

# On the structure of the monadic logic of the binary tree

D. Janin<sup>1</sup>, G. Lenzi<sup>1,2</sup>

LaBRI - Université de Bordeaux I  
351, cours de la libération  
F 33 405 Talence, Cedex

`janin,lenzi@labri.u-bordeaux.fr`

**Abstract.** Since the work of Rabin [9], it has been known that any monadic second order property of the (labeled) binary tree with successor functions (and not the prefix ordering) is a monadic  $\Delta_3$  property.

In this paper, we show this upper bound is optimal in the sense that there is a monadic  $\Sigma_2$  formula, stating the existence of a path where a given predicate holds infinitely often, which is not equivalent to any monadic  $\Pi_2$  formula. We even show that some monadic second order definable properties of the binary tree are not definable by any boolean combination of monadic  $\Sigma_2$  and  $\Pi_2$  formulas.

These results rely in particular on applications of Ehrenfeucht-Fraïssé like game techniques to the case of monadic  $\Sigma_2$  formulas.

## 1 Introduction

In this paper we are interested in a problem of *descriptive complexity*, an important and rapidly growing research area in theoretical computer science. Descriptive complexity was proposed by Fagin in the seventies (see [3]) as an approach to fundamental problems of complexity theory such as whether  $NP$  equals  $co-NP$ . While ordinary computational complexity theory is concerned with the amount of resources (such as time or space) necessary to solve a given problem, the idea of descriptive complexity is to study the expressibility of problems in some fixed logical formalism. For instance, in his seminal paper of 1974, Fagin shows that  $NP$  problems coincide (over finite structures) with the problems expressible in existential second order logic. Since then, there has been a large number of results in descriptive complexity. We note that most of these results concern finite structures (which are those interesting for the applications in computational complexity theory), but studying descriptive complexity also over infinite structures makes sense and may lead to a better comprehension of the expressiveness of various logical systems.

---

<sup>1</sup> supported by a TMR-GETGRATS grant of the University of Bordeaux I,

<sup>2</sup> supported by a TMR-Marie Curie grant of the EU

Here, we are concerned with the structure of the monadic second order logic (MSOL) on the infinite binary tree. It is a well known result of Rabin [9] that any monadic second order formula (MS-formula) is equivalent, on the binary tree, to a monadic  $\Sigma_2$  formula, that is a formula of the form

$$\exists X_1, \dots, X_m \forall Y_1, \dots, Y_n \psi$$

where  $\psi$  is a first order formula. This result comes from the translation of any MS-formula into tree-automata. It is also well-known [9] that there are formulas which are not equivalent to any monadic  $\Sigma_1$  formula (of the form  $\exists X_1, \dots, X_m \psi$ ). Since many logics of programs [2, 12] can be translated into MS-formulas on the binary tree, this difference of expressive power between monadic  $\Sigma_1$  and  $\Sigma_2$  allows a classification of these logics.

However, this upper bound obtained by Rabin works in the presence of a particular relation symbol, in the vocabulary of the logic considered, whose interpretation is the prefix ordering over nodes of the binary tree. If instead of this ordering we consider the two successor functions over nodes of the binary tree, Rabin's approach shows that every MS-formula is equivalent a monadic  $\Sigma_3$  formula, that is a formula of the form

$$\exists X_1, \dots, X_m \forall Y_1, \dots, Y_n \exists Z_1, \dots, Z_p \psi$$

where  $\psi$  is first order.

In this paper, we show that this upper bound is tight in the sense that, over the binary tree with successor functions instead of the prefix ordering, there are MS-formulas which are not equivalent to any monadic  $\Sigma_2$  formula. We even show that, again on the binary tree with successor functions, there are MS-formulas which are not equivalent to any Boolean combination of monadic  $\Sigma_2$  formulas. Notice that in many logics of programs the prefix ordering is (in some sense) definable (for instance with Kozen's propositional mu-calculus [7]) so it makes sense to be interested in the monadic second order logic of the binary tree without this ordering predefined.

A related result is established by Rabin [10]. He investigates there the expressive power of some restricted notion of tree-automata : the special or Büchi automata. He shows that there are MS-formulas which cannot be translated into this particular kind of tree automata. This result is extended by Hafer [5] who shows that there are MS-formulas that cannot be translated into any Boolean combination of Büchi automata.

Our result is possibly a consequence of Rabin's and Hafer's results provided monadic  $\Sigma_2$  formulas characterize exactly Buchi tree-automata. It is well known that Buchi tree-automata can be translated into monadic  $\Sigma_2$  formulas over the binary tree with successors functions. The converse is however an open question <sup>1</sup>.

---

<sup>1</sup> more recent works, yet unpublished, suggest unexpectedly that this converse may also be true...

## 2 Definitions and notations

Let  $T = \{0,1\}^*$  be the binary tree, i.e. the set of finite words over the alphabet  $\{0,1\}$ , equipped with (immediate) successor functions  $r_a : T \rightarrow T$  defined by  $r_a(w) = w.a$  for  $a = 0$  or  $1$ .

