On the Borel Inseparability of Game Tree Languages

Szczepan Hummel, Henryk Michalewski, Damian Niwinski

Symposium on Theoretical Aspects of Computer Science 2009 (Freiburg), pp. 565–576
www.stacs-conf.org

ON THE BOREL INSEPARABILITY OF GAME TREE LANGUAGES

SZCZEPAN HUMMEL [1] AND HENRYK MICHALEWSKI [1] AND DAMIAN NIWIŃSKI [1]

[1] Faculty of Mathematics, Informatics, and Mechanics
Warsaw University, Poland
E-mail address: {shummel,henrykm,niwinski}@mimuw.edu.pl

Abstract. The game tree languages can be viewed as an automata-theoretic counterpart of parity games on graphs. They witness the strictness of the index hierarchy of alternating tree automata, as well as the fixed-point hierarchy over binary trees.

We consider a game tree language of the first non-trivial level, where Eve can force that 0 repeats from some moment on, and its dual, where Adam can force that 1 repeats from some moment on. Both these sets (which amount to one up to an obvious renaming) are complete in the class of co-analytic sets. We show that they cannot be separated by any Borel set, hence *a fortiori* by any weakly definable set of trees.

This settles a case left open by L.Santocanale and A.Arnold, who have thoroughly investigated the separation property within the μ-calculus and the automata index hierarchies. They showed that separability fails in general for non-deterministic automata of type Σ^μ_n, starting from level $n = 3$, while our result settles the missing case $n = 2$.

Introduction

In 1970 Rabin [15] proved the following property: If a set of infinite trees can be defined both by an *existential* and by a *universal* sentence of monadic second order logic then it can also be defined in a weaker logic, with quantification restricted to *finite* sets. An automata-theoretic counterpart of this fact [15, 12] states that if a tree language, as well as its complement, are both recognizable by Büchi automata (called *special* in [15]) then they are also recognizable by weak alternating automata. Yet another formulation, in terms of the μ-calculus [3], states that if a tree language is definable both by a Π^μ_2-term (i.e., with a pattern $\nu\mu$) and a Σ^μ_2-term ($\mu\nu$), then it is also definable by an alternation free term, i.e., one in $Comp(\Pi^\mu_1 \cup \Sigma^\mu_1)$. This last formulation gives rise to a question if the equation

$$\Pi^\mu_n \cap \Sigma^\mu_n = Comp(\Pi^\mu_{n-1} \cup \Sigma^\mu_{n-1})$$

holds on all levels of the fixed-point hierarchy. Santocanale and Arnold showed [17], rather surprisingly, that it is not the case for $n \geq 3$. They exhibit a series of "ambiguous" properties, expressible by terms in Π^μ_n and in Σ^μ_n, but not in $Comp(\Pi^\mu_{n-1} \cup \Sigma^\mu_{n-1})$. On

1998 ACM Subject Classification: F.1.1 Automata, F.4.1 Set theory, F.4.3 Classes defined by grammars or automata.

Key words and phrases: Tree automata, Separation property, Borel sets, Parity games.

All authors are supported by Grant N206 008 32/0810.

SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

positive side however, they discover a more subtle generalization of Rabin's result, which continues to hold on the higher stages of the hierarchy.

Let us explain it at a more abstract level, with $\mathcal{L}$ ("large") and $\mathcal{S}$ ("small") being two classes of subsets of some universe U. Consider the following properties.

Simplification. Whenever L and its complement $\bar{L}$ are both in $\mathcal{L}$, they are also in $\mathcal{S}$.

Separation. Any two disjoint sets $L, M \in \mathcal{L}$ are separated by some set K in $\mathcal{S}$ (i.e., $L \subseteq K \subseteq U - M$).

Note that (given some $\mathcal{L}$ and $\mathcal{S}$) separation implies simplification, but in general not *vice versa*. In topology, it is well known (see, e.g., [11]) that the separation property holds for

$\mathcal{L}$ = analytic ($\mathbf{\Sigma}_1^1$) subsets of a Polish space (e.g., $\{0,1\}^\omega$),
$\mathcal{S}$ = Borel sets,

but fails for $\mathcal{L}$ = co-analytic sets ($\mathbf{\Pi}_1^1$) and $\mathcal{S}$ as above. On the other hand both classes enjoy the simplification property (which amounts to the Suslin Theorem).

In this setting, Rabin's result establishes the simplification property for

$\mathcal{L}$ = Büchi definable tree languages (Π_2^μ in the fixed-point hierarchy),
$\mathcal{S}$ = weakly definable tree languages ($Comp(\Pi_1^\mu \cup \Sigma_1^\mu)$).

