \subseteq Q$. □

6.2 Arrow languages and deterministic languages

We write $s \,\mathcal{R}\, t$ for monoid elements $s, t \in M$ if there exist $x, y \in M$ such that $s = ty$ and $t = sx$, i.e., if the *right-ideals* sM and tM are equal. The relation $\mathcal{R}$ is one of Green's relations, see e.g. [16].

The results of this section are very similar to results on deterministic and complement-deterministic languages which can be found in [15], too. Moreover, the conditions in Proposition 6.4 and Proposition 6.5 can be complemented by several other equivalent characterizations, see e.g. [15, Theorems VI.3.7]. One of them is the class of finite Boolean combinations of regular Cantor-open languages and another one is in terms of the second level of the Borel hierarchy over the Cantor topology.

Lemma 6.3 *Let $L \subseteq \Gamma^\infty$ be a deterministic language which is strongly recognized by some surjective homomorphism $h : \Gamma^* \to M$ onto a finite monoid M. Let $s, e, t, f, x, y, \in M$ such that $(s, e), (t, f)$ are linked pairs and $s = ty$ and $t = sx$ (thus, $s \,\mathcal{R}\, t$). Assume that*

$$[s][e]^\omega \cap L \cap \Gamma^\omega \neq \emptyset.$$

Then we have $[t][yexf]^\omega \subseteq L$.

Proof: Let $s_0, e_0, f_0, x_0, y_0 \in \Gamma^*$ be words which are mapped to the corresponding elements in $s, e, f, x, y \in M$. We choose $e_0 \neq 1$ nonempty, which we can do due to the assumption. Since L is deterministic, there exists a set $W \subseteq \Gamma^*$ such that $L \cap \Gamma^\omega = \overrightarrow{W} \cap \Gamma^\omega$. We are going to construct sequences of words $s_n \in [s]\big([xf][ye]\big)^n$ and $w_n \in W$ for $n \in \mathbb{N}$ such that

$$s_0 < w_0 < s_1 < w_1 < s_2 < w_2 < \cdots$$

where $<$ denotes the strict prefix order on words. Thus, the limit defines an infinite word α such that $\alpha \in [s]\big([xf][ye]\big)^{\omega} \cap \overrightarrow{W}$. In particular, $\alpha \in L$. Moreover, since $sxf = t$ we have $\alpha \in [t][yexf]^{\omega} \cap L$ and hence $[t][yexf]^{\omega} \subseteq L$ due to strong recognition.

Thus, it is enough to define the sequences s_n and w_n for $n \in \mathbb{N}$ as above. The condition $s_0 \in [s]\big([xf][ye]\big)^0$ is satisfied by definition. Let $n \in \mathbb{N}$. Inductively, we may assume that w_k and s_m are defined as desired for $k < n$ and $m \leq n$. We are going to define w_n and s_{n+1}. Infinitely many prefixes of $s_n x_0 f_0 y_0 e_0^{\omega}$ are in W, because $s_n x_0 f_0 y_0 e_0^{\omega} \in [s][e]^{\omega} \subseteq L$. Thus we find $w_n \in W$ and $\ell \geq 1$ such that

$$s_n < w_n < s_{n+1} = s_n x_0 f_0 y_0 e_0^{\ell}.$$

By induction we see that $s_{n+1} \in [s]\big([xf][ye]\big)^{n+1}$ because $x_0 f_0 y_0 e_0^{\ell} \in [xf][ye]$ since $e^2 = e$. □

Proposition 6.4 *Let $L \subseteq \Gamma^{\infty}$ be strongly recognized by some surjective homomorphism $h : \Gamma^* \to M$ onto a finite monoid M. Define*

$$W = \bigcup \{[s] \subseteq \Gamma^* \mid [s][e]^{\omega} \subseteq L \text{ for some linked pair } (s, e)\}.$$

Then the following four assertions are equivalent:

1. *$L = \overrightarrow{W}$.*
2. *For all linked pairs (s, e), (t, f) with $s \mathrel{\mathcal{R}} t$ we have*
$$[s][e]^{\omega} \subseteq L \iff [t][f]^{\omega} \subseteq L.$$
3. *For every linked pair (s, e) we have*
$$[s][e]^{\omega} \subseteq L \iff [s] \subseteq L.$$
4. *Both L and its complement are arrow languages.*

Proof: "1 ⇒ 2": Let $[s] \subseteq W$ and let (t, f) be a linked pair with $s \mathrel{\mathcal{R}} t$. It is enough to show $[t][f]^{\omega} \subseteq L$. If $s = t$, then $[t][f]^{\omega} \subseteq \overrightarrow{W} = L$. For $s \neq t$ we find $x \neq 1 \neq y$ with $s = ty$ and $t = sx$. It follows that $[s][xy]^{\omega} \cap L \cap \Gamma^{\omega} \neq \emptyset$. Lemma 6.3 yields $[t][yexf]^{\omega} \subseteq L$ for $e = xy$. But then $[t] \subseteq W$ and $[t][f]^{\omega} \subseteq \overrightarrow{W} = L$.

