Discussiones Mathematicae
General Algebra and Applications 36 (2016) 179–208
doi:10.7151/dmgaa.1259

VARIETIES OF REGULAR ALGEBRAS AND UNRANKED TREE LANGUAGES

Magnus Steinby, Eija Jurvanen

Department of Mathematics and Statistics
University of Turku, FI-20014 Turku

e-mail: steinby@utu.fi, jurvanen@utu.fi

and

Antonio Cano

Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia, E-46020 – Valencia

e-mail: antonio.cano.gomez@outlook.com

Abstract

In this paper we develop a variety theory for unranked tree languages and unranked algebras. In an unranked tree any symbol may label a node with any number of successors. Such trees appear in markup languages such as XML and as syntactic descriptions of natural languages. In the corresponding algebras each operation is defined for any number of arguments, but in the regular algebras used as tree recognizers the operations are finite-state computable. We develop the basic theory of regular algebras for a setting in which algebras over different operator alphabets are considered together. Using syntactic algebras of unranked tree languages we establish a bijection between varieties of unranked tree languages and varieties of regular algebras. As varieties of unranked tree languages are usually defined by means of congruences of term algebras, we introduce also varieties of congruences and a general device for defining such varieties. Finally, we show that the natural unranked counterparts of several varieties of ranked tree languages form varieties in our sense.

Keywords: unranked algebras, unranked trees, regular algebras, syntactic algebras, varieties of tree languages, varieties of regular algebras.

2010 Mathematics Subject Classification: 08A70, 68Q70, 08B99.

1. Introduction

In its prevalent form the theory of tree automata and tree languages (cf. [8, 12, 13]) deals with trees in which the nodes are labeled with symbols from a ranked alphabet that may be viewed as a finite set of operation symbols. When trees are defined as terms, finite tree automata become essentially finite algebras. Hence, Universal Algebra offers the theory a good formal framework and there is also a link to term rewriting. However, when trees are used as representations of XML documents or of parses of sentences of a natural language, fixing the possible ranks of symbols is awkward. It is in particular the study of XML that propels the current work on unranked tree languages (cf. [5, 8, 16, 17, 22]).

Actually unranked trees appeared in the theory of tree languages already in the 1960s when Thatcher [28] defined recognizable unranked tree languages and established a connection between them and context-free grammars. Recognizability was defined using “pseudoautomata” (attributed to J.R. Büchi and J.B. Wright). Here pseudoautomata reappear as our “regular algebras”. In [18] Pair and Quere consider hedges, i.e., finite sequences of unranked trees, and hedge languages recognized by a new type of algebras. Hedges have become a much used notion (cf. [4, 5, 27]), but we consider just unranked trees.

In Section 2 we introduce some general notation and unranked trees. Besides an unranked alphabet, called the operator alphabet, we also use a leaf alphabet for labeling leaves. The use of two alphabets is natural also in typical applications. In Section 3 we develop the basic theory of unranked algebras allowing algebras to have different operator alphabets. The next section introduces regular algebras in terms of which recognizability will be defined. We derive a representation for the variety of regular algebras (VRA) generated by a given class of regular algebras akin to Tarski’s classical HSP-theorem (cf. [3, 6, 7, 10]). At the end of Section 4 we introduce congruences that reflect the regularity of algebras.

In Section 5 we introduce syntactic congruences and syntactic algebras of subsets of unranked algebras. For unranked tree languages these notions are then obtained by viewing them as subsets of term algebras. In Section 6 we define the recognizable unranked tree languages. An unranked tree language is recognizable if and only if its syntactic algebra is regular, and the syntactic algebra is the least unranked algebra recognizing it. We also show that the syntactic algebra of any effectively given recognizable unranked tree language can be effectively constructed. This is not quite obvious as the operations are infinite objects and there are infinitely many trees of any given height ≥ 1.

In Section 7 we consider varieties of unranked tree languages (VUTs) and varieties of regular congruences (VRCs) by means of which VUTs can be defined. Then we introduce consistent systems of congruences (CSCs) that yield VRCs and VUTs that we call quasi-principal as they replace the principal varieties of the ranked case. In Section 8 we establish a bijection between the VRAs and VUTs. Section 9 contains several examples of VUTs that are natural unranked counterparts of some (general) varieties of ranked tree languages.

Parts of this work parallels the theory of general ranked varieties. Familiarity with the paper [25] may be helpful, but it is not assumed. Several straightforward proofs have been omitted.

2. General preliminaries and unranked trees

We may write $X := Y$ when X is defined to be Y, and $X :\Leftrightarrow Y$ means that X is defined by condition Y. For any integer $n \geq 0$, let $[n] := \{1, \ldots, n\}$. For a relation $\rho \subseteq A \times B$, the fact that $(a, b) \in \rho$ is also expressed by $a\,\rho\,b$. For any $a \in A$ and $A' \subseteq A$, let $a\rho := \{b \mid a\,\rho\,b\}$ and $A'\rho := \{b \in B \mid (\exists a \in A')\,a\,\rho\,b\}$. The *converse* of ρ is $\rho^{-1} := \{(b, a) \mid a\,\rho\,b\}$. The *composition* of $\rho \subseteq A \times B$ and $\rho' \subseteq B \times C$ is $\rho \circ \rho' := \{(a, c) \mid a \in A, c \in C, (\exists b \in B)\,a\,\rho\,b \text{ and } b\,\rho' c\}$. The set of equivalence relations on a set A is denoted by $\mathrm{Eq}(A)$, and for any $\theta \in \mathrm{Eq}(A)$, let $A/\theta := \{a\theta \mid a \in A\}$. Let $\Delta_A := \{(a, a) \mid a \in A\}$ and $\nabla_A := A \times A$. For a mapping $\varphi : A \to B$, the image $\varphi(a)$ of $a \in A$ is also denoted by $a\varphi$, and if $H \subseteq A$ and $K \subseteq B$, we may write $H\varphi$ and $K\varphi^{-1}$ for $\varphi(H)$ and $\varphi^{-1}(K)$, respectively. Especially homomorphisms are treated this way as right operators and the composition of $\varphi : A \to B$ and $\psi : B \to C$ is written as $\varphi\psi$. The *identity map* $A \to A, a \mapsto a$, is denoted by 1_A. For any sets $A_1, \ldots, A_n$ $(n \geq 1)$ and any $i \in [n]$, let π_i denote the i^{th} *projection* $A_1 \times \cdots \times A_n \to A_i$, $(a_1, \ldots, a_n) \mapsto a_i$.

For any finite alphabet X, the set X^* of all *words* (finite strings) over X forms with respect to concatenation the free monoid generated by X in which the *empty word* ε is the identity. Subsets of X^* are called *languages* over X. Here it is convenient to define a language $L \subseteq X^*$ to be *recognizable*, or *regular*, if $L = F\varphi^{-1}$ for a finite monoid M, a homomorphism $\varphi : X^* \to M$ and some $F \subseteq M$. For algebraic expositions of the theory of finite automata and regular languages cf. [11, 20, 21], for example.

The *unranked trees* to be considered are finite and node-labeled, and their branches have a specified left-to-right order. We use two alphabets, an *operator alphabet* and a *leaf alphabet*, for labeling our trees. A symbol from the operator alphabet may label any node of a tree, while the symbols of the leaf alphabet appear at leaves only. In what follows, Σ, Ω, Γ and Ψ denote operator alphabets, and X, Y and Z leaf alphabets. All alphabets are finite, operator alphabets are also nonempty, and leaf alphabets are disjoint from operator alphabets.

The set $T_\Sigma(X)$ of *unranked* ΣX*-trees* is the smallest set T of strings such that (1) $X \cup \Sigma \subseteq T$, and (2) $f(t_1, \ldots, t_m) \in T$ whenever $f \in \Sigma$, $m > 0$ and $t_1, \ldots, t_m \in T$. Subsets of $T_\Sigma(X)$ are called *unranked* ΣX*-tree languages*. Often we speak simply about ΣX*-trees* and ΣX*-tree languages*, or just about (*unranked*) *trees* and (*unranked*) *tree languages* without specifying the alphabets.

Any $u \in X \cup \Sigma$ represents a tree with just one node which is labeled with u, and $f(t_1, \ldots, t_m)$ is interpreted as a tree formed by adjoining the m trees represented by $t_1, \ldots, t_m$ to a new f-labeled root in this left-to-right order. The *height* $\mathrm{hg}(t)$ and the *root* $\mathrm{root}(t)$ of a ΣX-tree t are defined by (1) $\mathrm{hg}(u) = 0$ and $\mathrm{root}(u) = u$ for $u \in \Sigma \cup X$, and (2) $\mathrm{hg}(t) = \max\{\mathrm{hg}(t_1), \ldots, \mathrm{hg}(t_m)\} + 1$ and $\mathrm{root}(t) = f$ for $t = f(t_1, \ldots, t_m)$.

Let ξ be a special symbol not in Σ or X. A ΣX*-context* is a $\Sigma(X \cup \{\xi\})$-tree in which ξ appears exactly once. Let $C_\Sigma(X)$ denote the set of all ΣX-contexts. If $p, q \in C_\Sigma(X)$ and $t \in T_\Sigma(X)$, then $p(q) \in C_\Sigma(X)$ and $p(t) \in T_\Sigma(X)$ are obtained from p by replacing the ξ in it with q and t, respectively.

Let us demonstrate by a couple of examples that the use of two alphabets is quite natural. Figure 1 shows the tree representation of a small XML document. Here `invoices`, `invoice` and `line` belong to the operator alphabet while `text` is a generic name for leaf symbols.

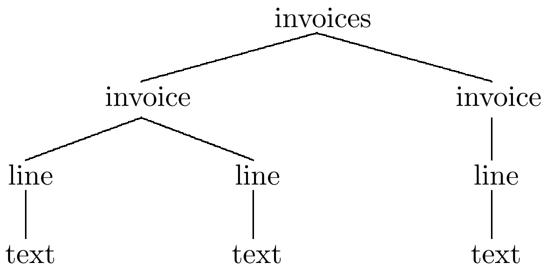

Figure 1. Unranked tree representing the structure of an XML document.

In Example 3.2 of [17] and Example 1 of [9] the unranked trees are Boolean expressions over the alphabet $\{\vee, \wedge, 0, 1\}$ in which disjunctions and conjunctions may appear with any arities. It would be natural to split the alphabet into $\Sigma = \{\vee, \wedge\}$ and $X = \{0, 1\}$; 0 and 1 may label leaves only.

3. Unranked algebras

In the following, the set A of elements of an algebra will also be regarded as an alphabet and the set of all finite sequences of elements of A is denoted by A^*. An m-tuple $(a_1, \ldots, a_m) \in A^m$ $(m \geq 0)$ may be written as the word $a_1 \ldots a_m$ and subsets of A^* are viewed as languages.

Definition 3.1. An *unranked* Σ*-algebra* $\mathcal{A}$ consists of a nonempty set A (of *elements* of $\mathcal{A}$) and an operation $f_\mathcal{A} : A^* \to A$ for each $f \in \Sigma$. We write simply

$\mathcal{A} = (A, \Sigma)$. The algebra $\mathcal{A}$ is *finite* if A is a finite set, and it is *trivial* if A is a singleton. We may also speak just about Σ-*algebras* or (*unranked*) *algebras*.

These are essentially the "pseudoalgebras" of Büchi and Wright (1960) used by Thatcher [28]. In what follows, $\mathcal{A} = (A, \Sigma)$, $\mathcal{B} = (B, \Sigma)$, $\mathcal{B} = (B, \Omega)$, $\mathcal{C} = (C, \Gamma)$, etc. are always unranked algebras with the operator alphabets shown. The classical counterparts of the following concepts, results and proofs, can be found in any universal algebra text such as [3, 6, 7] or [10], for example. The prefix g appearing in some names stands for "generalized".

If $\Omega \subseteq \Sigma$, an Ω-algebra $\mathcal{B} = (B, \Omega)$ is an Ω-*subalgebra* of $\mathcal{A} = (A, \Sigma)$ if $B \subseteq A$ and $f_{\mathcal{B}}(w) = f_{\mathcal{A}}(w)$ for all $f \in \Omega$ and $w \in B^*$. Then we also call $\mathcal{B}$ a *g-subalgebra* of $\mathcal{A}$ without specifying Ω, and B is an Ω-*closed subset* of $\mathcal{A}$, i.e., $f_{\mathcal{A}}(w) \in B$ for all $f \in \Omega$ and $w \in B^*$. Ω-closed subsets are nonempty since $\Omega \neq \emptyset$ and $f_{\mathcal{A}}(\varepsilon) \in B$ for any $f \in \Omega$. A Σ-subalgebra of a Σ-algebra is called just a *subalgebra*.

A pair of mappings $\iota : \Sigma \to \Omega$, $\varphi : A \to B$ forms a *g-morphism* from $\mathcal{A} = (A, \Sigma)$ to $\mathcal{B} = (B, \Omega)$, written as $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$, if

$$f_{\mathcal{A}}(a_1, \ldots, a_m)\varphi = \iota(f)_{\mathcal{B}}(a_1\varphi, \ldots, a_m\varphi)$$

for all $f \in \Sigma$, $m \geq 0$ and $a_1, \ldots, a_m \in A$. A g-morphism is a *g-epimorphism*, a *g-monomorphism* or a *g-isomorphism* if both maps are, respectively, surjective, injective or bijective. Two algebras $\mathcal{A}$ and $\mathcal{B}$ are *g-isomorphic*, $\mathcal{A} \cong_g \mathcal{B}$ in symbols, if there is a g-isomorphism $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$, and $\mathcal{B}$ is a *g-image* of $\mathcal{A}$, if there is a g-epimorphism from $\mathcal{A}$ onto $\mathcal{B}$. Furthermore, we write $\mathcal{A} \preceq_g \mathcal{B}$ if $\mathcal{A}$ is a g-image of a g-subalgebra of $\mathcal{B}$. If $\varphi_* : A^* \to B^*$ is the extension of φ to a monoid morphism, then (ι, φ) is a g-morphism if $f_{\mathcal{A}}(w)\varphi = \iota(f)_{\mathcal{B}}(w\varphi_*)$ for all $f \in \Sigma$ and $w \in A^*$.

For $\mathcal{A} = (A, \Sigma)$ and $\mathcal{B} = (B, \Sigma)$, a *morphism* $\varphi : \mathcal{A} \to \mathcal{B}$ is a map $\varphi : A \to B$ such that $f_{\mathcal{A}}(w)\varphi = f_{\mathcal{B}}(w\varphi_*)$ for all $f \in \Sigma$ and $w \in A^*$. It may be viewed as the g-morphism $(1_\Sigma, \varphi) : \mathcal{A} \to \mathcal{B}$. *Epi-*, *mono-* and *isomorphisms* are defined as usual. The algebras $\mathcal{A}$ and $\mathcal{B}$ are *isomorphic*, $\mathcal{A} \cong \mathcal{B}$, if there is an isomorphism $\varphi : \mathcal{A} \to \mathcal{B}$, and $\mathcal{A} \preceq \mathcal{B}$ if $\mathcal{A}$ is an epimorphic image of some subalgebra of $\mathcal{B}$.

A *g-congruence* of $\mathcal{A} = (A, \Sigma)$ is a pair (σ, θ), where $\sigma \in \mathrm{Eq}(\Sigma)$ and $\theta \in \mathrm{Eq}(A)$, such that for any $f, g \in \Sigma$, $m \geq 0$ and $a_1, \ldots, a_m, b_1, \ldots, b_m \in A$, if $f \,\sigma\, g$ and $a_1 \,\theta\, b_1, \ldots, a_m \,\theta\, b_m$, then $f_{\mathcal{A}}(a_1, \ldots, a_m) \,\theta\, g_{\mathcal{A}}(b_1, \ldots, b_m)$. The set $\mathrm{GCon}(\mathcal{A})$ of all g-congruences on $\mathcal{A}$ contains at least $(\Delta_\Sigma, \Delta_A)$ and (σ, ∇_A), where σ is any equivalence on Σ. With respect to the order defined by

$$(\sigma, \omega) \leq (\sigma', \omega') \;:\Leftrightarrow\; \sigma \subseteq \sigma' \text{ and } \omega \subseteq \omega' \qquad ((\sigma, \omega), (\sigma'\omega') \in \mathrm{GCon}(\mathcal{A})),$$

$\mathrm{GCon}(\mathcal{A})$ forms a complete lattice in which joins and meets are formed componentwise in $\mathrm{Eq}(\Sigma)$ and $\mathrm{Eq}(A)$, respectively. The *congruences* of $\mathcal{A}$ are the equivalences $\theta \in \mathrm{Eq}(A)$ such that $(\Delta_\Sigma, \theta) \in \mathrm{GCon}(\mathcal{A})$. Their set is denoted by $\mathrm{Con}(\mathcal{A})$. Note that $\theta \in \mathrm{Con}(\mathcal{A})$ if $(\omega, \theta) \in \mathrm{GCon}(\mathcal{A})$ for some $\omega \in \mathrm{Eq}(\Sigma)$.

