Fundamenta Informaticae XXI (2001) 1001–1017
IOS Press

Expressing Cardinality Quantifiers in Monadic Second-Order Logic over Trees

Vince Bárány
Oxford University Computing Laboratory
vbarany@comlab.ox.ac.uk

Łukasz Kaiser
RWTH Aachen, Mathematische Grundlagen der Informatik
kaiser@logic.rwth-aachen.de

Alex Rabinovich
Tel Aviv University, The Blavatnik School of Computer Science
rabinoa@post.tau.ac.il

Abstract. We study an extension of monadic second-order logic of order with the uncountability quantifier "there exist uncountably many sets". We prove that, over the class of finitely branching trees, this extension is equally expressive to plain monadic second-order logic of order.

Additionally we find that the continuum hypothesis holds for classes of sets definable in monadic second-order logic over finitely branching trees, which is notable for not all of these classes are analytic.

Our approach is based on Shelah's composition method and uses basic results from descriptive set theory. The elimination result is constructive, yielding a decision procedure for the extended logic.

Keywords: infinite trees; monadic second-order logic; cardinality quantifiers; Cantor topology

1. Introduction

Monadic second-order logic of order, MLO, extends first-order logic by allowing quantification over *subsets* of the domain. The binary relation symbol $<$ and unary predicate symbols P_i are its only non-logical relation symbols. MLO plays a very important role in mathematical logic and computer science.

Address for correspondence: Mathematische Grundlagen der Informatik RWTH Aachen, kaiser@logic.rwth-aachen.de

The fundamental connection between MLO and automata was discovered independently by Büchi, Elgot and Trakhtenbrot [7, 9, 23, 24] when the logic was proved to be decidable over the class of finite linear orders and over $(\omega, <)$. Moving away from linear orders, Rabin proved that the monadic second-order theory of the full binary tree, S2S for short, is decidable [17]. This celebrated theorem, obtained using the notion of tree automata, is often referred to as "the mother of all decidability results".

First-order cardinality quantifiers, studied by Mostowski and also by Magidor and Malitz in a topological setting, count the number of elements satisfying a given property inside a structure. Extensions of first-order logic with these quantifiers have been widely investigated over various natural classes of structures with respect to both decidability and the possibility of elimination. See for instance [2].

Second-order cardinality quantifiers in MLO, which we study in this paper, have been mostly considered in the context of automata and automatic structures [18]. The first observation of this nature, made in [5], was that the quantifier "there exist infinitely many words such that" can, in a certain sense, be eliminated on all automatic structures. More precisely, via the standard correspondence of automata with MLO, this amounts to eliminating the quantifier "there exist infinitely many (finite) sets such that" from *weak* MLO over $(\omega, <)$. The case of full MLO and the quantifier "there exist uncountably many sets such that" over $(\omega, <)$ corresponds to injectively presented ω-automatic structures and was solved in [13]. The structural properties of ω-regular languages identified in the latter work and its sequels have provided important insights into ω-automatic structures.

Motivated by previous work on $(\omega, <)$ that used word automata, we investigate second-order cardinality quantifiers over finitely branching trees, in particular, over the binary tree with arbitrary labelings, which corresponds to tree automata with additional parameters [8]. The parameterless question was previously studied by Niwiński, who in [15] proved that a regular language of infinite trees is uncountable if and only if it contains a non-regular tree.

We investigate over trees the expressive power of the extension of MLO by cardinality quantifiers $\exists^{\kappa} X$, with the interpretation "there exist at least κ many subsets X such that", for $\kappa \in \{\aleph_0, \aleph_1, 2^{\aleph_0}\}$. We denote this logic as $\mathrm{MLO}(\exists^{\aleph_0}, \exists^{\aleph_1}, \exists^{2^{\aleph_0}})$ and throughout the paper by *trees* we mean finitely-branching trees every branch of which is either finite or of order type ω. Our main results are summarized in the next two theorems.

Theorem 1.1. For every $\mathrm{MLO}(\exists^{\aleph_0}, \exists^{\aleph_1}, \exists^{2^{\aleph_0}})$ formula $\varphi(\overline{Y})$ there exists an MLO formula $\psi(\overline{Y})$, computable from φ, that is equivalent to $\varphi(\overline{Y})$ over trees.

In addition to the above, the reduction will show that over trees the quantifiers $\exists^{\aleph_1} X$ and $\exists^{2^{\aleph_0}} X$ are equivalent, i.e. that the continuum hypothesis holds for MLO-definable families of sets. Though not surprising, this is not obvious for it is known that in MLO one can define non-analytic classes of sets [16] and that CH is independent of ZFC already for co-analytic sets [14].

Theorem 1.2. On trees $\exists^{\aleph_1} X \varphi(X, \overline{Y})$ is equivalent to $\exists^{2^{\aleph_0}} X \varphi(X, \overline{Y})$ for every MLO formula $\varphi(X, \overline{Y})$.

Our results trivially extend to cardinality quantifiers $\exists^{\aleph_0} \overline{X}$, $\exists^{\aleph_1} \overline{X}$ and $\exists^{2^{\aleph_0}} \overline{X}$ counting (finite) tuples of sets using the simple fact that $\exists^{\kappa} (U, \overline{V})\ \varphi \ \equiv \ \exists^{\kappa} U \left(\exists \overline{V} \varphi\right) \ \vee \ \exists^{\kappa} \overline{V} \left(\exists U \varphi\right)$ for any cardinal $\kappa \geq \aleph_0$. Our theorems also supersede the previously mentioned results from [13] and generalize the theorem of Niwiński [15], which states that over the full binary tree the validity of $\exists^{\aleph_1} \overline{X}\, \varphi(\overline{X})$ is decidable and equivalent to that of $\exists^{2^{\aleph_0}} \overline{X}\, \varphi(\overline{X})$ for every MLO-formula $\varphi(\overline{X})$. Niwiński's theorem follows form the parameterless instances of our theorems. Certain structural insight gained from some of our intermediate lemmas might be of independent interest. More specifically we show that counting sets of nodes satisfying an MLO-formula on a tree can be effectively reduced to a combination of counting branches satisfying a certain MLO-formula, and counting chains with certain MLO-definable properties on individual branches. While the latter essentially amounts to dealing with the special case treated in [13], relying on basic results from descriptive set theory we show that counting of branches can also be formalized in MLO. An extended abstract of this paper was published in [3].

Organization

We begin by noting in Section 2 some observations regarding the second-order infinity quantifier $\exists^{\aleph_0} X$. In Section 3 we fix terminology and notation on trees and recollect some essentials of Shelah's composition method for MLO. The rest of the paper is devoted to the proof of Theorems 1.1 and 1.2.

In Section 4 we start by reducing the question of the existence of uncountably many sets X satisfying a given MLO formula $\varphi(X, \overline{Y})$ with parameters $\overline{Y}$ over a tree to a disjunction of three conditions: A, B and C. Condition A deals with MLO-properties of antichains; Condition C deals with a simpler version of the uncountability quantifier, namely with the quantifier "there exist uncountably many branches". Ultimately, condition B is concerned with the cardinality of chains with a specific MLO property on individual branches. It is postulated first in a broader form for deductive advantages.

In Section 5, we show that Condition B can be significantly weakened assuming that conditions A and C are not satisfied. Relying on the elimination results on $(\omega, <)$ from [13], we formalize this weakened form of Condition B in MLO and prove, that it guarantees the existence of continuum many sets satisfying φ.

In Section 6 we consider Condition C in the special case of the complete binary tree. The key theorem that we prove there, which might be of independent interest, is that MLO-definable sets of branches of the binary tree are Borel. This opens the way to formalizing Condition C in plain MLO, first over the binary tree and finally, in Section 7, over arbitrary trees.

The proofs of our main theorems are summarized in Section 8.

2. Infinity quantifier

With regard to the second-order infinity quantifier $\exists^{\aleph_0} X$ the following observations are worth making. While it clearly cannot be eliminated over all structures, it is easily expressible in monadic second-order logic (MSO) with the auxiliary predicate $\mathrm{Inf}(Z)$ asserting that the set Z is infinite, or equivalently, with the help of the first-order infinity quantifier $\exists^{\aleph_0} x$.

Proposition 2.1. For every $\mathrm{MSO}(\exists^{\aleph_0})$ formula $\varphi(\overline{Y})$ there exists an $\mathrm{MSO}(\mathrm{Inf})$ formula $\psi(\overline{Y})$ equivalent to $\varphi(\overline{Y})$ over all structures.

Proof:
Observe that the following are equivalent:

(1) There are only finitely many X which satisfy $\varphi(X, \overline{Y})$.