The set of formulas of MSOL we consider in the sequel is built from the function symbols  $r_0$  and  $r_1$  to be interpreted by the corresponding successor functions, equality predicate, Boolean connectives, existential and universal quantifications of first order variables  $x, y, \dots$ , and existential and universal quantifications of (monadic) second order variables  $X, Y, \dots$ .

We call monadic  $\Sigma_0$  and monadic  $\Pi_0$  the set of MS-formulas without set quantifiers and, for any integer  $n$ , we call monadic  $\Sigma_{n+1}$  (resp. monadic  $\Pi_{n+1}$ ) the set of MS-formulas formed by a sequence of existential (resp. universal) set quantifiers followed by a formula of monadic  $\Pi_n$  (resp. monadic  $\Sigma_n$ ).

Any MS-formula  $\varphi(x_1, \dots, x_m, X_1, \dots, X_n)$ , with free first order variables among  $\{x_1, \dots, x_m\}$  and with free second order variables among  $\{X_1, \dots, X_n\}$ , defines over the binary tree the set of tuples  $(t_1, \dots, t_m, R_1, \dots, R_n) \in T^m \times \mathcal{P}(T)^n$  that satisfy the formula  $\varphi$ . This is denoted by

$$(t_1, \dots, t_m, R_1, \dots, R_n) \models \varphi$$

(see [11], for instance, for a precise definition of the satisfaction relation).

Notice that any tuple  $(R_1, \dots, R_n) \in \mathcal{P}(T)^n$  corresponds to the function  $v$  from  $T$  to  $\mathcal{P}(\{X_1, \dots, X_n\})$  defined by  $v(t) = \{X_i : t \in R_i\}$ . In the sequel, such a function is called a  $\{X_1, \dots, X_n\}$ -colored tree (or  $\{X_1, \dots, X_n\}$ -tree for short).

Given  $R \subseteq T$  a set of nodes of the binary tree, we use the notation  $v|_R$  that stands for the restriction of  $v$  to  $R$ . Given  $v_1 : R_1 \rightarrow \mathcal{P}(C_1)$  and resp.  $v_2 : R_2 \rightarrow \mathcal{P}(C_2)$  two partially  $C_1$ -colored resp.  $C_2$ -colored trees, we say  $v_1$  and  $v_2$  are compatible when they agree on  $R_1 \cap R_2$ , i.e. for any  $t \in R_1 \cap R_2$ ,  $v_1(t) \cap C_1 = v_2(t) \cap C_2$ .

Given a partial  $C_1$ -tree  $v_1$  and a partial  $C_2$ -tree  $v_2$  compatible one with the other, we denote by  $(v_1, v_2)$  the partial  $C_1 \cup C_2$ -colored tree (also denoted by  $C_1, C_2$ -colored tree) given by the union of  $v_1$  and  $v_2$ .

## 3 An Ehrenfeucht-Fraïssé like game

Checking whether a colored binary tree satisfies some MS-formula can be seen as a game between two players : the first one trying to show the binary tree does satisfy the property, the second one trying to show the converse.

In the first order case, these game considerations lead to the standard notion of Ehrenfeucht-Fraïssé game (EF-game), a useful tool to show that a given property is not definable in first order logic (see for instance [1]).

Here, we define a second order version of this EF-game in order to handle monadic  $\Sigma_2$  formulas.

In order to investigate definability by monadic  $\Sigma_2$  formulas, we extend EF-games to the second order case as follows. This definition can be seen as a

generalization of the second order game (also called the Ajtai-Fagin game) introduced to handle monadic NP definability (see [4] for a discussion on these monadic NP games).

**Definition 1 (monadic  $\Sigma_2$  games).** *Given three integers  $c, d$  and  $e$ , given  $V_1$  and  $V_2$  two nonempty disjoint sets of  $B$ -colored trees, denoting by  $C$  and  $D$  two disjoint sets of set variables not in  $B$ , with  $c = |C|$  and  $d = |D|$ , we define the game  $EF_2(V_1, V_2, c, d, e)$  as a play in four (second order) rounds between two players (called Spoiler and Duplicator) as follows :*

1. for each  $v \in V_1$ , Spoiler chooses a  $C$ -coloring  $C_1(v)$ ;
2. Duplicator chooses one  $v_2 \in V_2$  and a  $C$ -coloring  $C_2(v_2)$ ;
3. Spoiler chooses a  $D$ -coloring  $D_2(v_2)$ ;
4. Duplicator chooses one  $v_1 \in V_1$  and a  $D$ -coloring  $D_1(v_1)$ .

We say Duplicator wins the game when the two  $B, C, D$ -colored trees

$$(v_1, C_1(v_1), D_1(v_1)) \text{ and } (v_2, C_2(v_2), D_2(v_2))$$

satisfy the same first order formulas of quantifier depth  $e$ . Otherwise, Spoiler wins.