A closer look at the original proof reveals that a (stronger) separation property also holds for these classes.

Santocanale and Arnold [17] showed in turn that the separation property holds for

$\mathcal{L}$ = tree languages recognizable by *non-deterministic* automata of level Π_n^μ,
$\mathcal{S}$ = tree languages definable by fixed-point terms in $Comp(\Pi_{n-1}^\mu \cup \Sigma_{n-1}^\mu)$,

for the remaining case of $n \geq 3$. On the negative side, they showed that the separation property fails for $\mathcal{L}$ consisting of tree languages recognizable by *non-deterministic* automata of level Σ_n^μ, for $n \geq 3$, leaving open the case of $n = 2$. In fact, their proof reveals that, in the case under consideration, even a (weaker) simplification property fails (see [17], section 2.2.3). As for Σ_2^μ however, the simplification property does hold, because of Rabin's result[1]. For this reason, the argument of Santocanale and Arnold cannot be extended to the class Σ_2^μ. In the present paper, we show that the separation property fails also in this case, completing the missing point in the classification of [17].

We use a topological argument and show in fact a somewhat stronger result, exhibiting two disjoint languages recognized by non-deterministic tree automata with co-Büchi condition (i.e., Σ_2^μ), which cannot be separated by any *Borel* set (in a standard Cantor-like topology on trees). The languages in question are the so-called *game tree languages* (of level (0,1)), which were used in [8] (and later also in [2]) in the proof of the strictness of the fixed-point hierarchy over binary trees. More specifically, one of these languages consists of the trees labeled in $\{0,1\} \times \{\exists, \forall\}$, such that in the induced game (see definition below) *Eve* has a strategy to force only 0's from some moment on. The second is the twin copy of the first and consists of those trees that *Adam* has a strategy to force only 1's from some moment on.

[1] If a set and its complement are recognized by non-deterministic co-Büchi automata then they are also both recognized by alternating Büchi automata [5], and hence by non-deterministic Büchi automata, and hence are weakly definable [15].

The wording introduced above differs slightly from the standard terminology of descriptive set theory, where a *separation property* of a class $\mathcal{L}$ means our property with $\mathcal{S} = \{X : X, \bar{X} \in \mathcal{L}\}$ (see [11]). To emphasize the distinction, following [1], we will refer to the latter as to the *first separation property*. In this setting, the first separation property holds for the class of Büchi recognizable tree languages, but it fails for the co-Büchi languages, similarly as it is the case of the analytic *vs.* co-analytic sets, mentioned above. This may be read as an evidence of a strong analogy between the Büchi class and $\mathbf{\Sigma}_1^1$. In fact, Rabin [15] early observed that the Büchi tree languages are definable by existential sentences of monadic logic, and hence analytic. We show however that, maybe surprisingly, the converse is not true, by exhibiting an analytic tree language, recognized by a parity (Rabin) automaton, but not by any Büchi automaton.

Note. The fixed-point hierarchy discussed above provides an obvious context of our results, but in the paper we do not rely on the μ-calculus concepts or methods. For definitions of relevant concepts, we refer an interested reader to the work by Santocanale and Arnold [17] or, e.g., to [4].

1. Basic concepts

Throughout the paper, ω stands for the set of natural numbers.

Metrics on trees. A full binary tree over a finite alphabet Σ (or shortly a tree, if confusion does not arise) is represented as a mapping $t : \{1,2\}^* \to \Sigma$.

We consider the classical topology *à la Cantor* on T_Σ induced by the metric

$$d(t_1,t_2) = \begin{cases} 0 & \text{if } t_1 = t_2 \\ 2^{-n} \text{ with } n = \min\{|w| : t_1(w) \neq t_2(w)\} & \text{otherwise} \end{cases} \tag{1.1}$$

It is well-known and easy to see that if Σ has at least two elements then T_Σ with this topology is homeomorphic to the Cantor discontinuum $\{0,1\}^\omega$. Indeed, it is enough to fix a bijection $\alpha : \omega \to \{1,2\}^*$, and a mapping (code) $C : \Sigma \to \{0,1\}^*$, such that $C(\Sigma)$ forms a maximal antichain w.r.t. the prefix ordering. Then $T_\Sigma \ni t \mapsto C \circ t \circ \alpha \in \{0,1\}^\omega$ is a desired homeomorphism. We assume that the reader is familiar with the basic concepts of set-theoretic topology (see, e.g., [11]). The *Borel sets* over T_Σ constitute the least family containing open sets and closed under complement and countable union. The *Borel relations* are defined similarly, starting with open relations (i.e., open subsets of T_Σ^n, for some n, considered with product topology). The *analytic* (or $\mathbf{\Sigma}_1^1$) sets are those representable by

