ISSN 1346-5597

National Institute of Informatics

NII Technical Report

Multi-dimensional Trees and a Chomsky-Schützenberger-Weir Representation Theorem for Simple Context-Free Tree Grammars

Makoto Kanazawa

NII-2013-003E
Nov. 2013

URL: http://www.nii.ac.jp/ Tokyo 101-8430, JAPAN +81-3-4212-2000 FAX: +81-3-3556-1916

Multi-dimensional Trees and a Chomsky-Schützenberger-Weir Representation Theorem for Simple Context-Free Tree Grammars

Makoto Kanazawa*

National Institute of Informatics, Tokyo, Japan
kanazawa@nii.ac.jp

Abstract. Weir [34] proved a Chomsky-Schützenberger-like representation theorem for the string languages of tree-adjoining grammars, where the Dyck language D_n in the Chomsky-Schützenberger characterization is replaced by the intersection $D_{2n} \cap g^{-1}(D_{2n})$, where g is a certain bijection on the alphabet consisting of $2n$ pairs of brackets. This paper presents a generalization of this theorem to the string languages that are the yield images of the tree languages generated by simple (i.e., linear and non-deleting) context-free tree grammars. This result is obtained through a natural generalization of the original Chomsky-Schützenberger theorem to the tree languages of simple context-free tree grammars. We use Rogers's [24,23] notion of multi-dimensional trees to state this latter theorem in a very general, abstract form.

Keywords: Context-free tree grammar; Multi-dimensional tree; Dyck language; Chomsky-Schützenberger theorem

1 Introduction

Weir [34] showed that every string language L generated by a tree-adjoining grammar [11] can be written as

$$L = h(R \cap D_{2n} \cap g^{-1}(D_{2n})),$$

where h is a homomorphism, R is a regular set, n is a positive integer, D_{2n} is the Dyck language over the alphabet Γ_{2n} consisting of $2n$ pairs of brackets $[_1,]_1, \dots, [_{2n},]_{2n}$, and g is the bijection on Γ_{2n} defined by

$$g([_{2i+1}) = [_{2i+1}, \quad g(]_{2i+1}) =]_{2i+2}, \quad g([_{2i+2}) =]_{2i+1}, \quad g(]_{2i+2}) = [_{2i+2},$$

for $i = 0, \dots, n-1$. The effect of the intersection with $g^{-1}(D_{2n})$ on the Dyck language D_{2n} is to make the consecutive odd-numbered and even-numbered

* This work was in part supported by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (KAKENHI) (25330020).

brackets $[_{2i+1},]_{2i+1}, [_{2i+2},]_{2i+2}$ always appear as a group, in the configuration $[_{2i+1}\, [_{2i+2}\,]_{2i+2}\,]_{2i+1}$. When two such groups, say, $[_1,]_1, [_2,]_2$ and $[_3,]_3, [_4,]_4$, overlap, the only possible configurations are

$$[_1\, [_3\, [_4\,]_4\,]_3\, [_2\,]_2\,]_1,$$
$$[_1\, [_2\,]_2\, [_3\, [_4\,]_4\,]_3\,]_1,$$
$$[_1\, [_3\, [_4\, [_2\,]_2\,]_4\,]_3\,]_1,$$

and those with the positions of the two groups interchanged. As Weir [34] showed, $D_{2n} \cap g^{-1}(D_{2n})$ is a non-context-free tree-adjoining language for every $n \geq 1$.

In this paper, we prove a generalization of Weir's theorem for *simple* (i.e., *linear* and *non-deleting*) *context-free tree grammars*[1] [26,8,15]: if L is the string language generated by a simple context-free tree grammar of rank $q-1$, then L can be written as

$$L = h(R \cap D_{qn} \cap g^{-1}(D_{qn})),$$

where h, R, and n are as before, D_{qn} is the Dyck language over the alphabet Γ_{qn} (containing qn pairs of brackets), and g is the bijection on Γ_{qn} defined by

$$g([_{qi+1}) = [_{qi+1}, \qquad g(]_{qi+1}) =]_{qi+q},$$
$$g([_{qi+j}) =]_{qi+j-1}, \qquad g(]_{qi+j}) = [_{qi+j},$$

for $i = 0, \ldots, n-1$ and $j = 2, \ldots, q$. This result generalizes Weir's [34] because tree-adjoining grammars generate the same string languages as simple context-free tree grammars that are *monadic* (i.e., of rank 1) [21,10,16]. As in the original Chomsky-Schützenberger theorem [4,3], we can take R to be a *local* set and h to be *alphabetic* in the sense that h maps each symbol either to a symbol or to the empty string.

It is known [12,13] that the string languages of simple context-free tree grammars are exactly those generated by *multiple context-free grammars* [29] that are *well-nested* in the sense of [13]. For multiple context-free grammars in general, Yoshinaka et al. [36] have proved a Chomsky-Schützenberger-like representation theorem, but the analogy to the Chomsky-Schützenberger theorem is somewhat weak because their notion of a *multiple Dyck language* is given only by reference to a certain multiple context-free grammar, and does not seem to have other independent characterizations, analogous to the characterization of ordinary Dyck languages in terms of the cancellation law $[_i\,]_i \rightsquigarrow \varepsilon$. Our result is obtained via a natural generalization of the Chomsky-Schützenberger theorem to the tree languages of simple context-free tree grammars. This intermediate result is stated in terms of *Dyck tree languages*, which are exactly analogous to the original Dyck languages in that they have two equivalent definitions, one in terms of inductive definitions and one in terms of rewriting with cancellation laws.

In order to emphasize the analogy between the string case and the tree case, we use the notion of a *multi-dimensional tree* introduced by Rogers [24,23]

[1] The term "simple context-free tree grammar" is taken from [7].

and state many lemmas as general facts about m-dimensional trees. We use 3-dimensional trees to represent derivation trees of simple context-free tree grammars. We note that our notion of the *yield* of an m-dimensional tree will be different from Rogers's, because Rogers's definition was specifically tailored for the monadic case.[2]

2 Preliminaries

2.1 First-Child-Next-Sibling Encoding of Ordered Unranked Trees

In an ordered unranked tree, a node may have any number of children, and the children of the same node are linearly ordered. We do not consider unordered trees in this paper, so we call ordered unranked trees simply unranked trees. In the usual term notation for unranked trees [32], the unranked trees over a set Σ of labels are defined inductively as follows:

- If $c \in \Sigma$, then c is an unranked tree over Σ.
- If $t_1, \dots, t_n$ are unranked trees over Σ ($n \geq 1$) and $c \in \Sigma$, then $c(t_1 \dots t_n)$ is an unranked tree over Σ.

There is a well-known way of encoding unranked trees into binary trees [17], often called the *first-child-next-sibling* encoding. We refer to a node in a binary tree by a string over the set $\{1, 2\}$. Thus, the set of nodes of a binary tree forms a prefix-closed subset T of $\{1, 2\}^*$. (Note that we do not assume binary trees to be *full* in the sense that each node has 0 or 2 children.) We write $u \cdot v$ for the concatenation of two strings $u, v \in \{1, 2\}^*$. In the first-child-next-sibling encoding of unranked trees, the relation $u \cdot 2 = v$ represents the relation "v is the first child of u", and the relation $u \cdot 1 = v$ represents the relation "v is the next sibling of u". The child relation is then represented by the first-child relation composed with the reflexive transitive closure of the next-sibling relation. In this way, any non-empty finite prefix-closed subset T of $\{1, 2\}^*$ such that $1 \notin T$ encodes the set of nodes of some unranked tree. In general, an arbitrary non-empty finite prefix-closed subset of $\{1, 2\}^*$ encodes the nodes of a *hedge*, a finite, non-empty sequence of unranked trees.[3] In this encoding, ε (empty string) is the

[2] When the present work was nearing completion, the author learned of a recent paper by Sorokin [30], in which he states (without proof) a result similar to Theorem 42 below (Theorem 3 of [30]). (The statement of his theorem is actually closer to Lemma 39 below.) As will be clear to the reader, the emphasis of the present paper is very different from Sorokin's. The merit of the present work lies not so much in Theorem 42 itself as in the method of obtaining it though a natural generalization of the constructions that can be used to prove the original Chomsky-Schützenberger theorem. (Sorokin's own emphasis is on the use of monoid automata to characterize the string languages of simple context-free tree grammars.)

[3] Sometimes the empty sequence of unranked trees is also allowed as a hedge, but we exclude it here in order to be able to encode all hedges into binary trees. Note that Knuth [17] and Takahashi [31] used *forest* instead of *hedge*, the term we adopt here following [5].

root of the first tree, 1 is the root of the second tree, $1 \cdot 1$ is the root of the third tree, and so on.

Trees and hedges we consider in this paper are all labeled. Labeled unranked trees and hedges over Σ are represented by pairs of the form $\boldsymbol{T} = (T, \ell)$, where T is a non-empty finite prefix-closed subset of $\{1, 2\}^*$ and ℓ is a function from T to Σ.

2.2 Dyck Languages

For $n \geq 1$, let $\Gamma_n = \bigcup_{i=1}^{n} \{ [_i,]_i \}$. For each i, the two symbols $[_i,]_i$ are regarded as a matching pair of brackets. Define a binary relation $\rightsquigarrow$ on Γ_n^* by

$$\rightsquigarrow = \{ (u \, [_i \,]_i \, v, uv) \mid u, v \in \Gamma_n^*, 1 \leq i \leq n \}.$$

The *Dyck language* D_n is defined by

$$D_n = \{ v \in \Gamma_n^* \mid v \rightsquigarrow^* \varepsilon \},$$

where $\rightsquigarrow^*$ denotes the reflexive transitive closure of the relation $\rightsquigarrow$. An alternative way of defining D_n is by the following context-free grammar:

$$S \to \varepsilon \mid AS,$$
$$A \to [_1 \; S \;]_1 \mid \cdots \mid [_n \; S \;]_n \, .$$

The set D'_n of *Dyck primes* is defined by

$$D'_n = (D_n - \{\varepsilon\}) - (D_n - \{\varepsilon\})^2.$$

Alternatively, the set D'_n is defined by the nonterminal A in the above context-free grammar.

Unranked trees and hedges can be represented by elements of Dyck languages. If Σ is a set of symbols, let

$$\Gamma_\Sigma = \bigcup_{c \in \Sigma} \{ [_c,]_c \}.$$

We write D_Σ and D'_Σ for the Dyck language and the set of Dyck primes over this alphabet. Using the standard term notation for labeled unranked trees, define the *string encoding* function **enc** from labeled unranked trees and hedges over Σ to strings over Γ_Σ by

$$\begin{aligned} \mathbf{enc}(c) &= [_c \;]_c, \\ \mathbf{enc}(c(t_1 \dots t_n)) &= [_c \; \mathbf{enc}(t_1 \dots t_n) \;]_c, \\ \mathbf{enc}(t_1 \dots t_n) &= \mathbf{enc}(t_1) \, \mathbf{enc}(t_2 \dots t_n) \quad \text{for } n \geq 2. \end{aligned}$$

It is clear that the function **enc** maps any unranked tree over Σ to an element of D'_Σ, and any hedge over Σ to an element of $D_\Sigma - \{\varepsilon\}$. Conversely, it is easy to see that any element of D'_Σ encodes a tree over Σ, and any element of $D_\Sigma - \{\varepsilon\}$ encodes a hedge over Σ. These correspondences are bijections.

2.3 Context-Free Tree Grammars

We deviate from the standard practice and let a context-free tree grammar generate a set of unranked trees. Thus, the terminal alphabet of a context-free tree grammar will be unranked. In contrast, the set of nonterminals will be a ranked alphabet, as in the standard definition.

A *ranked alphabet* is a union $\Upsilon = \bigcup_{r \in \mathbb{N}} \Upsilon^{(r)}$ of disjoint sets of symbols. If $f \in \Upsilon^{(r)}$, r is the *rank* of f. If Σ is an (unranked) alphabet and Υ a ranked alphabet, let $\mathbb{T}_{\Sigma,\Upsilon}$ be the set of trees $\boldsymbol{T} \in \mathbb{T}_{\Sigma \cup \Upsilon}$ such that whenever a node of $\boldsymbol{T}$ is labeled by some $f \in \Upsilon$, then the number of its children is equal to the rank of f.

For convenience, we use the term representation of trees. The set $\mathbb{T}_{\Sigma,\Upsilon}$ can be defined inductively as follows:

1. If $f \in \Sigma \cup \Upsilon^{(0)}$, then $f \in \mathbb{T}_{\Sigma,\Upsilon}$;
2. If $f \in \Sigma \cup \Upsilon^{(n)}$ and $t_1, \ldots, t_n \in \mathbb{T}_{\Sigma,\Upsilon}$ $(n \geq 1)$, then $f(t_1 \ldots t_n) \in \mathbb{T}_{\Sigma,\Upsilon}$.

In order to define the notion of a context-free tree grammar, we need a countably infinite supply of variables $x_1, x_2, x_3, \ldots$. The set consisting of the first n variables is denoted X_n (i.e., $X_n = \{x_1, \ldots, x_n\}$). The notation $\mathbb{T}_{\Sigma,\Upsilon}(X_n)$ denotes the set $\mathbb{T}_{\Sigma,\Upsilon \cup X_n}$, where members of X_n are all assumed to have rank 0. A tree in $\mathbb{T}_{\Sigma,\Upsilon}(X_n)$ is often written $t[x_1, \ldots, x_n]$, displaying the variables. If $t[x_1, \ldots, x_n] \in \mathbb{T}_{\Sigma,\Upsilon}(X_n)$ and $t_1, \ldots, t_n \in \mathbb{T}_{\Sigma,\Upsilon}$, then $t[t_1, \ldots, t_n]$ denotes the result of substituting $t_1, \ldots, t_n$ for $x_1, \ldots, x_n$, respectively, in $t[x_1, \ldots, x_n]$. An element $t[x_1, \ldots, x_n]$ of $\mathbb{T}_{\Sigma,\Upsilon}(X_n)$ is an *n-context* if for each $i = 1, \ldots, n$, x_i occurs exactly once in $t[x_1, \ldots, x_n]$. (In the literature, an n-context is sometimes called a *simple tree*.)

A *context-free tree grammar* [26,8] is a quadruple $G = (N, \Sigma, P, S)$, where

1. N is a finite ranked alphabet of nonterminals,
2. Σ is a finite unranked alphabet of terminals,
3. S is a nonterminal of rank 0, and
4. P is a finite set of productions of the form

$$B(x_1 \ldots x_n) \to t[x_1, \ldots, x_n],$$

   where $B \in N^{(n)}$ and $t[x_1, \ldots, x_n] \in \mathbb{T}_{\Sigma,N}(X_n)$.

The *rank* of G is $\max\{\, r \mid N^{(r)} \neq \varnothing \,\}$.

For every $s, s' \in \mathbb{T}_{\Sigma,N}$, $s \Rightarrow_G s'$ is defined to hold if and only if there is a 1-context $c[x_1] \in \mathbb{T}_{\Sigma,N}(1)$, a production $B(x_1 \ldots x_n) \to t[x_1, \ldots, x_n]$ in P, and trees $t_1, \ldots, t_n \in \mathbb{T}_{\Sigma,N}$ such that

$$s = c[B(t_1 \ldots t_n)],$$
$$s' = c[t[t_1, \ldots, t_n]].$$

The relation $\Rightarrow_G^*$ on $\mathbb{T}_{\Sigma,N}$ is defined as the reflexive transitive closure of $\Rightarrow_G$. The *tree langauge* generated by a context-free tree grammar G, denoted by $L(G)$, is defined as follows:

$$L(G) = \{\, t \in \mathbb{T}_\Sigma \mid S \Rightarrow_G^* t \,\}.$$

The *string language* generated by G is

$$\mathbf{y}(L(G)) = \{\, \mathbf{y}(t) \mid t \in L(G) \,\},$$

where $\mathbf{y}(t)$ is the yield of t in the usual sense.

A context-free tree grammar $G = (N, \Sigma, P, S)$ is said to be *simple* if for every production

$$B(x_1 \dots x_n) \to t[x_1, \dots, x_n]$$

in P, $t[x_1, \dots, x_n]$ is an n-context. We let $\mathrm{CFT}_{\mathrm{sp}}(r)$ stand for the family of tree languages L such that $L = L(G)$ for some simple context-free tree grammar G whose rank does not exceed r. We write $\mathrm{yCFT}_{\mathrm{sp}}(r)$ for the corresponding string languages $\{\, \mathbf{y}(L) \mid L \in \mathrm{CFT}_{\mathrm{sp}}(r) \,\}$.

Let $\{y_1, \dots, y_k\}$ be a ranked alphabet, where for $i = 1, \dots, k$, r_i is the rank of y_i. Let $t_i[x_1, \dots, x_{r_i}]$ be an r_i-context. For a tree $t \in \mathbb{T}_{\Sigma, \{y_1, \dots, y_k\}}(X_n)$, we define $t[t_i[x_1, \dots, x_{r_i}]/y_i]$ inductively as follows:

$$\begin{aligned}
c(u_1 \dots u_m)[t_i[x_1, \dots, x_{r_i}]/y_i] &= c(u_1[t_i[x_1, \dots, x_{r_i}]/y_i] \dots u_m[t_i[x_1, \dots, x_{r_i}]/y_i]) \\
&\qquad \text{if } c \in \Sigma, \\
x_j[t_i[x_1, \dots, x_{r_i}]/y_i] &= x_j, \\
y_i(u_1 \dots u_{r_i})[t_i[x_1, \dots, x_{r_i}]/y_i] &= t_i[u_1[t_i[x_1, \dots, x_{r_i}]/y_i], \dots, u_{r_i}[t_i[x_1, \dots, x_{r_i}]/y_i]], \\
y_j(u_1 \dots u_{r_j})[t_i[x_1, \dots, x_{r_i}]/y_i] &= y_j(u_1[t_i[x_1, \dots, x_{r_i}]/y_i] \dots u_{r_j}[t_i[x_1, \dots, x_{r_i}]/y_i]) \\
&\qquad \text{if } j \neq i.
\end{aligned}$$

(Here, the notation $c(u_1 \dots u_m)$ stands for c when $m = 0$, and likewise with $y_j(u_1 \dots u_{r_j})$.)

Let $G = (N, \Sigma, P, S)$ be a simple context-free tree grammar. The *derivation trees* of G and their *tree yield* are defined inductively as follows:

- Let $\pi = B(x_1 \dots x_n) \to t[x_1, \dots, x_n]$ be a production in P with no nonterminal occurring in $t[x_1, \dots, x_n]$. Then $\mathbf{d} = \pi$ is a derivation tree of sort B and its tree yield is $\mathbf{ty}(\mathbf{d}) = t[x_1, \dots, x_n]$.
- Let $\pi = B(x_1 \dots x_n) \to t[x_1, \dots, x_n]$ be a production in P with at least one nonterminal occurring in $t[x_1, \dots, x_n]$. Let $v_1, \dots, v_k$ be the pre-order listing of the nodes of $t[x_1, \dots, x_n]$ labeled by nonterminals. Let B_i be the label of v_i, for $i = 1, \dots, k$. If $\mathbf{d}_i$ is a derivation tree of sort B_i for $i = 1, \dots, k$, then

  $$\pi(\mathbf{d}_1 \dots \mathbf{d}_k)$$

  is a derivation tree of sort B. If $\bar{t}[x_1, \dots, x_n]$ is a tree just like $t[x_1, \dots, x_n]$ except that the label of v_i is changed to y_i, where $y_1, \dots, y_k$ are new symbols, then the tree yield $\mathbf{ty}(\mathbf{d})$ of $\mathbf{d} = \pi(\mathbf{d} \dots \mathbf{d}_k)$ is defined by

  $$\mathbf{ty}(\mathbf{d}) = \bar{t}[x_1, \dots, x_n][\mathbf{ty}(\mathbf{d}_1)/y_1] \dots [\mathbf{ty}(\mathbf{d}_k)/y_k].$$

Note that if $\mathbf{d}$ is a derivation tree of sort B and n is the rank of B, then $\mathbf{ty}(\mathbf{d})$ is an n-context, so the right-hand side of the above equation is well-defined if y_i

is regarded as a symbol whose rank equals the rank of B_i. It is well known that if $G = (N, \Sigma, P, S)$ is a simple context-free grammar, then

$$L(G) = \{\, \mathbf{ty}(\mathbf{d}) \mid \mathbf{d} \text{ is a derivation tree of } G \text{ of sort } S \,\}.$$

Example 1. Consider a simple context-free tree grammar $G = (N, \Sigma, P, S)$, where $N = N^{(0)} \cup N^{(2)} = \{S\} \cup \{B\}$, $\Sigma = \{a_1, a_2, a_3, a_4, a_5, a_6, e, g, h\}$, and P consists of the following rules:

$$\begin{aligned}
&\pi_1 \colon S \to B(ee),\\
&\pi_2 \colon B(x_1 x_2) \to h(a_1 B(h(a_2 x_1 a_3) h(a_4 x_2 a_5)) a_6),\\
&\pi_3 \colon B(x_1 x_2) \to g(x_1 x_2).
\end{aligned}$$

The following trees are derivation trees of this grammar:

$$\pi_1(\pi_2(\pi_3)) \qquad \pi_1(\pi_2(\pi_2(\pi_3)))$$

We have

$$\begin{aligned}
\mathbf{ty}(\pi_3) &= g(x_1 x_2),\\
\mathbf{ty}(\pi_2(\pi_3)) &= h(a_1 g(h(a_2 x_1 a_3) h(a_4 x_2 a_5)) a_6),\\
\mathbf{ty}(\pi_1(\pi_2(\pi_3))) &= h(a_1 g(h(a_2 e a_3) h(a_4 e a_5)) a_6),\\
\mathbf{ty}(\pi_2(\pi_2(\pi_3))) &= h(a_1 h(a_1 g(h(a_2 h(a_2 x_1 a_3) a_3) h(a_4 h(a_4 x_2 a_5) a_5)) a_6) a_6),\\
\mathbf{ty}(\pi_1(\pi_2(\pi_2(\pi_3)))) &= h(a_1 h(a_1 g(h(a_2 h(a_2 e a_3) a_3) h(a_4 h(a_4 e a_5) a_5)) a_6) a_6).
\end{aligned}$$

The string language generated by this grammar is

$$\mathbf{y}(L(G)) = \{\, a_1^n a_2^n e a_3^n a_4^n e a_5^n a_6^n \mid n \geq 0 \,\}.$$

The string languages of simple context-free grammars are the languages generated by *non-duplicating macro grammars* [9], studied by Seki and Kato [28].[4] They also coincide with the languages generated by *well-nested multiple context-free grammars* [13].

3 The Chomsky-Schützenberger Theorem

There are many different proofs of the Chomsky-Schützenberger theorem for context-free languages offered in the literature. Here, we give a proof based on the relation between context-free languages and local sets of unranked trees.[5]

[4] At the level of string languages, simple context-free tree grammars correspond to non-duplicating and argument-preserving (i.e., non-deleting) macro grammars, which are equivalent to non-duplicating macro grammars (Lemma 7 of [28]). Seki and Kato [28] called non-duplicating macro grammars *variable-linear*.

[5] Among the proofs found in well-known textbooks, the one closest to our proof seems to be the one given by Kozen [18].

This will serve as a starting point for our generalization of the theorem to simple context-free tree grammars.

Let $\boldsymbol{T} = (T, \ell)$ be a first-child-next-sibling encoding of an unranked tree. We define three binary relations on T:[6]

$$\begin{aligned} \prec_2^T &= \{ (u, v) \in T \times T \mid u \cdot 2 = v \}, \\ \prec_1^T &= \{ (u, v) \in T \times T \mid u \cdot 1 = v \}, \\ \lhd^T &= \{ (u, v) \in T \times T \mid u \prec_2^T \circ (\prec_1^T)^* \, v \}. \end{aligned}$$

The relation $\lhd^T$ is the child relation on the nodes of $\boldsymbol{T}$.

Let $\mathbb{T}_\Sigma$ be the set of all unranked trees over Σ, encoded as binary trees over Σ. If $A, Z \subseteq \Sigma$ and $I \subseteq \Sigma \times \Sigma^+$ are finite sets, define $\mathrm{Loc}(A, Z, I)$ to be the set of all trees $\boldsymbol{T} = (T, \ell)$ in $\mathbb{T}_\Sigma$ that satisfy the following conditions:

L1. $\ell(\varepsilon) \in A$.
L2. $u \in T - \mathrm{dom}(\prec_2^T)$ implies $\ell(u) \in Z$.
L3. $u \prec_2^T v_1 \prec_1^T \dots \prec_1^T v_n \notin \mathrm{dom}(\prec_1^T)$ $(n \geq 1)$ implies $(\ell(u), \ell(v_1) \dots \ell(v_n)) \in I$.

A set $L \subseteq \mathbb{T}_\Sigma$ is *local* [33,31] if there are finite sets $A, Z \subseteq \Sigma$ and $I \subseteq \Sigma \times \Sigma^+$ such that $L = \mathrm{Loc}(A, Z, I)$.

We introduce a more restrictive notion of locality. If $A, Z, Y \subseteq \Sigma$ and $K, J \subseteq \Sigma \times \Sigma$, define $\mathrm{SLoc}(A, Z, K, Y, J)$ to be the set of all trees $\boldsymbol{T} = (T, \ell)$ in $\mathbb{T}_\Sigma$ that satisfy the following conditions:

SL1. $\ell(\varepsilon) \in A$.
SL2. $u \in T - \mathrm{dom}(\prec_2^T)$ implies $\ell(u) \in Z$.
SL3. $u \prec_2^T v$ implies $(\ell(u), \ell(v)) \in K$.
SL4. $u \neq \varepsilon$ and $u \in T - \mathrm{dom}(\prec_1^T)$ imply $\ell(u) \in Y$.
SL5. $u \prec_1^T v$ implies $(\ell(u), \ell(v)) \in J$.

We call $L \subseteq \mathbb{T}_\Sigma$ *super-local* if there are finite sets $A, Z, Y \subseteq \Sigma$ and $K, J \subseteq \Sigma \times \Sigma$ such that $L = \mathrm{SLoc}(A, Z, K, Y, J)$.[7]

A set of strings $L \subseteq \Sigma^+$ is *local*[8] if there are finite sets $A, Z \subseteq \Sigma$ and $I \in \Sigma^2$ such that

$$L = A\Sigma^* \cap \Sigma^* Z - (\Sigma^*(\Sigma^2 - I)\Sigma^*).$$

In this paper, we allow the alphabet Σ to be infinite, but any local subset of Σ^+ is included in Σ_0^+ for some finite subset Σ_0 of Σ; likewise, any local or super-local subset of $\mathbb{T}_\Sigma$ is included in $\mathbb{T}_{\Sigma_0}$ for some finite $\Sigma_0 \subseteq \Sigma$.

We extend the string encoding function **enc** to a function from $\mathscr{P}(\mathbb{T}_\Sigma)$ to $\mathscr{P}(\Gamma_\Sigma^+)$ by $\mathbf{enc}(L) = \{ \mathbf{enc}(\boldsymbol{T}) \mid \boldsymbol{T} \in L \}$.

[6] When R and S are binary relations on some set, we write $R \circ S$ for the composition of R and S, and write R^* for the reflexive transitive closure of R.

[7] This notion of super-locality was called $\widetilde{F}$*-locality* by Takahashi [31].

[8] This notion of a local set is slightly different from McNaughton and Papert's [20] notion of a *strictly 2-testable* language. In the literature, a local set of strings is sometimes called *strictly 2-local* (for example, [25]). Eilenberg [6], Takahashi [31], and Perrin [22] use "local" in the present sense. Local sets of strings were called "standard regular events" by Chomsky and Schützenberger [3].

Lemma 2. *Let $L \subseteq \mathbb{T}_\Sigma$. If L is super-local, then there is a local set of strings $L' \subseteq \Gamma_\Sigma^+$ such that* $\mathbf{enc}(L) = L' \cap D'_\Sigma$.

Proof. Suppose that $A, Z, Y \subseteq \Sigma, K, J \subseteq \Sigma \times \Sigma$ are finite sets such that $L = \mathrm{SLoc}(A, Z, K, Y, J)$. Let

$$\begin{aligned} A' &= \{\, [_c \mid c \in A \,\}, \\ Z' &= \{\,]_c \mid c \in A \,\}, \\ I &= \{\, [_c\, [_d \mid (c,d) \in K \,\} \cup \{\, [_c\,]_c \mid c \in Z \,\} \cup \{\,]_c\, [_d \mid (c,d) \in J \,\} \cup \\ &\quad\ \{\,]_c\,]_d \mid c \in Y, d \in \Sigma \,\}. \end{aligned}$$

Let $L' \subseteq \Gamma_\Sigma^+$ be the local set of strings defined by

$$L' = A'\Gamma_\Sigma^* \cap \Gamma_\Sigma^* Z' - (\Gamma_\Sigma^*(\Gamma_\Sigma^2 - I)^*\Gamma_\Sigma^*).$$

We show that $\mathbf{enc}(L) = L' \cap D'_\Sigma$.

To prove $\mathbf{enc}(L) \subseteq L' \cap D'_\Sigma$, suppose $\boldsymbol{T} = (T, \ell) \in L$. Since we know that $\mathbf{enc}(\boldsymbol{T}) \in D'_\Sigma$, it suffices to show $\mathbf{enc}(\boldsymbol{T}) \in L'$. If c is the label of the root of $\boldsymbol{T}$, then by the definition of L, $c \in A$, so the first and last symbols of $\mathbf{enc}(\boldsymbol{T})$ must be $[_c \in A'$ and $]_c \in Z'$, respectively. So $\mathbf{enc}(\boldsymbol{T}) \in A'\Gamma_\Sigma^* \cap \Gamma_\Sigma^* Z'$. Now suppose $ab \in \Gamma_\Sigma^2$ is a substring of $\mathbf{enc}(\boldsymbol{T})$. We need to show $ab \in I$. From the definition of $\mathbf{enc}$, it is clear that there are two nodes u, v of $\boldsymbol{T}$ labeled c and d, respectively, such that one of the following conditions holds:

- $a = [_c$, $b = [_d$, and $u \prec_2^T v$. In this case, $\boldsymbol{T} \in L$ implies $(c,d) \in K$.
- $a = [_c$, $b =]_d$, $u = v \notin \mathrm{dom}(\prec_2^T)$. In this case, $\boldsymbol{T} \in L$ implies $c = d \in Z$.
- $a =]_c$, $b = [_d$, and $u \prec_1^T v$. In this case, $\boldsymbol{T} \in L$ implies $(c,d) \in J$.
- $a =]_c$, $b =]_d$, $u \neq \varepsilon$, and $u \notin \mathrm{dom}(\prec_1^T)$. In this case, $\boldsymbol{T} \in L$ implies $c \in Y$.

In each case, we have $ab \in I$. This shows that $\mathbf{enc}(\boldsymbol{T}) \notin \Gamma_\Sigma^*(\Gamma_\Sigma^2 - I)^*\Gamma_\Sigma^*$. We have shown that $\mathbf{enc}(\boldsymbol{T}) \in L'$.

To prove $L' \cap D'_\Sigma \subseteq \mathbf{enc}(L)$, suppose $s \in L' \cap D'_\Sigma$. Since $s \in D'_\Sigma$, there is some $\boldsymbol{T} \in \mathbb{T}_\Sigma$ such that $\mathbf{enc}(\boldsymbol{T}) = s$. We show that $\boldsymbol{T}$ satisfies the conditions SL1–SL5 for membership in $L = \mathrm{SLoc}(A, Z, K, Y, J)$.

SL1. Let c be the label of the root of $\boldsymbol{T}$. Then $s = \mathbf{enc}(\boldsymbol{T})$ starts with $[_c$ and ends with $]_c$. Since $s \in L'$, it must be that $[_c \in A'$ and $]_c \in Z'$. It follows that $c \in A$.

SL2. Let c be the label of any node $u \in T - \mathrm{dom}(\prec_2^T)$. Then from the definition of $\mathbf{enc}$, the string $[_c\,]_c$ must be a substring of s. Since $s \in L'$, it must be that $[_c\,]_c \in I$. It follows that $c \in Z$.

SL3. Let u, v be nodes of $\boldsymbol{T}$ labeled c and d, respectively, such that $u \prec_2^T v$. By the definition of $\mathbf{enc}$, it is easy to see that $[_c\, [_d$ is a substring of s and hence in I. It follows that $(c,d) \in K$.

SL4. Let u be any non-root node of $\boldsymbol{T}$ labeled c such that $u \notin \mathrm{dom}(\prec_1^T)$. Let d be the label of the parent of u. By the definition of $\mathbf{enc}$, $]_c\,]_d$ must be a substring of s and hence in I. It follows that $c \in Y$.

SL5. Let u, v be nodes of $\boldsymbol{T}$ labeled c and d, respectively such that $u \prec_1^T v$. By the definition of **enc**, $]_c\,[_d$ must be a substring of s and hence in I. It follows that $(c, d) \in J$.

We have shown that $\boldsymbol{T} \in L = \mathrm{SLoc}(A, Z, K, Y, J)$. ☐

Note that the converse of the above lemma does not necessarily hold, because L' can place a restriction on $]_d$ that can follow $]_c$. For example, $L = \{a(bc), a(bd), e(bc)\}$ is not super-local, even though $\mathbf{enc}(L)$ is local.

A mapping $\pi \colon \Sigma \to \Sigma'$ is called a *projection*. A projection π is extended to a function from $\mathbb{T}_\Sigma$ to $\mathbb{T}_{\Sigma'}$ and to a function from $\mathscr{P}(\mathbb{T}_\Sigma)$ to $\mathscr{P}(\mathbb{T}_{\Sigma'})$ in obvious ways.

Lemma 3. *Let $L \subseteq \mathbb{T}_\Sigma$ be a local set. There exist a finite alphabet Σ', a super-local set $L' \subseteq \mathbb{T}_{\Sigma'}$, and a projection $\pi \colon \Sigma' \to \Sigma$ such that $L = \pi(L')$. Moreover, π maps L' bijectively to L.*

Proof. Let $\boldsymbol{T} \in L$. We change the label of each node v of $\boldsymbol{T}$ by a pair $(c_1 \dots c_n, i)$, where $c_1 \dots c_n$ is the string of labels $c_1, \dots, c_n$ of the siblings of v, including v itself, in the order from left to right, and i is the position of v among its siblings. The relabeled trees obtained this way form a super-local set, and we can get back the original trees by a projection.

Formally, let[9]

$$\Sigma'' = \{\, (w, i) \mid w \in \Sigma^+, 1 \le i \le |w| \,\},$$

and define a projection $\pi \colon \Sigma'' \to \Sigma$ by

$$\pi((c_1 \dots c_n), i) = c_i.$$

Suppose that $A, Z \subseteq \Sigma$ and $I \in \Sigma \times \Sigma^+$ are finite sets such that $L = \mathrm{Loc}(A, Z, I)$. Let

$$\begin{aligned} F &= A \cup \{\, w \in \Sigma^+ \mid (c, w) \in I \,\}, \\ \Sigma' &= \{\, (w, i) \in \Sigma'' \mid w \in F \,\}. \end{aligned}$$

Note that Σ' is a finite subset of Σ''. Let

$$\begin{aligned} A' &= \{\, (c, 1) \mid c \in A \,\}, \\ Z' &= \{\, (c_1 \dots c_n, i) \in \Sigma' \mid c_i \in Z \,\}, \\ K &= \{\, ((d_1 \dots d_l, i), (c_1 \dots c_n, 1)) \mid (d_1 \dots d_l, i) \in \Sigma', (d_i, c_1 \dots c_n) \in I \,\} \\ Y &= \{\, (c_1 \dots c_n, i) \in \Sigma' \mid i = n \,\}, \\ J &= \{\, ((c_1 \dots c_n, i), (c_1 \dots c_n, i+1)) \mid (c_1 \dots c_n, i) \in \Sigma', i \le n - 1 \,\}. \end{aligned}$$

[9] If w is a string, we write $|w|$ for the length of w. We use $|\cdot|$ both for the length of a string and for the cardinality of a set. The context should make it clear which is intended.

These sets are all finite. Let $L' \subseteq \mathbb{T}_{\Sigma''}$ be the super-local set defined by $L' = \mathrm{SLoc}(A', Z', K, Y, J)$. Clearly, $L' \subseteq \mathbb{T}_{\Sigma'}$. We show that L' and π (restricted to Σ') satisfy the required properties.