**Theorem 1.** *For any sets  $V_1$  and  $V_2$  of  $B$ -colored trees, and for any integers  $c, d, e$ , if Duplicator has a winning strategy for the game  $EF_2(V_1, V_2, c, d, e)$  then there exists no  $\Sigma_2$  formula  $\varphi$  with free set variables among  $B$  and of the form :*

$$\exists X_1 \cdots X_c \forall Y_1 \cdots Y_d \psi$$

with  $\psi$  a first order sentence of quantifier depth  $e$  such that any  $v_1 \in V_1$  satisfies  $\varphi$  while no  $v_2 \in V_2$  satisfies  $\varphi$ .

*Proof.* A classical consequence of the standard definition of the satisfaction relation. □

In the sequel, in order to handle the winning condition given in the definition of  $EF_2$ -game above, we use the notion of  $r$ -type and  $r, m$ -equivalence and apply (the following non game theoretic version of) Hanf's theorem [6].

**Definition 2.** *Given integer  $r$  and node  $t$ , let  $S(t, r)$  be the set of nodes  $t'$  such that there is an undirected path between  $t$  and  $t'$  of length at most  $r$ . The  $r$ -type of some node  $t \in T$  in a  $C$ -colored tree  $v$  is the graph isomorphism class of  $v|S(t, r)$ . Given integer  $m$ , we say two  $C$ -colored trees  $v_1$  and  $v_2$  are  $r, m$ -equivalent if, for any  $r$ -type  $\tau$ , the number of nodes of  $r$ -type  $\tau$  in  $v_1$  equals the number of nodes of  $r$ -type  $\tau$  in  $v_2$ , or both these numbers are greater than  $m$ .*

**Theorem 2 (Hanf [6]).** *For any  $e$ , there exist  $r$  and  $m$  such that if two  $C$ -colored trees  $v_1$  and  $v_2$  are  $r, m$ -equivalent then they satisfy the same first-order formula of quantifier depth at most  $e$ .*

## 4 Monadic $\Sigma_2$ vs. monadic $\Pi_2$

In this section, we define the define a property, called the Büchi property, and show that it is not definable by a monadic  $\Pi_2$  formula (hence its complement is not definable by a monadic  $\Sigma_2$  formula). The same property is used by Rabin in [10] as a witness of properties definable by means of Büchi automata and whose complement is not definable by means of Büchi automata.

**Definition 3 (the Büchi property).** *Let  $X$  be a set variable. Given a  $\{X\}$ -colored tree  $v$ , we say  $v$  has the Büchi property  $B(X)$  if there is a directed path in  $T$  with infinitely many nodes in the interpretation of  $X$  in  $v$ .*

Equivalently, denoting by  $Acc(t, u)$  the property stating that there is a directed path from  $t$  to  $u$ , a  $\{X\}$ -colored tree satisfies the Büchi property if and only if the greatest solution of the set equation  $Y = F_X(Y)$  with

$$F_X(Y) = \{t \in T : \exists u \in Y, X \in v(u) \wedge t \neq u \wedge Acc(t, u)\}$$

is non empty. Although not very common, such a definition leads to the definition of the Büchi property by transfinite induction that we shall use later for our main result.

**Definition 4.** *For any countable ordinal  $\alpha$ , let us define  $F_X^\alpha(T)$  as the set  $T$  when  $\alpha = 0$ , the set  $F_X(F_X^{\alpha_1}(T))$  when  $\alpha = \alpha_1 + 1$  and the set  $\bigcap_{\alpha_1 < \alpha} F_X^{\alpha_1}(T)$  when  $\alpha$  is a limit ordinal. Let then  $B_\alpha(X)$  be the property of  $\{X\}$ -colored trees stating that  $F_X^\alpha(T)$  is non empty.*

**Proposition 1.** *(1) For any countable ordinal  $\alpha$  there are  $\{X\}$ -colored trees that do satisfy  $B_\alpha(X)$  while they do not satisfy  $B(X)$ . (2) A  $\{X\}$ -colored tree satisfies the property  $B(X)$  if and only if, for any countable ordinal  $\alpha$ , it does satisfy  $B_\alpha(X)$ .*

*Proof.* Fact (2) follows from the fact that the binary tree is countable. To prove fact (1), let us define, for each countable ordinal  $\alpha$ , the  $\{X\}$ -tree  $v_\alpha$  by the following induction. When  $\alpha = 0$ ,  $v_\alpha(t) = \emptyset$  for each  $t \in T$ . When  $\alpha = \alpha_1 + 1$ ,  $v_\alpha(\epsilon) = \{X\}$ , and for any  $t \in T$ ,  $v_\alpha(0.t) = \emptyset$  and  $v_\alpha(1.t) = v_{\alpha_1}(t)$ . When  $\alpha$  is a countable limit ordinal, given an arbitrary enumeration  $\alpha_0, \alpha_1, \dots$ , of all ordinals strictly smaller than  $\alpha$ ,  $v_\alpha(\epsilon) = \{X\}$ , and for any  $t \in T$ , any  $n \in \mathbb{N}$ ,  $v_\alpha(0^n.1.t) = v_{\alpha_n}(t)$ . None of these  $v_\alpha$  satisfies the Büchi property while, for each ordinal  $\alpha$ ,  $v_{\alpha+1}$  does satisfy  $B_\alpha(X)$ . □

Before stating and proving the main theorem below, we give some more technicalities in order to deal with  $EF_2$ -games. These definitions are inspired by similar techniques used by the second author in [8].

