

# Varieties of Data Languages

**Henning Urbat**

Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany  
henning.urbat@fau.de

**Stefan Milius**

Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany  
mail@stefan-milius.eu

---

## Abstract

We establish an Eilenberg-type correspondence for data languages, i.e. languages over an infinite alphabet. More precisely, we prove that there is a bijective correspondence between varieties of languages recognized by orbit-finite nominal monoids and pseudovarieties of such monoids. This is the first result of this kind for data languages. Our approach makes use of nominal Stone duality and a recent category theoretic generalization of Birkhoff-type theorems that we instantiate here for the category of nominal sets. In addition, we prove an axiomatic characterization of weak pseudovarieties as those classes of orbit-finite monoids that can be specified by sequences of nominal equations, which provides a nominal version of a classical theorem of Eilenberg and Schützenberger.

**2012 ACM Subject Classification** Theory of computation → Formal languages and automata theory

**Keywords and phrases** Nominal sets, Stone duality, Algebraic language theory, Data languages

**Digital Object Identifier** 10.4230/LIPIcs.ICALP.2019.130

**Category** Track B: Automata, Logic, Semantics, and Theory of Programming

**Related Version** <https://arxiv.org/abs/1903.08053>

**Funding** Henning Urbat acknowledges support by Deutsche Forschungsgemeinschaft (DFG) under project SCHR 1118/8-2. Stefan Milius acknowledges support by Deutsche Forschungsgemeinschaft (DFG) under project MI 717/5-1.

## 1 Introduction

In the algebraic theory of formal languages, one studies automata and the languages they represent in terms of associated algebraic structures. This approach has been successfully implemented for numerous types of languages and has proven extremely fruitful because it allows to import powerful algebraic methods into the realm of automata theory. As a prime example, regular languages can be described purely algebraically as the languages recognized by finite monoids, and a celebrated result by McNaughton, Papert, and Schützenberger [12, 19] asserts that a regular language is definable in first-order logic if and only if its syntactic monoid is aperiodic (i.e. it satisfies the equation  $x^{n+1} = x^n$  for sufficiently large  $n$ ). As an immediate application, this algebraic characterization yields an effective procedure for deciding first-order definability. The first systematic approach to correspondence results of this kind was initiated by Eilenberg [6] who proved that *varieties of languages* (i.e. classes of regular languages closed under the set-theoretic boolean operations, derivatives, and homomorphic preimages) correspond bijectively to *pseudovarieties of monoids* (i.e. classes of finite monoids closed under quotients monoids, submonoids, and finite products). Eilenberg’s result thus establishes a generic relation between properties of regular languages and properties of finite monoids. In addition, Eilenberg and Schützenberger [7] contributed a model-theoretic description



© Henning Urbat and Stefan Milius;

licensed under Creative Commons License CC-BY

46th International Colloquium on Automata, Languages, and Programming (ICALP 2019).

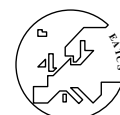
Editors: Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi;

Article No. 130; pp. 130:1–130:14



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



of pseudovarieties: they are those classes of finite monoids that can be axiomatized by a sequence  $(s_n = t_n)_{n \in \mathbb{N}}$  of equations, interpreted as “ $s_n = t_n$  holds for sufficiently large  $n$ ”. For instance, the pseudovariety of aperiodic finite monoids is axiomatized by  $(x^{n+1} = x^n)_{n \in \mathbb{N}}$ .

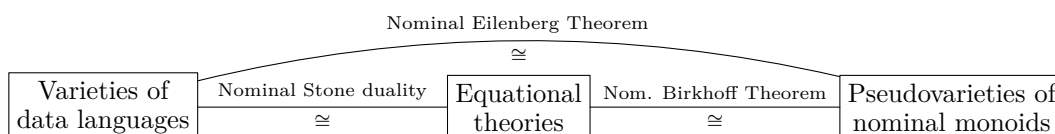
The goal of our present paper is to study *data languages*, i.e. languages over an infinite alphabet, from the perspective of algebraic language theory. Such languages have spurred significant interest in recent years, driven by practical applications in various areas of computer science, including efficient parsing of XML documents or software verification. Mathematically, data languages are modeled using *nominal sets*. Over the years, several machine models for handling data languages of different expressive power have been proposed; see [21, 20] for a comprehensive survey. The focus of this paper is on languages recognized by *orbit-finite nominal monoids*. They form an important subclass of the languages accepted by Francez and Kaminski’s *finite memory automata* [10] (which are expressively equivalent to orbit-finite automata in the category of nominal sets [5]) and have been characterized by a fragment of monadic second-order logic over data words called *rigidly guarded MSO* [17]. In addition, Bojańczyk [4] and Colcombet, Ley, and Puppis [17] established nominal versions of the McNaughton-Papert-Schützenberger theorem and showed that the first-order definable data languages are precisely the ones recognizable by aperiodic orbit-finite monoids.

In the light of these results, it is natural to ask whether a generic variety theory akin to Eilenberg’s seminal work can be developed for data languages. As the main contribution of our paper, we answer this positively by establishing nominal generalizations of two key results known from the algebraic theory of regular languages. The first one is a counterpart of Eilenberg’s variety theorem, which is the first result of this kind for data languages:

► **Nominal Eilenberg Theorem.** *Varieties of data languages correspond bijectively to pseudovarieties of nominal monoids.*

Here, the notion of a *pseudovariety of nominal monoids* is as expected: a class of orbit-finite nominal monoids closed under quotient monoids, submonoids, and finite products. In contrast, the notion of a *variety of data languages* requires two extra conditions unfamiliar from other Eilenberg-type correspondences, most notably a technical condition called *completeness* (Definition 4.13). Like the original Eilenberg theorem, its nominal version gives rise to a generic relation between properties of data languages and properties of nominal monoids. For instance, the aperiodic orbit-finite monoids form a pseudovariety, and the first-order definable data languages form a variety, and thus the equivalence of these concepts can be understood as an instance of the nominal Eilenberg correspondence.

On a conceptual level, our results crucially make use of *duality*, specifically an extension of Petrişan’s [15] nominal version of Stone duality which gives a dual equivalence between nominal sets and nominal complete atomic boolean algebras. To derive the nominal Eilenberg correspondence, we make two key observations. First, we show that varieties of data languages dualize (under nominal Stone duality) to the concept of an *equational theory* in the category of nominal sets. Second, we apply a recent categorical generalization of Birkhoff-type variety theorems [14] to show that equational theories correspond to pseudovarieties of nominal monoids. Our approach is summarized by the diagram below:



The idea that Stone-type dualities play a major role in algebraic language theory was firmly established by Gehrke, Grigorieff, and Pin [9]. It is also at the heart of our recent line of work [1, 2, 22, 3], which culminated in a uniform category theoretic proof of more than a dozen Eilenberg correspondences for various types of languages. A related, yet more abstract, approach was pursued by Salamanca [18]. The key insight of [22, 18] is that Eilenberg-type correspondences arise by combining a Birkhoff-type correspondence with a Stone-type duality. Our present approach to data languages is an implementation of this principle in the nominal setting. Since the existing categorical frameworks for algebraic language theory consider algebraic-like base categories (which excludes nominal sets) and the recognition of languages by finite structures, our Nominal Eilenberg Theorem is not covered by any previous categorical work and requires new techniques. However, our approach can be seen as an indication of the robustness of the duality-based methodology for algebraic recognition.

As our second main contribution, we complement the Nominal Eilenberg Theorem with a model-theoretic description of pseudovarieties of nominal monoids in terms of sequences of *nominal equations*, generalizing the classical result of Eilenberg and Schützenberger for ordinary monoids. Our result applies more generally to the class of *weak pseudovarieties* of nominal monoids, which are only required to be closed under support-reflecting (rather than arbitrary) quotients. We then obtain the

► **Nominal Eilenberg-Schützenberger Theorem.** *Weak pseudovarieties are exactly the classes of nominal monoids axiomatizable by sequences of nominal equations.*

While our main results apply to languages recognizable by orbit-finite monoids, the underlying methods are of fairly general nature and can be extended to other recognizing structures in the category of nominal sets. We illustrate this in section 5 by deriving a (local) Eilenberg correspondence for languages accepted by deterministic nominal automata.

Full proofs of all results can be found in the extended arXiv version [23] of our paper.

