Hilbert's Tenth Problem for Term Algebras with a Substitution Operator

Juvenal Murwanashyaka

Department of Mathematics, University of Oslo, Norway
juvenalm@math.uio.no

Abstract. We introduce a first-order theory of finite full binary trees and show that the analogue of Hilbert's Tenth Problem is undecidable by constructing a many-to-one reduction of Post's Correspondence Problem.

1 Introduction

Hilbert's Tenth Problem asks whether there exists an algorithm that given a polynomial $f \in \mathbb{Z}[x_1, x_2, \dots, x_n]$ decides whether f has a zero in $\mathbb{Z}^n$. In 1970, Yuri Matiyasevich proved that Hilbert's Tenth Problem is undecidable by showing that the exponential function is existentially definable in terms of addition and multiplication (see for example Davis [1]). After this, a standard technique for showing that a structure has undecidable existential theory has been to show that it existentially interprets the first-order structure of arithmetic $(\mathbb{N}, 0, 1, +, \times)$ (see sections 5.3 and 5.4a of Hodges [2] for more details). In this paper, we introduce a first-order structure $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ of finite full binary trees (see Section 2) and prove that the analogue of Hilbert's Tenth Problem for $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ is undecidable without interpreting arithmetic, that is, without relying on the solution to Hilbert's Tenth Problem (such a proof can also be produced by modifying slightly the coding in Section 5 to translate multiplication).

2 Preliminaries

We consider the first-order language $\mathcal{L}_{\mathsf{BT}} = \{\perp, \langle \cdot, \cdot \rangle, \cdot[\cdot \mapsto \cdot]\}$ where $\perp$ is a constant symbol, $\langle \cdot, \cdot \rangle$ is a binary function symbol and $\cdot[\cdot \mapsto \cdot]$ is a ternary function symbol. The intended structure $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ is a term model: The universe $\mathbf{H}$ is the set of all variable-free terms in the language $\{\perp, \langle \cdot, \cdot \rangle\}$ (equivalently, finite full binary trees). The constant symbol $\perp$ is interpreted as itself. The function symbol $\langle \cdot, \cdot \rangle$ is interpreted as the function that maps the pair (s, t) to the term $\langle s, t \rangle$. The function symbol $\cdot[\cdot \mapsto \cdot]$ is interpreted as a term substitution operator: $t[r \mapsto s]$ is the term we obtain by replacing each occurrence of r in t with s. We define $t[r \mapsto s]$ by recursion as follows: If $t = r$, then $t[r \mapsto s] = s$. If $r \neq \perp$, then $\perp [r \mapsto s] = \perp$. If $r \neq t = \langle t_1, t_2 \rangle$, then $t[r \mapsto s] = \big\langle t_1[r \mapsto s]\, , \, t_2[r \mapsto s] \big\rangle$.

To improve readability, it will occasionally be more convenient to represent finite binary trees using notation that is closer to their visual form: By recursion, for $n \geq 2$, let $\langle x_1, \dots, x_n, x_{n+1} \rangle$ be shorthand for $\langle \langle x_1, \dots, x_n \rangle \, , \, x_{n+1} \rangle$. By recursion, let $\perp^1 = \perp$ and $\perp^{n+1} = \langle \perp^n, \perp \rangle$.

We let $\mathsf{Th}^{\exists}(\mathcal{T}(\mathcal{L}_{\mathsf{BT}}))$ denote the set of all existential $\mathcal{L}_{\mathsf{BT}}$-sentences that are true in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$. We let $\mathsf{Th}^{\mathsf{H10}}(\mathcal{T}(\mathcal{L}_{\mathsf{BT}}))$ denote the set of all $\mathcal{L}_{\mathsf{BT}}$-sentences of the form $\exists \vec{x}\ [\ s = t\ ]$ that are true in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$. In Section 7, we prove that $\mathsf{Th}^{\exists}(\mathcal{T}(\mathcal{L}_{\mathsf{BT}}))$ is undecidable by constructing a reduction of Post's correspondence problem. The coding techniques that form the basis of the encoding are developed in sections 3, 4, 5, 6. In Section 8, we show that undecidability of $\mathsf{Th}^{\exists}(\mathcal{T}(\mathcal{L}_{\mathsf{BT}}))$ implies undecidability of $\mathsf{Th}^{\mathsf{H10}}(\mathcal{T}(\mathcal{L}_{\mathsf{BT}}))$.

Definition 1. *Let $\{0,1\}^+$ denote the set of all nonempty binary strings.* The Post Correspondence Problem (PCP) *is given by*

- *Instance: a list of pairs $\langle a_1, b_1\rangle, \ldots, \langle a_n, b_n\rangle$ where $a_i, b_i \in \{0,1\}^+$*
- *Solution: a finite nonempty sequence $i_1, ..., i_m$ of indexes such that we have the equality $a_{i_1} a_{i_2} \ldots a_{i_m} = b_{i_1} b_{i_2} \ldots b_{i_m}$.*

To analyze further what we can and cannot effectively decide over $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$, we introduce bounded quantifiers. We let $x \sqsubseteq t$ and $x \not\sqsubseteq t$ be shorthand for $t[\,x \mapsto \langle x, x\rangle\,] \neq t$ and $t[\,x \mapsto \langle x, x\rangle\,] = t$, respectively. Observe that $\sqsubseteq$ is the subtree relation on finite binary trees. In [5], Venkataraman shows that the existential theory of the structure we obtain by taking $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ and replacing the substitution operator with the subtree relation is decidable and the decision problem is NP-complete. Let $\forall x \sqsubseteq t\ \phi$ be shorthand for $\forall x\ [\ x \sqsubseteq t \rightarrow \phi\ ]$. Let $\Sigma_{1,0,1}^{\mathcal{T}(\mathcal{L}_{\mathsf{BT}})}$ denote the set of all $\mathcal{L}_{\mathsf{BT}}$-sentences that are true in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ and are of the form $\exists x\ \forall y \sqsubseteq x\ \phi$ where ϕ is quantifier-free. In Section 9, we show that $\Sigma_{1,0,1}^{\mathcal{T}(\mathcal{L}_{\mathsf{BT}})}$ is undecidable. We cannot prove this result by encoding Post's correspondence problem since this problem is about sequences of pairs and therefore necessitates the use of two bounded universal quantifiers. Instead, we encode the Modulo Problem of Kristiansen & Murwanashyaka [3].

Definition 2. *Let $f^0(x) = x$ and $f^{n+1}(x) = f(f^n(x))$.* The Modulo Problem *is given by*

- *Instance: a list of pairs $\langle A_0, B_0\rangle, \ldots, \langle A_{M-1}, B_{M-1}\rangle$ where $M > 1$ and $A_i, B_i \in \mathbb{N}$ for $i = 0, \ldots, M-1$.*
- *Solution: a natural number N such that $f^N(3) = 2$ where $f(x) = A_j z + B_j$ if there exists $j \in \{0, 1, \ldots, M-1\}$ such that $x = Mz + j$.*

3 Numbers

To encode Post's correspondence problem, we need to associate strings over a finite alphabet with finite binary trees. As a step towards this, we show that certain classes of number-like objects are existentially definable in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$.

Definition 3. *Let $\alpha \in \mathbf{H}$. Let $s_1, \ldots, s_n \in \mathbf{H}$ be such that α is not a subtree of s_i for all $i \leq n$ and $s_n \neq s_j$ for all $j < n$. Let*

$$\frac{1}{\alpha, \vec{s}} \equiv \langle \alpha, s_1, \ldots, s_n\rangle \quad \textit{and} \quad \frac{m+1}{\alpha, \vec{s}} \equiv \frac{1}{\alpha, \vec{s}}\big[\,\alpha \mapsto \frac{m}{\alpha, \vec{s}}\,\big]\ .$$

Let $\mathbb{N}_{\vec{s}}^{\alpha} = \{\, \frac{m}{\alpha, \vec{s}} \in \mathbf{H} :\ m \in \mathbb{N}\ \wedge\ m \geq 1\ \}$.

