Annales Mathematicae et Informaticae
56 (2022) pp. 58–70
DOI: https://doi.org/10.33039/ami.2022.12.007
URL: https://ami.uni-eszterhazy.hu

Tree generating context-free grammars and regular tree grammars are equivalent

Dávid Kószó

Department of Foundations of Computer Science
University of Szeged
Szeged, Hungary
koszod@inf.u-szeged.hu

Abstract. We show that it is decidable whether the language generated by a given context-free grammar over a tree alphabet is a tree language. Furthermore, if the answer to this question is "yes", then we can even effectively construct a regular tree grammar which generates that tree language.

Keywords: context-free grammar, regular tree grammar, tree language, parenthesis grammar, tree generating context-free grammar, decidability

AMS Subject Classification: 68Q45

1. Introduction

Context-free grammars (for short: cfg) were introduced in [3] in order to describe the structure of sentences and words in natural languages. Since then, a beautiful theory of cfg has been evolved, *cf. e.g.* [6, 7]. In computer science cfg are used to describe the structure of programming languages and play a crucial role in the Document Type Definition (DTD) of the Extensible Markup Language (XML) as well. The language generated by a Γ-cfg G, *i.e.*, a cfg over some alphabet Γ, is denoted by $\mathrm{L}(G)$ and called a context-free language.

In order to define well-formed terms, we use a special alphabet called a ranked alphabet and three further special symbols. A ranked alphabet Σ is an alphabet in which we associate with each symbol a unique rank. The three special symbols are the opening angle bracket "$\langle$", the closing angle bracket "$\rangle$", and the symbol "$\#$". The set of these special symbols is denoted by Ξ and the alphabet Σ^{Ξ} containing

Submitted: October 25, 2022
Accepted: December 3, 2022
Published online: December 28, 2022

the symbols of Σ and Ξ is called a tree alphabet. Using the three special symbols as separators, the Σ-terms are defined in the standard way, *i.e.*, each Σ-term is a string $\sigma\langle\xi_1\#\ldots\#\xi_k\rangle$ over the tree alphabet Σ^Ξ, where σ has rank k for some natural number k, and $\xi_1,\ldots,\xi_k$ are Σ-terms.

Since each Σ-term can be depicted as a tree-like directed labelled graph, we often refer to Σ-terms as Σ-trees. Moreover, a set of Σ-trees is called a (formal) Σ-tree language. We denote the set of all Σ-trees by T_Σ and we call a Σ^Ξ-cfg G tree generating if $\mathrm{L}(G) \subseteq \mathrm{T}_\Sigma$.

To generate Σ-tree languages, among others regular tree grammars (for short: Σ-rtg) were defined [2, 4, 5]. The Σ-tree language generated by a Σ-rtg $\mathcal{G}$, denoted by $\mathrm{L}(\mathcal{G})$, is called a regular Σ-tree language. The connection between context-free languages and regular tree languages has been thoroughly investigated. Among others, it was shown that, for each language L, the following statements are equivalent [2, 10]:

(1) L is a context-free language,

(2) L is the yield of a regular tree language.

Then several authors have exploited this strong connection, *cf.* *e.g.*, [4, 11, 12]. Furthermore, each Σ-rtg is evidently a tree generating Σ^Ξ-cfg. However, to the best of our knowledge, it has not been cleared yet whether there exists a Σ-tree language, which can be generated by a Σ^Ξ-cfg but it is not regular. Hence, here we deal with the following questions and answer them positively:

(Q1) Given a Σ^Ξ-cfg G, is it decidable whether G is tree generating?

(Q2) Given a Σ^Ξ-cfg G such that G is tree generating, is $\mathrm{L}(G)$ regular, and if yes, can we effectively construct a Σ-rtg $\mathcal{G}$ such that $\mathrm{L}(\mathcal{G}) = \mathrm{L}(G)$?

To answer the questions, we will consider the class of parenthesis grammars. A Γ-parenthesis grammar [9] is a Γ-cfg in which each rule has the form $A \to \langle\alpha\rangle$, where A is a nonterminal and α is a string over $\Gamma \setminus \{\langle,\rangle\}$. A language generated by a Γ-parenthesis grammar is called a Γ-parenthesis language. Interestingly, we can give a transduction φ such that, for each Σ-tree language L, the language $\varphi(L)$ is a Σ^Ξ-parenthesis language. (We note that there exists a Σ^Ξ-parenthesis language, which is not an image of any Σ-tree language under φ.) We prove our results by exploiting this connection between Σ-rtg and Σ^Ξ-parenthesis grammars and by applying Knuth's results [8]:

(R1) it is decidable, for a given Γ-cfg G, whether $\mathrm{L}(G)$ is a Γ-parenthesis language and

(R2) for a given Γ-cfg G such that $\mathrm{L}(G)$ is a Γ-parenthesis language, we can effectively construct a Γ-parenthesis grammar G' such that $\mathrm{L}(G') = \mathrm{L}(G)$.

We mention that, for unranked trees, question (Q1) was answered positively in [1].

Our paper is organized as follows. In Section 2 we recall the necessary notions and notations. In Section 3 we recall the concept of cfg and of rtg, and results on parenthesis grammars. In Section 4 we recall the concept of sequential transducer, which will be useful to prove our results. Finally, in Section 5 we give our results.

2. Preliminaries

2.1. Basic concepts

We denote the set $\{0, 1, 2, \ldots\}$ of nonnegative integers by $\mathbb{N}$ and we let $\mathbb{N}_+ = \mathbb{N}\setminus\{0\}$. For every $k \in \mathbb{N}$, we let $[k] = \{i \in \mathbb{N}_+ \mid i \leq k\}$. In particular, $[0] = \emptyset$. Furthermore, we denote the set of integers by $\mathbb{Z}$.

Let A be a set and $R, S \subseteq A \times A$ binary relations. The *composition of R and S*, denoted by $R \circ S$, is the set

$$R \circ S = \{(a, c) \in A \times A \mid (\exists b \in A) : (a, b) \in R \wedge (b, c) \in S\} \ .$$

We define, for each $n \in \mathbb{N}$, the *n-fold composition of R*, denoted by R^n, by $R^0 = \{(a, a) \mid a \in A\}$ and by $R^n = R^{n-1} \circ R$ for each $n \in \mathbb{N}_+$.

2.2. Strings and trees

We assume that the reader is familiar with the fundamental concepts and results of the theory formal languages [6, 7], and also of tree languages [4, 5].