## 2 Nominal Sets

We start by recalling basic definitions and facts from the theory of nominal sets [16]. Some of the concepts considered in this paper are most clearly and conveniently formulated in the language of category theory, but only very basic knowledge of category theory is required from the reader. Fix a countably infinite set  $\mathbb{A}$  of *atoms*, and denote by  $\text{Perm}(\mathbb{A})$  the group of finite permutations of  $\mathbb{A}$  (i.e. bijections  $\pi: \mathbb{A} \rightarrow \mathbb{A}$  that move only finitely many elements of  $\mathbb{A}$ ). A  $\text{Perm}(\mathbb{A})$ -set is a set  $X$  with an operation  $\text{Perm}(\mathbb{A}) \times X \rightarrow X$ , denoted as  $(\pi, x) \mapsto \pi \cdot x$ , such that  $(\sigma\pi) \cdot x = \sigma \cdot (\pi \cdot x)$  and  $id \cdot x = x$  for all  $\sigma, \pi \in \text{Perm}(\mathbb{A})$  and  $x \in X$ . If the group action is trivial, i.e.  $\pi \cdot x = x$  for all  $\pi \in \text{Perm}(\mathbb{A})$  and  $x \in X$ , we call  $X$  *discrete*. For any set  $S \subseteq \mathbb{A}$  of atoms, denote by  $\text{Perm}_S(\mathbb{A}) \subseteq \text{Perm}(\mathbb{A})$  the subgroup of all finite permutations  $\pi$  that fix  $S$ , i.e.  $\pi(a) = a$  for all  $a \in S$ . The set  $S$  is called a *support* of an element  $x \in X$  if for every  $\pi \in \text{Perm}_S(\mathbb{A})$  one has  $\pi \cdot x = x$ . The intuition is that  $x$  is some kind of syntactic object (e.g. a string, a tree, a term, or a program) whose free variables are contained in  $S$ . Thus, a variable renaming  $\pi$  that leaves  $S$  fixed does not affect  $x$ . A *nominal set* is a  $\text{Perm}(\mathbb{A})$ -set  $X$  such that every element of  $X$  has a finite support. This implies that every element  $x \in X$  has a least support, denoted by  $\text{supp}_X(x) \subseteq \mathbb{A}$ . A nominal set  $X$  is *strong* if, for every  $x \in X$  and  $\pi \in \text{Perm}(\mathbb{A})$ , one has  $\pi \cdot x = x$  if and only if  $\pi(a) = a$  for all  $a \in \text{supp}_X(x)$ . The *orbit* of an element  $x$  of a nominal set  $X$  is the set  $\{\pi \cdot x : \pi \in \text{Perm}(\mathbb{A})\}$ . The orbits form a partition of  $X$ . If  $X$  has only finitely many orbits, then  $X$  is called *orbit-finite*. More generally, for any finite set  $S \subseteq \mathbb{A}$  of atoms, the *S-orbit* of an element  $x \in X$  is the set  $\{\pi \cdot x : \pi \in \text{Perm}_S(\mathbb{A})\}$ , and the *S-orbits* form a partition of  $X$ .

► **Lemma 2.1.** *Let  $S$  be a finite subset of  $\mathbb{A}$ . Then every orbit-finite nominal set has only finitely many  $S$ -orbits.*

A map  $f: X \rightarrow Y$  between nominal sets is *equivariant* if  $f(\pi \cdot x) = \pi \cdot f(x)$  for all  $\pi \in \text{Perm}(\mathbb{A})$  and  $x \in X$ , and *finitely supported* if there exists a finite set  $S \subseteq \mathbb{A}$  such that  $f(\pi \cdot x) = \pi \cdot f(x)$  for all  $\pi \in \text{Perm}_S(\mathbb{A})$  and  $x \in X$ . Equivariant maps do not increase supports, i.e. one has  $\text{supp}_Y(f(x)) \subseteq \text{supp}_X(x)$  for all  $x \in X$ . We write  $\mathbf{Nom}_{\text{fs}}$  for the category of nominal sets and finitely supported maps, and  $\mathbf{Nom}$  for the (non-full) subcategory of nominal sets and equivariant maps. We shall use the following standard results about  $\mathbf{Nom}$ :

- (1)  $\mathbf{Nom}$  is complete and cocomplete. Finite limits and all colimits are formed on the level of underlying sets. In particular, finite products of nominal sets are given by cartesian products and coproducts by disjoint union.
- (2) For every pair  $X, Y$  of nominal sets, the *exponential*  $[X, Y]$  is the nominal set consisting of all finitely supported maps  $f: X \rightarrow Y$ , with the group action given by  $(\pi \cdot f)(x) = \pi \cdot f(\pi^{-1} \cdot x)$ . Moreover, for every nominal set  $X$ , the *nominal power set*  $\mathcal{P}X$  is carried by the set of all subsets  $X_0 \subseteq X$  with finite support; i.e. for which there exists a finite set  $S \subseteq \mathbb{A}$  of atoms such that  $\pi \cdot X_0 = X_0$  for  $\pi \in \text{Perm}_S(\mathbb{A})$ , where  $\pi \cdot X_0 = \{\pi \cdot x : x \in X_0\}$ . In particular, every singleton  $\{x\}$  is finitely supported by  $\text{supp}_X(x)$ . The group action on  $\mathcal{P}X$  is given by  $X_0 \mapsto \pi \cdot X_0$ , and we have  $\mathcal{P}X \cong [X, 2]$ , for the discrete nominal set  $2 = \{0, 1\}$ .
- (3) *Quotients* and *subobjects* in  $\mathbf{Nom}$  are represented by epimorphisms (= surjective equivariant maps) and monomorphisms (= injective equivariant maps), respectively.  $\mathbf{Nom}$  has image factorizations, i.e. every equivariant map  $f: X \rightarrow Y$  has a unique decomposition  $f = m \cdot e$  into a quotient  $e: X \twoheadrightarrow I$  followed by a subobject  $m: I \rightarrow Y$ . We call  $e$  the *coimage* of  $f$ .
- (4) Orbit-finite nominal sets are closed under quotients, subobjects, and finite products.
- (5) For each  $n \geq 0$ , the nominal set  $\mathbb{A}^{\#n} = \{(a_1, \dots, a_n) \in \mathbb{A}^n : a_i \neq a_j \text{ for } i \neq j\}$  with group action  $\pi \cdot (a_1, \dots, a_n) = (\pi(a_1), \dots, \pi(a_n))$  is strong and has a single orbit. More generally, the (orbit-finite) strong nominal sets are up to isomorphism exactly the (finite) coproducts of nominal sets of the form  $\mathbb{A}^{\#n}$ .

### 3 Pseudovarieties of Nominal Monoids

In this section, we investigate classes of orbit-finite nominal monoids and establish two characterizations of such classes: a categorical one, relating pseudovarieties of nominal monoids to equational theories in the category of nominal sets, and an axiomatic one, describing weak pseudovarieties in terms of sequences of nominal equations. The first of these results is the algebraic foundation of our subsequent treatment of varieties of data languages.

A *nominal monoid* is a monoid  $(M, \bullet, 1_M)$  in the category  $\mathbf{Nom}$ ; that is,  $M$  is equipped with the structure of a nominal set such that the multiplication  $\bullet: M \times M \rightarrow M$  is an equivariant map and the unit  $1_M \in M$  has empty support, i.e. it corresponds to an equivariant map  $1 \rightarrow M$ , where  $1$  is the nominal set with one element. We write  $\mathbf{nMon}$  for the category of nominal monoids and equivariant monoid morphisms (usually just called *morphisms*), and  $\mathbf{nMon}_{\text{of}}$  for the full subcategory of orbit-finite nominal monoids. The forgetful functor from  $\mathbf{nMon}$  to  $\mathbf{Nom}$  has a left adjoint assigning to each nominal set  $\Sigma$  the free nominal monoid  $\Sigma^*$  of all words over  $\Sigma$ , with monoid multiplication given by concatenation of words, unit  $\varepsilon$  (the empty word) and group action  $\pi \cdot (a_1 \cdots a_n) = \pi(a_1) \cdots \pi(a_n)$  for  $\pi \in \text{Perm}(\mathbb{A})$  and  $a_1 \cdots a_n \in \Sigma^*$ . The category  $\mathbf{nMon}$  has products (formed on the level of  $\mathbf{Nom}$ ), image factorizations, and surjective and injective morphisms represent *quotients* and *submonoids* of

nominal monoids. A quotient  $q: M \twoheadrightarrow M'$  is called *support-reflecting* if for every  $x' \in M'$  there exists an  $x \in M$  with  $q(x) = x'$  and  $\text{supp}_M(x) = \text{supp}_{M'}(x')$ . The following result characterizes the quotient monoids of  $\Sigma^*$  in terms of unary operations:

► **Proposition 3.1** (Unary presentation for nominal monoids). *For every nominal set  $\Sigma$  and every surjective equivariant map  $e: \Sigma^* \twoheadrightarrow M$ , the following statements are equivalent:*

