7

Cardinality Quantifiers in MSO on Trees

In this chapter, we extend the results on second-order cardinality quantifiers, shown for linear orders in the previous chapter, to trees. Our main result, obtained together with Vince Bárány and Alexander Rabinovich [8, 9], is that the uncountability quantifier can be eliminated from MSO over trees.

Theorem 7.1. *For every* $\mathsf{MSO}(\exists^{\aleph_0}, \exists^{\aleph_1}, \exists^{2^{\aleph_0}})$ *formula* $\varphi(\overline{Y})$ *there exists an* MSO *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 MSO-definable families of sets. Though not surprising, this is not obvious for it is known that in MSO one can define non-analytic classes of sets [70] and that the continuum hypothesis is independent of ZFC already for co-analytic sets [67].

Theorem 7.2. *On trees* $\exists^{\aleph_1} X\, \varphi(X, \overline{Y})$ *is equivalent to* $\exists^{2^{\aleph_0}} X\, \varphi(X, \overline{Y})$ *for every* MSO *formula* $\varphi(X, \overline{Y})$*.*

Our theorems translate to generalized-automatic structures, as formulated in the corollary below. They also supersede the previously mentioned results from [59] and generalize the theorem of Niwiński [69], 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 MSO-formula $\varphi(\overline{X})$.

Corollary 7.3. *Every expansion of an injectively generalized-automatic structure by a relation definable in first-order logic with (first-order) cardinality quantifiers is also an injectively generalized-automatic structure.*

7.1 D-Nodes versus U-Nodes and Relevant Branches

To eliminate the uncountability quantifier over trees, we will again define suitable notions of U-nodes and D-nodes, similar to U-intervals and D-intervals

Ł. Kaiser: Logic and Games on Automatic Structures, LNAI 6810, pp. 95–107, 2011.

used in the previous chapter. As our main tool, we will again use the composition method, in the form of the following theorem.

Theorem 7.4 (Composition Theorem for Trees II)
Let $\varphi(\overline{X})$ be an MSO*-formula in the signature of trees with l predicates, having m free variables and quantifier rank n. Given the enumeration $\tau_1(\overline{X}), \ldots, \tau_k(\overline{X})$ of $H_{n,l+m}$, there exists an* MSO*-formula $\theta(Q_1, \ldots, Q_k)$ computable from φ such that for every tree $\mathfrak{I} = (I, <^I)$ and family $\{\mathfrak{T}_i \mid i \in I\}$ of trees and subsets $V_1, \ldots, V_m$ of $\sum_{i \in \mathfrak{I}} \mathfrak{T}_i$,*

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

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

A *tree segment*, or *interval*, of a 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 $\mathfrak{T} = \sum_{i \in \mathfrak{I}} \mathfrak{T}_i$ are tree segments of $\mathfrak{T}$. The terms 'interval' and 'tree segment' are used interchangeably.

We denote by $\mathfrak{T}|_I$ the restriction of a 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 7.1.

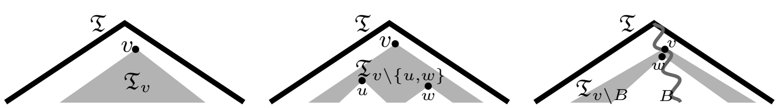

Fig. 7.1. A subtree $\mathfrak{T}_v$ and tree segments $\mathfrak{T}_{v \setminus \{u,w\}}$ and $\mathfrak{T}_{v \setminus B}$

Consider an MSO formula $\varphi(X, \overline{Y})$ over trees. To eliminate a single occurrence of the uncountability quantifier from $\exists^{\aleph_1} X\ \varphi(X, \overline{Y})$ over a tree $\mathfrak{T}$ we will make extensive use of the following notions for intervals. For the rest of this section we fix an MSO formula $\varphi(X, \overline{Y})$ over trees with l predicates and 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.

Definition 7.5. *Let $\mathfrak{T}$ be a tree, $X, \overline{Y}$ subsets of $\mathfrak{T}$ such that $\mathfrak{T} \models \varphi(X, \overline{Y})$, and I 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, $\overline{Y}$ if 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 $\varphi, X, \overline{Y}$, i.e. if $P \subseteq D(X)$.*

Again, 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 7.5 can be formalized in MSO. 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 1.10, 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

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

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

The following lemma is the first step in eliminating the $\exists^{\aleph_1}$ quantifier from MSO over trees. The three cases are depicted in Figure 7.2.