$$L = \{t : (\exists t')\, R(t,t')\}$$

where $R \subseteq T_\Sigma \times T_\Sigma$ is a Borel relation. The *co-analytic* (or $\mathbf{\Pi}_1^1$) sets are the complements of analytic sets. A continuous mapping $f : T_\Sigma \to T_\Sigma$ *reduces* a tree language $A \subseteq T_\Sigma$ to $B \subseteq T_\Sigma$ if $f^{-1}(B) = A$. As in complexity theory, a set $L \in \mathcal{K}$ is *complete* in class $\mathcal{K}$ if all sets in this class reduce to it.

Non-deterministic automata. A non-deterministic tree automaton over trees in T_Σ with a parity acceptance condition[2] is presented as $A = \langle \Sigma, Q, q_I, Tr, rank\rangle$, where Q is a finite set of *states* with an *initial state* q_I, $Tr \subseteq Q \times \Sigma \times Q \times Q$ is a set of *transitions*, and $rank : Q \to \omega$ is the *ranking* function. A transition (q, σ, p_1, p_2) is usually written $q \stackrel{\sigma}{\to} p_1, p_2$.

A *run* of A on a tree $t \in T_\Sigma$ is itself a Q–valued tree $\rho : \{1,2\}^* \to Q$ such that $\rho(\varepsilon) = q_I$, and, for each $w \in dom\,(\rho)$, $\rho(w) \stackrel{t(w)}{\to} \rho(w1), \rho(w2)$ is a transition in Tr. A *path* $P = p_0p_1 \ldots \in \{1,2\}^\omega$ in ρ is *accepting* if the highest rank occurring infinitely often along it is even, i.e., $\limsup_{n\to\infty} rank(\rho(p_0p_1 \ldots p_n))$ is even. A *run is accepting* if so are all its paths. A tree language $T(A)$ *recognized* by A consists of those trees in T_Σ which admit an accepting run.

The *Rabin–Mostowski index* of an automaton A is the pair $(\min(rank(Q)), \max(rank(Q)))$; without loss of generality, we may assume that $\min(rank(Q)) \in \{0,1\}$.

An automaton with the *Rabin–Mostowski index* $(1,2)$ is called a *Büchi automaton.* Note that a Büchi automaton accepts a tree t if, on each path, some state of rank 2 occurs infinitely often. We refer to the tree languages recognizable by Büchi automata as to *Büchi (tree) languages*. The *co-Büchi languages* are the complements of Büchi languages. It is known that if a tree language is recognized by a non-deterministic automaton of index $(0,1)$ then it is co-Büchi[3]; the converse is not true in general (see the languages $M_{i,k}$ in Example 1.1 below).

Example 1.1. Let

$$L = \{t \in T_{\{0,1\}} : (\exists P) \limsup_{n\to\infty} t(p_0p_1 \ldots p_n) = 1\}$$

This set is recognized by a Büchi automaton with transitions

$$\begin{array}{lll} q/p \stackrel{0}{\to} q, T; & q/p \stackrel{1}{\to} p, T; & T \stackrel{(0/1)}{\to} T, T; \\ q/p \stackrel{0}{\to} T, q; & q/p \stackrel{1}{\to} T, p; & \end{array}$$

with $rank(q) = 1$ and $rank(p) = rank(T) = 2$. Rabin [15] showed that its complement $\bar{L}$ cannot be recognized by any Büchi automaton, but it is recognizable by an (even deterministic) automaton of index $(0,1)$

$$0/1 \stackrel{0}{\to} 0, 0; \qquad 0/1 \stackrel{1}{\to} 1, 1; \qquad rank(i) = i; \quad for \quad i = 0, 1.$$

This last set can be generalized to the so-called parity languages (with $i \in \{0,1\}$)

$$M_{i,k} = \{t \in T_{\{i,\ldots,k\}} : (\forall P) \limsup_{n\to\infty} t(p_0p_1 \ldots p_n) \;\; is \;\; even\}$$

which are all co-Büchi but require arbitrary high indices [13]. It can also be showed that all languages $M_{i,k}$ (except for $(i,k) = (0,0), (1,1), (1,2)$) are complete in the class of co-analytic sets $\mathbf{\Pi}^1_1$ (see, e.g., [14]).