- (1)  *$e$  carries a quotient monoid of  $\Sigma^*$ , i.e. there exists a nominal monoid structure  $(M, \bullet, 1_M)$  on  $M$  such that  $e: \Sigma^* \twoheadrightarrow (M, \bullet, 1_M)$  is a morphism of nominal monoids;*
- (2) *the maps  $\Sigma^* \xrightarrow{w \cdot -} \Sigma^*$  and  $\Sigma^* \xrightarrow{- \cdot w} \Sigma^*$  ( $w \in \Sigma^*$ ) lift along  $e$ , i.e. there exist (necessarily unique) maps  $l_w$  and  $r_w$  making the following squares commute:*

$$\begin{array}{ccc} \Sigma^* & \xrightarrow{w \cdot -} & \Sigma^* \\ e \downarrow & & \downarrow e \\ M & \xrightarrow{l_w} & M \end{array} \quad \begin{array}{ccc} \Sigma^* & \xrightarrow{- \cdot w} & \Sigma^* \\ e \downarrow & & \downarrow e \\ M & \xrightarrow{r_w} & M \end{array} \quad \text{for every } w \in \Sigma^*.$$

In general, the maps  $w \cdot -$  and  $- \cdot w$  are not equivariant, but finitely supported (with support contained in the one of  $w$ ). This implies that also  $l_w$  and  $r_w$  in 2 are finitely supported.

### 3.1 Equational Theories

In previous work [14] we studied varieties of objects in a general category and their relation to an abstract form of equations. In the following, we instantiate these concepts to the category of nominal sets to derive a characterization of pseudovarieties of orbit-finite monoids.

► **Definition 3.2.** Let  $\Sigma$  be a nominal set. A  $\Sigma$ -generated nominal monoid is a nominal quotient monoid  $e: \Sigma^* \twoheadrightarrow M$  of the free monoid  $\Sigma^*$ . We denote by  $\Sigma^* \downarrow \mathbf{nMon}_{\text{of}}$  the poset of  $\Sigma$ -generated orbit-finite nominal monoids, ordered by  $e \leq e'$  iff  $e'$  factorizes through  $e$ .

► **Definition 3.3.** A *local pseudovariety of  $\Sigma$ -generated nominal monoids* is a filter  $\mathcal{T}_\Sigma \subseteq \Sigma^* \downarrow \mathbf{nMon}_{\text{of}}$  in the poset of  $\Sigma$ -generated orbit-finite nominal monoids; that is,  $\mathcal{T}_\Sigma$  is

- (1) *upwards closed:*  $e \in \mathcal{T}_\Sigma$  and  $e \leq e'$  implies  $e' \in \mathcal{T}_\Sigma$ , and
- (2) *downwards directed:* for each pair  $e_0, e_1 \in \mathcal{T}_\Sigma$  there exists  $e \in \mathcal{T}_\Sigma$  with  $e \leq e_0, e_1$ .

If we replace (1) by the weaker condition

- (1') for each  $e: \Sigma^* \twoheadrightarrow M$  in  $\mathcal{T}_\Sigma$  and each support-reflecting  $q: M \twoheadrightarrow N$  one has  $q \cdot e \in \mathcal{T}_\Sigma$ , then  $\mathcal{T}_\Sigma$  is called a *weak local pseudovariety of  $\Sigma$ -generated nominal monoids*.

► **Remark 3.4.** By Proposition 3.1, the definition of local pseudovariety can be equivalently stated as follows:

- (1)  $\mathcal{T}_\Sigma$  is a filter in the poset of orbit-finite quotients of  $\Sigma^*$  in  $\mathbf{Nom}$ ;
- (2) for every  $e \in \mathcal{T}_\Sigma$  and  $w \in \Sigma^*$ , the unary operations  $w \cdot -$  and  $- \cdot w$  on  $\Sigma^*$  lift along  $e$ .

Let  $\mathbf{Nom}_{\text{of},s}$  denote the full subcategory of  $\mathbf{Nom}$  on orbit-finite strong nominal sets.

► **Definition 3.5** (Equational Theory). A *(weak) equational theory* is a family

$$\mathcal{T} = (\mathcal{T}_\Sigma \subseteq \Sigma^* \downarrow \mathbf{nMon}_{\text{of}})_{\Sigma \in \mathbf{Nom}_{\text{of},s}}$$

of (weak) local pseudovarieties with the following two properties (see the diagrams below):

- (1) *Substitution invariance.* For each equivariant monoid morphism  $h: \Delta^* \rightarrow \Sigma^*$  with  $\Delta, \Sigma \in \mathbf{Nom}_{\text{of},s}$  and each  $e_\Sigma: \Sigma^* \twoheadrightarrow M_\Sigma$  in  $\mathcal{T}_\Sigma$ , the coimage  $e_\Delta$  of  $e_\Sigma \cdot h$  lies in  $\mathcal{T}_\Delta$ .
- (2) *Completeness.* For each  $\Sigma \in \mathbf{Nom}_{\text{of},s}$  and each quotient  $e: \Sigma^* \twoheadrightarrow M_\Sigma$  in  $\mathcal{T}_\Sigma$ , there exists  $\Delta \in \mathbf{Nom}_{\text{of},s}$  and a support-reflecting quotient  $e_\Delta: \Delta^* \twoheadrightarrow M_\Delta$  in  $\mathcal{T}_\Delta$  with  $M_\Delta = M_\Sigma$ .

$$\begin{array}{ccc}
 \Delta^* & \xrightarrow{\forall h} & \Sigma^* \\
 e_\Delta \downarrow & & \downarrow \forall e_\Sigma \\
 M_\Delta & \xrightarrow{\quad} & M_\Sigma
 \end{array}
 \qquad
 \begin{array}{ccc}
 \Delta^* & & \Sigma^* \\
 \exists e_\Delta \downarrow & & \downarrow \forall e_\Sigma \\
 M_\Delta & \equiv & M_\Sigma
 \end{array}$$

► **Remark 3.6.**

- (1) Local pseudovarieties were previously called *equations* [14]. In fact, in many instances of the framework in *op. cit.*, a filter of quotients can be represented as a single quotient of a free algebra on an object  $\Sigma$ , which in turn corresponds to a set of pairs of terms given by the kernel of the quotient, i.e. to the usual syntactic concept of an equation.
- (2) The restriction to *strong* nominal sets  $\Sigma$  as generators reflects that the latter are the “free” nominal sets [11], a property crucial for the proof of Theorem 3.8 below. More precisely, letting  $\mathcal{P}_f \mathbb{A}$  denote the set of finite subsets of  $\mathbb{A}$ , the forgetful functor  $U: \mathbf{Nom} \rightarrow \mathbf{Set}^{\mathcal{P}_f \mathbb{A}}$  mapping a nominal set  $X$  to the presheaf  $S \mapsto \{x \in X : \text{supp}_X(x) \subseteq S\}$  has a left adjoint  $F$ , and strong nominal sets are exactly the nominal sets of the form  $FP$  for  $P \in \mathbf{Set}^{\mathcal{P}_f \mathbb{A}}$ .
- (3) The somewhat technical completeness property cannot be avoided, i.e. a substitution-invariant family of local pseudovarieties is generally incomplete. Indeed, consider the family

$$\mathcal{T} = (\mathcal{T}_\Sigma \subseteq \Sigma^* \downarrow \mathbf{nMon}_{\text{of},s})_{\Sigma \in \mathbf{Nom}_{\text{of},s}},$$

where  $\mathcal{T}_\Sigma$  consists of all  $\Sigma$ -generated orbit-finite nominal monoids  $e: \Sigma^* \twoheadrightarrow M$  such that  $e$  maps each element of  $\Sigma^*$  with a support of size 1 to  $1_M$ .

To see that  $\mathcal{T}_\Sigma$  is a filter, suppose that  $e: \Sigma^* \twoheadrightarrow M$  and  $e': \Sigma^* \twoheadrightarrow M'$  are two quotients in  $\mathcal{T}_\Sigma$ . Form their subdirect product  $q$ , viz. the coimage of the morphism  $\langle e, e' \rangle: \Sigma^* \rightarrow M \times M'$ . Each  $w \in \Sigma^*$  with a support of size 1 is mapped by  $q$  to  $(e(w), e'(w)) = (1_M, 1_{M'}) = 1_{M \times M'}$ . Thus  $q \in \mathcal{T}_\Sigma$  and  $q \leq e, e'$ , i.e.  $\mathcal{T}_\Sigma$  is downwards directed. Clearly,  $\mathcal{T}_\Sigma$  is also upwards closed.