(2) There is a finite set Z such that any two distinct X_1, X_2 which both satisfy $\varphi(X_i, \overline{Y})$ differ on Z, i.e.

$$\exists Z \left(\neg \mathrm{Inf}(Z) \;\wedge\; \forall X_1 X_2 \;\left(\varphi(X_1, \overline{Y}) \wedge \varphi(X_2, \overline{Y}) \wedge X_1 \neq X_2 \;\rightarrow\; \exists z \in Z\,(z \in X_1 \leftrightarrow z \notin X_2)\right)\right).$$

Item (2) implies (1) as a collection of sets pairwise differing only on a finite set Z has cardinality at most $2^{|Z|}$. Conversely, if $X_1, \ldots, X_k$ are all the sets that satisfy $\varphi(X_i, \overline{Y})$, then choose for every pair of distinct sets X_i, X_j an element $z_{i,j}$ in the symmetric difference of X_i and X_j and define Z as the set of these chosen elements. □

Over finitely branching trees, $\mathrm{Inf}(Z)$ can of course be expressed in MLO. Indeed, with König's Lemma in mind, Z is infinite iff there is no downward closed set which includes Z and does not include an infinite, i.e., unbounded branch.

Corollary 2.1. $\mathrm{MLO}(\exists^{\aleph_0})$ collapses effectively to MLO over (finitely branching) trees.

Observe that the converse of Proposition 2.1 holds as well. In fact, the predicate $\mathrm{Inf}(Z)$ can be defined over all structures by the formula $\exists^{\kappa} Y\,(Y \subseteq Z)$ for any $\aleph_0 \leq \kappa \leq 2^{\aleph_0}$. Therefore, by Proposition 2.1, any of the quantifiers $\exists^{\kappa} Y$ with $\aleph_0 < \kappa \leq 2^{\aleph_0}$ can be used to define $\exists^{\aleph_0} X$ over arbitrary structures.

3. Preliminaries

For a given set A we denote by A^* the set of all finite sequences of elements of A, by A^ω the set of all infinite sequences of elements of A (i.e. functions $\omega \to A$), and $A^{\leq\omega} = A^* \cup A^\omega$. For any sequence $s = s_0 s_1 s_2 \ldots \in A^{\leq\omega}$ we denote by $|s|$ the length of s (either a natural number or ω) and by $s|_n = s_0 \ldots s_{n-1}$ the finite sequence composed of the first n elements of s, with $s|_0 = \varepsilon$, the empty sequence. We write $s[n]$ for the $(n+1)$st element of s (we count from 0), so $s[n] = s_n$ for $n \in \mathbb{N}$. Given a finite sequence s and a sequence $t \in A^{\leq\omega}$ we denote by $s \cdot t$ (or just st) the concatenation of s and t. Moreover, we write $s \preceq t$ if s is a prefix of t, i.e. if there exists a sequence r such that $t = sr$. A subset B of $A^{\leq\omega}$ is said to be prefix-closed if for every $t \in B$ and $s \preceq t$ it holds that $s \in B$.

3.1. Trees

For a number $l \in \mathbb{N}$, $l > 0$, an l-tree is a structure $\mathfrak{T} = (T, <, P_1, \ldots, P_l)$, where the P_i's are unary predicates and $<$ is the irreflexive and transitive binary *ancestor* relation with a least element called the *root of* $\mathcal{T}$ and such that for every $v \in T$ the set $\{u \in T \mid u < v\}$ of ancestors of v is finite and linearly ordered by $<$ and the number of $v \in T$ with at most n ancestors is finite for every natural n. Elements of a tree are referred to as *nodes*, maximal linearly ordered sets of nodes are called *branches*, ancestor-closed and linearly ordered sets of nodes are called *paths*, whereas *chains* are arbitrary linearly ordered subsets. An *antichain* is a set of pairwise incomparable nodes. Given a node v, the subtree of $\mathfrak{T}$ rooted in v is obtained by restricting the structure to the domain $T_v = \{u \in T \mid u \geq v\}$ and is denoted $\mathfrak{T}_v$.

Given a finite set A, we denote by $\mathfrak{T}(A)$ the full tree over A, which is a structure with the universe A^*, unary predicates $P_a = A^* a$ for each $a \in A$, and $<$ interpreted as the prefix ordering. For finite A with $|A| = n$, this structure is axiomatizable in MLO and its MLO theory is essentially the same as SnS, the monadic second-order theory of n successors (modulo trivial MLO-interpretations). We identify a path B of $\mathfrak{T}(A)$ with the sequence $\beta = a_0 a_1 a_2 \ldots \in A^{\leq\omega}$ such that $B = \{a_0 \ldots a_s \mid s \leq |\beta|\}$. Conversely, given a sequence $\beta \in A^{\leq\omega}$ we write $\mathrm{Pref}(\beta)$ for the corresponding path B.

Ordered sums of trees are defined as follows.

Definition 3.1. (Tree sum)
Let $l > 0$, $\mathfrak{I} = (I, <^{\mathcal{I}})$ be an unlabeled tree and let $\mathfrak{T}_i = (T_i, <^i, P_1^i, \ldots, P_l^i)$ be an l-tree for each $i \in I$. The *tree sum* of $(\mathfrak{T}_i)_{i\in\mathfrak{I}}$, denoted $\sum_{i\in\mathfrak{I}} \mathfrak{T}_i$, is the l-tree

$$\mathfrak{T} = \big(\bigcup_{i\in I} \{i\} \times T_i \ , <^{\mathfrak{T}}, \bigcup_{i\in I} \{i\} \times P_1{}^i, \ldots, \bigcup_{i\in I} \{i\} \times P_l{}^i \ \big) \, ,$$

where $(i, a) <^{\mathfrak{T}} (j, b)$ for $i, j \in I$, $a \in T_i$, $b \in T_j$ iff:

$$i <^{\mathfrak{I}} j \text{ and } a \text{ is the root of } \mathfrak{T}_i, \text{ or } \ i = j \text{ and } a <^i b \, .$$

Unless explicitly noted, we will not distinguish between $\mathfrak{T}_i$ and the isomorphic subtree $\{i\} \times \mathfrak{T}_i$ of $\mathfrak{T}$.

A particular special case of the sum we will be using is when the index structure $\mathfrak{I}$ consists of a single branch. Let $(I, <)$ be a linear order, which is finite or isomorphic to ω, and let $\langle \mathfrak{T}_i \mid i \in I \rangle$ be an I-indexed sequence of l-trees. Then the sum $\mathfrak{T} = \sum_{i\in I} \mathfrak{T}_i$ is well defined, and $(I, <)$ forms a path (not necessarily maximal) in $\mathfrak{T}$.

3.2. MLO and the composition method

We will work with labeled trees in the relational signature $\{<, P_1, \ldots, P_l\}$ where $<$ is a binary relation symbol denoting the ancestor relation of the tree, and the P_i's are unary predicates representing a labeling.

Monadic second-order logic of order, MLO for short, extends first-order logic by allowing quantification over *subsets* of the domain. MLO uses first-order variables $x, y, \ldots$ interpreted as elements, and set variables $X, Y, \ldots$ interpreted as subsets of the domain. Set variables will always be capitalized to distinguish them from first-order variables. The atomic formulas are of the form "$x < y$", "$x \in P_i$" or "$x \in X$". All other formulas are built from the atomic ones by applying Boolean connectives and the universal and existential quantifiers for both kinds of variables. Concrete formulas will be given in this syntax, taking the usual liberties and short-hands, such as $X \cup Y, X \cap Y, X \subseteq Y$, guarded quantifiers and relativization of formulas to a set.

The quantifier rank of a formula φ, denoted $\mathrm{qr}(\varphi)$, is the maximum depth of nesting of quantifiers in φ. For fixed n and l we denote by $\mathrm{Form}_{n,l}$ the set of formulas of quantifier depth $\leq n$ and with free variables among $X_1, \ldots, X_l$. Let $n, l \in \mathbb{N}$ and $\mathfrak{T}_1, \mathfrak{T}_2$ be l-trees. We say that $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are n-*equivalent*, denoted $\mathfrak{T}_1 \equiv^n \mathfrak{T}_2$, if for every $\varphi \in \mathrm{Form}_{n,l}$, $\mathfrak{T}_1 \models \varphi$ iff $\mathfrak{T}_2 \models \varphi$.

Clearly, $\equiv^n$ is an equivalence relation. For any $n \in \mathbb{N}$ and $l > 0$, the set $\mathrm{Form}_{n,l}$ is infinite. However, it contains only finitely many semantically distinct formulas, so there are only finitely many $\equiv^n$-classes of l-structures. In fact, we can compute representatives for these classes as follows.