For each $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}}) \in \mathbb{T}_{\Sigma}$, define a tree $\hat{\boldsymbol{T}} = (T, \ell^{\hat{\boldsymbol{T}}}) \in \mathbb{T}_{\Sigma''}$ by

$$\ell^{\hat{\boldsymbol{T}}}(\varepsilon) = (\ell^{\boldsymbol{T}}(\varepsilon), 1), \tag{1}$$

$$\ell^{\hat{\boldsymbol{T}}}(u \cdot 2 \cdot 1^{i-1}) = (\ell^{\boldsymbol{T}}(u \cdot 2)\ell^{\boldsymbol{T}}(u \cdot 2 \cdot 1) \dots \ell^{\boldsymbol{T}}(u \cdot 2 \cdot 1^{n-1}), i) \tag{2}$$
$$\text{if } u \cdot 2 \cdot 1^{n-1} \in T - \mathrm{dom}(\prec_1^T) \text{ and } 1 \leq i \leq n.$$

It is clear that $\pi(\hat{\boldsymbol{T}}) = \boldsymbol{T}$ for all $\boldsymbol{T} \in \mathbb{T}_{\Sigma}$. Our goal is to show

$$L' = \{\, \hat{\boldsymbol{T}} \mid \boldsymbol{T} \in L \,\}.$$

This clearly implies that π is a bijection from L' to L.

We begin by showing that for all $\boldsymbol{T} \in \mathbb{T}_{\Sigma}$,

$$\boldsymbol{T} \in L \text{ if and only if } \hat{\boldsymbol{T}} \in L'. \tag{3}$$

This follows from five observations. Firstly, note the following:

- Suppose $u \in T - \mathrm{dom}(\prec_1^T)$. Then $\ell^{\hat{\boldsymbol{T}}}(u)$ is of the form $(c_1 \dots c_n, n)$, which means that $\ell^{\hat{\boldsymbol{T}}}(u) \in Y$ if $\ell^{\hat{\boldsymbol{T}}}(u) \in \Sigma'$.
- Suppose $u \prec_1^T v$. Then $(\ell^{\hat{\boldsymbol{T}}}(u), \ell^{\hat{\boldsymbol{T}}}(v))$ is of the form $((c_1 \dots c_n, i), (c_1 \dots c_n, i+1))$, which means that $(\ell^{\hat{\boldsymbol{T}}}(u), \ell^{\hat{\boldsymbol{T}}}(v)) \in J$ if $\ell^{\hat{\boldsymbol{T}}}(u) \in \Sigma'$.

Thus, $\hat{\boldsymbol{T}}$ satisfies the last two conditions SL4 and SL5 for membership in $\mathrm{SLoc}(A', Z', K, Y, J)$ whenever $\hat{\boldsymbol{T}} \in \mathbb{T}_{\Sigma'}$. Secondly, the following biconditional always holds:

- $\ell^{\boldsymbol{T}}(\varepsilon) \in A$ if and only if $\ell^{\hat{\boldsymbol{T}}}(\varepsilon) \in A'$.

Thirdly, the following biconditional holds whenever $\ell^{\boldsymbol{T}}(v) \in \Sigma'$:

- $\ell^{\boldsymbol{T}}(v) \in Z$ if and only if $\ell^{\hat{\boldsymbol{T}}}(v) \in Z'$.

Fourthly, if $u \cdot 2 \cdot 1^{n-1} \in T - \mathrm{dom}(\prec_1^T)$ and $\ell^{\hat{\boldsymbol{T}}}(u) \in \Sigma'$, then the following biconditional holds:

- $(\ell^{\boldsymbol{T}}(u), \ell^{\boldsymbol{T}}(u \cdot 2)\ell^{\boldsymbol{T}}(u \cdot 2 \cdot 1) \dots \ell^{\boldsymbol{T}}(u \cdot 2 \cdot 1^{n-1})) \in I$ if and only if $(\ell^{\hat{\boldsymbol{T}}}(u), \ell^{\hat{\boldsymbol{T}}}(u \cdot 2)) \in K$.

Lastly, it is easy to see that $\boldsymbol{T} \in L$ implies $\hat{\boldsymbol{T}} \in \mathbb{T}_{\Sigma'}$. Combining these five observations, we get (3).

It follows from the "only if" direction of (3) that $\{\, \hat{\boldsymbol{T}} \mid \boldsymbol{T} \in L \,\} \subseteq L'$. To establish the converse inclusion, we show that

$$\text{if } \boldsymbol{T}' \in L' \text{ and } \boldsymbol{T} = \pi(\boldsymbol{T}'), \text{ then } \boldsymbol{T}' = \hat{\boldsymbol{T}}.$$

This together with the "if" direction of (3) clearly implies $L' \subseteq \{ \hat{\boldsymbol{T}} \mid \boldsymbol{T} \in L \}$.

So suppose $\boldsymbol{T}' = (T, \ell^{\boldsymbol{T}'}) \in L'$, and let $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}}) = \pi(\boldsymbol{T}')$. All we need to show is that the equations (1) and (2) hold with $\boldsymbol{T}'$ in place of $\hat{\boldsymbol{T}}$. As for (1), it follows from the fact that $\ell^{\boldsymbol{T}'}(\varepsilon) \in A'$. As for (2), suppose $u \cdot 2 \cdot 1^{n-1} \in T - \mathrm{dom}(\prec_2^T)$. Since $(\ell^{\boldsymbol{T}'}(u), \ell^{\boldsymbol{T}'}(u \cdot 2)) \in K$, we have $\ell^{\boldsymbol{T}'}(u \cdot 2) = (c_1 \dots c_m, 1)$ for some $c_1 \dots c_m \in F$. Since for all $i \leq n-1$ we must have $(\ell^{\boldsymbol{T}'}(u \cdot 2 \cdot 1^{i-1}), \ell^{\boldsymbol{T}'}(u \cdot 2 \cdot 1^i)) \in J$, we get $\ell^{\boldsymbol{T}'}(u \cdot 2 \cdot 1^{i-1}) = (c_1 \dots c_m, i) \in \Sigma'$ for $i = 1, \dots, n$. This implies $n \leq m$. But $\ell^{\boldsymbol{T}'}(u \cdot 2 \cdot 1^{n-1}) = (c_1 \dots c_m, n)$ must be in Y, so $m = n$. Since $\pi((c_1 \dots c_n, i)) = c_i = \ell^{\boldsymbol{T}}(u \cdot 2 \cdot 1^{i-1})$, it follows that (2) holds with $\boldsymbol{T}'$ in place of $\hat{\boldsymbol{T}}$. □

We assume the standard definition of the *yield* function $\mathbf{y} \colon \mathbb{T}_\Sigma \to \Sigma^+$. Using the tem notation for unranked trees, we can define it as follows:

$$\mathbf{y}(c) = c,$$
$$\mathbf{y}(c(t_1 \dots t_n)) = \mathbf{y}(t_1) \dots \mathbf{y}(t_n).$$

As is well known, a set of strings is a context-free language if and only if it is the yield image of a local set of trees.

Let us call a tree $\boldsymbol{T} = (T, \ell) \in \mathbb{T}_\Sigma$ *disjointly labeled with* Σ_0, Σ_1 if (i) Σ_0 and Σ_1 are disjoint subsets of Σ, (ii) $u \in \mathrm{dom}(\prec_2^T)$ implies $\ell(u) \in \Sigma_1$, and (iii) $u \in T - \mathrm{dom}(\prec_2^T)$ implies $\ell(u) \in \Sigma_0$. Let Σ_0, Σ_1 be disjoint sets and let

$$\mathbb{T}_{\Sigma_0}^{\Sigma_1} = \{ \boldsymbol{T} \in \mathbb{T}_{\Sigma_0 \cup \Sigma_1} \mid \boldsymbol{T} \text{ is disjointly labeled with } \Sigma_0, \Sigma_1 \}.$$

On $\mathbb{T}_{\Sigma_0}^{\Sigma_1}$, the yield function $\mathbf{y} \colon \mathbb{T}_{\Sigma_0}^{\Sigma_1} \to \Sigma_0^+$ can be expressed as the composition

$$\mathbf{y} = \mathbf{h}_{\Sigma_0, \Sigma_1} \circ \mathbf{enc}$$

of the string encoding function $\mathbf{enc}$ and an alphabetic homomorphism $\mathbf{h}_{\Sigma_0,\Sigma_1} \colon (\Gamma_{\Sigma_0 \cup \Sigma_1})^* \to \Sigma_0^*$ defined by

$$\mathbf{h}_{\Sigma_0,\Sigma_1}(\llbracket_c) = \begin{cases} c & \text{if } c \in \Sigma_0, \\ \varepsilon & \text{if } c \in \Sigma_1, \end{cases}$$
$$\mathbf{h}_{\Sigma_0,\Sigma_1}(\rrbracket_c) = \varepsilon.$$

Lemma 4. *Let $L \subseteq \Sigma^*$ be a context-free language. There exist a set Υ disjoint from Σ and a local set $K \subseteq \mathbb{T}_\Sigma^\Upsilon$ such that $L = \mathbf{y}(K) = h_{\Sigma,\Upsilon}(\mathbf{enc}(K))$.*

Proof. Let $G = (N, \Sigma, P, S)$ be a context-free grammar for L. Clearly, the parse trees of G form a local subset K of $\mathbb{T}_\Sigma^N$, and $L = \mathbf{y}(K) = \mathbf{h}_{\Sigma,N}(\mathbf{enc}(K))$.[10] □

[10] This also follows from the fact that a local set of trees is always obtained from a local set of disjointly labeled trees by a projection that does not change the labels of leaves.

Conversely, $h(\mathbf{enc}(K))$ is always a context-free language whenever K is a local set of trees and h is a homomorphism.

A projection $\pi\colon \Sigma' \to \Sigma$ induces a projection $\widehat{\pi}\colon \Gamma_{\Sigma'} \to \Gamma_{\Sigma}$ in an obvious way:

$$\widehat{\pi}([_c) = [_{\pi(c)}, \qquad \widehat{\pi}(]_c) =]_{\pi(c)} .$$

Lemma 5. *Let $\pi\colon \Sigma' \to \Sigma$ be a projection and $L \subseteq \mathbb{T}_{\Sigma'}$. Then $\mathbf{enc}(\pi(L)) = \widehat{\pi}(\mathbf{enc}(L))$.*

We can use Lemmas 2 through 5 to show that every context-free language L can be represented as $L = h(R \cap D'_n)$ with some alphabetic homomorphism h and local set R. The Chomsky-Schützenberger theorem, however, is stated in terms of the Dyck language D_n rather than the set D'_n of Dyck primes. The following lemma bridges the representation in terms of D'_n and that in terms of D_n.

Lemma 6. *Let $L \subseteq \Gamma^+_\Sigma$ be a local set of strings. Then there exist a finite alphabet Σ', a projection $\pi\colon \Sigma' \to \Sigma$, and a local set $L' \subseteq \Gamma^+_{\Sigma'}$ such that $L \cap D'_\Sigma = \widehat{\pi}(L' \cap D_{\Sigma'})$. Moreover, $\hat{\pi}$ maps $L' \cap D_{\Sigma'}$ bijectively to $L \cap D'_\Sigma$.*

Proof. Let $A, Z \subseteq \Gamma_\Sigma, I \subseteq \Gamma^2_\Sigma$ be finite sets such that $L = A\Gamma^*_\Sigma \cap \Gamma^*_\Sigma Z - (\Gamma^*_\Sigma(\Gamma^2_\Sigma - I)\Gamma^*_\Sigma)$. We may assume without loss of generality that Σ is finite. Let $\Sigma' = \Sigma \cup \{\, \bar{c} \mid c \in \Sigma \,\}$. Let

$$\begin{aligned}
A' &= \{\, [_{\bar{c}} \mid [_c \in A \,\},\\
Z' &= \{\,]_{\bar{c}} \mid]_c \in Z \,\},\\
I' &= I \cup \{\, [_{\bar{c}}\, d \mid [_c\, d \in I \,\} \cup \{\, d\,]_{\bar{c}} \mid d\,]_c \in I \,\} \cup \{\, [_{\bar{c}}\,]_{\bar{c}} \mid [_c\,]_c \in I \,\},
\end{aligned}$$

and put

$$L' = A'\Gamma^*_{\Sigma'} \cap \Gamma^*_{\Sigma'} Z' - (\Gamma^*_{\Sigma'}(\Gamma^2_{\Sigma'} - I')\Gamma^*_{\Sigma'}).$$

Define $\pi\colon \Sigma' \to \Sigma$ by

$$\pi(c) = c, \quad \pi(\bar{c}) = c$$

for each $c \in \Sigma$. Then it is clear that $\widehat{\pi}(D'_{\Sigma'}) = D'_\Sigma$ and $L' \cap D_{\Sigma'} = L' \cap D'_{\Sigma'}$. Also, since $\widehat{\pi}$ maps A', Z', I' to A, Z, I, respectively, it is easy to see that $\widehat{\pi}(L') \subseteq L$. This establishes $\widehat{\pi}(L' \cap D_{\Sigma'}) \subseteq L \cap D'_\Sigma$. Now suppose $w \in L \cap D'_\Sigma$. Then $w = [_c\, v\,]_c$ for some $c \in \Sigma$ and $v \in D_\Sigma$. Let $w' = [_{\bar{c}}\, v\,]_{\bar{c}}$. Then $w' \in D'_{\Sigma'}$ and $\widehat{\pi}(w') = w$. Since $w \in L$, $[_c \in A$ and $]_c \in Z$, so it follows that $[_{\bar{c}} \in A'$ and $]_{\bar{c}} \in Z'$. It is also easy to see that if $a_1 a_2$ is a substring of w', $a_1 a_2 \in I'$. Therefore, $w' \in L'$, and this shows that $L \cap D'_\Sigma \subseteq \widehat{\pi}(L' \cap D'_{\Sigma'}) = \widehat{\pi}(L' \cap D_{\Sigma'})$. It is also easy to see that w' is the unique element of L' that $\widehat{\pi}$ maps to w. ☐

Lemma 7. *If $L \subseteq \mathbb{T}_\Sigma$ is a local set, then there exist a finite alphabet Σ', a projection $\pi\colon \Sigma' \to \Sigma$, and a local set $L' \subseteq \Gamma^+_{\Sigma'}$ such that $\mathbf{enc}(L) = \widehat{\pi}(L' \cap D_{\Sigma'})$. Moreover, $\mathbf{enc}^{-1} \circ \widehat{\pi}$ maps $L' \cap D_{\Sigma'}$ bijectively to L.*

Proof. By Lemma 3, there exist a projection $\pi_1 \colon \Sigma_1 \to \Sigma$ and a super-local set $L_1 \subseteq \mathbb{T}_{\Sigma_1}$ such that $L = \pi_1(L_1)$ and π_1 is a bijection from L_1 to L. By Lemma 2, there exists a local set $L_2 \subseteq \Gamma^*_{\Sigma_1}$ such that $\mathbf{enc}(L_1) = L_2 \cap D'_{\Sigma_1}$. By Lemma 6, there exist a projection $\pi_3 \colon \Sigma' \to \Sigma_1$ and a local set $L' \subseteq \Gamma^*_{\Sigma'}$ such that $L_2 \cap D'_{\Sigma_1} = \widehat{\pi_3}(L' \cap D_{\Sigma'})$ and $\widehat{\pi_3}$ is a bijection from $L' \cap D_{\Sigma'}$ to $L_2 \cap D'_{\Sigma_1}$. By Lemma 5, $\mathbf{enc}(L) = \mathbf{enc}(\pi_1(L_1)) = \widehat{\pi_1}(\mathbf{enc}(L_1))$. Since $\mathbf{enc}$ is injective, $\widehat{\pi_1}$ is a bijection from $\mathbf{enc}(L_1)$ to $\mathbf{enc}(L)$. Taking these all together, we get

$$\begin{aligned}
\mathbf{enc}(L) &= \mathbf{enc}(\pi_1(L_1)) \\
&= \widehat{\pi_1}(\mathbf{enc}(L_1)) \\
&= \widehat{\pi_1}(L_2 \cap D'_{\Sigma_1}) \\
&= \widehat{\pi_1}(\widehat{\pi_3}(L' \cap D_{\Sigma'})) \\
&= (\widehat{\pi_1} \circ \widehat{\pi_3})(L' \cap D_{\Sigma'}) \\
&= \widehat{\pi}(L' \cap D_{\Sigma'}),
\end{aligned}$$

where $\pi = \pi_1 \circ \pi_3$. Since $\widehat{\pi_3}$ is a bijection from $L' \cap D_{\Sigma'}$ to $L_2 \cap D'_{\Sigma_1} = \mathbf{enc}(L_1)$ and $\widehat{\pi_1}$ is a bijection from $\mathbf{enc}(L_1)$ to $\mathbf{enc}(L)$, $\widehat{\pi}$ is a bijection from $L' \cap D_{\Sigma'}$ to $\mathbf{enc}(L)$, and the second statement of the lemma follows. ◻

Theorem 8 (Chomsky and Schützenberger). *A language $L \subseteq \Sigma^*$ is context-free if and only if there exist a positive integer n, a local set $R \subseteq \Gamma_n^+$, and an alphabetic homomorphism $h \colon \Gamma_n^* \to \Sigma^*$ such that $L = h(R \cap D_n)$.*

Proof. The "if" direction is by standard closure properties of the context-free languages. For the "only if" direction, let $L \subseteq \Sigma^*$ be a context-free language. Then Lemma 4 gives an alphabet Υ disjoint from Σ and a local set $K \subseteq \mathbb{T}_{\Sigma,\Upsilon}$ such that $L = h_{\Sigma,\Upsilon}(\mathbf{enc}(K))$. By Lemma 7, there are a projection $\pi \colon \Upsilon' \to \Sigma \cup \Upsilon$ and a local set $R \subseteq \Gamma^+_{\Upsilon'}$ such that $\mathbf{enc}(K) = \widehat{\pi}(R \cap D_{\Upsilon'})$. We have

$$\begin{aligned}
L &= h_{\Sigma,\Upsilon}(\mathbf{enc}(K)) \\
&= h_{\Sigma,\Upsilon}(\widehat{\pi}(R \cap D_{\Upsilon'})),
\end{aligned}$$

so the required condition holds with $n = |\Upsilon'|$ and $h = h_{\Sigma,\Upsilon} \circ \hat{\pi}$.[11] ◻

In the proof of Theorem 8, $\mathbf{enc}^{-1} \circ \widehat{\pi}$ is a bijection from $R \cap D_{\Upsilon'}$ to K. (See the second statement in Lemma 7.) If K is the set of derivation trees of a context-free grammar for L, then an element s of $R \cap D_{\Upsilon'}$ represents both the element $t = \mathbf{enc}^{-1}(\widehat{\pi}(s))$ of K and the element $h_{\Sigma,\Upsilon}(\widehat{\pi}(s)) = h_{\Sigma,\Upsilon}(\mathbf{enc}(t)) = \mathbf{y}(t)$ of L. Moreover, every pair $(t, \mathbf{y}(t))$ with $t \in K$ is so represented. This is an important consequence of the theorem explicitly noted by Chomsky [4, page 377], though rarely emphasized since.[12]

[11] Here, $|\Upsilon'|$ denotes the cardinality of the set Υ'. See footnote 9.

[12] Instead of a super-local set of trees, Chomsky [4] used the notion of a *modified normal grammar*, a restricted kind of grammar in Chomsky normal form.

We took a rather long route to the Chomsky-Schützenberger theorem. Our generalization of the theorem to multi-dimensional tree languages follows a similar path, except that an analogue of Lemma 4 is not needed, since the multi-dimensional counterpart of the function **enc** is not exactly a generalization of the usual notion.

4 Multi-dimensional Trees

Multi-dimensional trees were introduced by Rogers [24,23]. In an ordinary (labeled, ordered unranked) tree, the set of children of a node forms a linearly ordered sequence of labeled nodes, i.e., a string. In an *m-dimensional tree* ($m \geq 1$), the set of children of a node (if non-empty) forms an $(m-1)$-dimensional tree. A 0-dimensional tree just consists of a single labeled node.

Unlike Rogers [24,23], who introduces the higher-dimensional tree as a new kind of object, we prefer to define an m-dimensional tree as a special kind of ordinary m-ary tree.[13] The first-child-next-sibling encoding of unranked trees will be a special case of this definition for $m = 2$.

We use finite strings of elements of $[1, m] = \{1, \dots, m\}$ to represent nodes of m-ary trees. We write $u \cdot v$ for the concatenation of finite strings u, v over $[1, m]$, and write ε for the empty string.

An *m-ary tree domain* is any non-empty, finite, prefix-closed subset of $[1, m]^*$. (Since $\varnothing^* = \{\varepsilon\}$, the only 0-ary tree domain is $\{\varepsilon\}$.) If T is an m-ary tree domain, we write $u \prec_i^T v$ to mean $u, v \in T$ and $u \cdot i = v$. If Σ is a (possibly infinite) set of symbols, an *m-ary tree* over Σ is a pair (T, ℓ), where T is an m-ary tree domain and ℓ is a function from T to Σ.

If $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}})$ is an m-ary tree and $U \subseteq T$ is an m-ary tree domain, then the *restriction of* $\boldsymbol{T}$ *to* U is the m-ary tree

$$\boldsymbol{T} \upharpoonright U = (U, \ell^{\boldsymbol{T}} \upharpoonright U).$$

When $u \in T$, let

$$T/u = \{ v \mid uv \in T \}.$$

Then T/u is an m-ary tree domain and the *subtree of* $\boldsymbol{T}$ *rooted at* u is defined by

$$\boldsymbol{T}/u = (T/u, \ell),$$

where $\ell(v) = \ell^{\boldsymbol{T}}(uv)$.

Recall that a first-child-next-sibling encoding of an unranked tree is a binary tree (T, ℓ) such that $1 \notin T$. Analogously, an m-dimensional tree is an m-ary tree (T, ℓ) such that T is an m-ary tree domain included in a certain special subset of

[13] To be precise, our m-dimensional trees form a special class of m-ary *cardinal trees* in the sense of Benoit et al. [2]. In m-ary cardinal trees, each node has m slots for children, each of which may or may not be occupied, independently of the other slots. Cardinal trees are also known as *tries*.

$[1,m]^*$. For each natural number m, define a subset $\mathbb{P}_m$ of $[1,m]^*$ by recursion, as follows:

$$\mathbb{P}_0 = \{\varepsilon\},$$
$$\mathbb{P}_m = (m \cdot \mathbb{P}_{m-1})^* \quad \text{for } m > 1.$$

In other words, $w \in \mathbb{P}_m$ if and only if $w = i \cdot v$ implies $i = m$ and $w = u \cdot i \cdot j \cdot v$ implies $j \geq i - 1$. For $m \geq 0$, an *m-dimensional tree* (over Σ) is an m-ary tree (over Σ) $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}})$ such that $T \subseteq \mathbb{P}_m$. For $m \geq 1$, an *m-dimensional hedge* (over Σ) is an m-ary tree (over Σ) $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}})$ such that $T \subseteq \mathbb{P}_{m-1} \cdot \mathbb{P}_m$. We write $\mathbb{T}^m_\Sigma$ and $\mathbb{H}^m_\Sigma$ to denote the set of all m-dimensional trees over Σ and the set of all m-dimensional hedges over Σ, respectively. For $m \geq 1$, all m-dimensional trees are m-dimensional hedges. Note that a 1-dimensional hedge is just a 1-dimensional tree.

Note that a 0-dimensional tree is a structure $\boldsymbol{T}_c = (\{\varepsilon\}, \{(\varepsilon, c)\})$ consisting of a single node labeled by some $c \in \Sigma$. We may identify $\boldsymbol{T}_c$ with c; under this convention, $\mathbb{T}^0_\Sigma = \Sigma$. Note that if $m \neq n$, then $\mathbb{T}^m_\Sigma \cap \mathbb{T}^n_\Sigma = \mathbb{T}^0_\Sigma$.

A 1-dimensional tree is a non-empty, linearly ordered sequence of labeled nodes. We may use a string $c_1 \dots c_n \in \Sigma^+$ to denote the 1-dimensional tree $\boldsymbol{T}_{c_1 \dots c_n} = (\{\varepsilon, 1, \dots, 1^{n-1}\}, \ell)$, where $\ell(1^{i-1}) = c_i$ for $i = 1, \dots, n$. Under this convention, $\mathbb{T}^1_\Sigma = \Sigma^+$.

The first-child-next-sibling encodings of unranked trees and unranked hedges coincide with the 2-dimensional trees and the 2-dimensional hedges, respectively; we have $\mathbb{T}^2_\Sigma = \mathbb{T}_\Sigma$.

Henceforth, we use $\boldsymbol{T}, \boldsymbol{T}', \boldsymbol{U}$, etc., as variables ranging over m-dimensional trees and m-dimensional hedges. Unless we indicate otherwise, we assume $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}}), \boldsymbol{T}' = (T', \ell^{\boldsymbol{T}'}), \boldsymbol{U} = (U, \ell^{\boldsymbol{U}})$, etc.

Let $\boldsymbol{T}$ be an m-dimensional hedge. We can see that if $u \in T$, then

$$ST_i(\boldsymbol{T}, u) = (\boldsymbol{T}/u) \restriction \{\, v \in \mathbb{P}_i \mid uv \in T \,\}$$

is always an i-dimensional tree, and

$$SH_i(\boldsymbol{T}, u) = (\boldsymbol{T}/u) \restriction \{\, v \in \mathbb{P}_{i-1} \cdot \mathbb{P}_i \mid uv \in T \,\}$$

is always an i-dimensional hedge. In particular, when $u \cdot m \in T$, the subtree $\boldsymbol{T}/(u \cdot m) = SH_m(\boldsymbol{T}, u \cdot m)$ is always an m-dimensional hedge.

For $i \geq 1$, we write $u \lhd^T_i v$ to mean

$$v \in T \cap u \cdot i \cdot \mathbb{P}_{i-1}.$$

When $u \lhd^T_i v$, we say that v is a *child of u in the i-th dimension* (in $\boldsymbol{T}$). If $u \prec^T_i v$, then v is the *first* child of u in the i-th dimension. Define

$$C^T_i(u) = \{\, v \in \mathbb{P}_{i-1} \mid u \cdot i \cdot v \in T \,\} = \{\, v \mid u \lhd^T_i u \cdot i \cdot v \,\}.$$

If $u \cdot i \notin T$, that is, if $u \notin \mathrm{dom}(\prec^T_i)$, then $C^T_i(u) = \varnothing$. If $u \cdot i \in T$, define

$$\boldsymbol{C}^{\boldsymbol{T}}_i(u) = ST_{i-1}(\boldsymbol{T}, u \cdot i) = \boldsymbol{T}/(u \cdot i) \restriction C^T_i(u).$$

Then $\boldsymbol{C}_i^{\boldsymbol{T}}(u)$ is always an $(i-1)$-dimensional tree.

We assume that elements of $[1,m]^*$ are alphabetically ordered, with $k+1$ alphabetically *preceding* k. We write $u \lhd_{i,j}^T v$ to mean v is the j-th node, under this ordering, among the children of u in the i-th dimension. The *degree* of a node $v \in T$ is the number of children of v in the m-th (i.e., highest) dimension.

A subset of $\mathbb{T}_\Sigma^m$ is an *m-dimensional tree language*. We allow Σ to be an infinite set, but are usually interested in m-dimensional tree languages over some finite subset of Σ.

We call a set $L \subseteq \mathbb{T}_\Sigma^m$ *degree-bounded* if there exists a k such that for all $\boldsymbol{T} \in L$ and for all $v \in T$, the degree of v does not exceed k.

It is sometimes helpful to use term-like notations for m-dimensional hedges and trees. Let P be an $(m-1)$-ary tree domain included in $\mathbb{P}_{m-1}$ (i.e., a finite, non-empty, prefix-closed subset of $\mathbb{P}_{m-1}$), and let $u_1, \dots, u_k$ be the elements of P, in alphabetical order (which implies $u_1 = \varepsilon$). If $\boldsymbol{T}_1, \dots, \boldsymbol{T}_k \in \mathbb{T}_\Sigma^m$, then we write

$$P(\boldsymbol{T}_1, \dots, \boldsymbol{T}_k)$$

to denote an m-dimensional hedge $\boldsymbol{U} = (U, \ell^{\boldsymbol{U}}) \in \mathbb{H}_\Sigma^m$ such that

$$U = \bigcup_{i=1}^{k} u_i \cdot T_i, \qquad ST_m(\boldsymbol{U}, u_i) = \boldsymbol{T}_i.$$

(As a degenerate case, we have $\{\varepsilon\}(\boldsymbol{T}) = \boldsymbol{T}$ for any $\boldsymbol{T} \in \mathbb{T}_\Sigma^m$.) If $\boldsymbol{T} \in \mathbb{H}_\Sigma^m$ and $c \in \Sigma$, then we write

$$c -_m \boldsymbol{T}$$

to denote an m-dimensional tree $\boldsymbol{V} = (V, \ell^{\boldsymbol{V}}) \in \mathbb{T}_\Sigma^m$ such that

$$V = \{\varepsilon\} \cup m \cdot T, \qquad \ell^{\boldsymbol{V}}(\varepsilon) = c, \qquad SH_m(\boldsymbol{V}, m) = \boldsymbol{T}.$$

Combining the two notations,

$$c -_m P(\boldsymbol{T}_1, \dots, \boldsymbol{T}_k)$$

denotes the m-dimensional tree $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}})$ such that $\ell^{\boldsymbol{T}}(\varepsilon) = c$, $C_m^T(\varepsilon) = P$, and $ST_m(\boldsymbol{T}, u_i) = \boldsymbol{T}_i$, where $u_1, \dots, u_k$ is the alphabetical listing of the elements of P.[14]

Example 9. Derivation trees of a simple context-free tree grammar $G = (N, \Sigma, P, S)$ can be represented as 3-dimensional trees over the alphabet $N \cup \Sigma$. In these 3-dimensional trees, a node has children in the third dimension if and only if it is labeled by a nonterminal. For instance, the derivation tree $\pi_1(\pi_2(\pi_3))$ of the grammar from Example 1 may be represented by the 3-dimensional tree $\boldsymbol{T}$ in Fig. 1. In this tree, the node labeled by S is the root; the edges in the third dimension are colored red, those in the second dimension blue, and those in the first dimension green. For instance, the node 3 (i.e., the child of the root

[14] An equivalent notation for m-dimensional trees has been used by Kasprzik [14].

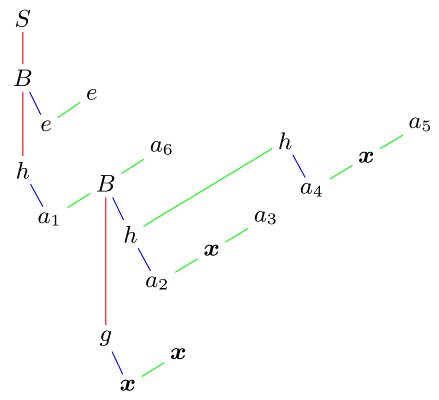

Fig. 1. A derivation tree of a simple context-free tree grammar represented as a 3-dimensional tree.

in the third dimension) is labeled by the nonterminal B, and its children in the third dimension form a 2-dimensional tree corresponding to the right-hand side of the rule $\pi_2 = B(x_1 x_2) \to h(a_1 B(h(a_2 x_1 a_3) h(a_4 x_2 a_5)) a_6)$. The numbering of variables in the rules are eschewed in favor of a single variable $\boldsymbol{x}$; the alphabetic ordering of the nodes labeled by $\boldsymbol{x}$ among the children in the third dimension of a nonterminal-labeled node is assumed to correspond to the numbering.[15] This tree can be represented in the term notation as follows (omitting the dots in the strings over $\{1, 2\}$ representing nodes):

$$
\begin{aligned}
& S -_3 \{\varepsilon, 2, 21\}(\\
& \quad B -_3 \{\varepsilon, 2, 21, 212, 2122, 21221, 212211, 2121, 21212, 212121, 2121211, 211\}(\\
& \quad\quad h, a_1, B -_3 \{\varepsilon, 2, 21\}(g, \boldsymbol{x}, \boldsymbol{x}), h, a_2, \boldsymbol{x}, a_3, h, a_4, \boldsymbol{x}, a_5, a_6 \\
& \quad), \\
& \quad e, \\
& \quad e \\
&).
\end{aligned}
$$

We have

$$
\begin{aligned}
\boldsymbol{C}_3^{\boldsymbol{T}}(3 \cdot 3 \cdot 2 \cdot 1) &= g -_2 \{\varepsilon, 1\}(\boldsymbol{x}, \boldsymbol{x}) \\
&= g(\boldsymbol{x}\boldsymbol{x}), \\
\boldsymbol{C}_2^{\boldsymbol{T}}(3 \cdot 3 \cdot 2 \cdot 1) &= h -_1 h \\
&= hh,
\end{aligned}
$$

[15] It is known that simple context-free tree grammars satisfying this condition constitute a normal form. This use of a single variable instead of numbered variables is not crucial for our purposes.

using both the notation introduced just above and the standard term and string representations for 2-dimensional and 1-dimensional trees.

5 Local and Super-local Sets of Multi-dimensional Trees

If $A, Z \subseteq \Sigma$ and $I \subseteq \Sigma \times \mathbb{T}_{\Sigma}^{m-1}$ are finite sets, we let $\mathrm{Loc}^m(A, Z, I)$ denote the set of all m-dimensional trees $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}})$ in $\mathbb{T}_{\Sigma}^{m}$ that satisfy the following conditions:

L1. $\ell^{\boldsymbol{T}}(\varepsilon) \in A$,
L2. $v \in T - \mathrm{dom}(\prec_m^T)$ implies $\ell^{\boldsymbol{T}}(v) \in Z$, and
L3. $v \in \mathrm{dom}(\prec_m^T)$ implies $(\ell^{\boldsymbol{T}}(v), \boldsymbol{C}_m^{\boldsymbol{T}}(v)) \in I$.

A set $L \subseteq \mathbb{T}_{\Sigma}^{m}$ is *local* [24,23] if there exist finite sets $A, Z \subseteq \Sigma$ and $I \subseteq \Sigma \times \mathbb{T}_{\Sigma}^{m-1}$ such that $L = \mathrm{Loc}^m(A, Z, I)$. Note that if $L \subseteq \mathbb{T}_{\Sigma}^{m}$ is local, then L must be degree-bounded; for, if $L = \mathrm{Loc}^m(A, Z, I)$, the degree of any node v of $\boldsymbol{T} \in L$ is bounded by the maximal size of $\boldsymbol{U}$ such that $(c, \boldsymbol{U}) \in I$ for some c. Clearly, the notion of locality coincides with the usual notion [22,33,31] when $m \in \{1, 2\}$.

Let $m \geq 2$. Write $\mathbb{N}_+$ for $\mathbb{N} - \{0\}$ (the set of positive integers). If $A, Z, Y \subseteq \Sigma$, $K \subseteq \Sigma \times \Sigma$, and $J \subseteq \Sigma \times \{\, P \subseteq \mathbb{P}_{m-2} \mid P \text{ is an } (m-2)\text{-ary tree domain} \,\} \times \mathbb{N}_+ \times \Sigma$ are finite sets, then we let $\mathrm{SLoc}^m(A, Z, K, Y, J)$ denote the set of all trees $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}})$ in $\mathbb{T}_{\Sigma}^{m}$ that satisfy the following conditions:

SL1. $\ell^{\boldsymbol{T}}(\varepsilon) \in A$,
SL2. $v \in T - \mathrm{dom}(\prec_m^T)$ implies $\ell^{\boldsymbol{T}}(v) \in Z$,
SL3. $u \prec_m^T v$ implies $(\ell^{\boldsymbol{T}}(u), \ell^{\boldsymbol{T}}(v)) \in K$,
SL4. $v \neq \varepsilon$ and $v \in T - \mathrm{dom}(\prec_{m-1}^T)$ imply $\ell^{\boldsymbol{T}}(v) \in Y$, and
SL5. $u \in \mathrm{dom}(\prec_{m-1}^T)$ and $u \lhd_{m-1,i}^T v$ imply $(\ell^{\boldsymbol{T}}(u), C_{m-1}^T(u), i, \ell^{\boldsymbol{T}}(v)) \in J$.

We call a set $L \subseteq \mathbb{T}_{\Sigma}^{m}$ *super-local* if there exist finite sets $A, Z, Y \subseteq \Sigma$, $K \subseteq \Sigma \times \Sigma$, and $J \subseteq \Sigma \times \{\, P \subseteq \mathbb{P}_{m-2} \mid P \text{ is an } (m-2)\text{-ary tree domain} \,\} \times \mathbb{N}_+ \times \Sigma$ such that $L = \mathrm{SLoc}^m(A, Z, K, Y, J)$. For $m = 2$, $\mathbb{P}_{m-2} = \mathbb{P}_0 = \{\varepsilon\}$, and $u \lhd_{1,i}^T v$ only if $i = 1$, so this definition generalizes our earlier definition of super-locality for subsets of $\mathbb{T}_{\Sigma} = \mathbb{T}_{\Sigma}^{2}$. It is easy to see that a degree-bounded super-local language must be local. Although we allow Σ to be infinite, any local or super-local set $L \subseteq \mathbb{T}_{\Sigma}^{m}$ must be an m-dimensional tree language over some finite subset of Σ.