For substitution invariance, let  $h: \Delta^* \rightarrow \Sigma^*$  be a morphism and  $e_\Sigma: \Sigma^* \twoheadrightarrow M_\Sigma$  a quotient in  $\mathcal{T}_\Sigma$ . Then  $e_\Sigma \cdot h$  maps each element with a support of size 1 to  $1_{M_\Sigma}$  since  $e_\Sigma$  does and the equivariant map  $h$  does not increase supports. Thus, the coimage of  $e_\Sigma \cdot h$  lies in  $\mathcal{T}_\Delta$ .

Finally, we show that  $\mathcal{T}$  is not complete. Fix an arbitrary orbit-finite nominal monoid  $M$  containing an element  $m$  with  $|\text{supp}_M m| = 1$ . Note that  $m \neq 1_M$  because  $1_M$  has empty support. Moreover, choose an orbit-finite strong nominal set  $\Sigma$  such that all elements of  $\Sigma$  have least support of size at least 2, and  $M$  can be expressed as a quotient  $e: \Sigma^* \twoheadrightarrow M$ . (For instance, one may take  $\Sigma = \coprod_{i < k} \mathbb{A}^{\#n}$  where  $k$  is the number of orbits of  $M$  and  $n = \max\{2, |\text{supp}_M(x)| : x \in M\}$ .) Since all nonempty words in  $\Sigma^*$  have a least support of size at least 2, one has  $e \in \mathcal{T}_\Sigma$ . For every  $\Delta \in \mathbf{Nom}_{\text{of},s}$  and every quotient  $q: \Delta^* \twoheadrightarrow M$  in  $\mathcal{T}_\Delta$ , the set  $q^{-1}\{m\} \subseteq \Delta^*$  contains no element with least support of size 1, since such elements are mapped by  $q$  to  $1_M \neq m$ . Consequently,  $q$  is not support-reflecting. This shows that  $M$  is not the codomain of any support-reflecting quotient in  $\mathcal{T}_\Delta$ .

► **Definition 3.7** (Pseudovariety and Weak Pseudovariety). A *pseudovariety of nominal monoids* is a nonempty class  $\mathcal{V}$  of orbit-finite nominal monoids closed under finite products, submonoids, and quotient monoids. That is,

- (1) for each  $M, N \in \mathcal{V}$  one has  $M \times N \in \mathcal{V}$ ;
- (2) for each  $M \in \mathcal{V}$  and each nominal submonoid  $N \twoheadrightarrow M$  one has  $N \in \mathcal{V}$ ;
- (3) for each  $M \in \mathcal{V}$  and each nominal quotient monoid  $M \twoheadrightarrow N$  one has  $N \in \mathcal{V}$ .

A *weak pseudovariety of nominal monoids* is a nonempty class of orbit-finite nominal monoids closed under finite products, submonoids, and support-reflecting quotient monoids.

The following result is a special case of the Generalized Variety Theorem [14, Theorem 3.15]. It asserts that equational theories and pseudovarieties are equivalent concepts. Note that (weak) equational theories form a poset ordered by  $\mathcal{T} \leq \mathcal{T}'$  iff  $\mathcal{T}_\Sigma \leq \mathcal{T}'_\Sigma$  for all  $\Sigma \in \mathbf{Nom}_{\text{of},s}$ , where  $\mathcal{T}_\Sigma \leq \mathcal{T}'_\Sigma$  holds iff for every  $e' \in \mathcal{T}'_\Sigma$  there exists an  $e \in \mathcal{T}_\Sigma$  with  $e \leq e'$ . Similarly, (weak) pseudovarieties of nominal monoids form a poset w.r.t. the inclusion ordering.

► **Theorem 3.8.** *(Weak) equational theories and (weak) pseudovarieties of nominal monoids form dually isomorphic complete lattices.*

The isomorphism maps a (weak) equational theory  $\mathcal{T}$  to the (weak) pseudovariety  $\mathcal{V}(\mathcal{T})$  of all orbit-finite monoids  $M$  such that each morphism  $h: \Sigma^* \rightarrow M$  with  $\Sigma \in \mathbf{Nom}_{\text{of},s}$  factorizes through some  $e_\Sigma \in \mathcal{T}_\Sigma$ . The inverse maps a (weak) pseudovariety  $\mathcal{V}$  to the (weak) equational theory  $\mathcal{T}(\mathcal{V})$  where  $\mathcal{T}(\mathcal{V})_\Sigma$  consists of all quotients  $e: \Sigma^* \twoheadrightarrow M$  with codomain  $M \in \mathcal{V}$ .

### 3.2 The Nominal Eilenberg-Schützenberger Theorem

In addition to their abstract category theoretic characterization in Theorem 3.8, weak pseudovarieties of nominal monoids admit an axiomatic description in terms of sequences of equations, analogous to the classical result of Eilenberg and Schützenberger [7] for pseudovarieties of ordinary monoids. The appropriate concept of equation is as follows:

► **Definition 3.9.**

- (1) An *equation* is a pair  $(s, t) \in X^* \times X^*$ , denoted as  $s = t$ , where  $X$  is an orbit-finite strong nominal set. A nominal monoid  $M$  *satisfies*  $s = t$  if for every equivariant map  $h: X \rightarrow M$  one has  $\widehat{h}(s) = \widehat{h}(t)$ , where  $\widehat{h}: X^* \rightarrow M$  denotes the unique extension of  $h$  to an equivariant monoid morphism.
- (2) Given a sequence  $E = (s_n = t_n)_{n \in \mathbb{N}}$  of equations (possibly taken over different orbit-finite strong nominal sets  $X$  of generators), a nominal monoid  $M$  *eventually satisfies*  $E$  if there exists an index  $n_0 \in \mathbb{N}$  such that  $M$  satisfies all the equations  $s_n = t_n$  with  $n \geq n_0$ . We denote by  $\mathcal{V}(E)$  the class of all orbit-finite nominal monoids that eventually satisfy  $E$ .

► **Remark 3.10.** Equations can be presented syntactically as expressions of the form

$$y_1 : S_1, \dots, y_n : S_n \vdash u = v, \quad (1)$$

where  $Y = \{y_1, \dots, y_n\}$  is a finite set of variables,  $S_1, \dots, S_n \subseteq \mathbb{A}$  are finite sets of atoms, and  $u, v$  are words in  $(\text{Perm}(\mathbb{A}) \times Y)^*$ . A nominal monoid  $M$  is said to *satisfy* (1) if for every *variable interpretation*, i.e. every map  $h: Y \rightarrow M$  with  $\text{supp}_M(h(y_i)) \subseteq S_i$  for  $i = 1, \dots, n$ , one has  $\bar{h}(u) = \bar{h}(v)$ . Here,  $\bar{h}: (\text{Perm}(\mathbb{A}) \times Y)^* \rightarrow M$  is the unique monoid morphism mapping  $(\pi, y_i)$  to  $\pi \cdot h(y_i)$ . Every equation can be transformed into an expressively equivalent syntactic equation, and vice versa [13, Lemma B.31].

► **Theorem 3.11 (Nominal Eilenberg-Schützenberger Theorem).** *A class  $\mathcal{V}$  of orbit-finite nominal monoids forms a weak pseudovariety iff  $\mathcal{V} = \mathcal{V}(E)$  for some sequence  $E$  of equations.*

**Proof sketch.** The proof proceeds along the lines of the one for ordinary monoids [7], although some subtle modifications are required. The “if” direction is a routine verification. For the “only if” direction, let  $\mathcal{V}$  be a weak pseudovariety. Using that there are only countably many orbit-finite monoids up to isomorphism, one can construct a sequence  $M_0, M_1, M_2, \dots$  of nominal monoids in  $\mathcal{V}$  such that each  $M \in \mathcal{V}$  is a quotient of all but finitely many  $M_n$ 's. Let  $X_0, X_1, X_2, \dots$  be the sequence of all (countably many) strong orbit-finite nominal sets up to isomorphism, and consider the equivariant congruence relation on  $X_n^*$  given by

$$s \equiv_n t \quad \text{iff} \quad M_n \text{ satisfies the equation } s = t.$$

One then shows that the congruence  $\equiv_n$  is generated by a finite subset  $W_n \subseteq \equiv_n$ , and that  $\mathcal{V} = \mathcal{V}(E)$  for every sequence  $E$  that lists all equations in the countable set  $\bigcup_n W_n$ . ◀

► **Example 3.12.** An orbit-finite nominal monoid  $M$  is called *aperiodic* [4, 17] if there exists a natural number  $n \geq 1$  such that  $x^{n+1} = x^n$  for all  $x \in M$ . The class of all orbit-finite aperiodic nominal monoids forms a pseudovariety. Taking the set  $Y = \{y\}$  of variables, it is not difficult to see that this pseudovariety is specified by the sequence of syntactic equations

$$y : S_n \vdash y^{n+1} = y^n \quad (n \in \mathbb{N}),$$

where  $S_n = \{a_0, a_1, \dots, a_{n-1}\}$  is the set of the first  $n$  atoms in the countably infinite set  $\mathbb{A} = \{a_0, a_1, a_2, \dots\}$  of all atoms, and we write  $y$  for  $(id, y) \in \text{Perm}(\mathbb{A}) \times Y$ .