**Definition 5 ( $r, m$ -type constraint).** *An  $r, m$ -type constraint is a function  $f$  which associates to each  $r$ -type  $\tau$  of its domain a number  $f(\tau) < m$ . We say  $v|R$  is compatible with type constraint  $f$  when, for any  $r$ -type  $\tau$  in the domain of  $f$ ,  $v|R$  has at most  $f(\tau)$  nodes of type  $\tau$ .*

**Definition 6 (*D*-closure of a *C*-colored finite region).** Given integers  $r$  and  $m$ , given a *C*-colored tree  $v$ , given a set  $D$  of variables disjoint from  $C$ , given finite regions  $R \subseteq R_1 \subseteq T$ , we say  $R_1$  is a *D*-closure of  $R$  (in  $v$ ) if, for any  $r, m$ -type constraint  $f$  and for every *D*-coloring  $v_0$  of  $R$ , if there is a *D*-coloring  $v_1$  of  $R_1$  which equals  $v_0$  on  $R$  and such that the *C, D*-coloring  $(v, v_1)$  of  $R_1$  is compatible with  $f$ , then there is a similar *D*-coloring of the complete binary tree.

We want to show that indeed *D*-closures exists; to this goal we begin with a lemma:

**Lemma 1.** Given  $R \subseteq T$ ,  $v$  a *C*-colored tree and  $v_1$  a *D*-coloring of  $R$ . Given  $f$  an  $r, m$ -type constraint. If every finite subset of the binary tree has a *D*-coloring which equals  $v_1$  on  $R$  and, together with  $v$ , is compatible with  $f$ , then the entire binary tree has a similar coloring.

*Proof.* By a classical compactness argument.

**Proposition 2.** For any integers  $r$  and  $m$ , for any *C*-colored tree  $v$ , for any finite region  $R \subseteq T$  colored by  $v$ , and for any set  $D$  of fresh variables, there exists a *D*-closure of  $R$ .

*Proof.* Let  $v$ ,  $R$  and  $D$  be as above. For each  $r, m$ -type constraint  $f$ , and for each *D*-coloring  $w$  of  $R$ , applying the previous Lemma, either there exists a *D*-coloring of the entire binary tree which, together with  $v$ , is compatible with the constraint  $f$  and extends  $w$ , or there exists a finite region  $R_{f,w}$  with no such *D*-coloring. One can check that the union of all these  $R_{f,w}$  is a *D*-closure of  $R$ .  $\square$

We can now state and prove our main theorem.

**Theorem 3.** Over  $\{X\}$ -colored binary trees with successor functions, the Büchi property  $B(X)$  is not definable by any monadic  $\Pi_2$  formula.

*Proof.* Assume  $B(X)$  is definable by a  $\Pi_2$  formula. This means  $\bar{B}(X)$ , the complement of  $B(X)$ , is definable by a  $\Sigma_2$  formula; henceforth, applying Theorem 1, Spoiler has a winning strategy on the game  $EF_2(\bar{B}(X), B(X), c, d, e)$  for some integer  $c, d$  and  $e$ . We show below that Duplicator has a winning strategy for this game hence this assumption is absurd.

Let  $C$  and  $D$  be disjoint sets of variables distinct from  $X$  with  $|C| = c$  and  $|D| = d$ , and let  $r$  and  $m$  be given by Hanf's theorem (Theorem 2) for classical EF-games in  $e$  rounds over  $\{X\}, C, D$ -colored binary trees. We want to show that the second order rounds can be played (by the Duplicator) so that the resulting pair of  $\{X\}, C, D$ -trees chosen at the end (by the Duplicator) are  $r, m$ -equivalent.

- (1) First, for each  $v \in \bar{B}(X)$ , Spoiler chooses a *C*-colored tree  $C_1(v)$ .
- (2) The duplicator's answer is the choice of an  $\{X\}$ -colored tree  $v_2 \in B(X)$  together with the choice of the *C*-colored tree  $C_2(v_2)$ . This answer is made using the following lemma.