Projections from Σ to Σ' are naturally extended to m-dimensional trees and hedges over Σ and to m-dimensional tree languages over Σ. The next lemma generalizes Lemma 3 to the higher-dimensional case. The proofs of two lemmas to follow (Lemmas 11 and 35) will be adaptations of the proof of this lemma.

Lemma 10. *Let $m \geq 2$. For any local m-dimensional tree language $L \subseteq \mathbb{T}_{\Sigma}^{m}$, there exist a finite alphabet Σ', a degree-bounded, super-local m-dimensional tree language $L' \subseteq \mathbb{T}_{\Sigma'}^{m}$, and a projection $\pi\colon \Sigma' \to \Sigma$ such that $L = \pi(L')$. Moreover, π maps L' bijectively to L.*

Proof. The proof parallels that of Lemma 3. The idea is to change the label of each non-root node v of $\boldsymbol{T} \in L$ to

$$(\boldsymbol{C}^{\boldsymbol{T}}_m(u), v'),$$

where $u \cdot m \cdot v' = v$ and $v' \in \mathbb{P}_{m-1}$. For uniformity, we change the label of the root from $c \in \Sigma$ to $(\boldsymbol{T}_c, \varepsilon)$, where $\boldsymbol{T}_c = (\{\varepsilon\}, \{(\varepsilon, c)\})$ is the single-node tree that we identified with c. The relabeled m-dimensional trees obtained this way form a super-local set, and we can get back the original m-dimensional trees by a projection.

Let

$$\Sigma'' = \{ (\boldsymbol{T}, v) \mid \boldsymbol{T} = (T, \ell^{\boldsymbol{T}}) \in \mathbb{T}^{m-1}_{\Sigma}, v \in T \},$$

and define a projection $\pi \colon \Sigma'' \to \Sigma$ by

$$\pi((\boldsymbol{T}, v)) = \ell^{\boldsymbol{T}}(v).$$

Suppose that $A, Z \subseteq \Sigma$ and $I \subseteq \Sigma \times \mathbb{T}^{m-1}_{\Sigma}$ are finite sets such that $L = \mathrm{Loc}^m(A, Z, I)$. Let

$$\begin{aligned} F &= \{ \boldsymbol{T}_c \mid c \in A \} \cup \{ \boldsymbol{T} \mid (c, \boldsymbol{T}) \in I \}, \\ \Sigma' &= \{ (\boldsymbol{T}, v) \in \Sigma'' \mid \boldsymbol{T} \in F \}. \end{aligned}$$

Note that Σ' is a finite subset of Σ''. Now define

$$\begin{aligned} A' &= \{ (\boldsymbol{T}_c, \varepsilon) \mid c \in A \}, \\ Z' &= \{ (\boldsymbol{T}, v) \in \Sigma' \mid \ell^{\boldsymbol{T}}(v) \in Z \}, \\ K &= \{ ((\boldsymbol{T}, v), (\boldsymbol{T}', \varepsilon)) \mid (\boldsymbol{T}, v) \in \Sigma', (\ell^{\boldsymbol{T}}(v), \boldsymbol{T}') \in I \}, \\ Y &= \{ (\boldsymbol{T}, v) \in \Sigma' \mid v \notin \mathrm{dom}(\prec^T_{m-1}) \}, \\ J &= \{ ((\boldsymbol{T}, u), C^T_{m-1}(u), i, (\boldsymbol{T}, v)) \mid (\boldsymbol{T}, u) \in \Sigma', u \lhd^T_{m-1,i} v \}. \end{aligned}$$

These are all finite sets. Let $L' \subseteq \mathbb{T}^m_{\Sigma''}$ be the super-local set defined by $L' = \mathrm{SLoc}^m(A', Z', K, Y, J)$. It is quite clear that $L' \subseteq \mathbb{T}^m_{\Sigma'}$. We show that L' and π (restricted to Σ') satisfy the required properties.

For each $\boldsymbol{T} \in \mathbb{T}^m_{\Sigma}$, define an m-dimensional tree $\hat{\boldsymbol{T}} = (T, \ell^{\hat{\boldsymbol{T}}}) \in \mathbb{T}^m_{\Sigma''}$ by

$$\ell^{\hat{\boldsymbol{T}}}(\varepsilon) = (\boldsymbol{T}_{\ell^{\boldsymbol{T}}(\varepsilon)}, \varepsilon), \tag{4}$$

$$\ell^{\hat{\boldsymbol{T}}}(u \cdot m \cdot v) = (\boldsymbol{C}^{\boldsymbol{T}}_m(u), v), \qquad \text{if } u \in \mathrm{dom}(\prec^T_m) \text{ and } v \in C^T_{m-1}(u). \tag{5}$$

It is clear that $\pi(\hat{\boldsymbol{T}}) = \boldsymbol{T}$ for all $\boldsymbol{T} \in \mathbb{T}^m_{\Sigma}$. Our goal is to show

$$L' = \{ \hat{\boldsymbol{T}} \mid \boldsymbol{T} \in L \}.$$

This clearly implies that π is a bijection from L' to L.

We show that for all $\boldsymbol{T} \in \mathbb{T}^m_{\Sigma}$,

$$\boldsymbol{T} \in L \text{ if and only if } \hat{\boldsymbol{T}} \in L'. \tag{6}$$

This follows from five observations. Firstly, note the following:

- Suppose $u \in T - \mathrm{dom}(\prec^T_{m-1})$. If $\ell^{\hat{\boldsymbol{T}}}(u) = (\boldsymbol{U}, v)$, then $v \notin \mathrm{dom}(\prec^U_{m-1})$. This means that $\ell^{\hat{\boldsymbol{T}}}(u) \in Y$ if $\ell^{\hat{\boldsymbol{T}}}(u) \in \Sigma'$.
- Suppose $s \lhd^T_m u = s \cdot m \cdot u' \lhd^T_{m-1,i} v = u \cdot (m-1) \cdot v'$. Then we have $u' \lhd^{C^T_m(s)}_{m-1,i} u' \cdot (m-1) \cdot v'$ and

$$\begin{aligned} \ell^{\hat{\boldsymbol{T}}}(u) &= (\boldsymbol{C}^{\boldsymbol{T}}_m(s), u'), \\ \ell^{\hat{\boldsymbol{T}}}(v) &= (\boldsymbol{C}^{\boldsymbol{T}}_m(s), u' \cdot (m-1) \cdot v'), \\ C^T_{m-1}(u) &= C^{C^T_m(s)}_{m-1}(u'). \end{aligned}$$

This means that $(\ell^{\hat{\boldsymbol{T}}}(u), C^T_{m-1}(u), i, \ell^{\hat{\boldsymbol{T}}}(v)) \in J$ if $\ell^{\hat{\boldsymbol{T}}}(u) \in \Sigma'$.

Thus, $\hat{\boldsymbol{T}}$ satisfies the last two conditions SL4 and SL5 for membership in $\mathrm{SLoc}^m(A', Z', K, Y, J)$ whenever $\hat{\boldsymbol{T}} \in \mathbb{T}^m_{\Sigma'}$. Secondly, the following biconditional always holds:

- $\ell^{\boldsymbol{T}}(\varepsilon) \in A$ if and only if $\ell^{\hat{\boldsymbol{T}}}(\varepsilon) \in A'$.

Thirdly, the following biconditional holds whenever $\ell^{\boldsymbol{T}}(v) \in \Sigma'$:

- $\ell^{\boldsymbol{T}}(v) \in Z$ if and only if $\ell^{\hat{\boldsymbol{T}}}(v) \in Z'$.

Fourthly, if $u \prec^T_m v$ and $\ell^{\hat{\boldsymbol{T}}}(u) \in \Sigma'$, then the following biconditional holds:

- $(\ell^{\boldsymbol{T}}(u), \boldsymbol{C}^{\boldsymbol{T}}_m(u)) \in I$ if and only if $(\ell^{\hat{\boldsymbol{T}}}(u), \ell^{\hat{\boldsymbol{T}}}(v)) \in K$.

Lastly, it is easy to see that $\boldsymbol{T} \in L$ implies $\hat{\boldsymbol{T}} \in \mathbb{T}^m_{\Sigma'}$. Combining these five observations, we get (6).

It follows from the "only if" direction of (6) that $\{\, \hat{\boldsymbol{T}} \mid \boldsymbol{T} \in L \,\} \subseteq L'$. To establish the converse inclusion, we show that

$$\text{if } \boldsymbol{T}' \in L' \text{ and } \boldsymbol{T} = \pi(\boldsymbol{T}'), \text{ then } \boldsymbol{T}' = \hat{\boldsymbol{T}}.$$

This together with the "if" direction of (6) clearly implies $L' \subseteq \{\, \hat{\boldsymbol{T}} \mid \boldsymbol{T} \in L \,\}$.

So suppose $\boldsymbol{T}' = (T, \ell^{\boldsymbol{T}'}) \in L'$, and let $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}}) = \pi(\boldsymbol{T}')$. All we need to show is that the equations (4) and (5) hold with $\boldsymbol{T}'$ in place of $\hat{\boldsymbol{T}}$.

As for (4), it follows from the fact that $\ell^{\boldsymbol{T}'}(\varepsilon) \in A'$.

As for (5), suppose $u \in \mathrm{dom}(\prec^T_m)$. Since $(\ell^{\boldsymbol{T}'}(u), \ell^{\boldsymbol{T}'}(u \cdot m)) \in K$, $\ell^{\boldsymbol{T}'}(u \cdot m) = (\boldsymbol{V}, \varepsilon)$ for some $\boldsymbol{V} \in F$. We show two things:

$$v \in C^T_m(u) \text{ implies } v \in V \text{ and } \ell^{\boldsymbol{T}'}(u \cdot m \cdot v) = (\boldsymbol{V}, v). \tag{7}$$

$$V \subseteq C^T_m(u). \tag{8}$$

It then easily follows that $\boldsymbol{V} = \boldsymbol{C}^{\boldsymbol{T}}_m(u)$ and (5) holds whenever $v \in C^T_m(u)$.

We show (7) by induction on $v \in C^T_m(u)$. For $v = \varepsilon$, we already know that $\varepsilon \in V$ and $\ell^{\boldsymbol{T}'}(u \cdot m) = (\boldsymbol{V}, \varepsilon)$. If $v \neq \varepsilon$, we can write $v = v' \cdot (m-1) \cdot v''$ with $v'' \in \mathbb{P}_{m-2}$. Suppose $u \cdot m \cdot v' \lhd^T_{m-1,i} u \cdot m \cdot v$. Since $v' \in C^T_m(u)$, by induction hypothesis, $v' \in V$ and $\ell^{\boldsymbol{T}'}(u \cdot m \cdot v') = (\boldsymbol{V}, v')$. Since $\boldsymbol{T}' \in L'$, $(\ell^{\boldsymbol{T}'}(u \cdot m \cdot v'), C^T_{m-1}(u \cdot m \cdot v'), i, \ell^{\boldsymbol{T}'}(u \cdot m \cdot v)) \in J$. By the definition of J, we must have $C^T_{m-1}(u \cdot m \cdot v') = C^V_{m-1}(v')$, which implies $v' \lhd^V_{m-1,i} v \in V$. The definition of J then implies $\ell^{\boldsymbol{T}'}(u \cdot m \cdot v) = (\boldsymbol{V}, v)$.

Having established (7), we proceed to show (8) by induction on $v \in V$. For $v = \varepsilon$, we have $\varepsilon \in C^T_m(u)$ since $u \in \mathrm{dom}(\prec^T_m)$. If $v = v' \cdot (m-1) \cdot v''$ with $v'' \in \mathbb{P}_{m-2}$, then $v' \in C^T_m(u)$ by induction hypothesis. Since $v' \in \mathrm{dom}(\prec^V_{m-1})$, $\ell^{\boldsymbol{T}'}(u \cdot m \cdot v') = (\boldsymbol{V}, v') \notin Y$. Since $\boldsymbol{T}' \in L'$, this means that $u \cdot m \cdot v' \in \mathrm{dom}(\prec^T_{m-1})$ and so $(\ell^{\boldsymbol{T}'}(u \cdot m \cdot v'), C^T_{m-1}(u \cdot m \cdot v'), 1, \ell^{\boldsymbol{T}'}(u \cdot m \cdot v' \cdot (m-1))) \in J$. Since $\ell^{\boldsymbol{T}'}(u \cdot m \cdot v') = (\boldsymbol{V}, v')$, the definition of J implies $C^T_{m-1}(u \cdot m \cdot v') = C^V_{m-1}(v')$. Since $v = v' \cdot (m-1) \cdot v'' \in V$, it follows that $v'' \in C^T_{m-1}(u \cdot m \cdot v')$ and hence $v = v' \cdot (m-1) \cdot v'' \in C^T_m(u)$.

This concludes the proof of the lemma. □

6 Encoding and Yield at Higher Dimensions

In order to prove an analogue of the Chomsky-Schützenberger theorem for the m-dimensional yields of local $(m+1)$-dimensional tree languages, we need to define the higher-dimensional counterparts of the mappings $\mathbf{enc}, \mathbf{y}$, and of the Dyck languages. Since we use 3-dimensional trees to represent derivation trees of simple context-free tree grammars, the yield function mapping 3-dimensional trees to 2-dimensional trees must be consistent with the relation between derivation trees and their tree yield of simple context-free tree grammars.

We set aside a special symbol $\boldsymbol{x}$ and use it to extend a given set Σ of symbols. The intended role of $\boldsymbol{x}$ in m-dimensional trees over $\Sigma \cup \{\boldsymbol{x}\}$ is that of a placeholder; the encoding function erases all occurrences of $\boldsymbol{x}$. We write $\mathbb{T}^m_\Sigma(n)$ to denote the set of m-dimensional trees in $\mathbb{T}^m_{\Sigma\cup\{\boldsymbol{x}\}}$ in which $\boldsymbol{x}$ labels exactly n nodes and none of these nodes have a child in the m-th dimension. Let $\boldsymbol{T} \in \mathbb{T}^m_\Sigma(n)$, $\boldsymbol{T}_1, \dots, \boldsymbol{T}_n \in \mathbb{T}^m_{\Sigma\cup\{\boldsymbol{x}\}}$, and let $u_1, \dots, u_n$ be the nodes of $\boldsymbol{T}$ labeled by $\boldsymbol{x}$, in the alphabetical order. Then we write

$$\boldsymbol{T}[\boldsymbol{T}_1, \dots, \boldsymbol{T}_n]$$

for the tree $\boldsymbol{T}' \in \mathbb{T}^m_{\Sigma\cup\{\boldsymbol{x}\}}$ such that

$$T' = T \cup u_1 \cdot T_1 \cup \dots \cup u_n \cdot T_n,$$

$$\ell^{\boldsymbol{T}'}(v) = \begin{cases} \ell^{\boldsymbol{T}}(v) & \text{if } v \in T - \{u_1, \dots, u_n\}, \\ \ell^{\boldsymbol{T}_i}(v') & \text{if } v = u_i \cdot v'. \end{cases}$$

Given an m-dimensional hedge $\boldsymbol{T} \in \mathbb{H}^m_{\Sigma\cup\{\boldsymbol{x}\}}$, define a binary relation $\blacktriangleleft^{\boldsymbol{T}}_{m,i}$ on T for each positive integer i, as follows: $u \blacktriangleleft^{\boldsymbol{T}}_{m,i} v$ if and only if v is alphabetically the i-th node in $\{\, w \mid u \lhd_m w, \ell^{\boldsymbol{T}}(w) = \boldsymbol{x} \,\}$.

Let $m \geq 2$. An m-dimensional hedge $\boldsymbol{T} \in \mathbb{H}^m_{\Sigma\cup\{\boldsymbol{x}\}}$ is *well-labeled* if

– for all $v \in T$, $\ell^{\boldsymbol{T}}(v) = \boldsymbol{x}$ implies $v \notin \mathrm{dom}(\prec_m^T) \cup \mathrm{dom}(\prec_{m-1}^T)$, and
– for all $v \in \mathrm{dom}(\prec_m^T)$, $\boldsymbol{C}_m^{\boldsymbol{T}}(v) \in \mathbb{T}_\Sigma^{m-1}(n)$ implies $|C_{m-1}(v)| = n$.

We write $\mathbb{H}^m_{\Sigma,\boldsymbol{x}}$ to denote the class of well-labeled m-dimensional hedges over $\Sigma \cup \{\boldsymbol{x}\}$. If $\boldsymbol{T} \in \mathbb{H}^m_{\Sigma,\boldsymbol{x}}$, then for each node $v \in \mathrm{dom}(\prec_m^T)$, there is a bijection between $\{\, u \in T \mid v \lhd_m^T u \text{ and } \ell^{\boldsymbol{T}}(u) = \boldsymbol{x} \,\}$ and $\{\, u \in T \mid v \lhd_{m-1}^T u \,\}$, namely, $\bigcup_{i \geq 1}((\blacktriangleleft^{\boldsymbol{T}}_{m,i})^{-1} \circ \lhd^T_{m-1,i})$. We write $\mathbb{H}^m_{\Sigma,\boldsymbol{x}}(n)$ for

$$\{\, \boldsymbol{T} \in \mathbb{H}^m_{\Sigma,\boldsymbol{x}} \mid \text{there are exactly } n \text{ nodes } v \in T \cap \mathbb{P}_{m-1} \text{ such that } \ell^{\boldsymbol{T}}(v) = \boldsymbol{x} \,\}.$$

Note that if $\boldsymbol{T} \in \mathbb{H}^m_{\Sigma,\boldsymbol{x}}(n)$, $\boldsymbol{T}$ may have more than n nodes labeled by $\boldsymbol{x}$.

We write $\mathbb{T}^m_{\Sigma,\boldsymbol{x}}$ for $\mathbb{H}^m_{\Sigma,\boldsymbol{x}}(0) \cap \mathbb{T}^m_{\Sigma\cup\boldsymbol{x}}$. We will give suitable definitions of encoding and yield for elements of $\mathbb{T}^m_{\Sigma,\boldsymbol{x}}$ shortly. Before that, here is a variant of Lemma 10 for languages consisting of well-labeled m-dimensional trees. If $\pi \colon \Sigma' \to \Sigma$ is a projection, we extend it to a projection $\pi \colon \Sigma' \cup \{\boldsymbol{x}\} \to \Sigma \cup \{\boldsymbol{x}\}$ by letting $\pi(\boldsymbol{x}) = \boldsymbol{x}$.

Lemma 11. *Let $m \geq 2$. For any local m-dimensional tree language $L \subseteq \mathbb{T}^m_{\Sigma,\boldsymbol{x}}$, there exist a finite alphabet Σ', a degree-bounded, super-local m-dimensional tree language $L' \subseteq \mathbb{T}^m_{\Sigma',\boldsymbol{x}}$, and a projection $\pi \colon \Sigma' \to \Sigma$ such that $L = \pi(L')$. Moreover, π maps L' bijectively L' to L.*

Proof. Since the case where $L \subseteq \mathbb{T}^m_\Sigma$ is covered by Lemma 10, we assume $L \not\subseteq \mathbb{T}^m_\Sigma$. Without loss of generality, we may assume that $L = \mathrm{Loc}^m(A, Z, I)$, for some $A \subseteq \Sigma$, $Z \subseteq \Sigma \cup \{\boldsymbol{x}\}$, $I \subseteq \Sigma \times \mathbb{T}^{m-1}_{\Sigma\cup\{\boldsymbol{x}\}}$ such that

– $\boldsymbol{x} \in Z$,
– $(c, \boldsymbol{T}) \in I$ implies $c \in \Sigma$ and $\ell^{\boldsymbol{T}}(v) \in \Sigma$ for all $v \in \mathrm{dom}(\prec_{m-1}^T)$, and
– there exist $(c, \boldsymbol{T}) \in I$ and $v \in T$ such that $\ell^{\boldsymbol{T}}(v) = \boldsymbol{x}$.

We modify the construction in the proof of Lemma 10 slightly. The difference is that where $(\boldsymbol{T}, v)$ would appear in the earlier construction, $\boldsymbol{x}$ appears instead just in case $\ell^{\boldsymbol{T}}(v) = \boldsymbol{x}$. Otherwise, the proof is essentially the same.

The definition of Σ'' is changed as follows:

$$\Sigma'' = \{\, (\boldsymbol{T}, v) \mid \boldsymbol{T} \in \mathbb{T}^{m-1}_{\Sigma\cup\boldsymbol{x}}, v \in T, \ell^{\boldsymbol{T}}(v) \in \Sigma \,\}.$$

The definition of $\pi \colon \Sigma'' \to \Sigma$ remains the same:

$$\pi((\boldsymbol{T}, v)) = \ell^{\boldsymbol{T}}(v).$$

As before, let

$$F = \{\, \boldsymbol{T}_c \mid c \in A \,\} \cup \{\, \boldsymbol{T} \mid (c, \boldsymbol{T}) \in I \,\},$$
$$\Sigma' = \{\, (\boldsymbol{T}, v) \in \Sigma'' \mid \boldsymbol{T} \in F \,\}.$$

The definitions of Z', K, Y, J are modified as follows:

$$A' = \{\, (\boldsymbol{T}_c, \varepsilon) \mid c \in A \,\},$$
$$Z' = \{\boldsymbol{x}\} \cup \{\, (\boldsymbol{T}, v) \in \Sigma' \mid \ell^{\boldsymbol{T}}(v) \in Z - \{\boldsymbol{x}\} \,\},$$

$$\begin{aligned}
K &= \{\, ((\boldsymbol{T}, v), (\boldsymbol{T}', \varepsilon)) \mid (\boldsymbol{T}, v) \in \Sigma', (\ell^{\boldsymbol{T}}(v), \boldsymbol{T}') \in I, \ell^{\boldsymbol{T}'}(\varepsilon) \in \Sigma \,\} \cup \\
&\quad\ \{\, ((\boldsymbol{T}, v), \boldsymbol{x}) \mid (\boldsymbol{T}, v) \in \Sigma', (\ell^{\boldsymbol{T}}(v), \boldsymbol{T}_{\boldsymbol{x}}) \in I \,\}, \\
Y &= \{\boldsymbol{x}\} \cup \{\, (\boldsymbol{T}, v) \in \Sigma' \mid v \notin \mathrm{dom}(\prec^T_{m-1}) \,\}, \\
J &= \{\, ((\boldsymbol{T}, u), C^T_{m-1}(u), i, (\boldsymbol{T}, v)) \mid (\boldsymbol{T}, u) \in \Sigma', u \lhd^T_{m-1,i} v, \ell^{\boldsymbol{T}}(v) \in \Sigma \,\} \cup \\
&\quad\ \{\, ((\boldsymbol{T}, u), C^T_{m-1}(u), i, \boldsymbol{x}) \mid (\boldsymbol{T}, u) \in \Sigma', u \lhd^T_{m-1,i} v, \ell^{\boldsymbol{T}}(v) = \boldsymbol{x} \,\}.
\end{aligned}$$

These are finite sets. As before, let $L' \subseteq \mathbb{T}^m_{\Sigma'' \cup \{\boldsymbol{x}\}}$ be the super-local set defined by $L' = \mathrm{SLoc}(A', Z', K, Y, J)$. It is easy to see that $L' \subseteq \mathbb{T}^m_{\Sigma' \cup \{\boldsymbol{x}\}}$.

For each $\boldsymbol{T} \in \mathbb{T}^m_{\Sigma \cup \{\boldsymbol{x}\}}$, define an m-dimensional tree $\hat{\boldsymbol{T}} = (T, \ell^{\hat{\boldsymbol{T}}}) \in \mathbb{T}^m_{\Sigma'' \cup \{\boldsymbol{x}\}}$ by

$$\ell^{\hat{\boldsymbol{T}}}(\varepsilon) = (\boldsymbol{T}_{\ell^{\boldsymbol{T}}(\varepsilon)}, \varepsilon), \tag{9}$$

$$\ell^{\hat{\boldsymbol{T}}}(u \cdot m \cdot v) = \begin{cases} (\boldsymbol{C}^{\boldsymbol{T}}_m(u), v) & \text{if } \ell^{\boldsymbol{T}}(u \cdot m \cdot v) \in \Sigma, \\ \boldsymbol{x} & \text{if } \ell^{\boldsymbol{T}}(u \cdot m \cdot v) = \boldsymbol{x}, \end{cases} \qquad \text{for } v \in C^T_{m-1}(u). \tag{10}$$

It is clear that $\pi(\hat{\boldsymbol{T}}) = \boldsymbol{T}$ for all $\boldsymbol{T} \in \mathbb{T}^m_{\Sigma \cup \{\boldsymbol{x}\}}$. Our goal is to show

$$L' = \{\, \hat{\boldsymbol{T}} \mid \boldsymbol{T} \in L \,\}.$$

This clearly implies that π is a bijection from L' to L.

We show that for all $\boldsymbol{T} \in \mathbb{T}^m_{\Sigma \cup \{\boldsymbol{x}\}}$,

$$\boldsymbol{T} \in L \text{ if and only if } \hat{\boldsymbol{T}} \in L'. \tag{11}$$

This follows from five observations. Firstly, note the following:

- Suppose $u \in T - \mathrm{dom}(\prec^T_{m-1})$. If $\ell^{\hat{\boldsymbol{T}}}(u) = (\boldsymbol{U}, v)$, then $v \notin \mathrm{dom}(\prec^U_{m-1})$. This means that $\ell^{\hat{\boldsymbol{T}}}(u) \in Y$ if $\ell^{\hat{\boldsymbol{T}}}(u) \in \Sigma' \cup \{\boldsymbol{x}\}$.
- Suppose $s \lhd^T_m u = s \cdot m \cdot u' \lhd^T_{m-1,i} v = u \cdot (m-1) \cdot v'$. Then we have $u' \lhd^{C^T_m(s)}_{m-1,i} u' \cdot (m-1) \cdot v'$ and

$$\begin{aligned}
\ell^{\hat{\boldsymbol{T}}}(u) &= \begin{cases} (\boldsymbol{C}^{\boldsymbol{T}}_m(s), u') & \text{if } \ell^{\boldsymbol{T}}(u) \in \Sigma, \\ \boldsymbol{x} & \text{if } \ell^{\boldsymbol{T}}(u) = \boldsymbol{x}, \end{cases} \\
\ell^{\hat{\boldsymbol{T}}}(v) &= \begin{cases} (\boldsymbol{C}^{\boldsymbol{T}}_m(s), u' \cdot (m-1) \cdot v') & \text{if } \ell^{\boldsymbol{T}}(v) \in \Sigma, \\ \boldsymbol{x} & \text{if } \ell^{\boldsymbol{T}}(v) = \boldsymbol{x}, \end{cases} \\
C^T_{m-1}(u) &= C^{C^T_m(s)}_{m-1}(u').
\end{aligned}$$

  This means that $(\ell^{\hat{\boldsymbol{T}}}(u), C^T_{m-1}(u), i, \ell^{\hat{\boldsymbol{T}}}(v)) \in J$ if $\ell^{\hat{\boldsymbol{T}}}(u) \in \Sigma'$.

Thus, $\hat{\boldsymbol{T}}$ satisfies the last two conditions SL4 and SL5 for membership in $\mathrm{SLoc}^m(A', Z', K, Y, J)$ whenever $\hat{\boldsymbol{T}} \in \mathbb{T}^m_{\Sigma', \boldsymbol{x}}$. Secondly, the following biconditional always holds:

– $\ell^{\boldsymbol{T}}(\varepsilon) \in A$ if and only if $\ell^{\hat{\boldsymbol{T}}}(\varepsilon) \in A'$.

Thirdly, the following biconditional holds whenever $\ell^{\boldsymbol{T}}(v) \in \Sigma' \cup \{\boldsymbol{x}\}$:

– $\ell^{\boldsymbol{T}}(v) \in Z$ if and only if $\ell^{\hat{\boldsymbol{T}}}(v) \in Z'$.

Fourthly, if $u \prec_m^T v$, $\ell^{\hat{\boldsymbol{T}}}(u) \in \Sigma'$, and either $v \notin \mathrm{dom}(\prec_{m-1}^T)$ or $\ell^{\hat{\boldsymbol{T}}}(v) \in \Sigma''$, then the following biconditional holds:

– $(\ell^{\boldsymbol{T}}(u), \boldsymbol{C}_m^{\boldsymbol{T}}(u)) \in I$ if and only if $(\ell^{\hat{\boldsymbol{T}}}(u), \ell^{\hat{\boldsymbol{T}}}(v)) \in K$.

Lastly, it is easy to see that $\boldsymbol{T} \in L$ implies $\hat{\boldsymbol{T}} \in \mathbb{T}^m_{\Sigma',\boldsymbol{x}}$. Combining these five observations, we get (11).

It follows from the "only if" direction of (11) that $\{\, \hat{\boldsymbol{T}} \mid \boldsymbol{T} \in L \,\} \subseteq L'$. To establish the converse inclusion, we show that

$$\text{if } \boldsymbol{T'} \in L' \text{ and } \boldsymbol{T} = \pi(\boldsymbol{T'}), \text{ then } \boldsymbol{T'} = \hat{\boldsymbol{T}}.$$

This together with the "if" direction of (11) clearly implies $L' \subseteq \{\, \hat{\boldsymbol{T}} \mid \boldsymbol{T} \in L \,\}$.

So suppose $\boldsymbol{T'} = (T, \ell^{\boldsymbol{T'}}) \in L'$, and let $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}}) = \pi(\boldsymbol{T'})$. All we need to show is that the equations (9) and (10) hold with $\boldsymbol{T'}$ in place of $\hat{\boldsymbol{T}}$.

As for (9), it follows from the fact that $\ell^{\boldsymbol{T'}}(\varepsilon) \in A'$.

As for (10), suppose $u \in \mathrm{dom}(\prec_m^T)$. Since $(\ell^{\boldsymbol{T'}}(u), \ell^{\boldsymbol{T'}}(u{\cdot}m)) \in K$, $\ell^{\boldsymbol{T'}}(u{\cdot}m) = (\boldsymbol{V}, \varepsilon)$ for some $\boldsymbol{V} \in F$. We show two things:

$$v \in C_m^T(u) \text{ implies } v \in V \text{ and } \ell^{\boldsymbol{T'}}(u \cdot m \cdot v) = \begin{cases} (\boldsymbol{V}, v) & \text{if } \ell^{\boldsymbol{V}}(v) \in \Sigma, \\ \boldsymbol{x} & \text{if } \ell^{\boldsymbol{V}}(v) = \boldsymbol{x}. \end{cases} \tag{12}$$

$$V \subseteq C_m^T(u). \tag{13}$$

It then easily follows that $\boldsymbol{V} = \boldsymbol{C}_m^{\boldsymbol{T}}(u)$ and (10) holds whenever $v \in C_m^T(u)$.

We show (12) by induction on $v \in C_m^T(u)$. For $v = \varepsilon$, we already know that $\varepsilon \in V$ and $\ell^{\boldsymbol{T'}}(u \cdot m) = (\boldsymbol{V}, \varepsilon)$. If $v \neq \varepsilon$, we can write $v = v' \cdot (m-1) \cdot v''$ with $v'' \in \mathbb{P}_{m-2}$. Suppose $u \cdot m \cdot v' \lhd_{m-1,i}^T u \cdot m \cdot v$. Since $v' \in C_m^T(u)$, by induction hypothesis, $v' \in V$ and $\ell^{\boldsymbol{T'}}(u \cdot m \cdot v')$ is either $(\boldsymbol{V}, v')$ or $\boldsymbol{x}$ depending on whether $\ell^{\boldsymbol{V}}(v') \in \Sigma$ or not. Since $\boldsymbol{T'} \in L'$, $(\ell^{\boldsymbol{T'}}(u{\cdot}m{\cdot}v'), C_{m-1}^T(u{\cdot}m{\cdot}v'), i, \ell^{\boldsymbol{T'}}(u{\cdot}m{\cdot}v)) \in J$. By the definition of J, we must have $\ell^{\boldsymbol{T'}}(u{\cdot}m{\cdot}v') = (\boldsymbol{V}, v')$ and $C_{m-1}^T(u{\cdot}m{\cdot}v') = C_{m-1}^V(v')$, which implies $v' \lhd_{m-1,i}^V v \in V$. The definition of J then implies $\ell^{\boldsymbol{T'}}(u \cdot m \cdot v)$ is either $(\boldsymbol{V}, v)$ or $\boldsymbol{x}$ depending on whether $\ell^{\boldsymbol{V}}(v) \in \Sigma$ or not.

Having established (12), we proceed to show (13) by induction on $v \in V$. For $v = \varepsilon$, we have $\varepsilon \in C_m^T(u)$ since $u \in \mathrm{dom}(\prec_m^T)$. If $v = v' \cdot (m-1) \cdot v''$ with $v'' \in \mathbb{P}_{m-2}$, then $v' \in C_m^T(u)$ by induction hypothesis. Since $\boldsymbol{V} \in F$ and $v' \in \mathrm{dom}(\prec_{m-1}^V)$, by the assumption about I, $\ell^{\boldsymbol{V}}(v') \in \Sigma$ and so $\ell^{\boldsymbol{T'}}(u \cdot m \cdot v') = (\boldsymbol{V}, v') \notin Y$. Since $\boldsymbol{T'} \in L'$, this means that $u \cdot m \cdot v' \in \mathrm{dom}(\prec_{m-1}^T)$ and so $(\ell^{\boldsymbol{T'}}(u \cdot m \cdot v'), C_{m-1}^T(u \cdot m \cdot v'), 1, \ell^{\boldsymbol{T'}}(u \cdot m \cdot v' \cdot (m-1))) \in J$. Since $\ell^{\boldsymbol{T'}}(u \cdot m \cdot v') = (\boldsymbol{V}, v')$, the definition of J implies $C_{m-1}^T(u \cdot m \cdot v') = C_{m-1}^V(v')$. Since $v = v' \cdot (m-1) \cdot v'' \in V$, it follows that $v'' \in C_{m-1}^T(u \cdot m \cdot v')$ and hence $v = v' \cdot (m-1) \cdot v'' \in C_m^T(u)$.

This concludes the proof of the lemma. □

Fix $m \geq 2$ and Σ. For each $c \in \Sigma$ and each (possibly empty) finite prefix-closed subset P of $\mathbb{P}_{m-1}$, define

$$\Gamma_{c,P} = \{\, (c, P, i) \mid 0 \leq i \leq |P| \,\}.$$

We consider $\Gamma_{c,P}$ to be a group of symbols that match with each other; this notion of a matching group of symbols generalizes the notion of a matching pair of brackets. Let

$$\widetilde{\Sigma} = \Sigma \cup \bigcup \{\, \Gamma_{c,P} \mid c \in \Sigma \text{ and } P \subseteq \mathbb{P}_{m-1} \text{ is finite and prefix-closed} \,\}.$$

Note that $\widetilde{\Sigma}$ is an infinite set.[16]

Let $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}$. For $u \in C^T_m(\varepsilon)$, let $T_u = \{\, v \in \mathbb{P}_m \cdot \mathbb{P}_{m+1} \mid m \cdot u \cdot v \in T \,\}$, i.e., the domain of $SH_{m+1}(\boldsymbol{T}, m \cdot u)$. Then we have

$$T = \begin{cases} \{\varepsilon\} \cup \bigcup_{u \in C^T_m(\varepsilon)} m \cdot u \cdot T_u & \text{if } m+1 \notin T, \\ \{\varepsilon\} \cup (m+1) \cdot (T/(m+1)) \cup \bigcup_{u \in C^T_m(\varepsilon)} m \cdot u \cdot T_u & \text{if } m+1 \in T. \end{cases}$$

Thus, $\boldsymbol{T}$ is completely determined by the following pieces of information:

- $\ell^{\boldsymbol{T}}(\varepsilon)$,
- $C^T_m(\varepsilon)$,
- $SH_{m+1}(\boldsymbol{T}, m \cdot u)$ for each $u \in C^T_m(\varepsilon)$,
- whether or not $m+1 \in T$, and
- in case $m+1 \in T$, the $(m+1)$-dimensional hedge $SH_{m+1}(\boldsymbol{T}, m+1) = \boldsymbol{T}/(m+1)$.