"2 ⇒ 3": If $[s][e]^{\omega} \subseteq L$ then by "2" we have $[s][1]^{\omega} \subseteq L$. Since $[s] \subseteq [s][1]^{\omega}$, it follows $[s] \subseteq L$. Conversely, if $[s] \subseteq L$, then strong recognition yields $[s][1]^{\omega} \subseteq L$; and hence $[s][e]^{\omega} \subseteq L$ by "2".

"3 ⇒ 4": The condition is symmetric in L and its complement. Therefore it is enough to show that L is an arrow language. We show $L = \overrightarrow{L \cap \Gamma^*}$. Let $[s][e]^{\omega} \subseteq L$. Then, by "3", we see that $[s] \subseteq L$ and hence $[s][e]^{\omega} \subseteq \overrightarrow{[s]} \subseteq \overrightarrow{L \cap \Gamma^*}$. For the other inclusion, let $\alpha \in \overrightarrow{L \cap \Gamma^*}$. Then $\alpha \in \overrightarrow{[s]}$ for some $s \in M$ with $[s] \cap L \neq \emptyset$. We can find a linked pair (s, e) such that $\alpha \in [s][e]^{\omega}$. By strong recognition, $[s] \subseteq [s][1]^{\omega} \subseteq L$. By "3" we conclude $[s][e]^{\omega} \subseteq L$ and $\alpha \in L$.

"4 ⇒ 1": Since L is an arrow language, it is enough to show $L \cap \Gamma^* = W$. The inclusion $L \cap \Gamma^* \subseteq W$ is trivial. For the converse assume by contradiction $[s] \cap L = \emptyset$, but $[s][e]^{\omega} \subseteq L$ for some linked pair (s, e). Then $[s] \subseteq \Gamma^* \setminus L$. Since the complement of L is an arrow language, we have $[s][e]^{\omega} \subseteq \overrightarrow{[s]} \subseteq \overrightarrow{\Gamma^* \setminus L} = \Gamma^{\infty} \setminus L$, which is a contradiction to $[s][e]^{\omega} \subseteq L$. Thus, $W \subseteq L \cap \Gamma^*$. □

The following result yields a simple proof for a *Landweber type* result in the special case of deterministic and complement-deterministic languages.

Proposition 6.5 *Let $L \subseteq \Gamma^{\omega}$ be a deterministic language which is strongly recognized by some surjective homomorphism $h : \Gamma^* \to M$ onto a finite monoid M. Let*

$$W = \bigcup \{[s] \subseteq \Gamma^* \mid [s][e]^{\omega} \subseteq L \text{ for some linked pair } (s, e)\}$$

and $U = \Gamma^* \setminus W$. Then $\overrightarrow{W} \cup \overrightarrow{U} = \Gamma^{\infty}$ and $\overrightarrow{W} \cap \overrightarrow{U} = \emptyset$, i.e., Γ^{∞} is a disjoint union of $\overrightarrow{W}$ and $\overrightarrow{U}$. Moreover, $\overrightarrow{W} \cap \Gamma^{\omega} = L$ if and only if L is complement-deterministic, too.

Proof: Clearly, $\overrightarrow{W} \cup \overrightarrow{U} = \Gamma^\infty$. Assume by contradiction that there is some $\alpha \in \overrightarrow{W} \cap \overrightarrow{U}$. Then $\alpha \in \Gamma^\omega$ with $\alpha \in \overrightarrow{[s]} \cap \overrightarrow{[t]}$ such that $[s] \subseteq W$ and $[t] \subseteq U$. Using the usual application of Ramsey's Theorem at those prefixes belonging to $[s]$ or $[t]$, respectively, we see that for some linked pairs (s, e), (t, f) we have $\alpha \in [s][e]^\omega$ and $\alpha \in [t][f]^\omega$. We have $s = ty$ and $t = sx$ with $x \neq 1 \neq y$ because $s \neq t$ as $U \cap W = \emptyset$. Since $[s][e]^\omega \cap L \cap \Gamma^\omega \neq \emptyset$, by Lemma 6.3 we have $[t][yexf]^\omega \subseteq L$. This contradicts $[t] \subseteq U = \Gamma^* \setminus W$.