## 4 Duality and the Nominal Eilenberg Correspondence

In this section, we establish our nominal version of Eilenberg's variety theorem. It is based on a dual interpretation of the concepts of a (local) pseudovariety of nominal monoids and of an equational theory, introduced in the previous section, under *nominal Stone duality*.

### 4.1 Nominal Stone Duality

A classical result from duality theory, known as *discrete Stone duality*, states that the category of sets is dually equivalent to the category of complete atomic boolean algebras, i.e. complete boolean algebras in which every non-zero element is above some atom. An analogous duality holds for the category  $\mathbf{Nom}_{\text{fs}}$  of nominal sets and finitely supported maps.

► **Definition 4.1.** A *nominal complete atomic boolean algebra (ncaba)* is a boolean algebra  $(B, \vee, \wedge, \neg, \perp, \top)$  in  $\mathbf{Nom}$  such that every finitely supported subset of  $B$  has a supremum, and for every element  $b \in B \setminus \{\perp\}$  there exists an atom (i.e. a minimal element)  $a \in B$  with  $a \leq b$ . Here, the partial order  $\leq$  is defined as usual by  $a \leq b$  iff  $a \wedge b = a$ . We denote by  $\mathbf{nCABA}_{\text{fs}}$  the category of ncabas and finitely supported morphisms (i.e. finitely supported maps preserving all the boolean operations and suprema of finitely supported subsets), and by  $\mathbf{nCABA}$  the (non-full) subcategory of ncabas and *equivariant* morphisms.

► **Theorem 4.2 (Nominal Stone Duality).** *The categories  $\mathbf{nCABA}_{\text{fs}}$  and  $\mathbf{Nom}_{\text{fs}}$  are dually equivalent. The duality restricts to one between the subcategories  $\mathbf{nCABA}$  and  $\mathbf{Nom}$ .*

The restricted duality is due to Petrişan [15, Prop. 5.3.11].

► **Remark 4.3.**

- (1) The equivalence functor  $\mathbf{Nom}_{\text{fs}} \xrightarrow{\cong} \mathbf{nCABA}_{\text{fs}}^{\text{op}}$  maps a nominal set  $X$  to the ncaba  $\mathcal{P}X$  of finitely supported subsets of  $X$  (equipped with the set-theoretic boolean operations), and a finitely supported map  $f: X \rightarrow Y$  to the morphism  $f^{-1}: \mathcal{P}Y \rightarrow \mathcal{P}X$  taking preimages. The inverse equivalence functor  $\mathbf{nCABA}_{\text{fs}}^{\text{op}} \xrightarrow{\cong} \mathbf{Nom}_{\text{fs}}$  maps an ncaba  $B$  to the equivariant subset  $\text{At}(B)$  of its atoms, with group action restricting the one of  $B$ .
- (2) The dual equivalence restricts to one between the full subcategories of orbit-finite nominal sets and *atom-finite* ncabas, i.e. ncabas whose set of atoms is orbit-finite. For atom-finite ncabas the property that every finitely supported subset has a supremum is equivalent to the weaker requirement that for every finite set  $S \subseteq \mathbb{A}$ , every  $S$ -orbit has a supremum. Indeed, given a finitely supported subset  $X \subseteq B$  (say with finite support  $S \subseteq \mathbb{A}$ ), put  $X' := \{a \in \text{At}(B) : a \leq x \text{ for some } x \in X\}$ . Since  $\leq$  is an equivariant relation,  $X' \subseteq \text{At}(B)$  is a subset with finite support  $S$ . Since  $\text{At}(B)$  is orbit-finite and thus has only finitely many  $S$ -orbits by Lemma 2.1, we can express  $X'$  as a finite union  $X' = X'_1 \cup \dots \cup X'_n$  of  $S$ -orbits. Using that every element of  $B$  is the join of the finitely supported set of all atoms below it, it follows that  $\bigvee X = \bigvee X' = \bigvee X'_1 \vee \dots \vee \bigvee X'_n$ , so  $\bigvee X$  is a finite join of joins of  $S$ -orbits.



## 4.2 Varieties of Data Languages

For the notion of a *language* over an alphabet  $\Sigma \in \mathbf{Nom}$  and the corresponding concept of algebraic recognition by nominal monoids, there are two natural choices: consider equivariant subsets  $L \subseteq \Sigma^*$  and their recognition by equivariant monoid morphisms [5, 17], or consider finitely supported subsets  $L \subseteq \Sigma^*$  and their recognition by finitely supported monoid morphisms [4]. For our duality-based approach to data languages, it turns out that we need to work with an intermediate concept: finitely supported languages recognizable by equivariant monoid morphisms (see the discussion in Remark 4.12 below). That is, we work with the following

► **Definition 4.4.** A *data language* over the alphabet  $\Sigma \in \mathbf{Nom}$  is a finitely supported map  $L: \Sigma^* \rightarrow 2$ . It is *recognized* by an equivariant monoid morphism  $e: \Sigma^* \rightarrow M$  if there exists a finitely supported map  $p: M \rightarrow 2$  with  $L = p \cdot e$ . In this case, we also say that  $M$  *recognizes*  $L$ . A data language is *recognizable* if it is recognized by some orbit-finite nominal monoid.

► **Remark 4.5.**

- (1) Identifying finitely supported maps into 2 with finitely supported subsets, Definition 4.4 can be restated: an equivariant monoid morphism  $e: \Sigma^* \rightarrow M$  *recognizes* a language  $L \subseteq \Sigma^*$  if there exists a finitely supported subset  $P \subseteq M^*$  with  $L = e^{-1}[P]$ .
- (2) If  $L$  is an equivariant recognizable language, then  $p$  in Definition 4.4 is also equivariant. Therefore, for equivariant languages we recover the notion of recognition of [5, 17].
- (3) If  $\Sigma$  is a finite set (viewed as an orbit-finite discrete nominal set), a data language is just an ordinary formal language over the alphabet  $\Sigma$ . Indeed, the free nominal monoid  $\Sigma^*$  is discrete, and thus every subset of  $\Sigma^*$  is finitely supported. Moreover, every orbit-finite nominal quotient monoid of  $\Sigma^*$  is discrete and finite. Hence, the above notion of language recognition coincides with the classical recognition by finite monoids. In particular, for finite  $\Sigma$ , a recognizable data language is the same as a regular language.

► **Example 4.6.** Examples of recognizable data languages over the alphabet  $\Sigma = \mathbb{A}$  include (1) every finite or cofinite subset  $L \subseteq \mathbb{A}^*$  (see Remark 4.8 below), (2)  $a\mathbb{A}^*$  for a fixed atom  $a \in \mathbb{A}$ , and (3)  $\bigcup_{a \in \mathbb{A}} \mathbb{A}^*aa\mathbb{A}^*$ . The languages (4)  $\{a_1 \dots a_n : a_i \neq a_j \text{ for } i \neq j\}$ , (5)  $\bigcup_{a \in \mathbb{A}} a\mathbb{A}^*a\mathbb{A}^*$ , and (6)  $\mathbb{A}^*a\mathbb{A}^*$  for a fixed atom  $a \in \mathbb{A}$  are not recognizable. The equivariant examples (3)–(5) are taken from [4, 5].

In previous work [1] we have given a categorical account of *local varieties of regular languages* [9], i.e. sets of regular languages over a fixed finite alphabet  $\Sigma$  closed under the set-theoretic boolean operations (finite union, finite intersection, complement) and derivatives. This concept can be generalized to data languages. The *derivatives* of a data language  $L \subseteq \Sigma^*$  with respect to a word  $w \in \Sigma^*$  are given by

$$w^{-1}L = \{v \in \Sigma^* : wv \in L\} \quad \text{and} \quad Lw^{-1} = \{v \in \Sigma^* : vw \in L\}.$$

Since  $\text{supp}(w^{-1}L), \text{supp}(Lw^{-1}) \subseteq \text{supp}(w) \cup \text{supp}(L)$ , the derivatives are again data languages.