The class of languages which are simultaneously Büchi and co-Büchi has numerous characterizations mentioned in the introduction; all these characterizations easily imply that such sets are Borel (even of finite Borel rank).

[2]Currently most frequently used in the literature, these automata are well-known to be equivalent to historically previous automata with the Muller or Rabin conditions [18].

[3]It follows, in particular, from the equivalence of the non-deterministic and alternating Büchi automata [5], mentioned in footnote 1.

Example 1.2. Consider the set $\bar{L} = M_{0,1}$ of Example 1.1, and its twin copy obtained by the renaming $0 \leftrightarrow 1$,

$$M'_{0,1} = \{t \in T_{\{0,1\}} : (\forall P) \liminf_{n\to\infty} t(p_0 p_1 \ldots p_n) = 1\}.$$

The sets $M_{0,1}$ and $M'_{0,1}$ are disjoint, co-Büchi and, as we have already noted, $\mathbf{\Pi}^1_1$ complete. They can be separated by a set K of trees[4], such that on the *rightmost* branch, there are only finitely many 1's

$$K = \{t \in T_{\{0,1\}} : \limsup_{n\to\infty} t(\underbrace{22\ldots2}_{n}) = 0\}$$

(i.e., $M_{0,1} \subseteq K \subseteq T_{\{0,1\}} - M'_{0,1}$). The set K can be presented as a countable union of closed sets

$$K = \bigcup_m \{t \in T_{\{0,1\}} : (\forall n \geq m)\, t(\underbrace{22\ldots2}_{n}) = 0\}$$

so it is on the level $\mathbf{\Sigma}^0_2$ (i.e., F_σ) of the Borel hierarchy. The membership in the Borel hierarchy can also be seen trough an automata-theoretic argument by showing that K is simultaneously Büchi and co-Büchi. Indeed it can be recognized by an (even deterministic) automaton with co-Büchi condition

$$0/1 \xrightarrow{0} T, 0; \qquad 0/1 \xrightarrow{1} T, 1; \qquad T \xrightarrow{(0/1)} T, T; \qquad \mathit{rank}(i) = i, \ \mathit{for}\ \ i = 0, 1, \quad \mathit{rank}(T) = 0,$$

as well as by a (non-deterministic) Büchi automaton

$$q \xrightarrow{(0/1)} T, q/p; \qquad p \xrightarrow{0} T, p; \qquad T \xrightarrow{(0/1)} T, T; \qquad \mathit{rank}(q) = 1, \qquad \mathit{rank}(p) = \mathit{rank}(T) = 2.$$

We will see in the next section that a Borel separation of co-Büchi languages is not always possible.

2. Inseparable pair

Let

$$\Sigma = \{\exists, \forall\} \times \{0, 1\},$$

we denote by π_i the projection on the ith component of Σ. With each $t \in T_\Sigma$, we associate a game $G(t)$, played by two players, *Eve* and *Adam*. The positions of Eve are those nodes v, for which $\pi_1(t(v)) = \exists$, the remaining nodes are positions of Adam. For each position v, it is possible to move to one of its successors, $v1$ or $v2$. The players start in the root and then move down the tree, thus forming an infinite path $P = (p_0 p_1 p_2 \ldots)$. The successor is selected by Eve or Adam depending on who is the owner of the position $p_0 p_1 \ldots p_{n-1}$. The play is won by Eve if

$$\limsup_{n\to\infty} \pi_2\,(t(p_0 p_1 \ldots p_n)) = 0$$

i.e., 1 occurs only finitely often, otherwise Adam is the winner. A strategy for Eve selects a move for each of her positions; it is winning if any play consistent with the strategy is won by Eve. We say that Eve *wins* the game $G(t)$ if she has a winning strategy. The analogous concepts for Adam are defined similarly.

A reader familiar with the *parity games* ([10], see also [18]) has noticed of course that the games $G(t)$ are a special case of these (with the index $(0, 1)$).

[4]This argument is due to Paweł Milewski.

Now let

$$W_{0,1} = \{t : \textit{Eve wins } G(t)\}$$

We also define a set $W'_{0,1} \subseteq T_\Sigma - W_{0,1}$, consisting of those trees t, where Adam has a strategy which guarantees him not only to win in $G(t)$, but also to force a stronger condition, namely

$$\liminf_{n\to\infty} \pi_2\left(t(p_0 p_1 \dots p_n)\right) = 1.$$

It should be clear that $W'_{0,1}$ can be obtained from $W_{0,1}$ by applying (independently on each component) a renaming $0 \leftrightarrow 1$, $\exists \leftrightarrow \forall$. Thus, the sets $W_{0,1}$ and $W'_{0,1}$ are disjoint, but have identical topological and automata-theoretic properties.