An *alphabet* is a finite set. Let Γ be an alphabet. A *string (over Γ)* is a finite sequence $a_1 \cdots a_k$ with $k \in \mathbb{N}$ and $a_i \in \Gamma$ for each $i \in [k]$. The *length of* $a_1 \cdots a_k$, denoted by $\mathrm{len}(a_1 \cdots a_k)$, is defined in the standard way. We denote by Γ^* the *set of all strings over* Γ and by ε the *empty string*. Each subset $L \subseteq \Gamma^*$ is called a *language over* Γ. Moreover, for all $v, w \in \Gamma^*$, we denote by vw the *concatenation of v and w*, and the *set of prefixes of v*, denoted by $\mathrm{prefix}(v)$, is defined by $\mathrm{prefix}(v) = \{u \in \Gamma^* \mid (\exists v' \in \Gamma^*) : v = uv'\}$.

A *ranked alphabet* is a tuple (Σ, rk), where Σ is an alphabet and $\mathrm{rk} : \Sigma \to \mathbb{N}$ is a mapping, called *rank mapping*, such that $\mathrm{rk}^{-1}(0) \neq \emptyset$. For all $k \in \mathbb{N}$, we let

$$\Sigma^{(k)} = \{\sigma \in \Sigma \mid \mathrm{rk}(\sigma) = k\} \ .$$

We always abbreviate (Σ, rk) by Σ.

Next we define Σ-trees. In the literature, Σ-trees are defined by using the opening and the closing parenthesis "(" and ")", respectively, and the comma "," as separators [4, 5]. In this paper, we will focus on these separators in trees frequently. Since it is easy to confuse these separators with the two parentheses in other formulas, we intentionally deviate and use the opening and the closing angle brackets "$\langle$" and "$\rangle$", respectively, and the symbol "#" to define Σ-trees.

Let Ξ be the set which consists of "$\langle$" and "$\rangle$" and "#". A *tree alphabet* Σ^Ξ is an alphabet consisting of symbols of Σ and Ξ, *i.e.*, $\Sigma^\Xi = \Sigma \cup \Xi$.

Let H be a set such that $H \cap \Sigma^\Xi = \emptyset$. The *set of Σ-trees over H*, denoted by $\mathrm{T}_\Sigma(H)$, is the smallest set $T \subseteq (\Sigma^\Xi \cup H)^*$ such that

(i) $H \subseteq T$ and

(ii) if $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, and $\xi_1, \ldots, \xi_k \in T$, then $\sigma\langle\xi_1\#\ldots\#\xi_k\rangle \in T$.

We abbreviate $\mathrm{T}_\Sigma(\emptyset)$ by T_Σ. A *Σ-tree language* (or just: *tree language*) is a subset of T_Σ.

> *From now on, we let Σ be an arbitrary ranked alphabet if not specified otherwise.*

3. Grammar models

3.1. Context-free grammars

Let Γ be an alphabet. A *context-free grammar over* Γ (for short: Γ-*cfg*) [6, 7] is a triple $G = (N, S, R)$ where N is a finite set (*nonterminals*) with $N \cap \Gamma = \emptyset$, $S \in N$ (*start symbol*), and R is a finite set (*rules*); each rule has the form $A \to \alpha$ where $A \in N$ and α is a string over $N \cup \Gamma$, *i.e.*, $\alpha \in (N \cup \Gamma)^*$. Furthermore, we call each element $a \in \Gamma$ a *terminal.*

Let $G = (N, S, R)$ be a Γ-cfg and let $r = (A \to \alpha)$ be a rule. We call A and α the *left-hand side* and the *right-hand side* of r, respectively. Moreover, we call r a *chain rule* (an ε-*rule*) if $\alpha \in N$ (if $\alpha = \varepsilon$, respectively). We say that G is *chain-free* (ε-*free*) if G does not have chain rules (ε-rules, respectively).

The *(leftmost) derivation relation* $\Rightarrow_G$ is defined such that, for every $u \in \Gamma^*$, $\gamma \in (N \cup \Gamma)^*$, and rule $A \to \alpha$ in R, we have $uA\gamma \Rightarrow_G u\alpha\gamma$. If G is clear from the context, then we abbreviate $\Rightarrow_G$ by $\Rightarrow$. For all $\gamma, \omega \in (N \cup \Gamma)^*$, if $\gamma \Rightarrow^n \omega$ for some $n \in \mathbb{N}$, then we say that this derivation has length n. As usual, we denote the reflexive and transitive closure of $\Rightarrow$ by $\Rightarrow^*$, *i.e.*, $\Rightarrow^* = \bigcup_{n \in \mathbb{N}} \Rightarrow^n$.

The *language generated by* G is the set

$$\mathrm{L}(G) = \{w \in \Gamma^* \mid S \Rightarrow^* w\} \ .$$

For each $L \subseteq \Gamma^*$, we call L a *context-free language* if there exists a Γ-cfg G such that $\mathrm{L}(G) = L$.

We call a nonterminal $A \in N$ *useful (in G)* if there exist $u, w \in \Gamma^*$ and $\gamma \in (N \cup \Gamma)^*$ such that $S \Rightarrow^* uA\gamma \Rightarrow^* w$. Moreover, if every $A \in N$ is useful, then we call G *reduced* [6, p. 78].

Lemma 3.1. [6, Thm. 3.2.3] If G is a Γ-cfg, then we can effectively construct a reduced Γ-cfg $\widehat{G}$ such that $\mathrm{L}(\widehat{G}) = \mathrm{L}(G)$.

Next we define parenthesis grammars and languages. They are normally defined by using the opening and the closing parenthesis "(" and ")". Later, in Section 5, we will relate Σ-tree languages and parenthesis languages. Therefore, we will consistently deviate from the convention and use the angle brackets "$\langle$" and "$\rangle$" instead of the usual "(" and ")", respectively; however we keep the notions parenthesis grammar and parenthesis language.

> *In the rest of this section, we let Γ be an alphabet which contains the angle brackets "$\langle$" and "$\rangle$".*

A Γ-*parenthesis grammar* [8] (or just: parenthesis grammar) is a Γ-cfg $G = (N, S, R)$ such that each rule in R has the form $A \to \langle\theta\rangle$ with $\theta \in (N \cup \Gamma \setminus \{\langle, \rangle\})^*$.

Table 1. Illustration of the content and the deficiency mappings, and the notion balanced.

w	$c(w)$	$d(w)$	balanced
$a\langle b\langle\rangle\rangle$	0	0	yes
$\langle\rangle\rangle$	-1	1	no
$\langle a\langle$	2	0	no
$\langle a\rangle b\rangle\rangle\langle b\rangle$	-2	2	no
$\rangle\rangle\rangle\langle\langle\langle$	0	3	no

Observation 3.2. If G is a Γ-parenthesis grammar, then G is chain-free and ε-free.

