Submitted to the Symposium on Theoretical Aspects of Computer Science
www.stacs-conf.org

FRAGMENTS OF FIRST-ORDER LOGIC OVER INFINITE WORDS

VOLKER DIEKERT [1] AND MANFRED KUFLEITNER [1]

[1] Universität Stuttgart, FMI, Universitätsstraße 38, D-70569 Stuttgart, Germany

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, Δ_2, $\mathrm{FO}^2 \cap \Sigma_2$ (and by duality $\mathrm{FO}^2 \cap \Pi_2$), and FO^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 [18] 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 for fragments of first-order logic over finite words has been very fruitful. For example, a result of Wilke and Thérien is that FO^2 and Δ_2 have the same expressive power [13], 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. [12] or [3] for surveys. Several results for finite words have been extended to other structures such as trees and other graphs, see [16] for a survey. 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, it is clear that 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.

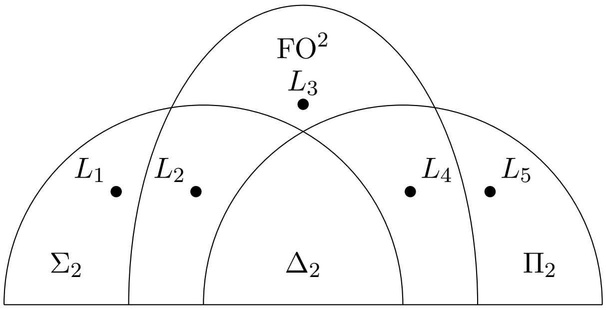
Our results deepen the understanding of first-order fragments over infinite words. A decidable characterization of the membership problem for FO^2 over infinite words has been given in the habilitation thesis of Wilke [18]. Recently, decidability for Σ_2 has been shown independently by Bojańczyk [2]. Language theoretic and decidable algebraic characterizations of the fragment Σ_1 and of its Boolean closure can be found in [7].

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 in terms of a property of its syntactic monoid and by requiring that L is open in some alphabetic topology. Both properties are decidable.

Received by the editors September 14, 2008.
1998 ACM Subject Classification: F.4.1 Mathematical Logic, F.4.3 Formal Languages.
Key words and phrases: infinite words, regular languages, first-order logic, automata theory, semigroups, topology.

Our second result is that a 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 [18]). 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.

Further main 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. In particular, 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 $\Delta_2 = \Sigma_2 \cap \Pi_2$:



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

L_1 = "there exists a factor ab" $= \Gamma^* ab \Gamma^\infty$
L_2 = "finitely many a's"
L_3 = "finitely many a's and infinitely many b's" $= L_2 \cap L_4$
L_4 = "infinitely many b's"
L_5 = "there is no factor ab" $= \Gamma^\infty \setminus L_1$

It will turn out that L_4 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 with respect to our topology is empty.

For basic notions on languages of infinite words we refer to standard references such as [7, 15]. **Proofs which are omitted can be found in the appendix of this submission.**

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 (finite) 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 different occurrences in α. The notation has been introduced in the framework of so called *complex traces*, see e.g. [5] for a detailed discussion of this concept. The notation $\mathrm{im}(\alpha)$ refers to the *imaginary part* and we adopt it here. A crucial role for us are sets of the form A^{im}, where, 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. All languages L here can be assumed to be *regular*. The finite part $L \cap \Gamma^*$ can be assumed to be specified by some NFA and infinite part $L \cap \Gamma^\omega$ can be assumed to be specified by some Büchi automaton. 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 less 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. 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 languages 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 [13] 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 it also coincides with unambiguous polynomials. This classical result is due to Schützenberger [9]. We refer to [11, 3] for more background on the class **DA**. It is a class of finite monoids defined e.g. by equations of type $(xy)^\omega = (xy)^\omega y (xy)^\omega$. We recall that the class **DA** can also be defined by equations of the form $e = ese$ for all idempotents e (i.e., $e^2 = e$) and for all s generated by factors of e.