For the second statement of the proposition: If $\overrightarrow{W} \cap \Gamma^\omega = L$, then by the first statement of this proposition we have $\Gamma^\omega \setminus L = \overrightarrow{U} \cap \Gamma^\omega$, i.e., L is complement-deterministic.

For the converse, let L be complement-deterministic. Clearly, $L \subseteq \overrightarrow{W} \cap \Gamma^\omega$. Assume by contradiction that there is some $\alpha \in \overrightarrow{W} \setminus L$ for some $\alpha \in \Gamma^\omega$. Then $\alpha \in \overrightarrow{[s]} \cap [t][f]^\omega$ for $[s] \subseteq W$ and (t, f) is a linked pair with $[t][f]^\omega \subseteq \Gamma^\omega \setminus L$. We have $s = ty$ and $t = sx$ for some $x, y \in M$. By definition of W, we find a linked pair (s, e) such that $[s][e]^\omega \subseteq L$. We have $[s][e]^\omega \cap L \cap \Gamma^\omega \neq \emptyset$ and $[t][f]^\omega \cap (\Gamma^\infty \setminus L) \cap \Gamma^\omega \neq \emptyset$. Since both L and $\Gamma^\infty \setminus L$ are deterministic, we can apply Lemma 6.3 and obtain $[t][yexf]^\omega \subseteq L$ and $[s][xfye]^\omega \subseteq \Gamma^\omega \setminus L$. This is a contradiction to strong recognizability, since $[t][yexf]^\omega \cap [s][xfye]^\omega \neq \emptyset$. □

6.3 Various characterizations of Δ_2

Theorem 6.6 *Let $L \subseteq \Gamma^\infty$ be a regular language. The following assertions are equivalent.*

1. *L is Δ_2-definable.*
2. *L is FO^2-definable and L is clopen in the alphabetic topology.*
3. *L is a finite union of unambiguous closed monomials $A_1^* a_1 \cdots A_k^* a_k A^\infty$, i.e., there is no $1 \leq i \leq k$ such that $\{a_i, \ldots, a_k\} \subseteq A_i$.*
4. *$\mathrm{Synt}(L) \in \mathbf{DA}$ and for all linked pairs (s, e), (t, f) with $s \mathrel{\mathcal{R}} t$ (i.e., there exist $x, y \in \mathrm{Synt}(L)$ such that $s = ty$ and $t = sx$) we have*
$$[s][e]^\omega \subseteq L \;\Leftrightarrow\; [t][f]^\omega \subseteq L.$$
5. *L is weakly recognized by $h : \Gamma^* \to M$ for some $M \in \mathbf{DA}$, and for all linked pairs (s, e), (t, f) with $s \mathrel{\mathcal{R}} t$ in M we have $[s][e]^\omega \subseteq L$ if and only if $[t][f]^\omega \subseteq L$.*
6. *$\mathrm{Synt}(L) \in \mathbf{DA}$ and both L and its complement $\Gamma^\infty \setminus L$ are arrow languages.*

Proof: "1 ⇒ 2": By Theorem 4.2 and its dual version for Π_2, we see that $\mathrm{Synt}(L) \in \mathbf{DA}$ and that L is clopen in the alphabetic topology. From Theorem 5.5 it follows that L is FO^2-definable.

"2 ⇒ 3": By Theorem 5.7, L is a finite union of unambiguous monomials. Property "3" now follows by Lemma 6.2 and Lemma 6.1.

"3 ⇒ 1": Theorem 5.7 and Theorem 5.8.

"2 ⇒ 4": By Theorem 5.5, we see that $\mathrm{Synt}(L) \in \mathbf{DA}$. Suppose $[s][e]^\omega \subseteq L$ and let $s = ty$ and $t = sx$. Since L is closed we see that $[s][exfy]^\omega \subseteq L$ and by strong recognition we conclude $[t][fyex]^\omega \subseteq L$. Let $A = \bigcup\{\mathrm{alph}(v) \mid v \in [f]\}$. Since L is open and by strong recognition, there exists $r \in \mathbb{N}$ such that $[t][fyex]^r A^\infty \subseteq L$. Moreover, $t = tfyex$ and thus, $[t]A^\infty \subseteq L$. In particular, $[t][f]^\omega \subseteq L$ because $[f] \subseteq A^*$.