Lemma 1. *Let $\alpha \in \mathbf{H}$. Let $s_1, \dots, s_n \in \mathbf{H}$ be such that α is not a subtree of s_i for all $i \leq n$ and $s_n \neq s_j$ for all $j < n$. Then, for all $T \in \mathbf{H}$*

$$T \in \mathbb{N}^{\alpha}_{\vec{s}} \;\Leftrightarrow\; T = \frac{1}{\alpha, \vec{s}} \vee \Big(\frac{2}{\alpha, \vec{s}} \sqsubseteq T \wedge T = \frac{1}{\alpha, \vec{s}}\Big[\, \alpha \mapsto T\Big[\, \frac{2}{\alpha, \vec{s}} \mapsto \frac{1}{\alpha, \vec{s}}\Big]\,\Big] \Big) \,.$$

Proof. The left-right implication of the claim is straightforward. Let the size of a binary tree T be the number of nodes in T. We prove by induction on the size of T that

$$T \;=\; \frac{1}{\alpha, \vec{s}} \;\;\vee\;\; \Big(\; \frac{2}{\alpha, \vec{s}} \sqsubseteq T \;\;\wedge\;\; T \;=\; \frac{1}{\alpha, \vec{s}}\Big[\; \alpha \;\mapsto\; T\Big[\; \frac{2}{\alpha, \vec{s}} \;\mapsto\; \frac{1}{\alpha, \vec{s}}\Big] \;\Big] \;\Big) \qquad (*)$$

implies $T \in \mathbb{N}^{\alpha}_{\vec{s}}$.

Assume T satisfies (*). We need to show that $T \in \mathbb{N}^{\alpha}_{\vec{s}}$. If $T = \frac{1}{\alpha,\vec{s}}$, then certainly $T \in \mathbb{N}^{\alpha}_{\vec{s}}$. Otherwise, by the second disjunct in (*), we have $\frac{2}{\alpha,\vec{s}} \sqsubseteq T$. Let $S = T[\, \frac{2}{\alpha,\vec{s}} \;\mapsto\; \frac{1}{\alpha,\vec{s}}]$. Then, S is strictly smaller than T. By the second disjunct in (*), we have $T = \frac{1}{\alpha,\vec{s}}[\; \alpha \;\mapsto\; S \;]$. By Definition 3, $\frac{1}{\alpha,\vec{s}} = \langle \alpha, s_1, \dots, s_n \rangle$. Since α is not a subtree of any s_i

$$T = \frac{1}{\alpha, \vec{s}}\big[\; \alpha \;\mapsto\; S \;\big] = \langle \alpha, s_1, \dots, s_n \rangle\big[\; \alpha \;\mapsto\; S \;\big] = \langle S, s_1, \dots, s_n \rangle \;. \qquad (**)$$

We know that $\frac{2}{\alpha,\vec{s}} \sqsubseteq T$. By Definition 3, $\frac{2}{\alpha,\vec{s}} = \langle \alpha, s_1, \dots, s_n, s_1, \dots, s_n \rangle$. Since $s_n \neq s_j$ for all $1 \leq j < n$, it follows from $\frac{2}{\alpha,\vec{s}} \sqsubseteq T$ and (**) that we have one of the following cases: (i) $S = \frac{1}{\alpha,\vec{s}}$, (ii) occurrences of $\frac{2}{\alpha,\vec{s}}$ in T can only be found in S. In case of (ii), we have

$$S = T\big[\, \tfrac{2}{\alpha,\vec{s}} \mapsto \tfrac{1}{\alpha,\vec{s}}\big] = \langle S, s_1, \dots, s_n \rangle\big[\, \tfrac{2}{\alpha,\vec{s}} \mapsto \tfrac{1}{\alpha,\vec{s}}\big] = \Big\langle S\big[\, \tfrac{2}{\alpha,\vec{s}} \mapsto \tfrac{1}{\alpha,\vec{s}}\big] \,, s_1, \dots, s_n \Big\rangle$$

$$= \langle \alpha, s_1, \dots, s_n \rangle\Big[\; \alpha \;\mapsto\; S\big[\, \tfrac{2}{\alpha,\vec{s}} \mapsto \tfrac{1}{\alpha,\vec{s}}\big] \;\Big] = \tfrac{1}{\alpha,\vec{s}}\Big[\; \alpha \mapsto S\big[\, \tfrac{2}{\alpha,\vec{s}} \mapsto \tfrac{1}{\alpha,\vec{s}}\big] \;\Big] \;.$$

We see that in case of either (i) or (ii), S satisfies (*). Thus, by the induction hypothesis, $S \in \mathbb{N}^{\alpha}_{\vec{s}}$. It then follows from (**) that $T \in \mathbb{N}^{\alpha}_{\vec{s}}$. □

4 Strings

Given a finite alphabet $A = \{a_1, \dots, a_m\}$, let ε denote the empty string and let A^* denote the set of all finite strings over A. Let $A^+ = A^* \setminus \{\varepsilon\}$. We will now associate A^* with an existentially definable class of finite binary trees.

Definition 4. *Let $A = \{a_1, \dots, a_m\}$ be a finite alphabet. For each natural number $i \geq 1$, let $\mathsf{g}_i \equiv \langle \bot^{3+i}, \bot^{3+i} \rangle$. Let $\alpha \in \mathbf{H}$ be incomparable with g_i with respect to the subtree relation for all i. We define a one-to-one map $\tau_\alpha : A^* \to \mathbf{H}$ by recursion*

$$\tau_\alpha(w) = \begin{cases} \alpha & \textit{if } w = \varepsilon \\ \langle \alpha, \mathsf{g}_i \rangle & \textit{if } w = a_i \\ \tau_\alpha(w_0)\big[\; \alpha \;\mapsto\; \tau_\alpha(w_1) \;\big] & \textit{if } w = w_0 w_1 \;\; \textit{and} \;\; w_0 \in A \;. \end{cases}$$

Given $s \in A^*$, we write $\frac{s}{\alpha}$ for $\tau_\alpha(s)$. Furthermore, we write a_i for g_i.

For example, $\frac{a_1 a_1 a_3 a_1 a_2}{\alpha} = \langle\, \alpha \,,\, a_2 \,,\, a_1 \,,\, a_3 \,,\, a_1 \,,\, a_1 \,\rangle$.

Lemma 2. *Let $A = \{a_1, \ldots, a_m\}$ be a finite alphabet. Then, $\tau_\alpha(A^*)$ is existentially definable in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$.*

Proof. We need the following property to prove that $\tau_\alpha(A^*)$ is existentially definable

(*) $\mathsf{g}_1, \ldots, \mathsf{g}_m$ are incomparable with respect to the subtree relation.

Lemma 1 tells us that the classes $\mathbb{N}^\alpha_{\mathsf{g}_i} \cup \{\alpha\}$ are existentially definable in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$. The idea is to show that $s \in \tau_\alpha(A^*)$ if and only if we can transform s into an element of $\mathbb{N}^\alpha_{\mathsf{g}_i} \cup \{\alpha\}$. We show that $\tau_\alpha(A^*)$ is defined by the formula $\phi(x) \equiv\ x[\,\mathsf{g}_2 \mapsto \mathsf{g}_1\,] \ldots [\,\mathsf{g}_m \mapsto \mathsf{g}_1\,] \in \mathbb{N}^\alpha_{\mathsf{g}_1} \cup \{\alpha\}$.

Clearly, each element in $\tau_\alpha(A^*)$ has the property $\phi(x)$. To see that the converse holds, assume $\phi(s)$. We need to show that $s \in \tau_\alpha(A^*)$. Since $\mathbb{N}^\alpha_{\mathsf{g}_1} \cup \{\alpha\} \subseteq \tau_\alpha(A^*)$, it suffices to show that for each $1 \leq i \leq n$ and each finite binary tree t, if $t[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] \in \tau_\alpha(A^*)$, then $t \in \tau_\alpha(A^*)$. We prove this by induction on the size of t.