► **Definition 4.7 (Local Variety of Data Languages).** Let  $\Sigma \in \mathbf{Nom}$ . A *local variety of data languages over  $\Sigma$*  is an equivariant set  $\mathcal{V}_\Sigma \subseteq \mathcal{P}\Sigma^*$  of recognizable data languages closed under the set-theoretic boolean operations, unions of  $S$ -orbits for every finite set  $S \subseteq \mathbb{A}$  of atoms (that is, for every  $L \in \mathcal{V}_\Sigma$  the language  $\bigcup_{\pi \in \text{Perm}_S(\mathbb{A})} \pi \cdot L$  lies in  $\mathcal{V}_\Sigma$ ), and derivatives.

► **Remark 4.8.**

- (1) If  $\Sigma$  is a finite set (viewed as a discrete nominal set), then by Remark 4.5 a local variety  $\mathcal{V}_\Sigma$  consists of regular languages, and the closure under unions of  $S$ -orbits is trivial: since  $\mathcal{P}\Sigma^*$  is discrete, every  $S$ -orbit has a single element. Thus, in this case, a local variety of data languages is precisely a local variety of regular languages.

## 130:10 Varieties of Data Languages

- (2) However, in general the closedness under unions of  $S$ -orbits cannot be dropped, as it is neither trivial nor implied by the other conditions. To see this, consider the alphabet  $\Sigma = \mathbb{A}$  and the equivariant set  $\mathcal{V}_{\mathbb{A}} \subseteq \mathcal{P}\mathbb{A}^*$  of all finite or cofinite subsets of  $\mathbb{A}^*$ . Note that every finite language  $L \subseteq \mathbb{A}^*$  is recognizable: let  $n \geq 1$  be an upper bound on the length of words in  $L$ , and take the orbit-finite monoid  $M = \mathbb{A}^{\leq n} \cup \{0\}$  consisting of all words over  $\mathbb{A}$  of length at most  $n$ , and a zero element  $0$ . The multiplication  $\bullet$  is defined as follows: given  $v, w \in \mathbb{A}^{\leq n}$ , if the word  $vw$  has length at most  $n$ , put  $v \bullet w = vw$ . Otherwise, put  $v \bullet w = 0$ . Then the equivariant monoid morphism  $e: \mathbb{A}^* \rightarrow M$  extending  $a \mapsto a$  recognizes  $L$  since  $L = e^{-1}[L]$ . It follows that also  $\mathbb{A}^* \setminus L = e^{-1}[M \setminus L]$ . This shows that every language in  $\mathcal{V}_{\mathbb{A}}$  is recognizable. Moreover, clearly  $\mathcal{V}_{\mathbb{A}}$  is closed under the set-theoretic boolean operations and derivatives. However, the languages  $\{a\}$ ,  $a \in \mathbb{A}$ , form an orbit in  $\mathcal{V}_{\mathbb{A}}$ , but their union  $\mathbb{A} = \bigcup_{a \in \mathbb{A}} \{a\}$  is not in  $\mathcal{V}_{\mathbb{A}}$ . Thus  $\mathcal{V}_{\mathbb{A}}$  is not a local variety of data languages in the sense of Definition 4.7.

A local variety  $\mathcal{V}_{\Sigma}$  is generally not a subobject of  $\mathcal{P}\Sigma^*$  in **nCABA**, because it is not required to be closed under unions of arbitrary finitely supported subsets and also not necessarily atomic as a boolean algebra. However, if the atomic languages in  $\mathcal{V}_{\Sigma}$  form an orbit-finite subset and every language in  $\mathcal{V}_{\Sigma}$  contains some atomic language, then  $\mathcal{V}_{\Sigma}$  is an atom-finite subobject of  $\mathcal{P}\Sigma^*$ , see Remark 4.32. In this case, we call  $\mathcal{V}_{\Sigma}$  an *atom-finite* local variety.

► **Theorem 4.9** (Finite Local Variety Theorem). *The lattice of atom-finite local varieties of data languages over  $\Sigma$  is dually isomorphic to the lattice of  $\Sigma$ -generated orbit-finite monoids.*

The isomorphism maps a  $\Sigma$ -generated orbit-finite monoid  $e: \Sigma^* \rightarrow M$  to the atom-finite local variety of all data languages recognized by  $e$ .

**Proof.** By the duality of **Nom** and **nCABA**, orbit-finite equivariant quotients  $e: \Sigma^* \rightarrow M$  of  $\Sigma^*$  in **Nom** correspond bijectively to atom-finite subobjects  $\mathcal{V}_{\Sigma} \hookrightarrow \mathcal{P}\Sigma^*$  in **nCABA**, i.e. atom-finite equivariant sets of languages closed under the set-theoretic boolean operations and unions of  $S$ -orbits for every finite  $S \subseteq \mathbb{A}$ . By Proposition 3.1 and the dual equivalence of **Nom<sub>fs</sub>** and **nCABA<sub>fs</sub>**, the map  $e$  represents a nominal quotient monoid of  $\Sigma^*$  if and only if  $\mathcal{V}_{\Sigma}$  is closed under derivatives, i.e. a local variety. The closure under left derivatives is illustrated by the two dual commutative squares below, where the left-hand one lives in **Nom<sub>fs</sub>** and the right-hand one in **nCABA<sub>fs</sub>**.

$$\begin{array}{ccc}
 \Sigma^* & \xrightarrow{w \cdot -} & \Sigma^* \\
 e \downarrow & & \downarrow e \\
 M & \xrightarrow{- \cdot -} & M
 \end{array}
 \qquad
 \begin{array}{ccc}
 \mathcal{P}\Sigma^* & \xleftarrow{w^{-1}(-)} & \mathcal{P}\Sigma^* \\
 \subseteq \uparrow & & \uparrow \subseteq \\
 \mathcal{V}_{\Sigma} & \xleftarrow{- \cdot -} & \mathcal{V}_{\Sigma}
 \end{array}$$

The elements of  $\mathcal{V}_{\Sigma}$  are precisely the languages recognized by  $e$ . Indeed, the former correspond to the morphisms  $\mathbf{1} \rightarrow \mathcal{V}_{\Sigma}$  in **nCABA<sub>fs</sub>**, where  $\mathbf{1}$  is the free boolean algebra on one generator, the latter to the finitely supported maps  $M \rightarrow 2$  in **Nom<sub>fs</sub>**, and  $\mathbf{1}$  and  $2$  are dual objects. ◀

Recall that an *ideal* in a poset is a downwards closed and upwards directed subset. For the lattice of local varieties of data languages over  $\Sigma$  (ordered by inclusion), we obtain

► **Lemma 4.10.** *The lattice of local varieties of data languages over  $\Sigma$  is isomorphic to the lattice of ideals in the poset of atom-finite local varieties over  $\Sigma$ .*

The isomorphism maps a local variety  $\mathcal{V}_\Sigma$  to the ideal of all atom-finite local subvarieties of  $\mathcal{V}_\Sigma$ . Its inverse maps an ideal  $\{\mathcal{V}_{\Sigma,i} : i \in I\}$  in the poset of atom-finite local varieties to the local variety  $\bigcup_{i \in I} \mathcal{V}_{\Sigma,i}$ . In order-theoretic terms, the above lemma states that local varieties of data languages form the ideal completion of the poset of atom-finite local varieties. Using Theorem 4.9, Lemma 4.10, and the fact that ideals are dual to filters, we obtain

► **Theorem 4.11** (Local Variety Theorem). *For each  $\Sigma \in \mathbf{Nom}$ , the lattice of local varieties of data languages over  $\Sigma$  is dually isomorphic to the lattice of local pseudovarieties of  $\Sigma$ -generated nominal monoids.*

► **Remark 4.12.**

- (1) Since data languages are morphisms  $L: \Sigma^* \rightarrow 2$  in  $\mathbf{Nom}_{\text{fs}}$ , the reader may wonder why we do not entirely work in this category and use monoids with finitely supported multiplication and finitely supported monoid morphisms (rather than the equivariant ones) for the recognition of languages. The reason lies on the dual side: in the proof of Theorem 4.9, we used that equivariant injective maps  $\mathcal{V}_\Sigma \rightarrow \mathcal{P}\Sigma^*$  can be uniquely identified with equivariant subsets of  $\mathcal{P}\Sigma^*$ . In contrast, finitely supported injective maps  $\mathcal{V}_\Sigma \rightarrow \mathcal{P}\Sigma^*$  do not correspond to the finitely supported (or any other kind of) subsets of  $\mathcal{P}\Sigma^*$ .
- (2) Similarly, we cannot restrict ourselves to the category  $\mathbf{Nom}$  and only consider equivariant languages  $L \subseteq \Sigma^*$  rather than finitely supported ones. Indeed, the Finite Local Variety Theorem then fails: the map sending a  $\Sigma$ -generated orbit-finite monoid  $e: \Sigma^* \rightarrow M$  to the set of equivariant languages it recognizes is no longer bijective. To see this, consider the nominal monoids  $M = \mathbb{A} \cup \{1\}$  with  $a \bullet b = a$  for  $a, b \in \mathbb{A}$ , and  $N = \{0, 1\}$  with  $0 \bullet 0 = 0 \bullet 1 = 1 \bullet 0 = 0$ . Then the two surjective morphisms  $e: \mathbb{A}^* \rightarrow M$ , extending  $a \mapsto a$ , and  $f: \mathbb{A}^* \rightarrow N$ , extending  $a \mapsto 0$ , recognize the same equivariant languages, namely  $\mathbb{A}^*$ ,  $\mathbb{A}^* \setminus \{\varepsilon\}$ ,  $\{\varepsilon\}$  and  $\emptyset$ .