Let us see that the set $W_{0,1}$ can be recognized by a non-deterministic automaton of index $(0,1)$; it is enough to take the states $\{0,1\} \cup \{T\}$, with $\mathit{rank}(T) = 0$, and $\mathit{rank}(\ell) = \ell$, for $\ell \in \{0,1\}$, the initial state 0, and transitions

$$\ell \overset{(\forall,m)}{\to} m, m; \qquad \ell \overset{(\exists,m)}{\to} m, T; \qquad \ell \overset{(\exists,m)}{\to} T, m; \qquad T \overset{(Q,m)}{\to} T, T,$$

with $m \in \{0,1\}$, and $Q \in \{\exists, \forall\}$. Hence, the sets $W_{0,1}$ and $W'_{0,1}$ are co-Büchi (c.f. the remark before Example 1.1).

We are ready to state the main result of this paper.

Theorem 2.1. *The sets $W_{0,1}$ and $W'_{0,1}$ cannot be separated by any Borel set.*

Proof. The proof relies on the following.

Lemma 2.2. *For any Borel set $B \subseteq T_\Sigma$, there is a continuous function $f_B : T_\Sigma \to T_\Sigma$, such that*

$$\begin{array}{lcl} u \in B & \Rightarrow & f_B(u) \in W_{0,1} \\ u \notin B & \Rightarrow & f_B(u) \in W'_{0,1} \end{array}$$

Proof. Note that f_B is required to reduce simultaneously B to $W_{0,1}$ and $T_\Sigma - B$ to $W'_{0,1}$. We proceed by induction on the complexity of the set B.

Note first that if B is clopen (simultaneously closed and open) then it is enough to fix two trees $t \in W_{0,1}$ and $t' \in W'_{0,1}$, and define f_B by

$$\begin{array}{lcl} u \in B & \Rightarrow & f_B(u) = t \\ u \notin B & \Rightarrow & f_B(u) = t' \end{array}$$

Also note that, by symmetry of the sets $W_{0,1}$ and $W'_{0,1}$, the claim for B readily implies the claim for the complement $T_\Sigma - B$. (Specifically, $f_{B'}$ is obtained by composing f_B with a suitable renaming.)

Finally note that the space $T_\Sigma \approx \{0,1\}^\omega$ has a countable basis consisting of clopen sets.

Then, in order to complete the proof, it remains to settle the induction step for $B = \bigcup_{n<\omega} B_n$. Assume that we have already the reductions f_{B_n} satisfying the claim, for $n < \omega$. Given $u \in T_\Sigma$, we construct a tree $f_B(u)$, by labeling the rightmost path by $(\exists, 1)$, and letting a subtree in the node $2^n 1$ be $f_{B_n}(u)$ (see Figure 1). In symbols,

$$\begin{array}{lcl} f_B(u)(2^n) & = & (\exists, 1) \\ f_B(u)(2^n 1 v) & = & f_{B_n}(u)(v), \quad \textit{for } \; n < \omega, \quad v \in \{1,2\}^* . \end{array}$$

Since all the functions f_{B_n} are continuous, the resulting f_B is continuous as well. Now, if $u \in B_m$, for some m, then Eve has an obvious winning strategy: follow the rightmost path and turn left in 2^m, then use the winning strategy on the subtree $f_{B_m}(u)$, which exists, by induction hypothesis.

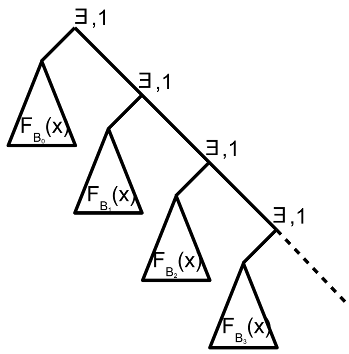

Figure 1: Induction step for $\bigcup_n B_n$.

If, however, $(\forall n)\, u \notin B_n$ then Adam can win the game with the stronger winning criterion, required in the definition of $W'_{0,1}$. Indeed, he can do so as soon as Eve enters any of the subtrees $f_{B_n}(u)$ (by induction hypothesis), but he also wins if Eve remains forever on the rightmost path.