Assume $t[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] \in \tau_\alpha(A^*)$. We need to show that $t \in \tau_\alpha(A^*)$. If g_i is not a subtree of t, then $t = t[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] \in \tau_\alpha(A^*)$. Assume now g_i is a subtree of t. Let $t = \langle t_0, t_1 \rangle$. We cannot have $t = \mathsf{g}_i$ since $\mathsf{g}_1 \notin \tau_\alpha(A^*)$. Hence, $t[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] = \big\langle t_0[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,], t_1[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,]\big\rangle$. By how the elements of $\tau_\alpha(A^*)$ are defined, $t_0[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] \in \tau_\alpha(A^*)$ and $t_1[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] = \mathsf{g}_j$ for some $1 \leq j \leq n$. Since $t_0[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] \in \tau_\alpha(A^*)$, by the induction hypothesis, $t_0 \in \tau_\alpha(A^*)$. If g_i is not a subtree of t_1, then $t_1 = t_1[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] = \mathsf{g}_j$. Assume now g_i is a subtree of t_1. Then, g_1 is a subtree of g_j since $t_1[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] = \mathsf{g}_j$. By (*), $\mathsf{g}_1 = \mathsf{g}_j$, which implies $t_1 = \mathsf{g}_i$. Hence, $t_0 \in \tau_\alpha(A^*)$ and $t_1 = \mathsf{g}_l$ for some $1 \leq l \leq n$. Then, $t = \langle t_0, t_1 \rangle \in \tau_\alpha(A^*)$ by how the elements of $\tau_\alpha(A^*)$ are defined.

Thus, by induction, if $t[\,\mathsf{g}_i \mapsto \mathsf{g}_1\,] \in \tau_\alpha(A^*)$, then $t \in \tau_\alpha(A^*)$. □

5 Sequences of Strings I

Recall that the instance $\langle a_1, b_1 \rangle, \ldots, \langle a_n, b_n \rangle$ of PCP has a solution if and only if there exist a finite nonempty sequence $i_1, ..., i_m$ of indexes such that we have $a_{i_1} a_{i_2} \ldots a_{i_m} = b_{i_1} b_{i_2} \ldots b_{i_m}$. So, given a finite sequence $C = \langle c_1, c_2, \ldots, c_n \rangle$ of nonempty binary strings, we need to express that a sequence $w_1, w_2, \ldots, w_k$ of binary strings satisfies the following two properties: (A) there exists $i \in \{1, \ldots, n\}$ such that $w_1 = c_i$, (B) for all $j \in \{1, \ldots, k-1\}$ there exists $i \in \{1, \ldots, n\}$ such that $w_{j+1} = w_j c_i$. In other words, we need to give an existential definition of the class $\mathbb{P}(C)$ of all sequences $w_1, w_2, \ldots, w_k$ that satisfy (A)-(B). In this section, we give a formal definition of $\mathbb{P}(C)$, as a class of finite binary trees, and show that it is existentially definable.

Since we are interested in describing sequences that satisfy (A)-(B), it is not the set $\{0, 1\}^*$ we are interested in, but rather the subset generated by

$\{c_1, c_2, \ldots, c_n\}$ under concatenation. We also need to treat the c_i's as distinct objects since we intend to replace C with one of the sequences $\langle a_1, \ldots, a_n\rangle$, $\langle b_1, \ldots, b_n\rangle$ where $\langle a_1, b_1\rangle, \ldots, \langle a_n, b_n\rangle$ is an instance of PCP. To capture this, we associate elements of $\{c_1, c_2, \ldots, c_n\}^+$ with strings over a larger alphabet $\{0, 1, \mu_1, \mu_2, \ldots, \mu_n\}$ where μ_i represents the last letter of c_i. Assume for example $c_1 = 110$, $c_2 = 011$ and $c_3 = 1010$. Then, we associate the binary string $c_2c_1c_3$ with the string $01\mu_2 11\mu_1 101\mu_3$.

Definition 5. *Let $C = \langle c_1, c_2, \ldots, c_n\rangle$ be a sequence of nonempty binary strings. We associate c_i with a finite binary tree in $\tau_\alpha(\{0, 1, \mu_1, \ldots, \mu_n\}^*)$ as follows*

$$\frac{c_i}{C, \alpha} \equiv \frac{w_i\mu_i}{\alpha} \quad \textit{where} \quad c_i = w_i d \ \wedge\ w_i \in \{0, 1\}^* \ \wedge\ d \in \{0, 1\} \ .$$

We let $\frac{\varepsilon}{C,\alpha} \equiv \alpha$. We associate the string $c_{i_1}c_{i_2}\ldots c_{i_m}$ with a finite binary tree in $\tau_\alpha(\{0, 1, \mu_1, \ldots, \mu_n\}^*)$ as follows

$$\frac{c_{i_1}c_{i_2}\ldots c_{i_m}}{C, \alpha} \equiv \frac{w_{i_1}\mu_{i_1}w_{i_2}\mu_{i_2}\ldots w_{i_m}\mu_{i_m}}{\alpha} \ .$$

We are finally ready to give a formal definition of the class of those finite binary trees that encode sequences that satisfy (A)-(B).

Definition 6. *Let $C = \langle c_1, c_2, \ldots, c_n\rangle$ be a sequence of nonempty binary strings. Let $\alpha, \gamma \in \mathbf{H}$ be incomparable with respect to the subtree relation. Assume α also satisfies the condition in Definition 4. Let $\mathbb{P}(C, \alpha, \gamma)$ be the smallest subset of $\mathbf{H}$ that satisfies*

- *$\langle \gamma\, ,\, \frac{c_i}{C,\alpha}\rangle \in \mathbb{P}(C, \alpha, \gamma)$ for all $i \in \{1, \ldots, n\}$*
- *if $T \in \mathbb{P}(C, \alpha, \gamma)$ where $T = \big\langle R\, ,\, \frac{c_{i_1}c_{i_2}\ldots c_{i_m}}{C,\alpha}\big\rangle$, then $\big\langle\, T\, ,\, \frac{c_{i_1}c_{i_2}\ldots c_{i_m}c_j}{C,\alpha}\,\big\rangle \in \mathbb{P}(C, \alpha, \gamma)$ for all $j \in \{1, \ldots, n\}$.*

Lemma 3. *Let $C = \langle c_1, c_2, \ldots, c_n\rangle$ be a sequence of nonempty binary strings. Let $\alpha, \gamma \in \mathbf{H}$ be incomparable with respect to the subtree relation. Assume α also satisfies the condition in Definition 4. Let $\delta = \langle \alpha, \alpha\rangle$. Let $F^\alpha_\delta(L) = L[\ \alpha \mapsto \delta\ ]$ for all $L \in \mathbf{H}$. Let $T \in \mathbf{H}$. Then, $T \in \mathbb{P}(C, \alpha, \gamma)$ if and only if*

(1) *$\delta \not\sqsubseteq T$*
(2) *there exists $m \in \{1, \ldots, n\}$ such that $\big\langle \gamma\, ,\, \frac{c_m}{C,\alpha}\big\rangle \sqsubseteq T$*
(3) *there exists $S \in \tau_\alpha(\{0, 1, \mu_1, \ldots, \mu_n\}^*)$ such that*

$$T = \Big\langle F^\alpha_\delta(T)\Big[\Big\langle \gamma\, ,\, \frac{c_m}{C, \delta}\Big\rangle \mapsto \gamma\ ,\ \frac{c_1}{C, \delta} \mapsto \alpha\ ,\ \ldots\ ,\ \frac{c_n}{C, \delta} \mapsto \alpha\Big]\ ,\ S\ \Big\rangle\ .$$

Before we prove the lemma, we illustrate why the left-right implication holds. First, observe that (1) holds if $T \in \mathbb{P}(C, \alpha, \gamma)$. Now, assume for example $T = \Big\langle\ \gamma\, ,\, \frac{c_2}{C,\alpha}\, ,\, \frac{c_2c_3}{C,\alpha}\, ,\, \frac{c_2c_3c_1}{C,\alpha}\ \Big\rangle$. The tree $F^\alpha_\delta(T)$ is just the tree we obtain by replacing each one of the three occurrences of α in T with δ. Hence, $F^\alpha_\delta(T) =$