Lemma 3.1. (Hintikka Lemma [11])
For $n, l \in \mathbb{N}$, we can compute a *finite* set $H_{n,l} \subseteq \mathrm{Form}_{n,l}$ such that:

- For every l-tree $\mathfrak{T}$ there is a *unique* $\tau \in H_{n,l}$ such that $\mathfrak{T} \models \tau$.
- If $\tau_1, \tau_2 \in H_{n,l}$ and $\tau_1 \neq \tau_2$ then $\tau_1 \wedge \tau_2$ is unsatisfiable.
- If $\tau \in H_{n,l}$ and $\varphi \in \mathrm{Form}_{n,l}$, then either $\tau \models \varphi$ or $\tau \models \neg\varphi$. Furthermore, there is an algorithm that, given such τ and φ, decides which of these two possibilities holds.

Elements of $H_{n,l}$ are called (n,l)-*Hintikka formulas*.

Given an l-tree $\mathfrak{T}$ we denote by $\mathrm{Tp}^n(\mathfrak{T})$ the unique element of $H_{n,l}$ satisfied in $\mathfrak{T}$ and call it the n-*type* of $\mathfrak{T}$. Thus, $\mathrm{Tp}^n(\mathfrak{T})$ effectively determines which formulas of quantifier-depth $\leq n$ are satisfied in $\mathfrak{T}$. We sometimes speak of the n-type of a tuple of subsets $\overline{V} = V_1, \ldots, V_m$ of a given l-tree $\mathfrak{T}$. This is synonymous with the n-type of the $(l+m)$-tree $(\mathfrak{T}, \overline{V})$ obtained by expansion of $\mathfrak{T}$ with the predicates $P_{l+1}, \ldots, P_{l+m}$ interpreted as the sets $V_1, \ldots, V_m$. This type will be denoted by $\mathrm{Tp}^n(\mathfrak{T}, \overline{V})$ and often referred to as an n-type in m variables, whereby the n-type of the $(l+m)$-tree $(\mathfrak{T}, \overline{V})$ is understood. To denote the n-type of $\overline{V}$ restricted to a substructure $\mathfrak{T}' \subseteq \mathfrak{T}$ we simply write $\mathrm{Tp}^n(\mathfrak{T}', \overline{V})$ instead of $\mathrm{Tp}^n(\mathfrak{T}', \overline{V} \cap \mathfrak{T}')$.

The essence of the composition method is that certain operations on structures, such as disjoint union and certain ordered sums, can be projected to n-types. A general composition theorem for MLO from which most others follow is due to Shelah [19]. In this paper we use the following form of composition, a more detailed presentation of the method can be found in [10, 21].

Theorem 3.1. (Composition Theorem for Trees)
For every MLO-formula $\varphi(\overline{X})$ in the signature of l-trees having m free variables and quantifier rank n, and given the enumeration $\tau_1(\overline{X}), \ldots, \tau_k(\overline{X})$ of $H_{n,l+m}$, there exists an MLO-formula $\theta(Q_1, \ldots, Q_k)$ computable from φ and such that for every tree $\mathfrak{I} = (I, <^I)$ and family $\{\mathfrak{T}_i \mid i \in I\}$ of l-trees and subsets $V_1, \ldots, V_m$ of $\sum_{i \in I} \mathfrak{T}_i$,

$$\sum_{i \in I} \mathfrak{T}_i \models \varphi(\overline{V}) \iff \mathfrak{I} \models \theta(Q_1, \ldots, Q_k)$$

where $Q_r = Q_r^{I;\overline{V}} = \{i \in I \mid \mathrm{Tp}^n(\mathfrak{T}_i, \overline{V}) = \tau_r\}$ for each $1 \leq r \leq k$.

4. D-nodes versus U-nodes and relevant branches

A *tree segment*, or *interval*, of an l-tree is a connected and convex set I of nodes, i.e. such that for every $u, w \in I$ if u and w are incomparable, then their greatest common ancestor is in I, and if $u < w$ then for every $u < v < w$ also $v \in I$. Every tree segment has a minimal element and every subtree $\mathfrak{T}_z$ of a tree $\mathfrak{T}$ is a tree segment. More generally, the summands $\mathfrak{T}_i$ of any tree sum $\mathfrak{T} = \sum_{i \in I} \mathfrak{T}_i$ are tree segments of $\mathfrak{T}$. The terms 'interval' and 'tree segment' are used interchangeably.

We denote by $\mathfrak{T}|_I$ the restriction of an l-tree $\mathfrak{T}$ to the interval I. Alternatively, given a node z and a set Z of nodes of $\mathfrak{T}$ we use the notation $\mathfrak{T}_{z \setminus Z}$ for the restriction of $\mathfrak{T}$ to the tree segment $\mathfrak{T}_z \setminus (\bigcup_{w \in Z, z < w} \mathfrak{T}_w)$. Any interval I with a minimal element z can be written in the form $\mathfrak{T}_{z \setminus Z}$, where $Z = \{u \mid u \geq z \wedge u \notin I\}$. In particular, if B is a branch, $v, w \in B$ such that w is the immediate successor of v on B, then $T_{v \setminus B} = T_v \setminus T_w$. These notations are schematically depicted in Figure 1.

Consider an MLO formula $\varphi(X, \overline{Y})$ of l-trees. To eliminate a single occurrence of the uncountability quantifier from $\exists^{\aleph_1} X\ \varphi(X, \overline{Y})$ over l-tree $\mathfrak{T}$ we will make extensive use of the following notions for intervals. For the rest of this section we fix $\varphi(X, \overline{Y})$: an MLO formula of l-trees with $1 + m$ free variables – of which $\overline{Y} = (Y_1, \ldots, Y_m)$ will often be regarded as parameters – and of quantifier rank n.

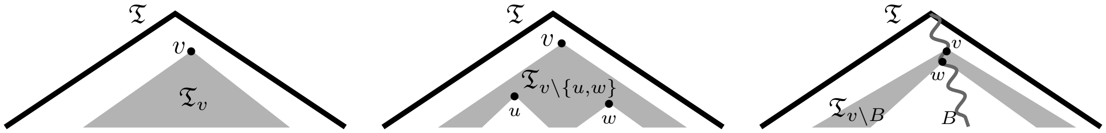


Figure 1. A subtree $\mathfrak{T}_v$ and tree segments $\mathfrak{T}_{v\setminus\{u,w\}}$ and $\mathfrak{T}_{v\setminus B}$.

Definition 4.1. Let $\mathfrak{T}$ be an l-tree, $X, \overline{Y}$ subsets such that $\mathfrak{T} \models \varphi(X, \overline{Y})$, and let I be an interval of $\mathfrak{T}$.

(1) We say that I is a *U-interval* for φ, X, $\overline{Y}$ whenever $X \cap I$ is the unique subset of its type on $\mathfrak{T}|_I$. More precisely, if $\mathfrak{T}|_I \models \forall Z\, \tau(Z, \overline{Y}) \rightarrow Z = X$, where $\tau(X, \overline{Y})$ is the n-type of $(\mathfrak{T}, X, \overline{Y})|_I$.

(2) I is a *D-interval* for φ, X, $\overline{Y}$ iff it is not a U-interval.

(3) In the special case of $I = \{u \mid u \geq z\}$ we say that the subtree $\mathfrak{T}_z$ is a *U-tree* or *D-tree*, respectively, and further say that z is a *U-node* or *D-node* for $\varphi, X, \overline{Y}$.

(4) The set of D-nodes for $\varphi, X, \overline{Y}$ is denoted $D(X)$.

(5) An infinite path P is called a *D-path* for $\varphi, X, \overline{Y}$ if every $v \in P$ is a D-node for $\varphi, X, \overline{Y}$. That is if $P \subseteq D(X)$.

The name "U-interval" attests to the fact that the set X in question is *uniquely* determined by its type on a given interval, as opposed to "D-intervals" offering two (or more) distinct choices for X with the same type on the interval, thus (at least) *doubling* the total number of choices for X over the entire domain. Whenever φ and $\overline{Y}$ are clear from the context we will write e.g. "D-interval for X" instead of "D-interval for $\varphi, X, \overline{Y}$", and similarly for the other notions above.

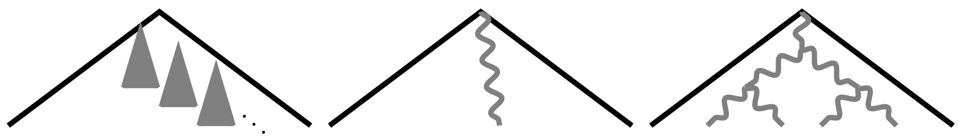
It is worth noting that each set $D(X)$ is prefix-closed since whenever $\mathfrak{T}_v$ is a D-tree and $u < v$, then $\mathfrak{T}_v$ is a subtree of $\mathfrak{T}_u$ and hence, by composition, $\mathfrak{T}_u$ is a D-tree as well. Thus $D(X)$ induces a tree whose infinite paths are precisely the D-paths for X.