This proves the claim for f_B, and thus completes the proof of the lemma. ∎

We are ready to complete the proof of the theorem. Suppose that there is a Borel set C, such that $W_{0,1} \subseteq C \subseteq T_\Sigma - W'_{0,1}$. The claim of the lemma immediately implies that

$$\begin{array}{lcl} u \in B & \Rightarrow & f_B(u) \in C \\ u \notin B & \Rightarrow & f_B(u) \in T_\Sigma - C. \end{array}$$

Thus any Borel set B over T_Σ is reducible to C, but this is clearly impossible, as it would contradict the strictness of the Borel rank hierarchy in the Cantor discontinuum $\{0,1\}^\omega$ (see, e.g., [11]). ∎

Since the sets $W_{0,1}$ and $W'_{0,1}$ are recognizable by non-deterministic automata of index $(0,1)$, Theorem 2.1 settles the case of $n = 2$, missing in Section 2.2.3 of [17], devoted to the failure of separation property for non-deterministic automata of type Σ^μ_n and the class $Comp(\Pi^\mu_{n-1} \cup \Sigma^\mu_{n-1})$.

In the terminology introduced at the end of introduction, we can state the following.

Corollary 2.3. *The class of co-Büchi tree languages does not have the first separation property.* ∎

This may be contrasted with the positive result of [15]. As we have mentioned in the introduction, Rabin's original proof essentially shows this property for the class of Büchi tree languages, although it is not explicitly stated there. For the sake of completeness, we sketch the argument below, following closely the μ-calculus version of [3] (based on the original proof of [15]).

Theorem 2.4 (Rabin). *The class of Büchi tree languages has the first separation property.*

Proof. Let A and B be two non-deterministic Büchi automata, such that $T(A) \cap T(B) = \emptyset$. We will refer to the states of rank 2 as to *accepting* states (of the corresponding automaton). A *cut* (of a tree) is a finite maximal antichain in $\{1,2\}^*$ with respect to the prefix ordering $\leq$. For two cuts X, Y we let $Y > X$ if Y lies below X, i.e., $(\forall y \in Y)\,(\exists x \in X)\, y > x$. It is easy to see that a run ρ of a Büchi automaton is accepting if, for each cut X, there is a cut $Y > X$, labeled by the accepting states (i.e., $(\forall y \in Y)\, rank(\rho(y)) = 2$). We inductively define a sequence of tree languages K^n_q, for each state q of A, and $n \geq 0$.

The set K^0_q consists of all trees t which admit some run (not necessarily accepting) of A starting from q (q-run, for short). The set K^{n+1}_q comprises those trees t, which admit a q-run ρ, such that, for each cut X, there exists a cut $X' > X$, and a run ρ', with the following properties:

- ρ' agrees with ρ until the cut X,
- all states in $\rho'(X')$ are accepting,
- $(\forall v \in X')$, the subtree of t rooted in v (in symbols $t.v$) belongs to K^n_p, where $p = \rho'(v)$.

It follows by induction on n that $T(A) \subseteq K^n_{q_I}$, where q_I is the initial state of A. Now let n_A and n_B be the numbers of states of A and B, respectively, and let $M = 2^{n_A \cdot n_B} + 1$. We claim that $K^M_{q_I}$ separates $T(A)$ and $T(B)$. We already know that $T(A) \subseteq K^M_{q_I}$. For the sake of contradiction, suppose that $t \in K^M_{q_I} \cap T(B)$, and let ρ' be an accepting run of B on t.

Using the inductive definition of $K^M_{q_I}$, we can construct a sequence of cuts $X_1 < X'_1 < \ldots < X_M < X'_M$, and a run ρ of A on t, such that

- $(\forall i \leq M)$ all states in $\rho(X_i)$ are accepting,
- $(\forall i \leq M, \forall v \in X_i)\, t.v \in K^{M-i}_{\rho(v)}$,
- $(\forall i \leq M)$ all states in $\rho'(X'_i)$ are accepting.

By the choice of M, there exist $1 \leq k < \ell \leq M$, such that

$$\{(\rho(u), \rho'(u)) : u \in X_k\} = \{(\rho(v), \rho'(v)) : v \in X_\ell\}$$

Note that, by construction,

$$X_k < X'_k < X_l$$

with all states in $\rho'(X'_k)$ accepting. Hence, by a standard tree-pumping argument, we can construct a new tree along with two accepting runs: by A and by B, contradicting $T(A) \cap T(B) = \emptyset$.