**Lemma 2.** *There exists a sequence  $\{(R_n, w_n, W_n, t_n)\}_{n \in \mathbb{N}}$  where  $R_n \subseteq T$  is a finite subset of the binary tree,  $W_n$  is a set of  $\{X\}$ ,  $C$ -colored trees of the form  $(v, C_1(v))$  with  $v \in \bar{B}(X)$  and  $t_n \in T$ , and such that  $T$  equals the union of all  $R_n$  and for each  $n$  :*

1.  $w_n : R_n \rightarrow \mathcal{P}(C \cup \{X\})$  is a  $\{X\}$ ,  $C$ -coloring of  $R_n$ ,
2. all  $\{X\}$ ,  $C$ -colored trees of  $W_n$  are compatible with  $w_n$  and color  $t_n$  by  $X$ , and  $W_{n+1} \subseteq W_n$ ,
3.  $R_{n+1}$  is a  $D$ -closure of  $R_n$  in every element of  $W_{n+1}$ ,
4.  $t_n$  is a proper prefix of  $t_{n+1}$  and belongs to  $R_{n+1}$ ,
5. for each countable ordinal  $\alpha$ , there exists  $w \in W_n$  such that the binary tree rooted at  $t_n$  and  $\{X\}$ ,  $C$ -colored by  $w$  satisfies  $B_\alpha(X)$ .

*Proof.* By induction on  $n$ . Let us take  $R_0 = \emptyset$ ,  $W_0$  the set of  $\{X\}$ ,  $C$ -trees  $(v, C_1(v))$  for each  $v \in \bar{B}(X)$  such that  $X \in v(\epsilon)$  for  $\epsilon$  the root of the binary tree and  $t_0 = \epsilon$ . Property 5 is ensured by Proposition 1.

Assume the construction has been made up to some integer  $n$ . For each  $t' \in T$  strictly above  $t_n$ , each finite  $R' \supseteq R$  (that also contains  $t_n$  and any node at a distance at most  $n$  from the root) and each  $\{X\}$ ,  $C$ -coloring  $w'$  of  $R'$ , let us define the set  $O(t', R', w')$  of all countable ordinals  $\alpha$  such that there exists  $(v, C_1(v)) \in W_n$  such that :

1.  $X \in v(t')$ ,
2.  $(v, C_1(v))|_{R'}$  and  $w'$  are equal,
3. the  $\{X\}$ -colored binary tree rooted at  $t'$  and colored by  $v$  satisfies  $B_\alpha(X)$ ,
4.  $R'$  is a  $D$ -closure of  $R_n$  w.r.t.  $(v, C_1(v))$ .

By the induction hypothesis, for each countable ordinal  $\alpha$  there exists  $w \in W_n$  such that the binary tree rooted at  $t_n$  colored by  $v$  has the  $B_{\alpha+1}(X)$  property. Also by the definition of  $B_{\alpha+1}$ , there exists some  $t'$  strictly above  $t_n$  as in item 3; hence we have

$$\bigcup_{(t', R', w')} O(t', R', w') = \aleph_1$$

Since there are  $\aleph_0$  possible values for  $t'$ ,  $R'$  and  $w'$ , one of these  $O(t', R', w')$  has size  $\aleph_1$  and we take in this case  $t_{n+1} = t'$ ,  $R_{n+1} = R'$  and  $w_{n+1} = w'$ . The set  $W_{n+1}$  is defined as the set of all  $w \in W_n$  such that  $w|_{R_{n+1}}$  equals  $w_{n+1}$ ,  $X \in w(t_{n+1})$  and  $R_{n+1}$  is a  $D$ -closure of  $R_n$  in  $w$ .

□

*Proof of Theorem 3 (continued).* Now, let us define the  $\{X\}$ ,  $C$ -tree  $(v_2, C_2(v_2))$  chosen by Duplicator as  $\bigcup_n w_n$ . Notice that  $v_2$  does satisfy the Büchi property since for each  $n$ ,  $X \in v_2(t_n)$ .

(3) Spoiler chooses a  $D$ -colored tree  $D_2(v_2)$ .

(4) Duplicator's answer is made as follows.

We note that there exists a finite subset  $R \subseteq T$  such that the  $\{X\}$ ,  $C$ ,  $D$ -colored tree  $(v_2, C_2(v_2), D_2(v_2))$  is  $r, m$ -equivalent to its restriction to  $R$ .

So let  $n$  be an integer such that  $R \subseteq R_{n-1}$  and let  $f$  be the minimal (w.r.t. the usual order on partial functions)  $r, m$ -type constraint satisfied by the  $\{X\}, C, D$ -colored tree above.

Duplicator chooses some (any)  $\{X\}, C$ -colored tree  $w \in W_n$  with  $w = (v_1, C_1(v_1))$  (henceforth  $(v_1|R_n)$  equals  $(v_2|R_n)$ ). An adequate  $D$ -coloring  $D_1(v_1)$  can be chosen as follows.

By the definition of  $D$ -closure, since  $R_n$  is the  $D$ -closure of  $R_{n-1}$  with respect to the  $\{X\}, C$ -coloring  $(v_1, C_1(v_1))$ , there exists a  $D$ -coloring  $D_1(v_1)$  of the binary tree that equals  $D_2(v_2)$  restricted to  $R_n$  and such that  $(v_1, C_1(v_1), D_1(v_1))$  is compatible with  $f$ .