Each of the notions introduced in Definition 4.1 can be formalized in MLO. Let us start by constructing the formula $\mathrm{DINT}_\varphi(I, X, \overline{Y})$, expressing that I is a D-interval for $\varphi, X, \overline{Y}$. By Lemma 3.1, the set of n-types $H_{n,l+m+1}$ is finite and can be computed. Take the formula

$$\psi_{\mathrm{eqtp}}(X, Z, \overline{Y}) \;=\; \bigwedge_{\tau \in H_{n,l+m+1}} \tau(X, \overline{Y}) \leftrightarrow \tau(Z, \overline{Y})$$

expressing that X and Z have the same n-type (on the tree at large), and let $\psi^{\mathrm{rel}}_{\mathrm{eqtp}}(X, Z, \overline{Y}, I)$ be the relativization of $\psi_{\mathrm{eqtp}}(X, Z, \overline{Y})$ to an interval I, thus asserting that X and Z have the same n-type on I. $\mathrm{DINT}_\varphi(I, X, \overline{Y})$ can now be written as

$$\varphi(X, \overline{Y}) \;\wedge\; \exists Z(\psi^{\mathrm{rel}}_{\mathrm{eqtp}}(X, Z, \overline{Y}, I) \;\wedge\; X \cap I \neq Z \cap I)\,.$$



(A) (B) (C)

Figure 2. The three conditions.

Using $\mathrm{DINT}_\varphi(I, X, \overline{Y})$ one can build formulas $\mathrm{DNODE}_\varphi(v, X, \overline{Y})$ and $\mathrm{DPATH}_\varphi(P, X, \overline{Y})$ expressing, respectively, that v is a D-node and that P is a D-path for $\varphi, X, \overline{Y}$; and also $\mathrm{DSET}_\varphi(D, X, \overline{Y})$ which holds iff $D = D(X)$.

The following lemma is the first step in eliminating the $\exists^{\aleph_1}$ quantifier from MLO over trees. The three cases are depicted in Figure 2.

Lemma 4.1. Let $\mathfrak{T}$ be an l-tree and $\varphi(X, \overline{Y})$ an MLO-formula in the signature of l-trees. Then, for every tuple of subsets $\overline{V}$ of $\mathfrak{T}$,

$$\mathfrak{T} \models \exists^{\aleph_1} X\, \varphi(X, \overline{V})$$

if and only if one of the following conditions is satisfied.

A. There is a set U satisfying $\mathfrak{T} \models \varphi(U, \overline{V})$ and there is an infinite antichain A of D-nodes for $\varphi, U, \overline{V}$.

B. There is an infinite branch B, which is a D-path for uncountably many U satisfying $\mathfrak{T} \models \varphi(U, \overline{V})$.

C. There are uncountably many branches B in $\mathfrak{T}$, each of which is a D-path for some U satisfying $\mathfrak{T} \models \varphi(U, \overline{V})$.

Proof:
Note that over finitely branching trees, where König's Lemma applies, condition A implies condition B and is enlisted here for deductive reasons only.

On the one hand, A is arguably the most natural and easily expressible condition sufficient for the existence of continuum many sets U satisfying $\mathfrak{T} \models \varphi(U, \overline{V})$. To see that, let U and A be as in A and let $I = \{w \in \mathfrak{T} \mid \neg\exists v\,(v \in A \wedge v < w)\}$ be the set of all nodes which are not below any of the nodes of A. Then $\mathfrak{T}$ can be decomposed with $(I, <)$ as index structure as $\mathfrak{T} = \sum_{w \in I \setminus A} [w] + \sum_{w \in A} \mathfrak{T}_w$. Here $[w]$ denotes a tree consisting of a single node bearing the same labels as w in $\mathfrak{T}$. We apply the Composition Theorem to this decomposition. Given that $\mathfrak{T} \models \varphi(U, \overline{V})$ using Theorem 3.1 we can ascertain that $\mathfrak{T} \models \varphi(U', \overline{V})$ for every U' such that $U' \cap (I \setminus A) = U \cap (I \setminus A)$ and $\mathrm{Tp}^n(\mathfrak{T}_w, U', \overline{V}) = \mathrm{Tp}^n(\mathfrak{T}_w, U, \overline{V})$ for all $w \in A$. By the choice of A such a U' can be independently chosen either to coincide or not to coincide with U on each subtree $\mathfrak{T}_w$ with $w \in A$ without changing its type. Hence there are continuum many different such U' and A is an antichain of D-nodes for every such U'. In a (finitely branching) tree with U and A fulfilling condition A there is also, by König's Lemma, an infinite branch B such that $\mathfrak{T}_v \cap A$ is infinite for all $v \in B$. In particular, B is a D-path for each U' obtained from U as above, implying condition B.

On the other hand, $\neg\mathsf{A}$ amounts to saying that for each U satisfying $\varphi(U,\overline{V})$ the set $D(U)$ induces a tree comprised of only finitely many branches. In particular, that there are only finitely many infinite D-paths for each such U.

Condition B explicitly requires the existence of uncountably many sets satisfying $\varphi(X,\overline{V})$, so it too is sufficient for $\exists^{\aleph_1} X\ \varphi(X,\overline{V})$ to hold. Hence it remains to be shown that when B fails then C is both sufficient and necessary hereto.

Assuming B does not hold in some $\mathfrak{T}$ then, as we have seen, A fails too and therefore there are only finitely many infinite D-paths for each U satisfying $\mathfrak{T} \models \varphi(U,\overline{V})$. Also by the failure of B every branch is a D-path for at most countably many U satisfying $\mathfrak{T} \models \varphi(U,\overline{V})$. It follows that for every such set U the collection $\{U' \mid D(U') = D(U)\,,\ \mathfrak{T} \models \varphi(U',\overline{V})\}$ is finite or countable. Indeed, this is clear from the above whenever $D(U)$ contains an infinite D-path. If on the other hand $D(U)$ is finite then U is fully determined by $U \cap D(U)$ and the n-types of all those U-nodes that are successors of some D-node,which only allows for a finite number of choices of U given that $\mathfrak{T}$ is finitely branching. Thus we have established that whenever B fails in some $\mathfrak{T}$ then there are uncountably many U satisfying $\mathfrak{T} \models \varphi(U,\overline{V})$ iff there are uncountably many sets $D(U)$ with $\mathfrak{T} \models \varphi(U,\overline{V})$ if and (because now each relevant $D(U)$ contains only finitely many branches) only if condition C holds. □

We remark that Lemma 4.1 fails for infinitely branching trees. Consider a tree of depth one with the root r having countably many successor nodes and the formula $\varphi(X,Y) = X \subseteq Y$ and fix a set V of successor nodes. Then $D(X) \subseteq \{r\}$ for every X satisfying $\varphi(X,V)$, hence conditions A, B and C all fail. Note that over infinitely branching trees even the predicate $\mathrm{Inf}(X)$ cannot be expressed in pure MLO. To extend our results to infinitely branching trees (reducing to $\mathrm{MLO}(\mathrm{Inf})$ instead of pure MLO) thus requires a fourth condition addressing such cases while making use of the Inf predicate.

Let us note again that if condition A holds then there are in fact continuum many sets X satisfying the formula $\varphi(X,\overline{Y})$. The description of Condition A can be directly formalized in $\mathrm{MLO}(\mathrm{Inf})$, hence, over (finitely branching) trees, also in MLO as follows:

$$\psi_{\mathsf{A}}(\overline{Y}) = \exists U\, \exists A\, \big(\varphi(U,\overline{Y}) \wedge \mathrm{Inf}(A) \wedge \mathrm{antichain}(A) \wedge \big(\forall w \in A\, \mathrm{DNODE}_{\varphi}(w,U,\overline{Y})\big)\big)\,,$$

where $\mathrm{antichain}(A) = \forall x, y \in A\, \neg(x < y \vee y < x)$.

5. Condition B

In this section, we show that a branch B is a witness for Condition B if and only if this branch satisfies a disjunction of three sub-conditions: Ba, Bb and Bc. Moreover, if both Condition A and Condition C fail, then already the sub-conditions Ba and Bc are sufficient. Finally, we express both Ba and Bc in MLO and show, that in fact both these sub-conditions guarantee the existence of continuum many sets X satisfying the formula $\varphi(X,\overline{Y})$ in consideration. As in the previous section, we fix an MLO-formula of l-trees $\varphi(X,\overline{Y})$ in $1+m$ many free variables and of quantifier rank n.