Let $P = C^T_m(\varepsilon)$, $k = |P|$, and for $i = 1, \dots, k$, $\varepsilon \lhd^T_{m,i} m \cdot u_i$ and $\boldsymbol{T}_i = SH_{m+1}(\boldsymbol{T}, m \cdot u_i)$. In case $m+1 \in T$ or $k \geq 1$ (i.e., $m \in T$), let $c = \ell^{\boldsymbol{T}}(\varepsilon) \in \Sigma$. The *$m$-dimensional encoding of $\boldsymbol{T}$*, $\mathbf{enc}_m(\boldsymbol{T})$ in symbols, is defined as follows:

$$\mathbf{enc}_m(\boldsymbol{T}) = \begin{cases} \boldsymbol{T}_{\ell^{\boldsymbol{T}}(\varepsilon)} & \text{if } m+1 \notin T \text{ and } k = 0, \\ c -_m P(\mathbf{enc}_m(\boldsymbol{T}_1), \dots, \mathbf{enc}_m(\boldsymbol{T}_k)) & \text{if } m+1 \notin T \text{ and } k \geq 1, \\ (c, P, 0) -_m (\mathbf{enc}_m(\boldsymbol{T}_0))[& \text{if } m+1 \in T \text{ and} \\ \qquad (c, P, 1) -_m \mathbf{enc}_m(\boldsymbol{T}_1), & \boldsymbol{T}_0 = SH_{m+1}(\boldsymbol{T}, m+1). \\ \qquad \dots, & \\ \qquad (c, P, k) -_m \mathbf{enc}_m(\boldsymbol{T}_k) & \\ \quad] & \end{cases}$$

(The substitution notation in the last clause presupposes $\mathbf{enc}_m(\boldsymbol{T}_0) \in \mathbb{T}^m_{\widetilde{\Sigma}}(k)$, and this is indeed the case as shown by the following lemma.)

---

[16] When we define $\widetilde{\Sigma}$ from Σ in this way, we assume that $\Sigma \cap \Gamma_{c,P} = \varnothing$ for all $c \in \Sigma$ and all finite and prefix-closed $P \subseteq \mathbb{P}_{m-1}$. Technically, this assumption may not always be satisfied; nevertheless, we always regard the symbols in $\Gamma_{c,P}$ as "new" symbols. If more rigor is desired, it can be achieved by complicating the definition of $\Gamma_{c,P}$.

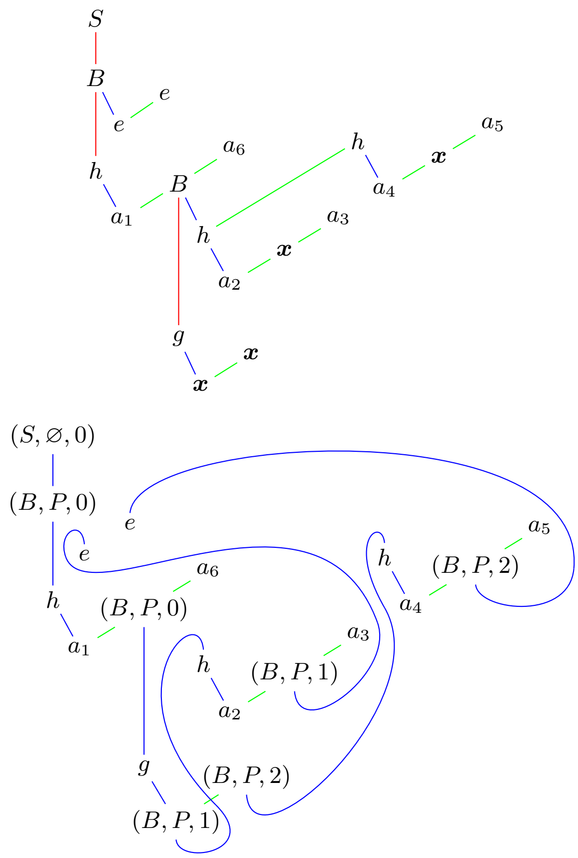

Fig. 2. A well-labeled 3-dimensional tree and its 2-dimensional encoding ($P = \{\varepsilon, 1\}$).

Example 12. Fig. 2 shows the 3-dimensional tree $\boldsymbol{T}$ from Example 9, which is in $\mathbb{T}^3_{N\cup\Sigma,\boldsymbol{x}}$, where $N = \{S, B\}$ and $\Sigma = \{h, g, a_1, a_2, a_3, a_4, a_5, a_6\}$, along with $\mathbf{enc}_2(\boldsymbol{T})$. Here, $P = \{\varepsilon, 1\}$. As before, the edges in the third dimension are colored red, those in the second dimension blue, and those in the first dimension green.

Lemma 13. *If* $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n)$*, then* $\mathbf{enc}_m(\boldsymbol{T}) \in \mathbb{T}^m_{\widetilde{\Sigma}}(n)$*.*

Proof. Induction on the size of $\boldsymbol{T}$. Let c, P, k, and $\boldsymbol{T}_i$ be as above. Suppose $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n)$. If $\ell^{\boldsymbol{T}}(\varepsilon) = \boldsymbol{x}$, then $\boldsymbol{T} = \boldsymbol{T}_{\boldsymbol{x}} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(1)$, and $\mathbf{enc}_m(\boldsymbol{T}) = \boldsymbol{T}_{\boldsymbol{x}} \in \mathbb{T}^m_{\widetilde{\Sigma}}(1)$. If $\ell^{\boldsymbol{T}}(\varepsilon) = c \in \Sigma$, then $\boldsymbol{T}_i \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n_i)$ for $i = 1, \dots, k$, where $n = n_1 + \dots + n_k$. By induction hypothesis, $\mathbf{enc}_m(\boldsymbol{T}_i) \in \mathbb{T}^m_{\widetilde{\Sigma}}(n_i)$. Suppose $m + 1 \notin T$. Then it is easy to see $\mathbf{enc}_m(\boldsymbol{T}) \in \mathbb{T}^m_{\widetilde{\Sigma}}(n)$. Now suppose $m + 1 \in T$. Since $\boldsymbol{T}$ is well-labeled, $\boldsymbol{T}_0 \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(k)$. By induction hypothesis, $\mathbf{enc}_m(\boldsymbol{T}_0) \in \mathbb{T}^m_{\widetilde{\Sigma}}(k)$.

Also, $(c, P, i) -_m \mathbf{enc}_m(\boldsymbol{T}_i)$ is in $\mathbb{T}^m_{\widetilde{\Sigma}}(n_i)$ for $i = 1, \ldots, k$. It easily follows that $\mathbf{enc}_m(\boldsymbol{T}) = (c, P, 0) -_m (\mathbf{enc}_m(\boldsymbol{T}_0))[(c, P, 1) -_m \mathbf{enc}_m(\boldsymbol{T}_1), \ldots, (c, P, k) -_m \mathbf{enc}_m(\boldsymbol{T}_k)] \in \mathbb{T}^m_{\widetilde{\Sigma}}(n)$. □

Note that in $\mathbf{enc}_m(\boldsymbol{T})$, every node with a label of the form (c, P, i) $(i \geq 0)$ has exactly one child in the m-th dimension. There is a simple way of deleting any collection of such nodes from an m-dimensional tree to produce another m-dimensional tree.

Let $\boldsymbol{T}$ be any m-dimensional tree, and assume that $U \subseteq T$ only consists of nodes v such that $|C^T_m(v)| = 1$. Define a function $f_U : T \to [1, m]^*$ by

$$\begin{aligned} f_U(\varepsilon) &= \varepsilon, \\ f_U(v \cdot i) &= f_U(v) \cdot i \qquad \text{for } i < m, \\ f_U(v \cdot m) &= \begin{cases} f_U(v) \cdot m & \text{if } v \notin U, \\ f_U(v) & \text{if } v \in U. \end{cases} \end{aligned}$$

Let $T' = \mathrm{ran}(f_U) = \{\, f_U(v) \mid v \in T \,\}$ and $f'_U = f_U \restriction (T - U)$. Then it is easy to see that $T' = \mathrm{ran}(f'_U)$, T' is a non-empty prefix-closed subset of $\mathbb{P}_m$, and f'_U is a bijection from $T - U$ to T'. Define

$$\mathbf{del}_m(\boldsymbol{T}, U) = (T', \ell'),$$

where

$$\ell'(v) = \ell^{\boldsymbol{T}}((f'_U)^{-1}(v)).$$

Since T' is a non-empty prefix-closed subset of $\mathbb{P}_m$, it follows that $\mathbf{del}_m(\boldsymbol{T}, U)$ is an m-dimensional tree.

Let $\Upsilon \subseteq \Sigma$ and $\boldsymbol{T} \in \mathbb{T}^m_{\Sigma}$. We define

$$\mathbf{del}_{m,\Upsilon}(\boldsymbol{T}) = \mathbf{del}_m(\boldsymbol{T}, U),$$

where

$$U = \{\, v \in T \mid \ell^{\boldsymbol{T}}(v) \in \Upsilon \text{ and } |C^T_m(v)| = 1 \,\}.$$

Now let $\boldsymbol{T} \in \mathbb{T}^{m+1}_{\Sigma,\boldsymbol{x}}$ $(m \geq 2)$. The *m-dimensional yield* of $\boldsymbol{T}$ is defined as follows:

$$\mathbf{y}_m(\boldsymbol{T}) = \mathbf{del}_{m,\widetilde{\Sigma}-\Sigma}(\mathbf{enc}_m(\boldsymbol{T})).$$

It is easy to see that $\mathbf{y}_m(\boldsymbol{T}) \in \mathbb{T}^m_{\Sigma}$.

It is of course straightforward to define $\mathbf{y}_m : \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n) \to \mathbb{T}^m_{\Sigma}(n)$ directly:

$$\mathbf{y}_m(\boldsymbol{T}) = \begin{cases} \boldsymbol{T}_c & \text{if } m+1 \notin T \text{ and } k = 0, \\ c -_m P(\mathbf{y}_m(\boldsymbol{T}_1), \ldots, \mathbf{y}_m(\boldsymbol{T}_k)) & \text{if } m+1 \notin T \text{ and } k \geq 1, \\ (\mathbf{y}_m(\boldsymbol{T}_0))[\mathbf{y}_m(\boldsymbol{T}_1), \ldots, \mathbf{y}_m(\boldsymbol{T}_k)] & \text{if } m+1 \in T \text{ and} \\ & \boldsymbol{T}_0 = SH_{m+1}(\boldsymbol{T}, m+1), \end{cases}$$

where, as before, $c = \ell^{\boldsymbol{T}}(\varepsilon)$, $P = C^T_m(\varepsilon)$, $k = |P|$, and for $i = 1, \ldots, k$, $\varepsilon \lhd^T_{m,i} m \cdot u_i$ and $\boldsymbol{T}_i = SH_{m+1}(\boldsymbol{T}, m \cdot u_i)$. The indirect definition through $\mathbf{enc}_m$, however, is

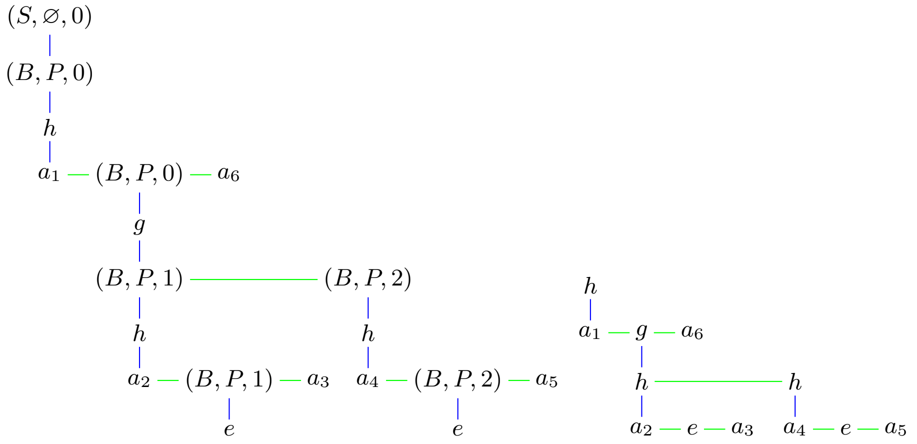

Fig. 3. The 2-dimensional encoding and 2-dimensional yield of a well-labeled 3-dimensional tree ($P = \{\varepsilon, 1\}$).

useful for our generalization of the Chomsky-Schützenberger theorem for multi-dimensional tree languages.

The case $m = 2$ of the above definition of $\mathbf{y}_m$ is meant to capture the notion of the (tree) yield of a derivation tree of a simple context-free tree grammar, which we represent as a (well-labeled) 3-dimensional tree. The definitions of $\mathbf{enc}_m, \mathbf{del}_{m,\Upsilon}, \mathbf{y}_m$ are all applicable to the case $m = 1$ as well, but the resulting definitions of $\mathbf{enc}_1$ and of $\mathbf{y}_1$ will not be equivalent to the standard ones, so we will continue to treat $m = 1$ as a special case.

Example 14. Fig. 3 shows $\mathbf{enc}_2(\boldsymbol{T})$ (the same tree as the lower tree in Fig. 2 with the nodes rearranged) and $\mathbf{y}_2(\boldsymbol{T})$, where $\boldsymbol{T}$ is the 3-dimensional tree from Example 9 (the upper tree in Fig. 2).

7 Multi-dimensional Dyck Languages

We continue to work with the alphabet

$$\widetilde{\Sigma} = \Sigma \cup \bigcup \{ \Gamma_{c,P} \mid c \in \Sigma \text{ and } P \text{ is a finite prefix-closed subset of } \mathbb{P}_{m-1} \},$$

as defined in the previous section. The range of the function $\mathbf{enc}_m \colon \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(0) \to \mathbb{T}^m_{\widetilde{\Sigma}}$ forms a special subset of $\mathbb{T}^m_{\widetilde{\Sigma}}$ similar to Dyck languages.

Let us define a rewriting relation $\rightsquigarrow$ on $\mathbb{T}^m_{\widetilde{\Sigma}}$:

$$\boldsymbol{T} \rightsquigarrow \boldsymbol{T}'$$

holds if there exist some $v_0, v_1, \dots, v_n \in T$ ($n \geq 0$), $c \in \Sigma$, and finite prefix-closed subset P of $\mathbb{P}_m$ such that[17]

[17] Note that for $u, v \in T$, $u\ (\lhd^T_m)^*\ v$ is equivalent to $v \in u \cdot \mathbb{P}_m$, and $u\ (\lhd^T_m)^+\ v$ is equivalent to $v \in u \cdot m \cdot \mathbb{P}_{m-1} \cdot \mathbb{P}_m$.

- $\boldsymbol{T}' = \mathbf{del}_m(\boldsymbol{T}, \{v_0, v_1, \dots, v_n\})$,
- $|C_m^T(v_i)| = 1$ for $i = 0, 1, \dots, n$,
- $\ell^{\boldsymbol{T}}(v_i) = (c, P, i)$ for $i = 0, 1, \dots, n$,
- $n = |P|$,
- $v_1, \dots, v_n$ is the alphabetical listing of $\{v_1, \dots, v_n\}$,
- $v_0\ (\lhd_m^T)^+\ v_i$ for $i = 1, \dots, n$,
- for every $i, j \in \{1, \dots, n\}$, if $v_i\ (\lhd_m^T)^*\ v_j$, then $i = j$, and
- for every $u \in T$, if $v_0\ (\lhd_m^T)^+\ u$ and there is no $i \in \{1, \dots, n\}$ such that $v_i\ (\lhd_m^T)^*\ u$, then $\ell^{\boldsymbol{T}}(u) \in \Sigma$.

Using the term notation, we can write

$$\begin{aligned}
&\boldsymbol{T} \rightsquigarrow \boldsymbol{T}' \text{ if and only if} \\
&\qquad \boldsymbol{T} = \boldsymbol{U}[(c, P, 0) -_m \boldsymbol{T}_0[(c, P, 1) -_m \boldsymbol{T}_1, \dots, (c, P, n) -_m \boldsymbol{T}_n]], \\
&\qquad \boldsymbol{T}' = \boldsymbol{U}[\boldsymbol{T}_0[\boldsymbol{T}_1, \dots, \boldsymbol{T}_n]]) \\
&\qquad\quad \text{for some } \boldsymbol{U} \in \mathbb{T}^m_{\widetilde{\Sigma}}(1), \boldsymbol{T}_0 \in \mathbb{T}^m_{\Sigma}(n), \boldsymbol{T}_i \in \mathbb{T}^m_{\widetilde{\Sigma}}\ (i = 1, \dots, n), \\
&\qquad\qquad\qquad c \in \Sigma, \text{ finite and prefix-closed } P \subseteq \mathbb{P}_{m-1} \text{ with } |P| = n.
\end{aligned}$$

Define the *m-dimensional Dyck tree language* over Σ by[18]

$$DT^m_\Sigma = \{\, \boldsymbol{T} \in \mathbb{T}^m_{\widetilde{\Sigma}} \mid \boldsymbol{T} \rightsquigarrow^* \boldsymbol{T}' \in \mathbb{T}^m_\Sigma \,\}.$$

Note that the alphabet of DT^m_Σ (i.e., the set of labels that appear in elements of DT^m_Σ) is infinite.

Just like the ordinary Dyck language D_n of strings over Γ_n has an alternative inductive definition in terms of a context-free grammar, so too the m-dimensional tree language DT^m_Σ admits an inductive definition. First, let us extend the definition of $\rightsquigarrow$ to a rewriting relation on $\mathbb{T}^m_{\widetilde{\Sigma}}(n)$ by taking the exact same definition as before, requiring $\ell^{\boldsymbol{T}}(u) \in \Sigma$ rather than $\ell^{\boldsymbol{T}}(u) \in \Sigma \cup \{\boldsymbol{x}\}$ in the consequent of the last condition. Note that this relation is confluent:

Lemma 15. *Let $\boldsymbol{T}, \boldsymbol{T}_1, \boldsymbol{T}_2 \in \mathbb{T}^m_{\widetilde{\Sigma}}(n)$. If $\boldsymbol{T} \rightsquigarrow \boldsymbol{T}_1$ and $\boldsymbol{T} \rightsquigarrow \boldsymbol{T}_2$, then there exists some $\boldsymbol{T}' \in \mathbb{T}^m_{\widetilde{\Sigma}}(n)$ such that $\boldsymbol{T}_1 \rightsquigarrow \boldsymbol{T}'$ and $\boldsymbol{T}_2 \rightsquigarrow \boldsymbol{T}'$.*

For each $n \in \mathbb{N}$, we define $DT^m_\Sigma(n)$ by

$$DT^m_\Sigma(n) = \{\, \boldsymbol{T} \in \mathbb{T}^m_{\widetilde{\Sigma}}(n) \mid \boldsymbol{T} \rightsquigarrow^* \boldsymbol{T}' \in \mathbb{T}^m_\Sigma(n) \,\}.$$

Clearly, $DT^m_\Sigma(0) = DT^m_\Sigma$.

Then we can prove that $(X_n)_{n\in\mathbb{N}} = (DT^m_\Sigma(n))_{n\in\mathbb{N}}$ is the least (in terms of the partial order defined by componentwise inclusion) sequence of sets that satisfies the following closure conditions:

[18] For dimension $m = 2$, analogous notions of Dyck tree language have been proposed by Matsubara and Kasai [19] and by Arnold and Dauchet [1] to capture the tree languages generated by tree-adjoining grammars and by (general) context-free tree grammars, respectively.

I1. $\boldsymbol{T}_c \in X_0$ for all $c \in \Sigma$.
I2. $\boldsymbol{T}_{\boldsymbol{x}} \in X_1$.
I3. If $c \in \Sigma$, P is a finite, non-empty, prefix-closed subset of $\mathbb{P}_{m-1}$, $k = |P|$, $\boldsymbol{T}_1 \in X_{n_1}, \ldots, \boldsymbol{T}_k \in X_{n_k}$, and $n = \sum_{i=1}^{k} n_i$, then

$$c -_m P(\boldsymbol{T}_1, \ldots, \boldsymbol{T}_k) \in X_n.$$

I4. If $c \in \Sigma$, P is a (possibly empty) finite prefix-closed subset of $\mathbb{P}_{m-1}$, $k = |P|$, $\boldsymbol{T}_1 \in X_{n_1}, \ldots, \boldsymbol{T}_k \in X_{n_k}$, $\boldsymbol{T}_0 \in X_k$, and $n = \sum_{i=1}^{k} n_i$, then

$$(c, P, 0) -_m \boldsymbol{T}_0[(c, P, 1) -_m \boldsymbol{T}_1, \ldots, (c, P, k) -_m \boldsymbol{T}_k] \in X_n.$$

Lemma 16. $(X_n)_{n\in\mathbb{N}} = (DT^m_\Sigma(n))_{n\in\mathbb{N}}$ *satisfies I1–I4.*

Lemma 17. *If $\boldsymbol{T} \in DT^m_\Sigma(n)$, then one of the following conditions holds:*

C1. $n = 0$ *and* $\boldsymbol{T} = \boldsymbol{T}_c$ *for some* $c \in \Sigma$.
C2. $n = 1$ *and* $\boldsymbol{T} = \boldsymbol{T}_{\boldsymbol{x}}$.
C3. $n = \sum_{i=1}^{k} n_i$ *for some* $k \geq 1, n_1, \ldots, n_k \geq 0$ *and*

$$\boldsymbol{T} = c -_m P(\boldsymbol{T}_1, \ldots, \boldsymbol{T}_k)$$

for some $c \in \Sigma$, some finite prefix-closed subset P of $\mathbb{P}_{m-1}$ such that $|P| = k$, and some $\boldsymbol{T}_1 \in DT^m_\Sigma(n_1), \ldots, \boldsymbol{T}_k \in DT^m_\Sigma(n_k)$.

C4. $n = \sum_{i=1}^{k} n_i$ *for some* $k \geq 0, n_1, \ldots, n_k \geq 0$ *and*

$$\boldsymbol{T} = (c, P, 0) -_m \boldsymbol{T}_0[(c, P, 1) -_m \boldsymbol{T}_1, \ldots, (c, P, k) -_m \boldsymbol{T}_k]$$

for some $c \in \Sigma$, some finite prefix-closed subset P of $\mathbb{P}_{m-1}$ such that $|P| = k$, and some $\boldsymbol{T}_1 \in DT^m_\Sigma(n_1), \ldots, \boldsymbol{T}_k \in DT^m_\Sigma(n_k), \boldsymbol{T}_0 \in DT^m_\Sigma(k)$.

Proof. Suppose $\boldsymbol{T} \in DT^m_\Sigma(n)$. If $|T| = 1$, then clearly, either C1 or C2 holds.

If $\ell^{\boldsymbol{T}}(\varepsilon) \in \Sigma$ and $C^T_m(\varepsilon) = P \neq \varnothing$, let $k = |P|$. Then $\boldsymbol{T} = c -_m P(\boldsymbol{T}_1, \ldots, \boldsymbol{T}_k)$ for some $\boldsymbol{T}_1 \in \mathbb{T}^m_{\widetilde{\Sigma}}(n_1), \ldots, \boldsymbol{T}_k \in \mathbb{T}^m_{\widetilde{\Sigma}}(n_k)$ such that $\sum_{i=1}^{k} n_i = n$. Since $\boldsymbol{T} \rightsquigarrow^* \boldsymbol{T}'$ for some $\boldsymbol{T}' \in \mathbb{T}^m_\Sigma(n)$, it is clear that for $i = 1, \ldots, k$, $\boldsymbol{T}_i \rightsquigarrow^* \boldsymbol{T}'_i$ for some $\boldsymbol{T}'_i \in \mathbb{T}^m_\Sigma(n_i)$. Therefore, C3 holds.

Now suppose $\ell^{\boldsymbol{T}}(\varepsilon) = (c, P, i)$. Since $\boldsymbol{T} \rightsquigarrow^* \boldsymbol{T}' \in \mathbb{T}^m_\Sigma(n)$, it is easy to see that $i = 0$ and $C^T_m(\varepsilon) = \{\varepsilon\}$. Let $k = |P|$ and let $\boldsymbol{T}'_0 \in \mathbb{T}^m_\Sigma(k)$, $\boldsymbol{T}'_1, \ldots, \boldsymbol{T}'_k$ be such that

$$\begin{aligned}\boldsymbol{T} &\rightsquigarrow^* (c, P, 0) -_m \boldsymbol{T}'_0[(c, P, 1) -_m \boldsymbol{T}'_1, \ldots, (c, P, k) -_m \boldsymbol{T}'_k] \\ &\rightsquigarrow \boldsymbol{T}'_0[\boldsymbol{T}'_1, \ldots, \boldsymbol{T}'_k] \\ &\rightsquigarrow^* \boldsymbol{T}'.\end{aligned}$$

Then it is easy to see that for $i = 1, \ldots, k$, $\boldsymbol{T}'_i \rightsquigarrow^* \boldsymbol{T}''_i \in \mathbb{T}^m_\Sigma(n_i)$ for some n_i such that $n = \sum_{i=1}^{k} n_i$. Also, we must have

$$\boldsymbol{T} = (c, P, 0) -_m \boldsymbol{T}_0[(c, P, 1) -_m \boldsymbol{T}_1, \ldots, (c, P, k) -_m \boldsymbol{T}_k]$$

for some $\boldsymbol{T}_0 \in \mathbb{T}^m_{\widetilde{\Sigma}}(k), \boldsymbol{T}_1 \in \mathbb{T}^m_{\widetilde{\Sigma}}(n_1), \ldots, \boldsymbol{T}_k \in \mathbb{T}^m_{\widetilde{\Sigma}}(n_k)$ such that $\boldsymbol{T}_0 \rightsquigarrow^* \boldsymbol{T}'_0$ and for $i = 1, \ldots, k$, $\boldsymbol{T}_i \rightsquigarrow^* \boldsymbol{T}'_i$. It follows that $\boldsymbol{T}_0 \in DT^m_\Sigma(k)$ and for $i = 1, \ldots, k$, $\boldsymbol{T}_i \in DT^m_\Sigma(n_i)$, i.e., C4 holds. □

Theorem 18. $(DT^m_\Sigma(n))_{n\in\mathbb{N}}$ *is the least sequence of sets that satisfies I1–I4.*

Proof. By Lemma 16, we know that $(DT^m_\Sigma(n))_{n\in\mathbb{N}}$ satisfies I1–I4. Let $(X_n)_{n\in\mathbb{N}}$ be any sequence of sets satisfying I1–I4. To establish that $DT^m_\Sigma(n) \subseteq X_n$ holds for all $n \in \mathbb{N}$, we prove by induction on the number of nodes of $\boldsymbol{T} \in \bigcup_n \mathbb{T}^m_\Sigma(n)$ that $\boldsymbol{T} \in DT^m_\Sigma(n)$ implies $\boldsymbol{T} \in X_n$. If $\boldsymbol{T} \in DT^m_\Sigma(n)$, by Lemma 17, one of C1–C4 holds. In case C1 or C2 holds, $\boldsymbol{T} \in X_n$ by I1 or I2. If C3 holds, then by induction hypothesis, $\boldsymbol{T} = c -_m P(\boldsymbol{T}_1, \dots, \boldsymbol{T}_k)$, where $\boldsymbol{T}_i \in X_{n_i}$ for $i = 1, \dots, k$. Then by I3, $\boldsymbol{T} \in X_n$. If C4 holds, then by induction hypothesis, $\boldsymbol{T} = (c, P, 0) -_m \boldsymbol{T}_0[(c, P, 1) -_m \boldsymbol{T}_1, \dots, (c, P, k) -_m \boldsymbol{T}_k]$, where $n = \sum_{i=1}^k n_i$, $\boldsymbol{T}_0 \in X_k$, and $\boldsymbol{T}_i \in X_{n_i}$ for $i = 1, \dots, k$. Then by I4, $\boldsymbol{T} \in X_n$. □

Just as in the case of ordinary Dyck languages, the inductive definition of $DT^m_\Sigma(n)$ in terms of the closure conditions I1–I4 is *unabmiguous* in the sense that every $\boldsymbol{T} \in DT^m_\Sigma(n)$ can be written in the form of one of the equations in C1–C4, in exactly one way. This follows from the next lemma:

Lemma 19. *Let* $\boldsymbol{U} \in DT^m_\Sigma(k), \boldsymbol{U}' \in DT^m_\Sigma(l)$. *If* $\boldsymbol{U}[(c, P, i_1) -_m \boldsymbol{T}_1, \dots, (c, P, i_k) -_m \boldsymbol{T}_k] = \boldsymbol{U}'[(c, P, j_1) -_m \boldsymbol{T}'_1, \dots, (c, P, j_l) -_m \boldsymbol{T}'_l]$ *with* $i_1, \dots, i_k, j_1, \dots, j_l \geq 1$, *then* $\boldsymbol{U} = \boldsymbol{U}'$.

Proof. This can be proved by straightforward induction on the size of $\boldsymbol{U}$, using Lemma 17. □

Lemma 20. $\{\, \mathbf{enc}_m(\boldsymbol{T}) \mid \boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n) \,\} = DT^m_\Sigma(n)$.

Proof. ($\subseteq$). We show by induction on the size of $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}$ that $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n)$ implies $\mathbf{enc}_m(\boldsymbol{T}) \in DT^m_\Sigma(n)$. Let $P = C^T_m(\varepsilon)$, and $k = |P|$. Assume that for $i = 1, \dots, k$, $\varepsilon \lhd^T_{m,i} m \cdot u_i$ and $\boldsymbol{T}_i = SH_{m+1}(\boldsymbol{T}, m \cdot u_i)$. Clearly, for each $i = 1, \dots, k$, $\boldsymbol{T}_i \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n_i)$ for some n_i such that $n = \sum_{i=1}^k n_i$. By induction hypothesis, $\mathbf{enc}_m(\boldsymbol{T}_i) \in DT^m_\Sigma(n_i)$.

Case 1. $m + 1 \in T$. Then $\ell^{\boldsymbol{T}}(\varepsilon) = c$ for some $c \in \Sigma$. Let $\boldsymbol{T}_0 = SH_{m+1}(\boldsymbol{T}, m + 1)$. Then $\boldsymbol{T}_0 \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(k)$. By induction hypothesis, we also have $\mathbf{enc}_m(\boldsymbol{T}_0) \in DT^m_\Sigma(k)$. Then $\mathbf{enc}_m(\boldsymbol{T}) = (c, P, 0) -_m (\mathbf{enc}_m(\boldsymbol{T}_0))[(c, P, 1) -_m \mathbf{enc}_m(\boldsymbol{T}_1), \dots, (c, P, k) -_m \mathbf{enc}_m(\boldsymbol{T}_k)] \in DT^m_\Sigma(n)$ by the closure condition I4.

Case 2. $m+1 \notin T$. If $P \neq \varnothing$, then $\ell^{\boldsymbol{T}}(\varepsilon) = c$ for some $c \in \Sigma$, and $\mathbf{enc}_m(\boldsymbol{T}) = c -_m P(\mathbf{enc}_m(\boldsymbol{T}_1), \dots, \mathbf{enc}_m(\boldsymbol{T}_k)) \in DT^m_\Sigma(n)$ by the closure condition I3. If $P = \varnothing$, then $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(0)$ or $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(1)$ depending on whether $\ell^{\boldsymbol{T}}(\varepsilon) = c \in \Sigma$ or $\ell^{\boldsymbol{T}}(\varepsilon) = \boldsymbol{x}$. In the former case, $\mathbf{enc}_m(\boldsymbol{T}) = \boldsymbol{T}_c \in DT^m_\Sigma(0)$ by the closure condition I1. In the latter case, $\mathbf{enc}_m(\boldsymbol{T}) = \boldsymbol{T}_{\boldsymbol{x}} \in DT^m_\Sigma(1)$ by the closure condition I2.

($\supseteq$). By Theorem 18, it suffices to prove that $(X_n)_{n\in\mathbb{N}} = (\{\, \mathbf{enc}_m(\boldsymbol{T}) \mid \boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n) \,\})_{n\in\mathbb{N}}$ satisfies the conditions I1–I4.

I1 is satisfied since $\boldsymbol{T}_c = \mathbf{enc}_m(\boldsymbol{T}_c)$ and $\boldsymbol{T}_c \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(0)$.

I2 is satisfied since $\boldsymbol{T}_{\boldsymbol{x}} = \mathbf{enc}_m(\boldsymbol{T}_{\boldsymbol{x}})$ and $\boldsymbol{T}_{\boldsymbol{x}} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(1)$.

To check that I3 is satisfied, let $c \in \Sigma$, P be a finite, non-empty, prefix-closed subset of $\mathbb{P}_{m-1}$, $k = |P|$, and for each $i = 1, \dots, k$, $\boldsymbol{T}_i \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n_i)$, where

$n = \sum_{i=1}^{k} n_i$. Let $u_1, \ldots, u_k$ be the elements of P, in alphabetical order. Define $\boldsymbol{T}$ by

$$\begin{aligned} T &= \{\varepsilon\} \cup \bigcup_{i=1}^{k} m \cdot u_i \cdot T_i, \\ \ell^{\boldsymbol{T}}(\varepsilon) &= c, \\ \ell^{\boldsymbol{T}}(m \cdot u_i \cdot v) &= \ell^{\boldsymbol{T}_i}(v) \quad \text{for } v \in T_i. \end{aligned}$$

Then $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n)$ and $\boldsymbol{T}_i = SH_{m+1}(\boldsymbol{T}, m \cdot u_i)$ for $i = 1, \ldots, k$. By the definition of $\mathbf{enc}_m$,

$$\mathbf{enc}_m(\boldsymbol{T}) = c -_m P(\mathbf{enc}_m(\boldsymbol{T}_1), \ldots, \mathbf{enc}_m(\boldsymbol{T}_k)).$$

This shows that I3 is satisfied.

To check that I4 is satisfied, let $c \in \Sigma$, P be a finite, possibly empty, prefix-closed subset of $\mathbb{P}_{m-1}$, $k = |P|$, $\boldsymbol{T}_0 \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(k)$, and for each $i = 1, \ldots, k$, $\boldsymbol{T}_i \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n_i)$, where $n = \sum_{i=1}^{k} n_i$. Let $u_1, \ldots, u_k$ list the elements of P, in alphabetical order. Define $\boldsymbol{T}$ by

$$\begin{aligned} T &= \{\varepsilon\} \cup (m+1) \cdot T_0 \cup \bigcup_{i=1}^{k} m \cdot u_i \cdot T_i, \\ \ell^{\boldsymbol{T}}(\varepsilon) &= c, \\ \ell^{\boldsymbol{T}}((m+1) \cdot v) &= \ell^{\boldsymbol{T}_0}(v) \quad \text{for } v \in T_0, \\ \ell^{\boldsymbol{T}}(m \cdot u_i \cdot v) &= \ell^{\boldsymbol{T}_i}(v) \quad \text{for } v \in T_i. \end{aligned}$$

Then it is easy to see $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(n)$, $\boldsymbol{T}_0 = SH_{m+1}(\boldsymbol{T}, m+1)$, and for $i = 1, \ldots, k$, $\boldsymbol{T}_i = SH_{m+1}(\boldsymbol{T}, m \cdot u_i)$. By the definition of $\mathbf{enc}_m$,

$$\mathbf{enc}_m(\boldsymbol{T}) =$$
$$(c, P, 0) -_m (\mathbf{enc}_m(\boldsymbol{T}_0))[(c, P, 1) -_m (\mathbf{enc}_m(\boldsymbol{T}_1)), \ldots, (c, P, k) -_m (\mathbf{enc}_m(\boldsymbol{T}_k))].$$

This shows that I4 is satisfied. □

Lemma 21. *For each $m \geq 2$, $\mathbf{enc}_m$ is an injection.*

Proof. This follows from the unambiguity of the inductive definition of $DT^m_\Sigma(n)$. □

It is useful to define a function $f^{\boldsymbol{T}}_{\mathbf{enc}_m}$ from the nodes of $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}$ to the nodes of $\boldsymbol{T}' = \mathbf{enc}_m(\boldsymbol{T})$. Let $P = C^T_m(\varepsilon)$ and $k = |P|$. Let $u_1, \ldots, u_k$ list the elements of P in alphabetical order, and let $\boldsymbol{T}_i = SH_{m+1}(\boldsymbol{T}, m \cdot u_i)$ for $i = 1, \ldots, k$. Define $f^{\boldsymbol{T}}_{\mathbf{enc}_m} : T \to T'$ by

(i) $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(\varepsilon) = \varepsilon$.
(ii) If $m+1 \in T$ and $\boldsymbol{T}_0 = SH_{m+1}(\boldsymbol{T}, m+1)$, then

$$\begin{aligned} f^{\boldsymbol{T}}_{\mathbf{enc}_m}((m+1) \cdot w) &= m \cdot f^{\boldsymbol{T}_0}_{\mathbf{enc}_m}(w) && \text{where } w \in T_0, \\ f^{\boldsymbol{T}}_{\mathbf{enc}_m}(m \cdot u_i \cdot w) &= m \cdot f^{\boldsymbol{T}_0}_{\mathbf{enc}_m}(v_i) \cdot m \cdot f^{\boldsymbol{T}_i}_{\mathbf{enc}_m}(w) && \text{where } w \in T_i \text{ and } \varepsilon \blacktriangleleft^{\boldsymbol{T}}_{m+1,i} v_i. \end{aligned}$$

(iii) If $m+1 \notin T$, then

$$f^{\boldsymbol{T}}_{\mathbf{enc}_m}(m \cdot u_i \cdot w) = m \cdot u_i \cdot f^{\boldsymbol{T}_i}_{\mathbf{enc}_m}(w) \quad \text{where } w \in T_i.$$

It is easy to check that $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \in T'$ indeed holds for all $v \in T$.