Saying that formulas are equivalent if they agree on all finite and infinite words 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 not 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 8.1 which means $\mathrm{FO}^2 = \Delta_2$ for finite words by choosing $A = \emptyset$. An important concept in this paper is topology.

3. The alphabetic topology and polynomials

We equip Γ^∞ with a refinement of the usual Cantor topology. As we will see, topological information is crucial in our characterization results. We define the *alphabetic topology* by its basis, which is given by all sets of the form uA^∞. Thus, a set L 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. All sets A^∞ are clopen. A set A^{im} is not open unless $A = \emptyset$, it is not closed unless $A = \Gamma$.

Remark 3.1. The space Γ^∞ with the alphabetic topology is Hausdorff, but not compact, in general (in contrast to the Cantor topology). To see that it is not compact for $\Gamma = \{a, b\}$ note that $\Gamma^\infty = a^\omega \cup \Gamma^* b \Gamma^\infty$. The singleton set a^ω is clopen, but for no finite subset $F \subseteq \Gamma^*$ we have $\Gamma^\infty = a^\omega \cup F b \Gamma^\infty$.

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 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 $L/A^\infty = \{u \in \Gamma^* \mid u\alpha \in L \text{ for } \alpha \in A^\infty\}$. For $L \subseteq \Gamma^*$ we define

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

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

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

Corollary 3.3. *Given a regular language $L \subseteq \Gamma^\infty$, we can decide whether L is closed (open resp., clopen resp.).*

Actually, we have a more precise statement than pure decidability.

Theorem 3.4. *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?

Remark 3.5. Neither languages of the form $\overrightarrow{L/A^\infty}$ nor $\overrightarrow{L/A^\infty} \cap \overline{A^{\mathrm{im}}}$ as in Proposition 3.2 need to be closed. Indeed, 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. Finally note that $\overline{L}$ is not closed in the Cantor topology since $a^\omega \notin \overline{L}$. Remember that a basis of the Cantor topology are the sets of the form $u\Gamma^\infty$.

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 $A^* = \emptyset^\infty \cup \bigcup_{a \in A} A^* a$, and polynomials are closed under intersection. Thus, $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^*$ is in our language a polynomial, but no 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.

By definition of the alphabetic topology, 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.

Lemma 3.6. *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*

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

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 $ua_i \cdots a_k \gamma \in P$, and hence $ua_i \cdots a_k \gamma \in uA^\infty \cap L$.

Let now $\alpha \in \overline{L}$ and write $\alpha \in uv_1 \cdots v_{k+1} A^\infty \cap A^{\mathrm{im}}$ with $\mathrm{alph}(v_j) = A$. There exists $\gamma \in A^\infty$ such that $uv_1 \cdots v_{k+1}\gamma \in P \cap B^{\mathrm{im}}$. This implies $B \subseteq A$. Since $uv_1 \cdots v_{k+1}\gamma \in A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ there are some $1 \leq i, j \leq k+1$ such that $uv_1 \cdots v_{j-1} \in A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^* \cap A^{\mathrm{im}}$, $v_j \in A_i^*$, 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 $\alpha \in A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty \cap A^{\mathrm{im}}$. ∎

4. 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 $M_e = \{s \in M \mid e \in MsM\}^*$. By definition, M_e is a submonoid of M. If M is generated by 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 $s \in M_e$.

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. 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 $v^2 \equiv_L v$.

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 $s \in M$. The following notations are used:

- $(s, e) \in M \times M$ 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 $t \leq s$.

Lemma 4.1. *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 $vy^\omega \in [t][f]^\omega \cap L$. ∎

For lack of space and in order to avoid too much machinery we do not treat ω-semigroups here [17, 8]. This is postponed to the journal version. However, let us define $tf^\omega \leq_L se^\omega$ for linked pairs 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 4.2. *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 $stf^\omega \leq_L se^\omega$.*

5. The fragment Σ_2

By a (slight extension of a) result of Thomas [14] 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 by Bojańczyk [2] using different techniques. We know that polynomials are open. Therefore, we concentrate on algebra.