Since  $(v_1, C_1(v_1), D_1(v_1))$  and  $(v_2, C_2(v_2), D_2(v_2))$  restricted to  $R, R_{n-1}$  or  $R_n$  are  $r, m$ -equivalent and since  $f$  has been chosen minimal w.r.t. the  $r$ -types of nodes of these partially colored trees, the full  $\{X\}, C, D$ -colored trees

$$(v_1, C_1(v_1), D_1(v_1)) \text{ and } (v_2, C_1(v_2), D_1(v_2))$$

are  $r, m$ -equivalent.

From this point, Hanf's theorem applies showing that Duplicator wins.  $\square$

## 5 The case of the boolean closure of $\Sigma_2$ and $\Pi_2$

In order to show that some  $MS$ -definable formulas are not definable by any Boolean combination of  $\Pi_2$  and  $\Sigma_2$  formulas, let us first prove a normal form theorem for these boolean combinations.

**Lemma 3.** *Any Boolean combination of  $\Sigma_2$  and  $\Pi_2$  formulas is equivalent to a formula of the form  $\pi_1 \wedge (\sigma_1 \vee (\pi_2 \wedge (\sigma_2 \vee \dots \vee (\pi_n \wedge \sigma_n) \dots)))$  with  $\pi_i$ s monadic  $\Pi_2$  formulas and  $\sigma_i$ s monadic  $\Sigma_2$  formulas.*

*Proof.* Since both  $\Sigma_2$  and  $\Pi_2$  are closed (up to equivalence) under conjunction and disjunction, Boolean combinations of  $\Sigma_2$  and  $\Pi_2$  formulas are equivalent to finite conjunctions of formulas of the form  $\pi \vee \sigma$  for  $\sigma \in \Sigma_2$  and  $\pi \in \Pi_2$ . It follows that it is sufficient to show that the set of formulas of the desired form is closed (up to equivalence) under conjunction with formulas of the form  $(\pi \vee \sigma)$ . We show this by induction over  $n$ .

For  $n = 0$  nothing has to be proved. Assume it is true up to rank  $n$  and let  $\varphi_{n+1} = \pi_1 \wedge (\sigma_1 \vee \varphi_n)$  be a formula of rank  $n + 1$  with  $\varphi_n$  a formula of rank  $n$ .

First, we may assume that  $\models (\neg\sigma_1) \Rightarrow \pi_1$  and  $\models \varphi_n \Rightarrow (\neg\sigma_1)$ . Otherwise, we can take  $\sigma_1 \vee \neg(\pi_1)$  instead of  $\sigma_1$  and  $\varphi_n \wedge \pi_1 \wedge \neg\sigma_1$  instead of  $\varphi_n$  obtaining thus an equivalent formula with the same shape and the desired property.

Given then  $\pi \in \Pi_2$  and  $\sigma \in \Sigma_2$ , one can check the formula  $\pi_1 \wedge ((\sigma \wedge \sigma_1) \vee ((\pi \vee (\neg\sigma_1)) \wedge (\sigma_1 \vee (\varphi_n \wedge (\pi \vee \sigma))))$  is equivalent to  $\varphi_{n+1} \wedge (\pi \vee \sigma)$ . Applying the induction hypothesis over  $\varphi_n \wedge (\pi \vee \sigma)$  concludes the induction step.  $\square$

Let us then define the predicate  $P(X, x)$  as the one stating that the binary tree rooted in node  $x$  satisfies the Buchi property with respect to the set variable  $X$ .

**Proposition 3.** *For any node  $t \in T$ , the property  $P(X, t)$  is not definable by a monadic  $\Pi_2$  formula.*

*Proof.* By applying the argument of Theorem 3 to the binary tree rooted in  $t$ .  $\square$

Now let us fix two variables  $X$  and  $Y$  and let us define the predicate  $A(X, Y, z)$  as the least <sup>2</sup> monadic predicate equivalent to  $P(X, z.0) \vee (\bar{P}(Y, z.0) \wedge A(X, Y, z.1))$ , where  $\bar{P}(Y, z.0)$  denotes the negation of  $P(Y, z.0)$ .

The property  $A(X, Y, \epsilon)$  is definable in MSOL since it is definable in the mu-calculus (see [7]). However :

**Theorem 4.** *Property  $A(X, Y, \epsilon)$  is not definable by any boolean combination of  $\Sigma_2$  and  $\Pi_2$  formulas.*

*Proof.* Let us assume the converse and let us show this leads to a contradiction. Assume  $A(X, Y, \epsilon)$  is equivalent to a Boolean combination of  $\Pi_2$  and  $\Sigma_2$  formulas. Applying Lemma 3 assume such a formula is of the form  $\pi_1 \wedge (\sigma_1 \vee (\pi_2 \wedge (\sigma_2 \vee (\dots \pi_n \wedge (\sigma_n) \dots))))$ .