$\Big\langle\ \gamma\,,\ \frac{c_2}{C,\delta}\,,\ \frac{c_2 c_3}{C,\delta}\,,\ \frac{c_2 c_3 c_1}{C,\delta}\ \Big\rangle$. Since there is only one occurrence of $\langle\gamma\,,\ \frac{c_2}{C,\delta}\rangle$ in $F^{\alpha}_{\delta}(T)$, we have $R_0 := F^{\alpha}_{\delta}(T)\big[\ \ \langle\gamma\,,\ \frac{c_2}{C,\delta}\rangle\ \mapsto\ \gamma\ \big] = \big\langle\ \gamma\,,\ \frac{c_2 c_3}{C,\delta}\,,\ \frac{c_2 c_3 c_1}{C,\delta}\ \big\rangle$. We replace the one occurrence of $\frac{c_1}{C,\delta}$ in R_0 and obtain $R_1 := R_0\big[\ \frac{c_1}{C,\delta}\ \mapsto\ \alpha\ \big] = \big\langle\ \gamma\,,\ \frac{c_2 c_3}{C,\delta}\,,\ \frac{c_2 c_3}{C,\alpha}\ \big\rangle$. Since $\frac{c_2 c_3}{C,\alpha}$ does not contain a subtree of the form $\frac{c_i}{C,\delta}$ by the choice of δ, there is no occurrence of $\frac{c_2}{C,\delta}$ in R_1. Hence, $R_2 := R_1\big[\ \frac{c_2}{C,\delta}\ \mapsto\ \alpha\ \big] = R_1$. We replace the occurrence of $\frac{c_3}{C,\delta}$ in R_2 and obtain $R_3 := R_2\big[\ \frac{c_3}{C,\delta}\ \mapsto\ \alpha\ \big] = \big\langle\ \gamma\,,\ \frac{c_2}{C,\alpha}\,,\ \frac{c_2 c_3}{C,\alpha}\ \big\rangle$. Now, observe that R_3 is the left subtree of T.

Proof (Proof of Lemma 3).

The left-right implication is obvious. We prove right-left implication by induction on the size of T. We need the following properties:

(A) Since γ and α are incomparable with respect to the subtree relation, the binary tree $\langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle$ is not a subtree of elements of $\tau_{\alpha}(\{0,1,\mu_1,\dots,\mu_n\}^*)$.

(B) Since γ and δ are incomparable with respect to the subtree relation, the binary tree $\langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle$ is not a subtree of elements of $\tau_{\delta}(\{0,1,\mu_1,\dots,\mu_n\}^*)$.

Assume T satisfies (1)-(3). We need to show that $T \in \mathbb{P}(C,\alpha,\gamma)$. By assumption, we have a natural number $m \in \{1,\dots,n\}$ and a string $s \in \{0,1,\mu_1,\dots,\mu_n\}^*$ such that the following three properties hold: (i) $\delta \not\sqsubseteq T$, (ii) $\langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle \sqsubseteq T$, (iii) $T = \Big\langle F^{\alpha}_{\delta}(T)\Big[\langle\gamma\,,\ \frac{c_m}{C,\delta}\rangle\ \mapsto\ \gamma\ ,\ \frac{c_1}{C,\delta}\ \mapsto\ \alpha\ ,\ \dots\ ,\ \frac{c_n}{C,\delta}\ \mapsto\ \alpha\Big]\ ,\ \frac{s}{\alpha}\ \Big\rangle$. Let $T_0 = F^{\alpha}_{\delta}(T)\Big[\langle\gamma\,,\ \frac{c_m}{C,\delta}\rangle\ \mapsto\ \gamma\ ,\ \frac{c_1}{C,\delta}\ \mapsto\ \alpha\ ,\ \dots\ ,\ \frac{c_n}{C,\delta}\ \mapsto\ \alpha\Big]$.

Assume $T_0 = \gamma$. By (ii), $\langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle \sqsubseteq T$. By (A), $\langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle \not\sqsubseteq \frac{s}{\alpha}$. Hence, $T = \langle T_0\,,\ \frac{s}{\alpha}\rangle = \langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle \in \mathbb{P}(C,\alpha,\gamma)$.

Assume now $T_0 \neq \gamma$. Since $T_0 \sqsubseteq T$, it follows from (i) that $\delta \not\sqsubseteq T_0$. Since $\langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle \sqsubseteq T$, $T \neq \langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle$ and $\langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle \not\sqsubseteq \frac{s}{\alpha}$, we have $\langle\gamma\,,\ \frac{c_m}{C,\alpha}\rangle \sqsubseteq T_0$. Finally, we have

$$\begin{aligned} T_0 &= F^{\alpha}_{\delta}(T)\Big[\langle\gamma\,,\ \tfrac{c_m}{C,\delta}\rangle\ \mapsto\ \gamma\ ,\ \tfrac{c_1}{C,\delta}\ \mapsto\ \alpha\ ,\ \dots\ ,\ \tfrac{c_n}{C,\delta}\ \mapsto\ \alpha\Big] \\ &= F^{\alpha}_{\delta}\big(\langle T_0\,,\ \tfrac{s}{\alpha}\rangle\big)\Big[\langle\gamma\,,\ \tfrac{c_m}{C,\delta}\rangle\ \mapsto\ \gamma\ ,\ \tfrac{c_1}{C,\delta}\ \mapsto\ \alpha\ ,\ \dots\ ,\ \tfrac{c_n}{C,\delta}\ \mapsto\ \alpha\Big] \\ &= \langle F^{\alpha}_{\delta}(T_0)\,,\ \tfrac{s}{\delta}\rangle\Big[\langle\gamma\,,\ \tfrac{c_m}{C,\delta}\rangle\ \mapsto\ \gamma\ ,\ \tfrac{c_1}{C,\delta}\ \mapsto\ \alpha\ ,\ \dots\ ,\ \tfrac{c_n}{C,\delta}\ \mapsto\ \alpha\Big] \\ &= \Big\langle F^{\alpha}_{\delta}(T_0)\Big[\langle\gamma\,,\ \tfrac{c_m}{C,\delta}\rangle\ \mapsto\ \gamma\ ,\ \tfrac{c_1}{C,\delta}\ \mapsto\ \alpha\ ,\ \dots\ ,\ \tfrac{c_n}{C,\delta}\ \mapsto\ \alpha\Big]\ ,\ S_0\ \Big\rangle \end{aligned}$$

where

$$\begin{aligned} S_0 &= \tfrac{s}{\delta}\Big[\langle\gamma\,,\ \tfrac{c_m}{C,\delta}\rangle\ \mapsto\ \gamma\ ,\ \tfrac{c_1}{C,\delta}\ \mapsto\ \alpha\ ,\ \dots\ ,\ \tfrac{c_n}{C,\delta}\ \mapsto\ \alpha\Big] \\ &= \tfrac{s}{\delta}\Big[\tfrac{c_1}{C,\delta}\ \mapsto\ \alpha\ ,\ \dots\ ,\ \tfrac{c_n}{C,\delta}\ \mapsto\ \alpha\Big] \qquad\qquad \text{(by } (B)\text{)} \\ &= \tfrac{s's''}{\delta}\Big[\tfrac{c_k}{C,\delta}\ \mapsto\ \alpha\Big] = \tfrac{s'}{\alpha} \in \tau_{\alpha}(\{0,1,\mu_1,\dots,\mu_n\}^*) \end{aligned}$$

where we have used that $s = s's''$ and $\frac{s''}{\delta} = \frac{c_k}{C,\delta}$ for some $k \in \{1, \ldots, n\}$ since we would otherwise have $\frac{s}{\delta}\Big[\frac{c_1}{C,\delta} \mapsto \alpha \ , \ \ldots \ , \ \frac{c_n}{C,\delta} \mapsto \alpha\Big] = \frac{s}{\delta}$ while $\delta \not\sqsubseteq T$ by (1). Since T_0 satisfies (1)-(3), $T_0 \in \mathbb{P}(C, \gamma)$ by the induction hypothesis. It then follows that $T \in \mathbb{P}(C, \gamma)$.