For any g-congruence (σ, θ) of $\mathcal{A} = (A, \Sigma)$, the *g-quotient algebra* $\mathcal{A}/(\sigma, \theta) = (A/\theta, \Sigma/\sigma)$ is defined by setting $(f\sigma)_{\mathcal{A}/(\sigma,\theta)}(a_1\theta, \ldots, a_m\theta) = f_{\mathcal{A}}(a_1, \ldots, a_m)\theta$ for all $f \in \Sigma$, $m \geq 0$, and $a_1, \ldots, a_m \in A$. In particular, $(f\sigma)_{\mathcal{A}/(\sigma,\theta)}(\varepsilon) = f_{\mathcal{A}}(\varepsilon)\theta$. The *quotient algebra* $\mathcal{A}/\theta = (A/\theta, \Sigma)$, where $\theta \in \mathrm{Con}(\mathcal{A})$ and $f_{\mathcal{A}/\theta}(a_1\theta, \ldots, a_m\theta) = f_{\mathcal{A}}(a_1, \ldots, a_m)\theta$ for all $f \in \Sigma$, $m \geq 0$ and $a_1, \ldots, a_m \in A$, may be regarded as a special g-quotient of $\mathcal{A}$; if we identify Σ and Σ/Δ_Σ, then $\mathcal{A}/\theta \cong \mathcal{A}/(\Delta_\Sigma, \theta)$.

The g-subalgebras, g-morphisms, g-congruences and g-quotients of unranked algebras have all the same basic properties as their classical counterparts, and they can be proved the same way. Some of them are listed in the following lemma.

Lemma 3.2. *Let $\mathcal{A} = (A, \Sigma)$, $\mathcal{B} = (B, \Omega)$ and $\mathcal{C} = (C, \Gamma)$ be unranked algebras, and $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$ and $(\varkappa, \psi) : \mathcal{B} \to \mathcal{C}$ be g-morphisms.*

(a) *The product $(\iota\varkappa, \varphi\psi) : \mathcal{A} \to \mathcal{C}$ is also a g-morphism. Moreover, if (ι, φ) and $(\varkappa, \psi)$ are g-epi-, g-mono- or g-isomorphisms, then so is $(\iota\varkappa, \varphi\psi)$.*

(b) *If R is a Ψ-subalgebra of $\mathcal{B}$, with $\Psi \subseteq \Omega$, then $R\varphi^{-1}$ is a $\iota^{-1}(\Psi)$-subalgebra of $\mathcal{A}$.*

(c) *If S is a Ψ-subalgebra of $\mathcal{A}$, with $\Psi \subseteq \Sigma$, then $S\varphi$ is a $\iota(\Psi)$-subalgebra of $\mathcal{B}$.*

(d) *For any g-congruence (σ, θ) of $\mathcal{A}$, the maps $\theta_\natural : A \to A/\theta, a \mapsto a\theta$, and $\sigma_\natural : \Sigma \to \Sigma/\sigma, f \mapsto f\sigma$, define a g-epimorphism $(\sigma_\natural, \theta_\natural) : \mathcal{A} \to \mathcal{A}/(\sigma, \theta)$.*

(e) *The* kernel $\ker(\iota, \varphi) := (\ker\iota, \ker\varphi)$ *of a g-morphism $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$ is a g-congruence of $\mathcal{A}$, and $\mathcal{A}/\ker(\iota, \varphi) \cong_g \mathcal{B}$ if (ι, φ) is a g-epimorphism.*

For any mapping $\varkappa : \Gamma \to \Sigma \times \Omega$, the *$\varkappa$-product* of $\mathcal{A} = (A, \Sigma)$ and $\mathcal{B} = (B, \Omega)$ is the Γ-algebra $\varkappa(\mathcal{A}, \mathcal{B}) = (A \times B, \Gamma)$ defined as follows: For any $f \in \Gamma$, $m \geq 0$ and $(a_1, b_1), \ldots, (a_m, b_m) \in A \times B$, let

$$f_{\varkappa(\mathcal{A},\mathcal{B})}((a_1, b_1), \ldots, (a_m, b_m)) = (g_{\mathcal{A}}(a_1, \ldots, a_m), h_{\mathcal{B}}(b_1, \ldots, b_m)),$$

where $(g, h) = \varkappa(f)$. The products $\varkappa(\mathcal{A}_1, \ldots, \mathcal{A}_n)$ of $n \geq 0$ unranked algebras are defined similarly. Without specifying $\varkappa$, we call any such product a *g-product*. For $n = 0$, the g-product $\varkappa(\mathcal{A}_1, \ldots, \mathcal{A}_n)$ is the appropriate trivial algebra.

The *direct product* $\mathcal{A}_1 \times \ldots \times \mathcal{A}_n$ of Σ-algebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$ may be reconstrued as the g-product $\varkappa(\mathcal{A}_1, \ldots, \mathcal{A}_n)$ with $\varkappa : \Sigma \to \Sigma \times \ldots \times \Sigma$, $f \mapsto (f, \ldots, f)$.

For any mapping $\iota : \Sigma \to \Omega$, the g-product $\iota(\mathcal{B}) = (B, \Sigma)$ of just one factor $\mathcal{B} = (B, \Omega)$, in which $f_{\iota(\mathcal{B})} = \iota(f)_{\mathcal{B}}$ for any $f \in \Sigma$, is called a *g-derived algebra* of $\mathcal{B}$. This notion has similar properties as the derived algebras considered in [10, 15], for example. In particular, we have the following obvious fact.

Lemma 3.3. *If $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$ is a g-morphism from a Σ-algebra $\mathcal{A} = (A, \Sigma)$ to an Ω-algebra $\mathcal{B} = (B, \Omega)$, then $\varphi : \mathcal{A} \to \iota(\mathcal{B})$ is a morphism of Σ-algebras.*

We define *generalized subdirect decompositions* with just finite algebras in mind. A *gsd-representation* of $\mathcal{A} = (A, \Sigma)$ with factors $\mathcal{A}_1 = (A_1, \Sigma_1), \ldots, \mathcal{A}_n = (A_n, \Sigma_n)$ is a g-monomorphism $(\iota, \varphi) : \mathcal{A} \to \varkappa(\mathcal{A}_1, \ldots, \mathcal{A}_n)$, where also $\varkappa : \Gamma \to \Sigma_1 \times \cdots \times \Sigma_n$ is injective, such that $A\varphi\pi_i = A_i$ and $\Sigma\iota\varkappa\pi_i = \Sigma_i$ for every $i \in [n]$. Note that π_i denotes both of the projections $\Sigma_1 \times \cdots \times \Sigma_n \to \Sigma_i$ and $A_1 \times \cdots \times A_n \to A_i$ ($i \in [n]$). Such a gsd-representation is *proper* if for no $i \in [n]$, both $\varphi\pi_i : A \to A_i$ and $\iota\varkappa\pi_i : \Sigma \to \Sigma_i$ are injective. A finite unranked algebra is *gsd-irreducible* if it has no proper gsd-representation.

The next two lemmas and Proposition 3.6 can be proved essentially the same way as their classical counterparts are proved in universal algebra.

Lemma 3.4. *Let $(\iota, \varphi) : \mathcal{A} \to \varkappa(\mathcal{A}_1, \ldots, \mathcal{A}_n)$ be a gsd-representation of an unranked algebra $\mathcal{A} = (A, \Sigma)$ with factors $\mathcal{A}_1 = (A_1, \Sigma_1), \ldots, \mathcal{A}_n = (A_n, \Sigma_n)$ ($n \geq 0$). For every $i \in [n]$, $(\iota\varkappa\pi_i, \varphi\pi_i) : \mathcal{A} \to \mathcal{A}_i$ is a g-epimorphism. Moreover, if we write $(\sigma_i, \theta_i) := \ker(\iota\varkappa\pi_i, \varphi\pi_i)$ for each $i \in [n]$, then*

(a) *$(\sigma_i, \theta_i) \in \mathrm{GCon}(\mathcal{A})$ and $\mathcal{A}/(\sigma_i, \theta_i) \cong_g \mathcal{A}_i$, and*

(b) *$(\sigma_1, \theta_1) \wedge \ldots \wedge (\sigma_n, \theta_n) = (\Delta_\Sigma, \Delta_A)$.*

If the representation is proper, then $(\sigma_i, \theta_i) > (\Delta_\Sigma, \Delta_A)$ for every $i \in [n]$.

Lemma 3.5. *If an unranked algebra $\mathcal{A} = (A, \Sigma)$ has g-congruences $(\sigma_1, \theta_1), \ldots, (\sigma_n, \theta_n)$ such that $(\sigma_1, \theta_1) \wedge \cdots \wedge (\sigma_n, \theta_n) = (\Delta_\Sigma, \Delta_A)$, then*

$$(1_\Sigma, \varphi) : \mathcal{A} \to \varkappa(\mathcal{A}/(\sigma_1, \theta_1), \ldots, \mathcal{A}/(\sigma_n, \theta_n))$$

is a gsd-representation of $\mathcal{A}$ for $\varkappa : \Sigma \to \Sigma/\sigma_1 \times \cdots \times \Sigma/\sigma_n$, $f \mapsto (f\sigma_1, \ldots, f\sigma_n)$, and $\varphi : A \to A/\theta_1 \times \cdots \times A/\theta_n$, $a \mapsto (a\theta_1, \ldots, a\theta_n)$. If $(\sigma_i, \theta_i) > (\Delta_\Sigma, \Delta_A)$ for every $i \in [n]$, then this gsd-representation is proper.

The next proposition corresponds to two fundamental results by G. Birkhoff.

Proposition 3.6. *Let $\mathcal{A} = (A, \Sigma)$ be a finite unranked algebra.*

(a) *$\mathcal{A}$ is gsd-irreducible if and only if $|A| = |\Sigma| = 1$ or it has a least nontrivial g-congruence, i.e., $\bigcap(\mathrm{GCon}(\mathcal{A}) \setminus \{(\Delta_\Sigma, \Delta_A)\}) > (\Delta_\Sigma, \Delta_A)$.*

(b) *$\mathcal{A}$ has a gsd-representation with finitely many factors each of which is a gsd-irreducible g-image of $\mathcal{A}$.*

Note that a trivial Σ-algebra $\mathcal{A} = (A, \Sigma)$ is gsd-irreducible exactly in case $|\Sigma| \leq 2$. If $|\Sigma| = 2$, then $(\nabla_\Sigma, \Delta_A)$ is the least nontrivial g-congruence of $\mathcal{A}$.

The *unranked ΣX-term algebra* $\mathcal{T}_\Sigma(X) = (T_\Sigma(X), \Sigma)$ is defined by $f_{\mathcal{T}_\Sigma(X)}(\varepsilon) = f$ for any $f \in \Sigma$, and $f_{\mathcal{T}_\Sigma(X)}(t_1, \ldots, t_m) = f(t_1, \ldots, t_m)$ for any $m > 0$ and $t_1, \ldots, t_m \in T_\Sigma(X)$. We may speak simply about (*unranked*) *term algebras*.

A g-morphism $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ between unranked term algebras replaces each label $f \in \Sigma$ with $\iota(f) \in \Omega$ and each leaf labeled with a symbol $x \in X$ with the ΩY-tree $x\varphi$. Thus they are the unranked analogs of the *inner alphabetic tree homomorphisms* of [14]. For any given $\iota : \Sigma \to \Omega$ and any leaf alphabet X, we define a mapping $\iota_X : T_\Sigma(X) \to T_\Omega(X)$ of this type as follows:

(1) $\iota_X(x) = x$ for $x \in X$, $\iota_X(f) = \iota(f)$ for $f \in \Sigma$, and

(2) $\iota_X(t) = \iota(f)(\iota_X(t_1), \ldots, \iota_X(t_m))$ for $t = f(t_1, \ldots, t_m)$.

Then $(\iota, \iota_X) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(X)$ is a g-morphism that transforms any ΣX-tree to an ΩX-tree by replacing any $f \in \Sigma$ with $\iota(f)$ but preserving every $x \in X$.

The following proposition can be proved by obvious modifications in the same way that usual term algebras are shown to be freely generated.

Proposition 3.7. *The term algebra $\mathcal{T}_\Sigma(X)$ is* freely generated *by X over the class of all unranked algebras, that is to say, X generates $\mathcal{T}_\Sigma(X)$, and if $\mathcal{A} = (A, \Omega)$ is any unranked algebra, then for any pair of mappings $\iota : \Sigma \to \Omega$ and $\alpha : X \to A$, there is a unique g-morphism $(\iota, \varphi_{\iota,\alpha}) : \mathcal{T}_\Sigma(X) \to \mathcal{A}$ such that $\varphi_{\iota,\alpha}\big|_X = \alpha$.*

The values of $\varphi_{\iota,\alpha}$ can be obtained by evaluation of term functions for the valuation $\alpha : X \to A$. The *term function* $t^{\mathcal{A}} \colon A^X \to A$ of a Σ-algebra $\mathcal{A} = (A, \Sigma)$ induced by a ΣX-tree $t \in T_\Sigma(X)$ is defined as follows: For any $\alpha : X \to A$,

(1) $x^{\mathcal{A}}(\alpha) = \alpha(x)$ for $x \in X$, $f^{\mathcal{A}}(\alpha) = f_{\mathcal{A}}(\varepsilon)$ for $f \in \Sigma$, and

(2) $t^{\mathcal{A}}(\alpha) = f_{\mathcal{A}}(t_1^{\mathcal{A}}(\alpha), \ldots, t_m^{\mathcal{A}}(\alpha))$ for $t = f(t_1, \ldots, t_m)$.

Then $t\varphi_{\iota,\alpha} = \iota_X(t)^{\mathcal{A}}(\alpha)$ for all $t \in T_\Sigma(X)$ with $\mathcal{A}$, ι and α as in Proposition 3.7.

A mapping $p : A \to A$ is an *elementary translation* of $\mathcal{A} = (A, \Sigma)$ if there exist $f \in \Sigma$, $m > 0$, and $i \in [m]$ and $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_m \in A$ such that $p(b) = f_{\mathcal{A}}(a_1 \cdots a_{i-1} b a_{i+1} \cdots a_m)$ for all $b \in A$. The set $\mathrm{Tr}(\mathcal{A})$ of *translations* of $\mathcal{A}$ is the smallest set of mappings $A \to A$ that contains the identity map 1_A and all elementary translations of $\mathcal{A}$, and is closed under composition.

Lemma 3.8. *Any congruence of an unranked algebra $\mathcal{A} = (A, \Sigma)$ is invariant with respect to all the translations of $\mathcal{A}$, and an equivalence on A is a congruence of $\mathcal{A}$ if it is invariant with respect to all the elementary translations of $\mathcal{A}$.*

Moreover, we have the following counterpart of Lemma 5.3 in [25].

Lemma 3.9. *Let $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$ be a g-morphism from a Σ-algebra $\mathcal{A} = (A, \Sigma)$ to an Ω-algebra $\mathcal{B} = (B, \Omega)$. For every translation $p \in \mathrm{Tr}(\mathcal{A})$, there is a translation $p_{\iota,\varphi}$ of $\mathcal{B}$ such that $p(a)\varphi = p_{\iota,\varphi}(a\varphi)$ for every $a \in A$. If (ι, φ) is a g-epimorphism, then every translation of $\mathcal{B}$ equals $p_{\iota,\varphi}$ for some $p \in \mathrm{Tr}(\mathcal{A})$.*

For each translation p of the term algebra $\mathcal{T}_\Sigma(X)$ there is a unique context $q \in C_\Sigma(X)$ such that $p(t) = q(t)$ for every $t \in T_\Sigma(X)$, and conversely.

4. Regular unranked algebras and regular congruences

Let us now introduce the unranked algebras that play the same role here as finite algebras in the ranked case. In [28] they were called "pseudoautomata".

Definition 4.1. An unranked algebra $\mathcal{A} = (A, \Sigma)$ is said to be *regular* if it is finite and $f_{\mathcal{A}}^{-1}(a)$ is a regular language over A for all $f \in \Sigma$ and $a \in A$. The class of all regular algebras is denoted by **Reg**.

That the sets $f_{\mathcal{A}}^{-1}(a) = \{w \in A^* \mid f_{\mathcal{A}}(w) = a\}$ are regular languages means that the functions $f_{\mathcal{A}} : A^* \to A$ can be computed by finite automata.