Applying the fixpoint definition of  $A(X, Y, z)$   $n$  times shows that  $A(X, Y, \epsilon)$  is equivalent to

$$P(X, 0) \vee (\bar{P}(Y, 0) \wedge (\dots P(X, 1^{n-1}.0) \vee (\bar{P}(Y, 1^{n-1}.0) \wedge A(X, Y, 1^n) \dots)))$$

denoting by  $a^k$ , as usual, the word composed of  $k$  times the letter  $a$ .

Let then  $X_1, \dots, X_n, X'$  (resp.  $Y_1, \dots, Y_n, Y'$ ) be some new fresh variables. There exists a monadic  $\Pi_1$  formula  $\varphi_X$  over the free variables  $X, X_1, \dots, X_n$  and  $X'$  that checks whether  $X$  equals the disjoint union of the intersection of  $X'$  with the subtree rooted in  $1^n$  and, for each  $k \in [1, n]$ , the intersection of  $X_k$  with the subtree rooted in  $1^{k-1}.0$  (and similarly a formula  $\varphi_Y$  for  $Y$ ). It follows that there also exists, for each  $k \in [1, n]$ , a  $\Pi_2$  formula  $\pi'_k$  (resp. the  $\Sigma_2$  formula  $\sigma'_k$ ) over the free variable  $X_1, \dots, X_n, X'$  and  $Y_1, \dots, Y_n, Y'$  that checks formula  $\pi_k$  (resp. formula  $\sigma_k$ ) does hold with  $X$  and  $Y$  implicitly defined by  $\varphi_X$  and  $\varphi_Y$ . From these definitions it follows that the new formula  $A'(X_1, \dots, X_n, X', Y_1, \dots, Y_n, Y', \epsilon)$  defined by

$$P(X_1, 0) \vee (\bar{P}(Y_1, 0) \wedge (\dots P(X_n, 1^{n-1}.0) \vee (\bar{P}(Y_n, 1^{n-1}.0) \wedge A(X', Y', 1^n) \dots)))$$

is equivalent to the formula  $\pi'_1 \wedge (\sigma'_1 \vee (\pi'_2 \wedge (\sigma'_2 \vee (\dots \pi'_n \wedge (\sigma'_n) \dots))))$ .

The next step is to show, by induction over  $k$ , that there exists a  $\{X_1, Y_1, \dots, X_k, Y_k\}$ -colored tree  $v_k$ , such that, for each  $i \in [1, k]$  :

$$v_k \models \forall X_{k+1}, Y_{k+1}, \dots, X_n, Y_n, X', Y', (\pi'_i \wedge \neg \sigma'_i)$$

while

$$v_k \models \neg P(X_i, 1^{i-1}.0) \wedge \bar{P}(Y_i, 1^{i-1}.0)$$

---

<sup>2</sup> or the greatest, it really does not matter in the following proof...

With  $k = n$  this leads to a contradiction since it says in particular that, for any  $i \in [1, n]$ ,  $v_n \models \forall X', Y', \varphi \Rightarrow \pi'_i$  from which it follows that

$$v_n \models \forall X', Y', \pi'_1 \wedge (\sigma'_1 \vee (\pi'_2 \wedge (\sigma'_2 \vee (\dots \pi'_n \wedge (\sigma'_n) \dots))))$$

hence, by the definition of the  $\sigma'_k$ s and  $\pi'_k$ s,

$$v_n \models \forall X', Y' A'(X_1, \dots, X_n, Y_1, \dots, Y_n, X', Y', \epsilon)$$

Also, for each  $i \in [1, n]$ ,  $v_n \models \neg P(X_i, 1^{i-1}.0) \wedge \bar{P}(Y_i, 1^{i-1}.0)$  hence,

$$v_n \models \forall X', Y' A(X', Y', 1^n)$$

This is obviously absurd since, given any  $\{X', Y'\}$ -tree  $v'$  that does not satisfy  $A(X', Y', 1^n)$  (and there are some) one has  $(v_n, v') \models \neg A(X', Y', 1^n)$  hence

$$v_n \models \exists X', Y' \neg A(X', Y', 1^n)$$

The proof by induction goes as follows. For  $k = 0$  nothing has to be proved. Assume the induction hypothesis for some  $k$ . In particular, the formulas

$$P(X_{k+1}, 1^k.0) \vee (\bar{P}(Y_{k+1}, 1^k.0) \wedge (\dots P(X_n, 1^{n-1}.0) \vee (\bar{P}(Y_n, 1^{n-1}.0) \wedge A(X', Y', 1^n) \dots)))$$

and  $\pi'_{k+1} \wedge (\sigma'_{k+1} \vee (\dots \pi'_n \wedge (\sigma'_n) \dots))$  are equivalent over any  $\{X_1, \dots, X_n, Y_1, \dots, Y_n, X', Y'\}$ -tree compatible with  $v_k$ .