"4 ⇒ 5": Trivial with $M = \mathrm{Synt}(L)$ and $h = h_L$.

"5 ⇒ 2": If $\alpha \in [s][e]^\omega \cap [t][f]^\omega$ for linked pairs $(s, e), (t, f)$, then $s \mathrel{\mathcal{R}} t$. Hence $[s][e]^\omega \subseteq L$ and $[s][e]^\omega \cap [t][f]^\omega \neq \emptyset$ implies $[t][f]^\omega \subseteq L$. In particular, h strongly recognizes L.

Definability in FO^2 follows by Theorem 5.5. By symmetry, it suffices to show that L is open. Let $\alpha \in [s][e]^\omega \subseteq L$ for some linked pair (s, e) and write $\alpha = u\beta$ with $u \in [s]$ and $\beta \in [e]^\omega \cap A^\infty \cap A^{\mathrm{im}}$ for some $A \subseteq \Gamma$. Let $v \leq \beta$ be a prefix such that $v \in [e]$ and $\mathrm{alph}(v) = \mathrm{alph}(\beta)$. We want to show $uvA^\infty \subseteq L$. Consider $uv\gamma \in \Gamma^\infty$ where $\gamma \in A^\infty$. We have $uv\gamma \in [t][f]^\omega$ for some linked pair (t, f). Let $v' \leq \gamma$ such that $uvv' \in [t]$. Since $\mathrm{Synt}(L) \in \mathbf{DA}$ we have $vv'v \in [e]$ and $s = t \cdot h(v)$. Together with $t = s \cdot h(v')$ it follows $s \mathrel{\mathcal{R}} t$ and by "4" we obtain $uv\gamma \in [t][f]^\omega \subseteq L$.

"4 ⇔ 6": This equivalence follows from Proposition 6.4. □

Corollary 6.7 *Let $L \subseteq \Gamma^\infty$ be a regular language such that* $\mathrm{Synt}(L) \in \mathbf{DA}$. *The following assertions are equivalent:*

1. *L is clopen in the alphabetic topology.*
2. *Both L and its complement $\Gamma^\infty \setminus L$ are arrow languages.*

Proof: The statement follows from the equivalence of "2" and "6" in Theorem 6.6 since by Theorem 5.5 the language L is FO^2-definable if and only if $\mathrm{Synt}(L) \in \mathbf{DA}$. □

Remark 6.8 *The statement of Corollary 6.7 does not need to hold outside the variety* $\mathbf{DA}$. *For example the aperiodic language $L = (ab)^\omega \cup (ab)^* a \subseteq \{a,b\}^\infty$ is an arrow language and its complement is also an arrow language, but it is not open.*

6.4 The intersection of Σ_2 and Π_2 over infinite words

The next corollary gives a characterization of the fragment Δ_2 for ω-languages, i.e., we consider the intersection of Σ_2 and Π_2 over infinite words (instead of finite and infinite words). Note that the language $\Gamma^\omega \subseteq \Gamma^\infty$ of all infinite words is Π_2-definable, but not Σ_2-definable as a subset of Γ^∞.

Corollary 6.9 *Let $L \subseteq \Gamma^\omega$ be an ω-regular language. The following assertions are equivalent:*

1. *$L \in \Pi_2$ and there exists a language $L_\sigma \in \Sigma_2$ such that $L = L_\sigma \cap \Gamma^\omega$.*
2. *There exist languages $L_\sigma \in \Sigma_2$ and $L_\pi \in \Pi_2$ such that $L = L_\sigma \cap \Gamma^\omega = L_\pi \cap \Gamma^\omega$.*
3. $\mathrm{Synt}(L) \in \mathbf{DA}$ *and L is deterministic and complement-deterministic.*
4. *There exists a language $L_\delta \in \Delta_2$ such that $L = L_\delta \cap \Gamma^\omega$.*

Proof: "1 $\Leftrightarrow$ 2": Trivial, since $L = L_\pi \cap \Gamma^\omega$ is Π_2-definable.