For example, if $\Sigma = \{f\}$ and $\mathcal{A} = (\{0,1\}, \Sigma)$ is defined by $f_{\mathcal{A}}(w) = 1$ if $w \in \{0^n 1^n \mid n \geq 0\}$, and $f_{\mathcal{A}}(w) = 0$ otherwise ($w \in A^*$), then $\mathcal{A}$ is not regular since $f_{\mathcal{A}}^{-1}(1) = \{0^n 1^n \mid n \geq 0\}$ is not a regular language, but $\mathcal{A}$ is regular if we set $f_{\mathcal{A}}(w) = a_1 + \cdots + a_n \pmod 2$ for all $w = a_1 \cdots a_n$, $(a_1, \ldots, a_n \in A)$.

Lemma 4.2. *The g-subalgebras and the g-images of a regular algebra are regular.*

Proof. Let $\mathcal{A} = (A, \Sigma)$ be a regular algebra and let $\mathcal{B} = (B, \Omega)$ be any unranked algebra. If $\mathcal{B}$ is an Ω-subalgebra of $\mathcal{A}$, then $f_{\mathcal{B}}^{-1}(b) = f_{\mathcal{A}}^{-1}(b) \cap B^*$ is a regular language for all $f \in \Omega$ and $b \in B$, and hence $\mathcal{B}$ is regular.

Next, let $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$ be a g-epimorphism. Consider any $g \in \Omega$, $b \in B$ and $w \in B^*$. If $\varphi_* : A^* \to B^*$ is the extension of φ to a monoid morphism, then $w = v\varphi_*$ for some $v \in A^*$. If $f \in \Sigma$ satisfies $\iota(f) = g$, then

$$\begin{aligned} w \in g_{\mathcal{B}}^{-1}(b) \;&\Leftrightarrow\; g_{\mathcal{B}}(w) = b \;\Leftrightarrow\; \iota(f)_{\mathcal{B}}(v\varphi_*) = b \;\Leftrightarrow\; f_{\mathcal{A}}(v)\varphi = b \;\Leftrightarrow v \in f_{\mathcal{A}}^{-1}(b\varphi^{-1}) \\ &\Rightarrow w \in f_{\mathcal{A}}^{-1}(b\varphi^{-1})\varphi_*, \end{aligned}$$

i.e., $g_{\mathcal{B}}^{-1}(b) \subseteq f_{\mathcal{A}}^{-1}(b\varphi^{-1})\varphi_*$. For the converse inclusion, let $w \in f_{\mathcal{A}}^{-1}(b\varphi^{-1})\varphi_*$. Then $w = v\varphi_*$ for some $a \in b\varphi^{-1}$ and $v \in f_{\mathcal{A}}^{-1}(a)$. This means that $b = a\varphi = f_{\mathcal{A}}(v)\varphi = g_{\mathcal{B}}(v\varphi_*) = g_{\mathcal{B}}(w)$, and hence $w \in g_{\mathcal{B}}^{-1}(b)$. We may conclude that $g_{\mathcal{B}}^{-1}(b)$ is regular because $f_{\mathcal{A}}^{-1}(b\varphi^{-1})$ is the union of the finitely many regular sets $f_{\mathcal{A}}^{-1}(a)$ with $a \in b\varphi^{-1}$. ∎

Lemma 4.3. *Any g-product of regular algebras is regular. In particular, any g-derived algebra of a regular algebra is regular.*

Proof. Consider a g-product $\varkappa(\mathcal{A}, \mathcal{B}) = (A \times B, \Gamma)$ of two regular algebras $\mathcal{A} = (A, \Sigma)$ and $\mathcal{B} = (B, \Omega)$. Let $f \in \Gamma$, $(a, b) \in A \times B$, and $\varkappa(f) = (g, h)$. If the morphisms $\varphi_1 : (A \times B)^* \to A^*$ and $\varphi_2 : (A \times B)^* \to B^*$ extend the projections $\pi_1 : A \times B \to A$ and $\pi_2 : A \times B \to B$, then $f_{\varkappa(\mathcal{A},\mathcal{B})}(w) = (g_{\mathcal{A}}(w\varphi_1), h_{\mathcal{B}}(w\varphi_2))$ for any $w \in (A \times B)^*$. Hence, $f_{\varkappa(\mathcal{A},\mathcal{B})}^{-1}(a, b) = g_{\mathcal{A}}^{-1}(a)\varphi_1^{-1} \cap h_{\mathcal{B}}^{-1}(b)\varphi_2^{-1}$ is regular. ∎

Let us say that a regular algebra $\mathcal{A} = (A, \Sigma)$ is *effectively given* if, for all $f \in \Sigma$ and $a \in A$, we are given a finite recognizer of $f_{\mathcal{A}}^{-1}(a)$.

Proposition 4.4. *For any effectively given regular algebra $\mathcal{A} = (A, \Sigma)$, the set* $\mathrm{Tr}(\mathcal{A})$ *of all translations of $\mathcal{A}$ is effectively computable.*

Proof. It suffices to show that for each $f \in \Sigma$, the set $E_f := \{f_{u,v} \mid u, v \in A^*\}$, where $f_{u,v} : A \to A, a \mapsto f_{\mathcal{A}}(uav)$, is effectively computable .

For each $a \in A$, we can find a finite monoid M_a, a morphism $\varphi_a : A^* \to M_a$ and a subset $F_a \subseteq M_a$ such that $f_{\mathcal{A}}^{-1}(a) = F_a\varphi_a^{-1}$. For any $u, v \in A^*$, let $u \sim v$ if and only if $u\varphi_a = v\varphi_a$ for every $a \in A$. Then $f_{u,v} = f_{u',v'}$ for all words $u, v, u', v' \in A^*$ such that $u \sim u'$ and $v \sim v'$ because, for all $a, b \in A$,

$$\begin{aligned} f_{u,v}(a) = b \Leftrightarrow f_{\mathcal{A}}(uav) = b \Leftrightarrow uav \in f_{\mathcal{A}}^{-1}(b) \Leftrightarrow (uav)\varphi_b \in F_b \\ \Leftrightarrow u\varphi_b \cdot a\varphi_b \cdot v\varphi_b \in F_b \Leftrightarrow u'\varphi_b \cdot a\varphi_b \cdot v'\varphi_b \in F_b \Leftrightarrow f_{u',v'}(a) = b. \end{aligned}$$

Let R be a set of representatives of the partition $A^*/\sim$. Such an R is finite and can be effectively formed using the regular sets $m\varphi_a^{-1}$ ($m \in M_a$, $a \in A$). Then E_f is obtained as the set $\{f_{u,v} \mid u, v \in R\}$. ∎

Since we consider classes that may contain unranked Σ-algebras for any Σ, also the operators S, H and P_f are applied to such classes.

Definition 4.5. For any class $\mathbf{K}$ of unranked algebras, let $S_g(\mathbf{K})$ be the class of algebras g-isomorphic to a g-subalgebra of a member of $\mathbf{K}$, $H_g(\mathbf{K})$ be the class of all g-images of members of $\mathbf{K}$, $P_g(\mathbf{K})$ be the class of algebras isomorphic to g-products of members of $\mathbf{K}$, $S(\mathbf{K})$ be the class of algebras isomorphic to a subalgebra of a member $\mathbf{K}$, $H(\mathbf{K})$ be the class of all epimorphic images of members of $\mathbf{K}$, and $P_f(\mathbf{K})$ be the class of algebras isomorphic to the direct product of a finite family of members of $\mathbf{K}$. A class $\mathbf{K}$ of regular algebras is a *variety of regular algebras (VRA)* if $S_g(\mathbf{K})$, $H_g(\mathbf{K})$, $P_g(\mathbf{K}) \subseteq \mathbf{K}$. The class of all VRAs is denoted by **VRA**. The *VRA generated* by a given class $\mathbf{K}$ of regular algebras is denoted by $V_g(\mathbf{K})$.

Since g-derived algebras are special g-products, the following fact is obvious.

Lemma 4.6. *Every VRA is closed under the forming of g-derived algebras.*

From Lemmas 4.2 and 4.3 we get the following proposition.

Proposition 4.7. **Reg** *is a VRA, and hence the greatest VRA.*

If P and Q are any algebra class operators, we denote by PQ the operator such that $PQ(\mathbf{K}) = P(Q(\mathbf{K}))$ for any $\mathbf{K}$. Moreover, $P \leq Q$ means that $P(\mathbf{K}) \subseteq Q(\mathbf{K})$ for every $\mathbf{K}$. The obvious facts that $\mathbf{K} \subseteq S(\mathbf{K}) \subseteq S_g(\mathbf{K})$, $\mathbf{K} \subseteq H(\mathbf{K}) \subseteq H_g(\mathbf{K})$, and $\mathbf{K} \subseteq P_f(\mathbf{K}) \subseteq P_g(\mathbf{K})$ for any $\mathbf{K}$, will be used without comment.

Lemma 4.8.

(a) $S_gS_g = S_gS = SS_g = S_g$, (b) $H_gH_g = H_gH = HH_g = H_g$,
(c) $P_gP_g = P_gP_f = P_fP_g = P_g$, (d) $S_gH \leq S_gH_g \leq HS_g \leq H_gS_g$,
(e) $P_gS \leq P_gS_g \leq SP_g \leq S_gP_g$, (f) $P_gH \leq P_gH_g \leq HP_g \leq H_gP_g$

Proof. Statements (a) and (b) hold because $S_gS_g = S_g$ and $H_gH_g = H_g$. For (c), it suffices to show that $P_gP_g \leq P_g$, and in (d), (e) and (f) all inequalities except for the second ones are obvious. Let us prove that $S_gH_g \leq HS_g$.

Let $\mathbf{K}$ be any class of unranked algebras. To construct a typical member $\mathcal{C} = (C, \Gamma)$ of $S_gH_g(\mathbf{K})$, let $\mathcal{A} = (A, \Sigma)$ be in $\mathbf{K}$, $(\iota, \varphi) : \mathcal{A} \to \mathcal{A}'$ be a g-epimorphism, $\mathcal{B} = (B, \Omega)$ be a g-subalgebra of $\mathcal{A}'$, and let $(\varkappa, \psi) : \mathcal{B} \to \mathcal{C}$ be a g-isomorphism. Now $\mathcal{B}\varphi^{-1} = (B\varphi^{-1}, \iota^{-1}(\Omega))$ is a g-subalgebra of $\mathcal{A}$. If we choose a subset Σ' of $\iota^{-1}(\Omega)$ in such a way that the restriction of ι to Σ' is a bijection $\iota' : \Sigma' \to \Omega$, then $\mathcal{D} = (B\varphi^{-1}, \Sigma')$ is a g-subalgebra of $\mathcal{A}$.

Define $\mathcal{E} = (B\varphi^{-1}, \Gamma)$ as follows: For each $g \in \Gamma$, let $g_{\mathcal{E}} = f_{\mathcal{D}}$ for the $f \in \Sigma'$ with $g = \varkappa(\iota'(f))$. Then $(\iota'\varkappa, 1_{B\varphi^{-1}}) : \mathcal{D} \to \mathcal{E}$ is a g-isomorphism. Indeed, if $f \in \Sigma'$ and $w \in (B\varphi^{-1})^*$, then $f_{\mathcal{D}}(w)1_{B\varphi^{-1}} = f_{\mathcal{D}}(w) = \varkappa(\iota'(f))_{\mathcal{E}}(w1_{B\varphi^{-1}})$. This means that $\mathcal{E} \in S_g(\mathbf{K})$. We show that $\varphi\psi : \mathcal{E} \to \mathcal{C}$ is an epimorphism. Clearly, $B\varphi^{-1}\varphi\psi = C$. Consider any $g \in \Gamma$ and $w \in (B\varphi^{-1})^*$. Let $f \in \Sigma'$ and $h \in \Omega$ be such that $\iota'(f) = h$ and $\varkappa(h) = g$. Then $g_{\mathcal{E}}(w)\varphi\psi = f_{\mathcal{D}}(w)\varphi\psi = f_{\mathcal{A}}(w)\varphi\psi = h_{\mathcal{A}'}(w\varphi)\psi = h_{\mathcal{B}}(w\varphi)\psi = g_{\mathcal{C}}(w\varphi\psi)$. Thus $\mathcal{C} \in HS_g(\mathbf{K})$. ∎

Now we get the following result in the usual way.

Proposition 4.9. $V_g = H_gS_gP_g$.

For a simpler representation of V_g, we need also the following two relations.

Lemma 4.10. (a) $H_gS \leq HS_g$, *and* (b) $S_gP_g \leq SP_g$.

Proof. Let $\mathbf{K}$ be any class of unranked algebras. To prove (a), let $\mathcal{A} = (A, \Sigma) \in \mathbf{K}$, $\mathcal{B} = (B, \Sigma)$ be a subalgebra of $\mathcal{A}$, and let $(\iota, \varphi) : \mathcal{B} \to \mathcal{C}$ be a g-epimorphism onto some $\mathcal{C} = (C, \Gamma)$. Let $\Omega \subseteq \Sigma$ be such that the restriction $\iota' : \Omega \to \Gamma$ of ι to Ω is a bijection. Then $\mathcal{B}' = (B, \Omega)$ is a g-subalgebra of $\mathcal{A}$. Define $\mathcal{B}'' = (B, \Gamma)$ by $g_{\mathcal{B}''} = h_{\mathcal{B}'}$ for each $g \in \Gamma$ and $h \in \Omega$ with $\iota'(h) = g$. Then $(\iota', 1_B) : \mathcal{B}' \to \mathcal{B}''$ is a g-isomorphism. Hence, $\mathcal{B}'' \in S_g(\mathbf{K})$. To prove $\mathcal{C} \in HS_g(\mathbf{K})$, we show that $\varphi : \mathcal{B}'' \to \mathcal{C}$ is an epimorphism. Let $g \in \Gamma$, $m \geq 0$, and $b_1, \ldots, b_m \in B$, and let $h \in \Omega$ be such that $\iota(h) = g$. Then $g_{\mathcal{B}''}(b_1, \ldots, b_m)\varphi = h_{\mathcal{B}'}(b_1, \ldots, b_m)\varphi = h_{\mathcal{B}}(b_1, \ldots, b_m)\varphi = g_{\mathcal{C}}(b_1\varphi, \ldots, b_m\varphi)$. Moreover, φ is surjective.

To prove (b), let $n \geq 0$, $\mathcal{A}_i = (A_i, \Sigma_i) \in \mathbf{K}$ for each $i \in [n]$, $\varkappa : \Omega \to \Sigma_1 \times \cdots \times \Sigma_n$ be a mapping, $\psi : \varkappa(\mathcal{A}_1, \ldots, \mathcal{A}_n) \to \mathcal{B}$ be an isomorphism to some $\mathcal{B} = (B, \Omega)$, $\mathcal{C} = (C, \Omega')$ be a g-subalgebra of $\mathcal{B}$, and (ι, φ) a g-isomorphism from $\mathcal{C}$ to $\mathcal{D} = (D, \Gamma)$. Then $\mathcal{D}$ is a typical representative of $S_gP_g(\mathbf{K})$.

Let $\lambda : \Gamma \to \Sigma_1 \times \cdots \times \Sigma_n$ be such that $\lambda(g) = \varkappa(\iota^{-1}(g))$ for each $g \in \Gamma$. Then $\mathcal{E} = (C\psi^{-1}, \Gamma)$ is a subalgebra of $\lambda(\mathcal{A}_1, \ldots, \mathcal{A}_n) = (A_1 \times \cdots \times A_n, \Gamma)$. Indeed, let $g \in \Gamma$, $m \geq 0$, and $\mathbf{a}_1 = (a_{11}, \ldots, a_{1n})$, $\ldots$, $\mathbf{a}_m = (a_{m1}, \ldots, a_{mn}) \in C\psi^{-1}$. If $\iota^{-1}(g) = h \in \Omega'$ and $\varkappa(h) = (f_1, \ldots, f_n)$, then

$$\begin{aligned} g_{\lambda(\mathcal{A}_1,\ldots,\mathcal{A}_n)}(\mathbf{a}_1, \ldots, \mathbf{a}_m)\psi &= ((f_1)_{\mathcal{A}_1}(a_{11}, \ldots, a_{m1}), \ldots, (f_n)_{\mathcal{A}_n}(a_{1n}, \ldots, a_{mn}))\psi \\ &= h_{\varkappa(\mathcal{A}_1,\ldots,\mathcal{A}_n)}(\mathbf{a}_1, \ldots, \mathbf{a}_m)\psi = h_{\mathcal{B}}(\mathbf{a}_1\psi, \ldots, \mathbf{a}_m\psi) = h_{\mathcal{C}}(\mathbf{a}_1\psi, \ldots, \mathbf{a}_m\psi) \end{aligned}$$

is in C since C is an Ω'-closed subset of $\mathcal{B}$. To prove $\mathcal{D} \in SP_g(\mathbf{K})$, we verify that $\psi\varphi : \mathcal{E} \to \mathcal{D}$, with ψ restricted to $C\psi^{-1}$, is an isomorphism. Consider any $g \in \Gamma$, $m \geq 0$ and $\mathbf{a}_1, \ldots, \mathbf{a}_m \in C\psi^{-1}$. For the $h \in \Omega'$ with $\iota(h) = g$, we get $g_{\mathcal{E}}(\mathbf{a}_1, \ldots, \mathbf{a}_m)\psi\varphi = g_{\lambda(\mathcal{A}_1,\ldots,\mathcal{A}_n)}(\mathbf{a}_1, \ldots, \mathbf{a}_m)\psi\varphi = h_{\mathcal{C}}(\mathbf{a}_1\psi, \ldots, \mathbf{a}_m\psi)\varphi = g_{\mathcal{D}}(\mathbf{a}_1\psi\varphi, \ldots, \mathbf{a}_m\psi\varphi)$. Moreover, it is clear that $\psi\varphi$ is bijective. ∎

Proposition 4.11. $V_g = HSP_g$.