Thus, by induction, $T \in \mathbb{P}(C, \alpha, \gamma)$ if T satisfies (1)-(3). □

6 Sequences of Strings II

Recall that the instance $\langle a_1, b_1\rangle, \ldots, \langle a_n, b_n\rangle$ of PCP has a solution if and only if there exist a finite nonempty sequence $i_1, ..., i_m$ of indexes such that we have $a_{i_1}a_{i_2}\ldots a_{i_m} = b_{i_1}b_{i_2}\ldots b_{i_m}$. Let $C = \langle c_1, c_2, \ldots, c_n\rangle$ be one of the sequences $\langle a_1, \ldots, a_n\rangle$, $\langle b_1, \ldots, b_n\rangle$. Each element $T \in \mathbb{P}(C, \alpha, \gamma)$ represents a sequence of the form $w_1, w_2, \ldots, w_m$ where $w_k = c_{i_1}c_{i_2}\ldots c_{i_k}$ and $i_j \in \{1, \ldots, n\}$ for all $j \in \{1, \ldots, m\}$. We need the sequence $i_1, i_2, \ldots, i_m$ to verify the equality $a_{i_1}a_{i_2}\ldots a_{i_m} = b_{i_1}b_{i_2}\ldots b_{i_m}$. We need an existential $\mathcal{L}_{\mathsf{BT}}$-formula that extracts this information from T. To achieve this, we need to encode sequences that are more complex than those we encountered in Section 5.

The class $\mathbb{P}(C, \alpha, \gamma)$ consists of finite binary trees that encode sequences of the form $w_1, w_2, \ldots, w_k$ where $w_i \in \tau_\alpha(\{0, 1, \mu_1, \ldots, \mu_n\}^*)$ for all $i \in \{1, \ldots, k\}$. We need to consider the class of those binary trees that encode sequences of the form $W_1, W_2, \ldots, W_k$ where $W_i \in \mathbb{P}(C, \alpha, \gamma)$ for all $i \in \{1, \ldots, k\}$. To illustrate how this helps us identify the sequence $i_1, i_2, \ldots, i_m$, let $T = \Big\langle \gamma \ , \ \frac{c_2}{C,\alpha} \ , \ \frac{c_2c_3}{C,\alpha} \ , \ \frac{c_2c_3c_1}{C,\alpha} \Big\rangle$ where $c_1 = 01$, $c_2 = 00$, $c_3 = 10$. We need to find an existential $\mathcal{L}_{\mathsf{BT}}$-formula $\Psi(T, X)$ that is true in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ if and only if X represents the string $\mu_2\mu_3\mu_1$. Instead of working with T, we work with the binary tree $W_1 = \Gamma_n^\alpha(T)$ in Figure 1. It contains the information μ_2, μ_3, μ_1 and has the advantage of having a simpler structure. We give a formal definition of the operator $\Gamma_n^\alpha : \mathbf{H} \to \mathbf{H}$ that takes T and gives us $\Gamma_n^\alpha(T)$. It is really the restriction of Γ_n^α to $\mathbb{P}(C, \alpha, \gamma)$ we are interested in. It will follow from the definition that Γ_n^α is existentially definable.

Definition 7. *Let $\alpha, 0, 1, \mu_1, \ldots, \mu_n$ be as in Definition 5. Let $\mu_{n+1}, \ldots, \mu_{2n}$ be distinct fresh letters. Let $\Gamma_n^\alpha : \mathbf{H} \to \mathbf{H}$ be the function defined by $\Gamma_n^\alpha(T) = T_2$ where*

$$T_0 = T\Big[\ \frac{\mu_1}{\alpha} \mapsto \frac{\mu_1}{\mu_{n+1}} \ , \ \ldots \ , \ \frac{\mu_n}{\alpha} \mapsto \frac{\mu_n}{\mu_{n+n}}\ \Big]$$

$$T_1 = T_0\Big[\ 1 \mapsto 0 \ , \ \mu_1 \mapsto 0 \ , \ \mu_2 \mapsto 0 \ , \ \ldots \ , \ \mu_n \mapsto 0\ \Big]$$

$$T_2 = T_1\Big[\ \mu_{n+1} \mapsto \mu_1 \ , \ \mu_{n+2} \mapsto \mu_2 \ , \ \ldots \ , \ \mu_{n+n} \mapsto \mu_n\ \Big] \ .$$

Recall that we are interested in specifying an existential $\mathcal{L}_{\mathsf{BT}}$-formula $\Psi(T, X)$ that is true if and only if X encodes the string $\mu_2\mu_3\mu_1$. As we have just seen,

$\mathit{\Gamma}_n^\alpha(T)$ contains also the information μ_2, μ_3, μ_1. So, we let $\mathit{\Psi}(T, X)$ be a formula of the form $\exists W \; \mathit{\Phi}(T, X, W)$ where W is a finite binary tree that encodes a sequence $W_1, W_2, \ldots, W_k$ where $W_1 = \mathit{\Gamma}_n^\alpha(T)$ and $W_k = X$. Before we give a formal definition of the class $\mathbb{P}_2(C, \alpha, \gamma)$ of all W with this property, we use the binary tree $T = \Big\langle \; \gamma \, , \, \frac{c_2}{C,\alpha} \, , \, \frac{c_2 c_3}{C,\alpha} \, , \, \frac{c_2 c_3 c_1}{C,\alpha} \; \Big\rangle$ to illustrate the form of W. Let $W_1, \ldots, W_7$ be the binary trees in Figure 1. Then, W can for example be the binary tree $\Big\langle \; \alpha \, , \, W_7 \, , \, W_6 \, , \, W_5 \, , \, W_4 \, , \, W_3 \, , \, W_2 \, , \, W_1 \; \Big\rangle$ or the binary tree $\Big\langle \; \alpha \, , \, W_7 \, , \, W_7 \, , \, W_6 \, , \, W_5 \, , \, W_4 \, , \, W_3 \, , \, W_2 \, , \, W_1 \; \Big\rangle$. It is not a problem that there are many choices for W. What is important is that $\mathit{\Gamma}_n^\alpha(T)$ is the unique right subtree of W, and W_7 encodes the information we need in a simple format and is the unique subtree X of W which is such that $\langle \alpha \, , \, X \rangle \sqsubseteq W$.