Consider the formula $\psi(X,\overline{Y},P)$ stating that P is an infinite D-path for X and that $\varphi(X,\overline{Y})$ holds.

$$\psi(X,\overline{Y},P) = \mathrm{DPATH}_{\varphi}(P,X,\overline{Y}) \wedge \mathrm{Inf}(P) \wedge \varphi(X,\overline{Y})$$

Note that a branch B witnesses Condition B in an l-tree $\mathfrak{T}$ if and only if $\mathfrak{T} \models \exists^{\aleph_1} U\ \psi(U,\overline{Y},B)$. To break up Condition B for a given branch B we therefore apply the Composition Theorem for the formula ψ

with the decomposition $\mathfrak{T} = \sum_{w \in B} \mathfrak{T}_{w\setminus B}$ along that branch. To that end let r be the number of $\mathrm{qr}(\psi)$-types in $l+m+2$ variables, which we enumerate as $\tau_1, \ldots, \tau_r$. Then Theorem 3.1 yields a formula θ such that

$$\mathfrak{T} \models \psi(X, \overline{Y}, B) \iff (B, <) \models \theta(P_1, \ldots, P_r) \tag{1}$$

with $P_i = \{w \in B \mid (\mathfrak{T}_{w\setminus B}, X, \overline{Y}, \{w\}) \models \tau_i\}$ for each $1 \leq i \leq r$. Note that we use the expansion of $\mathfrak{T}_{w\setminus B}$ by $\{w\}$ as w is the only element of $\mathfrak{T}_{w\setminus B}$ that belongs to B.

With this reformulation it is clear that a branch B witnesses Condition B in an l-tree $\mathfrak{T}$ if and only if there are uncountably many different $\overline{P}$ satisfying θ, or some $\overline{P}$ satisfying θ has uncountably many X corresponding to it. Taking advantage of the fact that, by virtue of the Composition Theorem, θ merely depends on ψ but not on $\mathfrak{T}$ nor the chosen branch B, we obtain the following breakdown of condition B.

Lemma 5.1. Let $\mathfrak{T}$ be an l-tree and B an infinite branch in $\mathfrak{T}$. There are uncountably many $X \subseteq \mathfrak{T}$ satisfying the formula $\psi(X, \overline{Y}, B)$ in $\mathfrak{T}$ iff one of the following sub-conditions holds.

(Ba) There exists a set X such that $\mathfrak{T}_{w\setminus B}$ is a D-interval for $\varphi, X, \overline{Y}$ for infinitely many $w \in B$.

(Bb) There exists a set X satisfying ψ and a $w \in B$ so that

$$\mathfrak{T}_{w\setminus B} \models \exists^{\aleph_1} X' \, \tau_i(X', \overline{Y} \cap \mathfrak{T}_{w\setminus B}, \{w\}),$$

where $\tau_i = \mathrm{Tp}^{\mathrm{qr}(\psi)}(\mathfrak{T}_{w\setminus B}, X, \overline{Y}, \{w\})$ for all $1 \leq i \leq r$.

(Bc) It holds that

$$(B, <) \models \exists^{\aleph_1} \overline{P} \left(\theta(\overline{P}) \wedge \bigwedge_{i=1}^{r} P_i \subseteq Q_i \wedge \forall x \bigvee_{i=1}^{r} \Big(x \in P_i \wedge \bigwedge_{j \neq i} x \notin P_j \Big) \right),$$

where for each $1 \leq i \leq r$, Q_i is the set of nodes on the branch B in which the type τ_i is satisfied by some set X, i.e.

$$Q_i = \{w \in B \mid \mathfrak{T}_{w\setminus B} \models \exists X \, \tau_i(X, \overline{Y} \cap \mathfrak{T}_{w\setminus B}, \{w\})\}.$$

Proof:
Recall that by (1) we have $\mathfrak{T} \models \psi(X, \overline{Y}, B)$ iff $(B, <) \models \theta(P_1, \ldots, P_r)$. We consider two cases.

Case 1: There exists a tuple $\overline{P}$ such that $(B, <) \models \theta(\overline{P})$ and there are uncountably many sets X for which $P_i = \{w \in B \mid (\mathfrak{T}_{w\setminus B}, X, \overline{Y}, \{w\}) \models \tau_i\}$ for each $1 \leq i \leq r$.
In this case the branch B witnesses Condition B, so we only need to show that one of the sub-conditions holds. Consider a set X_0 satisfying $\psi(X_0, \overline{Y}, B)$ and having $\mathrm{qr}(\psi)$-types on $\mathfrak{T}_{w\setminus B}$ for all $w \in B$ as described by $\overline{P}$. Assume that sub-condition (Ba) does not hold. Then the segment $\mathfrak{T}_{w\setminus B}$ is a U-interval for $\varphi, X_0, \overline{Y}$ for all but finitely many $w \in B$. Observe that $\mathrm{qr}(\psi) \geq \mathrm{qr}(\varphi)$. Therefore all of the uncountably many sets X that induce $\overline{P}$, i.e. have the same $\mathrm{qr}(\psi)$-type as X_0 on each segment $\mathfrak{T}_{w\setminus B}$, must be equal to X_0 on all but finitely many $\mathfrak{T}_{w\setminus B}$. So there is a $w \in B$ for which there are uncountably many different X having the same $\mathrm{qr}(\psi)$-type as X_0 on $\mathfrak{T}_{w\setminus B}$, and thus Condition (Bb) is satisfied.

Case 2: For each tuple $\overline{P}$ such that $(B,<) \models \theta(\overline{P})$ there are only countably many sets X for which $P_i = \{w \in B \mid (\mathfrak{T}_{w\setminus B}, X, \overline{Y}, \{w\}) \models \tau_i\}$.
In this case, we show that Condition (Bc) is both necessary and sufficient for the existence of uncountably many sets X satisfying ψ.

Necessity of Condition (Bc).
As a direct consequence of (1) and the condition of this case, if there are uncountably many sets X satisfying ψ then there are uncountably many corresponding tuples $\overline{P}$ for which $(B,<) \models \theta(\overline{P})$. Each P_i induced by some X as in (1) is, by definition, the set of w's for which $(\mathfrak{T}_{w\setminus B}, X, \overline{Y}, \{w\}) \models \tau_i$. So for every $w \in P_i$ we have, in particular, that $\mathfrak{T}_{w\setminus B} \models \exists X\, \tau_i(X, \overline{Y} \cap \mathfrak{T}_{w\setminus B}, \{w\})$. Thus $P_i \subseteq Q_i$ for every i. Since Hintikka formulas are mutually exclusive the P_i's are pairwise disjoint. This guarantees that the remaining conjunct $\forall x \big(\bigvee_{i=1}^{r} (x \in P_i \wedge \bigwedge_{s\neq r} x \notin P_s\big)$ of Condition (Bc) is also satisfied, and therefore Condition (Bc) holds.

Sufficiency of Condition (Bc).
By definition of the sets Q_i, for each $w \in Q_i$ there is a subset $X_{w,i} \subseteq \mathfrak{T}_{w\setminus B}$ such that $\mathfrak{T}_{w\setminus B} \models \tau_i(X_{w,i}, \overline{Y}, \{w\})$. Assuming that Condition (Bc) holds, let $\mathcal{P}$ be the uncountable set of tuples $\overline{P}$ that witness this condition. For each such tuple $\overline{P}$ and each $w \in B$ the last conjunct of Condition (Bc) guarantees that there is a unique $i = i(w, \overline{P})$ for which $w \in P_i$. Let $X_{\overline{P}} = \bigcup_{w\in B} X_{w,i(w,\overline{P})}$. Since $P_i \subseteq Q_i$, the tuple $\overline{P}$ describes indeed the types of the set $X_{\overline{P}}$ on the tree segments $\mathfrak{T}_{w\setminus B}$. According to (1) from $(B,<) \models \theta(\overline{P})$ we can infer that $\mathcal{T} \models \psi(X_{\overline{P}}, \overline{Y}, B)$. Clearly, for distinct tuples $\overline{P}_1$ and $\overline{P}_2$ the sets $X_{\overline{P_1}}$ and $X_{\overline{P_2}}$ are also distinct. Therefore $\{X_{\overline{P}} \mid \overline{P} \in \mathcal{P}\}$ constitutes an uncountable family of sets satisfying ψ. □

Observe that (Ba) already subsumes A in the sense that if condition A holds then there is a branch satisfying (Ba). Also observe that Condition (Bb) is itself just another instance of our initial problem. It is important to note, however, that the above cases classify conditions under which an *individual branch* may satisfy B. At closer inspection we find that if no branch satisfies either (Bc) or (Ba) (so that in particular A fails) and moreover condition C fails too, then (Bb) cannot hold either.