Proof. Since $\mathbf{K} \subseteq HSP_g(\mathbf{K}) \subseteq H_gS_gP_g(\mathbf{K}) = V_g(\mathbf{K})$ for any class of regular algebras $\mathbf{K}$, it suffices to show that $HSP_g(\mathbf{K})$ is a VRA, and this follows from Lemmas 4.8 and 4.10: $S_g(HSP_g) \leq HS_gSP_g \leq HS_gP_g \leq HSP_g$, $H_g(HSP_g) \leq H_gSP_g \leq HS_gP_g \leq HSP_g$ and $P_g(HSP_g) \leq HP_gSP_g \leq HSP_gP_g = HSP_g$. ∎

Finally, let us note the following important fact.

Lemma 4.12. *Let* $\mathbf{K}$ *be a VRA. If* (σ, θ) *is a g-congruence of an unranked algebra* $\mathcal{A} = (A, \Sigma)$*, then* $\mathcal{A}/\theta \in \mathbf{K}$ *if and only if* $\mathcal{A}/(\sigma, \theta) \in \mathbf{K}$.

Proof. Clearly, $(\sigma_\natural, 1_{A/\theta}) : \mathcal{A}/\theta \to \mathcal{A}/(\sigma, \theta)$ is a g-epimorphism. Hence, $\mathcal{A}/\theta \in \mathbf{K}$ implies $\mathcal{A}/(\sigma, \theta) \in \mathbf{K}$. Assume now that $\mathcal{A}/(\sigma, \theta) \in \mathbf{K}$. The g-derived algebra $\sigma_\natural(\mathcal{A}/(\sigma, \theta))$ is actually the algebra $\mathcal{A}/\theta$. Indeed, both are Σ-algebras with the same set A/θ of elements, and for any $f \in \Sigma$, $m \geq 0$ and $a_1, \ldots, a_m \in A$, $f_{\sigma_\natural(\mathcal{A}/(\sigma,\theta))}(a_1\theta, \ldots, a_m\theta) = (f\sigma)_{\mathcal{A}/(\sigma,\theta)}(a_1\theta, \ldots, a_m\theta) = f_{\mathcal{A}}(a_1, \ldots, a_m)\theta = f_{\mathcal{A}/\theta}(a_1\theta, \ldots, a_m\theta)$. Hence, $\mathcal{A}/\theta \in \mathbf{K}$ by Lemma 4.6. ∎

Definition 4.13. For any $\mathcal{A} = (A, \Sigma)$, let $\mathrm{FCon}(\mathcal{A}) := \{\theta \in \mathrm{Con}(\mathcal{A}) \mid A/\theta \text{ finite}\}$ and let $\mathrm{FGCon}(\mathcal{A}) := \{(\sigma, \theta) \in \mathrm{GCon}(\mathcal{A}) \mid \theta \in \mathrm{FCon}(\mathcal{A})\}$. A congruence θ of $\mathcal{A}$ is *regular* if $\theta \in \mathrm{FCon}(\mathcal{A})$ and $f^{-1}_{\mathcal{A}/\theta}(a\theta)$ is a regular language over A/θ for all $f \in \Sigma$ and $a \in A$. A g-congruence (σ, θ) of $\mathcal{A}$ is *regular* if θ is a regular congruence. Let $\mathrm{RCon}(\mathcal{A})$ and $\mathrm{RGCon}(\mathcal{A})$, respectively, denote the sets of regular congruences and regular g-congruences of $\mathcal{A}$.

For any $\theta \in \mathrm{Con}(\mathcal{A})$, let $\eta_\theta : A^* \to (A/\theta)^*$ be the morphism such that $a\eta_\theta = a\theta$ for each $a \in A$. For all $f \in \Sigma$, $a \in A$ and $w \in A^*$, $w\eta_\theta \in f^{-1}_{\mathcal{A}/\theta}(a\theta)$ if and only if $f_{\mathcal{A}}(w) \in a\theta$. Hence $f^{-1}_{\mathcal{A}}(a\theta) = f^{-1}_{\mathcal{A}/\theta}(a\theta)\eta_\theta^{-1}$. As η_θ is surjective, also $f^{-1}_{\mathcal{A}/\theta}(a\theta) = f^{-1}_{\mathcal{A}}(a\theta)\eta_\theta$ holds. These equalities yield the following lemma. Note that in $f^{-1}_{\mathcal{A}/\theta}(a\theta)$, $a\theta$ is an element of A/θ, but in $f^{-1}_{\mathcal{A}}(a\theta)$ it is a subset of A.

Lemma 4.14. *Let $\theta \in \mathrm{Con}(\mathcal{A})$. For any $f \in \Sigma$ and $a \in A$, $f_{\mathcal{A}/\theta}^{-1}(a\theta)$ is a regular language over A/θ if and only if $f_{\mathcal{A}}^{-1}(a\theta)$ is a regular language over A.*

If $\mathcal{A} = (A, \Sigma)$ is a regular algebra, then for any $\theta \in \mathrm{Con}(\mathcal{A})$, $f \in \Sigma$ and $a \in A$, $f_{\mathcal{A}}^{-1}(a\theta)$ is the union of the finitely many regular sets $f_{\mathcal{A}}^{-1}(b)$, where $b \in a\theta$. Hence, Lemma 4.14 yields the following proposition.

Proposition 4.15. *Every congruence of a regular algebra is regular.*

Lemma 4.16. *For any unranked algebra $\mathcal{A} = (A, \Sigma)$, $\mathrm{RCon}(\mathcal{A})$ is a filter of the lattice $\mathrm{Con}(\mathcal{A})$, and similarly, $\mathrm{RGCon}(\mathcal{A})$ is a filter of $\mathrm{GCon}(\mathcal{A})$.*

Proof. Since $\mathrm{RCon}(\mathcal{A})$ contains ∇_A, it is nonempty. If $\theta, \rho \in \mathrm{RCon}(\mathcal{A})$, then clearly $\theta \cap \rho \in \mathrm{FCon}(\mathcal{A})$. Moreover, for any $f \in \Sigma$ and $a \in A$, $f_{\mathcal{A}}^{-1}(a(\theta \cap \rho)) = f_{\mathcal{A}}^{-1}(a\theta) \cap f_{\mathcal{A}}^{-1}(a\rho)$, and hence also $f_{\mathcal{A}/\theta\cap\rho}^{-1}(a(\theta \cap \rho))$ is regular by Lemma 4.14.

Next, let $\theta \in \mathrm{RCon}(\mathcal{A})$, $\rho \in \mathrm{Con}(\mathcal{A})$ and $\theta \subseteq \rho$. Of course, $\rho \in \mathrm{FCon}(\mathcal{A})$. Moreover, for each $a \in A$ there is a finite set of elements $a_1, \ldots, a_k \in A$ ($k \geq 1$) such that $a\rho = a_1\theta \cup \ldots \cup a_k\theta$, and hence $f_{\mathcal{A}}^{-1}(a\rho) = f_{\mathcal{A}}^{-1}(a_1\theta) \cup \ldots \cup f_{\mathcal{A}}^{-1}(a_k\theta)$ is a regular language for every $f \in \Sigma$. Hence, ρ is regular by Lemma 4.14.

That $\mathrm{RGCon}(\mathcal{A})$ is a filter of $\mathrm{GCon}(\mathcal{A})$ follows immediately from the fact that $\mathrm{RCon}(\mathcal{A})$ is a filter of $\mathrm{Con}(\mathcal{A})$. ∎

The following facts are direct consequences of the relevant definitions.

Proposition 4.17. *If θ is a congruence of an unranked algebra $\mathcal{A}$, then $\mathcal{A}/\theta$ is a regular algebra exactly in case θ is a regular congruence. Similarly, if $(\sigma, \theta) \in \mathrm{GCon}(\mathcal{A})$, then $\mathcal{A}/(\sigma, \theta)$ is regular if and only if $(\sigma, \theta) \in \mathrm{RGCon}(\mathcal{A})$.*

5. Syntactic congruences and algebras

Syntactic algebras form a bridge between varieties of recognizable sets and varieties of finite algebras. In Eilenberg's [11] Variety Theory they are the syntactic monoids (or semigroups). We define syntactic congruences and syntactic algebras of subsets of unranked algebras similarly as they are defined for subsets of general ranked algebras (cf. [1, 23, 24, 26]), and the basic facts about them remain valid and can be proved similarly as in the ranked case.

Definition 5.1. The *syntactic congruence* θ_H of a subset $H \subseteq A$ of an unranked algebra $\mathcal{A} = (A, \Sigma)$ is defined by

$$a \,\theta_H\, b :\Leftrightarrow (\forall p \in \mathrm{Tr}(\mathcal{A}))(p(a) \in H \leftrightarrow p(b) \in H) \qquad (a, b \in A),$$

and $\mathrm{SA}(H) := \mathcal{A}/\theta_H$ is the *syntactic algebra* of H. The natural morphism $\varphi_H : \mathcal{A} \to \mathrm{SA}(H)$, $a \mapsto a\theta_H$, is called the *syntactic morphism* of H.

An equivalence $\theta \in \mathrm{Eq}(A)$ *saturates* $H \subseteq A$ if H is the union of θ-classes.

Lemma 5.2. *For any subset $H \subseteq A$ of an unranked algebra $\mathcal{A} = (A, \Sigma)$, θ_H is the greatest congruence of $\mathcal{A}$ that saturates H.*

The following fact is an immediate consequence of Proposition 4.4.

Proposition 5.3. *If $\mathcal{A} = (A, \Sigma)$ is an effectively given regular algebra, then the syntactic congruence θ_H and the syntactic algebra $\mathrm{SA}(H)$ of any effectively given subset $H \subseteq A$ can be effectively constructed.*

We call an unranked algebra $\mathcal{A} = (A, \Sigma)$ *syntactic* if it is isomorphic to the syntactic algebra of a subset of some unranked algebra. A subset $D \subseteq A$ of $\mathcal{A}$ is *disjunctive* if $\theta_D = \Delta_A$. The following facts can be proved similarly as Propositions 3.6 and 3.7 in [24].

Proposition 5.4. *An unranked algebra is syntactic if and only if it has a disjunctive subset. Every finite gsd-irreducible unranked algebra is syntactic. Hence every VRA is generated by regular syntactic algebras.*

It is easy to see that for any congruence θ of an unranked algebra $\mathcal{A} = (A, \Sigma)$, there is a greatest equivalence $\mathrm{M}(\theta)$ on Σ such that $(\mathrm{M}(\theta), \theta) \in \mathrm{GCon}(\mathcal{A})$. We shall need the following obvious properties of the M-operator.

Lemma 5.5. *Let $\mathcal{A} = (A, \Sigma)$ and $\mathcal{B} = (B, \Omega)$ be unranked algebras.*

(a) *If θ, $\theta' \in \mathrm{Con}(\mathcal{A})$ and $\theta \subseteq \theta'$, then $\mathrm{M}(\theta) \subseteq \mathrm{M}(\theta')$.*

(b) *For any set $\{\theta_i \mid i \in I\}$ of congruences of $\mathcal{A}$, $\mathrm{M}(\bigcap_{i \in I} \theta_i) = \bigcap_{i \in I} \mathrm{M}(\theta_i)$.*

(c) *$\iota \circ \mathrm{M}(\theta) \circ \iota^{-1} \subseteq \mathrm{M}(\varphi \circ \theta \circ \varphi^{-1})$ for any g-morphism $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$ and any $\theta \in \mathrm{Con}(\mathcal{B})$. If φ is surjective, then equality holds.*

We will also need the following operator-reduced versions of θ_H and $\mathrm{SA}(H)$.

Definition 5.6. The *reduced syntactic congruence* of a subset H of an unranked algebra $\mathcal{A} = (A, \Sigma)$ is the g-congruence (σ_H, θ_H) of $\mathcal{A}$, where θ_H is the syntactic congruence of H and $\sigma_H := \mathrm{M}(\theta_H)$, the *reduced syntactic algebra* $\mathrm{RA}(H)$ of H is the g-quotient $\mathcal{A}/(\sigma_H, \theta_H) = (A/\theta_H, \Sigma/\sigma_H)$, and the *syntactic g-morphism* $(\iota_H, \varphi_H) : \mathcal{A} \to \mathrm{RA}(H)$ is defined by $\iota_H : f \mapsto f\sigma_H$ and $\varphi_H : a \to a\theta_H$.

Proposition 5.7. *Let $\mathcal{A} = (A, \Sigma)$ and $\mathcal{B} = (B, \Omega)$ be unranked algebras.*

(a) *$\theta_{A \setminus H} = \theta_H$ for every $H \subseteq A$.*

(b) *$\theta_H \cap \theta_K \subseteq \theta_{H \cap K}$ for all $H, K \subseteq A$.*

(c) *$\theta_H \subseteq \theta_{p^{-1}(H)}$ for all $H \subseteq A$ and $p \in \mathrm{Tr}(\mathcal{A})$.*

(d) $\varphi \circ \theta_H \circ \varphi^{-1} \subseteq \theta_{H\varphi^{-1}}$ *and* $\iota \circ \sigma_H \circ \iota^{-1} \subseteq \sigma_{H\varphi^{-1}}$ *for any* $H \subseteq B$ *and any g-morphism* $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$, *and equalities hold if* (ι, φ) *is a g-epimorphism.*

Proof. Assertions (a)–(c) are obvious, so we prove just (d). For any a, $a' \in A$,

$$\begin{aligned} a\,\varphi \circ \theta_H \circ \varphi^{-1}\, a' &\Leftrightarrow (\forall q \in \mathrm{Tr}(\mathcal{B}))\,(q(a\varphi) \in H \leftrightarrow q(a'\varphi) \in H) \\ &\Rightarrow (\forall p \in \mathrm{Tr}(\mathcal{A}))\,(p_{\iota,\varphi}(a\varphi) \in H \leftrightarrow p_{\iota,\varphi}(a'\varphi) \in H) \\ &\Leftrightarrow (\forall p \in \mathrm{Tr}(\mathcal{A}))\,(p(a) \in H\varphi^{-1} \leftrightarrow p(a') \in H\varphi^{-1}) \\ &\Leftrightarrow a\,\theta_{H\varphi^{-1}}\, a'. \end{aligned}$$

Hence $\varphi\circ\theta_H\circ\varphi^{-1} \subseteq \theta_{H\varphi^{-1}}$, and $\iota\circ\sigma_H\circ\iota^{-1} \subseteq \sigma_{H\varphi^{-1}}$ now follows from Lemma 5.5:

$$\iota \circ \sigma_H \circ \iota^{-1} = \iota \circ \mathrm{M}(\theta_H) \circ \iota^{-1} \subseteq \mathrm{M}(\varphi \circ \theta_H \circ \varphi^{-1}) \subseteq \mathrm{M}(\theta_{H\varphi^{-1}}) = \sigma_{H\varphi^{-1}}.$$

If (ι, φ) is a g-epimorphism, the only "$\Rightarrow$" in the proof of the first inclusion can be replaced by "$\Leftrightarrow$", and all inclusions become equalities. ∎

Proposition 5.8. *Let* $\mathcal{A} = (A, \Sigma)$ *and* $\mathcal{B} = (B, \Omega)$ *be unranked algebras.*

(a) $\mathrm{SA}(A \setminus H) = \mathrm{SA}(H)$ *for every* $H \subseteq A$.

(b) $\mathrm{SA}(H \cap K) \preceq \mathrm{SA}(H) \times \mathrm{SA}(K)$ *for all* $H, K \subseteq A$.

(c) $\mathrm{SA}(p^{-1}(H))$ *is an epimorphic image of* $\mathrm{SA}(H)$ *for all* $H \subseteq A$ *and* $p \in \mathrm{Tr}(\mathcal{A})$.

(d) $\mathrm{RA}(H\varphi^{-1}) \preceq_g \mathrm{RA}(H)$ *for any g-morphism* $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$ *and any* $H \subseteq B$. *If* (ι, φ) *is a g-epimorphism, then* $\mathrm{RA}(H\varphi^{-1}) \cong_g \mathrm{RA}(H)$.