"2 $\Rightarrow$ 3": By Theorem 5.10 we see that L is FO^2-definable and by Theorem 5.5 we conclude $\mathrm{Synt}(L) \in \mathbf{DA}$. The complement of L_π is Σ_2-definable, hence L_π is closed by Theorem 4.2. Therefore, $L = L_\pi \cap \Gamma^\omega$ is closed too. By Corollary 3.3 it follows that L is deterministic. Symmetrically, we deduce that $\Gamma^\omega \setminus L$ is also deterministic.

"3 $\Rightarrow$ 4": Let $W = \bigcup \{[s] \subseteq \Gamma^* \mid [s][e]^\omega \subseteq L \text{ for some linked pair } (s,e)\}$ and set $L_\delta = \overrightarrow{W}$. By Proposition 6.5 we have $L = L_\delta \cap \Gamma^\omega$. Moreover, both L_δ and its complement are arrow languages. Since $\mathrm{Synt}(L_\delta) = \mathrm{Synt}(L)$ we can apply Theorem 6.6 and conclude $L_\delta \in \Delta_2$.

"4 $\Rightarrow$ 2": Trivial with $L_\sigma = L_\pi = L_\delta$. □

6.5 On the construction of examples

Let $h : \Gamma^* \to M$ be a surjective homomorphism onto a finite monoid M. By definition of weak recognition, for every linked pair (s,e) the language $[s][e]^\omega$ is weakly recognized by h and every language which is weakly recognized by h is a union of such languages. We say that two linked pairs (s,e), (t,f) are *conjugated*, if $e = xy$, $f = yx$, and $t = sx$ for some $x, y \in M$. It is easy to verify that conjugacy forms an equivalence relation on the set of linked pairs and that $[s][e]^\omega \cap [t][f]^\omega \neq \emptyset$ if and only if the linked pairs (s,e) and (t,f) are conjugated. We define for a linked pair (s,e) the class $[s,e]$ as a language by:

$$[s,e] \;=\; \bigcup \{[t][f]^\omega \mid (s,e) \text{ and } (t,f) \text{ are conjugated}\} \;\subseteq\; \Gamma^\infty.$$

The language $[s,e]$ is strongly recognized by h; and every language, which is strongly recognized by h, is a union of such languages.

The set $\overrightarrow{[s]}$ is an arrow language which is weakly recognized by h since

$$\overrightarrow{[s]} \;=\; \bigcup \{[s][f]^\omega \mid (s,f) \text{ is a linked pair for some } f \in M\}$$

If an arrow language $L \subseteq \Gamma^\infty$ is weakly recognized by h, then L is a union of languages of the form $\overrightarrow{[s]}$ since $L = \overrightarrow{L \cap \Gamma^*}$ and $L \cap \Gamma^* = \bigcup \{[s] \mid [s] \cap L \neq \emptyset\}$. In general, $\overrightarrow{[s]}$ is not strongly recognized by M.

For every $s \in M$ we denote by $\mathcal{R}_s$ the $\mathcal{R}$-class of s, i.e., $\mathcal{R}_s = \{t \in M \mid sM = tM\}$. We have

$$\overrightarrow{[\mathcal{R}_s]} = \bigcup \{[s,e] \mid \text{there exists } e \in M \text{ such that } (s,e) \text{ is a linked pair}\}\,.$$

By Proposition 6.4, both $\overrightarrow{[\mathcal{R}_s]}$ and its complement are arrow languages which are strongly recognized by h. Conversely, if L and its complement are arrow languages which are strongly recognized by h, then L is a union of languages of the form $\overrightarrow{[\mathcal{R}_s]}$. Moreover, as shown in the proof of Theorem 6.6, if L and its complement are arrow languages and if L is weakly recognized by h, then, in fact, L is strongly recognized by h.

Therefore, for some given $h : \Gamma^* \to M$ we find examples as follows:

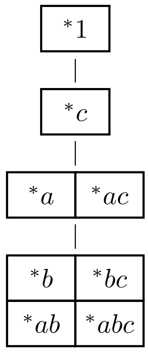
- $[s][e]^\omega$ are languages which are weakly recognizable by h.
- $[s,e]$ are languages which are strongly recognizable by h.
- $\overrightarrow{[s]}$ are arrow languages which are weakly recognizable by h.
- $\overrightarrow{[\mathcal{R}_s]}$ are arrow languages whose complement is also an arrow language, and which are strongly recognizable by h.

More concretely: If $M \in \mathbf{DA}$ then, by Theorem 5.5, the languages which are strongly recognizable by h are FO^2-definable, but by Example 5.6 weak recognition is not enough to guarantee FO^2-definability. By Theorem 6.6, languages L are Δ_2-definable if they are strongly recognizable by h and if both, L and $\Gamma^\infty \setminus L$ are arrow languages.