We call a language $L \subseteq \Gamma^*$ a Γ-*parenthesis language* (or just: parenthesis language) if there exists a Γ-parenthesis grammar G such that $\mathrm{L}(G) = L$.

Here we draw attention to the following phenomenon. Let G be a Γ-cfg such that $\mathrm{L}(G)$ is a parenthesis language. Then it does not follow that G is a parenthesis grammar. Rather, it follows that there exists a Γ-parenthesis grammar G' such that $\mathrm{L}(G') = \mathrm{L}(G)$. We will use this fact later.

The *content mapping* $c : \Gamma^* \to \mathbb{Z}$ and the *deficiency mapping* $d : \Gamma^* \to \mathbb{N}$ [8] are defined, for each $w \in \Gamma^*$, as follows:

(i) if $w = \varepsilon$, then we let $c(\varepsilon) = d(\varepsilon) = 0$,

(ii) if $w = a$ for some $a \in \Gamma$, then we let

$$c(a) = \begin{cases} 1 & \text{if } a = \langle \\ -1 & \text{if } a = \rangle \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad d(a) = \begin{cases} 1 & \text{if } a = \rangle \\ 0 & \text{otherwise} \end{cases}, \qquad \text{and}$$

(iii) if $w = va$ with $v \in \Gamma^*$ and $a \in \Gamma$, then we let $c(va) = c(v) + c(a)$ and $d(va) = \max\{d(v), d(a) - c(v)\}$.

Intuitively, for each string $w \in \Gamma^*$, the values $c(w)$ and $d(w)$ show us the excess of left parentheses over right parentheses in w and the greatest deficiency of left parentheses from right parentheses in any prefix of w, respectively. A string $w \in \Gamma^*$ is *balanced* if $c(w) = d(w) = 0$, and furthermore, a language $L \subseteq \Gamma^*$ is *balanced* if every $w \in L$ is balanced.

Observe that, for all balanced $u, v \in \Gamma^*$, also uv is balanced. Furthermore, each $u \in (\Gamma \setminus \{\langle, \rangle\})^*$ is balanced as well.

Example 3.3. Let $\Gamma = \{a, b, \langle, \rangle\}$. Table 1 shows, for some $w \in \Gamma^*$, the values of the content and the deficiency mappings, *i.e.*, $c(w)$ and $d(w)$, respectively, and whether w is balanced or not.

The next lemma shows an important property of parenthesis grammars and it will be useful to prove the results in Section 5.

Lemma 3.4. Let $G = (N, S, R)$ be a Γ-parenthesis grammar. Furthermore, let $A \in N$ and $w \in \Gamma^*$. If $A \Rightarrow^* w$, then $w = \langle u \rangle$ for some $u \in \Gamma^*$ such that u is balanced.

Proof. We prove our statement by induction on the length of the derivation $A \Rightarrow^* w$. Assume that $A \Rightarrow w$. Then, since G is a parenthesis grammar, we have $w = \langle u \rangle$ for some $u \in (\Gamma \setminus \{\langle, \rangle\})^*$. Hence u is balanced.

Now assume that $A \Rightarrow^{n+1} w$ for some $n \in \mathbb{N}_+$. Then, since G is a parenthesis grammar, there exist $k \in \mathbb{N}_+$, $v_0, v_1, v_2, \ldots, v_k$ in $(\Gamma \setminus \{\langle, \rangle\})^*$, $A_1, A_2, \ldots, A_k \in N$, $n_1, n_2, \ldots, n_k \in [n]$, and $w_1, w_2, \ldots, w_k \in \Gamma^*$ such that

- $w = \langle v_0 w_1 v_1 w_2 v_2 \cdots w_k v_k \rangle$,
- $A \to \langle v_0 A_1 v_1 A_2 v_2 \cdots A_k v_k \rangle$ is in R,
- for each $i \in [k]$ we have $A_i \Rightarrow^{n_i} w_i$,
- $n_1 + n_2 + \ldots + n_k = n$, and
- we have

$$A \Rightarrow^1 \langle v_0 A_1 v_1 A_2 v_2 \cdots A_k v_k \rangle \ \Rightarrow^{n_1} \langle v_0 w_1 v_1 A_2 v_2 \cdots A_k v_k \rangle \Rightarrow^{n_2} \cdots \Rightarrow^{n_k} w \ .$$

By I.H., for each $i \in [k]$, we may assume that there exists $u_i \in \Gamma^*$ such that $w_i = \langle u_i \rangle$ and u_i is balanced. Thus, for $u = v_0 w_1 v_1 w_2 v_2 \cdots w_k v_k$ it holds that $w = \langle u \rangle$ and u is balanced. This completes our proof. □

Let $w \in \Gamma^*$. For every $a, b \in \Gamma$, the terminals a, b are called *associates (in w)* [8] if $w = uavbv'$ for some $u, v, v' \in \Gamma^*$ and vb is balanced. A language $L \subseteq \Gamma^*$ is said to have *bound associates* if there exists a constant $K \in \mathbb{N}_+$ such that for all $w = uav$ in L with $u, v \in \Gamma^*$ and $a \in \Gamma$, the terminal a has at most K associates in w.

Example 3.5. Let $\Gamma = \{a, b, \langle, \rangle\}$. We consider the Γ-cfg

$$G = (\{S\}, S, \{\ S \to \varepsilon\ ,\ S \to aSb\ \})\ .$$

Then we have $\mathrm{L}(G) = \{a^n b^n \mid n \in \mathbb{N}\}$. Moreover, G is obviously not a parenthesis grammar. Now we consider the Γ-cfg

$$G' = (\{S'\}, S', \{\ S' \to \langle\rangle\ ,\ S' \to \langle aS'b \rangle\ \})\ .$$

Then, for each $n \in \mathbb{N}$, we have

$$S' \Rightarrow_{G'} \langle aS'b \rangle \Rightarrow^*_{G'} \langle a \langle \cdots \langle aS'b \rangle \cdots \rangle b \rangle \Rightarrow_{G'} \langle a \langle \cdots \langle a \langle\rangle b \rangle \cdots \rangle b \rangle\ ,$$

where both a and b occur n times. In particular, $\mathrm{L}(G')$ contains the string "$\langle\rangle$". Clearly, G' is a parenthesis grammar, and $\mathrm{L}(G')$ is balanced and has bounded associates.

Lemma 3.6. [8, Cor. 4] It is decidable, for arbitrary Γ-cfg G_1 and parenthesis grammar G_2, whether $\mathrm{L}(G_1) \subseteq \mathrm{L}(G_2)$.