Proof. Claims (a)–(c) follow from the corresponding parts of Proposition 5.7.

To prove (d), assume first that (ι, φ) is a g-epimorphism. It follows from Proposition 5.7(d) that the maps $\psi : A/\theta_{H\varphi^{-1}} \to B/\theta_H$, $a\theta_{H\varphi^{-1}} \mapsto (a\varphi)\theta_H$, and $\varkappa : \Sigma/\sigma_{H\varphi^{-1}} \to \Omega/\sigma_H$, $f\sigma_{H\varphi^{-1}} \mapsto \iota(f)\sigma_H$, are well-defined and injective. Clearly, they are also surjective, and for any $f \in \Sigma$, $m \geq 0$ and $a_1, \ldots, a_m \in A$,

$$\begin{aligned} (f\sigma_{H\varphi^{-1}})_{\mathrm{RA}(H\varphi^{-1})}(a_1\theta_{H\varphi^{-1}}, \ldots, a_m\theta_{H\varphi^{-1}})\psi &= (f_{\mathcal{A}}(a_1, \ldots, a_m)\theta_{H\varphi^{-1}})\psi \\ &= (f_{\mathcal{A}}(a_1, \ldots, a_m)\varphi)\theta_H = (\iota(f)_{\mathcal{B}}(a_1\varphi, \ldots, a_m\varphi))\theta_H \\ &= (\iota(f)\sigma_H)_{\mathrm{RA}(H)}((a_1\varphi)\theta_H, \ldots, (a_m\varphi)\theta_H) \\ &= \varkappa(f\sigma_{H\varphi^{-1}})_{\mathrm{RA}(H)}((a_1\theta_{H\varphi^{-1}})\psi, \ldots, (a_m\theta_{H\varphi^{-1}})\psi), \end{aligned}$$

which shows that $(\varkappa, \psi) : \mathrm{RA}(H\varphi^{-1}) \to \mathrm{RA}(H)$ is a g-isomorphism.

Consider now a general g-morphism $(\iota, \varphi) : \mathcal{A} \to \mathcal{B}$. Let $\mathcal{C} = (C, \iota(\Sigma)/\sigma_H)$, where $C = A\varphi\varphi_H$ and $\iota(\Sigma)/\sigma_H = \{\iota(f)\sigma_H \mid f \in \Sigma\}$, be the image of $\mathcal{A}$ in $\mathrm{RA}(H)$ under the g-morphism $(\iota\iota_H, \varphi\varphi_H) : \mathcal{A} \to \mathrm{RA}(H)$. The mappings

$$\varkappa : \Sigma \to \iota(\Sigma)/\sigma_H,\ f \mapsto \iota(f)\sigma_H,\ \text{and}\ \psi : A \to C,\ a \mapsto (a\varphi)\theta_H,$$

form a g-epimorphism $(\varkappa, \psi) : \mathcal{A} \to \mathcal{C}$, and thus $\mathrm{RA}(H\varphi^{-1}\psi\psi^{-1}) \cong_g \mathrm{RA}(H\varphi^{-1}\psi)$ by the previous part of the proof. Also, $\mathrm{RA}(H\varphi^{-1}\psi) \preceq_g \mathrm{RA}(H)$ as $\mathrm{RA}(H\varphi^{-1}\psi)$ is a g-image of the g-subalgebra $\mathcal{C}$ of $\mathrm{RA}(H)$. To obtain $\mathrm{RA}(H\varphi^{-1}) \preceq_g \mathrm{RA}(H)$ it therefore suffices to show that $H\varphi^{-1} = H\varphi^{-1}\psi\psi^{-1}$. Of course, $H\varphi^{-1} \subseteq H\varphi^{-1}\psi\psi^{-1}$, and on the other hand, by Proposition 5.7(d) and the equalities $\psi = \varphi\varphi_H$, $\varphi_H \circ \varphi_H^{-1} = \theta_H$ and $(H\varphi^{-1})\theta_{H\varphi^{-1}} = H\varphi^{-1}$, we get $H\varphi^{-1}\psi\psi^{-1} = (H\varphi^{-1})\varphi \circ \varphi_H \circ (\varphi \circ \varphi_H)^{-1} = (H\varphi^{-1})\varphi \circ \theta_H \circ \varphi^{-1} \subseteq (H\varphi^{-1})\theta_{H\varphi^{-1}} = H\varphi^{-1}$. ■

Simple modifications of the proofs of statements (d) of Propositions 5.7 and 5.8 yield the following specializations of those statements.

Corollary 5.9. *Let $\mathcal{A} = (A, \Sigma)$ and $\mathcal{B} = (B, \Sigma)$ be unranked Σ-algebras. Then $\varphi \circ \theta_H \circ \varphi^{-1} \subseteq \theta_{H\varphi^{-1}}$ and $\mathrm{SA}(H\varphi^{-1}) \preceq \mathrm{SA}(H)$ for any morphism $\varphi : \mathcal{A} \to \mathcal{B}$ and any $H \subseteq B$. If φ is an epimorphism, then $\varphi \circ \theta_H \circ \varphi^{-1} = \theta_{H\varphi^{-1}}$ and $\mathrm{SA}(H\varphi^{-1}) \cong \mathrm{SA}(H)$.*

For a ΣX-tree language T the *syntactic congruence* θ_T, the *syntactic algebra* $\mathrm{SA}(T)$, the *syntactic morphism* φ_T, the *reduced syntactic congruence* (σ_T, θ_T), the *reduced syntactic algebra* $\mathrm{RA}(T)$ and the *syntactic g-morphism* (ι_T, φ_T) are defined by regarding T as a subset of the term algebra $\mathcal{T}_\Sigma(X)$. Since the translations of $\mathcal{T}_\Sigma(X)$ are given by ΣX-contexts, we have

$$s\,\theta_T\,t \;\Leftrightarrow\; (\forall p \in C_\Sigma(X))(p(s) \in T \leftrightarrow p(t) \in T) \quad (s, t \in T_\Sigma(X)).$$

Let us note that the "top congruences" of [5] correspond to our syntactic congruences of unranked tree languages.

6. Recognizable unranked tree languages

The following definition agrees with that given by Thatcher [28], and it is also equivalent to the one arrived at via automata in [9, 16], for example.

Definition 6.1. An unranked Σ-algebra $\mathcal{A} = (A, \Sigma)$ *recognizes* an unranked ΣX-tree language T if $T = F\varphi^{-1}$ for some morphism $\varphi : \mathcal{T}_\Sigma(X) \to \mathcal{A}$ and a subset $F \subseteq A$, and we call T *recognizable* if it is recognized by a regular Σ-algebra. The set of all recognizable unranked ΣX-tree languages is denoted by $Rec(\Sigma, X)$.

Proposition 6.2. *An unranked Σ-algebra $\mathcal{A}$ recognizes an unranked ΣX-tree language T if and only if $\mathrm{SA}(T) \preceq \mathcal{A}$.*

Proof. It is clear that any tree language recognized by a subalgebra or an epimorphic image of an algebra $\mathcal{A}$, is recognized by $\mathcal{A}$, too. Since $\mathrm{SA}(T)$ recognizes T $(= T\varphi_T\varphi_T^{-1})$, this means that $\mathrm{SA}(T) \preceq \mathcal{A}$ implies that $\mathcal{A}$ recognizes T. The converse holds by Corollary 5.9: if $T = F\varphi^{-1}$ for some morphism $\varphi : T_\Sigma(X) \to \mathcal{A}$ and a subset F of $\mathcal{A}$, then $\mathrm{SA}(T) = \mathrm{SA}(F\varphi^{-1}) \preceq \mathrm{SA}(F) \preceq \mathcal{A}$. ■

Next we give a Myhill-Nerode theorem for unranked tree languages.

Proposition 6.3. *For any $T \subseteq T_\Sigma(X)$, the following statements are equivalent:*

(a) $T \in Rec(\Sigma, X)$*;*

(b) *T is saturated by a regular congruence of $\mathcal{T}_\Sigma(X)$;*

(c) *the syntactic congruence θ_T is regular.*

Proof. Let us first prove the equivalence of (a) and (b). If $T \in Rec(\Sigma, X)$, then $T = F\varphi^{-1}$ for a regular algebra $\mathcal{A} = (A, \Sigma)$, a morphism $\varphi : \mathcal{T}_\Sigma(X) \to \mathcal{A}$ and some $F \subseteq A$. We may assume that φ is surjective. It is clear that T is saturated by $\theta := \ker \varphi$, and $\theta \in \mathrm{RCon}(\mathcal{T}_\Sigma(X))$ by Proposition 4.17 as $\mathcal{T}_\Sigma(X)/\theta \cong \mathcal{A}$. On the other hand, if T is saturated by some $\theta \in \mathrm{RCon}(\mathcal{T}_\Sigma(X))$, then $T = T\theta_\natural\theta_\natural^{-1}$ means that T is recognized by the regular algebra $\mathcal{T}_\Sigma(X)/\theta$.

If T is saturated by a congruence $\theta \in \mathrm{RCon}(\mathcal{T}_\Sigma(X))$, then $\theta \subseteq \theta_T$ by Lemma 5.2, and hence θ_T is regular by Lemma 4.16. Therefore, (b) implies (c), and the converse holds by Lemma 5.2. ∎

In [5] it was stated (as Lemma 8.2), in different terms, that θ_T is of finite index if $T \in Rec(\Sigma, X)$, but the example meant to disprove the converse, appears incorrect. Nevertheless, their Theorem 1 essentially expresses the equivalence of (a) and (c) of our Proposition 6.3. From Propositions 6.3 and 4.17 we get:

Corollary 6.4. *An unranked tree language T is recognizable if and only if the syntactic algebra* $\mathrm{SA}(T)$ *is regular.*

Next we note that the family of recognizable unranked tree languages is closed under the operations that define our varieties of unranked tree languages.

Proposition 6.5. *The following hold for all alphabets Σ, Ω, X and Y.*

(a) *$\emptyset \in Rec(\Sigma, X)$, and $Rec(\Sigma, X)$ is closed under all Boolean operations.*

(b) *If $T \in Rec(\Sigma, X)$, then $p^{-1}(T) := \{t \in T_\Sigma(X) \mid p(t) \in T\} \in Rec(\Sigma, X)$ for every context $p \in C_\Sigma(X)$.*

(c) *If $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ is a g-morphism, then $T\varphi^{-1} \in Rec(\Sigma, X)$ for every $T \in Rec(\Omega, Y)$.*

Proof. Clearly, $\emptyset$ and $T_\Sigma(X)$ are recognized by any Σ-algebra, and the rest of the proposition follows from Corollary 6.4 and Propositions 4.7 and 5.8. ∎

We shall need the following fact about the sets $p^{-1}(T)$.

Lemma 6.6. *If $T \in Rec(\Sigma, X)$, then the set $\{p^{-1}(T) \mid p \in C_\Sigma(X)\}$ is finite.*

Proof. By Proposition 5.7(c) every set $p^{-1}(T)$ is saturated by θ_T. On the other hand, it follows from Proposition 6.3 that θ_T has just finitely many equivalence classes. Hence, the number of different sets $p^{-1}(T)$ must be finite, too. ∎

We say that $T \in Rec(\Sigma, X)$ is *effectively given* if $T = F\varphi^{-1}$, for an effectively given morphism $\varphi : \mathcal{T}_\Sigma(X) \to \mathcal{A}$, an effectively given regular algebra $\mathcal{A} = (A, \Sigma)$ and an effectively given subset $F \subseteq A$.

Proposition 6.7. *If $T \in Rec(\Sigma, X)$ is effectively given, then* $\mathrm{SA}(T)$ *can be effectively constructed.*

Proof. If $T = F\varphi^{-1}$ is effectively given as in the above definition, we may assume that $\varphi : \mathcal{T}_\Sigma(X) \to \mathcal{A}$ epimorphism. Then $\mathrm{SA}(T) \cong \mathrm{SA}(F)$ by Corollary 5.9, and $\mathrm{SA}(F)$ can be constructed by Proposition 5.3. ∎

7. Varieties of unranked tree languages

A *family of unranked tree languages* $\mathcal{V}$ assigns to each pair Σ, X a set $\mathcal{V}(\Sigma, X)$ of ΣX-tree languages. We write $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}_{\Sigma,X}$ with the understanding that Σ and X range over all operator alphabets and leaf alphabets, respectively. The inclusion relation, unions and intersections of these families are defined by the natural componentwise conditions. In particular, if $\mathcal{U} = \{\mathcal{U}(\Sigma, X)\}_{\Sigma,X}$ and $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}_{\Sigma,X}$ are two such families, then $\mathcal{U} \subseteq \mathcal{V}$ means that $\mathcal{U}(\Sigma, X) \subseteq \mathcal{V}(\Sigma, X)$ for all Σ and X, and $\mathcal{U} \cap \mathcal{V} = \{\mathcal{U}(\Sigma, X) \cap \mathcal{V}(\Sigma, X)\}_{\Sigma,X}$.

Definition 7.1. A *variety of unranked tree languages (VUT)* is a family of unranked tree languages $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}_{\Sigma,X}$ such that for all Σ, Ω, X and Y,

(V1) $\emptyset \neq \mathcal{V}(\Sigma, X) \subseteq Rec(\Sigma, X)$,

(V2) if $T \in \mathcal{V}(\Sigma, X)$, then also $T_\Sigma(X) \setminus T$ belongs to $\mathcal{V}(\Sigma, X)$,

(V3) if $T, U \in \mathcal{V}(\Sigma, X)$, then $T \cap U \in \mathcal{V}(\Sigma, X)$,

(V4) if $T \in \mathcal{V}(\Sigma, X)$, then $p^{-1}(T) \in \mathcal{V}(\Sigma, X)$ for every $p \in C_\Sigma(X)$, and

(V5) if $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ is a g-morphism, then $T\varphi^{-1} \in \mathcal{V}(\Sigma, X)$ for every $T \in \mathcal{V}(\Omega, Y)$.

Let **VUT** denote the class of all VUTs.

Clearly, the intersection of any family of VUTs and the union of any directed family of VUTs are VUTs. Hence $(\mathbf{VUT}, \subseteq)$ is an algebraic lattice. The least VUT is $Triv := \{\{\emptyset, T_\Sigma(X)\}\}_{\Sigma,X}$ and the greatest one is $Rec := \{Rec(\Sigma, X)\}_{\Sigma,X}$.

Proposition 7.2. *If $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}_{\Sigma,X}$ is a VUT and $T \in \mathcal{V}(\Sigma, X)$ for some Σ and X, then every θ_T-class is also in $\mathcal{V}(\Sigma, X)$.*

Proof. It follows from the definition of θ_T that for any $t \in T_\Sigma(X)$,

$$t\theta_T = \bigcap\{p^{-1}(T) \mid p \in C_\Sigma(X), p(t) \in T\} \setminus \bigcup\{p^{-1}(T) \mid p \in C_\Sigma(X), p(t) \notin T\}.$$

By Lemma 6.6, this shows that $t\theta_T$ is in $\mathcal{V}(\Sigma, X)$. ∎

Next we introduce systems of congruences that yield VUTs. For a nonempty subset H of a lattice L, let $[H)$ denote the filter generated by H, i.e., the set of all $b \in L$ such that $a_1 \wedge \cdots \wedge a_n \leq b$ for some $n \geq 1$ and $a_1, \dots, a_n \in H$. As special cases, we get the principal filters $[a) := [\{a\}] = \{b \in L \mid a \leq b\}$ $(a \in L)$.

By a *family of regular g-congruences* we mean a mapping $\mathcal{C}$ that assigns to each pair Σ, X a subset $\mathcal{C}(\Sigma, X)$ of $\mathrm{RGCon}(\mathcal{T}_\Sigma(X))$. Again, we write $\mathcal{C} = \{\mathcal{C}(\Sigma, X)\}_{\Sigma,X}$ and order these families by the componentwise inclusion relation.

Definition 7.3. A family of regular g-congruences $\mathcal{C} = \{\mathcal{C}(\Sigma, X)\}_{\Sigma,X}$ is a *variety of regular g-congruences (VRC)* if the following hold for all Σ, Ω, X and Y.

(C1) For every $\sigma \in \mathrm{Eq}(\Sigma)$, $\mathcal{C}(\Sigma, X)_\sigma := \{\theta \in \mathrm{RCon}(\mathcal{T}_\Sigma(X)) \mid (\sigma, \theta) \in \mathcal{C}(\Sigma, X)\}$ is a filter of $\mathrm{RCon}(\mathcal{T}_\Sigma(X))$.

(C2) If $(\sigma, \theta) \in \mathcal{C}(\Sigma, X)$, $\sigma' \in \mathrm{Eq}(\Sigma)$ and $\sigma' \subseteq \mathrm{M}(\theta)$, then $(\sigma', \theta) \in \mathcal{C}(\Sigma, X)$.