Lemma 7.6. *Let $\mathfrak{T}$ be a tree and $\varphi(X, \overline{Y})$ an* MSO*-formula. Then, for every tuple of subsets $\overline{V}$ of $\mathfrak{T}$,*

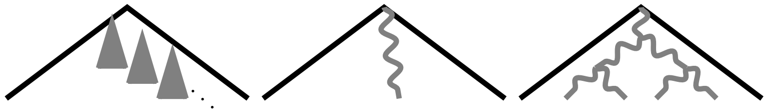

Fig. 7.2. The three conditions

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

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 Theorem 7.4 to this decomposition. Given that $\mathfrak{T} \models \varphi(U, \overline{V})$, 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, ¬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.

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 $\mathfrak{T} \models \varphi(U, \overline{V})$ if and only if there are uncountably many sets $D(U)$ with $\mathfrak{T} \models \varphi(U, \overline{V})$ if and only if Condition C holds. (The last "only if" holds because in that case each relevant $D(U)$ contains only finitely many branches.) □

We remark that Lemma 7.6 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 $\mathrm{Inf}(X)$, meaning that the set X is infinite, cannot be expressed in pure MSO.

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

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

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

7.2 Structure of D-Paths for Uncountably Many Sets

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 MSO 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 MSO formula $\varphi(X, \overline{Y})$ with $1 + m$ 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 a 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, assuming that l labels occur in $\mathfrak{T}$ (and φ), we fix r as the number of $\mathrm{qr}(\psi)$-types in $l + m + 2$ variables, which we enumerate as $\tau_1, \ldots, \tau_r$. Then Theorem 7.4 yields a formula θ such that

$$\mathfrak{T} \models \psi(X, \overline{Y}, B) \quad \Longleftrightarrow \quad (B, <) \models \theta(P_1, \ldots, P_r) \tag{7.1}$$

with $P_i = \{w \in B \mid (\mathfrak{T}_{w \setminus B}, X, \overline{Y}, \{w\}) \models \tau_i\}$ for each $i \in \{1, \ldots, r\}$. 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 a 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 7.7. *Let $\mathfrak{T}$ be a 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 $\mathfrak{T}$ if and only if 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 $i \in \{1, \ldots, r\}$.

Bc. *It holds that*

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

where for each $i \in \{1, \ldots, r\}$, 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 (7.1) we have $\mathfrak{T} \models \psi(X, \overline{Y}, B)$ if and only if $(B, <) \models \theta(P_1, \ldots, P_r)$. 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 $i \in \{1, \ldots, r\}$.
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 $\mathfrak{T}_{w\setminus B}$. Therefore 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 (7.1) 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 (7.1) 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 (7.1) from $(B,<) \models \theta(\overline{P})$ we can infer that $\mathfrak{T} \models \psi(X_{\overline{P}}, \overline{Y}, B)$. 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 7.8. *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 7.6 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 $\lambda^X(w) = \mathrm{Tp}^n(\mathfrak{T}_{w\setminus D}, X, \overline{Y})$ 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 $\lambda^X(w) = \mathrm{Tp}^n(\mathfrak{T}_{w\setminus B}, X, \overline{Y})$ 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 $\overline{P}^X = \langle P_1^X, \ldots P_r^X\rangle$ of B

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

with $D(X) = D$ and X satisfying φ. Using (7.1) 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 $X \cap \mathfrak{T}_{w\setminus D}$ 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 $\mathfrak{T}_{w\setminus D}$. However, by definition, every subtree of $\mathfrak{T}_{w\setminus D}$ is a U-tree relative to each of these X, because $D(X) = D$. Because $\mathfrak{T}$ is finitely branching, i.e. $\mathfrak{T}_{w\setminus D} \setminus \{w\}$ is a finite union of such U-trees, there can be only finitely many X as above and pairwise differing on $\mathfrak{T}_{w\setminus D}$, which is a contradiction. Therefore Condition (Ba) must hold. □

Next we will construct MSO 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 7.6.

7.2.1 Formalization of Condition Ba

Much like Condition A, (Ba) is naturally expressible in MSO(Inf) and thus, over trees, in pure MSO as well by the formula

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

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

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

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 7.6, because the set X can be left intact or changed to another one with the same type on any of the infinitely many trees $\mathfrak{T}_{w \setminus B}$ which are D-intervals for X.