Theorem 3.7. *cf.* [8, Thm. 4] The following statements hold true.

1. A context-free language is balanced and has bounded associates iff it is a parenthesis language.
2. For each Γ-cfg G, if $\mathrm{L}(G)$ is a parenthesis language, then we can effectively construct a Γ-parenthesis grammar G' such that $\mathrm{L}(G') = \mathrm{L}(G)$.

The next result is an easy consequence of Theorem 3.7(1).

Corollary 3.8. Let G be a Γ-parenthesis grammar and $L \subseteq \mathrm{L}(G)$ a context-free language. Then L is a parenthesis language.

Proof. Since $\mathrm{L}(G)$ is a parenthesis language, by Theorem 3.7(1), $\mathrm{L}(G)$ is balanced and has bounded associates. Clearly, also L is balanced. Moreover, since $\mathrm{L}(G)$ has bounded associates, there exists a constant $K \in \mathbb{N}_+$ such that for all $w = uav$ in $\mathrm{L}(G)$ with $u, v \in \Gamma^*$ and $a \in \Gamma$, the terminal a has at most K associates. Since $L \subseteq \mathrm{L}(G)$, for all $w = uav$ in L with $u, v \in \Gamma^*$ and $a \in \Gamma$, the terminal a has at most K associates, *i.e.*, also L has bounded associates. Hence, by Theorem 3.7(1), L is a parenthesis language as well. □

Now we define a new subclass of context-free grammars, which we call tree generating context-free grammars. Formally, for each Σ^Ξ-cfg G, we say that G is *tree generating* if $\mathrm{L}(G) \subseteq \mathrm{T}_\Sigma$.

In the next example we give a tree generating Σ^Ξ-cfg.

Example 3.9. Let $\Sigma = \{\omega^{(3)}, \beta^{(0)}\}$. We consider the Σ^Ξ-cfg

$$G = (\{S, A, B, C\}, S, R) \ ,$$

where

$$R = \{ \ S \to ASB\rangle \ , \ S \to ACB\rangle \ , \ A \to \omega\langle C\# \ , \ B \to \#C \ , \ C \to \beta\langle\rangle \ \} \ .$$

Then we have, *e.g.*,

$$\begin{aligned} S &\Rightarrow ASB\rangle \Rightarrow \omega\langle C\#SB\rangle \Rightarrow \omega\langle\beta\langle\rangle\#SB\rangle \\ &\Rightarrow \omega\langle\beta\langle\rangle\#ACB\rangle B\rangle \Rightarrow \omega\langle\beta\langle\rangle\#\omega\langle C\#CB\rangle B\rangle \\ &\Rightarrow \omega\langle\beta\langle\rangle\#\omega\langle\beta\langle\rangle\#CB\rangle B\rangle \Rightarrow \omega\langle\beta\langle\rangle\#\omega\langle\beta\langle\rangle\#\beta\langle\rangle B\rangle B\rangle \\ &\Rightarrow \omega\langle\beta\langle\rangle\#\omega\langle\beta\langle\rangle\#\beta\langle\rangle\#C\rangle B\rangle \Rightarrow \omega\langle\beta\langle\rangle\#\omega\langle\beta\langle\rangle\#\beta\langle\rangle\#\beta\langle\rangle\rangle B\rangle \\ &\Rightarrow \omega\langle\beta\langle\rangle\#\omega\langle\beta\langle\rangle\#\beta\langle\rangle\#\beta\langle\rangle\rangle\#C\rangle \Rightarrow \omega\langle\beta\langle\rangle\#\omega\langle\beta\langle\rangle\#\beta\langle\rangle\#\beta\langle\rangle\rangle\#\beta\langle\rangle\rangle \ . \end{aligned}$$

Evidently, $\mathrm{L}(G) \subseteq \mathrm{T}_\Sigma$, hence G is tree generating.

3.2. Regular tree grammars

A *regular tree grammar over* Σ (for short: Σ-*rtg*) [2, 4, 5] is a Σ^Ξ-cfg $\mathcal{G} = (N, S, R)$ such that each rule in R has the form $A \to \eta$ with $\eta \in \mathrm{T}_\Sigma(N)$. Obviously, if $A \Rightarrow^* \xi$ for some $\xi \in (\Sigma^\Xi)^*$, then $\xi \in \mathrm{T}_\Sigma$.

The Σ-*tree language generated by* $\mathcal{G}$ is the set

$$\mathrm{L}(\mathcal{G}) = \{\xi \in \mathrm{T}_\Sigma \mid S \Rightarrow^* \xi\} \ .$$

We call each $L \subseteq \mathrm{T}_\Sigma$ *regular* if there exists a Σ-rtg $\mathcal{G}$ such that $\mathrm{L}(\mathcal{G}) = L$. Observe that each Σ-rtg is a tree generating context-free grammar.

Example 3.10. Let $\Sigma = \{\omega^{(3)}, \beta^{(0)}\}$. We consider the Σ-rtg $\mathcal{G} = (\{S\}, S, R)$, where $R = \{\ S \to \omega\langle\beta\langle\rangle\#\beta\langle\rangle\#\beta\langle\rangle\rangle\ ,\ S \to \omega\langle\beta\langle\rangle\#S\#\beta\langle\rangle\rangle\ \}$. Fig. 1 shows, for each $n \in \mathbb{N}_+$, the tree ξ_n and the derivation of $\mathcal{G}$ for ξ_n. In fact, $\mathrm{L}(\mathcal{G}) = \{\xi_n \mid n \in \mathbb{N}_+\}$. One can show that, for the tree generating Σ^Ξ-cfg G defined in Example 3.9, we have $\mathrm{L}(\mathcal{G}) = \mathrm{L}(G)$.

4. Sequential transducers

To prove our results in the next section, it is necessary to recall the concept of sequential transducer and the Sequential Transducer Theorem.

Let Γ and Δ be two alphabets. A (Γ, Δ)-*sequential transducer* (or just sequential transducer) [6] is a tuple $\mathcal{S} = (Q, q_0, \delta)$ where Q is a finite nonempty set (*states*), $q_0 \in Q$ (*start state*), and δ is a finite subset of $Q \times \Gamma^* \times \Delta^* \times Q$ (*transitions*).