(C3) If $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ is a g-morphism and $(\omega, \theta) \in \mathcal{C}(\Omega, Y)$, then $(\iota \circ \omega \circ \iota^{-1}, \varphi \circ \theta \circ \varphi^{-1}) \in \mathcal{C}(\Sigma, X)$.

Clause (C3) anticipated the following fact. The lemma has a simple proof.

Lemma 7.4. $(\iota \circ \omega \circ \iota^{-1}, \varphi \circ \theta \circ \varphi^{-1}) \in \mathrm{RGCon}(\mathcal{T}_\Sigma(X))$ *for any g-morphism* $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ *and any* $(\omega, \theta) \in \mathrm{RGCon}(\mathcal{T}_\Omega(Y))$.

For any family $\mathcal{C} = \{\mathcal{C}(\Sigma, X)\}_{\Sigma,X}$ of regular g-congruences, let $\mathcal{C}^t$ be the family of recognizable unranked tree languages such that for all Σ and X,

$$\mathcal{C}^t(\Sigma, X) := \{T \subseteq T_\Sigma(X) \mid (\Delta_\Sigma, \theta_T) \in \mathcal{C}(\Sigma, X)\}.$$

Proposition 7.5. *If $\mathcal{C}$ is a VRC, then $\mathcal{C}^t$ is a VUT.*

Proof. Most of the proposition follows directly from the definitions involved and Proposition 5.7. Let us verify conditions (V1) and (V5) of Definition 7.1.

Firstly, for any Σ and X, $\mathcal{C}^t(\Sigma, X) \neq \emptyset$ as $(\Delta_\Sigma, \nabla_{T_\Sigma(X)}) \in \mathcal{C}(\Sigma, X)$ and $\theta_\emptyset = \nabla_{T_\Sigma(X)}$. For any $T \in \mathcal{C}^t(\Sigma, X)$, $\theta_T \in \mathrm{RCon}(\mathcal{T}_\Sigma(X))$ as $(\Delta_\Sigma, \theta_T) \in \mathcal{C}(\Sigma, X)$. By Proposition 6.3 this means that T is recognizable. Hence, $\mathcal{C}^t$ satisfies (V1).

If $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ is a g-morphism and $T \in \mathcal{C}^t(\Omega, Y)$, then $(\Delta_\Omega, \theta_T) \in \mathcal{C}(\Omega, Y)$. Hence $(\iota \circ \Delta_\Omega \circ \iota^{-1}, \varphi \circ \theta_T \circ \varphi^{-1}) \in \mathcal{C}(\Sigma, X)$ by condition (C3). Moreover, $\theta_T \in \mathrm{RCon}(\mathcal{T}_\Omega(Y))$ implies $(\Delta_\Sigma, \varphi \circ \theta_T \circ \varphi^{-1}) \in \mathrm{RGCon}(\mathcal{T}_\Sigma(X))$ by Lemma 7.4. Hence, $(\Delta_\Sigma, \varphi \circ \theta_T \circ \varphi^{-1}) \in \mathcal{C}(\Sigma, X)$ by (C2). On the other hand, $\varphi \circ \theta_T \circ \varphi^{-1} \subseteq \theta_{T\varphi^{-1}}$ by Proposition 5.7(d), and hence $(\Delta_\Sigma, \theta_{T\varphi^{-1}}) \in \mathcal{C}(\Sigma, X)$ by (C1). This means that $T\varphi^{-1} \in \mathcal{C}^t(\Sigma, X)$ and therefore $\mathcal{C}^t$ satisfies (V5). ∎

Many VUTs are unions of ascending chains or directed families of simpler VUTs. In the ranked case the basic varieties forming such a family are usually defined by so-called *principal varieties of congruences* [24, 26] that consist of principal filters. Here we need the following more general notion.

Definition 7.6. We call $\Theta = \{\theta(\Sigma, X)\}_{\Sigma,X}$, where $\theta(\Sigma, X) \in \mathrm{Con}(\mathcal{T}_\Sigma(X))$ for all Σ and X, a *consistent system of congruences (CSC)* if $\theta(\Sigma, X) \subseteq \varphi \circ \theta(\Omega, Y) \circ \varphi^{-1}$ for all Σ, Ω, X and Y, and every g-morphism $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$, and then let $\mathcal{C}_\Theta := \{\mathcal{C}_\Theta(\Sigma, X)\}_{\Sigma,X}$ be the family of regular congruences, where for all Σ and X, $\mathcal{C}_\Theta(\Sigma, X) := \{(\sigma, \theta) \in \mathrm{RGCon}(\mathcal{T}_\Sigma(X)) \mid \theta(\Sigma, X) \subseteq \theta\}$.

Proposition 7.7. *For any CSC Θ, $\mathcal{C}_\Theta$ is a VRC.*

Proof. (C1) If $\sigma \in \mathrm{Eq}(\Sigma)$, then $(\sigma, \nabla_{T_\Sigma(X)}) \in \mathcal{C}_\Theta(\Sigma, X)$, and hence $\mathcal{C}_\Theta(\Sigma, X)_\sigma \neq \emptyset$. If $\theta \subseteq \rho$ and $\theta \in \mathcal{C}_\Theta(\Sigma, X)_\sigma$, then $\theta(\Sigma, X) \subseteq \theta \subseteq \rho$. On the other hand, $(\sigma, \theta) \in \mathrm{RGCon}(\mathcal{T}_\Sigma(X))$ implies $(\sigma, \rho) \in \mathrm{RGCon}(\mathcal{T}_\Sigma(X))$ by Lemma 4.16. Hence $\rho \in \mathcal{C}_\Theta(\Sigma, X)_\sigma$. If $\theta, \rho \in \mathcal{C}_\Theta(\Sigma, X)_\sigma$, then $\theta(\Sigma, X) \subseteq \theta, \rho$ and $(\sigma, \theta), (\sigma, \rho) \in \mathrm{RGCon}(\mathcal{T}_\Sigma(X))$, and therefore $\theta(\Sigma, X) \subseteq \theta \cap \rho$ and – again by Lemma 4.16, $(\sigma, \theta \cap \rho) = (\sigma, \theta) \wedge (\sigma, \rho) \in \mathrm{RGCon}(\mathcal{T}_\Sigma(X))$. This means that $\theta \cap \rho \in \mathcal{C}_\Theta(\Sigma, X)_\sigma$, and thus we have shown that $\mathcal{C}_\Theta(\Sigma, X)_\sigma$ is a filter in $\mathrm{RCon}(\mathcal{T}_\Sigma(X))$.

(C2) If $(\sigma, \theta) \in \mathcal{C}_\Theta(\Sigma, X)$, $\sigma' \in \mathrm{Eq}(\Sigma)$ and $\sigma' \subseteq \mathrm{M}(\theta)$, then $(\sigma', \theta) \in \mathcal{C}_\Theta(\Sigma, X)$ because $\theta(\Sigma, X) \subseteq \theta$ by the first assumption.

(C3) If $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ is a g-morphism and $(\omega, \theta) \in \mathcal{C}_\Theta(\Omega, Y)$, then $(\iota \circ \omega \circ \iota^{-1}, \varphi \circ \theta \circ \varphi^{-1}) \in \mathrm{RGCon}(\mathcal{T}_\Sigma(X))$ by Lemma 7.4, and $\theta(\Omega, Y) \subseteq \theta$ implies $\theta(\Sigma, X) \subseteq \varphi \circ \theta(\Omega, Y) \circ \varphi^{-1} \subseteq \varphi \circ \theta \circ \varphi^{-1}$. So $(\iota \circ \omega \circ \iota^{-1}, \varphi \circ \theta \circ \varphi^{-1}) \in \mathcal{C}_\Theta(\Sigma, X)$. ■

Any VRC $\mathcal{C}_\Theta$ defined by a CSC Θ and the corresponding VUT $\mathcal{C}^t_\Theta$ are called *quasi-principal*. The following is a direct consequence of the definition of $\mathcal{C}^t_\Theta$.

Lemma 7.8. *For any CSC $\Theta = \{\theta(\Sigma, X)\}_{\Sigma,X}$ and all Σ and X, $\mathcal{C}^t_\Theta(\Sigma, X) = \{T \in Rec(\Sigma, X) \mid \theta(\Sigma, X) \subseteq \theta_T\}$.*

8. THE VARIETY THEOREM

We shall now prove that the following maps $\mathbf{K} \mapsto \mathbf{K}^t$ and $\mathcal{V} \mapsto \mathcal{V}^a$ form a pair of mutually inverse isomorphisms between the lattices $(\mathbf{VRA}, \subseteq)$ and $(\mathbf{VUT}, \subseteq)$.

Definition 8.1. For any VRA $\mathbf{K}$, let $\mathbf{K}^t = \{\mathbf{K}^t(\Sigma, X)\}$ be the family of recognizable unranked tree languages in which $\mathbf{K}^t(\Sigma, X) := \{T \subseteq T_\Sigma(X) \mid \mathrm{SA}(T) \in \mathbf{K}\}$ for all Σ and X. For any VUT $\mathcal{V} = \{\mathcal{V}(\Sigma, X)\}_{\Sigma,X}$, let $\mathcal{V}^a$ be the VRA generated by the algebras $\mathrm{SA}(T)$, where $T \in \mathcal{V}(\Sigma, X)$ for some Σ and X.

Note that $\mathcal{V}^a$ is a well-defined VRA for every VUT $\mathcal{V}$ because any algebra $\mathrm{SA}(T)$ with $T \in \mathcal{V}(\Sigma, X)$ is regular. By Lemma 4.12, the maps $\mathbf{K} \mapsto \mathbf{K}^t$ and $\mathcal{V} \mapsto \mathcal{V}^a$ could be defined equivalently using reduced syntactic algebras.

Lemma 8.2. *For any VRA* $\mathbf{K}$, $\mathbf{K}^t$ *is a VUT.*

Proof. It follows from Corollary 6.4 that $\mathbf{K}^t(\Sigma, X) \subseteq Rec(\Sigma, X)$ for all Σ and X. Moreover, $\mathbf{K}^t(\Sigma, X) \neq \emptyset$ because $\mathbf{K}$ contains at least the trivial Σ-algebras. Hence, $\mathbf{K}^t$ satisfies the first condition of Definition 7.1. Conditions (V2)–(V4) follow immediately from Proposition 5.8 and the fact that $\mathbf{K}$ is a VRA. As to (V5), we argue as follows. If $T \in \mathbf{K}^t(\Omega, Y)$, then $\mathrm{SA}(T) \in \mathbf{K}$, and hence $\mathrm{RA}(T) \in \mathbf{K}$ by Lemma 4.12. By Proposition 5.8(d) this means that $\mathrm{RA}(T\varphi^{-1}) \in \mathbf{K}$. Hence also $\mathrm{SA}(T\varphi^{-1}) \in \mathbf{K}$ by Lemma 4.12, from which $T\varphi^{-1} \in \mathbf{K}^t(\Sigma, X)$ follows. ■

Since it is clear that the maps $\mathbf{K} \mapsto \mathbf{K}^t$ and $\mathcal{V} \mapsto \mathcal{V}^a$ are order-preserving, it remains to be shown that they are inverses of each other.

Lemma 8.3. $\mathbf{K}^{ta} = \mathbf{K}$ *for every VRA* $\mathbf{K}$.

Proof. The VRA $\mathbf{K}^{ta}$ is generated by the algebras $\mathrm{SA}(T)$, where $T \in \mathbf{K}^t(\Sigma, X)$ for some Σ and X, but these algebras are also in $\mathbf{K}$. Hence, $\mathbf{K}^{ta} \subseteq \mathbf{K}$.

On the other hand, by Proposition 5.4, $\mathbf{K}$ is generated by regular syntactic algebras. Let $\mathcal{A}$ be any such generating algebra. If X is sufficiently large, there is an epimorphism $\varphi : \mathcal{T}_\Sigma(X) \to \mathcal{A}$. Furthermore, $\mathcal{A}$ has a disjunctive subset D by Proposition 5.4. The ΣX-tree language $T := D\varphi^{-1}$ is recognizable, and $\mathrm{SA}(T) \cong \mathrm{SA}(D)$ by Corollary 5.9. Also, $\mathcal{A} \cong \mathrm{SA}(D)$ because D is disjunctive, and therefore also $\mathrm{SA}(T) \in \mathbf{K}$, which shows that $T \in \mathbf{K}^t(\Sigma, X)$. As this means that $\mathrm{SA}(T) \in \mathbf{K}^{ta}$, we get $\mathcal{A} \in \mathbf{K}^{ta}$ and can conclude that $\mathbf{K} \subseteq \mathbf{K}^{ta}$. ■

Lemma 8.4. $\mathcal{V}^{at} = \mathcal{V}$ *for every VUT* $\mathcal{V}$.

Proof. If $T \in \mathcal{V}(\Sigma, X)$, then $\mathrm{SA}(T) \in \mathcal{V}^a$ implies $T \in \mathcal{V}^{at}(\Sigma, X)$, and hence $\mathcal{V} \subseteq \mathcal{V}^{at}$. If $T \in \mathcal{V}^{at}(\Sigma, X)$, then $\mathrm{SA}(T) \in \mathcal{V}^a$, and by Proposition 4.11

$$\mathrm{SA}(T) \preceq \varkappa(\mathrm{SA}(U_1), \ldots, \mathrm{SA}(U_n)),$$

for some $n \geq 0$, $U_1 \in \mathcal{V}(\Sigma_1, X_1)$, $\ldots$, $U_n \in \mathcal{V}(\Sigma_n, X_n)$ for some alphabets $\Sigma_1, \ldots, \Sigma_n$ and $X_1, \ldots, X_n$, and a mapping $\varkappa$ from Σ to $\Sigma_1 \times \cdots \times \Sigma_n$.

For each $i \in [n]$, denote $\mathcal{T}_{\Sigma_i}(X_i)$ by $\mathcal{T}_i$, and let $\mathrm{SA}(U_i) = (A_i, \Sigma_i)$. Furthermore, let $\varphi_i : \mathcal{T}_i \to \mathrm{SA}(U_i), t \mapsto t\theta_{U_i}$, be the syntactic morphism of U_i. By Proposition 6.2, there exist a morphism $\varphi : \mathcal{T}_\Sigma(X) \to \varkappa(\mathrm{SA}(U_1), \ldots, \mathrm{SA}(U_n))$ and a subset $F \subseteq A_1 \times \cdots \times A_n$ such that $T = F\varphi^{-1}$. For each $i \in [n]$, define $\lambda_i : \Sigma \to \Sigma_i$ by $\lambda_i(f) = f_i$ for any $f \in \Sigma$ if $\varkappa(f) = (f_1, \ldots, f_n)$. The syntactic morphisms φ_i yield an epimorphism

$$\eta : \varkappa(\mathcal{T}_1, \ldots, \mathcal{T}_n) \to \varkappa(\mathrm{SA}(U_1), \ldots, \mathrm{SA}(U_n)), (t_1, \ldots, t_n) \mapsto (t_1\varphi_1, \ldots, t_n\varphi_n),$$

and for each $i \in [n]$, we get the g-morphisms

$$(\lambda_i, \tau_i) : \varkappa(\mathcal{T}_1, \ldots, \mathcal{T}_n) \to \mathcal{T}_i \text{ and } (\lambda_i, \pi_i) : \varkappa(\mathrm{SA}(U_1), \ldots, \mathrm{SA}(U_n)) \to \mathrm{SA}(U_i),$$

where $\tau_i : (t_1, \ldots, t_n) \mapsto t_i$ and $\pi_i : (a_1, \ldots, a_n) \mapsto a_i$ are the respective i^{th} projections. Clearly, $\tau_i \varphi_i = \eta \pi_i$ for any $i \in [n]$. Since η is surjective, there is a $\psi_0 : X \to T(\Sigma_1, X_1) \times \cdots \times T(\Sigma_n, X_n)$ such that $x\psi_0\eta = x\varphi$ for all $x \in X$. If $\psi : \mathcal{T}_\Sigma(X) \to \varkappa(\mathcal{T}_1, \ldots, \mathcal{T}_n)$ is the homomorphic extension of ψ_0, then $\psi\eta = \varphi$.

Now, T is the union of finitely many sets $\mathbf{a}\varphi^{-1}$ with $\mathbf{a} = (a_1, \ldots, a_n) \in F$. Since $\varphi\pi_i = \psi\tau_i\varphi_i$ for each $i \in [n]$, we have

$$\mathbf{a}\varphi^{-1} = \bigcap\{a_i(\varphi\pi_i)^{-1} \mid i \in [n]\} = \bigcap\{(a_i\varphi_i^{-1})(\psi\tau_i)^{-1} \mid i \in [n]\},$$

where each $a_i\varphi_i^{-1}$ is a θ_{U_i}-class, and therefore belongs to $\mathcal{V}(\Sigma_i, X_i)$ by Proposition 7.2. This implies that $(a_i\varphi_i^{-1})(\psi\tau_i)^{-1} \in \mathcal{V}(\Sigma, X)$ for every $i \in [n]$, and hence also $T \in \mathcal{V}(\Sigma, X)$. This concludes the proof of $\mathcal{V}^{at} \subseteq \mathcal{V}$. ∎

The above results can be summed up as the following variety theorem.