Therefore we can produce examples along the following line: We start with some linked pair (s,e), this yields $[s][e]^\omega$ which is weakly recognizable and $[s,e]$ which is strongly recognized by h. The arrow language $\overrightarrow{[s]}$ is incomparable with $[s,e]$, in general. By definition, $\overrightarrow{[s]}$ is a deterministic language. Moving to $\overrightarrow{[\mathcal{R}_s]}$ yields an arrow language, where its complement is an arrow language, too. We have:

$$[s,e] \subseteq [s][e]^\omega \subseteq \overrightarrow{[s]} \qquad\qquad [s,e] \subseteq \overrightarrow{[\mathcal{R}_s]} \subseteq \overrightarrow{[s]}$$

Example 6.10 Let $\Gamma = \{a,b,c\}$ and $P = c^*a\Gamma^*b\Gamma^*c$. The syntactic monoid of P is in $\mathbf{DA}$, because P is FO^2-definable. We can write $\mathrm{Synt}(P) = \{1,a,b,ab,c,ac,bc,abc\}$ where the elements correspond to minimal length representatives of the classes induced by the syntactic congruence. To see this observe that $P = c^*a\Gamma^*b\Gamma^* \cap \Gamma^*c$. The syntactic monoid of $c^*a\Gamma^*b\Gamma^*$ has just the four elements in $\{1,a,b,ab\}$. For $\mathrm{Synt}(P)$ we copy these classes and add the information whether it represents a word ending in c. All elements of $\mathrm{Synt}(P)$ are idempotent and its egg-box representation (see e.g. [16]) is given by:



We have $P = [abc]$. The language $L = P^\omega = [abc]^\omega$ is weakly recognizable by $\mathrm{Synt}(P)$, too. All words in $\alpha \in L$ have infinitely many occurrences of the factor ca and $\mathrm{im}(\alpha) = \Gamma$. In particular, L is not open in the strict alphabetic topology. By Lemma 5.2, the language L is not strongly recognizable by any monoid in **DA**.

The conjugacy class of the linked pair (abc, abc) is $\{(ab, b), (ab, ab), (abc, bc), (abc, abc)\}$ and $[abc, abc] = c^*a\Gamma^*b\Gamma^\infty \cap (\Gamma^*b)^\omega$. The language $[abc, abc]$ is strongly recognizable by $\mathrm{Synt}(P) \in \mathbf{DA}$. By Theorem 5.5 it is FO^2-definable. The set $[abc, abc]$ is not open in the alphabetic topology. By Theorem 6.6, $[abc, abc]$ is not Δ_2-definable.

The set $\overrightarrow{[abc]} = \overrightarrow{P} = c^*a\Gamma^*b\Gamma^\infty \cap (\Gamma^*c)^\omega$ is an arrow language which is weakly recognizable by h. It is not strongly recognized by the syntactic homomorphism of P since $[abc][abc]^\omega \subseteq \overrightarrow{[abc]} \cap [abc, abc]$ but $[abc, abc] \not\subseteq \overrightarrow{[abc]}$. On the other hand, $\overrightarrow{[abc]}$ is FO^2-definable, and therefore, by Theorem 5.5, it is strongly recognizable by some other homomorphism onto a monoid in **DA**.

The $\mathcal{R}$-class of abc is $\mathcal{R}_{abc} = \{ab, abc\}$. Hence $\overrightarrow{[\mathcal{R}_{abc}]} = c^*a\Gamma^*b\Gamma^\infty$. By Proposition 6.4 the complement of $\overrightarrow{[\mathcal{R}_{abc}]}$ is also an arrow language; and by Theorem 6.6 the language $\overrightarrow{[\mathcal{R}_{abc}]}$ is Δ_2-definable. Indeed, for $\overrightarrow{[\mathcal{R}_{abc}]}$ it is enough to say that there is some b and there is some a with no b to its left. This is a Σ_2-sentence. The equivalent Π_2-sentence says that there is some b and for all b there exists some a to its left. It is also deterministic and complement-deterministic. ◇