Lemma 5.1. *Let $L \subseteq \Gamma^\infty$ be a polynomial. Then all idempotents of $\mathrm{Synt}(L)$ are locally top.*

Theorem 5.2. *Let $L \subseteq \Gamma^\infty$ be a regular language. The following 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 $\Leftrightarrow$ 2": This is a slight modification of a result by Thomas [14].
"2 $\Rightarrow$ 3": By definition, polynomials are open in the alphabetic topology. In Lemma 5.1 it has been shown that all idempotent elements are locally top.
"3 $\Leftrightarrow$ 4": The equivalence of L being open and "4a" is Lemma 4.2. Property "4b" is the definition of all elements being locally top.
"4 $\Rightarrow$ 5": Let $h = h_L$ be the syntactic homomorphism onto the syntactic monoid $M = \mathrm{Synt}(L)$. Applying Lemma 4.2, property "5a" follows from "4a" and "5b" trivially follows from "4b". The condition "5c" holds for $\mathrm{Synt}(L)$ by Lemma 4.1.
"5 $\Rightarrow$ 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 5.3. *It is decidable whether a regular language is Σ_2-definable.*

Remark 5.4. 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 5.3 transfers to ω-regular languages.

6. Two variable first-order logic

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 [18].

Lemma 6.1. *Let $L \subseteq \Gamma^\infty$ be FO^2-definable. Then the syntactic monoid $\mathrm{Synt}(L)$ is 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 proof the converse of Lemma 6.1.

Lemma 6.2. *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 $fef = f$. Since h strongly recognizes L, we can compute as follows:

$$\beta \in [sr][f]^\omega = [sr][fef]^\omega = [srfe][efe]^\omega = [s][e]^\omega \subseteq L$$

In particular, $\beta \in L$. ∎

Lemma 6.3. *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, we may assume that this vector 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 $|v_1 a_1 \cdots v_i a_i|$ 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 $u'wa_{i+1} \cdots a_k z^\omega \in [s][e]^\omega \subseteq L$. ∎

Lemma 6.4. *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.*

Theorem 6.5. *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 6.1. "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 6.2. "4 ⇒ 5": Lemma 6.3. "5 ⇒ 1": Lemma 6.4. ∎

Etessami, Vardi, and Wilke have given a characterization of FO^2 in terms of unary temporal logic [4]. Recall that if a language $L \subseteq \Gamma^\infty$ is weakly recognizable by a finite monoid, then it is also strongly recognizable by a finite monoid. The same holds for aperiodic monoids, but Theorem 6.5 suggests that this fails for $\mathbf{DA}$. Indeed, we have the following example.

Example 6.6. 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 $\mathbf{DA}$. 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 $m \geq 0$.

7. Unambiguous polynomials and the fragments $\mathrm{FO}^2 \cap \Sigma_2$ and $\mathrm{FO}^2 \cap \Pi_2$

Theorem 7.1. *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.*

Theorem 7.2. *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.*

Theorem 7.2 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 $L \in \Pi_2 \cap \mathrm{FO}^2$), 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 hopeless: Let $L = \Gamma^* a \cup \Gamma^\omega$, then L is closed and in $\Pi_2 \cap \mathrm{FO}^2$, 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 3.6, the closure of L is $\overline{L} = L \cup \Gamma^{\mathrm{im}}$ which is not Σ_2-definable.

8. The fragment $\Delta_2 = \Sigma_2 \cap \Pi_2$

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 8.1. *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 $\Rightarrow$ 2": By Theorem 6.5 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 7.1 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 $\Rightarrow$ 3": Trivial. "3 $\Rightarrow$ 1": If $L = L_\sigma \cap A^{\mathrm{im}}$, then a slight modification of the proof for Lemma 5.1 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 6.5 we see that L is FO^2-definable. ∎

Note that we cannot expect that $L_\sigma = L_\pi$ in the statement above, because L_σ is open and L_π is closed. Hence, a language in Δ_2 must be clopen. The first step for a convenient characterization on Δ_2 is therefore a description of clopen unambiguous monomials.