Example 22. Consider the 3-dimensional tree $\boldsymbol{T}$ and its 2-dimensional encoding $\mathbf{enc}_2(\boldsymbol{T})$, depicted in Fig. 2. In these diagrams, the nodes that are related by $f^{\boldsymbol{T}}_{\mathbf{enc}_m}$ are placed in roughly the same geometrical positions.

Lemma 23. *Let $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}$ and $\boldsymbol{T}' = \mathbf{enc}_m(\boldsymbol{T})$. For each $v \in T$, we have*

$$\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = \begin{cases} c & \text{if } v \notin \mathrm{dom}(\prec^T_{m+1}) \text{ and } \ell^{\boldsymbol{T}}(v) = c \in \Sigma, \\ (c, U, 0) & \text{if } v \in \mathrm{dom}(\prec^T_{m+1}),\ \ell^{\boldsymbol{T}}(v) = c, \text{ and } C^T_m(v) = U, \\ (c, U, i) & \text{if } \ell^{\boldsymbol{T}}(v) = \boldsymbol{x},\ u \blacktriangleleft^{\boldsymbol{T}}_{m+1,i} v,\ \ell^{\boldsymbol{T}}(u) = c, \text{ and } C^T_m(u) = U, \\ \boldsymbol{x} & \text{if } \ell^{\boldsymbol{T}}(v) = \boldsymbol{x} \text{ and } v \in T \cap \mathbb{P}_m. \end{cases}$$

Proof. This is easy to see from the definition of $\mathbf{enc}_m$. □

The function $f^{\boldsymbol{T}}_{\mathbf{enc}_m}$ allows an alternative definition by recursion with respect to the alphabetical order on the nodes of $\boldsymbol{T}$.

Lemma 24. *The function $f^{\boldsymbol{T}}_{\mathbf{enc}_m}$ satisfies the following equations:*

$$f^{\boldsymbol{T}}_{\mathbf{enc}_m}(\varepsilon) = \varepsilon,$$
$$f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u \cdot (m+1)) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u) \cdot m,$$

and for $v \in \mathbb{P}_{m-1}$,

$$f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u \cdot m \cdot v) = \begin{cases} f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u) \cdot m \cdot v & \text{if } u \notin \mathrm{dom}(\prec^T_{m+1}), \\ f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u \cdot (m+1) \cdot w) \cdot m & \text{if } u \in \mathrm{dom}(\prec^T_{m+1}), \\ & u \lhd^T_{m,i} v, \text{ and } u \blacktriangleleft^{\boldsymbol{T}}_{m+1,i} w. \end{cases}$$

Proof. The first equation is true by definition. The remaining two equations can be proved by induction on the length of u. □

Lemma 25. *For all $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}$, $f^{\boldsymbol{T}}_{\mathbf{enc}_m}$ is a bijection from the nodes of $\boldsymbol{T}$ to the nodes of $\mathbf{enc}_m(\boldsymbol{T})$.*

Proof. That $f^{\boldsymbol{T}}_{\mathbf{enc}_m}$ is injective can be shown by induction with respect to the alphabetical order on T. Suppose $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(t_1) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(t_2)$. If $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(t_1) = \varepsilon$, then clearly $t_1 = t_2 = \varepsilon$. If $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(t_1) \in \mathbb{P}_m \cdot m \cdot v$ for $v \in \mathbb{P}_{m-1} - \{\varepsilon\}$, then we must have $t_1 = u_1 \cdot m \cdot v$ and $t_2 = u_2 \cdot m \cdot v$ with $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_1) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_2)$. By the induction hypothesis, $u_1 = u_2$ and hence $t_1 = t_2$. If $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(t_1) \in \mathbb{P}_m \cdot m$, then for each $j \in \{1, 2\}$, one of the following holds:

(i) $t_j = u_j \cdot (m+1)$ and $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(t_j) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_j) \cdot m$.

(ii) $t_j = u_j \cdot m$ and $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(t_j) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_j) \cdot m$, where $u_j \notin \mathrm{dom}(\prec^T_{m+1})$.
(iii) $t_j = u_j \cdot m \cdot v_j$ and $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(t_j) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_j \cdot (m+1) \cdot w_j) \cdot m$, where $u_j \in \mathrm{dom}(\prec^T_{m+1})$, $u_j \lhd^T_{m,i} v_j$, and $u_j \blacktriangleleft^{\boldsymbol{T}}_{m+1,i} w_j$.

If the same case applies to both t_1 and t_2, then the induction hypothesis implies $t_1 = t_2$. If (i) applies to t_1 and (ii) applies to t_2, then $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_1) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_2)$, but since $u_1 \in \mathrm{dom}(\prec^T_{m+1})$ and $u_2 \notin \mathrm{dom}(\prec^T_{m+1})$, the induction hypothesis says this is impossible. If (i) applies to t_1 and (iii) applies to t_2, then $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_1) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_2 \cdot (m+1) \cdot w_2)$. Since $u_1 \in \mathrm{dom}(\prec^T_{m+1})$ and $u_2 \cdot (m+1) \cdot w_2 \notin \mathrm{dom}(\prec^T_{m+1})$ (by the fact that $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}$), the induction hypothesis says this is impossible. If (ii) applies to t_1 and (iii) applies to t_2, then $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_1) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u_2 \cdot (m+1) \cdot w_2)$. Since $u_1 \in \mathrm{dom}(\prec^T_m)$ and $u_2 \cdot (m+1) \cdot w_2 \notin \mathrm{dom}(\prec^T_m)$ (again by the fact that $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}$), the induction hypothesis says this is impossible. The remaining cases are symmetric.

We have shown that $f^{\boldsymbol{T}}_{\mathbf{enc}_m}$ is injective. Since $\boldsymbol{T}$ and $\boldsymbol{T}'$ have the same number of nodes, $f^{\boldsymbol{T}}_{\mathbf{enc}_m}$ is indeed a bijection. ◻

The m-dimensional counterpart DT'^m_Σ of the set of Dyck primes ($m \geq 2$) is defined by

$$DT'^m_\Sigma = \{\, (c, \varnothing, 0) -_m \boldsymbol{T} \mid c \in \Sigma, \boldsymbol{T} \in DT^m_\Sigma \,\} \cup \{\, \boldsymbol{T}_c \mid c \in \Sigma \,\}.$$

Lemma 26. *For all $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}(0)$, $\boldsymbol{T} \in \mathbb{T}^{m+1}_{\Sigma,\boldsymbol{x}}$ if and only if $\mathbf{enc}_m(\boldsymbol{T}) \in DT'^m_\Sigma$.*

Define a function $\rho \colon \widetilde{\Sigma} \to \Sigma \cup \{\boldsymbol{x}\}$ by

$$\begin{aligned} \rho(c) &= c \quad \text{for } c \in \Sigma, \\ \rho((c, U, 0)) &= c, \\ \rho((c, U, i)) &= \boldsymbol{x} \quad \text{for } 1 \leq i \leq |U|. \end{aligned}$$

Then for every $\boldsymbol{T} \in \mathbb{H}^{m+1}_{\Sigma,\boldsymbol{x}}$ and $v \in T$, if $\boldsymbol{T}' = \mathbf{enc}_m(\boldsymbol{T})$, we have

$$\rho(\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v))) = \ell^{\boldsymbol{T}}(v).$$

The following is a generalization of Lemma 2 to higher dimensions:

Lemma 27. *Let $L \subseteq \mathbb{T}^{m+1}_{\Sigma,\boldsymbol{x}}$. If L is super-local, then there exists a local set $L' \subseteq \mathbb{T}^m_{\widetilde{\Sigma}}$ such that $\mathbf{enc}_m(L) = L' \cap DT'^m_\Sigma$.*

Proof. Let L be a super-local subset of $\mathbb{T}^{m+1}_{\Sigma,\boldsymbol{x}}$. Without loss of generality, we may suppose that $L = \mathrm{SLoc}^{m+1}(A, Z, K, Y, J)$ for some finite sets

$A \subseteq \Sigma$,
$Z \subseteq \Sigma \cup \{\boldsymbol{x}\}$,
$K \subseteq \Sigma \times (\Sigma \cup \{\boldsymbol{x}\})$,
$Y \subseteq \Sigma \cup \{\boldsymbol{x}\}$,
$J \subseteq \Sigma \times \{\, P \subseteq \mathbb{P}_{m-1} \mid P \text{ is an } (m-1)\text{-ary tree domain} \,\} \times \mathbb{N}_+ \times (\Sigma \cup \{\boldsymbol{x}\})$.

Let

$$\Sigma' = (Z \cap \Sigma) \cup \bigcup \{\, \Gamma_{c,U} \mid \boldsymbol{x} \in Y \cap Z, (c,U,i,a) \in J, (c,b) \in K \,\} \cup \\ \{\, (c,\varnothing,0) \mid c \in A \cup Y, (c,a) \in K \,\}.$$

Note that Σ' is a finite subset of $\widetilde{\Sigma}$. Define finite sets $A', Z' \subseteq \Sigma'$ and $I \subseteq \Sigma' \times \mathbb{T}^{m-1}_{\Sigma'}$ by

$$\begin{aligned}
A' &= (A \cap Z) \cup \{\, (c,\varnothing,0) \mid c \in A, (c,a) \in K \,\},\\
Z' &= (A \cup Y) \cap Z \cap \Sigma,\\
I &= \{\, ((c,U,0),\boldsymbol{T}_d) \mid (c,U,0) \in \Sigma', d \in Z \cap \Sigma, (c,d) \in K \,\} \cup\\
&\quad \{\, ((c,U,0),\boldsymbol{T}_{(d,V,0)}) \mid (c,U,0),(d,V,0) \in \Sigma', (c,d) \in K, \text{and}\\
&\qquad\qquad \text{either } V \neq \varnothing \text{ or } d \in Y \,\} \cup\\
&\quad \{\, ((c,U,0),\boldsymbol{T}_{(c,U,1)}) \mid (c,U,1) \in \Sigma', |U| = 1, (c,\boldsymbol{x}) \in K \,\} \cup\\
&\quad \{\, ((c,U,i),\boldsymbol{T}_d) \mid (c,U,i) \in \Sigma', (c,U,i,d) \in J, d \in Z \cap \Sigma \,\} \cup\\
&\quad \{\, ((c,U,i),\boldsymbol{T}_{(d,V,0)}) \mid (c,U,i) \in \Sigma', (c,U,i,d) \in J, (d,V,0) \in \Sigma', \text{and}\\
&\qquad\qquad \text{either } V \neq \varnothing \text{ or } d \in Y \,\} \cup\\
&\quad \{\, ((c,U,i),\boldsymbol{T}_{(d,V,j)}) \mid (c,U,i) \in \Sigma', (c,U,i,\boldsymbol{x}) \in J, j \geq 1, (d,V,j) \in \Sigma' \,\} \cup\\
&\quad \{\, (c,\boldsymbol{U}) \mid c \in Z \cap \Sigma, \boldsymbol{U} \in \mathbb{T}^{m-1}_{\Sigma'},\\
&\qquad\qquad \text{for } i = 1,\dots,|U|, \text{ if } u_i \text{ is the } i\text{-th node of } \boldsymbol{U}, \text{ then}\\
&\qquad\qquad (c,U,i,\rho(\ell^{\boldsymbol{U}}(u_i))) \in J, \text{ and } \ell^{\boldsymbol{U}}(u_i) = (d,\varnothing,0) \text{ implies } d \in Y \,\}.
\end{aligned}$$

We show that $L' = \mathrm{Loc}^m(A', Z', I)$ satisfies the desired property.

To prove $\mathbf{enc}_m(L) \subseteq L' \cap DT'^m_{\Sigma}$, suppose $\boldsymbol{T} \in L$ and let $\boldsymbol{T}' = \mathbf{enc}_m(\boldsymbol{T})$. By Lemma 26, $\boldsymbol{T}' \in DT'^m_{\Sigma}$, so it suffices to show $\boldsymbol{T}' \in L'$.

L1. First we show $\ell^{\boldsymbol{T}'}(\varepsilon) \in A'$. Let $c = \ell^{\boldsymbol{T}}(\varepsilon)$. Since $\boldsymbol{T} \in L$, $c \in A$. Suppose first $\varepsilon \notin \mathrm{dom}(\prec^T_{m+1})$. Then since $\boldsymbol{T} \in L$, $c \in Z$. By the definition of $\mathbf{enc}_m$, $\ell^{\boldsymbol{T}'}(\varepsilon) = c \in A'$. Now suppose $\varepsilon \in \mathrm{dom}(\prec^T_{m+1})$. Then since $\boldsymbol{T} \in L$, $(c, \ell^{\boldsymbol{T}}(m+1)) \in K$. Since $\boldsymbol{T}$ is an $(m+1)$-dimensional tree, $C^T_m(\varepsilon) = \varnothing$. By the definition of $\mathbf{enc}_m$, $\ell^{\boldsymbol{T}'}(\varepsilon) = (c,\varnothing,0) \in A'$.

L2. Suppose $v \in T' - \mathrm{dom}(\prec^{T'}_m)$. Let $u = (f^{\boldsymbol{T}}_{\mathbf{enc}_m})^{-1}(v)$. By the definition of $f^{\boldsymbol{T}}_{\mathbf{enc}_m}$, it is clear that $u \notin \mathrm{dom}(\prec^T_{m+1}) \cup \mathrm{dom}(\prec^T_m)$ and $\ell^{\boldsymbol{T}}(u) \in \Sigma$. By Lemma 23, $\ell^{\boldsymbol{T}'}(v) = \ell^{\boldsymbol{T}}(u)$. Since $\boldsymbol{T} \in L$, $\ell^{\boldsymbol{T}}(u) \in (A \cup Y) \cap Z$, and this implies that $\ell^{\boldsymbol{T}'}(v) \in Z'$.

L3. Suppose $v \in \mathrm{dom}(\prec^{T'}_m)$ and $\boldsymbol{U}' = \boldsymbol{C}^{\boldsymbol{T}'}_m(v)$. Let $u = (f^{\boldsymbol{T}}_{\mathbf{enc}_m})^{-1}(v)$.

Case 1. $u \in \mathrm{dom}(\prec^T_{m+1})$. In this case, $\ell^{\boldsymbol{T}'}(v) = (c,U,0)$, where $c = \ell^{\boldsymbol{T}}(u)$ and $U = C^T_m(u)$. Since $\boldsymbol{T} \in L$, we have $(c, \ell^{\boldsymbol{T}}(u \cdot (m+1))) \in K$. If $u \notin \mathrm{dom}(\prec^T_m)$, then $U = \varnothing$, and either $u = \varepsilon$ and $c \in A$ or $c \in Y$. If $u \in \mathrm{dom}(\prec^T_m)$, then $(c,U,1,\ell^{\boldsymbol{T}}(u \cdot m)) \in J$. In either case, we have $(c,U,0) \in \Sigma'$. Note that $U' = \{\varepsilon\}$

and $v \cdot m = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u \cdot (m+1))$. We have

$$\ell^{\boldsymbol{T}'}(v \cdot m) = \begin{cases} d & \text{if } \ell^{\boldsymbol{T}}(u \cdot (m+1)) = d \in \Sigma \text{ and} \\ & u \cdot (m+1) \notin \mathrm{dom}(\prec^T_{m+1}), \\ (d, V, 0) & \text{if } \ell^{\boldsymbol{T}}(u \cdot (m+1)) = d \in \Sigma,\ u \cdot (m+1) \in \mathrm{dom}(\prec^T_{m+1}), \\ & \text{and } C^T_m(u \cdot (m+1)) = V, \\ (c, U, 1) & \text{if } \ell^{\boldsymbol{T}}(u \cdot (m+1)) = \boldsymbol{x}. \end{cases}$$

Since $\boldsymbol{T} \in L$, we have $d \in Z \cap \Sigma \subseteq \Sigma'$ in the first case. In the second case, $(d, \ell^{\boldsymbol{T}}(u \cdot (m+1) \cdot (m+1))) \in K$ and either $V = \varnothing$ and $d \in Y$ or $(d, V, 1, \ell^{\boldsymbol{T}}(u \cdot (m+1) \cdot m)) \in J$, which implies $(d, V, 0) \in \Sigma'$. In the third case, since $\boldsymbol{T}$ is well-labeled, we have $|U| = 1$ and $u \cdot (m+1) \notin \mathrm{dom}(\prec^T_m)$, and $\boldsymbol{T} \in L$ implies $\boldsymbol{x} \in Y \cap Z$ and $(c, U, 1, \ell^{\boldsymbol{T}}(u \cdot m)) \in J$. It follows that $(c, U, 1) \in \Sigma'$. In each case, we have $((c, U, 0), \boldsymbol{U}') = ((c, U, 0), \boldsymbol{T}_{\ell^{\boldsymbol{T}'}(v \cdot m)}) \in I$.

Case 2. $u \notin \mathrm{dom}(\prec^T_{m+1})$.

Case 2a. $\ell^{\boldsymbol{T}}(u) = c \in \Sigma$. In this case, $\ell^{\boldsymbol{T}'}(v) = c \in \Sigma \cap Z \subseteq \Sigma'$ and it is easy to see from Lemma 24 that $U' = C^T_m(u)$ and $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u \cdot m \cdot w) = v \cdot m \cdot w$ for all $w \in C^T_m(u)$. Since $\boldsymbol{T} \in L$, if $u \lhd^T_{m,i} u \cdot m \cdot u_i$, then $(c, C^T_m(u), i, \ell^{\boldsymbol{T}}(u \cdot m \cdot u_i)) = (c, U', i, \rho(\ell^{\boldsymbol{T}'}(v \cdot m \cdot u_i))) = (c, U', i, \rho(\ell^{\boldsymbol{U}'}(u_i))) \in J$. Also, if $\ell^{\boldsymbol{U}'}(u_i) = \ell^{\boldsymbol{T}'}(v \cdot m \cdot u_i) = (d, \varnothing, 0)$ for some $d \in \Sigma$, then $u \cdot m \cdot u_i \notin \mathrm{dom}(\prec^T_m)$ and hence $d \in Y$. It follows that $(c, \boldsymbol{U}') \in I$.

Case 2b. $\ell^{\boldsymbol{T}}(u) = \boldsymbol{x}$. In this case, for some $w, z \in T - \{\varepsilon\}$, $w \blacktriangleleft^{\boldsymbol{T}}_{m+1,i} u, w \lhd^T_{m,i} z$. By Lemma 24, we have $U' = \{\varepsilon\}$ and $v \cdot m = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(z)$. Let $c = \ell^{\boldsymbol{T}}(w) \in \Sigma$ and $U = C^T_m(w)$. Since $\boldsymbol{T} \in L$, $(c, \ell^{\boldsymbol{T}}(w \cdot m)) \in K$ and $(c, U, i, \ell^{\boldsymbol{T}}(z)) \in J$. Also, $\ell^{\boldsymbol{T}}(u) = \boldsymbol{x} \in Y \cap Z$, since $\boldsymbol{T}$ is well-labeled. Hence $\ell^{\boldsymbol{T}'}(v) = (c, U, i) \in \Sigma'$ and

$$\ell^{\boldsymbol{T}'}(v \cdot m) = \begin{cases} d & \text{if } \ell^{\boldsymbol{T}}(z) = d \in \Sigma \text{ and } z \notin \mathrm{dom}(\prec^T_{m+1}), \\ (d, V, 0) & \text{if } \ell^{\boldsymbol{T}}(z) = d \in \Sigma,\ z \in \mathrm{dom}(\prec^T_{m+1}), \text{ and } C^T_m(z) = V, \\ (d, V, j) & \text{if } \ell^{\boldsymbol{T}}(z) = \boldsymbol{x},\ t \blacktriangleleft^{\boldsymbol{T}}_{m+1,j} z,\ \ell^{\boldsymbol{T}}(t) = d, \text{ and } C^T_m(t) = V. \end{cases}$$

In the first case, since $\boldsymbol{T} \in L$, $d \in Z$. In the second case, $(d, \ell^{\boldsymbol{T}}(z \cdot (m+1))) \in K$, and if $V = \varnothing$, then $z \notin \mathrm{dom}(\prec^T_m)$ and hence $d \in Y$. In the third case, $(d, \ell^{\boldsymbol{T}}(t \cdot (m+1))) \in K$. It follows that in each case, $((c, U, i), \boldsymbol{U}') = ((c, U, i), \boldsymbol{T}_{\ell^{\boldsymbol{T}'}(v \cdot m)}) \in I$.

We have proved that $\boldsymbol{T}'$ satisfies the requirements L1–L3 for membership in $L' = \mathrm{Loc}^m(A', Z', I)$.

To show $L' \cap DT'^m_\Sigma \subseteq \mathbf{enc}_m(L)$, let $\boldsymbol{T}' \in L' \cap DT'^m_\Sigma$. By Lemma 26, there is a $\boldsymbol{T} \in \mathbb{T}^{m+1}_{\Sigma,\boldsymbol{x}}$ such that $\boldsymbol{T}' = \mathbf{enc}_m(\boldsymbol{T})$. It suffices to prove that $\boldsymbol{T} \in L$.

We show that $\boldsymbol{T}$ satisfies all requirements for membership in $L = \mathrm{SLoc}^{m+1}(A, Z, K, Y, J)$.

SL1. The fact that $\ell^{\boldsymbol{T}'}(\varepsilon) \in A'$ easily implies $\ell^{\boldsymbol{T}}(\varepsilon) = \rho(\ell^{\boldsymbol{T}'}(\varepsilon)) \in A$.

SL2. Let $v \in T - \mathrm{dom}(\prec^T_{m+1})$.

Case 1. $\ell^{\boldsymbol{T}}(v) = c \in \Sigma$. Then $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = c \in \Sigma' \subseteq Z$, and it follows that $c \in Z$.

Case 2. $\ell^{\boldsymbol{T}}(v) = \boldsymbol{x}$. Then $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = (d, V, i) \in \Sigma'$ for some $V \neq \varnothing$ and $i \geq 1$, and it follows that $\boldsymbol{x} \in Z$.

SL3. Let $v \in \mathrm{dom}(\prec^T_{m+1})$. We show $(\ell^{\boldsymbol{T}}(v), \ell^{\boldsymbol{T}}(v \cdot (m+1))) \in K$. By Lemma 24, we have $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v \cdot (m+1)) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \cdot m$. Let $c = \ell^{\boldsymbol{T}}(v)$ and $U = C^T_m(v)$. Then $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = (c, U, 0)$ and $\boldsymbol{C}^{\boldsymbol{T}'}_m(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = \boldsymbol{T}_{\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \cdot m)}$. Since $\boldsymbol{T}' \in L'$, it holds that $((c, U, 0), \boldsymbol{T}_{\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \cdot m)}) \in I$.

Case 1. $\ell^{\boldsymbol{T}}(v \cdot (m+1)) = d \in \Sigma$. Then $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \cdot m)$ is either d or $(d, V, 0)$ for some V depending on whether $v \cdot (m+1) \in \mathrm{dom}(\prec^T_{m+1})$ or not. Either way, $((c, U, 0), \boldsymbol{T}_{\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \cdot m)}) \in I$ implies $(\ell^{\boldsymbol{T}}(v), \ell^{\boldsymbol{T}}(v \cdot (m+1))) = (c, d) \in K$.

Case 2. $\ell^{\boldsymbol{T}}(v \cdot (m+1)) = \boldsymbol{x}$. Then $|U| = 1$ and $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \cdot m) = (c, U, 1)$. Since $((c, U, 0), \boldsymbol{T}_{\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \cdot m)}) \in I$, we have $(\ell^{\boldsymbol{T}}(v), \ell^{\boldsymbol{T}}(v \cdot (m+1))) = (c, \boldsymbol{x}) \in K$.

SL4. Suppose $v \neq \varepsilon$ and $v \in T - \mathrm{dom}(\prec^T_m)$. We show $\ell^{\boldsymbol{T}}(v) \in Y$.

Case 1. $\ell^{\boldsymbol{T}}(v) = c \in \Sigma$.

Case 1a. $v \in \mathrm{dom}(\prec^T_{m+1})$. Then $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = (c, \varnothing, 0)$. Since $v \neq \varepsilon$, $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \neq \varepsilon$ and there are some $u \in T'$ and $w \in \mathbb{P}_{m-1}$ such that $u \cdot m \cdot w = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)$. Since $\boldsymbol{T}' \in L'$, $(\ell^{\boldsymbol{T}'}(u), \boldsymbol{C}^{\boldsymbol{T}'}_m(u)) \in I$. Since $(c, \varnothing, 0) = \ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = \ell^{\boldsymbol{C}^{\boldsymbol{T}'}_m(u)}(w)$, the definition of I implies $c \in Y$.

Case 1b. $v \notin \mathrm{dom}(\prec^T_{m+1})$. Then $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = c$. Since $v \notin \mathrm{dom}(\prec^T_m)$ and $\ell^{\boldsymbol{T}}(v) \neq \boldsymbol{x}$, Lemma 24 implies $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) \notin \mathrm{dom}(\prec^{T'}_m)$. Since $\boldsymbol{T}' \in L'$, $c \in Z'$, which implies $c \in Y$.

Case 2. $\ell^{\boldsymbol{T}}(v) = \boldsymbol{x}$. Then $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = (d, V, j) \in \Sigma'$ for some $V \neq \varnothing$ and $j \geq 1$, which implies $\boldsymbol{x} \in Y$.

SL5. Let $u \lhd^T_{m,i} v$, $C^T_m(u) = U$, $c = \ell^{\boldsymbol{T}}(u)$. We show $(c, U, i, \ell^{\boldsymbol{T}}(v)) \in J$.

Case 1. $u \in \mathrm{dom}(\prec^T_{m+1})$. Then $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u)) = (c, U, 0)$ and for some w, $u \blacktriangleleft^{\boldsymbol{T}}_{m+1,i} w$, $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(w)) = (c, U, i)$, $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(w) \cdot m$, and $\boldsymbol{C}^{\boldsymbol{T}'}_m(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(w)) = \boldsymbol{T}_{\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v))}$. Since $\boldsymbol{T}' \in L'$, $((c, U, i), \boldsymbol{T}_{\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v))}) \in I$. We have

$$\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)) = \begin{cases} d & \text{if } \ell^{\boldsymbol{T}}(v) = d \in \Sigma \text{ and } v \notin \mathrm{dom}(\prec^T_{m+1}), \\ (d, V, 0) & \text{if } \ell^{\boldsymbol{T}}(v) = d \in \Sigma,\ v \in \mathrm{dom}(\prec^T_{m+1}), \text{ and } C^T_m(v) = V, \\ (d, V, j) & \text{if } \ell^{\boldsymbol{T}}(v) = \boldsymbol{x},\ z \blacktriangleleft^{\boldsymbol{T}}_{m+1,j} v,\ \ell^{\boldsymbol{T}}(z) = d \in \Sigma, \text{ and} \\ & \quad C^T_m(z) = V. \end{cases}$$

In each case, $((c, U, i), \boldsymbol{T}_{\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v))}) \in I$ implies $(c, U, i, \ell^{\boldsymbol{T}}(v)) \in J$.

Case 2. $u \notin \mathrm{dom}(\prec^T_{m+1})$. In this case, $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u)) = c$, $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u \cdot m \cdot t) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u) \cdot m \cdot t$ for all $t \in C^T_m(u)$, and $C^{T'}_m(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u)) = C^T_m(u) = U$. In particular, $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u) \lhd^{T'}_{m,i} f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)$. Since $\boldsymbol{T}' \in L'$, $(c, \boldsymbol{C}^{\boldsymbol{T}'}_m(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u))) \in I$. Let u_i be the i-th node of U, so that $f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v) = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u) \cdot m \cdot u_i$. Then $(c, U, i, \ell^{\boldsymbol{T}}(v)) = (c, U, i, \rho(\ell^{\boldsymbol{T}'}(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(v)))) = (c, C^{T'}_m(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u)), i, \rho(\ell^{\boldsymbol{C}^{\boldsymbol{T}'}_m(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u))}(u_i))) \in J$, by the definition of I.

This establishes $\boldsymbol{T} \in L = \mathrm{SLoc}^{m+1}(A, Z, K, Y, J)$. □

The converse of the above lemma does not hold for a reason similar to the one for the case of the standard **enc** function for dimension 1.[19]

A projection $\pi\colon \Sigma' \to \Sigma$ naturally induces a projection $\widetilde{\pi}\colon \widetilde{\Sigma'} \to \widetilde{\Sigma}$ in an obvious way:

$$\widetilde{\pi}(c) = \pi(c),$$
$$\widetilde{\pi}((c, P, i)) = (\pi(c), P, i).$$

Clearly, if $\boldsymbol{T}' \in DT^m_{\Sigma'}$, then $\widetilde{\pi}(\boldsymbol{T}') \in DT^m_{\Sigma}$. Also, if $\boldsymbol{T}' \in \mathbb{T}^{m+1}_{\Sigma',\boldsymbol{x}}$, then $\mathbf{enc}_m(\pi(\boldsymbol{T}')) = \widetilde{\pi}(\mathbf{enc}_m(\boldsymbol{T}'))$.

Here is a generalization of Lemma 6 to multi-dimensional Dyck languages:

Lemma 28. *Let $L \subseteq \mathbb{T}^m_{\widetilde{\Sigma}}$ be a local set. Then there exist a finite alphabet Σ', a projection $\pi\colon \Sigma' \to \Sigma$, and a local set $L' \subseteq \mathbb{T}^m_{\widetilde{\Sigma'}}$ such that $L \cap DT'^m_{\Sigma} = \widetilde{\pi}(L' \cap DT^m_{\Sigma'})$. Moreover, $\widetilde{\pi}$ maps $L' \cap DT^m_{\Sigma'}$ bijectively to $L \cap DT'^m_{\Sigma}$.*

Proof. Let $A, Z \subseteq \widetilde{\Sigma}$ and $I \subseteq \widetilde{\Sigma} \times \mathbb{T}^{m-1}_{\widetilde{\Sigma}}$ be finite sets such that $L = \mathrm{Loc}^m(A, Z, I)$. Since we are interested in the intersection of L and DT'^m_{Σ}, we may assume without loss of generality $Z \subseteq \Sigma$. Define

$$\begin{aligned}
\Sigma_0 &= Z \cup \{\, c \in \Sigma \mid (c, \boldsymbol{T}) \in I \,\},\\
\Sigma' &= \Sigma_0 \cup \{\, \bar{c} \mid c \in A \cap Z \,\},\\
A' &= \{\, \bar{c} \mid c \in A \cap Z \,\} \cup \{\, (c, \varnothing, 0) \mid c \in \Sigma_0, (c, \varnothing, 0) \in A \,\},\\
Z' &= Z \cup \{\, \bar{c} \mid c \in A \cap Z \,\}.
\end{aligned}$$

Then A' and Z' are finite subsets of $\widetilde{\Sigma'}$. Let $\pi\colon \Sigma' \to \Sigma$ be the projection defined by

$$\pi(c) = c, \quad \pi(\bar{d}) = d$$

for each $c \in \Sigma_0$ and $d \in A \cap Z$. Let

$$L' = \mathrm{Loc}^m(A', Z', I).$$

It is easy to see that $L \cap DT'^m_{\Sigma} = \widetilde{\pi}(L' \cap DT^m_{\Sigma'})$ and for each $\boldsymbol{T} \in L \cap DT'^m_{\Sigma}$, there is a unique $\boldsymbol{T}' \in L'$ such that $\pi(\boldsymbol{T}') = \boldsymbol{T}$. ☐

Lemma 29. *If $L \subseteq \mathbb{T}^{m+1}_{\Sigma,\boldsymbol{x}}$ is a local set, then there exist a finite alphabet Σ', a projection $\pi\colon \Sigma' \to \Sigma$, and a local set $L' \subseteq \mathbb{T}^m_{\widetilde{\Sigma'}}$ such that*

$$\mathbf{enc}_m(L) = \widetilde{\pi}(L' \cap DT^m_{\Sigma'}).$$

Moreover, $\mathbf{enc}_m^{-1} \circ \widetilde{\pi}$ maps $L' \cap DT^m_{\Sigma'}$ bijectively to L.

[19] There is also an additional reason. $L = \{a -_3 a\}$ is not super-local even though $\mathbf{enc}_2(L) = \{(a, \varnothing, 0) -_2 a\}$ is local.

Proof. By Lemma 11, there exist a projection $\pi_1 \colon \Sigma_1 \to \Sigma$ and a super-local $L_1 \subseteq \mathbb{T}^{m+1}_{\Sigma_1,\boldsymbol{x}}$ such that $L = \pi_1(L_1)$. By Lemma 27, there exist a local set $L_2 \subseteq \mathbb{T}^{m}_{\widetilde{\Sigma_1}}$ such that $\mathbf{enc}_m(L_1) = L_2 \cap DT'^{m}_{\Sigma_1}$. By Lemma 28, there exist a projection $\pi_2 \colon \Sigma' \to \Sigma_1$ and a local set $L' \subseteq \mathbb{T}^{m}_{\widetilde{\Sigma'}}$ such that $L_2 \cap DT'^{m}_{\Sigma_1} = \widetilde{\pi_2}(L' \cap DT^{m}_{\Sigma'})$. So

$$\begin{aligned} \mathbf{enc}_m(L) &= \mathbf{enc}_m(\pi_1(L_1)) \\ &= \widetilde{\pi_1}(\mathbf{enc}_m(L_1)) \\ &= \widetilde{\pi_1}(L_2 \cap DT'^{m}_{\Sigma_1}) \\ &= \widetilde{\pi_1}(\widetilde{\pi_2}(L' \cap DT^{m}_{\Sigma'})) \\ &= \widetilde{\pi}(L' \cap DT^{m}_{\Sigma'}), \end{aligned}$$

where $\pi = \pi_1 \circ \pi_2$. Since π_1 is a bijection from L_1 to L and $\mathbf{enc}_m$ is injective, $\widetilde{\pi_1}$ maps $\mathbf{enc}_m(L_1)$ bijectively to $\mathbf{enc}_m(L)$. Since $\widetilde{\pi_2}$ maps $L' \cap DT^{m}_{\Sigma'}$ bijectively to $L_2 \cap DT'^{m}_{\Sigma_1} = \mathbf{enc}_m(L_1)$, $\widetilde{\pi} = \widetilde{\pi_1} \circ \widetilde{\pi_2}$ maps $L' \cap DT^{m}_{\Sigma'}$ bijectively to $\mathbf{enc}(L)$. □

8 A Multi-dimensional Generalization of the Chomsky-Schützenberger Theorem

Let $m \geq 2$. We call $L \subseteq \mathbb{T}^{m}_{\Sigma}$ *simple context-free* if there exist a finite alphabet Υ and a local set $K \subseteq \mathbb{T}^{m+1}_{\Upsilon,\boldsymbol{x}}$ such that $L = \mathbf{y}_m(K)$.

For a finite alphabet Σ and $r \geq 0$, we define the finite alphabet

$$\widetilde{\Sigma}_r = \Sigma \cup \bigcup \{\, \Gamma_{c,P} \mid c \in \Sigma,\ P \text{ is finite and prefix-closed, } |P| \leq r \,\}.$$

For any alphabet Υ and $p \geq 1$, let

$$\mathbb{T}^{m}_{\Upsilon,p} = \{\, \boldsymbol{T} \in \mathbb{T}^{m}_{\Upsilon} \mid |C^{T}_{m}(v)| \leq p \text{ for all } v \in T \,\}.$$

Clearly, if Υ is finite, $\mathbb{T}^{m}_{\Upsilon,p}$ is a local subset of $\mathbb{T}^{m}_{\Upsilon}$. Also, any local subset L of $\mathtt{T}^{m}_{\Upsilon}$ is included in $\mathtt{T}^{m}_{\Upsilon,p}$ for some p, which is just another way of saying L is degree-bounded.