It remains to show that the language $K^M_{q_I}$ is both Büchi and co-Büchi. A direct construction of two Büchi automata would be somewhat cumbersome, but one can use here any of the characterizations of this intersection class mentioned above. In the proof given in [3], it is shown that the sets K^n_q are definable in the alternation-free μ-calculus. A reader familiar with monadic second-order logic can easily see that these languages are definable in its *weak* fragment, i.e., with quantifiers restricted to finite sets. This is enough as well, according to the characterization given by Rabin [15]. ∎

3. Broken analogy

A reader familiar with descriptive set theory may think of another inseparable pair of recognizable tree languages, induced by a classical example ([11], section 33.A). We will explain why it would not be useful for our purpose. Let us now consider non-labeled trees, i.e., subsets $T \subseteq \omega^*$ closed under initial segments. They can be viewed as elements of the Cantor discontinuum $\{0,1\}^\omega$ by fixing a bijection $\iota : \omega \to \omega^*$ and identifying a tree T with its characteristic function, given by $f_T(n) = 1$ iff $\iota(n) \in T$. In particular, we can discuss topological properties of sets of such trees. As before, $P \in \omega^\omega$ is a path in a tree T if all finite prefixes of P are in T. Let

$$\begin{aligned} \mathrm{WF} &= \{T : T \text{ has no infinite path }\} \\ \mathrm{UB} &= \{T : T \text{ has } \textit{exactly one} \text{ infinite path }\} \end{aligned}$$

Both sets are known to be $\mathbf{\Pi}^1_1$-complete, although the membership of UB in $\mathbf{\Pi}^1_1$ is not obvious, and is the subject of one of Lusin's theorems (Theorem 18.11 in [11]). WF and UB are also known to be inseparable by Borel sets ([11], section 35, see also [6]). Now, it is not difficult to "encode" these sets as languages of labeled binary trees, which turn out to be recognizable by parity automata. In [14] a continuous reduction of WF to $M_{0,1}$ was used to show that the latter set is complete in $\mathbf{\Pi}^1_1$ (Example 1.1 above). Let

$$\begin{aligned} \mathrm{UB}_{bin} \;=\; \{t \in T_{\{0,1\}} : \;& \text{there is } \textbf{exactly one} \text{ path } P \\ & \text{with } \limsup_{n\to\infty} t(p_0 p_1 \ldots p_n) = 1\} \end{aligned}$$

It is easy to construct a non-deterministic automaton accepting this language; one can also assure that this automaton is non-ambiguous, i.e., for each accepted tree, has exactly one accepting run. From considerations above, one can deduce that the sets $T_{0,1}$ and UB_{bin} are inseparable by Borel languages. However, the language UB_{bin} is not co-Büchi.

Proposition 3.1. *The language* $\overline{\mathrm{UB}_{bin}}$ *is recognizable and analytic, but not Büchi.*

Proof. Let us call a path with infinitely many 1's *bad*. So the above language consists of trees that have either none or at least two bad paths. Rabin [15] shows that the language $T_{0,1}$ (no bad paths) cannot be recognized by a Büchi automaton, by constructing a correct tree which by pumping argument can be transformed to a tree with *exactly one* bad path (mistakingly accepted by the hypothetical automaton). So this classical argument applies to the language UB_{bin} without any changes. ∎

As we have argued in the introduction, this example somehow breaks the analogy between the class of Büchi recognizable tree languages and that of analytic sets. It turns out that the topological complexity, and the automata-theoretic complexity, although closely related, do not always coincide.

4. Conclusion

The automata-theoretic hierarchies, in particular the index hierarchies for non-deterministic and alternating tree automata, are studied because of the issues of expressibility and complexity. Typically, the higher the level in the hierarchy, the higher the expressive power of automata, but also the complexity of the related algorithmic problems (like emptiness or inclusion). Once the strictness of the hierarchy is established [7, 8], the next important problem is an *effective simplification*, i.e., determining the exact level of an object (e.g., a tree language) in the hierarchy. The problem is generally unsolved (see [9] for a recent development in this direction). One may expect that a better understanding of structural properties of the hierarchy can bring a progress also in this problem. We believe that ideas coming from descriptive set theory, like separation and reduction properties, uniformization, or completeness, can be helpful here.

The inseparable pairs of co-analytic sets are common in mathematics. Natural examples include the set of all continuous real–valued functions on the unit interval $[0,1]$ which are everywhere differentiable together with the set of all continuous real–valued functions on the unit interval $[0,1]$ which are not differentiable in exactly one point, but as in this case, other examples usually reflect the same pattern of WF *vs.* UB (c.f. [6]). In contrast, our pair presented in Section 2 is very symmetric: the two sets are copies of each other up to a symbolic renaming. Recently, Saint Raymond [16] established that the pair WF *vs.* UB is complete (in the sense of Wadge) with respect to all coanalytic pairs in the Cantor set. In the proof he uses an interesting example of another complete coanalytic pair, which exhibits certain symmetric properties. Building on his results, in subsequent work, we show that the pair $W_{0,1}$, $W'_{0,1}$, has an analogous completeness property.