Lemma 8.2. *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 $1 \leq i \leq k$ 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.*

Lemma 8.3. *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 $1 \leq i \leq s$. 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 $1 \leq i \leq s$.
- $\{b_i, \ldots, b_t\} \not\subseteq B_i$ for all $s+1 \leq i \leq t$.
- $A = \{b_{s+1}, \ldots, b_t\}$.

It suffices to proof 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 8.2 there exists $1 \leq i \leq s$ 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 superset of $\{b_i, \ldots, b_t\}$ anymore. 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 $\alpha = u v_1 \cdots v_r \in B_1^* b_1 \cdots B_{i-1}^* b_{i-1} R^r$, 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

$$u v_1 \cdots v_{j_1 - 1} (v_{j_1} \cdots v_{j_2})^\omega \in L.$$

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 $\widetilde{P} \subseteq Q$. ∎

Theorem 8.4. *Let $L \subseteq \Gamma^\infty$. 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 $1 \le i \le k$ such that $\{a_i, \ldots, a_k\} \subseteq A_i$.*
4. *L is regular, $\mathrm{Synt}(L) \in \mathbf{DA}$, 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 $s = tx$ and $t = sy$) we have*

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

Proof. "1 ⇒ 2": By Theorem 5.2 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 6.5 it follows that L is FO^2-definable. "2 ⇒ 3": By Theorem 7.1, L is a finite union of unambiguous monomials. Property "3" now follows by Lemma 8.3 and Lemma 8.2. "3 ⇒ 1": Theorem 7.1 and Theorem 7.2.
"2 ⇒ 4": By Theorem 6.5, we see that $\mathrm{Synt}(L) \in \mathbf{DA}$. Suppose $[s][e]^\omega \subseteq L$ and let $s = tx$ and $t = sy$. Since L is closed we see that $[s][eyfx]^\omega \subseteq L$ and by strong recognition we conclude $[t][fxey]^\omega \subseteq L$. 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 $[t][fxey]^r A^\infty \subseteq L$. Moreover, $t = tfxey$ and thus, $[t]A^\infty \subseteq L$. In particular, $[t][f]^\omega \subseteq L$ because $[f] \subseteq A^*$.
"4 ⇒ 2": Definability in FO^2 follows by Theorem 6.5. 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 $v \le \beta$ 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 $v' \le \gamma$ 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 $uv\gamma \in [t][f]^\omega \subseteq L$. ∎

9. 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 "clopen in some appropriate 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.

Acknowledgement

We thank Luc Dartois from the ENS Cachan for helpful discussions during his internship in Stuttgart 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] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

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

[10] 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.

[11] 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.

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

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

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

[15] 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.

[16] 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.

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

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

Appendix A. Proofs from Section 3

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 the inclusion $\overline{L} \subseteq \bigcup_{A \subseteq \Gamma} (\overrightarrow{L/A^{\infty}} \cap A^{\mathrm{im}})$. 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 $\bigcup_{A \subseteq \Gamma} (\overrightarrow{L/A^{\infty}} \cap A^{\mathrm{im}}) \subseteq \bigcup_{A \subseteq \Gamma} (\overrightarrow{L/A^{\infty}} \cap \overline{A^{\mathrm{im}}})$ 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. ∎

Corollary 3.3. *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 together with some Büchi automaton. The construction of an NFA recognizing L/A^{∞} is clear. From that we obtain a (deterministic) Büchi automaton recognizing $\overrightarrow{L/A^{\infty}}$. Intersection with A^{im} yields the automaton for $\overline{L} \cap A^{\mathrm{im}}$. 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. ∎

The proof of Theorem 3.4 is given in the next two Propositions.