Lemma 5.2. If over a tree $\mathfrak{T}$ both Conditions A and C fail, then Condition B implies that some branch of $\mathfrak{T}$ satisfies Condition (Ba) or Condition (Bc).

One intuitive way to see this is that if all the conditions A, (Ba), (Bc) and C fail on a tree, and thereby also on every tree segment of that tree, then for (Bb) to hold for a proper tree segment that tree segment would have to contain a proper tree segment on which (Bb) holds, and so on indefinitely. This would ultimately trace an infinite branch witnessing (Ba) contrary to the initial assumption.

Proof:
It is easy to see that if conditions A and C fail then $\mathcal{D} = \{D(X) \mid \mathfrak{T} \models \varphi(X, \overline{Y})\}$ is countable. Indeed, in the proof of Lemma 4.1 we have already remarked that the failure of A implies that each $D \in \mathcal{D}$ is a union of finitely many paths and, by definition, C holds unless there are only countably many potential D-paths in total.

If Condition B holds then there are uncountably many sets X satisfying $\varphi(X, \overline{Y})$ and thus, as $\mathcal{D}$ is countable, there is a set D such that $D = D(X)$ for uncountably many X satisfying φ. Fix such a D and consider the set of labelings $\mathcal{L} = \{\lambda^X : D \to H_{n,l+m+1} \mid D(X) = D,\ \mathfrak{T} \models \varphi(X, \overline{Y})\}$, where

$\lambda^X(w) = \mathrm{Tp}^n(\mathfrak{T}_{w\backslash D}, X, \overline{Y})$ for all $w \in D$. We distinguish two cases.

Case 1: $\mathcal{L}$ is uncountable. Then, given that D contains only finitely many infinite paths and finitely many additional nodes, there is an infinite branch B in D such that $\{\lambda|_B \mid \lambda \in \mathcal{L}\}$ is uncountable. Observe that $\lambda^X(w) = \mathrm{Tp}^n(\mathfrak{T}_{w\backslash B}, X, \overline{Y})$ for all but finitely many nodes $w \in B$. Also observe that, since $\mathrm{qr}(\psi) \geq n$, each $\mathrm{qr}(\psi)$-type on the variables $X, \overline{Y}, B$ induces a unique n-type on the variables $X, \overline{Y}$. So there are necessarily uncountably many different partitions $\overline{P}^X = \langle P_1^X, \dots P_r^X \rangle$ of B

$$P_j^X = \{w \in B \mid \mathrm{Tp}^{\mathrm{qr}(\psi)}(\mathfrak{T}_{w\backslash B}, X, \overline{Y}, \{w\}) = \tau_j\} \;\; (1 \leq j \leq r)$$

with $D(X) = D$ and X satisfying φ. Using (1) we can check that Condition (Bc) is met for the branch B.

Case 2: $\mathcal{L}$ is countable. Then there is a type labeling $\lambda : D \to H_{n,l+m+1}$ such that $\lambda = \lambda^X$ for uncountably many X satisfying φ and having $D(X) = D$. Suppose that Condition (Ba) is not satisfied for any infinite branch B in D. Then $\lambda(w)$ uniquely determines $X \cap \mathfrak{T}_{w\backslash D}$ for all but finitely many $w \in D$ and all X satisfying φ and $D(X) = D$. Thus, there exists a $w \in D$ such that there are uncountably many X as above pairwise differing on the tree segment $\mathfrak{T}_{w\backslash D}$. However, by definition, every subtree of $\mathfrak{T}_{w\backslash D}$ is a U-tree relative to each of these X, because $D(X) = D$. Because $\mathfrak{T}$ is finitely branching, i.e. $\mathfrak{T}_{w\backslash D} \setminus \{w\}$ is a finite union of such U-trees, there can be only finitely many X as above and pairwise differing on $\mathfrak{T}_{w\backslash D}$, which is a contradiction. Therefore Condition (Ba) must hold. □

Next we will construct MLO formulas $\psi_{\mathsf{Ba}}(B, \overline{Y})$ and $\psi_{\mathsf{Bc}}(B, \overline{Y})$ formalizing sub-conditions (Ba) and (Bc), respectively. By the above, we can then use the formula $\psi_{\mathsf{B}}(\overline{Y}) = \exists B(\psi_{\mathsf{Ba}}(B, \overline{Y}) \vee \psi_{\mathsf{Bc}}(B, \overline{Y}))$ in place of Condition B in Lemma 4.1.

5.1. Formalization of Condition Ba

Much like condition A, (Ba) is naturally expressible in $\mathrm{MLO(Inf)}$ and thus, over trees, in pure MLO as well by the formula

$$\psi_{\mathsf{Ba}}(B, \overline{Y}) = \exists X \; \exists^{\aleph_0} w \; \mathrm{DINT}(T_{w\backslash B}, X, \overline{Y}),$$

where $T_{w\backslash B}$ is just a notation for the set defined by

$$x \in T_{w\backslash B} \iff w \leq x \,\wedge\, \neg \exists b \in B \, (b > w \wedge b \leq x).$$

The fact that Condition (Ba) is sufficient for the existence of continuum many sets U satisfying $\varphi(U, \overline{V})$ can be arrived at by appealing to the Composition Theorem in the same manner as for Condition A in the proof of Lemma 4.1, because the set X can be left intact or changed to another one with the same type on any of the infinitely many trees $\mathfrak{T}_{w\backslash B}$ which are D-intervals for X.

5.2. Formalization of Condition Bc

In order to eliminate the explicit use of the uncountability quantifier in Condition (Bc) over $(B, <) \cong (\omega, <)$, we make use of Proposition 2.5 from [13], which states that cardinality quantifiers can be eliminated over $(\omega, <)$, cf. also [4]. In [13] it was stated in automata theoretic language, we reformulate it in logical terms.

Proposition 5.1. For every MLO formula $\varphi(\overline{X},\overline{Y})$ there exists an effectively constructible formula $\psi(\overline{Y})$ such that over $(\omega,<)$ the following equivalence holds:

$$\psi(\overline{Y}) \;\equiv\; \exists^{\aleph_1}\overline{X}\,\varphi(\overline{X},\overline{Y}) \;\equiv\; \exists^{2^{\aleph_0}}\overline{X}\,\varphi(\overline{X},\overline{Y}).$$

Applying this result to the formula on the right hand side of Condition (Bc), with $\overline{Q}$ as parameters, we obtain a formula $\vartheta(\overline{Q})$ such that Condition (Bc) holds iff $(B,<)\models\vartheta(\overline{Q})$, with $\overline{Q}$ as specified there. By Proposition 5.1, if $\vartheta(\overline{Q})$ holds, then there are even continuum many sets $\overline{P}$ satisfying Condition (Bc). This in turn ensures the existence of continuum many sets X satisfying $\varphi(X,\overline{Y})$, because for each $\overline{P}$ accounted for in $\vartheta(\overline{Q})$ a corresponding X satisfying $\psi(X,\overline{Y},B)$ can be found and this association is necessarily injective.

To formalize Condition (Bc) in MLO over the tree $\mathfrak{T}$, we first define the sets Q_i on $\mathfrak{T}$. As the set of types is computable, we can compute each τ_i and thus effectively construct the formula $\alpha_i(w,B,\overline{Y})$ expressing that w is a node on the branch B such that $\mathfrak{T}_{w\setminus B}\models\exists X\;\tau_i(X,\overline{Y}\cap\mathfrak{T}_{w\setminus B},\{w\})$, i.e. $w\in Q_i$. Using this formula we can express Condition (Bc) as

$$\psi_{\mathsf{Bc}}(B,\overline{Y})=\exists\overline{Q}\;\left(\bigwedge_{i=1}^{r}\big(w\in Q_i\leftrightarrow\alpha_i(w,B,\overline{Y})\big)\;\wedge\;\vartheta^B(\overline{Q})\right)$$

where ϑ^B is a relativization of ϑ to the branch B.

6. The full binary tree and the Cantor space

In order to formalize Condition C in MLO over trees, we first analyze the problem only on the full binary tree and identify and prove the following key topological property that distinguishes counting branches from counting arbitrary sets.

On the full binary tree $\mathfrak{T}(2)=(\{0,1\}^*,\prec,S_0,S_1)$ where $\prec$ is the prefix-order and $S_i=\{0,1\}^*i$, we show that the set of branches satisfying any given MLO formula is a Borel set in the Cantor topology and hence it has the *perfect set property*: it is uncountable iff it contains a perfect subset iff it has the cardinality of the continuum. A *perfect set* is a closed set without isolated points.