Let $\mathcal{S} = (Q, q_0, \delta)$ be a (Γ, Δ)-sequential transducer. For all $w \in \Gamma^*$ and $u \in \Delta^*$, we have $u \in \mathcal{S}(w)$ iff there exist $k \in \mathbb{N}$, $w_1, \ldots, w_k \in \Gamma^*$, $u_1, \ldots, u_k \in \Delta^*$, and $q_1, \ldots, q_k \in Q$ such that $w = w_1 \cdots w_k$, $u = u_1 \cdots u_k$, and $(q_{i-1}, w_i, u_i, q_i) \in \delta$ for each $i \in [k]$. Moreover, for every $L \subseteq \Gamma^*$, we have

$$\mathcal{S}(L) = \bigcup_{w \in L} \mathcal{S}(w) \ .$$

We call a binary relation $\varphi \subseteq \Gamma^* \times \Delta^*$ a (Γ, Δ)-*transduction* (or just: transduction) if there exists a (Γ, Δ)-sequential transducer $\mathcal{S}$ such that $\mathcal{S}(w) = \varphi(w)$ for every $w \in \Gamma^*$.

Lemma 4.1. [6, Thm. 6.4.3] (The Sequential Transducer Theorem) Let $L \subseteq \Gamma^*$ be a context-free language and $\mathcal{S}$ be a (Γ, Δ)-sequential transducer. Then $\mathcal{S}(L) \subseteq \Delta^*$ is a context-free language as well.

5. Results

In this section we answer questions (Q1) and (Q2), which we proposed in the Introduction. To answer these questions the following steps are necessary.

Let $\varphi : (\Sigma^\Xi)^* \to (\Sigma^\Xi)^*$ be the mapping such that, for each string $w \in (\Sigma^\Xi)^*$, the mapping φ replaces every occurrence of $\sigma\langle$ in w into $\langle\sigma$ simultaneously for all $\sigma \in \Sigma$. Formally, for every string

$$w = v_0\sigma_1\langle v_1 \cdots \sigma_k\langle v_k \text{ over } \Sigma^\Xi$$

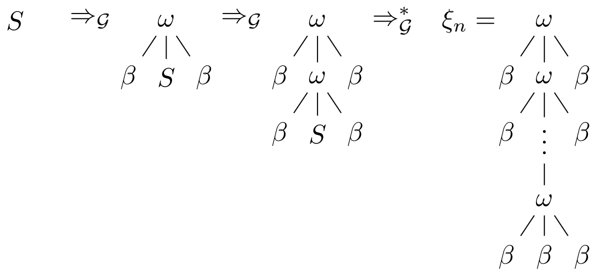

Figure 1. A derivation of the Σ-rtg $\mathcal{G}$ defined in Example 3.10 for $n \in \mathbb{N}_+$ and ξ_n, where ξ_n is the tree in which the symbol ω occurs n times.

Table 2. The illustration of the mapping φ.

w	$\varphi(w)$
$\omega\langle\beta\langle\rangle\#\beta\langle\rangle\#\beta\langle\rangle\rangle$	$\langle\omega\langle\beta\rangle\#\langle\beta\rangle\#\langle\beta\rangle\rangle$
$\omega\beta\langle\rangle\omega\langle\#$	$\omega\langle\beta\rangle\langle\omega\#$
$\langle\langle\rangle\rangle\langle\rangle$	$\langle\langle\rangle\rangle\langle\rangle$
$\langle\langle\#\rangle\langle\langle\langle\#\rangle$	$\langle\langle\#\rangle\langle\langle\langle\#\rangle$

with $k \in \mathbb{N}$, $v_0, v_1, \ldots, v_k \in (\Sigma^\Xi)^*$, $\sigma_1, \ldots, \sigma_k \in \Sigma$ such that, for each $i \in \{0, \ldots, k\}$, there do not exist $u, v \in (\Sigma^\Xi)^*$ and $\sigma \in \Sigma$ such that $v_i = u\sigma\langle v$, we have

$$\varphi(w) = v_0\langle\sigma_1 v_1 \cdots \langle\sigma_k v_k \quad .$$

Example 5.1. Let $\Sigma = \{\omega^{(3)}, \beta^{(0)}\}$. Table 2 shows $\varphi(w)$ for some particular w over Σ^Ξ.

Now we give a (Σ^Ξ, Σ^Ξ)-sequential transducer $\mathcal{S}$ such that, for all strings w over Σ^Ξ, we have $\mathcal{S}(w) = \varphi(w)$. Fig. 2 depicts that sequential transducer $\mathcal{S} = (\{p, q\}, p, \delta)$ as follows. We represent every state $q' \in \{p, q\}$ as a circle with q' in its center, the start state p by an ingoing directed edge with the label "start", and each transition $(p', u, v, q') \in \delta$ by a directed edge from p' to q' with the label u/v. In order to make our figure compact, we add the quantifications "$(\forall\sigma \in \Sigma)$:", "$(\forall a \in \Xi)$:", or "$(\forall a \in \Xi \backslash \{\langle\})$:" to omit a few edges. Furthermore, the label of the edge from q to p consists of two lines representing concisely that $(q, \sigma\langle, \langle\sigma, p) \in \delta$ for every $\sigma \in \Sigma$ and $(q, a, a, p) \in \delta$ for each a in $\Xi \setminus \{\langle\}$, respectively. Observe that, for each w over Σ^Ξ, the set $\mathcal{S}(w)$ is a singleton set, and thus, we sometimes identify $\mathcal{S}(w)$ with its one and only element.

The following result shows that φ is a (Σ^Ξ, Σ^Ξ)-transduction.

Lemma 5.2. For each w over Σ^Ξ, we have $\mathcal{S}(w) = \varphi(w)$.

Proof. We prove our statement by induction on the length of w. Clearly, for each w in $\Sigma^\Xi \cup \{\varepsilon\} \cup \{\sigma\langle \mid \sigma \in \Sigma\}$, we have $\mathcal{S}(w) = \varphi(w)$.

Now let $w = w'b$ for some $w' \in (\Sigma^\Xi)^*$ and $b \in \Sigma^\Xi$. By I.H., we may assume that $\mathcal{S}(w') = \varphi(w')$. By the construction of $\mathcal{S}$, there exist $k \in \mathbb{N}_+$ and $w_1, \dots, w_k \in (\Sigma^\Xi)^*$ such that $w' = w_1 \cdots w_k$ and $1 \le \mathrm{len}(w_i) \le 2$ for all $i \in [k]$. Furthermore, there exist $u_1, \dots, u_k \in (\Sigma^\Xi)^*$ and $q_0, q_1, \dots, q_k \in \{p, q\}$ such that $\mathcal{S}(w') = u_1 \cdots u_k$, $q_0 = p$, and $(q_{i-1}, w_i, u_i, q_i) \in \delta$ for each $i \in [k]$. We consider the next cases.