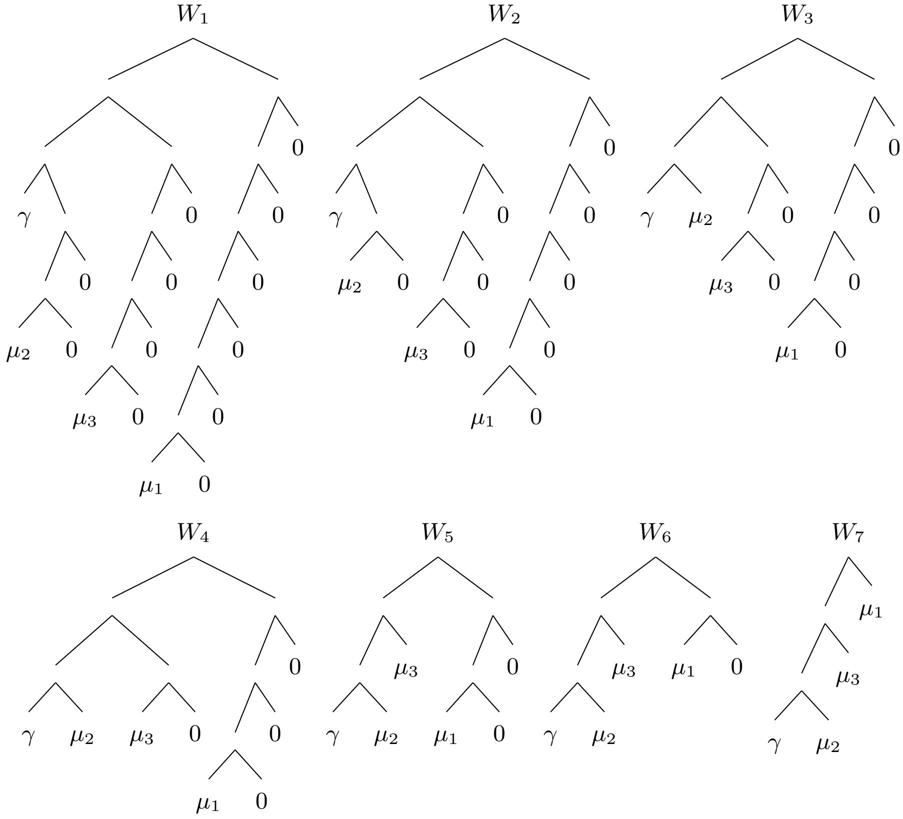

Fig. 1. Let $T = \Big\langle \; \gamma \, , \, \frac{c_2}{C,\alpha} \, , \, \frac{c_2 c_3}{C,\alpha} \, , \, \frac{c_2 c_3 c_1}{C,\alpha} \; \Big\rangle$. Then, $W_1 = \mathit{\Gamma}_n^\alpha(T)$. Binary trees of the form $W = \Big\langle \; \alpha \, , \, W_7 \, , \, \ldots \, , \, W_7 \, , \, W_6 \, , \, W_5 \, , \, W_4 \, , \, W_3 \, , \, W_2 \, , \, W_1 \; \Big\rangle$ are elements of $\mathbb{P}_2(C, \alpha, \gamma)$.

Definition 8. *Let $C = \langle c_1, c_2, \ldots, c_n \rangle$ be a sequence of nonempty binary strings. Let $\alpha, \gamma \in \mathbf{H}$ be incomparable with respect to the subtree relation. Assume α satisfies the condition in Definition 4. Assume γ is not a subtree of μ_i for all $i \in \{1, \ldots, n\}$. Let $W \in \mathbb{P}_2(C, \alpha, \gamma)$ if and only if there exists a sequence $W_1, W_2, \ldots, W_k \in \mathbf{H}$ such that there exists $T \in \mathbb{P}(C, \alpha, \gamma)$ such that $W_1 = \Gamma_n^\alpha(T)$, $W = \big\langle\, \alpha\,,\, W_k, W_{k-1}\,,\, \ldots\,,\, W_1\, \big\rangle$, $W_k \in \tau_\gamma(\{\mu_1, \ldots, \mu_n\}^+)$ and $W_{i+1} = W_i\Big[\, \frac{0}{\mu_1} \mapsto \mu_1\,,\, \frac{0}{\mu_2} \mapsto \mu_2\,,\, \ldots\,,\, \frac{0}{\mu_n} \mapsto \mu_n \,\Big]$ for all $i \in \{1, 2, \ldots, k-1\}$.*

We prove that $\mathbb{P}_2(C, \alpha, \gamma)$ is existentially definable.

Lemma 4. *Let $C = \langle c_1, c_2, \ldots, c_n \rangle$ be a sequence of nonempty binary strings. Let $\alpha, \gamma \in \mathbf{H}$ be incomparable with respect to the subtree relation. Assume α satisfies the condition in Definition 4. Assume γ is not a subtree of μ_i for all $i \in \{1, \ldots, n\}$. Let $W \in \mathbf{H}$. Then, $W \in \mathbb{P}_2(C, \alpha, \gamma)$ if and only if*

(1) *there exists $X \in \tau_\gamma(\{\mu_1, \ldots, \mu_n\}^+)$ such that $\langle \alpha\,,\, X \rangle \sqsubseteq W$*
(2) *there exists $T \in \mathbb{P}(C, \alpha, \gamma)$ such that $W = \langle V\,,\, \Gamma_n^\alpha(T) \rangle$ where*

$$V = W\Big[\, \langle \alpha\,,\, X \rangle \mapsto \alpha\,,\, \frac{0}{\mu_1} \mapsto \mu_1\,,\, \frac{0}{\mu_2} \mapsto \mu_2\,,\, \ldots\,,\, \frac{0}{\mu_n} \mapsto \mu_n \,\Big]\,.$$

Proof. The left-right implication is a straightforward consequence of Definition 8. We focus on proving the right-left implication.

Assume W satisfies (1)-(2). We need to show that $W \in \mathbb{P}_2(C, \alpha, \gamma)$. By Definition 8, we need to show that there exist $W_1, \ldots, W_k \in \mathbf{H}$ such that: (A) $W = \big\langle\, \alpha\,,\, W_k, W_{k-1}\,,\, \ldots\,,\, W_2\,,\, W_1\, \big\rangle$, (B) $W_k \in \tau_\gamma(\{\mu_1, \ldots, \mu_n\}^+)$, (C) $W_{i+1} = W_i\Big[\, \frac{0}{\mu_1} \mapsto \mu_1\,,\, \frac{0}{\mu_2} \mapsto \mu_2\,,\, \ldots\,,\, \frac{0}{\mu_n} \mapsto \mu_n \,\Big]$ for all $i \in \{1, 2, \ldots, k-1\}$, (D) there exists $T \in \mathbb{P}(C, \alpha, \gamma)$ such that $W_1 = \Gamma_n^\alpha(T)$.

Let X and T be binary trees that satisfy clauses (1)-(2). First, we prove by (backward) induction that if $\langle \alpha\,,\, X \rangle \sqsubseteq U \sqsubseteq W$ and $U = \langle U_0\,,\, U_1 \rangle$, then

$$U_0 = U\Big[\, \langle \alpha\,,\, X \rangle \mapsto \alpha\,,\, \frac{0}{\mu_1} \mapsto \mu_1\,,\, \ldots\,,\, \frac{0}{\mu_n} \mapsto \mu_n \,\Big] \quad \text{and} \quad \alpha \not\sqsubseteq U_1\,.$$

We let (*) refer to the equality, and we let (**) refer to $\alpha \not\sqsubseteq U_1$. The base case $U = W$ is Clause (2). So, assume $U = \langle V\,,\, U_1 \rangle$, $V = \langle V_0\,,\, V_1 \rangle$, $\langle \alpha\,,\, X \rangle \sqsubseteq V \sqsubseteq U \sqsubseteq W$ and U satisfies (*) and (**). We need to show that V satisfies (*) and (**). Since U satisfies (**), $\langle \alpha\,,\, X \rangle \not\sqsubseteq U_1$. Since α is incomparable with 0 and μ_i with respect to $\sqsubseteq$, the binary tree $\frac{0}{\mu_i}$ cannot equal a binary tree that has α as subtree. Furthermore, if $\alpha \sqsubseteq R$, then $\alpha \sqsubseteq R[\, \frac{0}{\mu_i} \mapsto \mu_i \,]$. Hence, by (*)

$$\begin{aligned} V &= U\Big[\, \langle \alpha\,,\, X \rangle \mapsto \alpha\,,\, \tfrac{0}{\mu_1} \mapsto \mu_1\,,\, \ldots\,,\, \tfrac{0}{\mu_n} \mapsto \mu_n \,\Big] \\ &= \langle V\,,\, U_1 \rangle\Big[\, \langle \alpha\,,\, X \rangle \mapsto \alpha\,,\, \tfrac{0}{\mu_1} \mapsto \mu_1\,,\, \ldots\,,\, \tfrac{0}{\mu_n} \mapsto \mu_n \,\Big] = \Big\langle\, U'\,,\, U'' \,\Big\rangle \end{aligned}$$

where by (**)

$$U'' = U_1\Big[\, \tfrac{0}{\mu_1} \mapsto \mu_1\,,\, \ldots\,,\, \tfrac{0}{\mu_n} \mapsto \mu_n \,\Big] \not\sqsupseteq \alpha$$

$$U' = V\Big[\, \langle \alpha\,,\, X \rangle \mapsto \alpha\,,\, \tfrac{0}{\mu_1} \mapsto \mu_1\,,\, \ldots\,,\, \tfrac{0}{\mu_n} \mapsto \mu_n \,\Big]\,.$$

Thus, V satisfies (*) and (**). Thus, by induction, if $\langle \alpha \, , \, X \rangle \sqsubseteq U \sqsubseteq W$ and $U = \langle U_0 \, , \, U_1 \rangle$, then U satisfies (*) and (**).