7.2.2 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 [59], more directly proven in the previous chapter, which states that cardinality quantifiers can be eliminated over $(\omega, <)$.

Proposition 7.9. *For every* MSO *formula* $\varphi(\overline{X}, \overline{Y})$ *there exists an effectively constructable 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 if and only if $(B, <) \models \vartheta(\overline{Q})$, with $\overline{Q}$ as specified there.

By Proposition 7.9, 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 MSO 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

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

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

7.3 The Full Binary Tree and the Cantor Space

In order to formalize Condition C in MSO 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 MSO 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.

7.3.1 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 [67], 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 MSO-definable tree languages and the complexity of tree-automata recognizing them is well understood [85, 70]. By Rabin's complementation theorem, all MSO-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 [3, 70]. There are regular tree languages on arbitrary finite levels of the Borel hierarchy [81].

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. MSO-definable sets of ω-words, which are, by McNaughton's theorem, Boolean combinations of $\mathbf{\Pi}^0_2$ sets [85].

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 7.10 ([52, 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 [67]. 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 [67]. A key observation that our formalization will exploit is that, even though there are non-analytic sets of trees definable in MSO, sets of definable paths are Borel.

7.3.2 Definable Sets of Branches are Borel

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 can be identified with the set of prefixes of π. The following theorem was recently strengthened in [15].

Theorem 7.11 (MSO 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 MSO 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 $Q^\pi_r = \{v \in B \mid \mathrm{Tp}^n(\mathfrak{T}^\pi_{v\setminus B}) = \tau_r\}$ for each $r \in \{1, \ldots, k\}$ in the enumeration of appropriate types. Note that θ does not depend on π and $(B, <) \cong (\omega, <)$.

By the well-known correspondence of MSO 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 holds $(\omega,<) \models \theta(\overline{Q})$. In particular, by McNaughton's theorem, $L_\theta \in \mathbf{\Sigma}^0_3$ [85].

Consider now the mapping f assigning to each $\pi \in \{0,1\}^\omega$ the sequence $\rho \in (\{0,1\}^k)^\omega$ with $\rho[n] = \langle Q^\pi_r(\pi|_n) \mid r \in \{1,\dots,k\}\rangle$. Note that if $\pi|_{n+1} = \pi'|_{n+1}$ then $Q^\pi_r(\pi|_n) \leftrightarrow Q^{\pi'}_r(\pi'|_n)$ for all $r \in \{1,\dots,k\}$, 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. □

7.4 Formalizing Existence of Uncountably Many Branches

The perfect set property established in Theorem 7.11 provides an MSO-definable characterization of Condition C of Lemma 7.6 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 7.12 (Eliminating uncountably-many-branches quantifier). *For every* MSO *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 7.11 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 tree $\mathfrak{T}$ is isomorphic to some $(T,\prec,P_1,\dots,P_l)$ 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 $\mu : \mathbb{N}^* \to \{0,1\}^*$

$$(n_0,n_1,\dots,n_s) \mapsto 0^{n_0}10^{n_1}1\dots0^{n_s}1,$$

and set $S = \mu(T)$ and $Q_i = \mu(P_i)$ for each $i = 1\dots l$.

Given that $v \prec w$ in $\mathfrak{T}$ if and only if $\mu(v) \prec \mu(w)$ in $\mathfrak{T}(2)$, this defines an interpretation of $\mathfrak{T}$ inside $(\mathfrak{T}(2),S,Q_1,\dots,Q_l)$. In particular, for every MSO-formula $\vartheta(\overline{X})$ over trees with l predicates,

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

where ϑ^* is obtained from ϑ by interpreting each P_i with Q_i and relativizing all quantifiers to subsets or 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 7.11, $\mathcal{D}^*$ contains a perfect set of branches. The inverse image of this set under the continuous mapping μ^* is a perfect set of branches in $\mathcal{D}$. □

Towards an MSO 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 7.13. *Over trees, Condition* C *is expressible in* MSO *as*

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

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

As we have shown above, each of the conditions of Lemma 7.6 can be formalized in MSO 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 MSO($\exists^{\aleph_1}$) to an MSO formula equivalent over the class of trees by inductively eliminating the inner-most occurrence of a cardinality quantifier. Theorem 7.1 follows. Moreover, as we have shown in the corresponding sections, each of the conditions of Lemma 7.6 implies the existence of continuum many sets X satisfying $\varphi(X,\overline{Y})$, thus Theorem 7.2 follows as well.