Assume that $w_k = a\sigma$ for some a in $\Sigma^\Xi \cup \{\varepsilon\}$ and $\sigma \in \Sigma$ and $b = \langle$. We have $(q_{k-1}, a, a, q) \in \delta$ if $a \in \Sigma$; and $(q_{k-1}, a, a, p) \in \delta$ if ($a \in \Xi$ and $q_{k-1} = p$) or ($a \in \Xi \setminus \{\langle\}$ and $q_{k-1} = q$). Since $\mathcal{S}(w') = \varphi(w')$, we may assume that $q_{k-1} \neq q$ or $a \neq \langle$. Moreover, both $(p, \sigma\langle, \langle\sigma, p) \in \delta$ and $(q, \sigma\langle, \langle\sigma, p) \in \delta$. Hence, $\mathcal{S}(w) = u_1 \cdots u_{k-1} a \langle\sigma$, and furthermore, $\mathcal{S}(w) = \varphi(w)$.

Otherwise, *i.e.*, $w_k \neq a\sigma$ or $b \neq \langle$, we have $(q_k, b, b, q') \in \delta$ for some $q' \in \{p, q\}$, and thus, $\mathcal{S}(w) = \varphi(w)$. □

The next result is an immediate consequence of Lemma 4.1 using the (Σ^Ξ, Σ^Ξ)-sequential transducer $\mathcal{S}$ given at the beginning of this section.

Corollary 5.3. Let G be a Σ^Ξ-cfg. There exists a Σ^Ξ-cfg $G_\mathcal{S}$ such that $\mathrm{L}(G_\mathcal{S}) = \mathcal{S}(\mathrm{L}(G))$.

Next we show that $\mathcal{S}(\mathrm{T}_\Sigma)$ is a parenthesis language by constructing a Σ^Ξ-parenthesis grammar G_Σ such that $\mathrm{L}(G_\Sigma) = \mathcal{S}(\mathrm{T}_\Sigma)$. Let $G_\Sigma = (\{S\}, S, R)$ be the Σ^Ξ-cfg such that

$$R = \{S \to \langle\sigma \underbrace{S\#S\#\dots\#S}_{k\text{-times}}\rangle \mid k \in \mathbb{N}, \sigma \in \Sigma^{(k)}\} \ .$$

Clearly, G_Σ is a parenthesis grammar.

Lemma 5.4. $\mathrm{L}(G_\Sigma) = \mathcal{S}(\mathrm{T}_\Sigma)$.

Proof. It is sufficient to prove that, for each $w \in (\Sigma^\Xi)^*$, the following statements are equivalent.

1. $S \Rightarrow^*_{G_\Sigma} w$.
2. There exists $\xi \in \mathrm{T}_\Sigma$ such that $w = \mathcal{S}(\xi)$.

($1 \Rightarrow 2$). We prove it by induction on the length of the derivation. If $S \Rightarrow_{G_\Sigma} w$, then $w = \langle\alpha\rangle$ for some $\alpha \in \Sigma^{(0)}$, and, clearly, for $\xi = \alpha\langle\rangle$, we have $\langle\alpha\rangle = \mathcal{S}(\alpha\langle\rangle)$.

Now assume that $S \Rightarrow^{n+1}_{G_\Sigma} w$ for some $n \in \mathbb{N}$. This derivation can be written in the form

$$S \Rightarrow_{G_\Sigma} \langle\sigma S\#S\#\dots\#S\rangle \Rightarrow^*_{G_\Sigma} \langle\sigma w_1\#w_2\#\dots\#w_k\rangle = w \ ,$$

where $S \to \langle\sigma S\#S\#\dots\#S\rangle$ is in R, and by I.H., for each $i \in [k]$, there exists $\xi_i \in \mathrm{T}_\Sigma$ such that $w_i = \mathcal{S}(\xi_i)$. Then, for the tree $\xi = \sigma\langle\xi_1\#\xi_2\#\dots\#\xi_k\rangle$, we have $w = \mathcal{S}(\xi)$.

($2 \Rightarrow 1$). We prove it by structural induction on ξ. If $\xi = \alpha\langle\rangle$ for some $\alpha \in \Sigma^{(0)}$, then $w = \langle\alpha\rangle$. Since $S \to \langle\alpha\rangle$ is in R, we have $S \Rightarrow_{G_\Sigma} w$.

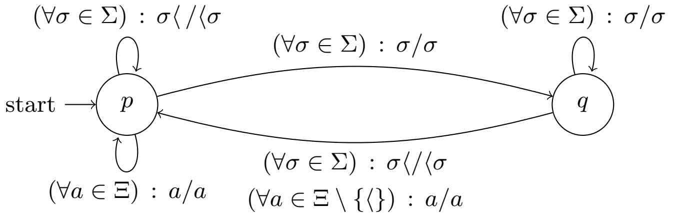
$(\forall\sigma\in\Sigma)\,:\,\sigma\langle/\langle\sigma$
$(\forall\sigma\in\Sigma)\,:\,\sigma/\sigma$
$(\forall\sigma\in\Sigma)\,:\,\sigma/\sigma$
start
p
q
$(\forall a\in\Xi)\,:\,a/a$
$(\forall\sigma\in\Sigma)\,:\,\sigma\langle/\langle\sigma$
$(\forall a\in\Xi\setminus\{\langle\})\,:\,a/a$

Figure 2. Illustration of the $(\Sigma^{\Xi},\Sigma^{\Xi})$-sequential transducer $\mathcal{S}$ given at the beginning of Section 5.

Now let $\xi=\sigma\langle\xi_1\#\xi_2\#\dots\#\xi_k\rangle$ for some $k\in\mathbb{N}_+$, $\sigma\in\Sigma^{(k)}$, and $\xi_1,\xi_2,\dots,\xi_k\in\mathrm{T}_\Sigma$. Observe that we have $\mathcal{S}(\xi)=\langle\sigma\mathcal{S}(\xi_1)\#\mathcal{S}(\xi_2)\#\dots\#\mathcal{S}(\xi_k)\rangle$. By I.H., for each $i\in[k]$, we have $S\Rightarrow^*_{G_\Sigma}\mathcal{S}(\xi_i)$. Since the rule $S\to\langle\sigma S\#S\#\dots\#S\rangle$ is in R, we have

$$S\Rightarrow_{G_\Sigma}\langle\sigma S\#S\#\dots\#S\rangle\Rightarrow^*_{G_\Sigma}\langle\sigma\mathcal{S}(\xi_1)\#\mathcal{S}(\xi_2)\#\dots\#\mathcal{S}(\xi_k)\rangle=\mathcal{S}(\xi)=w\ .$$

□

Now we are ready to answer question (Q1) as follows.