Our example shows that the first separation property fails for the co-Büchi class (Σ_2^μ in the fixed-point hierarchy) while, by Rabin results [15], it holds for the Büchi class (Π_2^μ). By this we have also settled a missing case in a classification by Santocanale and Arnold [17]. However, these authors were interested in the *relative* separation property (as explained in our introduction), as they primarily wanted to find if the *ambiguous class* $\Pi_n^\mu \cap \Sigma_n^\mu$ can be effectively captured by $Comp(\Pi_{n-1}^\mu \cup \Sigma_{n-1}^\mu)$. As this coincidence turned out to fail for $n \geq 3$, it is meaningful to ask if the status of the first separation property established for the Büchi/co-Büchi classes, continues to hold for the higher-level classes Π_n^μ/Σ_n^μ. That is, if two disjoint sets definable in Π_n^μ can always be separated by a set in $\Pi_n^\mu \cap \Sigma_n^\mu$. (A similar question for Σ_n^μ, with expected answer negative.) In our opinion, it is an interesting problem, which may challenge for a better understanding of the topological structure of recognizable languages above $\mathbf{\Pi}_1^1 \cup \mathbf{\Sigma}_1^1$.

References

[1] J. W. Addison. Tarski's theory of definability: common themes in descriptive set theory, recursive function theory, classical pure logic, and finite-universe logic. *Annals of Pure and Applied Logic*, 126 (2004), 77–92.

[2] A. Arnold. The μ-calculus alternation-depth hierarchy is strict on binary trees. *RAIRO-Theoretical Informatics and Applications*, 33 (1999), 329–339.

[3] A. Arnold and D. Niwiński. Fixed point characterization of weak monadic logic definable sets of trees. In M.Nivat, A.Podelski, editors, *Tree Automata and Languages*, Elsevier, 1992, 159-188.

[4] A. Arnold and D. Niwiński. Rudiments of μ-Calculus. Elsevier Science, Studies in Logic and the Foundations of Mathematics, 146, North–Holland, Amsterdam, 2001.

[5] A. Arnold and D. Niwiński. Fixed point characterization of Büchi automata on infinite trees. J. Inf. Process. Cybern. EIK, 26, 1990, 453–461.

[6] H. Becker. Some Examples of Borel Inseparable Pairs of Co-analytic Sets. Mathematika 33, 1986, 72–79.

[7] J. C. Bradfield. The modal mu-calculus alternation hierarchy is strict. *Theoret. Comput. Sci.*, 195 (1997), 133–153.

[8] J. C. Bradfield. Simplifying the modal mu-calculus alternation hierarchy. In: *Proc. STACS'98*, Lect. Notes Comput. Sci. 1373 (1998), 39–49.

[9] T. Colcombet and C. Löding. The Non-deterministic Mostowski Hierarchy and Distance-Parity Automata. *Proc. ICALP 2008*, Lect. Notes Comput. Sci. 5126 (2008), 398-409.

[10] E. A. Emerson and C. S. Jutla. Tree automata, mu-calculus and determinacy. In: *Proceedings 32th Annual IEEE Symp. on Foundations of Comput. Sci.* (1991), 368–377.

[11] A. S. Kechris. Classical descriptive set theory. Springer-Verlag, New York, 1995.

[12] D. E. Muller, A. Saoudi, and P. E. Schupp. Alternating Automata, the Weak Monadic Theory of Trees and its Complexity. *Theoret. Comput. Sci.* 97(2), (1992), 233-244.

[13] D. Niwiński. On fixed point clones. In: ICALP'86, Lect. Notes Comput. Sci. 226, Springer-Verlag, 1986, 464–473.

[14] D. Niwiński and I. Walukiewicz. A gap property of deterministic tree languages. *Theoret. Comput. Sci.* 303 (2003), 215–231.

[15] M. O. Rabin. Weakly definable relations and special automata. In: *Mathematical Logic and Foundations of Set Theory*, Y. Bar-Hillel ed., 1970, 1-23.

[16] J. Saint Raymond. Complete pairs of coanalytic sets. *Fundamenta Mathematicae* 194 (2007), 267–281.

[17] L. Santocanale and A. Arnold. Ambiguous classes in μ-calculi hierarchies. *Theoret. Comput. Sci.* 333 (2005), 265-296.

[18] W. Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 3, Springer-Verlag, 1997, pp. 389–455.