Lemma 30. *Let Σ be a finite set. For $m \geq 2$, $DT^{m}_{\Sigma} \cap \mathbb{T}^{m}_{\widetilde{\Sigma}_r,p}$ is simple context-free.*

Proof. We adapt the inductive definition of $DT^{m}_{\Sigma}(n)$ to obtain the required local set. Let $\Upsilon = \widetilde{\Sigma}_r \cup \{X_0, \dots, X_r\}$. We write U_k for the set $\{\varepsilon, (m-1), \dots, (m-1)^{k-1}\} \subseteq \mathbb{P}_{m-1}$. Let

$$\begin{aligned} A_n &= \{X_n\} \quad \text{for } n = 0, \dots, r, \\ Z &= \widetilde{\Sigma}_r \cup \{\boldsymbol{x}\}, \end{aligned}$$

$$I = \{(X_0, \boldsymbol{T}_c), (X_1, \boldsymbol{T}_{\boldsymbol{x}})\} \cup$$
$$\left\{ \left(X_n, \begin{array}{l} c -_m P(\\ \quad X_{n_1} -_m U_{n_1}(\boldsymbol{x}, \ldots, \boldsymbol{x}), \\ \quad \ldots, \\ \quad X_{n_k} -_m U_{n_k}(\boldsymbol{x}, \ldots, \boldsymbol{x}) \\ ) \end{array} \right) \middle| \begin{array}{l} P \subseteq \mathbb{P}_{m-1}, \\ P \text{ is finite and prefix-closed,} \\ 1 \leq |P| = k \leq p, \\ 0 \leq n = n_1 + \cdots + n_k \leq r \end{array} \right\} \cup$$
$$\left\{ \left(X_n, \begin{array}{l} (c, P, 0) -_m X_k -_m U_k(\\ \quad (c, P, 1) -_m X_{n_1} -_m U_{n_1}(\boldsymbol{x}, \ldots, \boldsymbol{x}), \\ \quad \ldots, \\ \quad (c, P, k) -_m X_{n_k} -_m U_{n_k}(\boldsymbol{x}, \ldots, \boldsymbol{x}) \\ ) \end{array} \right) \middle| \begin{array}{l} P \subseteq \mathbb{P}_{m-1}, \\ P \text{ is finite and prefix-closed,} \\ 0 \leq |P| = k \leq r, \\ 0 \leq n = n_1 + \cdots + n_k \leq r \end{array} \right\}.$$

Here, the number of occurrences of $\boldsymbol{x}$ in $U_{n_i}(\boldsymbol{x}, \ldots, \boldsymbol{x})$ is $|U_{n_i}| = n_i$. When $j = 0$, we understand the notation $X_j -_m U_j(\boldsymbol{x}, \ldots, \boldsymbol{x})$ to mean X_0, i.e., the tree consisting of a single node labeled by X_0. Note that A_n and Z are (finite) subsets of Υ and I is a finite subset of $\Upsilon \times \mathbb{T}^m_{\Upsilon \cup \boldsymbol{x}}$. It is straightforward to prove that

$$DT^m_{\Sigma}(n) \cap \mathbb{T}^m_{\widetilde{\Sigma}_r, p}(n) = \mathbf{y}_m(\mathrm{Loc}^{m+1}(A_n, Z, I))$$

holds for $n = 0, \ldots, r$. The case of $n = 0$ gives the statement of the lemma. We omit the details. ◻

The following lemma is straightforward.

Lemma 31. *If L is simple context-free, there are finite sets A, Z, I such that $L = \mathbf{y}_m(\mathrm{Loc}^{m+1}(A, Z, I))$ and $Z \cap \{ c \mid (c, \boldsymbol{T}) \in I \} = \varnothing$.*

Recall the definition of $\mathbf{del}_m(\boldsymbol{T}, U)$ for $U \subseteq \{ v \in T \mid |C^T_m(v)| = 1 \}$, which was given in terms of the function $f_U \colon T \to [1, m]^*$:

$$f_U(\varepsilon) = \varepsilon,$$
$$f_U(v \cdot i) = f_U(v) \cdot i \qquad \text{for } i < m,$$
$$f_U(v \cdot m) = \begin{cases} f_U(v) \cdot m & \text{if } v \notin U, \\ f_U(v) & \text{if } v \in U. \end{cases}$$

For $\boldsymbol{T}^m_{\Sigma}$ and $\Upsilon \subseteq \Sigma$, $\mathbf{del}_{m,\Upsilon}(\boldsymbol{T})$ was defined to be $\mathbf{del}_m(\boldsymbol{T}, U)$, with $U = \{ v \in T \mid \ell^{\boldsymbol{T}}(v) \in \Upsilon, |C^T_m(v)| = 1 \}$.

For $\boldsymbol{T} \in \mathbb{T}^{m+1}_{\Sigma, \boldsymbol{x}}$, let $f^{\boldsymbol{T}}_{\mathbf{y}_m}$ be the mapping from the nodes of $\boldsymbol{T}$ to the nodes of $\mathbf{y}_m(\boldsymbol{T})$ defined by

$$f^{\boldsymbol{T}}_{\mathbf{y}_m}(u) = f_U(f^{\boldsymbol{T}}_{\mathbf{enc}_m}(u)),$$

where

$$U = f^{\boldsymbol{T}}_{\mathbf{enc}_m}(\{ u \in T \mid u \in \mathrm{dom}(\prec^T_{m+1}) \text{ or } \ell^{\boldsymbol{T}}(u) = \boldsymbol{x} \}).$$

It is clear that if $u \in T - \mathrm{dom}(\prec_{m+1})$ and $\ell^{\boldsymbol{T}}(u) \neq \boldsymbol{x}$, then $\ell^{\boldsymbol{T}}(u) = \ell^{\mathbf{y}_m(\boldsymbol{T})}(f^{\boldsymbol{T}}_{\mathbf{y}_m}(u))$.[20]

[20] Here and in the proof of the next lemma, we use the notations $\ell^{\mathbf{y}_m(\boldsymbol{T})}$ and $\ell^{\mathbf{enc}_m(\boldsymbol{T})}$ with their obvious meaning.

Lemma 32. *Let $L \subseteq \mathbb{T}^m_\Sigma$ be a simple context-free set.*

(i) *If $L' \subseteq \mathbb{T}^m_\Sigma$ is local, then $L \cap L'$ is simple context-free.*
(ii) *For every projection $\pi \colon \Sigma \to \Sigma'$, $\pi(L)$ is simple context-free.*
(iii) *If $\Sigma' \subseteq \Sigma$, then $\mathbf{del}_{m,\Sigma'}(L)$ is simple context-free.*

Proof. Let $L = \mathbf{y}_m(\mathrm{Loc}^{m+1}(A, Z, I))$, where $\Sigma \subseteq \Upsilon$ and $A \subseteq \Upsilon$, $Z \subseteq \Upsilon \cup \{\boldsymbol{x}\}$, $I \subseteq \Upsilon \times \mathbb{T}^m_{\Upsilon \cup \{\boldsymbol{x}\}}$ are finite sets. By Lemma 31, we may assume $Z \subseteq \Sigma \cup \{\boldsymbol{x}\}$ and $\{\, c \mid (c, \boldsymbol{T}) \in I \,\} \subseteq \Upsilon - \Sigma$. Let $r = \max\{\, n \mid (c, \boldsymbol{T}) \in I, \boldsymbol{T} \in \mathbb{T}^m_\Upsilon(n) \,\}$. We only prove (i), since (ii) and (iii) are straightforward.[21]

Let $L' = \mathrm{Loc}^m(A', Z', I')$, where $A', Z' \subseteq \Sigma$, $I' \subseteq \Sigma \times \mathbb{T}^{m-1}_\Sigma$ are finite sets. Let

$$\Upsilon_1 = \{\, (c, d, e_1, \dots, e_n) \mid c \in \Upsilon - \Sigma, d, e_1, \dots, e_n \in \Sigma, n \leq r \,\}.$$

Define projections $\pi_1 \colon \Sigma \cup \Upsilon_1 \to \Upsilon$ and $\pi_2 \colon \Sigma \cup \Upsilon_1 \to \Sigma$ by

$$\begin{aligned} \pi_1(c) &= c \quad \text{for } c \in \Sigma, \\ \pi_1((c, d, e_1, \dots, e_n)) &= c \quad \text{for } (c, d, e_1, \dots, e_n) \in \Upsilon_1, \\ \pi_2(c) &= c \quad \text{for } c \in \Sigma, \\ \pi_2((c, d, e_1, \dots, e_n)) &= d \quad \text{for } (c, d, e_1, \dots, e_n) \in \Upsilon_1. \end{aligned}$$

Let

$$\begin{aligned} &A''_d = \pi_2^{-1}(\{d\}) \quad \text{for } d \in \Sigma, \\ &Z'' = Z' \cup \{\, (c, d, e_1, \dots, e_n) \in \Upsilon_1 \mid n = 0 \,\}, \\ &I'' = \{\, (c, \boldsymbol{T}) \in \Sigma \times \mathbb{T}^{m-1}_{\Sigma \cup \Upsilon_1} \mid (c, \pi_2(\boldsymbol{T})) \in I' \,\} \cup \\ &\qquad \{\, ((c, d, e_1, \dots, e_n), \boldsymbol{T}) \in \Upsilon_1 \times \mathbb{T}^{m-1}_{\Sigma \cup \Upsilon_1} \mid \\ &\qquad\qquad |T| = n \text{ and if } u_i \text{ is the } i\text{-th node of } T, \text{ then } \pi_2(\ell^{\boldsymbol{T}}(u_i)) = e_i \,\}. \end{aligned}$$

Define

$$\begin{aligned} &A_1 = (A \cap A' \cap Z') \cup \{\, (c, d) \in \Upsilon_1 \mid c \in A, d \in A' \,\}, \\ &Z_1 = Z, \\ &I_1 = \{\, ((c, d, e_1, \dots, e_n), \boldsymbol{T}) \mid (c, d, e_1, \dots, e_n) \in \Upsilon_1, \boldsymbol{T} \in \mathbb{T}^m_{\Sigma \cup \Upsilon_1}(n), (c, \pi_1(\boldsymbol{T})) \in I, \\ &\qquad\qquad \boldsymbol{T}[(c, e_1), \dots, (c, e_n)] \in \mathrm{Loc}^m(A''_d, Z'', I'') \,\}. \end{aligned}$$

Note that A_1, Z_1, I_1 are finite sets. We show $\mathbf{y}_m(\mathrm{Loc}^{m+1}(A_1, Z_1, I_1)) = L \cap L'$.

We first note some useful facts about members of $\mathrm{Loc}^{m+1}(A_1, Z_1, I_1)$. To state this, we need another projection and a certain relabeling defined on some subset of $\mathbb{T}^{m+1}_{\Sigma \cup \Upsilon_1, \boldsymbol{x}}$. Define a projection $\pi_3 \colon \Sigma \cup (\widetilde{\Upsilon_1} - \Upsilon_1) \to \Sigma$ by

$$\begin{aligned} \pi_3(c) &= c \quad \text{if } c \in \Sigma, \\ \pi_3(((c, d, e_1, \dots, e_n), P, 0)) &= d, \\ \pi_3(((c, d, e_1, \dots, e_n), P, i)) &= e_i \quad \text{if } 1 \leq i \leq n. \end{aligned}$$

[21] This proof is very long and laborious. An alternative approach would be to use the notion of a recognizable (equivalently, MSO-definable) set of m-dimensional trees [24,23] and rely on the fact that the yield mapping is an MSO-definable transduction.

Let

$$M = \{\, \boldsymbol{T} \in \mathbb{T}^{m+1}_{\Sigma \cup \Upsilon_1, \boldsymbol{x}} \mid \text{for all } v \in T, \text{ (i) } v \in \mathrm{dom}(\prec^T_{m+1}) \text{ if and only if } \ell^{\boldsymbol{T}_1}(v) \in \Upsilon_1 \text{ and (ii) } \ell^{\boldsymbol{T}_1}(v) = (c, d, e_1, \dots, e_n) \text{ implies } \boldsymbol{C}^{\boldsymbol{T}_1}_{m+1}(v) \in \mathbb{T}^m_{\Sigma \cup \Upsilon_1}(n) \,\}.$$

It is clear that $\mathrm{Loc}^{m+1}(A_1, Z_1, I_1) \subseteq M$. For $\boldsymbol{T}_1 = (T, \ell^{\boldsymbol{T}_1}) \in M$, define $\boldsymbol{T}_1^\dagger = (T, \ell^{\boldsymbol{T}_1^\dagger})$ by

$$\ell^{\boldsymbol{T}_1^\dagger}(v) = \begin{cases} \ell^{\boldsymbol{T}_1}(v) & \text{if } \ell^{\boldsymbol{T}_1}(v) \in \Sigma \cup \Upsilon_1, \\ (c, e_i) & \text{if } \ell^{\boldsymbol{T}_1}(v) = \boldsymbol{x},\ u \blacktriangleleft_{m+1,i} v, \text{ and } \ell^{\boldsymbol{T}_1}(u) = (c, d, e_1, \dots, e_n). \end{cases} \tag{14}$$

It is easy to see that for all $u \in T$, we have

$$\pi_2(\ell^{\boldsymbol{T}_1^\dagger}(u)) = \pi_3(\ell^{\mathbf{enc}_m(\boldsymbol{T}_1)}(f^{\boldsymbol{T}_1}_{\mathbf{enc}_m}(u))). \tag{15}$$

Now let $\boldsymbol{T}_1 = (T, \ell^{\boldsymbol{T}_1}) \in \mathrm{Loc}^{m+1}(A_1, Z_1, I_1)$. Since $\boldsymbol{T}_1 \in \mathrm{Loc}^{m+1}(A_1, Z_1, I_1)$, it is also easy to see the following:

$$u \in \mathrm{dom}(\prec^T_{m+1}) \text{ implies } \pi_2(\ell^{\boldsymbol{T}_1^\dagger}(u)) = \pi_2(\ell^{\boldsymbol{T}_1^\dagger}(u \cdot (m+1))), \tag{16}$$

$$u \blacktriangleleft^{\boldsymbol{T}_1}_{m+1,i} v \text{ and } u \lhd^T_{m,i} w \text{ imply } \pi_2(\ell^{\boldsymbol{T}_1^\dagger}(v)) = \pi_2(\ell^{\boldsymbol{T}_1^\dagger}(w)). \tag{17}$$

It follows from (15), (16), and (17) that if $u \in \mathrm{dom}(\prec^T_{m+1})$ or $\ell^{\boldsymbol{T}_1}(u) = \boldsymbol{x}$, then

$$\pi_3(\ell^{\boldsymbol{T}_1}(f^{\boldsymbol{T}_1}_{\mathbf{enc}_m}(u)) = \pi_3(\ell^{\boldsymbol{T}_1}(f^{\boldsymbol{T}_1}_{\mathbf{enc}_m}(u) \cdot m)). \tag{18}$$

Let $U = f^{\boldsymbol{T}_1}_{\mathbf{enc}_m}(\mathrm{dom}(\prec^T_{m+1}) \cup \{\, v \in T \mid \ell^{\boldsymbol{T}_1}(v) = \boldsymbol{x} \,\})$. Note that if $u \in T$ and $\ell^{\boldsymbol{T}_1}(u) \in \Sigma$ (or, equivalently, if $f^{\boldsymbol{T}_1}_{\mathbf{enc}_m}(u) \notin U$), then $\ell^{\boldsymbol{T}_1}(u) = \ell^{\mathbf{enc}_m(\boldsymbol{T}_1)}(f^{\boldsymbol{T}_1}_{\mathbf{enc}_m}(u)) = \ell^{\mathbf{y}_m(\boldsymbol{T}_1)}(f^{\boldsymbol{T}_1}_{\mathbf{y}_m}(u))$. By (18), we can see that for all nodes w, t of $\mathbf{enc}_m(\boldsymbol{T}_1)$, $f_U(w) = f_U(t)$ implies $\pi_3(\ell^{\mathbf{enc}_m(\boldsymbol{T}_1)}(w)) = \pi_3(\ell^{\mathbf{enc}_m(\boldsymbol{T}_1)}(t))$. So if w' is the unique node such that $w' \notin U$ and $f_U(w) = f_U(w')$, then

$$\begin{aligned} \pi_3(\ell^{\mathbf{enc}_m(\boldsymbol{T}_1)}(w)) &= \pi_3(\ell^{\mathbf{enc}_m(\boldsymbol{T}_1)}(w')) \\ &= \pi_3(\ell^{\mathbf{y}_m(\boldsymbol{T}_1)}(f_U(w'))) \\ &= \ell^{\mathbf{y}_m(\boldsymbol{T}_1)}(f_U(w')) \\ &= \ell^{\mathbf{y}_m(\boldsymbol{T}_1)}(f_U(w)). \end{aligned}$$

It follows from this and (15) that for all $u \in T$, we have

$$\pi_2(\ell^{\boldsymbol{T}_1^\dagger}(u)) = \ell^{\mathbf{y}_m(\boldsymbol{T}_1)}(f^{\boldsymbol{T}_1}_{\mathbf{y}_m}(u)). \tag{19}$$

Now for $\boldsymbol{T} \in \mathrm{Loc}^{m+1}(A, Z, I)$, define $\hat{\boldsymbol{T}} = (T, \ell^{\hat{\boldsymbol{T}}}) \in \mathbb{T}^{m+1}_{\Sigma \cup \Upsilon_1, \boldsymbol{x}}$ by

$$\ell^{\hat{\boldsymbol{T}}}(u) = \begin{cases} \ell^{\boldsymbol{T}}(u) & \text{if } u \notin \mathrm{dom}(\prec^T_{m+1}), \\ (c, d, e_1, \dots, e_n) & \text{if } u \in \mathrm{dom}(\prec^T_{m+1}),\ c = \ell^{\boldsymbol{T}}(u), \\ & d = \ell^{\mathbf{y}_m(\boldsymbol{T})}(f^{\boldsymbol{T}}_{\mathbf{y}_m}(u)),\ |C^T_m(u)| = n,\ u \lhd^T_{m,i} w_i, \\ & e_i = \ell^{\mathbf{y}_m(\boldsymbol{T})}(f^{\boldsymbol{T}}_{\mathbf{y}_m}(w_i)) \text{ for } i = 1, \dots, n. \end{cases} \tag{20}$$

Then for all $\boldsymbol{T} \in \mathrm{Loc}^{m+1}(A, Z, I)$, we have $\hat{\boldsymbol{T}} \in M$, $\pi_1(\hat{\boldsymbol{T}}) = \boldsymbol{T}$, and $\mathbf{y}_m(\boldsymbol{T}) = \mathbf{y}_m(\hat{\boldsymbol{T}})$.

We claim that

$$\text{if } \boldsymbol{T}_1 \in \mathrm{Loc}^{m+1}(A_1, Z_1, I_1) \text{ and } \boldsymbol{T} = \pi_1(\boldsymbol{T}_1), \text{ then } \boldsymbol{T} \in \mathrm{Loc}^{m+1}(A, Z, I) \text{ and } \boldsymbol{T}_1 = \hat{\boldsymbol{T}}. \quad (21)$$

The definition of A_1, Z_1, I_1 easily implies $\pi_1(\mathrm{Loc}^{m+1}(A_1, Z_1, I_1)) \subseteq \mathrm{Loc}^{m+1}(A, Z, I)$. Suppose $\boldsymbol{T}_1 \in \mathrm{Loc}^{m+1}(A_1, Z_1, I_1)$ and let $\boldsymbol{T} = \pi_1(\boldsymbol{T}_1)$. Clearly, we have $\mathbf{y}_m(\boldsymbol{T}) = \mathbf{y}_m(\boldsymbol{T}_1)$ and $f^{\boldsymbol{T}_1}_{\mathbf{y}_m} = f^{\boldsymbol{T}}_{\mathbf{y}_m}$. To prove $\boldsymbol{T}_1 = \hat{\boldsymbol{T}}$, all we need to show is that (20) holds with $\boldsymbol{T}_1$ in place of $\hat{\boldsymbol{T}}$. If $u \notin \mathrm{dom}(\prec^T_{m+1})$, then $\ell^{\boldsymbol{T}_1}(u) \in Z_1 = Z \subseteq \Sigma \cup \{\boldsymbol{x}\}$, so $\ell^{\boldsymbol{T}}(u) = \pi_1(\ell^{\boldsymbol{T}_1}(u)) = \ell^{\boldsymbol{T}_1}(u)$. Suppose $u \in \mathrm{dom}(\prec^T_{m+1})$, $|C^T_m(u)| = n$, and for $i = 1, \dots, n$, $u \blacktriangleleft^{\boldsymbol{T}_1}_{m+1,i} v_i$, $u \lhd^T_{m,i} w_i$. Since $\boldsymbol{T}_1$ is well-labeled, $\boldsymbol{C}^{\boldsymbol{T}_1}_{m+1}(u) \in \mathbb{T}^m_{\Sigma \cup \Upsilon_1}(n)$. Since $(\ell^{\boldsymbol{T}_1}(u), \boldsymbol{C}^{\boldsymbol{T}_1}_{m+1}(u)) \in I_1$, $\ell^{\boldsymbol{T}_1}(u) = (c, d, e_1, \dots, e_n)$ for some $c \in \Upsilon - \Sigma$ and $d, e_1, \dots, e_n \in \Sigma$. By (19), $d = \pi_2(\ell^{\boldsymbol{T}_1}(u)) = \pi_2(\ell^{\boldsymbol{T}_1^\dagger}(u)) = \ell^{\mathbf{y}_m(\boldsymbol{T}_1)}(f^{\boldsymbol{T}_1}_{\mathbf{y}_m}(u)) = \ell^{\mathbf{y}_m(\boldsymbol{T})}(f^{\boldsymbol{T}}_{\mathbf{y}_m}(u))$. By (17), we have $e_i = \pi_2(\ell^{\boldsymbol{T}_1^\dagger}(v_i)) = \pi_2(\ell^{\boldsymbol{T}_1^\dagger}(w_i))$. By (19) again, $\pi_2(\ell^{\boldsymbol{T}_1^\dagger}(w_i)) = \ell^{\mathbf{y}_m(\boldsymbol{T}_1)}(f^{\boldsymbol{T}_1}_{\mathbf{y}_m}(w_i))$, so $e_i = \ell^{\mathbf{y}_m(\boldsymbol{T}_1)}(f^{\boldsymbol{T}_1}_{\mathbf{y}_m}(w_i)) = \ell^{\mathbf{y}_m(\boldsymbol{T})}(f^{\boldsymbol{T}}_{\mathbf{y}_m}(w_i))$. We have shown that (20) holds with $\boldsymbol{T}_1$ in place of $\hat{\boldsymbol{T}}$. This proves (21).

We next prove that the following equivalence holds for all $\boldsymbol{T} \in L = \mathrm{Loc}^{m+1}(A, Z, I)$:

$$\mathbf{y}_m(\boldsymbol{T}) \in L' \text{ if and only if } \hat{\boldsymbol{T}} \in \mathrm{Loc}^{m+1}(A_1, Z_1, I_1). \quad (22)$$

It easily follows from (21) and (22) that $L \cap L' = \mathbf{y}_m(\mathrm{Loc}^{m+1}(A_1, Z_1, I_1))$ and hence $L \cap L'$ is simple context-free.

We prove (22). Let $\boldsymbol{T} \in \mathrm{Loc}^{m+1}(A, Z, I)$ and $\boldsymbol{V} = \mathbf{y}_m(\boldsymbol{T}) = \mathbf{y}_m(\hat{\boldsymbol{T}})$. Since $\hat{\boldsymbol{T}} \in M$, the relabeling $\hat{\boldsymbol{T}}^\dagger = (T, \ell^{\hat{\boldsymbol{T}}^\dagger})$ of $\hat{\boldsymbol{T}}$ given by (14) is defined. From the way $\hat{\boldsymbol{T}}$ is defined, it is clear that for all $u \in T$, we have

$$\pi_2(\ell^{\hat{\boldsymbol{T}}^\dagger}(u)) = \ell^{\mathbf{y}_m(\boldsymbol{T})}(f^{\boldsymbol{T}}_{\mathbf{y}_m}(u)). \quad (23)$$

For the "only if" direction of (22), suppose $\boldsymbol{V} \in L'$. We show that $\hat{\boldsymbol{T}}$ satisfies the requirements L1–L3 for membership in $\mathrm{Loc}^{m+1}(A_1, Z_1, I_1)$.

L1. We have $\ell^{\hat{\boldsymbol{T}}}(\varepsilon) \in A_1$ if and only if either $\varepsilon \notin \mathrm{dom}(\prec^T_{m+1})$ and $\ell^{\boldsymbol{T}}(\varepsilon) = \ell^{\boldsymbol{V}}(\varepsilon) \in A \cap A' \cap Z'$ or $\varepsilon \in \mathrm{dom}(\prec^T_{m+1})$, $\ell^{\boldsymbol{T}}(\varepsilon) \in A$, and $\ell^{\mathbf{y}_m(\boldsymbol{T})}(f^{\boldsymbol{T}}_{\mathbf{y}_m}(\varepsilon)) = \ell^{\boldsymbol{V}}(\varepsilon) \in A'$. Since $\boldsymbol{T} \in \mathrm{Loc}^{m+1}(A, Z, I)$ and $\boldsymbol{V} \in L' = \mathrm{Loc}^m(A', Z', I')$, clearly one of these conditions holds.

L2. Suppose $u \in T - \mathrm{dom}(\prec^T_{m+1})$. Then $\ell^{\hat{\boldsymbol{T}}}(u) = \ell^{\boldsymbol{T}}(u) \in Z = Z_1$.

L3. Suppose $u \in \mathrm{dom}(\prec^T_{m+1})$. Let $(c, d, e_1, \dots, e_n) = \ell^{\hat{\boldsymbol{T}}}(u)$ and $\boldsymbol{W} = \boldsymbol{C}^{\hat{\boldsymbol{T}}}_{m+1}(u)[(c, e_1), \dots, (c, e_n)] = \boldsymbol{C}^{\hat{\boldsymbol{T}}^\dagger}_{m+1}(u)$. We need to show $\boldsymbol{W} \in \mathrm{Loc}^m(A''_d, Z'', I'')$.

- We show $\ell^{\boldsymbol{W}}(\varepsilon) \in A''_d$. By (23), $\pi_2(\ell^{\boldsymbol{W}}(\varepsilon)) = \pi_2(\ell^{\hat{\boldsymbol{T}}^\dagger}(u \cdot (m+1)) = \ell^{\mathbf{y}_m(T)}(f^{\boldsymbol{T}}_{\mathbf{y}_m}(u \cdot (m+1))) = \ell^{\mathbf{y}_m(\boldsymbol{T})}(f^{\boldsymbol{T}}_{\mathbf{y}_m}(u)) = d$. So $\ell^{\boldsymbol{W}}(\varepsilon) \in A''_d$.

– Let $w \in W - \mathrm{dom}(\prec_m^W)$. We show $\ell^{\boldsymbol{W}}(w) \in Z''$. Note that since $w \notin \mathrm{dom}(\prec_m^W)$, we have $u \cdot (m+1) \cdot w \notin \mathrm{dom}(\prec_m^T)$.
  Case 1. $u \cdot (m+1) \cdot w \in \mathrm{dom}(\prec_{m+1}^T)$. Then $\ell^{\hat{\boldsymbol{T}}}(u \cdot (m+1) \cdot w) \neq \boldsymbol{x}$ and $\ell^{\boldsymbol{W}}(w) = \ell^{\hat{\boldsymbol{T}}}(u \cdot (m+1) \cdot w)$. Since $C_m^T(u \cdot (m+1) \cdot w) = \varnothing$, $\ell^{\boldsymbol{W}}(w) = \ell^{\hat{\boldsymbol{T}}}(u \cdot (m+1) \cdot w)$ is of the form (c', d') and hence is in Z''.
  Case 2. $u \cdot (m+1) \cdot w \notin \mathrm{dom}(\prec_{m+1}^T)$. In this case, either $\ell^{\boldsymbol{W}}(w) = \ell^{\hat{\boldsymbol{T}}}(u \cdot (m+1) \cdot w) = \ell^{\boldsymbol{T}}(u \cdot (m+1) \cdot w) = c'$ for some $c' \in \Sigma$ or $\ell^{\hat{\boldsymbol{T}}}(u \cdot (m+1) \cdot w) = \boldsymbol{x}$ and $\ell^{\boldsymbol{W}}(w) = (c, e_i)$ for some i. In the latter case, $\ell^{\boldsymbol{W}}(w) \in Z''$ by the definition of Z''. In the former case, $C_m^T(u \cdot (m+1) \cdot w) = C_m^V(f_{\mathbf{y}_m}^{\boldsymbol{V}}(f \cdot (m+1) \cdot w))$, and since $u \cdot (m+1) \cdot w \notin \mathrm{dom}(\prec_m^T)$, $f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w) \notin \mathrm{dom}(\prec_m^V)$. Since $\boldsymbol{V} \in L'$, $\ell^{\boldsymbol{W}}(w) = \ell^{\boldsymbol{T}}(u \cdot (m+1) \cdot w) = \ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w)) \in Z' \subseteq Z''$.

– Let $w \in \mathrm{dom}(\prec_m^W)$. Then $u \cdot (m+1) \cdot w \in \mathrm{dom}(\prec_m^T)$ and $\ell^{\hat{\boldsymbol{T}}}(u \cdot (m+1) \cdot w) \neq \boldsymbol{x}$, so $\ell^{\boldsymbol{W}}(w) = \ell^{\hat{\boldsymbol{T}}}(u \cdot (m+1) \cdot w)$. We show $(\ell^{\boldsymbol{W}}(w), \boldsymbol{C}_m^{\boldsymbol{W}}(w)) \in I''$. Let $p = |C_m^W(w)|$ and for $i = 1, \dots, p$, let w_i be the i-th node of $C_m^W(w)$.
  Case 1. $u \cdot (m+1) \cdot w \notin \mathrm{dom}(\prec_{m+1}^T)$. Then $\ell^{\boldsymbol{W}}(w) \in \Sigma$ and $\ell^{\boldsymbol{W}}(w) = \ell^{\mathbf{y}_m(\hat{\boldsymbol{T}})}(f_{\mathbf{y}_m}^{\hat{\boldsymbol{T}}}(u \cdot (m+1) \cdot w)) = \ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w))$. Also, $C_m^W(w) = C_m^T(u \cdot (m+1) \cdot w) = C_m^V(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w))$. For $i = 1, \dots, p$, $\pi_2(\ell^{\boldsymbol{W}}(w \cdot m \cdot w_i)) = \pi_2(\ell^{\hat{\boldsymbol{T}}^\dagger}(u \cdot (m+1) \cdot w \cdot m \cdot w_i)) = \ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w \cdot m \cdot w_i))$ by (23). Since $u \cdot (m+1) \cdot w \notin \mathrm{dom}(\prec_{m+1}^T)$, $f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w \cdot m \cdot w_i) = f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w) \cdot m \cdot w_i$. Hence $\pi_2(\boldsymbol{C}_m^{\boldsymbol{W}}(w)) = \boldsymbol{C}_m^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w))$. Since $\boldsymbol{V} \in L' = \mathrm{Loc}^m(A', Z', I')$, $(\ell^{\boldsymbol{W}}(w), \pi_2(\boldsymbol{C}_m^{\boldsymbol{W}}(w))) = (\ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w)), \boldsymbol{C}_m^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w))) \in I'$. Therefore, $(\ell^{\boldsymbol{W}}(w), \boldsymbol{C}_m^{\boldsymbol{W}}(w)) \in I''$.
  Case 2. $u \cdot (m+1) \cdot w \in \mathrm{dom}(\prec_{m+1}^T)$. Then by the definition of $\hat{\boldsymbol{T}}$, $\ell^{\boldsymbol{W}}(w) = \ell^{\hat{\boldsymbol{T}}}(u \cdot (m+1) \cdot w) = (c', d', e_1', \dots, e_p')$, where for $i = 1, \dots, p$, $e_i' = \ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w \cdot m \cdot w_i))$. We have $\ell^{\boldsymbol{C}_m^{\boldsymbol{W}}(w)}(w_i) = \ell^{\boldsymbol{W}}(w \cdot m \cdot w_i) = \ell^{\hat{\boldsymbol{T}}^\dagger}(u \cdot (m+1) \cdot w \cdot m \cdot w_i)$, and by (23), $\pi_2(\ell^{\boldsymbol{C}_m^{\boldsymbol{W}}(w)}(w_i)) = \ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w \cdot m \cdot w_i)) = e_i'$, so $(\ell^{\boldsymbol{W}}(w), \boldsymbol{C}_m^{\boldsymbol{W}}(w)) \in I''$.

For the "if" direction of (22), suppose $\hat{\boldsymbol{T}} \in \mathrm{Loc}^{m+1}(A_1, Z_1, I_1)$. We show that $\mathbf{y}_m(\boldsymbol{T}) = \boldsymbol{V}$ satisfies the requirements L1–L3 for membership in $\mathrm{Loc}^m(A', Z', I')$.

For each $v \in V$, let $h(v)$ be the unique node in $T - \mathrm{dom}(\prec_{m+1}^T)$ such that $\ell^{\boldsymbol{T}}(h(v)) \neq \boldsymbol{x}$ and $f_{\mathbf{y}_m}^{\boldsymbol{T}}(h(v)) = v$. We have $\ell^{\boldsymbol{T}}(h(v)) = \ell^{\hat{\boldsymbol{T}}}(h(v)) = \ell^{\boldsymbol{V}}(v) \in \Sigma$ and $C_m^T(h(v)) = C_m^V(v)$.

L1. Since $\ell^{\hat{\boldsymbol{T}}}(\varepsilon) \in A_1$, either $\varepsilon \notin \mathrm{dom}(\prec_{m+1}^T)$ and $\ell^{\boldsymbol{T}}(\varepsilon) = \ell^{\boldsymbol{V}}(\varepsilon) \in A \cap A' \cap Z'$, or $\varepsilon \in \mathrm{dom}(\prec_{m+1}^T)$, $\ell^{\boldsymbol{T}}(\varepsilon) \in A$, and $\ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(\varepsilon)) = \ell^{\boldsymbol{V}}(\varepsilon) \in A'$. So in either case, we have $\ell^{\boldsymbol{V}}(\varepsilon) \in A'$.

L2. Let $v \in V - \mathrm{dom}(\prec_m^V)$. If $h(v) = \varepsilon$, then $\varepsilon \notin \mathrm{dom}(\prec_{m+1}^T)$ and $\ell^{\hat{\boldsymbol{T}}}(\varepsilon) = \ell^{\boldsymbol{V}}(v) \in A \cap A' \cap Z'$, so $\ell^{\boldsymbol{V}}(v) \in Z'$. If $h(v) \neq \varepsilon$, let $u \in T$ be such that $u \lhd_{m+1}^T h(v) = u \cdot (m+1) \cdot w$. Let $\ell^{\hat{\boldsymbol{T}}}(u) = (c, d, e_1, \dots, e_n)$ and $\boldsymbol{W} = \boldsymbol{C}_{m+1}^{\hat{\boldsymbol{T}}}(u)[(c, e_1), \dots, (c, e_n)]$. Since $\varnothing = C_m^V(v) = C_m^T(h(v)) = C_m^W(w)$, $w \notin \mathrm{dom}(\prec_m^W)$. Since $((c, d, e_1, \dots, e_n), \boldsymbol{W}) \in I_1$, we have $\boldsymbol{W} \in \mathrm{Loc}^m(A_d'', Z'', I'')$, and hence $\ell^{\boldsymbol{V}}(v) = \ell^{\hat{\boldsymbol{T}}}(h(v)) = \ell^{\boldsymbol{W}}(w) \in Z''$. Since $\ell^{\boldsymbol{V}}(v) \in \Sigma$, it follows that $\ell^{\boldsymbol{V}}(v) \in Z'$.

L3. Let $v \in \mathrm{dom}(\prec_m^V)$. Since $C_m^V(v) = C_m^T(h(v)) \neq \varnothing$, $h(v) \neq \varepsilon$. Let $u \in T$ be such that $u \lhd_{m+1}^T = h(v) = u \cdot (m+1) \cdot w$. Let $\ell^{\hat{\boldsymbol{T}}}(u) = (c, d, e_1, \dots, e_n)$. Then $\boldsymbol{W} = \boldsymbol{C}_{m+1}^{\hat{\boldsymbol{T}}}(u)[(c, e_1), \dots, (c, e_n)] \in \mathrm{Loc}^m(A''_d, Z'', I'')$. We have $C_m^V(v) = C_m^T(h(v)) = C_m^W(w)$. Let $p = |C_m^V(v)|$ and for $i = 1, \dots, p$, let v_i be such that $v \lhd_{m,i}^V v \cdot m \cdot v_i$. We must have $(\ell^{\boldsymbol{W}}(w), \boldsymbol{C}_m^{\boldsymbol{W}}(w)) \in I''$, and since $\ell^{\boldsymbol{W}}(w) = \ell^{\boldsymbol{V}}(v) \in \Sigma$, we get $(\ell^{\boldsymbol{W}}(w), \pi_2(\boldsymbol{C}_m^{\boldsymbol{W}}(w))) \in I'$. By (23),

$$\begin{aligned}\pi_2(\ell^{\boldsymbol{W}}(w \cdot m \cdot v_i)) &= \pi_2(\ell^{\hat{\boldsymbol{T}}^\dagger}(u \cdot (m+1) \cdot w \cdot m \cdot v_i)) \\ &= \ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(u \cdot (m+1) \cdot w \cdot m \cdot v_i)) \\ &= \ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(h(v) \cdot m \cdot v_i)) \\ &= \ell^{\boldsymbol{V}}(f_{\mathbf{y}_m}^{\boldsymbol{T}}(h(v)) \cdot m \cdot v_i) \\ &= \ell^{\boldsymbol{V}}(v \cdot m \cdot v_i).\end{aligned}$$

So $(\ell^{\boldsymbol{V}}(v), \boldsymbol{C}_m^{\boldsymbol{V}}(v)) = (\ell^{\boldsymbol{W}}(w), \pi_2(\boldsymbol{C}_m^{\boldsymbol{W}}(w))) \in I'$.