Theorem 5.5. *It is decidable, for an arbitrary Σ^{Ξ}-cfg G, whether G is tree generating.*

Proof. By Corollary 5.3, there exists a Σ^{Ξ}-cfg $G_{\mathcal{S}}$ such that $\mathrm{L}(G_{\mathcal{S}})=\mathcal{S}(\mathrm{L}(G))$. Then we have

$$\mathrm{L}(G)\subseteq\mathrm{T}_\Sigma\quad\text{iff}\quad\mathcal{S}(\mathrm{L}(G))\subseteq\mathcal{S}(\mathrm{T}_\Sigma)\quad\text{iff}\quad\mathrm{L}(G_{\mathcal{S}})\subseteq\mathrm{L}(G_\Sigma),\tag{5.1}$$

where the second equivalence follows from Lemma 5.4. By Lemma 3.6 (for $G_1=G_{\mathcal{S}}$ and $G_2=G_\Sigma$), it is decidable whether $\mathrm{L}(G_{\mathcal{S}})\subseteq\mathrm{L}(G_\Sigma)$. Hence, by (5.1), it is decidable whether $\mathrm{L}(G)\subseteq\mathrm{T}_\Sigma$ as well. □

Built upon the preceding result, we give an answer to question (Q2).

Theorem 5.6. *Let G be a Σ^{Ξ}-cfg such that G is tree generating. We can effectively construct a Σ-rtg $\mathcal{G}$ such that $\mathrm{L}(\mathcal{G})=\mathrm{L}(G)$.*

Proof. If G is a Σ-rtg, then we let $\mathcal{G}=G$ and we are done, otherwise we proceed as follows.

By Corollary 5.3, there exists a Σ^{Ξ}-cfg $G_{\mathcal{S}}$ such that $\mathrm{L}(G_{\mathcal{S}})=\mathcal{S}(\mathrm{L}(G))$. Moreover, by (5.1), we have $\mathrm{L}(G_{\mathcal{S}})\subseteq\mathrm{L}(G_\Sigma)$.

Since $\mathrm{L}(G_\Sigma)$ is a parenthesis language, by Corollary 3.8, also $\mathrm{L}(G_{\mathcal{S}})$ is a parenthesis language. By Theorem 3.7(2), we can effectively construct a Σ^{Ξ}-parenthesis grammar $G'=(N',S',R')$ such that $\mathrm{L}(G')=\mathrm{L}(G_{\mathcal{S}})$. Recall that, since G' is a parenthesis grammar, each rule in R' has the form $A\to\langle\theta\rangle$ such that $A\in N'$ and

θ is a string over $N' \cup \Sigma \cup \{\#\}$. We note that, by Observation 3.2, G' is chain-free and ε-free. Furthermore, by Lemma 3.1, we may assume that G' is reduced.

Let $A \in N'$, θ be a string over $N' \cup \Sigma \cup \{\#\}$, and $\xi = \sigma\langle\xi_1\#\xi_2\#\dots\#\xi_k\rangle$ in T_Σ for some $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, and $\xi_1, \xi_2, \dots, \xi_k \in \mathrm{T}_\Sigma$. We claim that

$$\begin{aligned}&\text{if } A \Rightarrow_{G'} \langle\theta\rangle \Rightarrow^*_{G'} \mathcal{S}(\xi) \ \text{, then } \theta = \sigma A_1\#A_2\#\dots\#A_k \\ &\text{for some } A_1, A_2, \dots, A_k \in N' \text{ with } A_i \Rightarrow^*_{G'} \mathcal{S}(\xi_i) \text{ for all } i \in [k] \ .\end{aligned} \tag{5.2}$$

Now we prove (5.2). Since G' is a parenthesis grammar, by Lemma 3.4, there do not exist $B \in N'$ and $\gamma \in \mathrm{prefix}(\sigma\mathcal{S}(\xi_1)\#\mathcal{S}(\xi_2)\#\dots\#\mathcal{S}(\xi_k))$ such that $B \Rightarrow^*_{G'} \gamma$, and thus, $\theta = \sigma\theta'$ for some string θ' over $N' \cup \Sigma \cup \{\#\}$. We proceed by case analysis.

Assume that $k = 0$. Then $\sigma = \alpha$ and $\xi = \alpha\langle\rangle$ for some $\alpha \in \Sigma^{(0)}$, and hence, $\mathcal{S}(\alpha\langle\rangle) = \langle\alpha\rangle$. Furthermore, since G' is a parenthesis grammar, we have $\theta = \alpha$ and $\theta' = \varepsilon$.

Now assume that $k > 0$. Then, since $\langle$ is in $\mathrm{prefix}(\mathcal{S}(\xi_1)\#\mathcal{S}(\xi_2)\#\dots\#\mathcal{S}(\xi_k))$ and G' is a parenthesis grammar, we must have $\theta' = A_1\theta''$ for some $A_1 \in N'$ and string θ'' over $N' \cup \Sigma \cup \{\#\}$. Since G' is a parenthesis grammar, by Lemma 3.4, for all $w \in (\Sigma^\Xi)^*$, if $A_1 \Rightarrow^*_{G'} w$, then $w = \langle u\rangle$ for some $u \in (\Sigma^\Xi)^*$ such that u is balanced. The one and only way to satisfy the aforementioned requirement on A_1 with respect to $A \Rightarrow_{G'} \langle\sigma A_1\theta''\rangle \Rightarrow^*_{G'} \mathcal{S}(\xi)$ is that if $A_1 \Rightarrow^*_{G'} \mathcal{S}(\xi_1)$. (Observe that, since G' is a parenthesis grammar, we have $A_1 \Rightarrow_{G'} \langle\theta_1\rangle \Rightarrow^*_{G'} \mathcal{S}(\xi_1)$ for some string θ_1 over $N' \cup \Sigma \cup \{\#\}$, which satisfies the condition of (5.2) as well.) Then, since G' is a parenthesis grammar, by Lemma 3.4, there do not exist $C \in N'$ and $v \in \mathrm{prefix}(\#\mathcal{S}(\xi_2)\#\dots\#\mathcal{S}(\xi_k))$ such that $C \Rightarrow^*_{G'} v$, and hence, $\theta'' = \#\hat{\theta}$ for some string $\hat{\theta}$ over $N' \cup \Sigma \cup \{\#\}$. Putting these together, we currently have $\theta = \langle\sigma A_1\#\hat{\theta}\rangle$. Clearly, by continuing our argumentation in a similar way, we can show that $\theta = \sigma A_1\#A_2\#\dots\#A_k$ and that $A_i \Rightarrow^*_{G'} \mathcal{S}(\xi_i)$ for all $i \in [k]$. This completes the proof of (5.2).