In the following, we consider data languages whose alphabet  $\Sigma$  is an orbit-finite strong nominal set (see Remark 3.62). By dualizing the concept of an equational theory, we obtain

► **Definition 4.13** (Variety of data languages). *A variety of data languages is a family*

$$\mathcal{V} = (\mathcal{V}_\Sigma \subseteq \mathcal{P}\Sigma^*)_{\Sigma \in \mathbf{Nom}_{\text{of},s}}$$

of local varieties of data languages with the following two properties:

- (1) *Closedness under preimages.* For each equivariant monoid morphism  $h: \Delta^* \rightarrow \Sigma^*$  with  $\Sigma, \Delta \in \mathbf{Nom}_{\text{of},s}$  and each  $L \in \mathcal{V}_\Sigma$ , one has  $h^{-1}[L] \in \mathcal{V}_\Delta$ .
- (2) *Completeness.* For each atom-finite local subvariety  $\mathcal{V}'_\Sigma \rightarrow \mathcal{V}_\Sigma$ , there exists an equivariant monoid morphism  $h: \Sigma^* \rightarrow \Delta^*$  and an atom-finite local subvariety  $\mathcal{V}'_\Delta \rightarrow \mathcal{V}_\Delta$  such that
  - a. the map  $L \mapsto h^{-1}[L]$  defines a bijection between  $\mathcal{V}'_\Delta$  and  $\mathcal{V}'_\Sigma$ , and
  - b. every atomic language  $L \in \mathcal{V}'_\Delta$  contains a word  $w \in \Delta^*$  with  $\text{supp}_{\mathcal{P}\Delta^*}(L) = \text{supp}_{\Delta^*}(w)$ .

► **Remark 4.14.** Except for the completeness condition, the above concept is analogous to Eilenberg's original notion of a variety of regular languages (i.e. a family of local varieties of regular languages closed under preimages of monoid morphisms). In fact, if  $\Sigma$  is a finite alphabet, and thus  $\mathcal{V}_\Sigma$  is just a local variety of regular languages, completeness is trivial: given any finite local subvariety  $\mathcal{V}'_\Sigma$  of  $\mathcal{V}_\Sigma$ , choose  $\Delta = \Sigma$ ,  $\mathcal{V}'_\Delta = \mathcal{V}'_\Sigma$ , and  $h = \text{id}: \Sigma^* \rightarrow \Delta^*$ . Then 2a is clear, and 2b holds because each  $L \in \mathcal{P}\Sigma^*$  and each  $w \in \Sigma^*$  has empty support.

In general, however, the completeness property cannot be dropped. This follows from Remark 3.63 and the observation that the completeness of a variety dualizes to the completeness of the corresponding equational theory (see the proof of Theorem 4.15).

We are ready to state the main result of our paper:

► **Theorem 4.15** (Nominal Eilenberg Theorem). *Varieties of data languages and pseudovarieties of nominal monoids form isomorphic complete lattices.*

The isomorphism maps a variety  $\mathcal{V}$  of data languages to the pseudovariety  $\mathcal{V}$  of all orbit-finite nominal monoids that recognize only languages from  $\mathcal{V}$ . Its inverse maps a pseudovariety  $\mathcal{V}$  to the variety  $\mathcal{V}$  of all data languages recognized by some monoid in  $\mathcal{V}$ .

**Proof sketch.** We observe that the concept of a variety is dual to that of an equational theory. Indeed, by the Local Variety Theorem 4.11, a family  $\mathcal{V} = (\mathcal{V}_\Sigma \subseteq \mathcal{P}\Sigma^*)_{\Sigma \in \mathbf{Nom}_{\text{of},s}}$  of local varieties of data languages bijectively corresponds to a family  $\mathcal{T} = (\mathcal{T}_\Sigma \subseteq \Sigma^* \downarrow \mathbf{nMon}_{\text{of}})_{\Sigma \in \mathbf{Nom}_{\text{of},s}}$  of local pseudovarieties of nominal monoids. One then shows that (1)  $\mathcal{T}$  is substitution-invariant if and only if  $\mathcal{V}$  is closed under preimages, and (2)  $\mathcal{T}$  is complete if and only if  $\mathcal{V}$  is complete. In particular,  $\mathcal{T}$  is a theory if and only if  $\mathcal{V}$  is a variety of data languages. Since theories correspond to pseudovarieties by Theorem 3.8, this proves the theorem. ◀

## 5 Adding Expressivity: Regular Data Languages

As recognizing structures for data languages, nominal monoids are of limited expressivity; in particular, they are strictly weaker than deterministic automata in the category of nominal sets [5, 4]. Therefore, we now show how to extend our results for monoid-recognizable data languages and establish a local variety theorem for languages accepted by nominal automata.

► **Definition 5.1.** Fix an input alphabet  $\Sigma \in \mathbf{Nom}$ . A *nominal  $\Sigma$ -automaton*  $A = (Q, \delta, q_0)$  consists of a nominal set  $Q$  of states, an equivariant transition map  $\delta: Q \times \Sigma \rightarrow Q$ , and an initial state  $q_0 \in Q$  with empty support. It is called *orbit-finite* if  $Q$  is orbit-finite. A *morphism* between nominal automata  $A = (Q, \delta, q_0)$  and  $A' = (Q', \delta', q'_0)$  is an equivariant map  $h: Q \rightarrow Q'$  such that  $\delta(h(q), a) = h(\delta(q, a))$  for all  $q \in Q$  and  $a \in \Sigma$ , and  $h(q_0) = q'_0$ .

The *initial nominal  $\Sigma$ -automaton* is given by  $I = (\Sigma^*, \delta, \varepsilon)$  with transition map  $\delta(w, a) = wa$  for  $w \in \Sigma^*$  and  $a \in \Sigma$ . It is characterized by the universal property that for every nominal  $\Sigma$ -automaton  $A = (Q, \delta, q_0)$ , there exists a unique morphism  $e_A: I \rightarrow A$ , sending a word  $w \in \Sigma^*$  to the state reached from  $q_0$  after reading  $w$ . The automaton  $A$  is called *reachable* if  $e_A$  is surjective. A data language  $L \subseteq \Sigma^*$  is *accepted* by  $A$  if there exists a finitely supported subset  $F \subseteq Q$  with  $L = e_A^{-1}[F]$ . This corresponds to the usual notion of acceptance of an automaton with final states  $F$ : the language  $L$  consists of all words  $w \in \Sigma^*$  such that  $A$  reaches a state in  $F$  after reading  $w$ . A data language  $L \subseteq \Sigma^*$  is called *regular* if there exists an orbit-finite nominal automaton accepting it. In analogy to Proposition 3.1, we get

► **Proposition 5.2** (Unary presentation for nominal automata). *For every surjective equivariant map  $e: \Sigma^* \twoheadrightarrow Q$ , the following statements are equivalent:*

- (1) *there exists a nominal automaton  $A = (Q, \delta, q_0)$  with states  $Q$  such that  $e = e_A$ ;*
- (2) *the maps  $\Sigma^* \xrightarrow{- \cdot w} \Sigma^*$  ( $w \in \Sigma^*$ ) lift along  $e$ , i.e. there exist (necessarily unique) maps  $r_w: Q \rightarrow Q$  such that  $e \cdot (- \cdot w) = r_w \cdot e$  for all  $w \in \Sigma^*$ .*