Theorem 8.5. *The mappings* $\mathbf{K} \to \mathbf{K}^t$ *and* $\mathcal{V} \mapsto \mathcal{V}^a$ *define mutually inverse isomorphisms between the lattices* $(\mathbf{VRA}, \subseteq)$ *and* $(\mathbf{VUT}, \subseteq)$.

9. Some varieties of unranked tree languages

We shall now introduce several varieties of unranked tree languages that correspond to some known general varieties of ranked tree languages, most of which can be found in [25]. The following obvious facts are helpful in many of the examples. Note that (b) does not follow directly from (a) as in the ranked case.

Lemma 9.1. *Let* $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ *be a g-morphism.*

(a) $\mathrm{hg}(t\varphi) \geq \mathrm{hg}(t)$ *for every* $t \in T_\Sigma(X)$.

(b) $t\varphi^{-1}$ *is finite for every* $t \in T_\Omega(Y)$.

9.1. Nilpotency

We begin with the simplest nontrivial VUT. For any Σ and X, let $Nil(\Sigma, X)$ consist of all finite ΣX-tree languages and their complements in $T_\Sigma(X)$, and let $Nil := \{Nil(\Sigma, X)\}_{\Sigma, X}$. In view of Proposition 6.5(a), $Nil \subseteq Rec$ since any singleton $\{t\} \subseteq T_\Sigma(X)$ is obviously recognizable. Clearly, every $Nil(\Sigma, X)$ is closed under all Boolean operations, and $p^{-1}(T)$ and $T\varphi^{-1}$ are finite if T is finite; for $T\varphi^{-1}$ this follows from Lemma 9.1(b).

Nilpotent unranked algebras cannot be defined in terms of the height of trees, as in the ranked case, since there are infinitely many trees of any height ≥ 1. Let the *size* $\mathrm{size}(t)$ of a tree $t \in T_\Sigma(X)$ be the number of its nodes. We call $\mathcal{A} = (A, \Sigma)$ *nilpotent* if there exist a $k \geq 1$ and an element $a_0 \in A$, the *absorbing state*, such that for any X and $t \in T_\Sigma(X)$, if $\mathrm{size}(t) \geq k$, then $t^{\mathcal{A}}(\alpha) = a_0$ for every $\alpha : X \to A$. The least k for which this holds is its *degree (of nilpotency)*. For each $k \geq 1$, let $\mathbf{Nil}_k$ be the class of regular nilpotent algebras of degree $\leq k$, and let $\mathbf{Nil} = \bigcup_{k \geq 1} \mathbf{Nil}_k$. Obviously, $\mathbf{Nil}_1 \subset \mathbf{Nil}_2 \subset \mathbf{Nil}_3 \subset \ldots$. It is easy to show that each $\mathbf{Nil}_k$, and hence also $\mathbf{Nil}$, is a VRA.

To prove $\mathbf{Nil}^t \subseteq Nil$, let $T = F\varphi^{-1}$ for some $\mathcal{A} = (A, \Sigma)$ in $\mathbf{Nil}_k$, a morphism $\varphi : \mathcal{T}_\Sigma(X) \to \mathcal{A}$ and $F \subseteq A$. If a_0 is the absorbing state of $\mathcal{A}$ and $\alpha : X \to A$ is the restriction of φ to X, then $t^{\mathcal{A}}(\alpha) = a_0$ whenever $\mathrm{size}(t) \geq k$. Hence T is finite if $a_0 \notin F$ and co-finite if $a_0 \in F$.

To prove $Nil \subseteq \mathbf{Nil}^t$, consider any finite ΣX-tree language T. Let $k := \max\{\mathrm{size}(t) \mid t \in T\} + 1$ (for $T = \emptyset$, let $k = 1$). We construct a nilpotent algebra $\mathcal{A} = (A, \Sigma)$ recognizing T as follows. Let $B := \{t \in T_\Sigma(X) \mid \mathrm{size}(t) < k\}$ and $A := B \cup \{a_0\}$ (with $a_0 \notin B$), and for all $f \in \Sigma$, $m \geq 0$ and $b_1, \ldots, b_m \in A$ set

$$f_{\mathcal{A}}(b_1, \ldots, b_m) = \begin{cases} f(b_1, \ldots, b_m) & \text{if } f(b_1, \ldots, b_m) \in B; \\ a_0 & \text{otherwise.} \end{cases}$$

It is clear that $\mathcal{A}$ is regular and that $t^{\mathcal{A}}(\alpha) = a_0$ for every $\alpha : X \to A$ whenever $\mathrm{size}(t) \geq k$. If $\varphi : \mathcal{T}_\Sigma(X) \to \mathcal{A}$ is the morphism such that $x\varphi = x$ for all $x \in X$, then $t\varphi = t$ if $\mathrm{size}(t) < k$ and $t\varphi = a_0$ otherwise, and therefore $T = T\varphi^{-1}$. For a co-finite T, we construct such an $\mathcal{A}$ for $S := T_\Sigma(X) \setminus T$ and get T as $(A \setminus S)\varphi^{-1}$. This completes the proof of the following result.

Proposition 9.2. *Nil is the VUT corresponding to the VRA* **Nil**.

9.2. Definiteness

Next we consider tree languages determined by root segments of some given height. The *k-root* $\mathrm{rt}_k(t)$ of a ΣX-tree t is defined as follows:

(0) $\mathrm{rt}_0(t) = \varepsilon$, where ε represents the empty root segment, for all $t \in T_\Sigma(X)$;

(1) $\mathrm{rt}_1(t) = \mathrm{root}(t)$ for every $t \in T_\Sigma(X)$;

(2) for $k \geq 2$, $\mathrm{rt}_k(t) = t$ if $\mathrm{hg}(t) < k$, and $\mathrm{rt}_k(t) = f(\mathrm{rt}_{k-1}(t_1), \ldots, \mathrm{rt}_{k-1}(t_m))$ if $\mathrm{hg}(t) \geq k$ and $t = f(t_1, \ldots, t_m)$.

A recognizable unranked ΣX-tree language T is *k-definite* if for all $s, t \in T_\Sigma(X)$, if $\mathrm{rt}_k(s) = \mathrm{rt}_k(t)$ and $s \in T$, then $t \in T$, and it is *definite* if it is k-definite for some $k \geq 0$. Let $Def_k = \{Def_k(\Sigma, X)\}_{\Sigma,X}$ and $Def = \{Def(\Sigma, X)\}_{\Sigma,X}$ be the families of k-definite ($k \geq 0$) and all definite tree languages, respectively.

Clearly $Def_0 \subset Def_1 \subset Def_2 \subset \ldots$ and $Def = \bigcup_{k \geq 0} Def_k$. We could verify the conditions (V1)–(V5) directly, but let us present a CSC for Def_k.

For any $k \geq 0$, Σ and X, define the relation $\delta_k(\Sigma, X)$ on $T_\Sigma(X)$ by $s\, \delta_k(\Sigma, X)\, t$ if and only if $\mathrm{rt}_k(s) = \mathrm{rt}_k(t)$ for all $s, t \in T_\Sigma(X)$. Note that for every $k \geq 2$, there are infinitely many $\delta_k(\Sigma, X)$-classes. Let $\Delta(k) := \{\delta_k(\Sigma, X)\}_{\Sigma,X}$. The following lemma can be verified by induction on $k \geq 1$.

Lemma 9.3. *Let $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ be a g-morphism. Then $\mathrm{rt}_k(t\varphi) = \mathrm{rt}_k(\mathrm{rt}_k(t)\varphi)$ for all $t \in T_\Sigma(X)$ and $k \geq 1$.*

Proposition 9.4. *For each $k \geq 0$, $\Delta(k)$ is a CSC and Def_k is the quasi-principal VUT defined by it. Hence, Def is also a VUT.*

Proof. Fix a $k \geq 0$. That $\Delta(k)$ is a CSC that defines Def_k is shown as follows. Here Σ, Ω, X and Y are any alphabets.

Firstly, $\delta_k(\Sigma, X)$ is a congruence of $\mathcal{T}_\Sigma(X)$. For $k = 0$, this is obvious. For $k > 0$, we use the simple fact that $\mathrm{rt}_k(s) = \mathrm{rt}_k(t)$ implies $\mathrm{rt}_{k-1}(s) = \mathrm{rt}_{k-1}(t)$.

Secondly, to prove that $\delta_k(\Sigma, X) \subseteq \varphi \circ \delta_k(\Omega, Y) \circ \varphi^{-1}$ for any g-morphism $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$, it suffices to show that if $s, t \in T_\Sigma(X)$ and $\mathrm{rt}_k(s) = \mathrm{rt}_k(t)$, then $\mathrm{rt}_k(s\varphi) = \mathrm{rt}_k(t\varphi)$, and this follows from Lemma 9.3.

Thirdly, if $T \in Rec(\Sigma, X)$, then $T \in Def_k(\Sigma, X)$ if and only if T is saturated by $\delta_k(\Sigma, X)$, and by Lemma 5.2 this is the case if and only if $\delta_k(\Sigma, X) \subseteq \theta_T$.

Finally, as the union of a chain of VUTs, also Def is a VUT. ∎

9.3. Reverse definiteness

Next we consider tree languages defined by the subtrees of a given height of their trees. A ΣX-tree s is a *subtree* of a ΣX-tree t if $t = p(s)$ for some $p \in C_\Sigma(X)$. For any $t \in T_\Sigma(X)$, let $\mathrm{st}(t)$ denote the set of subtrees of t, and for each $k \geq 0$, let $\mathrm{st}_k(t) = \{s \in \mathrm{st}(t) \mid \mathrm{hg}(s) < k\}$. Note that $\mathrm{st}_0(t) = \emptyset$ for every t.

We call a recognizable unranked ΣX-tree language T *reverse k-definite* if for all $s, t \in T_\Sigma(X)$, if $\mathrm{st}_k(s) = \mathrm{st}_k(t)$ and $s \in T$, then $t \in T$, and it is *reverse definite* if it is reverse k-definite for some $k \geq 0$. Let $RDef_k = \{RDef_k(\Sigma, X)\}_{\Sigma,X}$ and $RDef = \{RDef(\Sigma, X)\}_{\Sigma,X}$ be the families of the reverse k-definite ($k \geq 0$) and the reverse definite tree languages. Clearly $RDef_k \subset RDef_{k+1}$ for all $k \geq 0$.

For any $k \geq 0$, Σ, X and $s, t \in T_\Sigma(X)$, let $s\, \rho_k(\Sigma, X)\, t$ if and only if $\mathrm{st}_k(s) = \mathrm{st}_k(t)$, and let $\mathrm{P}(k) := \{\rho_k(\Sigma, X)\}_{\Sigma,X}$. The following lemma can easily be proved by induction on k.

Lemma 9.5. *Let $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ be a g-morphism. Then $\mathrm{st}_k(t\varphi) = \bigcup\{\mathrm{st}_k(s\varphi) \mid s \in \mathrm{st}_k(t)\}$ for all $t \in T_\Sigma(X)$ and $k \geq 0$.*

The following proposition can now be proved similarly as Proposition 9.4.

Proposition 9.6. *For each $k \geq 0$, $\mathrm{P}(k)$ is a CSC and $RDef_k$ is the corresponding quasi-principal VUT. Hence, $RDef$ is also a VUT.*

9.4. Generalized definiteness

Let us now combine root and subtree tests. For any $h, k \geq 0$, an unranked ΣX-tree language T is *h, k-definite* if for all $s, t \in T_\Sigma(X)$, if $\mathrm{st}_h(s) = \mathrm{st}_h(t)$ and $\mathrm{rt}_k(s) = \mathrm{rt}_k(t)$, then $s \in T$ if and only if $t \in T$, and it is *generalized definite* if it is h, k-definite for some $h, k \geq 0$. Let $GDef_{h,k} = \{GDef_{h,k}(\Sigma, X)\}_{\Sigma,X}$ and $GDef = \{GDef(\Sigma, X)\}_{\Sigma,X}$ be the families of all recognizable h, k-definite ($h, k \geq 0$) and all recognizable generalized definite tree languages. Clearly $GDef_{h,k} \subseteq GDef_{h',k'}$ whenever $h \leq h'$ and $k \leq k'$, and $GDef = \bigcup_{h,k\geq 0} GDef_{h,k}$.

For any $h, k \geq 0$, Σ and X, let $\gamma_{h,k}(\Sigma, X) = \rho_h(\Sigma, X) \cap \delta_k(\Sigma, X)$, and let $\Gamma(h, k) := \{\gamma_{h,k}(\Sigma, X))\}_{\Sigma,X}$. The following proposition can be proved simply by combining the arguments used in the previous two examples.

Proposition 9.7. *For all $h, k \geq 0$, $\Gamma(h, k)$ is a CSC and $GDef_{h,k}$ is the quasi-principal VUT defined by it. Hence, $GDef$ is also a VUT.*

9.5. Local testability

For defining local testability we need an appropriate notion of "local pattern". For any $k \geq 2$, Σ and X, the set $\mathrm{fork}_k(t)$ of *k-forks* of $t \in T_\Sigma(X)$ is defined as follows:

(1) $\mathrm{fork}_k(t) = \emptyset$ if $\mathrm{hg}(t) < k - 1$;

(2) $\mathrm{fork}_k(t) = \{\mathrm{rt}_k(t)\} \cup \mathrm{fork}_k(t_1) \cup \cdots \cup \mathrm{fork}_k(t_m)$ if $\mathrm{hg}(t) \geq k - 1$ and $t = f(t_1, \ldots, t_m)$.

Clearly, $\mathrm{fork}_k(t)$ is a finite set of ΣX-trees of height $k - 1$. For example, if $t = f(x, f(y))$, then $\mathrm{fork}_2(t) = \{f(x, f), f(y)\}$, $\mathrm{fork}_3(t) = \{t\}$ and $\mathrm{fork}_k(t) = \emptyset$ for all $k \geq 4$. Note that the set of all possible k-forks of ΣX-trees is infinite.

Now, let $\lambda_k(\Sigma, X)$ be the relation on $T_\Sigma(X)$ such that for any $s, t \in T_\Sigma(X)$,

$$s\, \lambda_k(\Sigma, X)\, t \;:\Leftrightarrow\; \mathrm{st}_{k-1}(s) = \mathrm{st}_{k-1}(t), \mathrm{rt}_{k-1}(s) = \mathrm{rt}_{k-1}(t), \mathrm{fork}_k(s) = \mathrm{fork}_k(t).$$

It is easy to see that $\lambda_k(\Sigma, X) \in \mathrm{Con}(\mathcal{T}_\Sigma(X))$. An unranked ΣX-tree language is *k-testable* if it is saturated by $\lambda_k(\Sigma, X)$, and it is *locally testable* if it is k-testable for some $k \geq 2$. Let $Loc_k(\Sigma, X)$ be the set of all recognizable k-testable ΣX-tree languages, and let $Loc(\Sigma, X) := \bigcup_{k\geq 2} Loc_k(\Sigma, X)$. Note that although $\mathrm{fork}_k(s) = \mathrm{fork}_k(t)$ does not imply $\mathrm{fork}_{k-1}(s) = \mathrm{fork}_{k-1}(t)$, we have $\lambda_k(\Sigma, X) \subseteq \lambda_{k-1}(\Sigma, X)$ for every $k \geq 3$, and hence $Loc_2 \subseteq Loc_3 \subseteq \cdots$.

If t is a string represented as a unary tree, then $\mathrm{st}_{k-1}(t)$ consists of the prefix of t of length $\leq k - 1$, $\mathrm{rt}_{k-1}(t)$ is the suffix of t of length $k - 1$, and $\mathrm{fork}_k(t)$ is the set of its substrings of length k. Hence, our unranked k-testable tree languages are natural counterparts of the k-testable string languages (cf. [11], for example).

To show that the families $Loc_k := \{Loc_k(\Sigma, X)\}_{\Sigma,X}$ $(k \geq 2)$ and $Loc := \{Loc(\Sigma, X)\}_{\Sigma,X}$ are varieties, we prove that the systems of congruences $\Lambda(k) := \{\lambda_k(\Sigma, X)\}_{\Sigma,X}$ $(k \geq 2)$ are consistent. For this we need the following fact. As the lemma itself is intuitively quite obvious, we omit the rather technical proof.

Lemma 9.8. *If $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ is a g-morphism and $k \geq 2$, then*

$$\mathrm{fork}_k(t\varphi) = \bigcup\{\mathrm{rt}_k(u\varphi) \mid u \in \mathrm{fork}_k(t)\} \cup \bigcup\{\mathrm{fork}_k(s\varphi) \mid s \in \mathrm{st}_{k-1}(t)\}$$

for every $t \in T_\Sigma(X)$.