We have established (22). This concludes the proof. □

Clearly, $\boldsymbol{T} \in DT_\Sigma^m$ implies $\mathbf{del}_{m,\widetilde{\Sigma}-\Sigma}(\boldsymbol{T}) \in \mathbb{T}_\Sigma^m$. We obtain the following generalization of the Chomsky-Schützenberger theorem:

Theorem 33. *Let $L \subseteq \mathbb{T}_\Sigma^m$. The following are equivalent:*

(i) *L is simple context-free.*
(ii) *There exist finite alphabets Υ, Υ', a projection $\pi\colon \Upsilon' \to \Upsilon$, and a local set $R \subseteq \mathbb{T}_{\widetilde{\Upsilon'}}^m$ such that $L = \mathbf{del}_{m,\widetilde{\Upsilon}-\Upsilon}(\widetilde{\pi}(R \cap DT_{\Upsilon'}^m))$.*

Proof. (ii) ⇒ (i). Suppose $R \subseteq \mathbb{T}_{\widetilde{\Upsilon'}}^m$ is a local set. Clearly, $R \subseteq \mathbb{T}_{\widetilde{\Upsilon'}_{q,p}}^m$ for some p, q. So $R \cap DT_{\Upsilon'}^m = R \cap DT_{\Upsilon'}^m \cap \mathbb{T}_{\widetilde{\Upsilon'}_{q,p}}^m$. By Lemma 30, $DT_{\Upsilon'}^m \cap \mathbb{T}_{\widetilde{\Upsilon'}_{q,p}}^m$ is simple context-free. It then follows by Lemma 32 that $L = \mathbf{del}_{m,\widetilde{\Upsilon}-\Upsilon}(\widetilde{\pi}(R \cap DT_{\Upsilon'}^m)) = \mathbf{del}_{m,\widetilde{\Upsilon}-\Upsilon}(\widetilde{\pi}(R \cap DT_{\Upsilon'}^m \cap \mathbb{T}_{\widetilde{\Upsilon'}_{q,p}}^m))$ is simple context-free.

(i) ⇒ (ii). Let $K \subseteq \mathbb{T}_{\Upsilon,\boldsymbol{x}}^{m+1}$ be a local set such that $L = \mathbf{y}_m(K)$. By Lemma 29, there exist a projection $\pi\colon \Upsilon' \to \Upsilon$, and a local set $R \subseteq \mathbb{T}_{\widetilde{\Upsilon'}}^m$ such that $\mathbf{enc}_m(K) = \widetilde{\pi}(R \cap DT_{\Upsilon'}^m)$. So

$$\begin{aligned}L &= \mathbf{y}_m(K) \\ &= \mathbf{del}_{m,\widetilde{\Upsilon}-\Upsilon}(\mathbf{enc}_m(K)) \\ &= \mathbf{del}_{m,\widetilde{\Upsilon}-\Upsilon}(\widetilde{\pi}(R \cap DT_{\Upsilon'}^m)).\end{aligned}$$ □

As was the case with the original Chomsky-Schützenberger Theorem, in the proof of Theorem 33, $\mathbf{enc}_m^{-1} \circ \widetilde{\pi}$ is a bijection from $R \cap DT_{\Upsilon'}^m$ to K. (See the second statement in Lemma 29.)

9 A Chomsky-Schützenberger-Weir Representation Theorem for Simple Context-Free Tree Grammars

We are now going to use Theorem 33 to obtain a generalization of Weir's representation theorem about the string languages of tree-adjoining grammars to the string languages of simple context-free tree grammars. First, we prove a lemma that generally holds of m-dimensional Dyck tree languages.

The following lemma is straightforward.

Lemma 34. *Let Σ' be a finite alphabet and $\pi \colon \Sigma' \to \Sigma$ be a projection. If L is a super-local subset of $\mathbb{T}^m_\Sigma$, then $\pi^{-1}(L)$ is a super-local subset of $\mathbb{T}^m_{\Sigma'}$.*

Lemma 35. *Let $m \geq 2$. For any local set $L \subseteq \mathbb{T}^m_{\widetilde{\Sigma}}$, there exist a finite alphabet Σ', a degree-bounded, super-local $L' \subseteq \mathbb{T}^m_{\widetilde{\Sigma'}}$, and a projection $\pi \colon \Sigma' \to \Sigma$ that satisfy the following conditions:*

(i) $\widetilde{\pi}(L') \subseteq L$.
(ii) $L \cap DT^m_\Sigma = \widetilde{\pi}(L' \cap DT^m_{\Sigma'})$. *Moreover, $\widetilde{\pi}$ maps $L' \cap DT^m_{\Sigma'}$ bijectively to $L \cap DT^m_\Sigma$.*

Proof. By Lemma 10, there are a finite alphabet Σ_1, a super-local subset L_1 of $\mathbb{T}^m_{\Sigma_1}$, and a projection $\pi_1 \colon \Sigma_1 \to \widetilde{\Sigma}$ such that π_1 maps L_1 bijectively to L. Since $\widetilde{\pi_1}^{-1}(L \cap DT^m_\Sigma)$ is not a subset of an m-dimensional Dyck tree language, we have to relabel some nodes of $\widetilde{\pi_1}^{-1}(\boldsymbol{T})$ for $\boldsymbol{T} \in L \cap DT^m_\Sigma$ to get a set of the form $L' \cap DT^m_{\Sigma'}$.

For $d \in \Sigma$, P a finite prefix-closed subset of $\mathbb{P}_{m-1}$, and $i \in [0, |P|]$, let

$$\Delta_{d,P,i} = \{\, \delta \in \Sigma_1 \mid \pi_1(\delta) = (d, P, i) \,\}.$$

Define

$$\begin{aligned}
\Sigma_2 &= \Sigma_1 - \bigcup_{d,P,i} \Delta_{d,P,i}, \\
\Delta_{d,P} &= \{\, (\delta_0, \delta_1, \dots, \delta_{|P|}) \mid \delta_i \in \Delta_{d,P,i} \text{ for } i = 1, \dots, |P| \,\}, \\
\Delta &= \bigcup_{d,P} \Delta_{d,P}, \\
\Sigma' &= \Sigma_2 \cup \Delta.
\end{aligned}$$

Note that Σ' is a finite alphabet. Define a projection $\pi \colon \Sigma' \to \Sigma$ by

$$\begin{aligned}
\pi(c) &= \pi_1(c) && \text{if } c \in \Sigma_2, \\
\pi(\boldsymbol{\delta}) &= d && \text{if } \boldsymbol{\delta} \in \Delta_{d,P}.
\end{aligned}$$

Then $\widetilde{\pi}$ maps m-dimensional trees over $\widetilde{\Sigma'}$ to m-dimensional trees over $\widetilde{\Sigma}$. Let

$$\Delta' = \{\, (\boldsymbol{\delta}, P, i) \in \widetilde{\Sigma'} \mid \boldsymbol{\delta} \in \Delta_{d,P}, 0 \leq i \leq |P| \,\},$$

and define a projection $\pi_2 \colon \Sigma_2 \cup \Delta' \to \Sigma_1$ by

$$\begin{aligned} \pi_2(c) &= c && \text{if } c \in \Sigma_2, \\ \pi_2(((\delta_0, \delta_1, \dots, \delta_{|P|}), P, i)) &= \delta_i && \text{if } (\delta_0, \delta_1, \dots, \delta_k) \in \Delta_{d,P} \text{ and } 0 \leq i \leq |P|. \end{aligned}$$

Then for $\boldsymbol{T} \in \mathbb{T}^m_{\Sigma_2 \cup \Delta'}$,

$$\widetilde{\pi}(\boldsymbol{T}) = \pi_1(\pi_2(\boldsymbol{T})).$$

Let

$$L' = \widetilde{\pi}^{-1}(L) \cap \mathbb{T}^m_{\Sigma_2 \cup \Delta'}.$$

Then

$$\begin{aligned} L' &= \pi_2^{-1}(\pi_1^{-1}(L)) \\ &= \pi_2^{-1}(L_1). \end{aligned}$$

By Lemma 34, L' is a super-local subset of $\mathbb{T}^m_{\Sigma_2 \cup \Delta'}$ and hence of $\mathbb{T}^m_{\Sigma'}$.

Clearly, $\widetilde{\pi}(L') \subseteq L$, so (i) holds. Since $\widetilde{\pi}(DT^m_{\Sigma'}) \subseteq DT^m_{\Sigma}$ always holds for any projection $\pi \colon \Sigma' \to \Sigma$, we also have $\widetilde{\pi}(L' \cap DT^m_{\Sigma'}) \subseteq L \cap DT^m_{\Sigma}$.

It remains to show that for each $\boldsymbol{T} \in L \cap DT^m_{\Sigma}$, there is a unique $\boldsymbol{T}' \in L' \cap DT^m_{\Sigma'}$ such that $\widetilde{\pi}(\boldsymbol{T}') = \boldsymbol{T}$.

Let $\boldsymbol{T} \in L \cap DT^m_{\Sigma}$. We relabel the nodes of $\boldsymbol{T}$ to turn it into a $\hat{\boldsymbol{T}} \in DT^m_{\Sigma'}$. Recall that π_1 maps L_1 bijectively to L, so we have $\boldsymbol{T}_1 = \pi_1^{-1}(\boldsymbol{T}) \in L_1$. Let $\boldsymbol{V} = (V, \ell^{\boldsymbol{V}}) = \mathbf{enc}_m^{-1}(\boldsymbol{T})$. Define $\hat{\boldsymbol{V}} = (V, \ell^{\hat{\boldsymbol{V}}})$ by

$$\ell^{\hat{\boldsymbol{V}}}(v) = \begin{cases} \ell^{\boldsymbol{T}_1}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v)) & \text{if } v \in V - \mathrm{dom}(\prec^V_{m+1}) \text{ and } \ell^{\boldsymbol{V}}(v) \neq \boldsymbol{x}, \\ \boldsymbol{x} & \text{if } \ell^{\boldsymbol{V}}(v) = \boldsymbol{x}, \\ (\delta_0, \delta_1, \dots, \delta_k) & \text{if } v \in \mathrm{dom}(\prec^V_{m+1}),\ |C^V_m(v)| = k, \\ & \delta_0 = \ell^{\boldsymbol{T}_1}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v)), \text{ and} \\ & v \blacktriangleleft^{\boldsymbol{V}}_{m+1,i} v_i, \delta_i = \ell^{\boldsymbol{T}_1}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v_i)) \text{ for } i = 1, \dots, k. \end{cases}$$

Let

$$\hat{\boldsymbol{T}} = \mathbf{enc}_m(\hat{\boldsymbol{V}}).$$

If $v \in V - \mathrm{dom}(\prec^V_{m+1})$ and $\ell^{\hat{\boldsymbol{V}}}(v) \neq \boldsymbol{x}$, it is easy to see that $\ell^{\hat{\boldsymbol{V}}}(v) = \ell^{\hat{\boldsymbol{T}}}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v)) \in \Sigma_2 \subseteq \Sigma'$. Let $v \in \mathrm{dom}(\prec^V_{m+1}), P = C^V_m(v), k = |P|$, and $v \blacktriangleleft^{\boldsymbol{V}}_{m+1,i} v_i$ for $i = 1, \dots, k$. Let $(\delta_0, \delta_1, \dots, \delta_k) = \ell^{\hat{\boldsymbol{V}}}(v)$ and $d = \ell^{\boldsymbol{V}}(v)$. Then

$$\pi_1(\delta_0) = \ell^{\boldsymbol{T}}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v)) = (d, P, 0)$$

and for $i = 1, \dots, k$,

$$\pi_1(\delta_i) = \ell^{\boldsymbol{T}}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v_i)) = (d, P, i).$$

This implies that $(\delta_0, \delta_1, \dots, \delta_k) = \ell^{\hat{\boldsymbol{V}}}(v) \in \Delta_{d,P} \subseteq \Delta \subseteq \Sigma'$. We have $\ell^{\hat{\boldsymbol{T}}}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v)) = ((\delta_0, \delta_1, \dots, \delta_k), P, 0) \in \Delta'$ and for $i = 1, \dots, k$, $\ell^{\hat{\boldsymbol{T}}}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v_i)) = ((\delta_0, \delta_1, \dots, \delta_k), P, i) \in \Delta'$. Therefore, $\hat{\boldsymbol{V}} \in \mathbb{T}^{m+1}_{\Sigma', \boldsymbol{x}}$ and $\hat{\boldsymbol{T}} \in DT^m_{\Sigma'} \cap \mathbb{T}^m_{\Sigma_2 \cup \Delta'}$. It is also easy to see that $\pi_2(\hat{\boldsymbol{T}}) = \boldsymbol{T}_1$, so $\hat{\boldsymbol{T}} \in \pi_2^{-1}(L_1) = L'$. We have shown $\hat{\boldsymbol{T}} \in L' \cap DT^m_{\Sigma'}$. Since $\pi_2(\hat{\boldsymbol{T}}) = \boldsymbol{T}_1$, $\widetilde{\pi}(\hat{\boldsymbol{T}}) = \pi_1(\pi_2(\hat{\boldsymbol{T}})) = \pi_1(\boldsymbol{T}_1) = \boldsymbol{T}$.

Now to show uniqueness, suppose $\boldsymbol{T}' \in L' \cap DT^m_{\Sigma'}$ and $\boldsymbol{T} = \widetilde{\pi}(\boldsymbol{T}')$. Let $\boldsymbol{T}_1 = \pi_1^{-1}(\boldsymbol{T})$. Then we have $\boldsymbol{T}_1 = \pi_2(\boldsymbol{T}')$. We prove $\boldsymbol{T}' = \hat{\boldsymbol{T}}$. Let $\boldsymbol{V}' = (V, \ell^{\boldsymbol{V}'}) = \mathbf{enc}_m^{-1}(\boldsymbol{T}')$ and $\boldsymbol{V} = (V, \ell^{\boldsymbol{V}}) = \pi(\boldsymbol{V}')$. Then it is clear that $\mathbf{enc}_m(\boldsymbol{V}) = \boldsymbol{T}$. So it suffices to prove $\boldsymbol{V}' = \hat{\boldsymbol{V}}$. Let $v \in V$. If $\ell^{\boldsymbol{V}}(v) = \boldsymbol{x}$, then clearly, $\ell^{\boldsymbol{V}'}(v) = \boldsymbol{x} = \ell^{\hat{\boldsymbol{V}}}(v)$. If $v \in V - \mathrm{dom}(\prec^V_{m+1})$ and $\ell^{\boldsymbol{V}}(v) \neq \boldsymbol{x}$, then $\ell^{\boldsymbol{V}'}(v) = \ell^{\boldsymbol{T}'}(f^{\boldsymbol{V}'}_{\mathbf{enc}_m}(v)) \in \Sigma_2$, since $\boldsymbol{V}' \in \mathbb{T}^{m+1}_{\Sigma',\boldsymbol{x}}$ and $\boldsymbol{T}' \in L' \subseteq \mathbb{T}^m_{\Sigma_2 \cup \Delta'}$. So $\ell^{\boldsymbol{V}'}(v) = \ell^{\boldsymbol{T}'}(f^{\boldsymbol{V}'}_{\mathbf{enc}_m}(v)) = \pi_2(\ell^{\boldsymbol{T}'}(f^{\boldsymbol{V}'}_{\mathbf{enc}_m}(v))) = \ell^{\boldsymbol{T}_1}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v)) = \ell^{\hat{\boldsymbol{V}}}(v)$. If $v \in \mathrm{dom}(\prec^V_{m+1})$ and $C^V_m(v) = P$, then $\ell^{\boldsymbol{T}'}(f^{\boldsymbol{V}'}_{\mathbf{enc}_m}(v)) = (\ell^{\boldsymbol{V}'}(v), P, 0) \in \Delta'$, so $\ell^{\boldsymbol{V}'}(v) = (\delta_0, \delta_1, \dots, \delta_k)$, where $k = |P|$ and $(\delta_0, \delta_1, \dots, \delta_k) \in \Delta_{d,P}$ for some d. For $i = 1, \dots, k$, let v_i be such that $v \blacktriangleleft^{\boldsymbol{V}'}_{m+1,i} v_i$, or, equivalently, $v \blacktriangleleft^{\boldsymbol{V}}_{m+1,i} v_i$. Then for $i = 1, \dots, k$, $\delta_i = \pi_2((\delta_0, \delta_1, \dots, \delta_k), P, i) = \pi_2(\ell^{\boldsymbol{T}'}(f^{\boldsymbol{V}'}_{\mathbf{enc}_m}(v_i))) = \ell^{\boldsymbol{T}_1}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v_i))$. We also have $\delta_0 = \pi_2((\delta_0, \delta_1, \dots, \delta_k), P, 0) = \pi_2(\ell^{\boldsymbol{T}'}(f^{\boldsymbol{V}'}_{\mathbf{enc}_m}(v))) = \ell^{\boldsymbol{T}_1}(f^{\boldsymbol{V}}_{\mathbf{enc}_m}(v))$. So $\ell^{\hat{\boldsymbol{V}}}(v) = (\delta_0, \delta_1, \dots, \delta_k) = \ell^{\boldsymbol{V}'}(v)$. □

Next we prove a lemma about DT^2_{Υ}. Recall that there was an implicit dependence on the dimension m in the definition of $\widetilde{\Upsilon}$, which is the alphabet of the language DT^m_{Υ}; when a symbol of the form (c, P, i) is in $\widetilde{\Upsilon}$, it is assumed that P is a finite, possibly empty, prefix-closed subset of $\mathbb{P}_{m-1}$. In what follows, we assume that the alphabet $\widetilde{\Upsilon}$ is defined from Υ with respect to dimension $m = 2$, so that $(c, P, i) \in \widetilde{\Upsilon}$ implies $P = \{\varepsilon, 1, \dots, 1^{k-1}\}$ for some $k \geq 0$. We abbreviate (c, P, i) by (c, k, i), where $|P| = k$. Under this convention, $\widetilde{\Upsilon}_q = \Upsilon \cup \{\, (c, k, i) \mid c \in \Upsilon, 0 \leq k \leq q, 0 \leq i \leq k \,\}$.

Recall that for any alphabet Σ, the alphabet Γ_Σ consists of symbols of the form $[_c$ or $]_c$ with $c \in \Sigma$.

We let $\mathbb{T}_{\Upsilon, \widetilde{\Upsilon} - \Upsilon}$ stand for the set of trees $\boldsymbol{T} \in \mathbb{T}_{\widetilde{\Upsilon}}$ such that for every $v \in T$, $\ell^{\boldsymbol{T}}(v) \in \widetilde{\Upsilon} - \Upsilon$ implies $|C^T_2(v)| = 1$. (In other words, $\widetilde{\Upsilon} - \Upsilon$ is regarded as a ranked alphabet all of whose symbols have rank 1.)

Lemma 36. *Let $\eta\colon \Gamma^*_{\widetilde{\Upsilon}} \to \Gamma^*_{\widetilde{\Upsilon}}$ be the alphabetic homomorphism defined as follows:*

$$\begin{aligned}
\eta([_c) &= \varepsilon, \\
\eta(]_c) &= \varepsilon \qquad \textit{for } c \in \Upsilon, \\
\eta([_{(c,k,0)}) &= [_{(c,k,0)}, \\
\eta(]_{(c,k,0)}) &=]_{(c,k,k)}, \\
\eta([_{(c,k,i)}) &=]_{(c,k,i-1)}, \\
\eta(]_{(c,k,i)}) &= [_{(c,k,i)} \qquad \textit{for } 1 \leq i \leq k.
\end{aligned}$$

Then

$$DT^2_{\Upsilon} = \mathbb{T}_{\Upsilon, \widetilde{\Upsilon} - \Upsilon} \cap \mathbf{enc}^{-1}(\eta^{-1}(D_{\widetilde{\Upsilon}})).$$

(Here, **enc** *is the standard encoding function defined on ordinary 2-dimensional trees.)*

Proof. ($\subseteq$). Suppose $\boldsymbol{T} \in DT^2_{\Upsilon}$. Clearly, $\boldsymbol{T} \in \mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon}$, so it suffices to prove $\eta(\mathbf{enc}(\boldsymbol{T})) \in D_{\widetilde{\Upsilon}}$. This is proved by induction on the length of the reduction $\boldsymbol{T} \rightsquigarrow^* \boldsymbol{T}' \in \mathbb{T}_{\Upsilon}$. If $\boldsymbol{T} \in \mathbb{T}_{\Upsilon}$, then clearly, $\eta(\mathbf{enc}(\boldsymbol{T})) = \varepsilon \in D_{\widetilde{\Upsilon}}$. Suppose $\boldsymbol{T} \rightsquigarrow \boldsymbol{T}'' \rightsquigarrow^* \boldsymbol{T}' \in \mathbb{T}_{\Upsilon}$. Then $\boldsymbol{T} = \boldsymbol{U}[(c,k,0) -_2 \boldsymbol{T}_0[(c,k,1) -_2 \boldsymbol{T}_1, \dots, (c,k,k) -_2 \boldsymbol{T}_k]]$ and $\boldsymbol{T}' = \boldsymbol{U}[\boldsymbol{T}_0[\boldsymbol{T}_1, \dots, \boldsymbol{T}_k]]$ for some $c \in \Upsilon$, $k \geq 0$, $\boldsymbol{U} \in \mathbb{T}_{\widetilde{\Upsilon}}(1)$, $\boldsymbol{T}_0 \in \mathbb{T}_{\Upsilon}(k)$, and $\boldsymbol{T}_i \in \mathbb{T}_{\widetilde{\Upsilon}}$ $(i = 1, \dots, k)$. Then $\eta(\mathbf{enc}(\boldsymbol{T})) = z_1 \,[_{(c,k,0)}\,]_{(c,k,0)}\, y_1 \,[_{(c,k,1)}\,]_{(c,k,1)}\, y_2 \dots y_k \,[_{(c,k,k)}\,]_{(c,k,k)}\, z_2$ and $\eta(\mathbf{enc}(\boldsymbol{T}'')) = z_1 y_1 y_2 \dots y_k z_2$ for some $z_1, z_2, y_1, y_2, \dots, y_k \in (\Gamma_{\widetilde{\Upsilon}-\Upsilon})^*$. By induction hypothesis, $\eta(\mathbf{enc}(\boldsymbol{T}'')) \in D_{\widetilde{\Upsilon}}$, and this easily implies $\eta(\mathbf{enc}(\boldsymbol{T})) \in D_{\widetilde{\Upsilon}}$.

($\supseteq$). Suppose $\boldsymbol{T} \in \mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon}$ and $\eta(\mathbf{enc}(\boldsymbol{T})) \in D_{\widetilde{\Upsilon}}$. If $\boldsymbol{T} \in \mathbb{T}_{\Upsilon}$, then $\boldsymbol{T} \in DT^2_{\Upsilon}$. Suppose that $\boldsymbol{T} \notin \mathbb{T}_{\Upsilon}$. First, we claim that $\boldsymbol{T}$ must have a node labeled by $(c,k,0)$ for some $c \in \Upsilon$ and $k \geq 0$. For, if $\boldsymbol{T}$ has a node v such that $\ell^{\boldsymbol{T}}(v) = (c,k,i)$ for some $c \in \Upsilon$, $k \geq 1$, and $i \geq 1$, then $\mathbf{enc}(\boldsymbol{T}) = x_1 \,[_{(c,k,i)}\, x_2 \,]_{(c,k,i)}\, x_3$ and $\eta(\mathbf{enc}(\boldsymbol{T})) = \eta(x_1) \,]_{(c,k,i-1)}\, \eta(x_2) \,[_{(c,k,i)}\, \eta(x_3)$ for some $x_1, x_2, x_3 \in \Gamma^*_{\widetilde{\Upsilon}}$. Since $\eta(\mathbf{enc}(\boldsymbol{T})) \in D_{\widetilde{\Upsilon}}$, $[_{(c,k,i-1)}$ must occur in $\eta(x_1)$. If $i - 1 = 0$, this implies that $[_{(c,k,0)}$ occurs in x_1 and it follows that $\boldsymbol{T}$ has a node labeled by $(c,k,0)$. Otherwise, $]_{(c,k,i-1)}$ occurs in x_1, and it follows that $\boldsymbol{T}$ has a node labeled by $(c,k,i-1)$. Repeating this reasoning, we see that $\boldsymbol{T}$ must have a node labeled by $(c,k,0)$.

We show that $\boldsymbol{T} \in DT^2_{\Upsilon}$ by induction on the number of nodes of $\boldsymbol{T}$ that are labeled by a symbol of the form $(c,k,0)$. Let v be a node of $\boldsymbol{T}$ labeled by $(c,k,0)$ such that no node v' with $v \,(\lhd_2^T)^+\, v'$ is labeled by a symbol of the form $(d,l,0)$. Then $\mathbf{enc}(\boldsymbol{T}) = x_1 \,[_{(c,k,0)}\, y \,]_{(c,k,0)}\, x_2$ for some $x_1, x_2 \in \Gamma^*_{\widetilde{\Upsilon}}$ and $y \in \mathbf{enc}(\mathbb{T}^2_{\Upsilon,\widetilde{\Upsilon}-\Upsilon})$. (Note that $|C^T_2(v)| = 1$.)

Case 1. $k = 0$. We show $y \in D'_{\Upsilon}$, i.e., the subtree $\boldsymbol{T}_0$ of $\boldsymbol{T}$ rooted at $v \cdot 2$ is in $\mathbb{T}_{\Upsilon}$. Suppose otherwise, and take the alphabetically first node of $\boldsymbol{T}_0$ labeled by some $(d,l,j) \in \widetilde{\Upsilon} - \Upsilon$. Then $y = y' \,[_{(d,l,j)}\, y''$ for some $y' \in \Gamma^*_{\Upsilon}$ and $y'' \in \Gamma^*_{\widetilde{\Upsilon}}$. By our assumption about v, $j \geq 1$ and $l \geq 1$. We have $\eta(\mathbf{enc}(\boldsymbol{T})) = \eta(x_1) \,[_{(c,0,0)}\, \eta(y) \,]_{(c,0,0)}\, \eta(x_2) = \eta(x_1) \,[_{(c,0,0)}\,]_{(d,l,j-1)}\, \eta(y'') \,]_{(c,0,0)}\, \eta(x_2)$, which contradicts the assumption that $\eta(\mathbf{enc}(\boldsymbol{T})) \in D_{\widetilde{\Upsilon}}$. So $\boldsymbol{T}_0$ is in $\mathbb{T}_{\Upsilon}$, and we can write $\boldsymbol{T} = \boldsymbol{U}[(c,0,0) -_2 \boldsymbol{T}_0]$. So $\boldsymbol{T} \rightsquigarrow \boldsymbol{U}[\boldsymbol{T}_0] \in \mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon}$. We have $\eta(\mathbf{enc}(\boldsymbol{T})) = \eta(x_1) \,[_{(c,0,0)}\,]_{(c,0,0)}\, \eta(x_2)$ and $\eta(\mathbf{enc}(\boldsymbol{U}[\boldsymbol{T}_0])) = \eta(x_1)\eta(x_2)$. Since $\eta(\mathbf{enc}(\boldsymbol{T})) \in D_{\widetilde{\Upsilon}}$, $\eta(x_1)\eta(x_2) \in D_{\widetilde{\Upsilon}}$ as well, and $\boldsymbol{U}[\boldsymbol{T}_0] \in DT^2_{\Upsilon}$ by induction hypothesis. Since $\boldsymbol{T} \rightsquigarrow \boldsymbol{U}[\boldsymbol{T}_0]$, we conclude $\boldsymbol{T} \in DT^2_{\Upsilon}$.

Case 2. $k \geq 1$. We show

$$y = z_0 \,[_{(c,k,1)}\, y_1 \,]_{(c,k,1)}\, z_1 \,[_{(c,k,2)}\, y_2 \,]_{(c,k,2)} \dots z_{k-1} \,[_{(c,k,k)}\, y_k \,]_{(c,k,k)}\, z_k$$

for some $z_0, z_1, \dots, z_k \in \Gamma^*_{\Upsilon}$ and $y_1, y_2, \dots, y_k \in \mathbf{enc}(\mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon})$. First, we show by induction that the following condition holds for $i = 0, \dots, k$:

$$y \text{ has a prefix of the form } z_0 \,[_{(c,k,1)}\, y_1 \,]_{(c,k,1)} \dots z_{i-1} \,[_{(c,k,i)}\, y_i \,]_{(c,k,i)}. \quad (24)$$

The case of $i = 0$ is trivial. Suppose we have shown (24) for $i < k$, i.e., $y = z_0 \,[_{(c,k,1)}\, y_1 \,]_{(c,k,1)} \dots z_{i-1} \,[_{(c,k,i)}\, y_i \,]_{(c,k,i)}\, y'$ with $z_0, \dots, z_{i-1} \in \Gamma^*_{\Upsilon}$ and

$y_1, \dots, y_i \in \mathbf{enc}(\mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon})$. Since

$$\begin{aligned}\eta(\mathbf{enc}(\boldsymbol{T})) &= \eta(x_1)\ [_{(c,k,0)}\ \eta(y)\ ]_{(c,k,k)}\ \eta(x_2)\\ &= \eta(x_1)\ [_{(c,k,0)}\ ]_{(c,k,0)}\ \eta(y_1)\ [_{(c,k,1)} \dots]_{(c,k,i-1)}\ \eta(y_i)\ [_{(c,k,i)}\ \eta(y')\ ]_{(c,k,k)},\end{aligned}$$

we must have $\eta(y') \neq \varepsilon$. Let z_i be the longest prefix of y' in Γ_Υ^*. Since $z_0, \dots, z_{i-1} \in \Gamma_\Upsilon^*$ and $y_1, \dots, y_i$ and y are all in $\mathbf{enc}(\mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon})$, it is easy to see that $y' = z_i\ [_{(d,l,j)}\ y_{i+1}\ ]_{(d,l,j)}\ y''$ for some $y_{i+1} \in \mathbf{enc}(\mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon})$ and $l, j \geq 1$. We have $\eta(y') =]_{(d,l,j-1)}\ \eta(y_{i+1})\ [_{(d,l,j)}\ \eta(y'')$, and so

$$\begin{aligned}&\eta(\mathbf{enc}(\boldsymbol{T})) =\\ &\eta(x_1)\ [_{(c,k,0)}\ ]_{(c,k,0)}\ \eta(y_1)\ [_{(c,k,1)} \dots]_{(c,k,i-1)}\ \eta(y_i)\ [_{(c,k,i)}\ ]_{(d,l,j-1)}\ \eta(y_{i+1})\ [_{(d,l,j)}\ \eta(y'')\ ]_{(c,k,k)},\end{aligned}$$

which implies $d = c$, $l = k$, and $j = i + 1$. This shows that (24) holds with $i + 1$ in place of i. By induction, (24) holds with $i = k$.

We have

$$y = z_0\ [_{(c,k,1)}\ y_1\ ]_{(c,k,1)} \dots z_{k-1}\ [_{(c,k,k)}\ y_k\ ]_{(c,k,k)}\ y'$$

with $z_0, \dots, z_{k-1} \in \Gamma_\Upsilon^*$ and $y_1, \dots, y_k \in \mathbf{enc}(\mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon})$. We show that $y' \in \Gamma_\Upsilon^*$. Suppose otherwise. Then we must have $y' = z_k\ [_{(d,l,j)}\ y_{k+1}\ ]_{(d,l,j)}\ y''$ for some $d \in \Upsilon$, $j, l \geq 1$, $z_k \in \Gamma_\Upsilon^*$, and $y_{k+1} \in \mathbf{enc}(\mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon})$. Then $\eta(\mathbf{enc}(\boldsymbol{T}))$ contains as a substring $[_{(c,k,k)}\ ]_{(d,l,j-1)}$, which is a contradiction since $\eta(\mathbf{enc}(\boldsymbol{T})) \in D_{\widetilde{\Upsilon}}$ and $j - 1 < l$. Therefore,

$$y = z_0\ [_{(c,k,1)}\ y_1\ ]_{(c,k,1)} \dots z_{k-1}\ [_{(c,k,k)}\ y_k\ ]_{(c,k,k)}\ z_k$$

with $z_0, \dots, z_k \in \Gamma_\Upsilon^*$ and $y_1, \dots, y_k \in \mathbf{enc}(\mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon})$. This means that $\boldsymbol{T}$ is of the form

$$\boldsymbol{T} = \boldsymbol{U}[(c,k,0) -_2 \boldsymbol{T}_0[(c,k,1) -_2 \boldsymbol{T}_1, \dots, (c,k,k) -_2 \boldsymbol{T}_k]],$$

where $\boldsymbol{T}_0 \in \mathbb{T}_\Upsilon(k)$. So

$$\boldsymbol{T} \rightsquigarrow \boldsymbol{T}' = \boldsymbol{U}[\boldsymbol{T}_0[\boldsymbol{T}_1, \dots, \boldsymbol{T}_k]].$$

Clearly, $\boldsymbol{T}' \in \mathbb{T}_{\Upsilon,\widetilde{\Upsilon}-\Upsilon}$, and

$$\eta(\mathbf{enc}(\boldsymbol{T}')) = \eta(x_1)\eta(y_1) \dots \eta(y_k)\eta(x_2).$$

Since

$$\begin{aligned}&\eta(\mathbf{enc}(\boldsymbol{T})) =\\ &\eta(x_1)\ [_{(c,k,0)}\ ]_{(c,k,0)}\ \eta(y_1)\ [_{(c,k,1)}\ ]_{(c,k,1)}\ \eta(y_2) \dots [_{(c,k,k-1)}\ ]_{(c,k,k-1)}\ \eta(y_k)\ [_{(c,k,k)}\ ]_{(c,k,k)}\ \eta(x_2)\end{aligned}$$

is in $D_{\widetilde{\Upsilon}}$, it follows that $\eta(\mathbf{enc}(\boldsymbol{T}'))$ is in $D_{\widetilde{\Upsilon}}$ as well, and the induction hypothesis gives $\boldsymbol{T}' \in DT_\Upsilon^2$. Since $\boldsymbol{T} \rightsquigarrow \boldsymbol{T}'$, we conclude $\boldsymbol{T} \in DT_\Upsilon^2$. □

Lemma 37. *If $L \subseteq \mathbb{T}^3_{\Sigma,\boldsymbol{x}}$ is a local set, then there exist a finite alphabet Υ, a projection $\pi\colon \Upsilon \to \Sigma$, and a local set $R \subseteq \Gamma^+_{\widetilde{\Upsilon}}$ such that*

$$\mathbf{enc}(\mathbf{enc}_2(L)) = \widehat{\widetilde{\pi}}(R \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}})),$$

where η is the alphabetic homomorphism defined in Lemma 36. Moreover, $\mathbf{enc}_2^{-1} \circ \mathbf{enc}^{-1} \circ \widehat{\widetilde{\pi}}$ maps $R \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}})$ bijectively to L.