Define a *local pseudovariety of nominal  $\Sigma$ -automata* to be a class  $\mathcal{V}_\Sigma$  of orbit-finite reachable nominal  $\Sigma$ -automata such that (1)  $\mathcal{V}_\Sigma$  is closed under quotients (represented by surjective automata morphisms), and (2) for every pair  $A, B \in \mathcal{V}_\Sigma$ , the reachable part of the product  $A \times B$  lies in  $\mathcal{V}_\Sigma$ . Here, the product of two nominal automata  $A = (Q, \delta, q_0)$  and  $B = (Q', \delta', q'_0)$  is given by  $A \times B = (Q \times Q', \delta, (q_0, q'_0))$  with  $\delta((q, q'), a) = (\delta(q, a), \delta'(q', a))$  for  $(q, q') \in Q \times Q'$  and  $a \in \Sigma$ , and the reachable part  $R$  of  $A \times B$  is the coimage  $e: \Sigma^* \twoheadrightarrow R$  of the unique

morphism  $e_{A \times B}: \Sigma^* \rightarrow A \times B$ . Note that a local pseudovariety corresponds precisely to a filter in the poset  $\Sigma^* \downarrow \Sigma\text{-nAut}_{\text{of}}$  of orbit-finite reachable nominal  $\Sigma$ -automata. The dual version of this concept is the one of a *local variety of regular data languages over  $\Sigma$* : an equivariant set  $\mathcal{V}_\Sigma \subseteq \mathcal{P}\Sigma^*$  of such languages closed under the set-theoretic boolean operations, unions of  $S$ -orbits for every finite set  $S \subseteq \mathbb{A}$  of atoms, and *right* derivatives. The following theorem, and its proof, are completely analogous to Theorem 4.11:

► **Theorem 5.3 (Local Variety Theorem for Regular Data Languages).** *For each  $\Sigma \in \mathbf{Nom}$ , the lattice of local varieties of regular data languages over  $\Sigma$  is dually isomorphic to the lattice of local pseudovarieties of nominal  $\Sigma$ -automata.*

## 6 Conclusions and Future Work

We have demonstrated that two cornerstones of the algebraic theory of regular languages, Eilenberg’s variety theorem and Eilenberg and Schützenberger’s axiomatic characterization of pseudovarieties, can be generalized to data languages recognizable by orbit-finite monoids. Our results are the first of this type for data languages, and thus the present work makes a contribution towards developing a fully fledged algebraic theory of such languages. In a broader sense, the approach taken in this paper can be seen as a further illustration of the power of duality in formal language theory: we believe that without the guidance given by nominal Stone duality, it would have been significantly harder to even come up with the suitable notion of a variety of data languages that makes the nominal Eilenberg correspondence work. The duality-based approach thus adds much conceptual clarity and simplicity. There remain several research questions and interesting directions for future work.

As indicated in section 5, the techniques used in our paper can be adapted without much effort to languages recognized by nominal algebraic structures other than monoids, including deterministic nominal automata. As a first step, we aim to extend the local variety theorem for regular data languages (Theorem 5.3) to a full Eilenberg correspondence. It remains an important goal to further extend our results to more powerful classes of data languages.

Our proof of the (local) Eilenberg correspondence rests on the observation that a local variety of data languages can be expressed as the directed union of its atom-finite subvarieties. From a category theoretic perspective, this suggests that local varieties are formed within the *Ind-completion* (i.e. the free completion under directed colimits) of the category of atom-finite nominal complete atomic boolean algebras. We conjecture that this completion can be described as a category of nominal boolean algebras with joins of  $S$ -orbits for each finite set  $S$  of atoms. On the dual side, we expect that the *Pro-completion* (i.e. the free completion under codirected limits) of the category of orbit-finite nominal sets consists of some form of nominal Stone spaces. The approach of working with free completions should lead to a topological version of nominal Stone duality similar to the one established by Gabbay, Litak, and Petrişan [8]. More importantly, it might pave the way to the introduction of pro-(orbit-)finite methods for the theory of data languages.

---

## References

- 1 Jiří Adámek, Stefan Milius, Robert S.R. Myers, and Henning Urbat. Generalized Eilenberg Theorem I: Local Varieties of Languages. In Anca Muscholl, editor, *Proc. Foundations of Software Science and Computation Structures (FoSSaCS)*, volume 8412 of *Lecture Notes Comput. Sci.*, pages 366–380. Springer, 2014.

- 2 Jiří Adámek, Stefan Milius, Robert S.R. Myers, and Henning Urbat. Varieties of Languages in a Category. In Catuscia Palamidessi, editor, *Proc. 30th Annual Symposium on Logic in Computer Science (LICS'15)*, pages 414–425. IEEE Computer Society, 2015.
- 3 Jiří Adámek, Stefan Milius, Robert S.R. Myers, and Henning Urbat. Generalized Eilenberg Theorem: Varieties of Languages in a Category. *ACM Trans. Comput. Log.*, 20(1):3:1–3:47, 2019.
- 4 Mikołaj Bojańczyk. Nominal Monoids. *Theory of Computing Systems*, 53(2):194–222, 2013.
- 5 Mikołaj Bojańczyk, Bartek Klin, and Sławomir Lasota. Automata theory in nominal sets. *Log. Methods Comput. Sci.*, 10(3:4):44 pp., 2014.
- 6 Samuel Eilenberg. *Automata, Languages, and Machines Vol. B*. Academic Press, 1976.
- 7 Samuel Eilenberg and Marcel-Paul Schützenberger. On pseudovarieties. *Advances Math.*, 10:413–418, 1976.
- 8 Murdoch J. Gabbay, Tadeusz Litak, and Daniela Petrişan. Stone Duality for Nominal Boolean Algebras with New. In A. Corradini, B. Klin, and C. Cîrstea, editors, *Proc. CALCO'11*, volume 6859 of *LNCS*, pages 192–207. Springer, 2011.
- 9 Mai Gehrke, Serge Grigorieff, and Jean-Éric Pin. Duality and equational theory of regular languages. In *Proc. ICALP'08, Part II*, volume 5126 of *LNCS*, pages 246–257. Springer, 2008.
- 10 Michael Kaminski and Nissim Francez. Finite-memory automata. *Theoret. Comput. Sci.*, 134(2):329–363, 1994.
- 11 Alexander Kurz and Daniela Petrişan. On universal algebra over nominal sets. *Mathematical Structures in Computer Science*, 20(2):285–318, 2010.
- 12 Robert McNaughton and Seymour A. Papert. *Counter-Free Automata (M.I.T. Research Monograph No. 65)*. The MIT Press, 1971.
- 13 Stefan Milius and Henning Urbat. Equational Axiomatization of Algebras with Structure. Full version; available online at [arXiv:1812.02016](https://arxiv.org/abs/1812.02016), 2018.
- 14 Stefan Milius and Henning Urbat. Equational axiomatization of algebras with structure. In *Proc. 22nd International Conference on Foundations of Software Science and Computation Structures*, 2019. To appear.
- 15 Daniela Petrişan. *Investigations into Algebra and Topology over Nominal Sets*. PhD thesis, University of Leicester, 2012.
- 16 Andrew M. Pitts. *Nominal Sets: Names and Symmetry in Computer Science*. Cambridge University Press, 2013.
- 17 Gabriele Puppis, Thomas Colcombet, and Clemens Ley. Logics with rigidly guarded data tests. *Log. Methods Comput. Sci.*, 11(3:10):56 pp., 2015.
- 18 Julian Salamanca. Unveiling Eilenberg-type Correspondences: Birkhoff's Theorem for (finite) Algebras + Duality, February 2017. [arXiv:1702.02822](https://arxiv.org/abs/1702.02822).
- 19 Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Inform. and Control*, 8:190–194, 1965.
- 20 Thomas Schwentick. Automata for XML – A survey. *J. Comput. System Sci.*, 73(3):289–315, 2007.
- 21 Luc Segoufin. Automata and Logics for Words and Trees over an Infinite Alphabet. In *Computer Science Logic, 20th International Workshop, CSL 2006, 15th Annual Conference of the EACSL, Szeged, Hungary, September 25-29, 2006, Proceedings*, pages 41–57, 2006.
- 22 Henning Urbat, Jiří Adámek, Liang-Ting Chen, and Stefan Milius. Eilenberg Theorems for Free. In Kim G. Larsen, Hans L. Bodlaender, and Jean-François Raskin, editors, *Proc. 42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017)*, volume 83 of *LIPICs*. Schloss Dagstuhl, 2017. EATCS Best Paper Award.
- 23 Henning Urbat and Stefan Milius. Varieties of Data Languages. *CoRR*, abs/1903.08053, 2019. [arXiv:1903.08053](https://arxiv.org/abs/1903.08053).