It follows that

$$v_k \models \forall X_{k+1}, Y_{k+1}, \dots, X_n, Y_n, X', Y' (P(X_{k+1}, 1^k.0) \Rightarrow \pi'_{k+1})$$

That is

$$v_k \models \forall X_{k+1} (P(X_{k+1}, 1^k.0) \Rightarrow (\forall Y_{k+1}, \dots, X_n, Y_n, X', Y' \pi'_{k+1}))$$

Since  $P(X_{k+1}, 1^k.0)$  is not definable by a monadic  $\Pi_2$  formula, the above implication is strict in the sense that there exists a  $\{X_{k+1}\}$ -tree  $v'$  such that

$$(v_k, v') \models (\forall Y_{k+1}, \dots, X_n, Y_n, X', Y' \pi'_{k+1}) \wedge \neg P(X_{k+1}, 1^k.0)$$

In other words, for any colored tree  $w$  compatible with  $(v_k, v')$  one has

$$\bar{P}(Y_{k+1}, 1^k.0) \wedge (\dots P(X_n, 1^{n-1}.0) \vee (\bar{P}(Y_n, 1^{n-1}.0) \wedge A(X', Y, 1^n) \dots))$$

equivalent to  $\sigma'_{k+1} \vee (\dots \pi'_n \wedge (\sigma'_n) \dots)$ . Again, in particular,

$$(v_k, v') \models \forall Y_{k+1} \dots, X_n, Y_n, X', Y' (\sigma'_{k+1} \Rightarrow \bar{P}(Y_{k+1}, 1^k.0))$$

hence

$$(v_k, v') \models \forall Y_{k+1} (\neg \bar{P}(Y_{k+1}, 1^k.0) \Rightarrow (\forall X_{k+2}, Y_{k+2}, \dots, X_n, Y_n, X', Y' \neg \sigma'_{k+1}))$$

Now, since  $\neg \bar{P}(Y_{k+1}, 1^k.0)$  is not equivalent to a monadic  $\Pi_2$  formula, there exists a  $\{Y_k\}$ -tree  $v''$  such that

$$(v_k, v', v'') \models \bar{P}(Y_{k+1}, 1^k.0) \wedge (\forall X_{k+2}, Y_{k+2}, \dots, X_n, Y_n, X', Y' \neg \sigma'_{k+1})$$

Putting  $v_{k+1} = (v_k, v', v'')$  concludes the induction step.  $\square$

## 6 As a conclusion...

One may notice that all the material presented in last section can be generalized. The following exercise is one possible formulation of such a generalization.

*Exercise.* Let  $C$  be the class of (even only finite) directed graphs with a source such that all nodes are reachable from the source. For any integer  $n \geq 2$ , if there exists a monadic  $\Sigma_n$  definable property of graphs of  $C$  which is not monadic  $\Pi_n$  definable, then there also exists a monadic  $\Delta_{n+1}$  graph property (that is a property both definable by a monadic  $\Sigma_{n+1}$  formula and  $\Pi_{n+1}$  formula) which is not definable by a (finite) Boolean combination of monadic  $\Pi_n$  and  $\Sigma_n$  formula.

*Hint : think of directed comb-like constructions.*

## Acknowledgement

Thanks to Damian Niwiński, André Arnold and Mike Robson for their help and comment on earlier version of this communication.

## References

1. H.D. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical Logic*. Undergraduate Texts in Mathematics. Springer-verlag, 1984.
2. E. Emerson. Temporal and modal logic. In J. Van Leeuwen, editor, *Handbook of Theor. Comp. Science (vol. B)*, pages 995–1072. Elsevier, 1990.
3. R. Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In *Complexity of Computation*, volume 7. SIAM-AMS, 1974.
4. R. Fagin. Comparing the power of monadic NP games. *Mathematical Logic Quarterly*, 43(4):431–455, 1997.
5. T. Hafer. On the boolean closure of Büchi tree automaton definable sets of  $\omega$ -trees. Technical report, Aachener Infor.Ber. Nr. 87-16, RWTH Aachen, 1987.
6. W. Hanf. Model-theoretic methods in the study of elementary logic. In L. Henkin J. Addison and A. Tarski, editors, *The theory of models*, pages 132–145, Amsterdam, 1965. North Holland.
7. D. Kozen. Results on the propositional  $\mu$ -calculus. *Theoretical Comp. Science*, 27:333–354, 1983.
8. G. Lenzi. *The  $\mu$ -calculus and the Hierarchy Problem*. PhD thesis, Scuola Normale Superiore, Pisa, 1997.
9. M. O. Rabin. Decidability of second order theories and automata on infinite trees. *Trans. Amer. Math. Soc.*, 141, 1969.
10. M. O. Rabin. Weakly definable relations and special automata. In *Mathematical Logic and Foundation of Set Theory*. North Holland, 1970.
11. J. R. Shoenfield. *Mathematical Logic*. Addison-Wesley, Reading, Mass., 1967.
12. W. Thomas. Automata on infinite objects. In J. Van Leeuwen, editor, *Handbook of Theor. Comp. Science (vol. B)*, pages 133–191. Elsevier, 1990.