7 Summary

We have given language-theoretic, algebraic and topological characterizations for several first-order fragments over infinite words. Since FO^2 and Δ_2 have the same expressive power only when restricted to some fixed set of letters occurring infinitely often (Thm. 5.10), the picture becomes more complex than in the case of finite words. By Pol we denote the language class of polynomials, UPol are unambiguous polynomials, and *restricted* UPol is a proper subclass of UPol. *Simple polynomials* are finite unions of languages of the form $\Gamma^*a_1 \cdots \Gamma^*a_n\Gamma^\infty$. A language $L \subseteq \Gamma^\infty$ is *piecewise testable* if there exists some $k \in \mathbb{N}$ such that for every $\alpha \in \Gamma^\infty$ membership in L only depends on the set of scattered subwords of α length $\leq k$. The first-order fragment Σ_1 consist of first-order sentences in prenex normal without universal quantifiers. Its Boolean closure is $\mathbb{B}\Sigma_1$.

All of the below-mentioned algebraic properties are decidable, since the syntactic monoid of a regular language is effectively computable [15, 24]. Together with the PSPACE-completeness of the problem whether a language is closed in the alphabetic topology (Thm. 3.5), this yields decidability of the membership problem for the respective first-order fragments as a corollary. Decidability was shown before by Wilke [28] for FO^2 and by Bojańczyk [2] for Σ_2. The characterization for the fragment Σ_1 is due to Pin [17]; see also [15]. The same holds for the Boolean closure of Σ_1 except for the topological part of the "Algebra + Topology" characterization, which is a consequence of Corollary 6.7.

Logic	Languages	Algebra		Topology	
Σ_2	Pol	$eM_e e \leq e$	+	open alphabetic	Thm. 4.2
FO^2	UPol + A^{im}	**DA**			Thm. 5.5
		weak **DA**	+	closed strict alphabetic	
$\mathrm{FO}^2 \cap \Sigma_2$	UPol	**DA**	+	open alphabetic	Thm. 5.7
$\mathrm{FO}^2 \cap \Pi_2$		**DA**	+	closed alphabetic	Thm. 5.8
Δ_2	restricted UPol	**DA**	+	clopen alphabetic	Thm. 6.6
Σ_1	simple Pol	$x \leq 1$	+	open Cantor	[15]
$\mathbb{B}\Sigma_1$	piecewise testable	$\mathcal{J}$-trivial	+	clopen alphabetic	Cor. 6.7 and [15]

8 Outlook and open problems

By definition, Σ_1-definable languages are open in the Cantor topology. We introduced an alphabetic topology such that Σ_2-definable languages are open in this topology. Therefore, an interesting question is whether it is possible to extend this topological approach to higher levels of the first-order alternation hierarchy. To date, even over finite words no decidable characterization of the Boolean closure of Σ_2 is known. In case that a decidable criterion is found, it might lead to a decidable criterion for infinite words simply by adding a condition of the form "L and its complement are in the second level of the Borel hierarchy of the alphabetic topology". Another possible way to generalize our approach might be combinations of algebraic and topological characterizations for fragments with successor predicate suc such as $\mathrm{FO}^2[<, \mathrm{suc}]$ or $\Sigma_2[<, \mathrm{suc}]$. A characterization of those languages which are weakly recognizable by monoids in **DA** is also open.

Acknowledgements

We thank the anonymous referees of our STACS 2009 paper for many useful suggestions and we thank Luc Dartois from the ENS Cachan for helpful discussions during his internship in Stuttgart in 2008 and for his firm conviction that FO^2 should coincide with **DA**.

References

[1] M. Bojańczyk. Two-way unary temporal logic over trees. In *LICS'07*, pages 121–130. IEEE Computer Society, 2007.

[2] M. Bojańczyk. The common fragment of ACTL and LTL. In *Foundations of Software Science and Computational Structures, 11th International Conference, FoSSaCS 2008, Proceedings*, volume 4962 of *Lecture Notes in Computer Science*, pages 172–185. Springer-Verlag, 2008.

[3] M. Bojańczyk and L. Segoufin. Tree languages defined in first-order logic with one quantifier alternation. In *ICALP'08*, volume 5126 of *Lecture Notes in Computer Science*, pages 233–245. Springer-Verlag, 2008.

[4] M. Bojańczyk, L. Segoufin, and H. Straubin. Piecewise testable tree languages. In *LICS'08*, pages 442–451. IEEE Computer Society, 2008.

[5] J. A. Brzozowski and F. E. Fich. On generalized locally testable languages. *Discrete Mathematics*, 50:153–169, 1984.

[6] L. E. Dickson. Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors. *American Journal of Mathematics*, 35(4):413–422, 1913.