Overview of topological notions

The argument we present is based on basic results of descriptive set theory and the theory of finite automata on infinite words in connection with monadic second-order logic and the Borel hierarchy of the Cantor space. Let us recall a few basic notions from descriptive set theory. A thorough introduction to descriptive set theory can be found in [14], we only mention a few basic facts.

The Cantor space is the topological space with the product topology on $\{0,1\}^\omega$. It is a Polish space with the topology generated by basic neighborhoods $w\{0,1\}^\omega$ with the prefix $w\in\{0,1\}^*$. Alternatively, it can be defined by the metric $d(\alpha,\beta)=2^{-\min\{n\,:\,\alpha[n]\neq\beta[n]\}}$.

The hierarchy of Borel sets is generated starting from open sets, i.e. unions of basic neighborhoods, denoted $\mathbf{\Sigma}^0_1$, and closed sets, which are complements of open sets and denoted $\mathbf{\Pi}^0_1$. Further on by transfinite induction for any countable ordinal α, $\mathbf{\Sigma}^0_\alpha$ is defined as $\{\bigcup_{i\in\omega}A_i\mid\forall i\;\exists\beta_i<\alpha\;A_i\in\mathbf{\Pi}^0_{\beta_i}\}$ and the $\mathbf{\Pi}^0_\alpha$-sets are the complements of $\mathbf{\Sigma}^0_\alpha$-sets. Each class $\mathbf{\Sigma}^0_\alpha$ and $\mathbf{\Pi}^0_\alpha$ is closed under taking inverse images by continuous functions. In fact there are complete languages in each class with respect to continuous reductions. The projective hierarchy is built on top of the Borel hierarchy, starting with $\mathbf{\Sigma}^1_0 = \mathbf{\Pi}^1_0$ as the class of Borel sets. On the first level one has the class $\mathbf{\Sigma}^1_1$ of *analytic sets*, which are projections of Borel sets, and the class $\mathbf{\Pi}^1_1$ of *co-analytic sets*, whose complements are analytic. The hierarchy is built in this manner with sets in $\mathbf{\Sigma}^1_{\alpha+1}$ being projections of $\mathbf{\Pi}^1_\alpha$-sets, and $\mathbf{\Pi}^1_{\alpha+1}$ sets being complements of $\mathbf{\Sigma}^1_\alpha$ sets.

The connection between the topological complexity of MLO-definable tree languages and the complexity of tree-automata recognizing them is well understood [22, 16]. By Rabin's complementation theorem, all MLO-definable tree languages are in $\mathbf{\Sigma}^1_2 \cap \mathbf{\Pi}^1_2$. There are $\mathbf{\Sigma}^1_1$-complete as well as $\mathbf{\Pi}^1_1$-complete regular tree languages. For instance, the set of $\{a, b\}$-labeled binary trees, which have on every path only finitely many a's, is $\mathbf{\Pi}^1_1$-complete [1, 16]. There are regular tree languages on arbitrary finite levels of the Borel hierarchy [20]. There also exist regular tree languages not contained in $\mathbf{\Sigma}^1_1 \cup \mathbf{\Pi}^1_1$, however, languages accepted by deterministic tree automata do belong to $\mathbf{\Pi}^1_1$.

This is in stark contrast to the situation of ω-regular languages, i.e. MLO-definable sets of ω-words, which are, by McNaughton's theorem, Boolean combinations of $\mathbf{\Pi}^0_2$ sets [22].

The Cantor-Bendixson Theorem states that closed subsets of a Polish space have the *perfect set property*: they are either countable or contain a perfect subset and thus have cardinality continuum. A set P is *perfect* if it is closed and if it has no isolated points, i.e. if every open neighborhood of every point $p \in P$ contains another point of P. We shall rely on the following fundamental result on Borel sets.

Proposition 6.1. ([12, Theorem 13.6])
Every uncountable Borel subset of a Polish space contains a perfect subset.

In fact, Souslin has proved that all analytic sets have the perfect set property [14]. It is, however, independent of ZFC whether all co-analytic sets, or all sets on higher levels of the projective hierarchy, satisfy the continuum hypothesis [14]. A key observation that our formalization will exploit is that, even though there are non-analytic sets of trees definable in MLO, sets of definable paths are Borel. Recall that for a sequence $\pi \in \{0,1\}^\omega$ we denote by $\mathrm{Pref}(\pi)$ the path through the full binary tree $\mathfrak{T}(2)$ that corresponds to this sequence, which formally is the set of prefixes of π.

Theorem 6.1. (MLO definable sets of branches are Borel)
Let $U_1, \ldots, U_m$ be subsets of $\mathfrak{T}(2)$ and let $\psi(X, \overline{Y})$ be an MLO formula over $\mathfrak{T}(2)$. Then the set

$$\mathcal{X} = \{\, \pi \in \{0,1\}^\omega \mid \mathfrak{T}(2) \models \psi(\mathrm{Pref}(\pi), \overline{U}) \,\}$$

of branches of the binary tree satisfying $\psi(X, \overline{U})$ is on the third level of the Borel hierarchy, in particular, it has the perfect set property.

Proof:
Given a path $\pi \in \{0,1\}^\omega$ let $B = \mathrm{Pref}(\pi)$ be the corresponding infinite branch and consider the labeled tree $\mathfrak{T}^\pi = (\mathfrak{T}(2), \mathrm{Pref}(\pi), \overline{U})$, and its decomposition as a tree sum along π: $\mathfrak{T}^\pi = \sum_{v \in B} \mathfrak{T}^\pi_{v \setminus B}$. Applying the Composition Theorem to $\mathfrak{T}^\pi$ and φ we find θ such that

$$\mathfrak{T}(2) \models \varphi(\mathrm{Pref}(\pi), \overline{U}) \iff \sum_{v \in B} \mathfrak{T}^\pi_{v \setminus B} \models \varphi \iff (B, <) \models \theta(Q^\pi_1, \ldots, Q^\pi_k)$$

where $Q_r^\pi = \{v \in B \mid \mathrm{Tp}^n(\mathfrak{T}^\pi_{v\setminus B}) = \tau_r\}$ for each $1 \leq r \leq k$ in the enumeration of appropriate types, θ does not depend on π and $(B, <) \cong (\omega, <)$.

By the well-known correspondence of MLO and finite automata there is an ω-regular language $L_\theta \subseteq (\{0,1\}^k)^\omega$ consisting of precisely those ω-words representing the characteristic sequences of predicates $\overline{Q}$ on ω for which $(\omega, <) \models \theta(\overline{Q})$. In particular, by McNaughton's theorem, $L_\theta \in \mathbf{\Sigma}^0_3$ [22].

Consider now the mapping f assigning to each $\pi \in \{0,1\}^\omega$ the sequence $\rho \in (\{0,1\}^k)^\omega$ with $\rho[n] = \langle Q_r^\pi(\pi|_n) \mid 1 \leq r \leq k\rangle$. Note that if $\pi|_{n+1} = \pi'|_{n+1}$ then $Q_r^\pi(\pi|_n) \leftrightarrow Q_r^{\pi'}(\pi'|_n)$ for all $1 \leq r \leq k$, in other words, $\rho|_n = \rho'|_n$. Therefore f is continuous with respect to the Cantor topology. By the above, $\mathcal{X} = f^{-1}(L_\theta)$ and therefore also $\mathcal{X} \in \mathbf{\Sigma}^0_3$ as claimed. □

Theorem 6.1 was recently strengthened in [6].

7. Formalizing Condition C

The perfect set property established in Theorem 6.1 provides an MLO-definable characterization of Condition C of Lemma 4.1 over the full binary tree (with arbitrary labeling). Via interpretations, this can be extended to all (finitely branching) trees to yield the following characterization.

Proposition 7.1. (Eliminating uncountably-many-branches quantifier)
For every MLO formula $\varphi(X, \overline{Y})$ the assertion "$\exists^{\aleph_1} B$ branch$(B) \wedge \varphi(B, \overline{Y})$" is equivalent over all trees to the existence of a perfect set of branches B, each satisfying $\varphi(B, \overline{Y})$. The latter ensures that there are in fact continuum many such branches.

Proof:
Perfect sets of branches are of continuum cardinality, hence the condition is clearly sufficient. Conversely, Theorem 6.1 shows that over the full binary tree with arbitrary additional unary predicates this condition is also necessary. We can transfer this result to all trees as follows.