It follows from (5.2) that each rule in R' has the form $A \to \langle\sigma A_1\#A_2\#\dots\#A_k\rangle$ with $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, and $A, A_1, A_2, \dots, A_k \in N'$.

Next we can effectively construct the Σ-rtg $\mathcal{G} = (N', S', R'')$ such that $A \to \sigma\langle A_1\#A_2\#\dots\#A_k\rangle$ is in R'' iff $A \to \langle\sigma A_1\#A_2\#\dots\#A_k\rangle$ is in R'.

We claim that, for all $A \in N'$ and $\xi \in \mathrm{T}_\Sigma$, we have

$$A \Rightarrow^*_{G'} \mathcal{S}(\xi) \quad \text{iff} \quad A \Rightarrow^*_{\mathcal{G}} \xi \ . \tag{5.3}$$

Next we prove (5.3) by structural induction on ξ. Let $\xi = \alpha\langle\rangle$ for some $\alpha \in \Sigma^{(0)}$. Clearly, we have $\mathcal{S}(\alpha\langle\rangle) = \langle\alpha\rangle$. Moreover, we have

$$A \Rightarrow^*_{G'} \langle\alpha\rangle \quad \text{iff} \quad A \to \langle\alpha\rangle \text{ is in } R' \quad \text{iff} \quad A \to \alpha\langle\rangle \text{ is in } R'' \quad \text{iff} \quad A \Rightarrow^*_{\mathcal{G}} \alpha\langle\rangle \ .$$

Now let $\xi = \sigma\langle\xi_1\#\xi_2\#\dots\#\xi_k\rangle$ with $k \in \mathbb{N}_+$, $\sigma \in \Sigma^{(k)}$, and $\xi_1, \xi_2, \dots, \xi_k \in \mathrm{T}_\Sigma$. For every $A_1, A_2, \dots, A_k \in N'$, the rule $A \to \langle\sigma A_1\#A_2\#\dots\#A_k\rangle$ exists in R' iff the rule $A \to \sigma\langle A_1\#A_2\#\dots\#A_k\rangle$ exists is in R''. Moreover, by I. H., for each $i \in [k]$, we have $A_i \Rightarrow^*_{G'} \mathcal{S}(\xi_i)$ iff $A_i \Rightarrow^*_{\mathcal{G}} \xi_i$. So, we have

$$A \Rightarrow_{G'} \langle\sigma A_1\#A_2\#\dots\#A_k\rangle \Rightarrow^*_{G'} \langle\sigma\mathcal{S}(\xi_1)\#A_2\#\dots\#A_k\rangle$$

$$\Rightarrow^*_{G'} \langle \sigma \mathcal{S}(\xi_1)\#\mathcal{S}(\xi_2)\# \dots \#\mathcal{S}(\xi_k)\rangle = \mathcal{S}(\xi)$$

if and only if

$$A \Rightarrow_{\mathcal{G}} \sigma\langle A_1\#A_2\# \dots \#A_k\rangle \Rightarrow^*_{\mathcal{G}} \sigma\langle \xi_1\#A_2\# \dots \#A_k\rangle$$
$$\Rightarrow^*_{\mathcal{G}} \sigma\langle \xi_1\#\xi_2\# \dots \#\xi_k\rangle = \xi \ .$$

Therefore, for each $\xi \in \mathrm{T}_\Sigma$, we have

$$\mathcal{S}(\xi) \in \mathrm{L}(G') \text{ iff } S' \Rightarrow^*_{G'} \mathcal{S}(\xi) \text{ iff}^{(*)} S' \Rightarrow^*_{\mathcal{G}} \xi \text{ iff } \xi \in \mathrm{L}(\mathcal{G}) \ ,$$

where at $(*)$ we used the fact that $S' \Rightarrow^*_{G'} \mathcal{S}(\xi)$ iff $S' \Rightarrow^*_{\mathcal{G}} \xi$ by (5.3). □

Acknowledgements. The author would like to thank Joost Engelfriet for highlighting the close connection between parenthesis languages and tree languages and for pointing out the paper [8], and the anonymous referees of this paper for their work and useful suggestions.

References

[1] J. Berstel, L. Boasson: *Formal properties of XML grammars and languages*, Acta Inform. 38 (2002), pp. 649–671, doi: https://doi.org/10.1007/s00236-002-0085-4.

[2] W. S. Brainerd: *Tree Generating Regular Systems*, Inf. Control 14.2 (1969), pp. 217–231, doi: https://doi.org/10.1016/S0019-9958(69)90065-5.

[3] N. Chomsky: *Three models for the description of language*, IEEE Trans. Inf. Theory 2.3 (1956), pp. 113–124, doi: https://doi.org/10.1109/TIT.1956.1056813.

[4] J. Engelfriet: *Tree automata and tree grammars*, tech. rep. DAIMI FN-10, see also: arXiv:1510.02036v1 [cs.FL] 7 Oct 2015, Inst. of Mathematics, University of Aarhus, Department of Computer Science, Denmark, 1975.

[5] F. Gécseg, M. Steinby: *Tree Automata*, see also: arXiv:1509.06233v1 [cs.FL] 21 Sep 2015, Akadémiai Kiadó, Budapest, 1984.

[6] M. Harrison: *Introduction to Formal Language Theory*, Addison-Wesley, 1978.

[7] J. Hopcroft, J. Ullman: *Introduction to automata theory, languages, and computation*, Addison-Wesley, 1979.

[8] D. E. Knuth: *A Characterization of Parenthesis Languages*, Inf. Control 11.3 (1967), pp. 269–289, doi: https://doi.org/10.1016/S0019-9958(67)90564-5.

[9] R. McNaughton: *Parenthesis Grammars*, J. Assoc. Computing Mach. 14 (1967), pp. 490–500.

[10] J. Mezei, J. B. Wright: *Algebraic automata and context-free sets*, Inf. Control 11.1-2 (1967), pp. 3–29, doi: https://doi.org/10.1016/S0019-9958(67)90353-1.

[11] W. C. Rounds: *Tree-oriented proofs of some theorems on context-free and indexed languages*, in: Proceedings of the second annual ACM symposium on Theory of computing, 1970, pp. 109–116, doi: https://doi.org/10.1145/800161.805156.

[12] J. Thatcher: *Tree automata: an informal survey*, in: Currents in the Theory of Computing, ed. by A. V. Aho, Englewood Cliffs, N. J.: Prentice-Hall, 1973, pp. 143–172.