Proposition A.1. *We can check in* PSPACE *whether a regular language $L \subseteq \Gamma^{\omega}$ is closed.*

Proof. 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. For all $A \subseteq \Gamma$ we can check in PSPACE whether $L(\mathcal{A}) \cap A^{\mathrm{im}} = \overrightarrow{L/A^{\infty}} \cap A^{\mathrm{im}}$, see [10]. ∎

Proposition A.2. *It is* PSPACE*-hard to decide whether a regular language $L \subseteq \Gamma^{\omega}$ is closed.*

Proof. We use a reduction of the problem whether $L(\mathcal{A}) = \Gamma^*$ for some NFA $\mathcal{A}$, see [6]. 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 $L(\mathcal{B}) = \{w_1 c w_2 c \cdots \in (\Gamma \cup \{c\})^{\omega} \mid \forall i \colon w_i \in \Gamma^* \text{ and } \exists i \colon w_i \in L(\mathcal{A})\}$. 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. ∎

Appendix B. Proofs from Section 4

Lemma 4.2. *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 $\mathrm{alph}(t) \subseteq \mathrm{alph}(tf) \subseteq A$. Hence $stf^\omega \subseteq L$. 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. ∎

Appendix C. Proofs from Section 5

Lemma 5.1. *Let $L \subseteq \Gamma^\infty$ be 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 $e^n \equiv e$. 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. Now 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 $[e]_L$ 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., $h(e)_L$ is locally top. ∎

Corollary 5.3. *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 5.2 are decidable. ∎

Appendix D. Proofs from Section 6

Lemma 6.1. *Let $L \subseteq \Gamma^\infty$ be* FO^2*-definable. Then the syntactic monoid* $\mathrm{Synt}(L)$ *is in* **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 $\alpha = xw^nuw^nyz^\omega$, $\alpha' = xw^nyz^\omega$ and $\beta = x(w^nuw^ny)^\omega$, $\beta' = x(w^ny)^\omega$. 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, $\mathrm{Synt}(L) \in$ **DA**. ∎

Lemma 6.4. *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 $\alpha \in L(\varphi) a L(\psi) \subseteq P$ and $L(\psi) \subseteq (\Gamma \setminus \{a\})^\infty$. Now, the last a-position x in every $\alpha \in L(\varphi) a L(\psi)$ 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 $L(\varphi) a L(\psi) = L\big(\varphi_{<a} \wedge \exists x \colon \xi(x) \wedge \psi_{>a}\big)$. 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). ∎

Appendix E. Proofs from Section 7

Theorem 7.1. *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 5.2.
"2 ⇒ 3": Let $\alpha \in L \in \mathrm{FO}^2 \cap \Sigma_2$. By Theorem 6.5 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 $uv_1 \cdots v_\ell v^\infty \in PA^\infty \cap A^{\mathrm{im}} \subseteq L$. 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 $\Rightarrow$ 1": Theorem 5.2 and Theorem 6.5.
"1 $\Leftrightarrow$ 4": This is the dual statement of Theorem 7.2. The proof of this theorem in turn uses "2 $\Rightarrow$ 1", but this has just been shown. ∎

We have the following characterization of the class $\mathrm{FO}^2 \cap \Pi_2$ which also yields the missing part "1 $\Leftrightarrow$ 4" in Theorem 7.1.

Theorem 7.2. *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 $\Rightarrow$ 2": This is the dual statement of "1 $\Rightarrow$ 2" in Theorem 6.5.
"2 $\Rightarrow$ 3" is trivial.
"3 $\Rightarrow$ 1": By Theorem 6.5 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 3.6 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 7.1, "2 $\Rightarrow$ 1", and therefore L is Π_2-definable. ∎

Appendix F. Proofs from Section 8

Lemma 8.2. *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 $1 \leq i \leq k$ 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 3.6 we know 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 $1 \leq i \leq k$, 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 $1 \leq i \leq k$. We have $a_1 \cdots a_{i-1}(a_i \cdots a_k)^m \in P$ for all $m \geq 1$. 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. ∎