Now, to prove that (A)-(D) hold, it suffices to prove by induction on the size of finite binary trees that if U is a subtree of W which is such that $\langle \alpha \, , \, X \rangle \sqsubseteq U$, then there exists a sequence $U_1, \ldots , U_m$ such that: (i) $U = \Big\langle \; \alpha, U_m, U_{m-1}, \ldots , U_1 \; \Big\rangle$, (ii) $U_m = X$, (iii) $U_{i+1} = U_i \Big[\; \frac{0}{\mu_1} \; \mapsto \; \mu_1 \; , \; \ldots \; , \; \frac{0}{\mu_n} \; \mapsto \; \mu_n \; \Big]$ for all $i \in \{1, 2, \ldots , m-1\}$.

So, assume $\langle \alpha \, , \, X \rangle \sqsubseteq U \sqsubseteq W$. If $U = \langle \alpha \, , \, X \rangle$, then U satisfies (i)-(iii) trivially. Otherwise, by (**), there exist V and U_1 such that $U = \langle V \, , \, U_1 \rangle$ and $\langle \alpha \, , \, X \rangle \sqsubseteq V$. By the induction hypothesis, there exists a sequence $V_1, \ldots , V_m$ such that the following holds: (iv) $V = \Big\langle \; \alpha, V_m, V_{m-1}, \ldots , V_1 \; \Big\rangle$, (v) $V_m = X$, (vi) $V_{i+1} = V_i \Big[\; \frac{0}{\mu_1} \; \mapsto \; \mu_1 \; , \; \ldots \; , \; \frac{0}{\mu_n} \; \mapsto \; \mu_n \; \Big]$ for all $i \in \{1, 2, \ldots , m-1\}$. In particular, $U = \langle V \, , \, U_1 \rangle = \Big\langle \; \alpha \, , \, V_m \, , \, V_{m-1} \, , \, \ldots , V_1 \, , \, U_1 \; \Big\rangle$. By (v)-(vi) and (**), there can only be one occurrence of α in U. Hence $U \Big[\; \langle \alpha \, , \, X \rangle \; \mapsto \; \alpha \; \Big] = \Big\langle \; \alpha \, , \, V_{m-1} \, , \, \ldots , V_1 \, , \, U_1 \; \Big\rangle$. Then, by (*) and (vi)

$$\begin{aligned} \Big\langle \; \alpha \, , \, V_m \, , \, V_{m-1} \, , \, \ldots , V_1 \; \Big\rangle = V = \\ U \Big[\; \langle \alpha \, , \, X \rangle \; \mapsto \; \alpha \; , \; \frac{0}{\mu_1} \; \mapsto \; \mu_1 \; , \; \ldots \; , \; \frac{0}{\mu_n} \; \mapsto \; \mu_n \; \Big] = \\ \Big\langle \; \alpha \, , \, V_{m-1} \, , \, \ldots , V_1 \, , \, U_1 \; \Big\rangle \Big[\; \frac{0}{\mu_1} \; \mapsto \; \mu_1 \; , \; \ldots \; , \; \frac{0}{\mu_n} \; \mapsto \; \mu_n \; \Big] = \\ \Big\langle \; \alpha \, , \, V_m \, , \, \ldots \, , \, V_2 \, , \, U_1' \; \Big\rangle \end{aligned}$$

where $U_1' = U_1 \Big[\; \frac{0}{\mu_1} \; \mapsto \; \mu_1 \; , \; \ldots \; , \; \frac{0}{\mu_n} \; \mapsto \; \mu_n \; \Big]$. Hence

$$U = \Big\langle \; \alpha \, , \, V_m \, , \, V_{m-1} \, , \, \ldots , V_1 \, , \, U_1 \; \Big\rangle \; \text{ and } \; V_1 = U_1 \Big[\; \frac{0}{\mu_1} \; \mapsto \; \mu_1 \, , \; \ldots \; , \; \frac{0}{\mu_n} \; \mapsto \; \mu_n \; \Big] \, .$$

Thus, U satisfies (i)-(iii).

Thus, by induction, if U is a subtree of W which is such that $\langle \alpha \, , \, X \rangle \sqsubseteq U$, then U satisfies (i)-(iii). □

7 Reduction of Post's Correspondence Problem

We are ready to specify a many-to-one reduction of Post's Correspondence Problem.

Theorem 1. *The Post Correspondence Problem is many-to-one reducible to the fragment* $\mathsf{Th}^{\exists}(\mathcal{T}(\mathcal{L}_{\mathsf{BT}}))$.

Proof. Consider an instance $\langle a_1, b_1 \rangle, \ldots , \langle a_n, b_n \rangle$ of PCP. We need to construct an existential $\mathcal{L}_{\mathsf{BT}}$-sentence ϕ that is true in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ if and only if $\langle a_1, b_1 \rangle, \ldots , \langle a_n, b_n \rangle$ has a solution. The instance $\langle a_1, b_1 \rangle, \ldots , \langle a_n, b_n \rangle$ has a solution if and only if there exist two sequences $u_1, u_2, \ldots , u_k$ and $v_1, v_2, \ldots , v_m$ such that

(I) there exists $f_1 \in \{1, \ldots, n\}$ such that $u_1 = a_{f_1}$ and for all $j \in \{1, \ldots, k-1\}$ there exist $f_{j+1} \in \{1, \ldots, n\}$ such that $u_{j+1} = u_j a_{f_{j+1}}$
(II) there exists $g_1 \in \{1, \ldots, n\}$ such that $v_1 = b_{g_1}$ and for all $j \in \{1, \ldots, m-1\}$ there exist $g_{j+1} \in \{1, \ldots, n\}$ such that $v_{j+1} = v_j b_{g_{j+1}}$
(III) $k = m$ and $f_j = g_j$ for all $j \in \{1, \ldots, k\}$
(IV) $u_k = v_m$.

Let $\alpha = \langle \bot\,,\, \bot^2 \rangle$ and $\gamma = \langle \bot\,,\, \bot^3 \rangle$. Then, α and γ satisfy the conditions in Definition 6 and Definition 8. Let $A = \langle a_1, a_2, \ldots, a_n \rangle$ and let $B = \langle b_1, b_2, \ldots, b_n \rangle$. Definition 6 tells us that the sequence $u_1, u_2, \ldots, u_k$ is encoded by a binary tree $L \in \mathbb{P}(A, \alpha, \gamma)$ and the right subtree of L, denoted U, encodes u_k. Similarly, the sequence $v_1, v_2, \ldots, v_m$ is encoded by a binary tree $R \in \mathbb{P}(B, \alpha, \gamma)$ and the right subtree of R, denoted V, encodes v_m. Lemma 3 tells us that $\mathbb{P}(A, \alpha, \gamma)$ and $\mathbb{P}(B, \alpha, \gamma)$ are existentially definable.

Definition 8 gives us binary trees X_L and $W_L \in \mathbb{P}_2(A, \alpha, \gamma)$ such that $\mathit{\Gamma}_n^\alpha(L)$ is the right subtree of W_L, $\langle \alpha\,,\, X_L \rangle \sqsubseteq W_L$ and X_L encodes the sequence $f_1, f_2, \ldots, f_k$. The existentially definable operator $\mathit{\Gamma}_n^\alpha$ is defined in Definition 7. Similarly, there exist X_R and $W_R \in \mathbb{P}_2(B, \alpha, \gamma)$ such that $\mathit{\Gamma}_n^\alpha(R)$ is the right subtree of W_R, $\langle \alpha\,,\, X_R \rangle \sqsubseteq W_R$ and X_R encodes the sequence $g_1, g_2, \ldots, g_m$. Lemma 4 tells us that $\mathbb{P}_2(A, \alpha, \gamma)$ and $\mathbb{P}_2(B, \alpha, \gamma)$ are existentially definable.