Every l-tree $\mathfrak{T}$ is isomorphic to some $(T, \prec, P_1, \ldots, P_l)$ where $T \subseteq \mathbb{N}^*$ is a prefix-closed subset of finite sequences of natural numbers and $\prec$ is the prefix relation. Consider the following encoding $\mu : \mathbb{N}^* \to \{0,1\}^*$

$$(n_0, n_1, \ldots, n_s) \mapsto 0^{n_0}10^{n_1}1 \ldots 0^{n_s}1,$$

and set $S = \mu(T)$ and $Q_i = \mu(P_i)$ for each $i = 1 \ldots l$. Given that $v \prec w$ in $\mathfrak{T}$ iff $\mu(v) \prec \mu(w)$ in $\mathfrak{T}(2)$, this defines an interpretation of $\mathfrak{T}$ inside $(\mathfrak{T}(2), S, Q_1, \ldots, Q_l)$. In particular, for every MLO-formula $\vartheta(\overline{X})$ of l-trees

$$\mathfrak{T} \models \vartheta(\overline{U}) \iff (\mathfrak{T}(2), S, Q_1, \ldots, Q_l) \models \vartheta^*(\mu(\overline{U})),$$

where ϑ^* is obtained from ϑ by interpreting each P_i with Q_i and relativizing all quantifiers to subsets/elements of S.

The embedding μ induces an injective mapping μ^* of the set of infinite branches of $\mathfrak{T}$ to infinite branches of $\mathfrak{T}(2)$. It is easy to check that μ^* is continuous.

Consider the formula $\varphi(B, \overline{Y})$ defining an uncountable set $\mathcal{D}$ of branches B of $\mathfrak{T}$ with parameters $\overline{V}$. Then $\mathcal{D}^* = \{\mu^*(B) \mid B \in \mathcal{D}\}$ is an uncountable set of branches of $\mathfrak{T}(2)$, which is defined by the formula "branch$(B) \wedge \exists$ infinite $P \subseteq B\ \varphi^*(P, \mu(\overline{V}))$" over $(\mathfrak{T}(2), S, Q_1, \ldots, Q_l)$. Hence, by Theorem 6.1, $\mathcal{D}^*$ contains a perfect set of branches, the inverse image of which under the continuous mapping μ^* is a perfect set of branches in $\mathcal{D}$. □

Towards an MLO formulation, note that the collection of nodes of a perfect set of branches induces a perfect tree, and vice versa. Let perfect(P) be a formula that expresses that P is a perfect subset, i.e. that P is prefix closed and for every $u \in P$ there are incomparable $v, w > u$ such that $v \in P$ and $w \in P$.

Corollary 7.1. Over trees Condition C is expressible in MLO as

$$\psi_{\mathsf{C}}(\overline{Y}) = \exists P \text{ perfect}(P) \ \wedge \ \forall B \subset P \text{ branch}(B) \rightarrow \ \exists X \ \varphi(X, \overline{Y}) \wedge \mathrm{DPATH}_{\varphi}(B, X, \overline{Y}) \, .$$

In particular, Condition C entails the existence of continuum many D-paths of sets X satisfying $\varphi(X, \overline{Y})$.

8. Summary

As we have shown above, each of the conditions of Lemma 4.1 can be formalized in MLO over trees. Thus we can again state the conclusion of this lemma: $\mathfrak{T} \models \exists^{\aleph_1} X \ \varphi(X, \overline{Y})$ holds if and only if

$$\mathfrak{T} \models \psi_{\mathsf{A}}(\overline{Y}) \ \vee \ \exists B \, (\, \psi_{\mathsf{Ba}}(B, \overline{Y}) \ \vee \ \psi_{\mathsf{Bc}}(B, \overline{Y}) \,) \ \vee \ \psi_{\mathsf{C}}(\overline{Y}).$$

Using the above, we can reduce any formula of $\mathrm{MLO}(\exists^{\aleph_1})$ to an MLO formula equivalent over the class of trees by inductively eliminating the inner-most occurrence of a cardinality quantifier. Theorem 1.1 follows. Moreover, as we have shown in the corresponding sections, each of the conditions of Lemma 4.1 implies the existence of continuum many sets X satisfying $\varphi(X, \overline{Y})$, whence Theorem 1.2.

We remark that the same technique employed here can be adapted to obtain similar results on eliminating cardinality quantifiers over several classes of linear orders, such as on the class of all ordinals and on the class of countable linear orders. These findings will appear in [4].

Acknowledgment

We are very grateful to Sasha Rubin for insightful discussions at an earlier stage of this work, and to the anonymous referee for pointed remarks that helped us improve the presentation. This research was facilitated by the ESF project AutoMathA. The first author was supported in part by ANR-06-MDCA-05 (2007-2009) DocFlow, and by the EPSRC grant EP/E010865/1. The third author was partially supported by ESF project Games and EPSRC grant EP/H018581/1.

References

[1] Arnold, A., Duparc, J., Murlak, F., Niwiński, D.: On the topological complexity of tree languages. In *Logic and Automata: History and Perspectives*, volume 2 of *Texts in Logic and Games*, pages 9–29. Amsterdam University Press, 2007.

[2] Baudisch, A., Seese, D., Tuschik, H. P., Weese, M.: *Decidability and Generalized Quantifiers*. Akademie-Verlag Berlin, 1980.

[3] Bárány, V., Kaiser, Ł., Rabinovich, A.: *Cardinality Quantifiers in MLO over Trees*. In *Proc. 18th EACSL Annual Conference on Computer Science Logic, CSL'09*, (E. Grädel and R. Kahle, Eds.), vol. 5771 of LNCS, pp. 117–132. Springer, 2009.

[4] Bárány, V., Kaiser, Ł., Rabinovich, A.: *Expressing Cardinality Quantifiers in Monadic Second-Order Logic over Chains*. to appear in J. Symb. Logic.

[5] Blumensath, A.: Automatic structures. Diploma thesis, RWTH Aachen, 1999.

[6] Bojańczyk, M., Niwiński, D., Rabinovich, A., Radziwończyk-Syta, A., Skrzypczak, M.: On the Borel Complexity of MSO Definable Sets of Branches. *Fundamenta Informaticae*, 98(4):337–349, 2010.

[7] Büchi, J. R.: Weak second-order arithmetic and finite automata. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 6:66–92, 1960.

[8] Colcombet, Th., Löding, Ch.: Transforming structures by set interpretations. *Logical Methods in Computer Science*, 3(2:4):1–36, 2007.

[9] Elgot, C. C.: Decision problems of finite automata design and related arithmetics. *Transactions of the American Mathematical Society*, 98:21–51, 1961.

[10] Gurevich, Y.: Monadic second-order theories. In Jon Barwise and Solomon Feferman, editors, *Model-Theoretical Logics*, pages 479–506. Springer, 1985.

[11] Hintikka, J.: Distributive normal forms in the calculus of predicates. *Acta Philosophica Fennica* 6 (1953).

[12] Kechris, A. S.: *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, 1995.

[13] Kuske, D., Lohrey, M.: First-order and counting theories of omega-automatic structures. *Journal of Symbolic Logic*, 73:129–150, 2008.

[14] Moschovakis, Y. N.: *Descriptive Set Theory*, volume 100 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Company, 1980.

[15] Niwiński, D.: On the cardinality of sets of infinite trees recognizable by finite automata. In *Proceedings of Mathematical Foundations of Computer Science, MFCS'91*, volume 520 of *LNCS*, pages 367–376. Springer, 1991.

[16] Niwiński, D., Walukiewicz, I.: A gap property of deterministic tree languages. *Theoretical Computer Science*, 1(303):215–231, 2003.

[17] Rabin, M. O.: Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969.

[18] Rubin, S.: Automata Presenting Structures: A Survey of the Finite String Case. *Bulletin of Symbolic Logic*, 14(2):169–209, 2008.

[19] Shelah, S.: The monadic theory of order. *Annals of Mathematics*, 102:379–419, 1975.

[20] Skurczyński J.: The Borel hierarchy is infinite in the class of regular sets of trees. *Theoretical Computer Science* 112(2): 413–418, 1993.

[21] Thomas, W.: Ehrenfeucht games, the composition method, and the monadic theory of ordinal words. In *Structures in Logic and Computer Science*, volume 1261 of *Lecture Notes in Computer Science*, pages 118–143. Springer, 1997.

[22] Thomas, W., Lescow, H.: Logical specifications of infinite computations. In J. W. de Bakker, W. P. de Roever, and G. Rozenberg, editors, *REX School/Symposium*, volume 803 of *LNCS*, pages 583–621. Springer, 1993.

[23] Trakhtenbrot, B. A.: Finite automata and the logic of monadic predicates. *Dokl. Akad. Nauk SSSR*, 140:326–329, 1961.

[24] Trakhtenbrot, B. A.: Finite automata and the logic of one-place predicates. *Sibirian Mathematical Journal*, 13:103–131, 1962. (in Russian).