Proposition 9.9. *For every $k \geq 2$, $\Lambda(k)$ is a CSC and Loc_k is the quasi-principal VUT defined by it. Hence, Loc is also a VUT.*

Proof. Fix a $k \geq 2$ and any g-morphism $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$, and let $s, t \in T_\Sigma(X)$ satisfy $s\, \lambda_k(\Sigma, X)\, t$. We should show that $s\varphi\, \lambda_k(\Omega, Y)\, t\varphi$.

We know that $\mathrm{st}_{k-1}(s\varphi) = \mathrm{st}_{k-1}(t\varphi)$ and $\mathrm{rt}_{k-1}(s\varphi) = \mathrm{rt}_{k-1}(t\varphi)$ follow from $\mathrm{st}_{k-1}(s) = \mathrm{st}_{k-1}(t)$ and $\mathrm{rt}_{k-1}(s) = \mathrm{rt}_{k-1}(t)$, respectively. Similarly, $\mathrm{fork}_k(s) = \mathrm{fork}_k(t)$ and $\mathrm{st}_{k-1}(s) = \mathrm{st}_{k-1}(t)$ imply $\mathrm{fork}_k(s\varphi) = \mathrm{fork}_k(t\varphi)$ by Lemma 9.8. Hence $s\varphi\, \lambda_k(\Omega, Y)\, t\varphi$. That Loc_k is the quasi-principal VUT defined by $\Lambda(k)$ is obvious, and Loc is a VUT as the union of the chain $Loc_2 \subseteq Loc_3 \subseteq \ldots$. ∎

9.6. Aperiodicity

To show that the natural unranked counterparts of the aperiodic tree languages [29] form a variety is as easy as in the ranked case [25].

For any $p, q \in C_\Sigma(X)$ and $t \in T_\Sigma(X)$, let $p \cdot q := q(p)$ and $t \cdot p := p(t)$. Obviously, $(C_\Sigma(X), \cdot, \xi)$ is a monoid and the powers p^n $(n \geq 0)$ of a ΣX-context p are defined as usual. For any $n \geq 0$, an unranked tree language $T \subseteq T_\Sigma(X)$ is called *n-aperiodic* if for all $q, r \in C_\Sigma(X)$ and $t \in T_\Sigma(X)$,

$$t \cdot q^{n+1} \cdot r \in T \;\Leftrightarrow\; t \cdot q^n \cdot r \in T,$$

and T is *aperiodic* if it is n-aperiodic for some $n \geq 0$. Let $Ap(\Sigma, X)$ be the set of all recognizable aperiodic ΣX-tree languages, and let $Ap := \{Ap(\Sigma, X)\}_{\Sigma,X}$.

Proposition 9.10. *Ap is a VUT.*

Proof. Conditions (V1)–(V3) are easy to verify. For (V4), let $T \in Ap(\Sigma, X)$ be n-aperiodic and consider any $p \in C_\Sigma(X)$. For all $q, r \in C_\Sigma(X)$ and $t \in T_\Sigma(X)$,

$$t \cdot q^{n+1} \cdot r \in p^{-1}(T) \;\Leftrightarrow\; t \cdot q^{n+1} \cdot (r \cdot p) \in T \;\Leftrightarrow t \cdot q^n \cdot (r \cdot p) \in T \;\Leftrightarrow\; t \cdot q^n \cdot r \in p^{-1}(T),$$

which shows that $p^{-1}(T) \in Ap(\Sigma, X)$.

Let $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ be a g-morphism and $T \in Ap(\Omega, Y)$ be n-aperiodic. Let $\hat{\varphi} : C_\Sigma(X) \to C_\Omega(Y)$ be the monoid morphism such that $\xi\hat{\varphi} := \xi$, and $f(t_1, \ldots, q, \ldots, t_m)\hat{\varphi} := \iota(f)(t_1\varphi, \ldots, q\hat{\varphi}, \ldots, t_m\varphi)$ for any $f \in \Sigma$, $m \geq 1$, $t_1, \ldots, t_m \in T_\Sigma(X)$ and $q \in C_\Sigma(X)$. It is clear that $(t \cdot p)\varphi = t\varphi \cdot p\hat{\varphi}$ for all $t \in T_\Sigma(X)$ and $p \in C_\Sigma(X)$. This implies that, for all $q, r \in C_\Sigma(X)$ and $t \in T_\Sigma(X)$,

$$t \cdot q^{n+1} \cdot r \in T\varphi^{-1} \Leftrightarrow t\varphi \cdot (q\hat{\varphi})^{n+1} \cdot r\hat{\varphi} \in T \Leftrightarrow t\varphi \cdot (q\hat{\varphi})^{n} \cdot r\hat{\varphi} \in T \Leftrightarrow t \cdot q^{n} \cdot r \in T\varphi^{-1},$$

which shows that Ap satisfies (V5), too. ■

9.7. Piecewise testability

Natural definitions of piecewise subtrees and piecewise testability of tree languages can be based on the well-known homeomorphic embedding order of trees (cf. [2]). In the ranked case this is done in [19], and a corresponding order underlies the definition of the piecewise testable hedge languages considered in [4]. For any Σ and X, the *homeomorphic embedding order* $\trianglelefteq$ on $T_\Sigma(X)$ is defined so that for any $s, t \in T_\Sigma(X)$, $s \trianglelefteq t$ if and only if

(1) $s = t$, or

(2) $s = f(s_1, \ldots, s_m)$ and $t = f(t_1, \ldots, t_m)$ where $s_1 \trianglelefteq t_1, \ldots, s_m \trianglelefteq t_m$, or

(3) $t = f(t_1, \ldots, t_m)$ and $s \trianglelefteq t_i$ for some $i \in [m]$.

For any $k \geq 0$ and $t \in T_\Sigma(X)$, let $P_k(t) := \{s \in T_\Sigma(X) \mid s \trianglelefteq t, \mathrm{hg}(s) < k\}$ and

$$\tau_k(\Sigma, X) := \{(s, t) \mid s, t \in T_\Sigma(X), P_k(s) = P_k(t)\}.$$

An unranked ΣX-tree language is *piecewise k-testable* if it is saturated by $\tau_k(\Sigma, X)$, and it is *piecewise testable* if it is piecewise k-testable for some $k \geq 0$. Let $Pwt_k(\Sigma, X)$ be the set of all recognizable piecewise k-testable unranked ΣX-tree languages. To prove that $Pwt_k := \{Pwt_k(\Sigma, X)\}_{\Sigma, X}$ ($k \geq 0$) and $Pwt := \bigcup_{k \geq 0} Pwt_k$ are VUTs, it suffices to show that $\mathrm{T}(k) := \{\tau_k(\Sigma, X)\}_{\Sigma, X}$ is a CSC for Pwt_k. It is easy to see that $\tau_k(\Sigma, X) \in \mathrm{Con}(\mathcal{T}_\Sigma(X))$ for every $k \geq 0$. For showing that the system $\mathrm{T}(k)$ is consistent, we need the following lemma.

Lemma 9.11. *Let $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ be a g-morphism. For any $k \geq 0$, $s \in T_\Sigma(X)$ and $t \in P_k(s\varphi)$, there exists an $s' \in P_k(s)$ such that $t \in P_k(s'\varphi)$.*

Proof. The proof goes by induction on k. The case $k = 0$ is trivial. If $t \in P_1(s\varphi)$, then $t \in \Omega \cup Y$, and now we proceed by induction on s. If $s \in \Sigma \cup X$, we may let s' be s. If $s = f(s_1, \ldots, s_m)$, then $s\varphi = \iota(f)(s_1\varphi, \ldots, s_m\varphi)$ and $t \in P_1(s_i\varphi)$ for some $i \in [m]$, and hence there is an $s' \in P_1(s_i) \subseteq P_1(s)$ such that $t \in P_1(s'\varphi)$.

Assume now that $k \geq 2$ and that the lemma holds for all smaller values of k. If $s \in \Sigma \cup X$, then $\mathrm{hg}(s) < k$, and we may set $s' := s$. Let $s = f(s_1, \ldots, s_m)$ and suppose that the claim holds for all smaller trees. Since $s\varphi = \iota(f)(s_1\varphi, \ldots, s_m\varphi)$, there are two possibilities. If $t \in P_k(s_i\varphi)$ for some $i \in [m]$, the required s' can be found as a piecewise subtree of s_i. Otherwise, $t = \iota(f)(t_1, \ldots, t_m)$ for some $t_1 \in P_{k-1}(s_1\varphi), \ldots, t_m \in P_{k-1}(s_m\varphi)$. By the main inductive assumption, there are trees $s'_1 \in P_{k-1}(s_1), \ldots, s'_m \in P_{k-1}(s_m)$ such that $t_1 \in P_{k-1}(s'_1\varphi), \ldots, t_m \in P_{k-1}(s'_m\varphi)$. Then $t \in P_k(s'\varphi)$ for $s' := f(s'_1, \ldots, s'_m) \in P_k(s)$. ■

Proposition 9.12. *For each $k \geq 0$, $\mathrm{T}(k)$ is a CSC and Pwt_k is the quasi-principal VUT defined by it. Hence, Pwt is also a VUT.*

Proof. That the system $\mathrm{T}(k)$ is consistent follows from Lemma 9.11. Indeed, if $(\iota, \varphi) : \mathcal{T}_\Sigma(X) \to \mathcal{T}_\Omega(Y)$ is a g-morphism and $s\,\tau_k(\Sigma, X)\,t$, then $s\varphi\,\tau_k(\Omega, Y)\,t\varphi$ because $P_k(s) = P_k(t)$ implies $P_k(s\varphi) = P_k(t\varphi)$ by that lemma.

By the definition of $\tau_k(\Sigma, X)$, a recognizable ΣX-tree language T is piecewise k-testable if and only if T is saturated by $\tau_k(\Sigma, X)$, and this is the case exactly when $\tau_k(\Sigma, X) \subseteq \theta_T$. This means that $\mathcal{C}^t_{\mathrm{T}(k)} = Pwt_k$. ■

10. Concluding remarks

We have defined and studied some basic algebraic notions for unranked algebras. In particular, we considered finite unranked algebras in which the operations are controlled by regular languages, and defined the recognizability of unranked tree languages in terms of them. We introduced varieties of unranked tree languages and established a bijection between them and varieties of regular algebras using syntactic algebras. We have demonstrated that the natural unranked counterparts of several known varieties of ranked tree languages form varieties in our sense. Thus it seems that we achieved a good general framework for an algebraic classification theory of unranked tree languages. We also presented a general scheme by which many varieties were obtained from certain systems of congruences of term algebras. Of course, much remains to be done. For example, the varieties of regular algebras corresponding to our example varieties of unranked tree languages should be identified and studied. Concerning the formal language aspect, one could try to characterize some of the VUTs by suitable tree logics.

Acknowledgement

We thank the careful referee for many helpful suggestions.

References

[1] J. Almeida, *On pseudovarieties, varieties of languages, filters of congruences, pseudo-identities and related topics*, Algebra Universalis **27** (1990) 333–350. doi:10.1007/BF01190713

[2] F. Baader and T. Nipkow, Term Rewriting and All That (Cambridge University Press, Cambridge, 1998).
doi:10.1017/CBO9781139172752

[3] C. Bergman, Universal Algebra. Fundamentals and Selected Topics (CRC Press, Taylor & Francis Group, Boca Raton, Fl, 2012).

[4] M. Bojańczyk, L. Segoufin and H. Straubing, *Piecewise testable tree languages*, Log. Methods in Comput. Sci. **8** (2012) 1–32.
doi:10.2168/LMCS-8(3:26)2012

[5] A. Brüggemann-Klein, M. Murata and D. Wood, Regular tree and hedge languages over unranked alphabets, Version 1, HKUST Theoretical Computer Science Center Research Report HKUST-TCS-2001-05 (Hong Kong, 2001).

[6] S. Burris and H.P. Sankappanavar, A Course in Universal Algebra (Springer-Verlag, New York, 1981).

[7] P.M. Cohn, Universal Algebra, 2. revised ed. (D. Reidel Publishing Company, Dordrecht, 1981).

[8] H. Comon, M. Dauchet, R. Gilleron, C. Löding, F. Jacquemard, D. Lugiez, S. Tison and M. Tommasi, Tree Automata Techniques and Applications, available online: http://www.grappa.univ-lille3.fr/tata, 2007.

[9] J. Cristau, C. Löding and W. Thomas, *Deterministic automata on unranked trees*, in: M. Liśkiewicz, R. Reischuk (Eds.), Fundamentals of Computation Theory, Proc. 15th Internat. Symp. FCT 2005, Lübeck 2005, Lect. Notes Comput. Sci. 3623 (Springer, Berlin, 2005) pp. 68–79.
doi:10.1007/11537311_7

[10] K. Denecke and S.L. Wismath, Universal Algebra and Applications in Theoretical Computer Science (Chapman & Hall/CRC, Boca Raton, Fl, 2002).

[11] S. Eilenberg, Automata, Languages, and Machines (Academic Press, New York, Vol. A 1974, Vol. B 1976).

[12] F. Gécseg and M. Steinby, Tree Automata, Akadémiai Kiadó, Budapest 1984. 2. ed. downloadable from arXiv.org as arXiv:1509.06233, September 2015.

[13] F. Gécseg and M. Steinby, Tree languages, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Vol. 3 (Springer, Berlin, 1997) pp. 1–69.

[14] E. Jurvanen, A. Potthoff and W. Thomas, *Tree languages recognizable by regular frontier check*, in: G. Rozenberg, A. Salomaa (Eds.), Developments in Language Theory, World Scientific (Singapore, 1994) pp. 3–17.

[15] J. Koppitz and K. Denecke, M-Solid Varieties of Algebras (Springer-Verlag, New York, 2006).
doi:10.1007/0-387-30806-7

[16] W. Martens and J. Niehren, *On the mininization of XML Schemas and tree automata for unranked trees*, J. Comput. Syst. Sci. **73** (2007) 550–583.
doi:10.1016/j.jcss.2006.10.021

[17] F. Neven, Automata, Logic, and XML, in: J. Bradfield (Ed.), Computer Science Logic, Proc. 16th Internat. Workshop, CSL 2002, Edinburgh, UK, 2002, Lect. Notes Comput. Sci. 2471 (Springer, Berlin, 2002) pp. 2–26.

[18] C. Pair and A. Quere, *Définition et etude bilangages réguliers*, Inform. Control **13** (1968) 565–593.
doi:10.1016/S0019-9958(68)90999-6

[19] V. Piirainen, Piecewise testable tree languages, TUCS Technical Report 634 (Turku Centre for Computer Science, Turku, 2004).

[20] J.E. Pin, Varieties of Formal Languages (North Oxford Academic Publishers, London, 1986).

[21] J. Sakarovitch, Elements of Automata Theory (Cambridge University Press, Cambridge, 2009).

[22] T. Schwentick, *Automata for XML – A survey*, J. Comput. Syst. Sci. **73** (2007) 289–315.
doi:10.1016/j.jcss.2006.10.003

[23] M. Steinby, *Syntactic algebras and varieties of recognizable sets*, in: M.C. Gaudel, J.P. Jouannaud (Eds.), Les Arbres en Algèbre et en Programmation, Proc. 4th CAAP (University of Lille, Lille, 1979) pp. 226–240.

[24] M. Steinby, *A theory of tree language varieties*, in: M. Nivat, A. Podelski, (Eds.), Tree Automata and Languages (North-Holland, Amsterdam, 1992) pp. 57–81.

[25] M. Steinby, *General varieties of tree languages*, Theor. Comput. Sci. **205** (1998) 1–43. doi:10.1016/S0304-3975(98)00010-3

[26] M. Steinby, *Algebraic classifications of regular tree languages*, in: V.B. Kudryavtsev, I.G. Rosenberg (Eds.), Structural Theory of Automata, Semigroups, and Universal Algebra (Springer, Dordrecht, 2005) pp. 381–432.

[27] M. Takahashi, *Generalizations of regular sets and their application to a study of context-free languages*, Inform. Control **27** (1975) 1–36.
doi:10.1016/S0019-9958(75)90058-3

[28] J.W. Thatcher, *Characterizing derivation trees of context-free grammars through a generalization of finite automata theory*, J. Comput. Syst. Sci. **1** (1967) 317–322.
doi:10.1016/S0022-0000(67)80022-9

[29] W. Thomas, *Logical aspects in the study of tree languages*, in: B. Courcelle (Ed.), 9th Colloquium on Trees in Algebra and Programming, Proc. Conf., Bordeaux, 1984 (Cambridge University Press, London, 1984) pp. 31–49.

Received 4 July 2016
I Revised 20 October 2016
II Revised 5 November 2016