Now, encoding (III) corresponds to requiring that $X_L = X_R$ holds. To encode (IV), we cannot simply require that $U = V$ holds since U is the representation of u_k when viewed as an element of $\{0, 1, \mu_1, \ldots, \mu_n\}^+$ and V is the representation of v_m when viewed as an element of $\{0, 1, \mu_1, \ldots, \mu_n\}^+$. So, let $\mathit{\Theta}_n^A(U)$ be the binary tree we obtain by replacing μ_i with the last letter of a_i and let $\mathit{\Theta}_n^B(V)$ be the binary tree we obtain by replacing μ_j with the last letter of b_j. Then, encoding (IV) corresponds to requiring that $\mathit{\Theta}_n^A(U) = \mathit{\Theta}_n^B(V)$ holds.

Let $\mathit{\Theta}_n^A(U) = U\big[\, \mu_1 \mapsto d_1 \,,\, \ldots \,,\, \mu_n \mapsto d_n \big]$ where d_i is the last letter of a_i. Let $\mathit{\Theta}_n^B(V) = V\big[\, \mu_1 \mapsto e_1 \,,\, \ldots \,,\, \mu_n \mapsto e_n \big]$ where e_j is the last letter of b_j. Let

$$\begin{aligned} \phi \equiv\ & \exists L \in \mathbb{P}(A, \alpha, \gamma)\ \exists U, U'\ \exists R \in \mathbb{P}(B, \alpha, \gamma)\ \exists V, V' \\ & \exists W_L \in \mathbb{P}_2(A, \alpha, \gamma)\ \exists X_L, S_L\ \exists W_R \in \mathbb{P}_2(B, \alpha, \gamma)\ \exists X_R, S_R\ \Big[\\ & L = \langle U'\,,\, U \rangle \ \wedge\ R = \langle V'\,,\, V \rangle \ \wedge\ \langle \alpha\,,\, X_L \rangle \sqsubseteq W_L \ \wedge\ W_L = \langle S_L\,,\, \mathit{\Gamma}_n^\alpha(L) \rangle \ \wedge \\ & \langle \alpha\,,\, X_R \rangle \sqsubseteq W_R \ \wedge\ W_R = \langle S_R\,,\, \mathit{\Gamma}_n^\alpha(R) \rangle \ \wedge\ \mathit{\Theta}_n^A(U) = \mathit{\Theta}_n^B(V) \ \wedge\ X_L = X_R \ \Big] . \end{aligned}$$

Then, ϕ is true in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ if and only if $\langle a_1, b_1 \rangle, \ldots, \langle a_n, b_n \rangle$ has a solution. □

8 Analogue of Hilbert's Tenth Problem

In this section, we show that the analogue of Hilbert's Tenth Problem for $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ is undecidable.

Theorem 2. *The fragment* $\mathsf{Th}^{\mathsf{H10}}(\mathcal{T}(\mathcal{L}_{\mathsf{BT}}))$ *is undecidable.*

Proof. Since $\mathsf{Th}^{\exists}(\mathcal{T}(\mathcal{L}_{\mathsf{BT}}))$ is undecidable, it suffices to show that given an existential $\mathcal{L}_{\mathsf{BT}}$-sentence ϕ, we can compute a finite number of $\mathcal{L}_{\mathsf{BT}}$-sentences $\phi_1, \ldots, \phi_n$ of the form $\exists \vec{x}\ [\ s = t\ ]$ such that $\mathcal{T}(\mathcal{L}_{\mathsf{BT}}) \models \phi \leftrightarrow \bigvee_{i=1}^{n} \phi_i$. Since $\mathcal{T}(\mathcal{L}_{\mathsf{BT}}) \models (\ s_1 = t_1 \wedge s_2 = t_2\ ) \leftrightarrow \langle s_1, s_2 \rangle = \langle t_1, t_2 \rangle$, it suffices to show that given a $\mathcal{L}_{\mathsf{BT}}$-formula of the form $s \neq t$, we can compute a finite number of atomic $\mathcal{L}_{\mathsf{BT}}$-formulas $s_1 = t_1, \ldots, s_k = t_k$ such that we have $\mathcal{T}(\mathcal{L}_{\mathsf{BT}}) \models s \neq t \leftrightarrow \bigvee_{j=1}^{k} s_j = t_j$. This is the case since $s \neq t \ \Leftrightarrow\ t[\, s \mapsto \langle s, s \rangle\,] = t \ \vee\ s[\, t \mapsto \langle t, t \rangle\,] = s$. □

9 Bounded Quantifiers

We end this paper by showing that $\Sigma_{1,0,1}^{\mathcal{T}(\mathcal{L}_{\mathsf{BT}})}$ is undecidable. We prove this by encoding the Modulo Problem.

Theorem 3. *The fragment* $\Sigma_{1,0,1}^{\mathcal{T}(\mathcal{L}_{\mathsf{BT}})}$ *is undecidable.*

Proof. We encode natural numbers as follows: $n \equiv \perp^{n+2}$. The next step is to associate linear polynomials in one variable with $\mathcal{L}_{\mathsf{BT}}$-terms. We let $L(z) \equiv z[0 \mapsto z]$. If z represents the natural number q, then $L(z)$ represents the natural number $2q$ since 0 has exactly one occurrence in z. Recall that $L^0(z) = z$ and $L^{k+1}(z) = L(L^k(z))$. Hence, if $n > 0$, then $L^{n-1}(z)$ represents the natural number nq. If $n > 0$, then the term $m[0 \mapsto L^{n-1}(z)]$ represents the natural number $nq{+}m$. We complete our translation of linear polynomials in one variable as follows: For any formula $\phi(x)$ where x is a free variable, $\phi(nz + m) = \phi(m)$ if $n = 0$ and $\phi(nz + m) = \phi(m[0 \mapsto L^{n-1}(z)])$ if $n > 0$.

Given an instance $\langle A_0, B_0 \rangle, \ldots, \langle A_{M-1}, B_{M-1} \rangle$, we need to compute a $\Sigma_{1,0,1}$-sentence ψ that is true in $\mathcal{T}(\mathcal{L}_{\mathsf{BT}})$ if and only if the instance has a solution. Let $x \in y$ be shorthand for $\langle x, \alpha \rangle \sqsubseteq y \ \wedge\ \alpha \not\sqsubseteq x$ where $\alpha \equiv \ \langle \perp, \perp^2 \rangle$. The sentence ψ needs to say that there exists a finite set T such that $3 \in T$, $2 \in T$ and if $2 \neq Mz + j \in T \ \wedge\ 0 \leq j < M$, then $A_j z + B_j \in T$. With this in mind, we let ψ be the sentence $\exists T\ \forall z \sqsubseteq T\ [\ 3 \in T \ \wedge\ \psi_0\ ]$ where ψ_0 is $\bigwedge_{j=0}^{M-1} \Big(\ (\ Mz + j \in T \ \wedge\ Mz + j \neq 2\ ) \rightarrow A_j z + B_j \in T\ \Big)$. □

References

1. Davis, M.: Hilbert's Tenth Problem is Unsolvable. The American Mathematical Monthly 80 (3), pp. 233–269 (1973) .
2. Hodges, W.: Model Theory. Cambridge University Press (1993).
3. Kristiansen, L., Murwanashyaka, J.: First-Order Concatenation Theory with Bounded Quantifiers. Archive for Mathematical Logic 60 no. 1-2, pp. 77–104 (2021).
4. Post, E.L.: A Variant of a Recursively Unsolvable Problem. Bulletin (new Series) of the American Mathematical Society 52 no. 4 (1946) 264–268.
5. Venkataraman, K.: Decidability of the Purely Existential Fragment of the Theory of Term Algebras. Journal of the ACM 34 no. 2, pp. 492–510 (1987).