[7] V. Diekert and P. Gastin. First-order definable languages. In J. Flum, E. Grädel, and Th. Wilke, editors, *Logic and Automata: History and Perspectives*, Texts in Logic and Games, pages 261–306. Amsterdam University Press, 2008.

[8] V. Diekert, P. Gastin, and M. Kufleitner. A survey on small fragments of first-order logic over finite words. *International Journal of Foundations of Computer Science*, 19(3):513–548, June 2008. Special issue DLT 2007.

[9] V. Diekert, M. Horsch, and M. Kufleitner. On first-order fragments for Mazurkiewicz traces. *Fundamenta Informaticae*, 80(1-3):1–29, 2007.

[10] V. Diekert and M. Kufleitner. Fragments of first-order logic over infinite words (extended abstract). In S. Albers and J.-Y. Marion, editors, *26th International Symposium on Theoretical Aspects of Computer Science, STACS 2009, Freiburg, Germany, February 26-28, 2009, Proceedings*, pages 325–336. Online proceedings at DROPS and HAL, 2009.

[11] K. Etessami, M. Y. Vardi, and Th. Wilke. First-order logic with two variables and unary temporal logic. *Information and Computation*, 179(2):279–295, 2002.

[12] P. Gastin and A. Petit. Infinite traces. In V. Diekert and G. Rozenberg, editors, *The Book of Traces*, chapter 11, pages 393–486. World Scientific, Singapore, 1995.

[13] M. Kufleitner. Polynomials, fragments of temporal logic and the variety DA over traces. In O. H. Ibarra and Z. Dang, editors, *Proc. of the 10th Int. Conf. on Developments in Language Theory (DLT'06)*, volume 4036 of *LNCS*, pages 37–48. Springer-Verlag, 2006.

[14] A. R. Meyer and L. J. Stockmeyer. The equivalence problem for regular expressions with squaring requires exponential space. In *13th Annual Symposium on Switching and Automata Theory*, pages 125–129. IEEE Computer Society, 1972.

[15] D. Perrin and J.-É. Pin. *Infinite words*, volume 141 of *Pure and Applied Mathematics*. Elsevier, Amsterdam, 2004.

[16] J.-É. Pin. *Varieties of Formal Languages*. North Oxford Academic, London, 1986.

[17] J.-É. Pin. Positive varieties and infinite words. In C. Lucchesi and A. Moura, editors, *Latin'98*, volume 1380 of *Lecture Notes in Computer Science*, pages 76–87. Springer-Verlag, 1998.

[18] M. P. Schützenberger. Sur le produit de concaténation non ambigu. *Semigroup Forum*, 13:47–75, 1976.

[19] A. P. Sistla, M. Y. Vardi, and P. L. Wolper. The complementation problem for Büchi automata with applications to temporal logic. *Theoretical Computer Science*, 49(2-3):217–237, 1987.

[20] P. Tesson and D. Thérien. Diamonds are Forever: The Variety DA. In G. M. dos Gomes Moreira da Cunha, P. V. A. da Silva, and J.-É. Pin, editors, *Semigroups, Algorithms, Automata and Languages, Coimbra (Portugal) 2001*, pages 475–500. World Scientific, 2002.

[21] P. Tesson and D. Thérien. Logic meets algebra: The case of regular languages. *Logical Methods in Computer Science*, 3(1):1–37, 2007.

[22] D. Thérien and Th. Wilke. Over words, two variables are as powerful as one quantifier alternation. In *STOC*, pages 234–240, 1998.

[23] W. Thomas. Classifying regular events in symbolic logic. *Journal of Computer and System Sciences*, 25:360–376, 1982.

[24] W. Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, chapter 4, pages 133–191. Elsevier Science Publishers B. V., 1990.

[25] P. Weil. Some results on the dot-depth hierarchy. *Semigroup Forum*, 46:352–370, 1993.

[26] P. Weil. Algebraic recognizability of languages. In *Mathematical Foundations of Computer Science 2004*, volume 3153 of *Lecture Notes in Computer Science*, pages 149–174. Springer, Berlin, 2004.

[27] Th. Wilke. An Eilenberg theorem for ∞-languages. In *ICALP'91*, volume 510 of *Lecture Notes in Computer Science*, pages 588–599. Springer-Verlag, 1991.

[28] Th. Wilke. *Classifying Discrete Temporal Properties*. Habilitationsschrift, Universität Kiel, April 1998.