Proof. By Lemma 29, there exist a finite alphabet Υ_1, a projection $\pi_1\colon \Upsilon_1 \to \Sigma$, and a local set $L_1 \subseteq \mathbb{T}^2_{\widetilde{\Upsilon_1}}$ such that $\mathbf{enc}_2(L) = \widetilde{\pi_1}(L_1 \cap DT^2_{\Upsilon_1})$ and $\widetilde{\pi_1}$ is a bijection from $L_1 \cap DT^2_{\Upsilon_1}$ to $\mathbf{enc}_2(L)$. We may assume $L_1 \subseteq \mathbb{T}_{\Upsilon_1, \widetilde{\Upsilon_1} - \Upsilon_1}$. By Lemma 35, there exist a finite alphabet Υ_2, a projection $\pi_2\colon \Upsilon_2 \to \Upsilon_1$, and a super-local set $L_2 \subseteq \mathbb{T}^2_{\widetilde{\Upsilon_2}}$ such that $\widetilde{\pi_2}(L_2) \subseteq L_1$ and $\widetilde{\pi_2}(L_2 \cap DT^2_{\Upsilon_2}) = L_1 \cap DT^2_{\Upsilon_1}$. Since $L_1 \subseteq \mathbb{T}_{\Upsilon_1, \widetilde{\Upsilon_1} - \Upsilon_1}$, it follows that $L_2 \subseteq \mathbb{T}_{\Upsilon_2, \widetilde{\Upsilon_2} - \Upsilon_2}$. We have

$$\begin{aligned}
\mathbf{enc}(\mathbf{enc}_2(L)) &= \mathbf{enc}(\widetilde{\pi_1}(L_1 \cap DT^2_{\Upsilon_1})) \\
&= \mathbf{enc}(\widetilde{\pi_1}(\widetilde{\pi_2}(L_2 \cap DT^2_{\Upsilon_2}))) \\
&= \widehat{\widetilde{\pi_1}}(\widehat{\widetilde{\pi_2}}(\mathbf{enc}(L_2 \cap DT^2_{\Upsilon_2}))) \\
&= \widehat{\widetilde{\pi_1}}(\widehat{\widetilde{\pi_2}}(\mathbf{enc}(L_2) \cap \mathbf{enc}(DT^2_{\Upsilon_2}))),
\end{aligned} \tag{25}$$

since $\mathbf{enc}$ is injective. By Lemma 36,

$$\mathbf{enc}(DT^2_{\Upsilon_2}) = \mathbf{enc}(\mathbb{T}_{\Upsilon_2, \widetilde{\Upsilon_2} - \Upsilon_2}) \cap \eta_2^{-1}(D_{\widetilde{\Upsilon_2}}),$$

where $\eta_2\colon \Gamma^*_{\widetilde{\Upsilon_2}} \to \Gamma^*_{\widetilde{\Upsilon_2}}$ is an alphabetic homomorphism defined like η. Since $L_2 \subseteq \mathbb{T}_{\Upsilon_2, \widetilde{\Upsilon_2} - \Upsilon_2}$,

$$\mathbf{enc}(L_2) \cap \mathbf{enc}(DT^2_{\Upsilon_2}) = \mathbf{enc}(L_2) \cap \eta_2^{-1}(D_{\widetilde{\Upsilon_2}}).$$

By Lemma 2, $\mathbf{enc}(L_2) = R_2 \cap D'_{\widetilde{\Upsilon_2}}$ for some local set $R_2 \subseteq \Gamma^+_{\widetilde{\Upsilon_2}}$. So we have

$$\mathbf{enc}(L_2) \cap \mathbf{enc}(DT^2_{\Upsilon_2}) = R_2 \cap D'_{\widetilde{\Upsilon_2}} \cap \eta_2^{-1}(D_{\widetilde{\Upsilon_2}}). \tag{26}$$

Given (25) and (26), all we need is to turn $R_2 \cap D'_{\widetilde{\Upsilon_2}} \cap \eta_2^{-1}(D_{\widetilde{\Upsilon_2}})$ into the form $\widehat{\widetilde{\pi_3}}(R \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}}))$. For this, we can use a method similar to the one we used in the proof of Lemma 6. Let

$$\Upsilon = \Upsilon_2 \cup \{\, \bar{c} \mid c \in \Upsilon_2 \,\}$$

and define $\pi_3\colon \Upsilon \to \Upsilon_2$ by

$$\pi_3(c) = c, \quad \pi_3(\bar{c}) = c,$$

for each $c \in \Upsilon_2$. Let

$$\begin{aligned}
\Delta_1 &= \{\, \bar{c} \mid c \in \Upsilon_2 \,\} \cup \{\, (\bar{c}, P, 0) \mid (c, P, 0) \in \widetilde{\Upsilon_2} \,\}, \\
\Delta_2 &= \widetilde{\Upsilon_2} \cup \{\, (\bar{c}, P, i) \mid (c, P, i) \in \widetilde{\Upsilon_2}, i \geq 1 \,\}.
\end{aligned}$$

Then Δ_1, Δ_2 is a partition of $\widetilde{\Upsilon}$. Let

$$R = (\{\, [_d \mid d \in \Delta_1 \,\}\, \Gamma^*_{\Delta_2}\, \{\,]_d \mid d \in \Delta_1 \,\}) \cap \widehat{\widehat{\pi_3}}^{-1}(R_2).$$

Then R is a local subset of $\Gamma^+_{\widetilde{\Upsilon}}$,[22] and it is easy to see

$$\widehat{\widehat{\pi_3}}(R \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}})) = R_2 \cap D'_{\widetilde{\Upsilon}_2} \cap \eta_2^{-1}(D_{\widetilde{\Upsilon}_2}), \tag{27}$$

where η^{-1} is as defined in Lemma 36. It is also easy to see that $\widehat{\widehat{\pi_3}}$ maps $R \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}})$ bijectively to $R_2 \cap D'_{\widetilde{\Upsilon}_2} \cap \eta_2^{-1}(D_{\widetilde{\Upsilon}_2})$.

We obtain the statement of the lemma from (25), (26), and (27) by letting $\pi = \pi_3 \circ \pi_2 \circ \pi_1$. □

Recall that when Σ_0 and Σ_1 are disjoint alphabets, $\mathbb{T}^{\Sigma_1}_{\Sigma_0}$ consists of all (ordinary 2-dimensional) trees in $\mathbb{T}_{\Sigma_0 \cup \Sigma_1}$ that are disjointly labeled with Σ_0, Σ_1.

Lemma 38. *If $L \subseteq \mathbb{T}^3_{\Sigma,\boldsymbol{x}}$ is a local set, then there exist an alphabet Σ' disjoint from Σ, a projection $\pi\colon \Sigma \cup \Sigma' \to \Sigma$, and a local set $L' \subseteq \mathbb{T}^3_{\Sigma \cup \Sigma',\boldsymbol{x}}$ that satisfy the following conditions:*

(i) $\pi(c) = c$ *for all* $c \in \Sigma$.
(ii) $L = \pi(L')$. *Moreover,* π *maps* L' *bijectively to* L.
(iii) $\mathbf{y}_2(L') \subseteq \mathbb{T}^{\Sigma'}_{\Sigma}$.
(iv) $\mathbf{y}_2(L) = \pi(\mathbf{y}_2(L'))$. *Moreover,* π *maps* $\mathbf{y}_2(L')$ *bijectively to* $\mathbf{y}_2(L)$.
(v) $\mathbf{y}(\mathbf{y}_2(L)) = \mathbf{y}(\mathbf{y}_2(L'))$.

Proof. Let $L = \mathrm{Loc}^3(A, Z, I)$ be a local subset of $\mathbb{T}^3_{\Sigma,\boldsymbol{x}}$. Let $\Sigma' = \{\, \bar{c} \mid c \in \Sigma \,\}$. Define a projection $\pi\colon \Sigma \cup \Sigma' \to \Sigma$ by

$$\pi(c) = c, \quad \pi(\bar{c}) = c,$$

for each $c \in \Sigma$. Let

$$Z' = Z \cup \{\, \bar{c} \mid c \in Z \,\}$$
$$I' = \{\, (c, \boldsymbol{T}') \mid (c, \boldsymbol{T}) \in I, \boldsymbol{T} \in \mathbb{T}^2_{\Sigma}(0) \,\} \cup \{\, (\bar{c}, \boldsymbol{T}') \mid (c, \boldsymbol{T}) \in I, \boldsymbol{T} \in \mathbb{T}^2_{\Sigma}(n), n \geq 1 \,\},$$

where for $\boldsymbol{T} = (T, \ell^{\boldsymbol{T}}) \in \mathbb{T}^2_{\Sigma \cup \{\boldsymbol{x}\}}$, $\boldsymbol{T}' = (T, \ell^{\boldsymbol{T}'})$ is defined by

$$\ell^{\boldsymbol{T}'}(v) = \begin{cases} \bar{c} & \text{if } \ell^{\boldsymbol{T}}(v) = c \in \Sigma \text{ and } v \in \mathrm{dom}(\prec^T_2), \\ \ell^{\boldsymbol{T}}(v) & \text{otherwise.} \end{cases}$$

Note that $\boldsymbol{T}' \in \mathbb{T}^2_{\Sigma \cup \Sigma' \cup \{\boldsymbol{x}\}}$. Define a local subset L' of $\mathbb{T}^3_{\Sigma \cup \Sigma' \cup \{\boldsymbol{x}\}}$ by $L' = \mathrm{Loc}^3(A, Z', I')$. Then it is easy to see that π and L' satisfy the required properties. □

[22] Although $\Gamma_{\widetilde{\Upsilon}}$ is an infinite alphabet, only finitely many symbols in it appear in $\widehat{\widehat{\pi_3}}^{-1}(R_2)$ since R_2 is local.

Lemma 39. *If* $L \subseteq \mathbb{T}^3_{\Sigma,\boldsymbol{x}}$ *is a local set, then there exist a finite alphabet* Υ*, a local set* $R \subseteq \Gamma^+_{\widetilde{\Upsilon}}$*, and an alphabetic homomorphism* $h\colon \Gamma^*_{\widetilde{\Upsilon}} \to \Sigma^*$ *such that*

$$\mathbf{y}(\mathbf{y}_2(L)) = h(R \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}})),$$

where $\widetilde{\Upsilon}$ *is defined with respect to dimension 2 and* η *is the alphabetic homomorphism defined in Lemma 36.*

Proof. Let $L \subseteq \mathbb{T}^3_{\Sigma,\boldsymbol{x}}$ be a local set. Applying Lemma 38 to L, we obtain a projection $\pi'\colon \Sigma \cup \Sigma' \to \Sigma$ and a local set $L' \subseteq \mathbb{T}^3_{\Sigma\cup\Sigma',\boldsymbol{x}}$ such that π' maps L' bijectively to L, $\mathbf{y}_2(L') \subseteq \mathbb{T}^{\Sigma'}_{\Sigma}$, and $\mathbf{y}(\mathbf{y}_2(L)) = \mathbf{y}(\mathbf{y}_2(L'))$. By Lemma 37,

$$\mathbf{enc}(\mathbf{enc}_2(L')) = \widehat{\widehat{\pi}}(R \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}}))$$

for some finite set Υ, projection $\pi\colon \Upsilon \to \Sigma \cup \Sigma'$, and local set $R \subseteq \Gamma^+_{\widetilde{\Upsilon}}$. We have

$$\begin{aligned}\mathbf{y}(\mathbf{y}_2(L)) &= \mathbf{y}(\mathbf{y}_2(L'))\\ &= h_{\Sigma_0,\Sigma_1}(\mathbf{enc}(\mathbf{del}_{2,\widetilde{\Upsilon}-\Upsilon}(\mathbf{enc}_2(L')))))\\ &= h_{\Sigma_0,\Sigma_1}(h_{\Gamma_\Upsilon,\Gamma_{\widetilde{\Upsilon}-\Upsilon}}(\mathbf{enc}(\mathbf{enc}_2(L'))))\\ &= h_{\Sigma_0,\Sigma_1}(h_{\Gamma_\Upsilon,\Gamma_{\widetilde{\Upsilon}-\Upsilon}}(\widehat{\widehat{\pi}}(R \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}})))),\end{aligned}$$

so the statement of the lemma holds with $h = h_{\Sigma_0,\Sigma_1} \circ h_{\Gamma_\Upsilon,\Gamma_{\widetilde{\Upsilon}-\Upsilon}} \circ \widehat{\widehat{\pi}}$. □

Note that in the above proof, the set $R \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}})$ is mapped bijectively to L by $\pi' \circ \mathbf{enc}_2^{-1} \circ \mathbf{enc}^{-1} \circ \widehat{\widehat{\pi}}$.

The following lemma is analogous to the corresponding characterization of context-free (string) languages.

Lemma 40. *Let* $L \subseteq \mathbb{T}_\Sigma$*. Then* $L \in \mathrm{CFT}_{\mathrm{sp}}(r)$ *if and only if there exist a finite alphabet* Υ *and a local set* $K \subseteq \mathbb{T}^3_{\Upsilon,\boldsymbol{x}}$ *such that*

(i) $L = \mathbf{y}_2(K)$*, and*
(ii) *for all* $\boldsymbol{T} \in K$ *and all* $v \in T$*, if* $v \in \mathrm{dom}(\prec_3^T)$*, then* $|C_2^T(v)| \le r$*.*

Lemma 41. *For every* $L \in \mathrm{CFT}_{\mathrm{sp}}(r)$*, there is an* $L' \in \mathrm{CFT}_{\mathrm{sp}}(r)$ *such that* $\mathbf{enc}(L) = \mathbf{y}(L')$*.*

Proof. Let $K \subseteq \mathbb{T}^3_{\Upsilon,\boldsymbol{x}}$ be a local set satisfying condition (ii) of Lemma 40. By Lemma 38, we may assume $K = \mathrm{Loc}^3(A, Z, I)$ with $Z \cap \{\, c \mid (c, \boldsymbol{T}) \in I \,\} = \varnothing$. Let

$$\begin{aligned}A' &= A \cup \{\, \bar{c} \mid c \in A \cap Z \,\},\\ Z' &= Z \cup \bigcup \{\, \{ [_c,]_c \} \mid c \in Z \,\},\\ I' &= \{\, (c, \varphi(\boldsymbol{T})) \mid (c, \boldsymbol{T}) \in I \,\} \cup \{\, (\bar{c}, c([_c\,]_c)) \mid c \in A \cap Z \,\}.\end{aligned}$$

where for each $c \in A \cap Z$, $\bar{c}$ is a new symbol, and

$$\varphi(\boldsymbol{x}) = \boldsymbol{x},$$
$$\varphi(c) = \begin{cases} c(\llbracket_c \rrbracket_c) & \text{if } c \in Z, \\ c & \text{otherwise,} \end{cases}$$
$$\varphi(c(\boldsymbol{T}_1 \dots \boldsymbol{T}_n)) = \begin{cases} c(\llbracket_c \varphi(\boldsymbol{T}_1) \dots \varphi(\boldsymbol{T}_n) \rrbracket_c) & \text{if } c \in Z, \\ c(\varphi(\boldsymbol{T}_1) \dots \varphi(\boldsymbol{T}_n)) & \text{otherwise.} \end{cases}$$

Then it is easy to see $K' = \mathrm{Loc}^3(A, Z', I')$ also satisfies condition (ii) of Lemma 40, and we have $\mathbf{enc}(\mathbf{y}_2(K)) = \mathbf{y}(\mathbf{y}_2(K'))$. We omit the details. □

Recall that $\Gamma_n = \{ \llbracket_1, \rrbracket_1, \dots, \llbracket_n, \rrbracket_n \}$ and D_n is the Dyck language over Γ_n, where $\llbracket_i$ and $\rrbracket_i$ form a matching pair of brackets for $i = 1, \dots, n$.

Theorem 42. *Let $q \geq 1$ and $M \subseteq \Sigma^*$. The following are equivalent:*

(i) $M \in \mathrm{yCFT}_{\mathrm{sp}}(q-1)$.
(ii) *There exist a positive integer n, a local set $R \subseteq \Gamma_{qn}^*$, and an alphabetic homomorphism $h \colon \Gamma_{qn}^* \to \Sigma^*$ such that $M = h(R \cap D_{qn} \cap g^{-1}(D_{qn}))$, where g is the bijection on Γ_{qn} defined by*

$$g(\llbracket_{qi+1}) = \llbracket_{qi+1}, \qquad g(\rrbracket_{qi+1}) = \rrbracket_{qi+q},$$
$$g(\llbracket_{qi+j}) = \rrbracket_{qi+j-1}, \qquad g(\rrbracket_{qi+j}) = \llbracket_{qi+j},$$

for $i = 0, \dots, n-1$ and $j = 2, \dots, q$.

Proof. (ii) ⇒ (i). By Lemma 41 and the fact that $\mathrm{yCFT}_{\mathrm{sp}}(q-1)$ is a substitution-closed full abstract family of languages [28], it suffices to show that there are some $L \in \mathrm{CFT}_{\mathrm{sp}}(q-1)$ and homomorphism h such that $h(\mathbf{enc}(L)) = D_{qn} \cap g^{-1}(D_{qn})$. Let $m = 2, \Sigma = \{c_1, \dots, c_n\}, r = q-1, p = 2$, and $\Upsilon = \widetilde{\Sigma}_{q-1} \cup \{X_0, \dots, X_{q-1}\}$. Lemma 30 gives finite sets $A \subseteq \Upsilon, Z = \widetilde{\Sigma}_{q-1}, I \subseteq \{X_0, \dots, X_{q-1}\} \times \mathbb{T}^2_\Upsilon$ such that

$$DT^2_\Sigma \cap \mathbb{T}^2_{\widetilde{\Sigma}_{q-1},2} = \mathbf{y}_2(\mathrm{Loc}^3(A, Z, I)).$$

Let $L = DT^2_\Sigma \cap \mathbb{T}^2_{\widetilde{\Sigma}_{q-1},2}$. Inspection of the proof of Lemma 30 also shows that for all $\boldsymbol{T} \in \mathrm{Loc}^3(A, Z, I)$ and all $v \in T$, $v \in \mathrm{dom}(\prec_3^T)$ implies $|C_2^T(v)| \leq q-1$. So $L \in \mathrm{CFT}_{\mathrm{sp}}(q-1)$. Let $\Phi = \{ (c_i, q-1, j) \mid 1 \leq i \leq n, 0 \leq j \leq q-1 \}$. We identify $\Gamma_\Phi = \bigcup \{ \{ \llbracket_d, \rrbracket_d \} \mid d \in \Phi \}$ with Γ_{qn}. Let $h \colon (\Gamma_{\widetilde{\Sigma}_{q-1}})^* \to (\Gamma_{\widetilde{\Sigma}_{q-1}})^*$ be the homomorphism that erases all symbols that are not in Γ_{qn}. Our goal is to show

$$h(\mathbf{enc}(L)) = D_{qn} \cap g^{-1}(D_{qn}).$$

To show $h(\mathbf{enc}(L)) \subseteq D_{qn} \cap g^{-1}(D_{qn})$, suppose $\boldsymbol{T} \in L$. Since $L \in \mathbb{T}^2_{\widetilde{\Sigma}_{q-1},2}$, it is clear that

$$h(\mathbf{enc}(\boldsymbol{T})) \in D_{qn}. \tag{28}$$

Since $L \subseteq DT^2_{\Sigma}$, by Lemma 36,

$$\eta(\mathbf{enc}(\boldsymbol{T})) \in D_{\widetilde{\Sigma}_{q-1}},$$

where $\eta \colon \Gamma^*_{\widetilde{\Sigma}} \to \Gamma^*_{\widetilde{\Sigma}}$ is as defined in Lemma 36, with Σ in place of Υ. So

$$h(\eta(\mathbf{enc}(\boldsymbol{T})) \in h(D_{\widetilde{\Sigma}_{q-1}}) = D_{qn}.$$

Note that η restricted to Γ_{qn} coincides with g. So we have

$$\begin{aligned} h(\eta(\mathbf{enc}(\boldsymbol{T})) &= \eta(h(\mathbf{enc}(\boldsymbol{T}))) \\ &= g(h(\mathbf{enc}(\boldsymbol{T}))). \end{aligned}$$

This shows that

$$h(\mathbf{enc}(\boldsymbol{T})) \in g^{-1}(D_{qn}). \tag{29}$$

By (28) and (29), $h(\mathbf{enc}(L)) \subseteq D_{qn} \cap g^{-1}(D_{qn})$.

Now we show the converse inclusion. Let $s \in D_{qn} \cap g^{-1}(D_{qn})$. Then there is a hedge $\boldsymbol{T} \in \mathbb{H}^2_{\Phi}$ such that $s = \mathbf{enc}(\boldsymbol{T})$. We turn $\boldsymbol{T}$ into a tree $\boldsymbol{T}' = \varphi(\boldsymbol{T}) \in \mathbb{T}_{\{c_1\},\Phi} \cap \mathbb{T}^2_{\widetilde{\Sigma}_{q-1},2}$, where symbols in Φ are assumed to have rank 1:

$$\begin{aligned} \varphi((c_i, q-1, j)) &= (c_i, q-1, j)(c_1), \\ \varphi((c_i, q-1, j)(\boldsymbol{T}_1 \dots \boldsymbol{T}_n)) &= (c_i, q-1, j)(\varphi(\boldsymbol{T}_1 \dots \boldsymbol{T}_n)), \\ \varphi(\boldsymbol{T}_1 \dots \boldsymbol{T}_n) &= c_1(\varphi(\boldsymbol{T}_1)\, \varphi(\boldsymbol{T}_2 \dots \boldsymbol{T}_n)) \quad \text{where } n \geq 2. \end{aligned}$$

Then $s = h(\mathbf{enc}(\boldsymbol{T}'))$. We have

$$\eta(\mathbf{enc}(\boldsymbol{T}')) = g(s) \in D_{qn} \subseteq D_{\widetilde{\Sigma}}.$$

Since $\boldsymbol{T}' \in \mathbb{T}_{\{c_1\},\Phi} \subseteq \mathbb{T}_{\Sigma,\widetilde{\Sigma}-\Sigma}$, Lemma 36 implies that $\boldsymbol{T}' \in DT^2_{\Sigma}$. So $\boldsymbol{T}' \in L$ and $s = h(\mathbf{enc}(\boldsymbol{T}')) \in h(\mathbf{enc}(L))$. We conclude $D_{qn} \cap g^{-1}(D_{qn}) \subseteq h(\mathbf{enc}(L))$.

We have shown $h(\mathbf{enc}(L)) = D_{qn} \cap g^{-1}(D_{qn})$.

(i) $\Rightarrow$ (ii). Let $L \in \mathrm{CFT_{sp}}(q-1)$ be such that $M = \mathbf{y}(L)$. By Lemma 40, $L = \mathbf{y}_2(K)$ and $\mathbf{enc}_2(K) \subseteq \mathbb{T}^2_{\widetilde{\Psi}_{q-1}}$ for some finite set Ψ and some local set $K \subseteq \mathbb{T}^3_{\Psi,\boldsymbol{x}}$. By Lemma 37, $\mathbf{enc}(\mathbf{enc}_2(K)) = \widehat{\widehat{\pi}}(R' \cap D_{\widetilde{\Upsilon}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}}))$ for some local set $R' \subseteq \Gamma_{\widetilde{\Upsilon}}$ and projection $\pi \colon \Upsilon \to \Psi$. Since $\mathbf{enc}_2(K) \subseteq \mathbb{T}^2_{\widetilde{\Psi}_{q-1}}$, it easily follows that $\mathbf{enc}(\mathbf{enc}_2(K)) = \widehat{\widehat{\pi}}(R'' \cap D_{\widetilde{\Upsilon}_q} \cap \eta^{-1}(D_{\widetilde{\Upsilon}_q}))$, where $R'' = R' \cap (\Gamma_{\widetilde{\Upsilon}_{q-1}})^+$, which is a local subset of $(\Gamma_{\widetilde{\Upsilon}_{q-1}})^+$. Using this in the proof of Lemma 39, we easily obtain

$$M = h'(R'' \cap D_{\widetilde{\Upsilon}_{q-1}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}_{q-1}})), \tag{30}$$

where $h' \colon (\Gamma_{\widetilde{\Upsilon}_{q-1}})^* \to \Sigma^*$ is an alphabetic homomorphism. In order to obtain the statement of the lemma, there are three things we need to fix:

– for $c \in \Upsilon$, η erases $[_c$ and $]_c$,

– when $q \geq 2$, the number of pairs of brackets in the group $[_c,]_c$ is $1 < q$, and
– when $k < q - 1$, the number of pairs of brackets in the group $[_{(c,k,0)},]_{(c,k,0)}, [_{(c,k,1)},]_{(c,k,1)}, \ldots, [_{(c,k,k)},]_{(c,k,k)}$ is $k + 1 < q$.

We introduce the following new brackets:

$$[_{c,1},]_{c,1}, \ldots, [_{c,q-1},]_{c,q-1},$$
$$[_{(c,k,k+1)},]_{(c,k,k+1)}, \ldots, [_{(c,k,q-1)},]_{(c,k,q-1)},$$

for each $c \in \Upsilon$ and $k < q - 1$. We now have an alphabet Γ_{qn} consisting of n groups of q pairs of brackets:

$$[_c,]_c, [_{c,1},]_{c,1}, \ldots, [_{c,q-1},]_{c,q-1},$$
$$[_{(c,k,0)},]_{(c,k,0)}, \ldots, [_{(c,k,q-1)},]_{(c,k,q-1)},$$

where $n = |\Upsilon| \times (q + 1)$. Define a homomorphism $\psi \colon (\Gamma_{\widetilde{\Upsilon}_{q-1}})^* \to \Gamma_{qn}^*$ by

$$\begin{aligned}
\psi([_c) &= [_c \, [_{c,1}, \\
\psi(]_c) &=]_{c,1} \, [_{c,2} \,]_{c,2} \ldots [_{c,q-1} \,]_{q-1} \,]_c, \\
\psi([_{(c,k,0)}) &= [_{(c,k,0)}, \\
\psi(]_{(c,k,0)}) &= [_{(c,k,k+1)} \,]_{(c,k,k+1)} \ldots [_{(c,k,q-1)} \,]_{(c,k,q-1)} \,]_{(c,k,0)}, \\
\psi([_{(c,k,i)}) &= [_{(c,k,i)}, \\
\psi(]_{(c,k,i)}) &=]_{(c,k,i)} \qquad \text{for } 1 \leq i \leq k.
\end{aligned}$$

Now it is easy to see that ψ is injective and $\psi(R'')$ is a local subset of Γ_{qn}^*. Also, we can show that all $x \in (\Gamma_{\widetilde{\Upsilon}_{q-1}})^*$ satisfy the following properties:

$$x \in D_{\widetilde{\Upsilon}_{q-1}} \text{ if and only if } \psi(x) \in D_{qn}, \qquad (31)$$
$$g(\psi(x)) \rightsquigarrow^* \psi(\eta(x)).$$

These properties combined ensure that

$$\eta(x) \in D_{\widetilde{\Upsilon}_{q-1}} \text{ if and only if } g(\psi(x)) \in D_{qn}. \qquad (32)$$

Let $\chi \colon \Gamma_{qn}^* \to (\Gamma_{\widetilde{\Upsilon}_{q-1}})^*$ be the alphabetic homomorphism that erases all new brackets. Clearly, χ restricted to the range of ψ is the inverse of ψ. Now observe

$$\chi(\psi(R'') \cap D_{qn} \cap g^{-1}(D_{qn})) = R'' \cap D_{\widetilde{\Upsilon}_{q-1}} \cap \eta^{-1}(D_{\widetilde{\Upsilon}_{q-1}}). \qquad (33)$$

Indeed, if $x \in R''$, $\psi(x) \in D_{qn}$, and $g(\psi(x)) \in D_{qn}$, then $\chi(\psi(x)) = x \in R'' \cap D_{\widetilde{\Upsilon}_{q-1}}$ by (31), and $\eta(\chi(\psi(x)) = \eta(x) \in D_{\widetilde{\Upsilon}_{q-1}}$ by (32). Conversely, if $x \in R'' \cap D_{\widetilde{\Upsilon}_{q-1}}$ and $\eta(x) \in D_{\widetilde{\Upsilon}_{q-1}}$, then $\psi(x) \in \psi(R'') \cap D_{qn}$ by (31) and $g(\psi(x)) \in D_{qn}$ by (32), and so $x = \chi(\psi(x)) \in \chi(\psi(R'') \cap D_{qn} \cap g^{-1}(D_{qn}))$.

We obtain the statement of the lemma from (30) and (33) by taking $R = \psi(R'')$ and $h = h' \circ \chi$. □

As before, in the direction (i) ⇒ (ii) of the above proof, $R \cap D_{qn} \cap g^{-1}(D_{qn}) = \psi(R'') \cap D_{qn} \cap g^{-1}(D_{qn})$ stands in one-one correspondence with the local set $K \subseteq \mathbb{T}^3_{\Psi,\boldsymbol{x}}$. Thus, each derivation tree $\boldsymbol{T}$ of a simple context-free tree grammar for M is uniquely represented by an element s of $R \cap D_{qn} \cap g^{-1}(D_{qn})$ such that $\mathbf{enc}(\mathbf{enc}_2(\boldsymbol{T}))$ is the image of s under a certain alphabetic homomorphism, and vice versa. This is exactly analogous to the situation with the original Chomsky-Schützenberger representation theorem.

As in the case of context-free languages, we can take a fixed Dyck language D_{2q}, instead of D_{qn} with varying n, and use a rational transduction to represent any string language that is the yield image of some $L \in \mathrm{CFT}_{\mathrm{sp}}(q-1)$:[23]

Corollary 43. *For any $M \in \mathrm{yCFT}_{\mathrm{sp}}(q-1)$, there is a rational transduction τ such that $M = \tau(D_{2q} \cap g^{-1}(D_{2q}))$, where g is as defined in Theorem 42.*

10 Conclusion

We have generalized Weir's [34] characterization of the string languages of tree-adjoining grammars to the string languages of simple context-free tree grammars of arbitrary fixed rank. We obtained this result via a natural generalization of the original Chomsky-Schützenberger theorem to simple context-free tree grammars. We represented derivation trees of simple context-free tree grammars as 3-dimensional trees, and proved this latter result as a general fact about simple context-free sets of m-dimensional trees, for arbitrary $m \geq 2$. This generality is of course an overkill for the purpose of obtaining our generalization of Weir's theorem, but it may be of independent interest. Moreover, all the complexity of the general case is essentially already present in the 3-dimensional case; proving only the special case of our lemmas that are needed for the generalization of Weir's theorem will not be substantially simpler.

In order to define the m-dimensional yield of an $(m+1)$-dimensional tree, we placed a restriction on the occurrences of a special label $\boldsymbol{x}$ that serve as targets for substitution. If we wish to iterate the process of taking the yield, i.e., if we are interested in the yield of the yield of an $(m+2)$-dimensional tree, etc., we will need more than one variable as placeholders, with each variable providing targets for substitution at a different step of the iterative process of taking yields. Although we did not attempt to do so in this paper, it may be interesting to study the resulting hierarchy of classes of tree languages (and their yield images), the first three levels of the hierarchy being the local tree languages,

[23] Also, when string languages over a fixed alphabet Σ with $|\Sigma| = k$ are considered, we can use $D_{q(k+2)}$ and a fixed alphabetic homomorphism h so that every $M \in \mathrm{yCFT}_{\mathrm{sp}}(q-1)$ can be written as $M = h(R \cap D_{q(k+2)} \cap g^{-1}(D_{q(k+2)}))$ for some *regular* set R. (This will require modification of the constructions used in several lemmas.) See [36] for an analogous characterization of string languages of q-MCFGs of rank r, and, e.g., [27] for the case of context-free languages.

the simple context-free tree languages, and the yields of the simple context-free sets of (well-labeled) 3-dimensional trees.[24]

References

1. Arnold, A., Dauchet, M.: Un theoreme de Chomsky-Schützenberger pour les forets algebriques. CALCOLO 14(2), 161–184 (1977)
2. Benoit, D., Demaine, E.D., Munro, I., Raman, R., Raman, V., Rao, S.S.: Representing trees of higher degree. Algorithmica 43, 275–292 (2005)
3. Chomsky, N., Schützeberger, M.P.: The algebraic theory of context-free languages. In: Braffort, P., Hirschberg, D. (eds.) Computer Programming and Formal Systems, pp. 118–161. North-Holland, Amsterdam (1963)
4. Chomsky, N.: Formal properties of grammars. In: Luce, R.D., Bush, R.R., Galanter, E. (eds.) Handbook of Mathematical Psychology, Volume II, pp. 323–418. John Wiley and Sons, New York (1963)
5. Comon, H., Dauchet, M., Gilleron, R., Jacquemard, F., Lugiez, D., Löding, C., Tison, S., Tommasi, M.: *Tree Automata Techniques and Applications.* Available on `http://tata.gforge.inria.fr/` (2008), release November 18, 2008
6. Eilenberg, S.: Automata, Languages, and Machines, vol. A. Academic Press, New York (1974)
7. Engelfreit, J., Maneth, S.: Tree languages generated by context-free graph grammars. In: Ehrig, H., Engels, G., Kreowski, H.J., Rozenberg, G. (eds.) Theory and Application of Graph Transformations: 6th International Workshop, TAGT'98. pp. 15–59. Springer, Berlin (2000)
8. Engelfriet, J., Schmidt, E.M.: IO and OI, part I. The Journal of Computer and System Sciences 15, 328–353 (1977)
9. Fischer, M.J.: Grammars with Macro-Like Productions. Ph.D. thesis, Harvard University (1968)
10. Fujiyoshi, A., Kasai, T.: Spinal-formed context-free tree grammars. Theory of Computing Systems 33, 59–83 (2000)
11. Joshi, A.K., Schabes, Y.: Tree-adjoining grammars. In: Rozenberg, G., Salomaa, A. (eds.) Handbook of Formal Languages, vol. 3, pp. 69–123. Springer, Berlin (1997)
12. Kanazawa, M.: The convergence of well-nested mildly context-sensitive grammar formalisms (July 2009), an invited talk given at the 14th Conference on Formal Grammar, Bordeaux, France. Slides available at `http://research.nii.ac.jp/~kanazawa/`.
13. Kanazawa, M.: The pumping lemma for well-nested multiple context-free languages. In: Diekert, V., Nowotka, D. (eds.) Developments in Language Theory: 13th International Conference, DLT 2009. pp. 312–325. Springer, Berlin (2009)
14. Kasprzik, A.: Making finite-state methods applicable to languages beyond context-freeness via multi-dimensional trees. In: Piskorski, J., Vatson, B.W., Yli-Jyrä, A. (eds.) Finite-State Methods and Natural Language Processing, pp. 98–109. IOS Press, Amsterdam (2009)
15. Kepser, S., Mönnich, U.: Closure properties of linear context-free tree languages with an application to optimality theory. Theoretical Computer Science 354(1), 82–97 (2006)

[24] Rogers [23] looked at a connection between multi-dimensional trees (with his notion of yield) and Weir's [35] control language hierarchy.

16. Kepser, S., Rogers, J.: The equivalence of tree adjoining grammars and monadic linear context-free tree grammars. Journal of Logic, Language and Information 20, 361–384 (2011)
17. Knuth, D.E.: The Art of Computer Programming, Vol. I: Fundamental Algorithms. Addison-Wesley, Reading, Mass. (1997)
18. Kozen, D.C.: Automata and Computability. Springer, New York (1997)
19. Matsubara, S., Kasai, T.: A characterization of TALs by the generalized Dyck language. IEICE Transactions on Information and Systems (Japanese Edition) J90-D, 1417–1427 (2007)
20. McNaughton, R., Papert, S.A.: Counter-Free Automata. MIT Press, Cambridge, Mass. (1971)
21. Mönnich, U.: Adjunction as substitution: An algebraic formulation of regular, context-free and tree adjoining languages (1997), arXiv:cmp-lg/9707012v1
22. Perrin, D.: Finite automata. In: van Leeuwen, J. (ed.) Handbook of Theoretical Computer Science, vol. B, pp. 1–57. Elsevier, Amsterdam (1990)
23. Rogers, J.: Syntactic structures as multi-dimensional trees. Research on Language and Computation 1, 265–305 (2003)
24. Rogers, J.: wMSO theories as grammar formalisms. Theoretical Computer Science 293, 291–320 (2003)
25. Rogers, J., Pullum, G.K.: Aural pattern recognition experiments and the subregular hierarchy. Journal of Logic, Language and Information 20, 329–342 (2011)
26. Rounds, W.: Mappings and grammars on trees. Mathematical Systems Theory 4, 257–287 (1970)
27. Salomaa, A.: Formal Languages. Academic Press, Orlando, Florida (1973)
28. Seki, H., Kato, Y.: On the generative power of multiple context-free grammars and macro grammars. IEICE Transactions on Information and Systems E91–D, 209–221 (2008)
29. Seki, H., Matsumura, T., Fujii, M., Kasami, T.: On multiple context-free grammars. Theoretical Computer Science 88(2), 191–229 (1991)
30. Sorokin, A.: Monoid automata for displacement context-free languages. In: ESSLLI Student Session 2013 Preproceedings. pp. 158–167 (2013)
31. Takahashi, M.: Generalizations of regular sets and their application to a study of context-free languages. Information and Control 27, 1–36 (1975)
32. Thatcher, J.W.: Characterizing derivation trees of context-free grammars through a generalization of finite automata theory. Journal of Computer and System Sciences 1, 317–322 (1967)
33. Thatcher, J.W.: Generalized^2 sequential machine maps. Journal of Computer and System Sciences 24, 339–367 (1970)
34. Weir, D.J.: Characterizing Mildly Context-Sensitive Grammar Formalisms. Ph.D. thesis, University of Pennsylvania (1988)
35. Weir, D.J.: A geometric hierarchy beyond context-free languages. Theoretical Computer Science 104(2), 235–261 (1992)
36. Yoshinaka, R., Kaji, Y., Seki, H.: Chomsky-Schützenberger-type characterization of multiple context-free languages. In: Dediu, A.H., Fernau, H., Martín-Vide, C. (eds.) Language and Automata Theory and Applications, Fourth International Conference, LATA 2010. pp. 596–607. Springer, Berlin (2010)